



VISIGRAPP 2024

19th International Joint Conference on Computer Vision,
Imaging and Computer Graphics Theory and Applications

Final Program and Book of Abstracts

27 - 29 February, 2024

<https://visigrapp.scitevents.org>

SPONSORED BY



PAPERS AVAILABLE AT



VISIGRAPP 2024

Final Program and Book of Abstracts

19th International Joint Conference on Computer Vision, Imaging
and Computer Graphics Theory and Applications

Rome - Italy
February 27 - 29, 2024

Sponsored by

INSTICC - Institute for Systems and Technologies of Information, Control and Communication

ACM In Cooperation

SIGGRAPH - ACM Special Interest Group on Graphics and Interactive Techniques

Endorsed by

IAPR - International Association for Pattern Recognition

In Cooperation with

IS&T - Society for Imaging Science and Technology

EG - European Association for Computer Graphics

VRVis - Center for Virtual Reality and Visualization Forschungs-GmbH

GPCG - Grupo Português de Computação Gráfica

AFIG - French Association for Computer Graphics

Table of Contents

| | |
|---------------------------------------|-----|
| Foreword | 5 |
| Social Event and Banquet | 7 |
| Important Information | 9 |
| General Information | 10 |
| Rooms Layout | 11 |
| Program Layout | 12 |
| Tuesday Sessions: February 27 | 31 |
| Wednesday Sessions: February 28 | 63 |
| Thursday Sessions: February 29 | 97 |
| Author Index | 131 |

Foreword

This book contains the final program and paper abstracts of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024), organized and sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC). This event is endorsed by the International Association for Pattern Recognition (IAPR), and in cooperation with the ACM Special Interest Group on Graphics and Interactive Techniques, the Society for Imaging Science and Technology (IS&T), the European Association for Computer Graphics (EUROGRAPHICS), the VRVis Center for Virtual Reality and Visualization Forschungs-GmbH, the EUROGRAPHICS Portuguese Chapter, and the French Association for Computer Graphics (AFIG).

These proceedings present new and innovative solutions while highlighting technical problems in each field, both challenging and worthy of being disseminated to research audiences.

VISIGRAPP 2024 promotes discussions about the conference's various research topics between researchers, developers, manufacturers and end-users, and establishes guidelines in the development of advanced solutions.

This event was held in Rome, Italy, from the 27th to the 29th of February 2024.

We received a high number of paper submissions for this edition of VISIGRAPP, 431 in total, with contributions from 52 countries, demonstrating the success and global scope of VISIGRAPP. To evaluate each submission, we used a hierarchical process of double-blind evaluations where each paper was reviewed by two to six experts from the International Program Committee (IPC).

The IPC selected for oral presentation and publication 16 full papers from GRAPP, 5 from HUCAPP, 5 from IVAPP, and 64 from VISAPP, which led to a full-paper acceptance ratio of 21% and a high-quality program. Apart from the above full papers, the conference program also features 157 short papers and 85 poster presentations. These conference proceedings, which are submitted for indexing to Web of Science/Conference Proceedings Citation Index, SCOPUS, DBLP, Semantic Scholar, Google Scholar, and EI, will help the Computer Vision, Imaging, Visualization, Computer Graphics and Human-Computer Interaction communities in developing their future research.

Moreover, we are proud to announce that the program also includes four plenary keynote lectures, given by internationally distinguished researchers, namely Fahad Khan (Linköping University, Sweden), Mel Slater (University of Barcelona, Spain), Gerhard Rigoll (Technical University of Munich, Germany), and Alvitta Ottley (Washington University, United States).

Authors of selected papers will be invited to submit extended versions for inclusion in a forthcoming book of VISIGRAPP Selected Papers to be published by Springer in 2024 as part of the CCIS series. Some papers will also be selected for publication of extended and revised versions in a special issue of the Springer Nature Computer Science journal and a special issue of the Image and Vision Computing journal. All papers presented at this conference will be available at the SCITEPRESS Digital Library.

Three awards are delivered during the closing session: Best Conference Paper, Best Student Paper and Best Poster for each of the four conferences. An award has also been dedicated to the best industrial paper of VISAPP, delivered during the closing session.

We would like to express our thanks, first of all, to the authors, whose work and dedication made it possible to put together a program that we believe to be exciting and of high technical quality. Next, we would like to thank the Associate Chairs, all the members of the program committee and auxiliary reviewers, who helped us with their expertise and time. We would also like to thank the invited speakers for their invaluable contribution and for sharing their vision.

Finally, we gratefully acknowledge the professional support of the INSTICC team for all organizational processes.

We wish you all an exciting conference. We hope to meet you again for the next edition of VISIGRAPP, details of which are available at: <http://www.visigrapp.scitevents.org>

Thomas Bashford-Rogers, University of Warwick, United Kingdom
Daniel Meneveaux, Computer Graphics, University of Poitiers, France
Mehdi Ammi, University of Paris 8, France
Mounia Ziat, Bentley University, United States
Stefan Jänicke, Department of Mathematics and Computer Science, University of Southern Denmark, Denmark
Helen Purchase, Monash University, Australia
Petia Radeva, Mathematics and Computer Science, Universitat de Barcelona, Spain
Antonino Furnari, Mathematics and Computer Science, University of Catania, Italy
Kadi Bouatouch, IRISA, University of Rennes 1, France
A. Augusto Sousa, DEI, FEUP/INESC TEC, Portugal

Social Event and Banquet

Unforgettable Evening at Castello di Decima

Wednesday 28, 19:00 – 23:30

Castello di Decima is located at the tenth mile of the ancient Roman consular road leading to Naples. The lava stone slabs of the consular road are still hidden under the tarmac road of the old Via Pontina, now via Clarice Tartufari.

Important archaeological findings of the Latin people were discovered near the junction for Pratica di Mare: a "grave with chariot" dating 750-725 B.C., a "female tomb" from mid seventh century B.C. and other findings are now preserved at the High Middle Age Museum in Rome.

In Roman times, there was a place for changing horses at the tenth mile. The tower of the fortress and the church rest on tufa rocks and have Roman underground walls as their foundations – these were the remains of a Roman pool. On an adjacent tufa plateau, there are Latin walls, maybe remnants of ancient Politorium, destroyed by Anco Marzio.

Although the minor consular roads fell into disuse during the lower Empire, the village was still always inhabited. Many famous people owned the Castle over several centuries. In 1768 Cardinal Luigi Torrigiani, Secretary of State of Pope Clement XIII, built the Palace in Decima on previous constructions, and restructured the parish church dedicated to St. Antonio Abate, with various houses and warehouses in Decima Bassa.



In 1938 the estate was bought by Count Romolo Vaselli, one of the most important and modern Roman entrepreneurs of the early twentieth century, and he immediately committed to developing the land. Important canals and a network of internal roads were constructed and many houses with water and electric light were built.

At the same time, he set up an innovative tobacco growing activity and planted a modern peach orchard. Many trees and a pine forest were planted. Great improvement was initiated, granting privilege to the agricultural character of the place. His heirs continued the work with important restoration works and on-going improvements in the park.

The church, headed to St. Andrew the Apostle, and part of the parish of Tor de 'Cenci, contains an interesting coffered ceiling with the coat of arms of the Torrigiani family.



The eighteenth-century Castello di Decima still maintains all the charm of its ancient past. The Castle is located within the nature reserve of Decima Malafede. With its 6,145 hectares, the estate is the largest among

the parks protected by Roma Natura. Nestled in the green hills and in a beautiful pine forest, the castle is an island away from the world, a serene and exclusive venue for our evening.



Today the ancient castle of Decima represents a perfect combination of history, beauty and nature. The wonderful halls of the castle, with ancient painted ceilings and antique furnishings are the perfect location for our dinner. The music combined with fantastic food will provide our guests a memorable and unforgettable evening.

Important Information

Internet Access

Please check at the welcome desk the information to connect to the wireless network.

Event App

Download the Event App from the Play Store and App Store now, to have mobile access to the technical program and also to get notifications and reminders concerning your favorite sessions.

Create Your Own Schedule *

The option "My Program" gives you the possibility of creating a selection of the sessions that you plan to attend. This service also allows you to print-to-pdf all papers featured in your selection thus creating a pdf file per conference day.

Online Access to the Proceedings *

In the option "Proceedings and Final Program" you cannot only download the proceedings but also access the digital version of the book of abstracts with the final program.

Digital Access to the Receipt *

By clicking on the option "Delegate Home" and then "Registration Documents" it will enable you to access the final receipt which confirms the registration payment.

Photos Availability

The photos taken at the venue will be shared with you shortly after the event is finished. There will be an option entitled "Photo Gallery" in PRIMORIS. There, besides having access to the photos, you can also create your own personal albums by selecting "My Albums "Create New Album" and also be able to tag yourself in those photos, using the option "Tag Me".

Keynotes Videos

The keynote lectures will also be available on video on the website after the event, as long as the appropriate authorization from the keynote is received, so you will be able to see them again or watch them should you have missed one.

Survey

Every year we conduct a survey to access the participants' satisfaction with the conference and gather the suggestions. You will receive an e-mail after the event with the detailed information. Your contribution will be carefully analysed and a serious effort to react appropriately will be made.

* Please login to PRIMORIS (www.insticc.org/Primoris), select the role "Delegate" and the correct event.

If you have any doubt, we will be happy to help you at the Welcome Desk.

General Information

Welcome Desk/On-site Registration

Monday, February 26 – Open from 16:00 to 18:00

Tuesday, February 27 – Open from 08:15 to 18:00

Wednesday, February 28 – Open from 08:30 to 18:00

Thursday, February 29 – Open from 08:30 to 18:15

Opening Session

Tuesday, February 27, at 09:00 in the Plenary room.

Welcome Drink

Tuesday, February 27, at 18:00.

Closing Session & Awards Ceremony

Thursday, February 29, at 18:15 in the Plenary room.

Farewell Drink

Thursday, February 29, at 18:45.

Meals

Coffee-breaks will be served in the Foyer to all registered participants.

Lunches will be served in the Restaurant to all registered participants. Please check the hours in the Program Layout.

Communications

Wireless access will be provided free of charge to all registered participants.

Secretariat Contacts

VISIGRAPP Secretariat

Address: Avenida de S. Francisco Xavier, Lote 7 Cv. C

2900-616 Setúbal, Portugal

Tel.: +351 265 520 185

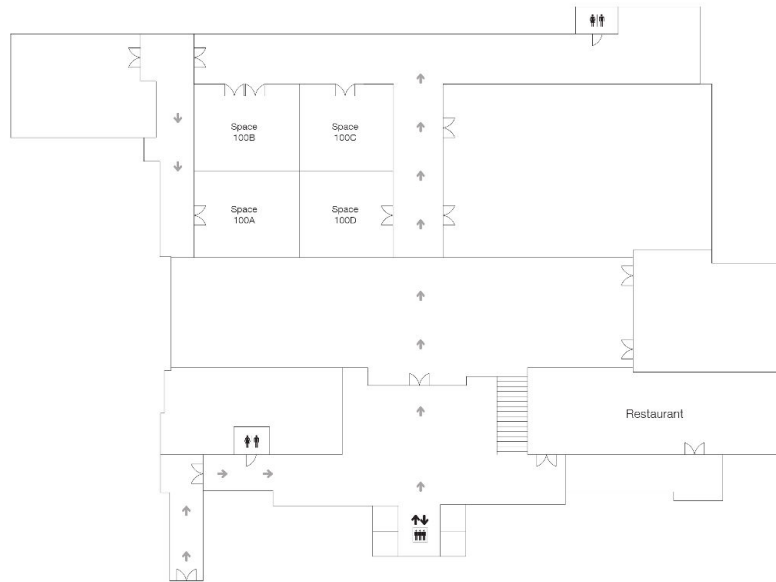
Fax: +351 265 520 186

e-mail: visigrapp.secretariat@insticc.org

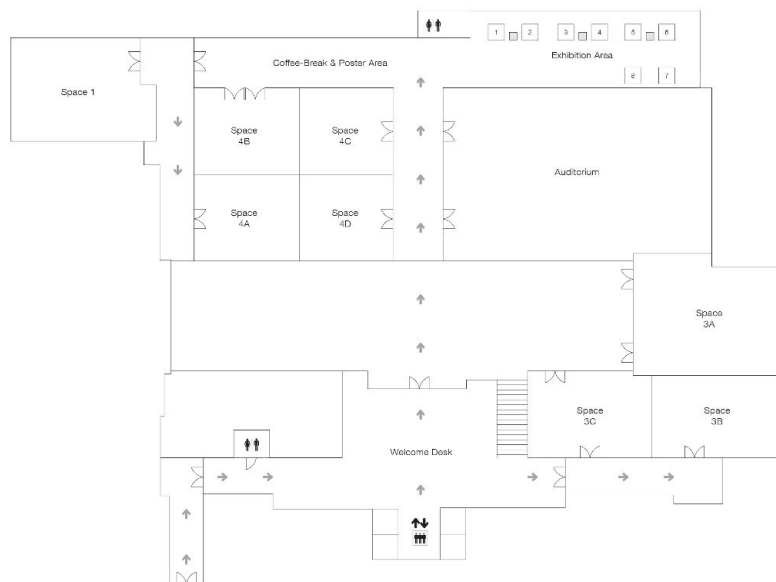
website: <https://visigrapp.scitevents.org>

Rooms Layout

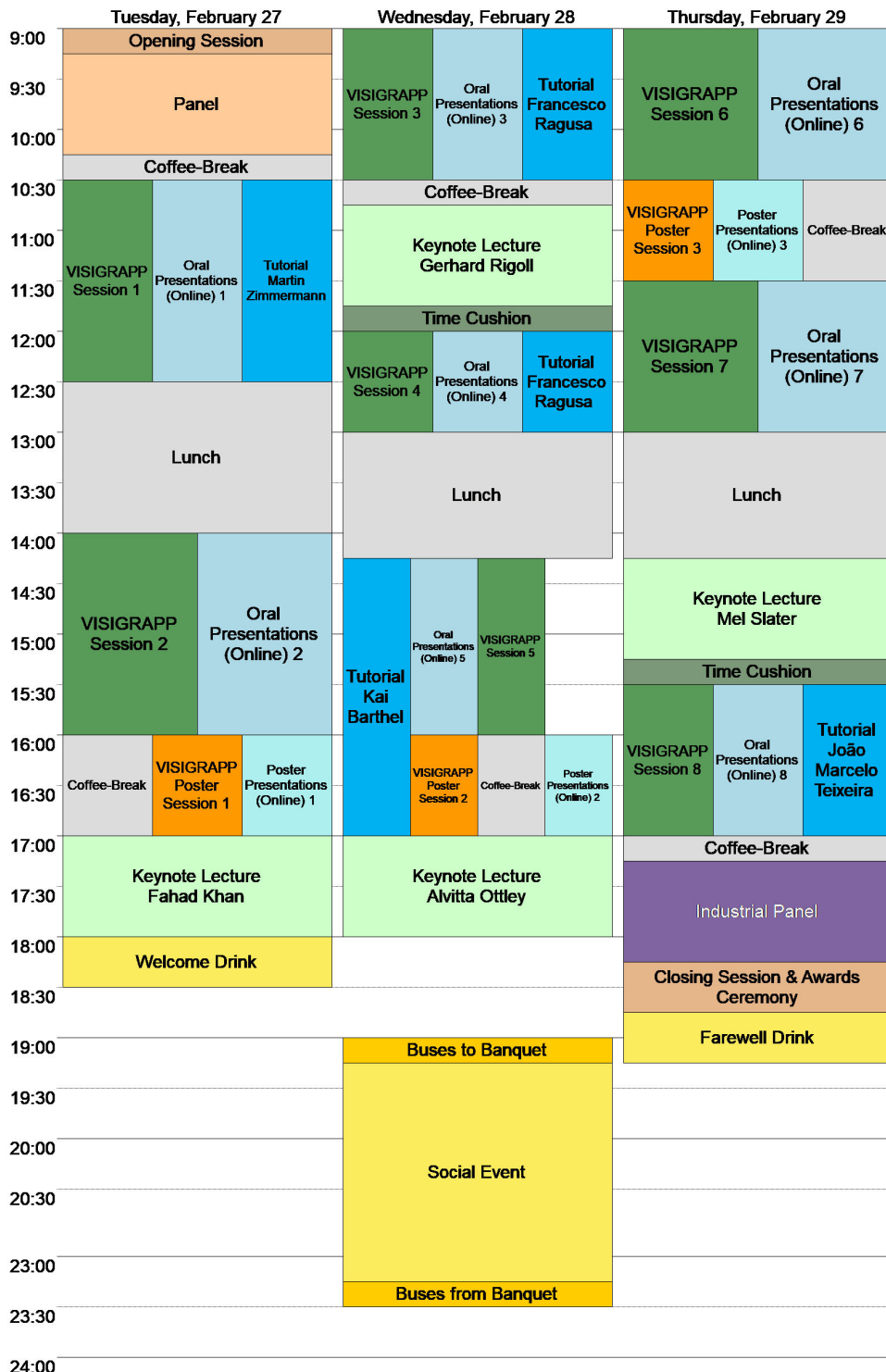
Floor -1



Floor -2



Program Layout



Final Program and Book of Abstracts

Contents

Tuesday Sessions: February 27

Opening Session (09:00 - 09:15)

| | |
|------------------------------------|----|
| Room Plenary - VISIGRAPP | 33 |
|------------------------------------|----|

Panel (09:15 - 10:15)

| | |
|------------------------------------|----|
| Room Plenary - VISIGRAPP | 33 |
|------------------------------------|----|

Session 1 (10:30 - 12:30)

| | |
|---|----|
| Room Space 1 - VISAPP: Event and Human Activity Recognition | 33 |
| Complete Paper #62: S3Aug: Segmentation, Sampling, and Shift for Action Recognition, <i>by Taiki Sugiura and Toru Tamaki</i> | 33 |
| Complete Paper #178: Mediapi-RGB: Enabling Technological Breakthroughs in French Sign Language (LSF) Research Through an Extensive Video-Text Corpus, <i>by Yanis Ouakrim, Hannah Bull, Michèle Gouiffès, Denis Beautemps, Thomas Hueber and Annelies Braffort</i> | 33 |
| Complete Paper #267: Calisthenics Skills Temporal Video Segmentation, <i>by Antonio Finocchiaro, Giovanni Farinella and Antonino Furnari</i> | 33 |
| Complete Paper #242: AR-VPT: Simple Auto-Regressive Prompts for Adapting Frozen ViTs to Videos, <i>by Muhammad Yousuf, Syed Wasim, Syed Hasany and Muhammad Farhan</i> | 33 |
| Complete Paper #247: Benchmarking Sensitivity of Continual Graph Learning for Skeleton-Based Action Recognition, <i>by Wei Wei, Tom De Schepper and Kevin Mets</i> | 34 |
| Room Space 3B - VISAPP: Image Enhancement and Restoration | 34 |
| Complete Paper #115: Simple Base Frame Guided Residual Network for RAW Burst Image Super-Resolution, <i>by Anderson Cotrim, Gerson Barbosa, Cid Santos and Helio Pedrini</i> | 34 |
| Complete Paper #165: Training Methods for Regularizing Gradients on Multi-Task Image Restoration Problems, <i>by Samuel Willingham, Mårten Sjöström and Christine Guillemot</i> | 34 |
| Complete Paper #250: Beyond Variational Models and Self-Similarity in Super-Resolution: Unfolding Models and Multi-Head Attention, <i>by Ivan Pereira-Sánchez, Eloi Sans, Julia Navarro and Joan Duran</i> | 34 |
| Complete Paper #258: Avoiding Undesirable Solutions of Deep Blind Image Deconvolution, <i>by Antonie Brožová and Václav Šmídl</i> | 35 |
| Complete Paper #265: SWViT-RRDB: Shifted Window Vision Transformer Integrating Residual in Residual Dense Block for Remote Sensing Super-Resolution, <i>by Mohamed Ibrahim, Robert Benavente, Daniel Ponsa and Felipe Lumbreras</i> | 35 |
| Room Space 3C - VISAPP: Visual Attention and Image Saliency | 35 |
| Complete Paper #209: When Medical Imaging Met Self-Attention: A Love Story That Didn't Quite Work out, <i>by Tristan Piater, Niklas Penzel, Gideon Stein and Joachim Denzler</i> | 35 |
| Complete Paper #387: Frames Preprocessing Methods for Chromakey Classification in Video, <i>by Evgeny Bessonitsyn, Artyom Chebykin, Grigorii Stafeev and Valeria Efimova</i> | 36 |
| Complete Paper #49: Iterative Saliency Enhancement over Superpixel Similarity, <i>by Leonardo Joao and Alexandre Falcao</i> | 36 |
| Complete Paper #161: Comprehensive Evaluation of End-to-End Driving Model Explanations for Autonomous Vehicles, <i>by Chenkai Zhang, Daisuke Deguchi, Jialei Chen and Hiroshi Murase</i> | 36 |
| Complete Paper #385: A Learning Paradigm for Interpretable Gradients, <i>by Felipe Figueroa, Hanwei Zhang, Ronan Sicre, Yannis Avrithis and Stephane Ayache</i> | 36 |
| Room Space 4A - VISAPP: Deep Learning for Visual Understanding | 37 |
| Complete Paper #124: Pixel-Wise Gradient Uncertainty for Convolutional Neural Networks Applied to Out-of-Distribution Segmentation, <i>by Kira Maag and Tobias Riedlinger</i> | 37 |
| Complete Paper #288: Detecting Anomalies in Textured Images Using Modified Transformer Masked Autoencoder, <i>by Afshin Dini and Esa Rahtu</i> | 37 |
| Complete Paper #331: Alias-Free GAN for 3D-Aware Image Generation, <i>by Attila Szabó, Yevgeniy Puzikov, Sahan Ayvaz, Sonia Aurelio, Peter Gehler, Reza Shirvany and Malte Alf</i> | 37 |
| Complete Paper #157: GAF-Net: Video-Based Person Re-Identification via Appearance and Gait Recognitions, <i>by Moncef Boujou, Rabah Iguernaissi, Lionel Nicod, Djamel Merad and Séverine Dubuisson</i> | 37 |

| | |
|--|----|
| Complete Paper #406: Enhanced Deepfake Detection Using Frequency Domain Upsampling, <i>by Mamadou Bah and Mohamed Dahmane</i> | 38 |
| Room Space 4B - GRAPP: Visualization and Exploration | 38 |
| Complete Paper #367: Thinking on Your Feet: Enhancing Foveated Rendering in Virtual Reality During User Activity, <i>by David Petrescu, Paul Warren, Zahra Montazeri, Gabriel Strain and Steve Pettifer</i> | 38 |
| Complete Paper #213: Viewpoint Selection for Molecular Visualization: Analysis and Applications, <i>by Vincent Larroque, Maxime Maria, Stéphane Mérillou and Matthieu Montes</i> | 38 |
| Complete Paper #390: Exploring the Effect of Display Type on Co-Located Multiple Player Gameplay Performance, Immersion, Social Presence, and Behavior Patterns, <i>by Wenge Xu, Ruichen Zheng, Diego Monteiro, Vijayakumar Nanjappan, Yihong Wang and Hai-Ning Liang</i> | 38 |
| Complete Paper #260: Handing Pedagogical Scenarios Back over to Domain Experts: A Scenario Authoring Model for VR with Pedagogical Objectives, <i>by Mathieu Risy, Valérie Gouranton and Bruno Arnaldi</i> | 39 |
| Oral Presentations (Online) 1 (10:30 - 12:30) | |
| Room Online 1 - VISAPP: Machine Learning Technologies for Vision | 39 |
| Complete Paper #111: Important Pixels Sampling for NeRF Training Based on Edge Values and Squared Errors Between the Ground Truth and the Estimated Colors, <i>by Kohei Fukuda, Takio Kurita and Hiroaki Aizawa</i> | 39 |
| Complete Paper #67: Attacking the Loop: Adversarial Attacks on Graph-Based Loop Closure Detection, <i>by Jonathan Kim, Martin Urschler, Patricia Riddle and Jörg Wicker</i> | 39 |
| Complete Paper #144: Concept Basis Extraction for Latent Space Interpretation of Image Classifiers, <i>by Alexandros Doumanoglou, Dimitrios Zarpalas and Kurt Driessens</i> | 39 |
| Complete Paper #192: A Fusion Approach for Enhanced Remote Sensing Image Classification, <i>by Vian Ahmed, Khaled Jouini, Amel Tuama and Ouajdi Korbaa</i> | 40 |
| Complete Paper #225: Ontology-Driven Deep Learning Model for Multitask Visual Food Analysis, <i>by Daniel Ponte, Eduardo Aguilar, Mireia Ribera and Petia Radeva</i> | 40 |
| Complete Paper #268: A Cascade Methodology to Evaluate Black-Box Recognition Systems Based on a Copycat Algorithm, <i>by Dinh Nguyen, Nhan Le, Van Mai, Tuong Nguyen, Van Nguyen and The Nguyen</i> | 40 |
| Room Online 2 - HUCAPP: Human Computer Interaction Theory and Applications | 41 |
| Complete Paper #275: I Feel Safe with the Prediction: The Effect of Prediction Accuracy on Trust, <i>by Lisa Graichen and Matthias Graichen</i> | 41 |
| Complete Paper #276: Virtual Reality and Autism Spectrum Disorder: Emergence of Sensory-Motor and Olfactory Potentialities in an Anthropocentric Epistemological Approach, <i>by Cécile Lacôte-Coquereau, Paul Richard, Emmanuelle Richard and Patrice Bourdon</i> | 41 |
| Complete Paper #397: Towards Automated Decision Making in Dating Apps Through Pupillary Responses, <i>by Jan Ehlers, Sebastian Alfonso and Arup Mazumder</i> | 41 |
| Complete Paper #94: Applying Cognitive and Cultural Frameworks to mHealth Application Design for Elderly Users, <i>by Mia Forbes, Joyram Chakraborty and Johannes Badejo</i> | 41 |
| Complete Paper #14: Modelling Cognitive Workload to Build Multimodal Voice Interaction in the Car, <i>by Sylvia Bhattacharya and J. Higgins</i> | 42 |
| Complete Paper #48: Look-over-there: Real-World Co-Located Cross-Referencing Using Augmented Reality, <i>by Yuqi Zhou and Voicu Popescu</i> | 42 |
| Tutorial (10:30 - 12:30) | |
| Room Space 4C | 42 |
| Context-Aware Applications Using Visual Programming: Case Studies on Mobile Apps and Humanoid Robot Applications, <i>by Martin Zimmermann</i> | 42 |
| Session 2 (14:00 - 16:00) | |
| Room Space 1 - VISAPP: Transfer Learning | 42 |
| Complete Paper #44: Synthesizing Classifiers from Prior Knowledge, <i>by G. Burghouts, K. Schutte, M. Kruithof, W. Huizinga, F. Ruis and H. Kuijff</i> | 42 |
| Complete Paper #153: Enabling RAW Image Classification Using Existing RGB Classifiers, <i>by Rasmus Munkso, Mathias Andersen, Lau Nørgaard, Andreas Møgelmoose and Thomas Moeslund</i> | 43 |
| Complete Paper #32: Classification of Towels in a Robotic Workcell Using Deep Neural Networks, <i>by Jens Rossen, Patrick Terp, Norbert Krüger, Laus Bigum and Tudor Morar</i> | 43 |
| Complete Paper #279: Unsupervised Annotation and Detection of Novel Objects Using Known Objectness, <i>by Harsh Jadon, Jagdish Deshmukh, Kamakshya Nayak, Kamalakar Thakare and Debi Dogra</i> | 43 |
| Complete Paper #413: Evaluating Multiple Combinations of Models and Encoders to Segment Clouds in Satellite Images, <i>by Jocsan Ferreira, Leandro Silva, Mauricio Escarpinati, André Backes and João Mari</i> | 43 |
| Room Space 3B - VISAPP: Generative AI | 44 |
| Complete Paper #45: StyleHumanCLIP: Text-Guided Garment Manipulation for StyleGAN-Human, <i>by Takato Yoshikawa, Yuki Endo and Yoshihiro Kanamori</i> | 44 |

| | |
|--|----|
| Complete Paper #399: Image Generation from Hyper Scene Graphs with Trinomial Hyperedges Using Object Attention, by Ryosuke Miyake, Tetsu Matsukawa and Einoshin Suzuki | 44 |
| Complete Paper #121: Efficient Posterior Sampling for Diverse Super-Resolution with Hierarchical VAE Prior, by Jean Prost, Antoine Houdard, Andrés Almansa and Nicolas Papadakis | 44 |
| Complete Paper #270: Generating Videos from Stories Using Conditional GAN, by Takahiro Kozaki, Fumihiko Sakaue and Jun Sato | 44 |
| Complete Paper #408: Image Augmentation for Object Detection and Segmentation with Diffusion Models, by Leon Useinov, Valeria Efimova and Sergey Muravyov | 45 |
| Room Space 3C - VISAPP: Image-Based Modeling and 3D Reconstruction | 45 |
| Complete Paper #182: Animating NeRFs from Texture Space: A Framework for Pose-Dependent Rendering of Human Performances, by Paul Knoll, Wieland Morgenstern, Anna Hilsmann and Peter Eisert | 45 |
| Complete Paper #261: ImgAdaPoinTr: Improving Point Cloud Completion via Images and Segmentation, by Vaagn Chopuryan, Mikhail Kuznetsov, Vasilii Latonov and Natalia Semenova | 45 |
| Complete Paper #299: Evaluation of 3D Point Cloud Distances: A Comparative Study in Multi-Point Cloud Fusion Environments, by Ulugbek Alibekov, Vanessa Staderini, Geetha Ramachandran, Philipp Schneider and Doris Antensteiner | 45 |
| Complete Paper #342: Toward Multi-Plane Image Reconstruction from a Casually Captured Focal Stack, by Shiori Ueda, Hideo Saito and Shohei Mori | 46 |
| Room Space 4A - VISAPP: Motion, Tracking and Stereo Vision | 46 |
| Complete Paper #95: BEVSeg2TP: Surround View Camera Bird's-Eye-View Based Joint Vehicle Segmentation and Ego Vehicle Trajectory Prediction, by Sushil Sharma, Arindam Das, Ganesh Sistu, Mark Halton and Ciarán Eising | 46 |
| Complete Paper #125: Are Semi-Dense Detector-Free Methods Good at Matching Local Features?, by Matthieu Vilain, Rémi Giraud, Hugo Germain and Guillaume Bourmaud | 46 |
| Complete Paper #222: BASE: Probably a Better Approach to Visual Multi-Object Tracking, by Martin Larsen, Sigmund Rolfsjord, Daniel Gusland, Jörgen Ahlberg and Kim Mathiassen | 46 |
| Complete Paper #370: A Multilevel Strategy to Improve People Tracking in a Real-World Scenario, by Cristiano B. de Oliveira, Joao Neves, Rafael Ribeiro and David Menotti | 47 |
| Complete Paper #405: Self-Mounted Motion Capture System Using Mutual Projection of Asynchronous Cameras, by Kazusa Ozaki, Fumihiko Sakaue and Jun Sato | 47 |
| Room Space 4B - GRAPP: Animation and Simulation | 47 |
| Complete Paper #75: Real-Time Desertscares Simulation with CUDA, by Alexander Nilles, Lars Günther and Stefan Müller | 47 |
| Complete Paper #143: Using the Polynomial Particle-in-Cell Method for Liquid-Fabric Interaction, by Robert Dennison and Steve Maddock | 47 |
| Complete Paper #256: Data-Driven Viscosity Solver for Fluid Simulation, by Wonjung Park, Hyunsoo Kim and Jinah Park | 48 |
| Complete Paper #334: Transformer-Based Two-level Approach for Music-driven Dance Choreography, by Yanbo Cheng and Yingying Wang | 48 |
| Room Space 4C - IVAPP: Visualization Techniques | 48 |
| Complete Paper #194: Scale and Time Independent Clustering of Time Series Data, by Florian Steinwider, Istvan Szilagy, Eva Eggeling and Torsten Ullrich | 48 |
| Complete Paper #120: Fundamental Limitations of Inverse Projections and Decision Maps, by Yu Wang and Alexandru Telea | 48 |
| Complete Paper #455: Understanding How Different Visual Aids for Augmented Reality Influence Tool-Patient Alignment in Surgical Tasks: A Preliminary Study, by Stefano Stradiotti, Nicolas Emiliani, Emanuela Marcelli and Laura Cercenelli | 49 |
| Oral Presentations (Online) 2 (14:00 - 16:00) | |
| Room Online 2 - VISAPP: Features Extraction | 49 |
| Complete Paper #140: Deep Learning-Based Models for Performing Multi-Instance Multi-Label Event Classification in Gameplay Footage, by Etienne Julia, Marcelo Zanchetta do Nascimento, Matheus Faria and Rita Julia | 49 |
| Complete Paper #234: Robust Denoising and DenseNet Classification Framework for Plant Disease Detection, by Kevin Zhou and Dimah Dera | 49 |
| Complete Paper #317: Facial Point Graphs for Amyotrophic Lateral Sclerosis Identification, by Nicolas Gomes, Arissa Yoshida, Mateus Roder, Guilherme Camargo de Oliveira and João Papa | 49 |
| Complete Paper #315: On the Use of Visual Transformer for Image Complexity Assessment, by Luigi Celona, Gianluigi Ciocca and Raimondo Schettini | 50 |
| Complete Paper #410: Enhancement-Driven Pretraining for Robust Fingerprint Representation Learning, by Ekta Gavas, Kaustubh Olpadkar and Anoop Namboodiri | 50 |
| Room Online 1 - VISAPP: Deep Learning for Visual Understanding | 50 |
| Complete Paper #155: Image Inpainting on the Sketch-Pencil Domain with Vision Transformers, by Jose Campana, Luís Decker, Marcos Souza, Helena Maia and Helio Pedrini | 50 |
| Complete Paper #384: Single-Class Instance Segmentation for Vectorization of Line Drawings, by Rhythm Vohra, Amanda Dash and Alexandra Branzan Albu | 50 |

| | |
|--|----|
| Complete Paper #78: Data Quality Aware Approaches for Addressing Model Drift of Semantic Segmentation Models, <i>by Samiha Mirza, Vuong Nguyen, Pranav Mantini and Shishir Shah</i> | 51 |
| Complete Paper #303: Transformer-Based Video Saliency Prediction with High Temporal Dimension Decoding, <i>by Morteza Moradi, Simone Palazzo and Concetto Spampinato</i> | 51 |
| Complete Paper #430: Decoding Visual Stimuli and Visual Imagery Information from EEG Signals Utilizing Multi-Perspective 3D-CNN Based Hierarchical Deep-Fusion Learning Network, <i>by Fatma Emanet and Kazim Sekeroglu</i> | 51 |

Poster Session 1 (16:00 - 17:00)

| | |
|--|----|
| Foyer - GRAPP | 51 |
| Complete Paper #73: ML-Tree and MRL-Tree: Combining Mass-Spring System, Rigid-Body Dynamics and L-Systems to Model Physical Effects on Trees, <i>by See Min Lim and Like Gobeawan</i> | 51 |
| Complete Paper #91: Pure Physics-Based Hand Interaction in VR, <i>by Mohammed-Bashir Mahdi, Erwan Guillou, Alexandre Meyer, Arash Habibi and Saïda Bouakaz</i> | 52 |
| Complete Paper #226: Automatic Registration of 3D Point Cloud Sequences, <i>by Natálie Vítová, Jakub Frank and Libor Váša</i> | 52 |
| Foyer - IVAPP | 52 |
| Complete Paper #16: Flowstrates++: An Approach to Visualize Multi-Dimensional OD Data, <i>by Nicolas Fuchs, Pierre Vanhulst, Raphaël Tuor and Denis Lalanne</i> | 52 |
| Complete Paper #141: Dashboard Design: Interactive and Visual Exploration of Spotify Songs, <i>by Sarah Clavdetscher, Michael Schlotter, Nadine Christen, Juliane Streitberg and Michael Burch</i> | 52 |
| Complete Paper #444: Simultaneous Optimization of Edge Bundling and Node Layout Using Genetic Algorithm, <i>by Junsei Meikari and Ryosuke Saga</i> | 53 |
| Foyer - HUCAPP | 53 |
| Abstract #30: Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality, <i>by Dawid Ratajczyk, Monika Żyła, Piort Jaworski and Paweł Łupkowski</i> | 53 |
| Complete Paper #42: Hôsea: A Touch Table for Cognitive and Motor Rehabilitation for the Elderly - A Preliminary Study, <i>by Maxime Macé, Lise Macé, Emmanuelle Ménétrier, Paul Richard and Tassadit Amghar</i> | 53 |
| Complete Paper #285: ARTISTA: Redefining Pottery Design with Virtual Reality and Physically Simulated Clay, <i>by Gilda Manfredi, Gabriele Gilio, Nicola Capece, Ugo Erra and Vincenzo Baldi</i> | 53 |
| Complete Paper #350: Assessment of the Detectability of Vulnerable Road Users: An Empirical Study, <i>by Wentong Yang, Shota Matsubayashi, Kazuhisa Miwa, Shinya Kitayama, Manabu Otsuka and Koji Hamada</i> | 54 |
| Complete Paper #409: Visual Behavior Based on Information Foraging Theory Toward Designing of Auditory Information, <i>by Yuta Kurihara, Motoki Shino, Katsuko Nakahira and Muneo Kitajima</i> | 54 |
| Complete Paper #428: A Comparative Study: Augmented and Virtual Reality Applications for Improving Comprehension of Abstract Programming Concepts, <i>by Omer Cinar, Karen Rafferty, David Cutting and Hui Wang</i> | 54 |
| Foyer - VISAPP | 55 |
| Complete Paper #27: Synthetic Data-Driven Approach for Missing Nut and Bolt Classification in Flange Joints, <i>by Frankly Toro, Hassane Trigui, Yazeed Alnumay, Siddharth Mishra and Sahejad Patel</i> | 55 |
| Complete Paper #28: Finding and Navigating to Humans in Complex Environments for Assistive Tasks, <i>by Asfand Yaar, Antonino Furnari, Marco Rosano, Aki Härmä and Giovanni Farinella</i> | 55 |
| Complete Paper #36: Towards Better Morphed Face Images Without Ghosting Artifacts, <i>by Clemens Seibold, Anna Hilsman and Peter Eisert</i> | 55 |
| Complete Paper #47: Teeth Localization and Lesion Segmentation in CBCT Images Using SpatialConfiguration-Net and U-Net, <i>by Arnela Hadzic, Barbara Kirnbauer, Darko Štern and Martin Urschler</i> | 55 |
| Complete Paper #58: Multi-View 3D Reconstruction for Construction Site Monitoring, <i>by Guangan Chen, Michiel Vlaminc, Wilfried Philips and Hiep Luong</i> | 56 |
| Complete Paper #60: Pedestrian's Gaze Object Detection in Traffic Scene, <i>by Hiroto Murakami, Jialei Chen, Daisuke Deguchi, Takatsugu Hirayama, Yasutomo Kawanishi and Hiroshi Murase</i> | 56 |
| Complete Paper #64: Improved Pest Detection in Insect Larvae Rearing with Pseudo-Labeling and Spatio-Temporal Masking, <i>by Paweł Majewski, Piotr Lampa, Robert Burduk and Jacek Reiner</i> | 56 |
| Complete Paper #71: Online Human Activity Recognition Using Efficient Neural Architecture Search with Low Environmental Impact, <i>by Nassim Mokhtari, Alexis Nédélec, Marlène Gilles and Pierre De Loor</i> | 56 |
| Complete Paper #74: Towards Rapid Prototyping and Comparability in Active Learning for Deep Object Detection, <i>by Tobias Riedlinger, Marius Schubert, Karsten Kahl, Hanno Gottschalk and Matthias Rottmann</i> | 57 |
| Complete Paper #83: A Study of Real World Information Mapping for Information Sharing Using Edge Devices, <i>by Takenori Hara and Hideo Saito</i> | 57 |
| Complete Paper #105: Sumo Action Classification Using Mawashi Keypoints, <i>by Yuki Utsuro, Hidehiko Shishido and Yoshinari Kameda</i> | 57 |
| Complete Paper #189: Automatic Error Correction of GPT-Based Robot Motion Generation by Partial Affordance of Tool, <i>by Takahiro Suzuki, Yuta Ando and Manabu Hashimoto</i> | 57 |
| Complete Paper #259: SynthRSF: A Novel Photorealistic Synthetic Dataset for Adverse Weather Condition Denoising, <i>by Angelos Kanlis, Vazgken Vanian, Sotiris Karvarsamis, Ioanna Gkika, Konstantinos Konstantoudakis and Dimitrios Zarpalas</i> | 58 |

Poster Presentations (Online) 1 (16:00 - 17:00)

| | |
|--|----|
| Room Online 1 - VISIGRAPP | 58 |
| Room Online 1 - IVAPP | 58 |
| Complete Paper #246: Bringing Objects to Life: Supporting Program Comprehension Through Animated 2.5D Object Maps from Program Traces, <i>by Christoph Thiede, Willy Scheibel and Jürgen Döllner</i> | 58 |
| Complete Paper #269: Human-Machine Collaboration for the Visual Exploration and Analysis of High-Dimensional Spatial Simulation Ensembles, <i>by Mai Dahshan, Nicholas Polys, Leanna House, Karim Youssef and Ryan Pollyea</i> | 58 |
| Room Online 1 - VISAPP | 58 |
| Complete Paper #39: Anisotropic Diffusion for Depth Estimation in Shape from Focus Systems, <i>by Bilal Ahmad, Ivar Farup and Pål Floor</i> | 58 |
| Complete Paper #41: Informative Rays Selection for Few-Shot Neural Radiance Fields, <i>by Marco Orsingher, Anthony Dell'Eva, Paolo Zani, Paolo Medici and Massimo Bertozzi</i> | 59 |
| Complete Paper #170: PIRO: Permutation-Invariant Relational Network for Multi-Person 3D Pose Estimation, <i>by Nicolas Ugrinovic, Adria Ruiz, Antonio Agudo, Alberto Sanfeliu and Francesc Moreno-Noguer</i> | 59 |
| Complete Paper #205: NeRF-Diffusion for 3D-Consistent Face Generation and Editing, <i>by Héctor Laria, Kai Wang, Joost van de Weijer, Bogdan Raducanu and Yaxing Wang</i> | 59 |
| Room Online 2 - VISIGRAPP | 59 |
| Room Online 2 - VISAPP | 59 |
| Complete Paper #158: Assessing the Performance of Autoencoders for Particle Density Estimation in Acoustofluidic Medium: A Visual Analysis Approach, <i>by Lucas Massa, Tiago Vieira, Allan Martins and Bruno Ferreira</i> | 59 |
| Complete Paper #160: Image Edge Enhancement for Effective Image Classification, <i>by Bu Tianhao, Michalis Lazarou and Tania Stathaki</i> | 60 |
| Complete Paper #333: A Computer Vision Approach to Compute Bubble Flow of Offshore Wells, <i>by Rogerio Hart and Aura Conci</i> | 60 |
| Complete Paper #442: Selection of Backbone for Feature Extraction with U-Net in Pancreas Segmentation, <i>by Alexandre Araújo, Joao Sousa de Almeida, Anselmo Cardoso de Paiva and Geraldo Braz Junior</i> | 60 |
| Complete Paper #310: Influence of Pixel Perturbation on eXplainable Artificial Intelligence Methods, <i>by Juliana Feitosa, Mateus Roder, João Papa and José Brega</i> | 60 |
| Complete Paper #318: Convolutional Neural Networks and Image Patches for Lithological Classification of Brazilian Pre-Salt Rocks, <i>by Mateus Roder, Leandro Passos, Clayton Pereira, João Papa, Altanir Mello Junior, Marcelo Fagundes de Rezende, Yaro Silva and Alexandre Vidal</i> | 61 |

Keynote Lecture (17:00 - 18:00)

| | |
|--|----|
| Room Plenary - VISIGRAPP | 61 |
| Towards Detailed Understanding of the Visual World, <i>by Fahad Khan</i> | 61 |

Wednesday Sessions: February 28**Session 3 (09:00 - 10:30)**

| | |
|---|----|
| Room Space 1 - VISAPP: Categorization and Scene Understanding | 65 |
| Complete Paper #23: Enabling On-Device Continual Learning with Binary Neural Networks and Latent Replay, <i>by Lorenzo Vorabbi, Davide Maltoni, Guido Borghi and Stefano Santi</i> | 65 |
| Complete Paper #365: Anomaly Detection on Roads Using an LSTM and Normal Maps, <i>by Yusuke Nonaka, Hideo Saito, Hideaki Uchiyama, Kyota Higa and Masahiro Yamaguchi</i> | 65 |
| Complete Paper #51: Cybersecurity Intrusion Detection with Image Classification Model Using Hilbert Curve, <i>by Punyawat Jaroensiripong, Karin Sumongkayothin, Prarinya Siritanawan and Kazunori Kotani</i> | 65 |
| Complete Paper #407: SMART-RD: Towards a Risk Assessment Framework for Autonomous Railway Driving, <i>by Justin Bescop, Nicolas Goeman, Amel Aissaoui, Benjamin Allaert and Jean-Philippe Vandeborre</i> | 65 |
| Room Space 3B - VISAPP: Machine Learning Technologies for Vision | 66 |
| Complete Paper #255: Investigating the Corruption Robustness of Image Classifiers with Random p -norm Corruptions, <i>by Georg Siedel, Weijia Shao, Silvia Vock and Andrey Morozov</i> | 66 |
| Complete Paper #368: Beyond the Known: Adversarial Autoencoders in Novelty Detection, <i>by Muhammad Asad, Ihsan Ullah, Ganesh Sistu and Michael Madden</i> | 66 |
| Complete Paper #173: Large Filter Low-Level Processing by Edge TPU, <i>by Gerald Krell and Thilo Pionteck</i> | 66 |
| Complete Paper #272: Anomaly Detection and Localization for Images of Running Paper Web in Paper Manufacturing, <i>by Afshin Dini, Marja Mettänen and Esa Rahtu</i> | 66 |
| Room Space 3C - VISAPP: Multimedia Forensics | 67 |
| Complete Paper #264: Defying Limits: Super-Resolution Refinement with Diffusion Guidance, <i>by Marcelo Santos, João Neves, Hugo Proença and David Menotti</i> | 67 |
| Complete Paper #293: On Feasibility of Transferring Watermarks from Training Data to GAN-Generated Fingerprint Images, <i>by Venkata Mannam, Andrey Makrushin and Jana Dittmann</i> | 67 |

| | |
|--|----|
| Complete Paper #302: Towards the Detection of Diffusion Model Deepfakes, <i>by Jonas Ricker, Simon Damm, Thorsten Holz and Asja Fischer</i> | 67 |
| Complete Paper #239: Diffusion-Inspired Dynamic Models for Enhanced Fake Face Detection, <i>by Mamadou Bah, Rajjeshwar Ganguly and Mohamed Dahmane</i> | 67 |
| Room Space 4A - VISAPP: Object Detection and Localization | 68 |
| Complete Paper #171: End-to-End Chess Recognition, <i>by Athanasios Masouris and Jan van Gemert</i> | 68 |
| Complete Paper #424: CSE: Surface Anomaly Detection with Contrastively Selected Embedding, <i>by Simon Thomine and Hichem Snoussi</i> | 68 |
| Complete Paper #220: Understanding Marker-Based Normalization for FLIM Networks, <i>by Leonardo Joao, Matheus Cerqueira, Barbara Benato and Alexandre Falcao</i> | 68 |
| Complete Paper #326: AbSynth: Using Abstract Image Synthesis for Synthetic Training, <i>by Dominik Penk, Maik Horn, Christoph Strohmeier, Bernhard Egger, Marc Stamminger and Frank Bauer</i> | 69 |
| Room Space 4B - GRAPP: Rendering | 69 |
| Complete Paper #98: Real-Time Editing of Path-Traced Scenes with Prioritized Re-Rendering, <i>by Annalena Ulschmid, Bernhard Kerbl, Katharina Krösl and Michael Wimmer</i> | 69 |
| Complete Paper #301: Hierarchical Bitmask Implicit Grids for Efficient Point-in-Volume Queries on the GPU, <i>by Julius Ikkala, Tuomas Lantti, Pekka Jääskeläinen and Markku Mäkitalo</i> | 69 |
| Complete Paper #451: Non-Photorealistic Rendering of 3D Point Clouds Using Segment-Specific Image-Space Effects, <i>by Ole Wegen, Josafat-Mattias Burmeister, Max Reimann, Rico Richter and Jürgen Döllner</i> | 69 |
| Complete Paper #320: Sparse Spatial Shading in Augmented Reality, <i>by Rikard Olajos and Michael Doggett</i> | 70 |
| Room Space 4D - HUCAPP: Theories, Models and User Evaluation | 70 |
| Complete Paper #59: Automatic Viewpoint Selection for Interactive Motor Feedback Using Principle Component Analysis, <i>by Florian Diller, Alexander Wiebel and Gerik Scheuermann</i> | 70 |
| Complete Paper #37: AR Authoring: How to Reduce Errors from the Start?, <i>by Camille Truong-Allié, Martin Herbeth and Alexis Paljic</i> | 70 |
| Complete Paper #112: Using Fitts' Law to Compare Sonification Guidance Methods for Target Reaching Without Vision, <i>by Coline Fons, Sylvain Huet, Denis Pellerin and Christian Graff</i> | 70 |
| Oral Presentations (Online) 3 (09:00 - 10:30) | |
| Room Online 2 - IVAPP: Information Visualization Theory and Applications | 71 |
| Complete Paper #295: Using Retrieval Augmented Generation to Build the Context for Data-Driven Stories, <i>by Angelica Lo Duca</i> | 71 |
| Complete Paper #372: Particle-Wise Higher-Order SPH Field Approximation for DVR, <i>by Jonathan Fischer, Martin Schulze, Paul Rosenthal and Lars Linsen</i> | 71 |
| Room Online 1 - VISAPP: Machine Learning Technologies for Vision | 71 |
| Complete Paper #52: Augmenting Human-Robot Collaboration Task by Human Hand Position Forecasting, <i>by Shyngyskhan Abilkassov, Michael Gentner and Mirela Popa</i> | 71 |
| Complete Paper #69: Vision-Perceptual Transformer Network for Semantic Scene Understanding, <i>by Mohamad Alansari, Hamad AlRemeithi, Bilal Hassan, Sara Alansari, Jorge Dias, Majid Khonji, Naoufel Werghi and Sajid Javed</i> | 72 |
| Complete Paper #183: Social Distancing Monitoring by Human Detection Through Bird's-Eye View Technique, <i>by Gona Rozhbayani, Amel Tuama and Fadwa Al-Azzo</i> | 72 |
| Complete Paper #356: Visual Perception of Obstacles: Do Humans and Machines Focus on the Same Image Features?, <i>by Constantinos Kyriakides, Marios Thoma, Zenonas Theodosiou, Harris Partaourides, Loizos Michael and Andreas Lanitis</i> | 72 |
| Tutorial (09:00 - 10:30) | |
| Room Space 4C | 72 |
| First Person (Egocentric) Vision: History and Applications, <i>by Francesco Ragusa</i> | 72 |
| Keynote Lecture (10:45 - 11:45) | |
| Room Plenary - VISIGRAPP | 73 |
| Multi-Modal Human-Machine Interaction: Joint Optimization of Single Modalities and Automatic Learning of Communication Channel Fusion, <i>by Gerhard Rigoll</i> | 73 |
| Session 4 (12:00 - 13:00) | |
| Room Space 1 - VISAPP: Segmentation and Grouping | 73 |
| Complete Paper #166: Semantic State Estimation in Robot Cloth Manipulations Using Domain Adaptation from Human Demonstrations, <i>by Georgies Tzelepis, Eren Aksoy, Júlia Borràs and Guillem Alenyà</i> | 73 |
| Complete Paper #103: Stereo-Event-Camera-Technique for Insect Monitoring, <i>by Regina Pohle-Fröhlich, Colin Gebler and Tobias Bolten</i> | 73 |
| Complete Paper #168: Instance Segmentation of Event Camera Streams in Outdoor Monitoring Scenarios, <i>by Tobias Bolten, Regina Pohle-Fröhlich and Klaus Tönnies</i> | 73 |

| | |
|---|----|
| Room Space 3B - VISAPP: <i>Assistive Computer Vision</i> | 74 |
| Complete Paper #43: Enhanced Multimodal Timely Prediction of Pulmonary Fibrosis Progression with Uncertainty Estimation from Chest CT Images and Clinical Metadata, <i>by Mohamed Dahmane</i> | 74 |
| Complete Paper #244: Physical Ergonomics Anticipation with Human Motion Prediction, <i>by Mattias Billast, Jonas De Bruyne, Klaas Bombeke, Tom De Schepper and Kevin Mets</i> | 74 |
| Complete Paper #453: Automatic Assessment of Skill and Performance in Fencing Footwork, <i>by Filip Malawski, Marek Krupa and Ksawery Kapela</i> | 74 |
| Room Space 3C - VISAPP: <i>Self-supervised Learning</i> | 74 |
| Complete Paper #76: Deep Active Learning with Noisy Oracle in Object Detection, <i>by Marius Schubert, Tobias Riedlinger, Karsten Kahl and Matthias Rottmann</i> | 74 |
| Complete Paper #88: A Comparative Evaluation of Self-Supervised Methods Applied to Rock Images Classification, <i>by Van Nguyen, Dominique Fourer, Désiré Sidibé, Jean-François Lecomte and Souhail Youssef</i> | 75 |
| Complete Paper #169: Diverse Data Selection Considering Data Distribution for Unsupervised Continual Learning, <i>by Naoto Hayashi, Naoki Okamoto, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi</i> | 75 |
| Room Space 4A - VISAPP: <i>Image Formation, Acquisition Devices and Sensors</i> | 75 |
| Complete Paper #298: The Risk of Image Generator-Specific Traces in Synthetic Training Data, <i>by Georg Wimmer, Dominik Söllinger and Andreas Uhl</i> | 75 |
| Complete Paper #312: FuDensityNet: Fusion-Based Density-Enhanced Network for Occlusion Handling, <i>by Zainab Ouardirhi, Otmane Amel, Mostapha Zbakh and Sidi Mahmoudi</i> | 75 |
| Complete Paper #325: Multidimensional Compressed Sensing for Spectral Light Field Imaging, <i>by Wen Cao, Ehsan Miandji and Jonas Unger</i> | 76 |
| Room Space 4B - GRAPP: <i>Information and Virtual Reality</i> | 76 |
| Complete Paper #262: Real-Time 3D Information Visualization on Mobile Devices: Efficient Occlusion Detection for Geospatial Applications, <i>by Agata Migalska</i> | 76 |
| Complete Paper #363: Exploring Foveation Techniques for Virtual Reality Environments, <i>by Razeen Hussain, Manuela Chessa and Fabio Solari</i> | 76 |
| Abstract #458: Domain Specific Language for Type Casting in Image Processing, <i>by Yamato Kanetaka, Haruki Nogami, Yuki Naganawa and Norishige Fukushima</i> | 76 |
| Room Space 4D - IVAPP: <i>Text Visualization and Storytelling</i> | 77 |
| Complete Paper #237: Quantifying Topic Model Influence on Text Layouts Based on Dimensionality Reductions, <i>by Daniel Atzberger, Tim Cech, Willy Scheibel, Jürgen Döllner and Tobias Schreck</i> | 77 |
| Complete Paper #436: A Survey on Storytelling Techniques for Heritage on Nazi Persecution, <i>by Niek Meffert, Camilla Østergaard, Stefan Jänicke, Richard Khulusi, Esther Rachow and Nicklas Andersen</i> | 77 |
| Oral Presentations (Online) 4 (12:00 - 13:00) | |
| Room Online 2 - HUCAPP: <i>Human Computer Interaction Theory and Applications</i> | 77 |
| Complete Paper #129: Real-Time Heart Rate Visualization for Individuals with Autism Spectrum Disorder: An Evaluation of Technology Assisted Physical Activity Application to Increase Exercise Intensity, <i>by Bo Fu, Katrina Orevillo, Dennis Lo, Andrew Bae and Melissa Bittner</i> | 77 |
| Room Online 1 - VISAPP: <i>Object Detection and Localization</i> | 78 |
| Complete Paper #184: Attentive-YOLO: On-Site Water Pipeline Inspection Using Efficient Channel Attention and Reduced ELAN-Based YOLOv7, <i>by Atanu Roy and Priyanka Bagade</i> | 78 |
| Complete Paper #212: High Precision Single Shot Object Detection in Automotive Scenarios, <i>by Soumya A, C Krishna Mohan and Linga Reddy Cenkeramaddi</i> | 78 |
| Complete Paper #433: Variational Autoencoders for Pedestrian Synthetic Data Augmentation of Existing Datasets: A Preliminary Investigation, <i>by Ivan Nikolov</i> | 78 |
| Tutorial (12:00 - 13:00) | |
| Room Space 4C | 79 |
| First Person (Egocentric) Vision: History and Applications, <i>by Francesco Ragusa</i> | 79 |
| Session 5 (14:15 - 16:00) | |
| Room Space 1 - VISAPP: <i>Features Extraction</i> | 79 |
| Complete Paper #56: Calibration-Accuracy Measurement in Railway Overlapping Multi-Camera Systems, <i>by Martí Sánchez, Nerea Aranjuelo, Jon Iñiguez de Gordo, Pablo Alonso, Mikel García, Marcos Nieto and Mikel Labayen</i> | 79 |
| Complete Paper #123: Kore Initial Clustering for Unsupervised Domain Adaptation, <i>by Kyungsik Lee, Youngmi Jun, Eunji Kim, Suhyun Kim, Seong Hwang and Jonghyun Choi</i> | 79 |
| Complete Paper #201: Feature Selection Using Quantum Inspired Island Model Genetic Algorithm for Wheat Rust Disease Detection and Severity Estimation, <i>by Sourav Samanta, Sanjay Chatterji and Sanjoy Pratihar</i> | 79 |
| Complete Paper #359: ELSA: Expanded Latent Space Autoencoder for Image Feature Extraction and Classification, <i>by Emerson Vilar de Oliveira, Dunfrey Aragão and Luiz Gonçalves</i> | 80 |
| Complete Paper #366: On Granularity Variation of Air Quality Index Visualization from Sentinel-5, <i>by Jordan Cuno, Arthur Bezerra, Aura Conci and Luiz Gonçalves</i> | 80 |

| | |
|---|----|
| Room Space 3B - VISAPP: Deep Learning for Visual Understanding | 80 |
| Complete Paper #154: Challenging the Black Box: A Comprehensive Evaluation of Attribution Maps of CNN Applications in Agriculture and Forestry, by <i>Lars Nieradzik, Henrike Stephani, Jördis Sieburg-Rockel, Stephanie Helmling, Andrea Olbrich and Janis Keuper</i> | 80 |
| Complete Paper #200: Scale Learning in Scale-Equivariant Convolutional Networks, by <i>Mark Basting, Robert-Jan Bruinjtjes, Thaddäus Wiedemer, Matthias Kümmerer, Matthias Bethge and Jan van Gemert</i> | 80 |
| Complete Paper #327: AV-PEA: Parameter-Efficient Adapter for Audio-Visual Multimodal Learning, by <i>Abduljalil Radman and Jorma Laaksonen</i> | 81 |
| Complete Paper #369: Improving Low-Light Image Recognition Performance Based on Image-Adaptive Learnable Module, by <i>Seitaro Ono, Yuka Ogino, Takahiro Toizumi, Atsushi Ito and Masato Tsukada</i> | 81 |
| Complete Paper #423: Fact-Checked Claim Detection in Videos Using a Multimodal Approach, by <i>Frédéric Rayar</i> | 81 |
| Room Space 3C - VISAPP: 3D Deep Learning | 81 |
| Complete Paper #174: Hand Mesh and Object Pose Reconstruction Using Cross Model Autoencoder, by <i>Chaitanya Bandi and Ulrike Thomas</i> | 81 |
| Complete Paper #188: HD-VoxelFlex: Flexible High-Definition Voxel Grid Representation, by <i>Igor Vozniak, Pavel Astreika, Philipp Müller, Nils Lipp, Christian Müller and Philipp Slusallek</i> | 82 |
| Complete Paper #248: RailCloud-HdF: A Large-Scale Point Cloud Dataset for Railway Scene Semantic Segmentation, by <i>Mahdi Abid, Mathis Teixeira, Ankur Mahtani and Thomas Laurent</i> | 82 |
| Complete Paper #445: Vehicle Pose Estimation: Exploring Angular Representations, by <i>Ivan Orlov, Marco Buzzelli and Raimondo Schettini</i> | 82 |
| Room Space 4A - VISAPP: Machine Learning Technologies for Vision | 82 |
| Complete Paper #128: Multispectral Stereo-Image Fusion for 3D Hyperspectral Scene Reconstruction, by <i>Eric Wisotzky, Jost Triller, Anna Hilsmann and Peter Eisert</i> | 82 |
| Complete Paper #191: Deep Learning-Powered Assembly Step Classification for Intricate Machines, by <i>Luca Rodiga, Eva Eggeling, Ulrich Krispel and Torsten Ullrich</i> | 83 |
| Complete Paper #308: Recognizing Actions in High-Resolution Low-Framerate Videos: A Feasibility Study in the Construction Sector, by <i>Benjamin Vandersmissen, Arian Sabaghi, Phil Reiter and Jose Oramas</i> | 83 |
| Complete Paper #322: Large Scale Graph Construction and Label Propagation, by <i>Z. Ibrahim, A. Bosaghzadeh and F. Dornaika</i> | 83 |
| Complete Paper #323: Image Augmentation Preserving Object Parts Using Superpixels of Variable Granularity, by <i>D. Sun and F. Dornaika</i> | 83 |
| Room Space 4B - HUCAPP: Agents and Human Interaction | 84 |
| Complete Paper #87: Evaluating Blink Rate as a Dynamic Indicator of Mental Workload in a Flight Simulator, by <i>Simon Schwed and Axel Schulte</i> | 84 |
| Complete Paper #20: Gaia: A Social Robot to Help Connect Humans and Plants, by <i>Christopher Xenophonos, Teresa Clark, Michael Seals, Cole Lampman, Ilyas Tursynbek and Mounia Ziat</i> | 84 |
| Abstract #21: Exploring the Uncanny Valley Phenomenon for Hand Prosthetics: Stimuli Presentation Factor, by <i>Pawel Lupkowski, Aleksandra Wasielewska and Marcin Jukiewicz</i> | 84 |
| Complete Paper #147: A Digital Inclusion and Technological Barriers: Investigating the Challenges Faced by Formerly Incarcerated Populations in Adopting and Accessing Technology, by <i>Johannes Badejo, Joyram Chakraborty and Mia Forbes</i> | 85 |
| Complete Paper #450: Performance Evaluation of Visual Analytics Framework for Monitoring Neuromotor Rehabilitation, by <i>Maryam Boumrah, Samir Garbaya and Amina Radgui</i> | 85 |
| Oral Presentations (Online) 5 (14:15 - 16:00) | |
| Room Online 1 - VISAPP: Computer Vision Theory and Applications | 85 |
| Complete Paper #196: Symmetry Completion Test: A Novel Approach for Visual Distortion Mapping and Correction Using Symmetry Constraints, by <i>Ye Ling, David Frohlich, Tom Williamson and Jean-Yves Guillemaut</i> | 85 |
| Complete Paper #278: Deep Discriminative Feature Learning for Document Image Manipulation Detection, by <i>Kais Rouis, Petra Gomez-Krämer, Mickaël Coustaty, Saddock Kébairi and Vincent Poulain d'Andecy</i> | 85 |
| Complete Paper #307: Explainability and Interpretability for Media Forensic Methods: Illustrated on the Example of the Steganalysis Tool Stegdetect, by <i>Christian Kraetzer and Mario Hildebrandt</i> | 86 |
| Complete Paper #392: A Tool for 3D Representation of the 2D Thermographic Breast Acquisitions, by <i>Eudoxia Moura, Gleidson Costa, Tiago Borchardt and Aura Conci</i> | 86 |
| Complete Paper #432: Automatic Computation of the Posterior Nipple Line from Mammographies, by <i>Quynh Tran, Tina Santner and Antonio Rodríguez-Sánchez</i> | 86 |
| Room Online 2 - GRAPP: Computer Graphics Theory and Applications | 86 |
| Complete Paper #330: Automated Palette Cycling Animations, by <i>Ali Javid and David Mould</i> | 86 |
| Complete Paper #347: Adaptation Speed for Exposure Control in Virtual Reality, by <i>Claus Madsen and Johan Kristensen</i> | 87 |
| Complete Paper #345: Virtual Reality for Detailed Visualization and Generation of Proximal and Distal Bone Fracture Patterns, by <i>J. J. Jiménez-Delgado, F. D. Pérez-Cano, G. Parra-Cabrera and I. Remolar-Quintana</i> | 87 |
| Tutorial (14:15 - 17:00) | |

| | |
|--|----|
| Room Space 4C | 87 |
| Grid-based Layout Algorithms for Visual Sorting, Dimensionality Reduction, and Optimization, <i>by Kai U. Barthel</i> | 87 |
| Poster Session 2 (16:00 - 17:00) | |
| Foyer - GRAPP | 87 |
| Complete Paper #202: Efficiency of 3D Fractal Generation Through Raymarching, <i>by Anna Semrau and Dariusz Sawicki</i> | 87 |
| Complete Paper #357: A Virtual Reality Prototype as a Tool Against Verbal Abuse in Classrooms: A Multidisciplinary Approach, <i>by Irene Fondon, Mar Elena, Irene Lagares and Susana Gaytan</i> | 88 |
| Foyer - HUCAPP | 88 |
| Abstract #102: Are Attitudes Towards Robots Universal? Corpus of Speech About Robots (COSAR) Study of Human Attitudes Towards Social Robots, <i>by Aleksandra Wasielewska</i> | 88 |
| Complete Paper #396: From Surveys to TV Series: Study of the Role of Haptic in Social Interaction, <i>by Amine Frihi and Bob-Antoine Ménélas</i> | 88 |
| Foyer - VISAPP | 89 |
| Complete Paper #281: Learning 3D Human UV with Loose Clothing from Monocular Video, <i>by Meng-Yu Kuo, Jingfan Guo and Ryo Kawahara</i> | 89 |
| Complete Paper #305: Deep Learning Model Compression for Resource Efficient Activity Recognition on Edge Devices: A Case Study, <i>by Dieter Balemans, Benjamin Vandersmissen, Jan Steckel, Siegfried Mercelis, Phil Reiter and José Oramas</i> | 89 |
| Complete Paper #382: RecViT: Enhancing Vision Transformer with Top-Down Information Flow, <i>by Štefan Pócoš, Iveta Bečková and Igor Farkaš</i> | 89 |
| Complete Paper #96: Estimation of the Inference Quality of Machine Learning Models for Cutting Tools Inspection, <i>by Kacper Marciniak, Paweł Majewski and Jacek Reiner</i> | 89 |
| Complete Paper #118: Diffusion-Based Image Generation for In-Distribution Data Augmentation in Surface Defect Detection, <i>by Luigi Capogrosso, Federico Girella, Francesco Taioli, Michele Chiara, Muhammad Aqeel, Franco Fummi, Francesco Setti and Marco Cristani</i> | 90 |
| Complete Paper #119: Class Weighted Focal Loss for Improving Class Imbalance in Semi-Supervised Object Detection, <i>by Shinichi Hoketsu, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi</i> | 90 |
| Complete Paper #126: Character Identification in Images Extracted from Portuguese Manuscript Historical Documents, <i>by Gustavo Lacerda and Raimundo Vasconcelos</i> | 90 |
| Complete Paper #127: Identifying Representative Images for Events Description Using Machine Learning, <i>by Marcos Soares de Sousa and Raimundo Vasconcelos</i> | 90 |
| Complete Paper #139: Association of Grad-CAM, LIME and Multidimensional Fractal Techniques for the Classification of H&E Images, <i>by Thales Lopes, Guilherme Roberto, Carlos Soares, Thaína Tosta, Adriano Silva, Adriano Loyola, Sérgio Cardoso, Paulo R. de Faria, Marcelo Z. do Nascimento and Leandro Neves</i> | 91 |
| Complete Paper #146: A Comparative Analysis of the Three-Alternative Forced Choice Method and the Slider-Based Method in Subjective Experiments: A Case Study on Contrast Preference Task, <i>by Olga Cherepkova, Seyed Ali Amirshahi and Marius Pedersen</i> | 91 |
| Complete Paper #149: Comparative Study Between Object Detection Models, for Olive Fruit Fly Identification, <i>by Margarida Victoriano, Lino Oliveira and Hélder Oliveira</i> | 91 |
| Complete Paper #159: Applying Prompts and Parameter-Efficient Methods to Enhance Single-Stream Vision-Language Transformers, <i>by Xuehao Liu, Sarah Jane Delany and Susan McKeever</i> | 91 |
| Complete Paper #251: Coreset Based Medical Image Anomaly Detection and Segmentation, <i>by Ciprian-Mihai Ceaușescu, Bogdan Alexe and Riccardo Volpi</i> | 92 |
| Complete Paper #332: Error Analysis of Aerial Image-Based Relative Object Position Estimation, <i>by Zsombor Páncsics, Nelli Nyisztor, Tekla Tóth, Imre Juhász, Gergely Treplán and Levente Hajder</i> | 92 |
| Complete Paper #355: An Assistive Technology Based on Object Detection for Automated Task List Generation, <i>by Frédéric Rayar</i> | 92 |
| Complete Paper #371: Microplankton Discrimination in FlowCAM Images Using Deep Learning, <i>by Francisco Bonin-Font, Gorka Buenvaron, Mary Kane and Idan Tuval</i> | 92 |
| Complete Paper #376: Depth Estimation Using Weighted-Loss and Transfer Learning, <i>by Muhammad Hafeez, Michael Madden, Ganesh Sistu and Ihsan Ullah</i> | 93 |
| Complete Paper #379: Optical Illusion in Which Line Segments Continue to Grow or Shrink by Displaying Two Images Alternately, <i>by Kazuhisa Yanaka and Sota Mihara</i> | 93 |
| Complete Paper #425: Evaluating the Usability of a 3D Map Visualizer Augmented Reality Application, <i>by Franklin Fracaro, Fabiana Peres, Claudio Mauricio and João Teixeira</i> | 93 |
| Poster Presentations (Online) 2 (16:00 - 17:00) | |
| Room Online 1 - VISIGRAPP | 93 |
| Room Online 1 - IVAPP | 93 |
| Complete Paper #136: Visualizing Plasma Physics Simulations in Immersive Environments, <i>by Nuno Verdelho Trindade, Óscar Amaro, David Brás, Daniel Gonçalves, João Madeiras Pereira and Alfredo Ferreira</i> | 93 |
| Room Online 1 - HUCAPP | 94 |

| | |
|--|----|
| Complete Paper #187: Enhancing the User Interaction of Online Students: Analysis of an Interaction Concept for a Learner Dashboard, <i>by Thorleif Harder and Gilbert Drzyzga</i> | 94 |
| Room Online 1 - VISAPP | 94 |
| Complete Paper #198: Robotics and Computer Vision in the Brazilian Electoral Context: A Case Study, <i>by Jairton Falcão Filho, Matheus Costa, Jonas Silva, Cecília Santos da Silva, Felipe Mendonca, Jefferson Norberto, Riei Rodrigues and Marcondes Silva Júnior</i> | 94 |
| Complete Paper #211: Nearest Neighbor-Based Data Denoising for Deep Metric Learning, <i>by George Galanakis, Xenophon Zabulis and Antonis Argyros</i> | 94 |
| Complete Paper #214: XYZ Unsupervised Network: A Robust Image Dehazing Approach, <i>by Percy Maldonado-Quispe and Helio Pedrini</i> | 95 |
| Complete Paper #386: Strategies for Classifier Selection Based on Genetic Programming for Multimedia Data Recognition, <i>by Rafael Martarelli, Douglas Rodrigues, Clayton Pereira, Jurandy Almeida and João Papa</i> | 95 |

Keynote Lecture (17:00 - 18:00)

| | |
|--|----|
| Room Plenary - VISIGRAPP | 95 |
| The Predictable Side of Unpredictable Humans, <i>by Alvitta Ottley</i> | 95 |

Thursday Sessions: February 29**Session 6 (09:00 - 10:30)**

| | |
|--|-----|
| Room Space 3A - VISAPP: Object Detection and Localization | 99 |
| Complete Paper #156: Non-Local Context-Aware Attention for Object Detection in Remote Sensing Images, <i>by Yassin Terraf, El Mercha and Mohammed Erradi</i> | 99 |
| Complete Paper #61: Lens Flare-Aware Detector in Autonomous Driving, <i>by Shanxing Ma and Jan Aelterman</i> | 99 |
| Complete Paper #277: World-Map Misalignment Detection for Visual Navigation Systems, <i>by Rosario Forte, Michele Mazzamuto, Francesco Ragusa, Giovanni Farinella and Antonino Furnari</i> | 99 |
| Complete Paper #344: Effects of Model Drift on Ship Detection Models, <i>by Namita Agarwal, Anh Vo, Michela Bertolotto, Alan Barnett, Ahmed Khalid and Merry Globin</i> | 99 |
| Room Space 3B - VISAPP: Image Enhancement and Restoration | 100 |
| Complete Paper #243: Efficient and Accurate Hyperspectral Image Demosaicing with Neural Network Architectures, <i>by Eric Wisotzky, Lara Wallburg, Anna Hilsmann, Peter Eisert, Thomas Wittenberg and Stephan Göb</i> | 100 |
| Complete Paper #254: Two Nonlocal Variational Models for Retinex Image Decomposition, <i>by Frank Hammond, Catalina Sbert and Joan Duran</i> | 100 |
| Complete Paper #289: An Image Sharpening Technique Based on Dilated Filters and 2D-DWT Image Fusion, <i>by Victor Bogdan, Cosmin Bonchiş and Ciprian Orhei</i> | 100 |
| Complete Paper #353: Fast and Reliable Inpainting for Real-Time Immersive Video Rendering, <i>by Jakub Stankowski and Adrian Dziembowski</i> | 101 |
| Room Space 3C - VISAPP: Medical Image Applications | 101 |
| Complete Paper #85: Pair-GAN: A Three-Validated Generative Model from Single Pairs of Biomedical and Ground Truth Images, <i>by Clara Brémond-Martin, Huaqian Wu, Cédric Clouchoux and Kévin François-Bouaou</i> | 101 |
| Complete Paper #99: CaRe-CNN: Cascading Refinement CNN for Myocardial Infarct Segmentation with Microvascular Obstructions, <i>by Franz Thaler, Matthias Gsell, Gernot Plank and Martin Urschler</i> | 101 |
| Complete Paper #185: Comparing 3D Shape and Texture Descriptors Towards Tourette's Syndrome Prediction Using Pediatric Magnetic Resonance Imaging, <i>by Murilo Costa de Barros, Kaue Duarte, Chia-Jui Hsu, Wang-Tso Lee and Marco Garcia de Carvalho</i> | 101 |
| Complete Paper #228: Investigation of Deep Neural Network Compression Based on Tucker Decomposition for the Classification of Lesions in Cavity Oral, <i>by Vitor Fernandes, Adriano Silva, Danilo Pereira, Sérgio Cardoso, Paulo R. de Faria, Adriano Loyola, Thaína Tosta, Leandro Neves and Marcelo Z. do Nascimento</i> | 102 |
| Room Space 4A - VISAPP: Features Extraction | 102 |
| Complete Paper #218: Feature Selection for Unsupervised Anomaly Detection and Localization Using Synthetic Defects, <i>by Lars Heckler and Rebecca König</i> | 102 |
| Complete Paper #240: SIDAR: Synthetic Image Dataset for Alignment & Restoration, <i>by Monika Kwiatkowski, Simon Matern and Olaf Hellwich</i> | 102 |
| Complete Paper #35: Machine Learning in Industrial Quality Control of Glass Bottle Prints, <i>by Maximilian Bundscherer, Thomas Schmitt and Tobias Bocklet</i> | 103 |
| Complete Paper #79: Privacy Preservation in Image Classification Using Seam Doppelgänger, <i>by Nishitha Prakash and James Pope</i> | 103 |
| Room Space 4B - GRAPP: Interactive 3D Graphics and Immersive Systems | 103 |
| Complete Paper #377: An Evaluation Research on Dynamic Hit Stop Using Eye Gaze, <i>by Rena Tomizawa and Tomokazu Ishikawa</i> | 103 |
| Complete Paper #401: HandWindowTeleportation: Locomotion with Hand Gestures for Virtual Reality Games, <i>by Hibiki Kirihata and Tomokazu Ishikawa</i> | 103 |

| | |
|---|-----|
| Complete Paper #29: Trajectory Augmentation for Robust Neural Locomotion Controllers, <i>by Dhruv Agrawal, Mathias König, Jakob Buhmann, Robert Sumner and Martin Guay</i> | 104 |
| Room Space 4D - HUCAPP: Multimodal Systems and Application | 104 |
| Complete Paper #435: Emotional and Meditative States in Interactive Media Access with a Positive Computing Perspective, <i>by Carla Bernardino, Sílvia Martins, Hugo Ferreira and Teresa Chambel</i> | 104 |
| Complete Paper #343: AI-Supported Diagnostic of Depression Using Clinical Interviews: A Pilot Study, <i>by Bakir Hadžić, Julia Ohse, Michael Danner, Nicolina Peperkorn, Parvez Mohammed, Youssef Shibani and Matthias Rättsch</i> | 104 |
| Complete Paper #434: Exploring Multimodal Interactions with a Robot Assistant in an Assembly Task: A Human-Centered Design Approach, <i>by Simona D'Attanasio, Théo Alabert, Clara Francis and Anna Studzinska</i> | 104 |
| Oral Presentations (Online) 6 (09:00 - 10:30) | |
| Room Online 2 - VISAPP: Segmentation and Grouping | 105 |
| Complete Paper #337: BEVFastLine: Single Shot Fast BEV Line Detection for Automated Parking Applications, <i>by Praveen Narasappareddygar, Venkatesh Karunamoorthy, Venkatesh Karunamoorthy, Shubham Sonarghare, Ganesh Sistu, Prasad Deshpande and Prasad Deshpande</i> | 105 |
| Complete Paper #54: Estimation of Package-Boundary Confidence for Object Recognition in Rainbow-SKU Depalletizing Automation, <i>by Kento Sekiya, Taiki Yano, Nobutaka Kimura and Kiyoto Ito</i> | 105 |
| Complete Paper #101: SAMMI: Segment Anything Model for Malaria Identification, <i>by Luca Zedda, Andrea Loddo and Cecilia Di Ruberto</i> | 105 |
| Complete Paper #203: Unsupervised Few-Shot Image Segmentation with Dense Feature Learning and Sparse Clustering, <i>by Kuangdai Leng, Robert Atwood, Winfried Kockelmann, Deniza Chekrygina and Jeyan Thiyagalingam</i> | 105 |
| Room Online 1 - VISAPP: Deep Learning for Visual Understanding | 106 |
| Complete Paper #25: Improving Pseudo-Labeling and Enhancing Robustness for Semi-Supervised Domain Generalization, <i>by Adnan Khan, Mai A. Shaaban and Muhammad Haris Khan</i> | 106 |
| Complete Paper #135: Interpretable Anomaly Analysis for Surveillance Video, <i>by Meng Dong</i> | 106 |
| Complete Paper #374: Word and Image Embeddings in Pill Recognition, <i>by Richárd Rádl, Zsolt Vörösházi and László Czúni</i> | 106 |
| Complete Paper #426: Multimodal Crowd Counting with Pix2Pix GANs, <i>by Muhammad Asif Khan, Hamid Menouar and Ridha Hamila</i> | 107 |
| Poster Session 3 (10:30 - 11:30) | |
| Foyer - IVAPP | 107 |
| Abstract #456: Naming Categorical Palette Colors, <i>by Martijn Tennekes and Marco Puts</i> | 107 |
| Complete Paper #398: Navigating the Trade-Off Between Explainability and Privacy, <i>by Johanna Schmidt, Verena Pietsch, Martin Nocker, Michael Rader and Alessio Montuoro</i> | 107 |
| Complete Paper #400: Evaluation of Approximate Reflectional Symmetry, <i>by Martin Maňák, David Podgorelec and Ivana Kolingerová</i> | 107 |
| Complete Paper #454: Visual Analysis of Military Diving Incident Reports, <i>by G. Walsh, N. Andersen, J. Kusnick, E. Sørensen and S. Jänicke</i> | 108 |
| Foyer - VISAPP | 108 |
| Complete Paper #33: Evaluating Learning Potential with Internal States in Deep Neural Networks, <i>by Shogo Takasaki and Shuichi Enokida</i> | 108 |
| Complete Paper #456: Semantic Image Synthesis for Realistic Image Generation in Robotic Assisted Partial Nephrectomy, <i>by Stefano Mazzocchetti, Laura Cercenelli, Lorenzo Bianchi, Riccardo Schiavina and Emanuela Marcelli</i> | 108 |
| Complete Paper #162: Modification of DDIM Encoding for Generating Counterfactual Pathology Images of Malignant Lymphoma, <i>by Ryoichi Koga, Mauricio Kugler, Tatsuya Yokota, Kouichi Ohshima, Hiroaki Miyoshi, Miharu Nagaishi, Noriaki Hashimoto, Ichiro Takeuchi and Hidekata Hontani</i> | 108 |
| Complete Paper #186: Relationship Between Semantic Segmentation Model and Additional Features for 3D Point Clouds Obtained from on-Vehicle LIDAR, <i>by Hisato Hashimoto and Shuichi Enokida</i> | 109 |
| Complete Paper #193: Teacher-Student Models for AI Vision at the Edge: A Car Parking Case Study, <i>by Mbaso Molo, Emanuele Carlini, Luca Ciampi, Claudio Gennaro and Lucia Vadicamo</i> | 109 |
| Complete Paper #199: Most Relevant Viewpoint of an Object: A View-Dependent 3D Saliency Approach, <i>by Marie Pelissier-Combesure, Sylvie Chambon and Geraldine Morin</i> | 109 |
| Complete Paper #221: Combining Total Variation and Nonlocal Variational Models for Low-Light Image Enhancement, <i>by Daniel Torres, Catalina Sbert and Joan Duran</i> | 110 |
| Complete Paper #229: Oral Dysplasia Classification by Using Fractal Representation Images and Convolutional Neural Networks, <i>by Rafael Carvalho, Adriano Silva, Alessandro Martins, Sérgio Cardoso, Guilherme Freire, Paulo R. de Faria, Adriano Loyola, Thaína Tosta, Leandro Neves and Marcelo Z. do Nascimento</i> | 110 |
| Complete Paper #253: Multi-Task Planar Reconstruction with Feature Warping Guidance, <i>by Luan Wei, Anna Hilsmann and Peter Eisert</i> | 110 |

| | |
|--|-----|
| Complete Paper #294: Learning Projection Patterns for Direct-Global Separation, <i>by Takaoki Ueda, Ryo Kawahara and Takahiro Okabe</i> | 110 |
| Complete Paper #306: Combining Progressive Hierarchical Image Encoding and YOLO to Detect Fish in Their Natural Habitat, <i>by Antoni Burguera</i> | 111 |
| Complete Paper #338: Blind Deblurring of THz Time-Domain Images Based on Low-Rank Representation, <i>by Marina Ljubenović, Mário Figueiredo and Arianna Traviglia</i> | 111 |
| Complete Paper #373: Identification of Honeybees with Paint Codes Using Convolutional Neural Networks, <i>by Gabriel Santiago-Plaza, Luke Meyers, Andrea Gomez-Jaime, Rafael Meléndez-Ríos, Fanfan Noel, Jose Agosto, Tugrul Giray, Josué Rodríguez-Cordero and Rémi Mégret</i> | 111 |
| Complete Paper #381: SAM-Based Detection of Structural Anomalies in 3D Models for Preserving Cultural Heritage, <i>by David Jurado-Rodríguez, Alfonso López, J. Jiménez, Antonio Garrido, Francisco Feito and Juan Jurado</i> | 111 |
| Complete Paper #391: A Generative Model for Guided Thermal Image Super-Resolution, <i>by Patricia Suárez and Angel Sappa</i> | 112 |
| Complete Paper #417: Colorectal Image Classification Using Randomized Neural Network Descriptors, <i>by Jarbas Sá Junior and André Backes</i> | 112 |
| Complete Paper #431: Deformable Pose Network: A Multi-Stage Deformable Convolutional Network for 2D Hand Pose Estimation, <i>by Sartaj Salman, Ali Zakir and Hiroki Takahashi</i> | 112 |
| Complete Paper #437: Unsupervised Domain Adaptation for Human Pose Action Recognition, <i>by Mattias Billast, Tom De Schepper and Kevin Mets</i> | 112 |
| Complete Paper #446: Exploring Unsupervised Domain Adaptation Approaches for Water Parameters Estimation from Satellite Images, <i>by Mauren Coelho de Andrade, Anderson Souza, Bruno Oliveira, Maria Starling, Camila Amorim and Jefersson Santos</i> | 113 |
| Complete Paper #449: M&M: Multimodal-Multitask Model Integrating Audiovisual Cues in Cognitive Load Assessment, <i>by Long Nguyen-Phuoc, Renald Gaboriau, Dimitri Delacroix and Laurent Navarro</i> | 113 |
| Poster Presentations (Online) 3 (10:30 - 11:30) | |
| Room Online 2 - VISIGRAPP | 113 |
| Room Online 2 - VISAPP | 113 |
| Complete Paper #114: Enhancing Object Detection Accuracy with Variational Autoencoders as a Filter in YOLO, <i>by Shubham Dubey, J. Satyanarayana and C. Mohan</i> | 113 |
| Complete Paper #152: Body-Part Enabled Wildlife Detection and Tracking in Video Sequences, <i>by Alberto Lee, Kofi Appiah and Sze Kwok</i> | 113 |
| Complete Paper #177: Real-Time Detection and Mapping of Crowd Panic Emergencies, <i>by Ilias Lazarou, Anastasios Kesidis and Andreas Tsatsaris</i> | 114 |
| Complete Paper #236: Automated Brain Lobe Segmentation and Feature Extraction from Multiple Sclerosis Lesions Using Deep Learning, <i>by Nada Haj Messaoud, Rim Ayari, Asma Ben Abdallah and Mohamed Hedi Bedoui</i> | 114 |
| Complete Paper #266: A Mobile-Phone Pose Estimation for Gym-Exercise Form Correction, <i>by Matthew Turner, Kofi Appiah and Sze Kwok</i> | 114 |
| Complete Paper #282: Curriculum for Crowd Counting: Is It Worthy?, <i>by Muhammad Asif Khan, Hamid Menouar and Ridha Hamila</i> | 114 |
| Room Online 1 - VISIGRAPP | 114 |
| Room Online 1 - GRAPP | 115 |
| Complete Paper #72: Towards Generating 3D City Models with GAN and Computer Vision Methods, <i>by Sarun Poolkrajang and Anand Bhojan</i> | 115 |
| Room Online 1 - VISAPP | 115 |
| Complete Paper #216: Direct 3D Body Measurement Estimation from Sparse Landmarks, <i>by David Bojanić, Kristijan Bartol, Tomislav Petković and Tomislav Pribanić</i> | 115 |
| Complete Paper #273: Conic Linear Units: Improved Model Fusion and Rotational-Symmetric Generative Model, <i>by Changqing Fu and Laurent Cohen</i> | 115 |
| Complete Paper #443: CL-FedFR: Curriculum Learning for Federated Face Recognition, <i>by Devilliers Dube, Çiğdem Erdem and Ömer Korçak</i> | 115 |
| Complete Paper #452: RetailKLIP: Finetuning OpenCLIP Backbone Using Metric Learning on a Single GPU for Zero-Shot Retail Product Image Classification, <i>by Muktabh Srivastava</i> | 116 |
| Oral Presentations (Online) 7 (11:30 - 13:00) | |
| Room Online 1 - VISAPP: Image and Video Understanding | 116 |
| Complete Paper #304: Parts-Based Implicit 3D Face Modeling, <i>by Yajie Gu and Nick Pears</i> | 116 |
| Complete Paper #362: Occlusion-Robust and Efficient 6D Pose Estimation with Scene-Level Segmentation Refinement and 3D Partial-to-6D Full Point Cloud Transformation, <i>by Sukhan Lee, Soojin Lee and Yongjun Yang</i> | 116 |
| Room Online 2 - VISAPP: Computer Vision Theory and Applications | 116 |
| Complete Paper #57: Event-Based Semantic-Aided Motion Segmentation, <i>by Chenao Jiang, Julien Moreau and Franck Davoine</i> | 116 |

| | |
|---|-----|
| Complete Paper #104: PanoTherm: Panoramic Thermal Imaging for Object Detection and Tracking, <i>by Thomas Kernbauer, Philipp Fleck and Clemens Arth</i> | 117 |
| Complete Paper #339: Camera Self-Calibration from Two Views with a Common Direction, <i>by Yingna Su, Xinnian Guo and Yang Shen</i> | 117 |
| Session 7 (11:30 - 13:00) | |
| Room Space 3A - VISAPP: Object and Face Recognition | 117 |
| Complete Paper #163: EBA-PRNetCC: An Efficient Bridge Attention-Integration PoseResNet for Coordinate Classification in 2D Human Pose Estimation, <i>by Ali Zakir, Sartaj Salman, Gibran Benitez-Garcia and Hiroki Takahashi</i> | 117 |
| Complete Paper #328: Robust Long-Tailed Image Classification via Adversarial Feature Re-Calibration, <i>by Jinghao Zhang, Zhenhua Feng and Yaochu Jin</i> | 117 |
| Complete Paper #180: Privacy-Preserving Face Recognition in Hybrid Frequency-Color Domain, <i>by Dong Han, Yong Li and Joachim Denzler</i> | 118 |
| Complete Paper #335: Exploring the Impact of Knowledge Graphs on Zero-Shot Visual Object State Classification, <i>by Filippos Gouidis, Konstantinos Papoutsakis, Theodore Patkos, Antonis Argyros and Dimitris Plexousakis</i> | 118 |
| Room Space 3B - VISAPP: Color and Texture Analyses | 118 |
| Complete Paper #38: Generative Texture Super-Resolution via Differential Rendering, <i>by Milena Bagdasarian, Peter Eisert and Anna Hilsmann</i> | 118 |
| Complete Paper #117: CAVC: Cosine Attention Video Colorization, <i>by Leandro Stival, Ricardo Torres and Helio Pedrini</i> | 119 |
| Complete Paper #300: Using Extended Light Sources for Relighting from a Small Number of Images, <i>by Toshiki Hirao, Ryo Kawahara and Takahiro Okabe</i> | 119 |
| Complete Paper #416: Optimization and Learning Rate Influence on Breast Cancer Image Classification, <i>by Gleidson Barbosa, Larissa Moreira, Pedro Moises de Sousa, Rodrigo Moreira and André Backes</i> | 119 |
| Room Space 3C - VISAPP: Deep Learning for Visual Understanding | 119 |
| Complete Paper #40: Uncertainty-Based Detection of Adversarial Attacks in Semantic Segmentation, <i>by Kira Maag and Asja Fischer</i> | 119 |
| Complete Paper #175: Multi-View Inversion for 3D-aware Generative Adversarial Networks, <i>by Florian Barthel, Anna Hilsmann and Peter Eisert</i> | 120 |
| Complete Paper #197: Agrinet: A Hyperspectral Image Based Precise Crop Classification Model, <i>by Aditi Palit, Himanshu Dolekar and Kalidas Yeturu</i> | 120 |
| Complete Paper #354: Minimalist CNN for Medical Imaging Classification with Small Dataset: Does Size Really Matter and How?, <i>by Marie Économidès and Pascal Desbarats</i> | 120 |
| Room Space 4A - VISAPP: Image Formation, Acquisition Devices and Sensors | 120 |
| Complete Paper #65: Investigating Color Illusions from the Perspective of Computational Color Constancy, <i>by Oguzhan Ulucan, Diclehan Ulucan and Marc Ebner</i> | 120 |
| Complete Paper #402: Analysis of Scattering Media by High-Frequency Polarized Light Projection Using Polarizing Projector, <i>by Aigo Ohno, Fumihiko Sakaue and Jun Sato</i> | 121 |
| Complete Paper #403: Probabilistic NeRF for 3D Shape Recovery in Scattered Medium, <i>by Yoshiki Ono, Fumihiko Sakaue and Jun Sato</i> | 121 |
| Complete Paper #411: Dense Light Field Imaging with Mixed Focus Camera, <i>by Masato Hirose, Fumihiko Sakaue and Jun Sato</i> | 121 |
| Room Space 4B - GRAPP: Geometry | 121 |
| Complete Paper #19: Polyline Simplification with Predefined Edge Directions by Mixed Integer Linear Programs, <i>by Steffen Goebbels and Jochen Rethmann</i> | 121 |
| Complete Paper #404: Detection of Local Symmetry Polylines of Polygons Based on Sweeping Paradigm, <i>by Martin Safko, Luka Lukač, Borut Žalik and Ivana Kolingerová</i> | 122 |
| Complete Paper #447: Pseudo-Curvature of Fractal Curves for Geometric Control of Roughness, <i>by Mohamad Janbein, Christian Gentil, Céline Roudet and Clement Poull</i> | 122 |
| Complete Paper #122: A Predictor for Triangle Mesh Compression Working in Tangent Space, <i>by Petr Vaněček, Filip Hácha and Libor Váša</i> | 122 |
| Room Space 4D - IVAPP: Visualization Short Papers | 122 |
| Complete Paper #97: Increasing User Engagement with a Tracking App Through Data Visualizations, <i>by Daniela Nickmann and Victor Oliveira</i> | 122 |
| Complete Paper #263: Visualization of Swedish News Articles: A Design Study, <i>by Kostiantyn Kucher, Nellie Engström, Wilma Axelsson, Berkant Savas and Andreas Kerren</i> | 123 |
| Complete Paper #324: Visualizing Group Structure in Compound Graphs: The Current State, Lessons Learned, and Outstanding Opportunities, <i>by Henry Ehlers, Diana Marin, Hsiang-Yun Wu and Renata Raidou</i> | 123 |
| Complete Paper #346: A Review on Data Terminology in Visual Analytics Tools, <i>by Johanna Schmidt and Milena Vuckovic</i> | 123 |
| Keynote Lecture (14:15 - 15:15) | |
| Room Plenary - VISIGRAPP | 123 |
| Virtual Reality in Mental Health: A Self-Counselling Approach, <i>by Mel Slater</i> | 123 |

Session 8 (15:30 - 17:00)

| | |
|--|-----|
| Room Space 3A - VISAPP: Color and Texture Analyses | 124 |
| Complete Paper #360: Conditional Vector Graphics Generation for Music Cover Images, <i>by Ivan Jarsky, Valeria Efimova, Ilya Bizyaev and Andrey Filchenkov</i> | 124 |
| Complete Paper #340: Neural Style Transfer for Vector Graphics, <i>by Ivan Jarsky, Valeria Efimova, Artyom Chebykin, Viacheslav Shalamov and Andrey Filchenkov</i> | 124 |
| Complete Paper #378: Reliability and Stability of Mean Opinion Score for Image Aesthetic Quality Assessment Obtained Through Crowdsourcing, <i>by Egor Ershov, Artyom Panshin, Ivan Ermakov, Nikola Banić, Alex Savchik and Simone Bianco</i> | 124 |
| Room Space 3B - VISAPP: 3D Deep Learning | 124 |
| Complete Paper #427: Fooling Neural Networks for Motion Forecasting via Adversarial Attacks, <i>by Edgar Medina and Leyong Loh</i> | 124 |
| Complete Paper #137: Analysis of Point Cloud Domain Gap Effects for 3D Object Detection Evaluation, <i>by Aitor Iglesias, Mikel García, Nerea Aranjuelo, Ignacio Arganda-Carreras and Marcos Nieto</i> | 125 |
| Complete Paper #148: Incorporating Temporal Information into 3D Hand Pose Estimation Using Scene Flow, <i>by Niklas Hermes, Alexander Bigalke and Mattias Heinrich</i> | 125 |
| Complete Paper #383: Detecting Anomalous 3D Point Clouds Using Pre-Trained Feature Extractors, <i>by Dario Mantegazza and Alessandro Giusti</i> | 125 |
| Room Space 3C - VISAPP: Domain Adaptation | 125 |
| Complete Paper #110: Reducing Bias in Pre-Trained Models by Tuning While Penalizing Change, <i>by Niklas Penzel, Gideon Stein and Joachim Denzler</i> | 125 |
| Complete Paper #138: Pre-Training and Fine-Tuning Attention Based Encoder Decoder Improves Sea Surface Height Multi-Variate Inpainting, <i>by Théo Archambault, Arthur Filoche, Anastase Charantonis and Dominique Béréziat</i> | 126 |
| Complete Paper #22: Learning End-to-End Deep Learning Based Image Signal Processing Pipeline Using a Few-Shot Domain Adaptation, <i>by Georgy Perevozchikov and Egor Ershov</i> | 126 |
| Complete Paper #142: How Quality Affects Deep Neural Networks in Fine-Grained Image Classification, <i>by Joseph Smith, Zheming Zuo, Jonathan Stonehouse and Boguslaw Obara</i> | 126 |
| Room Space 4A - VISAPP: Segmentation and Grouping | 127 |
| Complete Paper #100: Efficiency Optimization Strategies for Point Transformer Networks, <i>by Jannis Unkrig and Markus Friedrich</i> | 127 |
| Complete Paper #448: FingerSeg: Highly-Efficient Dual-Resolution Architecture for Precise Finger-Level Semantic Segmentation, <i>by Gibran Benitez-Garcia and Hiroki Takahashi</i> | 127 |
| Complete Paper #89: Automated Generation of Instance Segmentation Labels for Traffic Surveillance Models, <i>by D. Scholte, T. T. G. Urselmann, M. H. Zwemer, E. Bondarev and P. H. N. de With</i> | 127 |
| Room Space 4B - GRAPP: Modeling and Natural Environments | 127 |
| Complete Paper #219: Between Gaming and Microclimate Simulations: Temperature Estimation of an Urban Area, <i>by Eva Strauss and Dimitri Bulatov</i> | 127 |
| Complete Paper #231: TreeSpecies-PC2DT: Automated Tree Species Modeling from Point Clouds to Digital Twins, <i>by Like Gobeawan, Xuan Liu, Chi Lim, Venugopalan Raghavan, Joyjit Chatteraj, Jan Schindler and Feng Yang</i> | 128 |
| Complete Paper #249: Parameter-Free Connectivity for Point Clouds, <i>by Diana Marin, Stefan Ohrhallinger and Michael Wimmer</i> | 128 |
| Room Space 4D - HUCAPP: User Performance Evaluation | 128 |
| Complete Paper #18: Sense of Presence, Realism, and Simulation Sickness in Operational Tasks: A Comparative Analysis of Virtual and Mixed Reality, <i>by Giorgio Ballestin and Heike Diepeveen</i> | 128 |
| Complete Paper #438: Perception of a Spatial Implausibility Caused by Seamless Covert Teleportation, <i>by Mathieu Lutfallah, Dylan Pires, Valentina Gorobets and Andreas Kunz</i> | 128 |

Oral Presentations (Online) 8 (15:30 - 17:00)

| | |
|---|-----|
| Room Online 1 - VISAPP: Deep Learning for Visual Understanding | 129 |
| Complete Paper #77: Attention-Based Shape and Gait Representations Learning for Video-Based Cloth-Changing Person Re-Identification, <i>by Vuong Nguyen, Samiha Mirza, Pranav Mantini and Shishir Shah</i> | 129 |
| Complete Paper #82: Multi-Task Learning Based on Log Dynamic Loss Weighting for Sex Classification and Age Estimation on Panoramic Radiographs, <i>by Igor Prado, David Lima, Julian Liang, Ana Hougaz, Bernardo Peters and Luciano Oliveira</i> | 129 |
| Complete Paper #151: Efficient Parameter Mining and Freezing for Continual Object Detection, <i>by Angelo Menezes, Augusto Peterlevitz, Mateus Chinelatto and André de Carvalho</i> | 129 |
| Complete Paper #309: Region-Transformer: Self-Attention Region Based Class-Agnostic Point Cloud Segmentation, <i>by Dipesh Gyawali, Jian Zhang and Bijaya Karki</i> | 129 |

Tutorial (15:30 - 17:00)

| | |
|---|-----|
| Room Plenary | 130 |
| Efficient 3D Facial Animation using Low Cost Motion Capture Solutions, <i>by João Marcelo N. Teixeira</i> | 130 |

Industrial Panel (17:15 - 18:15)

Room Plenary - VISIGRAPP 130

Closing Session & Awards Ceremony (18:15 - 18:45)

Room Plenary - VISIGRAPP 130



Tuesday Sessions: February 27

Tuesday Sessions: February 27 Program Layout

| | Coffee-Break | Foyer | Online 1 | Online 2 | Plenary | Restaurant | Space 1 | Space 3B | Space 3C | Space 4A | Space 4B | Space 4C |
|-------|--------------|----------------------------|---------------------------------|---------------------------------|----------------------------|------------|------------------------------------|-------------------------------------|------------------------------------|-------------------------------------|------------------------------|----------------------------|
| 9:00 | | | | | Opening Session | | | | | | | |
| 9:30 | | | | | Panel | | | | | | | |
| 10:00 | | | | | | | | | | | | |
| 10:30 | Coffee-Break | | | | | | | | | | | |
| 11:00 | | | Oral Presentations (Online) 1 | Oral Presentations (Online) 1 | | | VISIGRAPP Session 1 | VISIGRAPP Session 1 | VISIGRAPP Session 1 | VISIGRAPP Session 1 | VISIGRAPP Session 1 | Tutorial Martin Zimmermann |
| 11:30 | | | | | | | VISAPP #62, #178, #242, #247, #267 | VISAPP #115, #165, #250, #258, #265 | VISAPP #49, #161, #209, #385, #387 | VISAPP #124, #157, #288, #331, #406 | GRAPP #213, #260, #367, #390 | |
| 12:00 | | | | | | | | | | | | |
| 12:30 | | | | | | | | | | | | |
| 13:00 | | | | | | Lunch | | | | | | |
| 13:30 | | | | | | | | | | | | |
| 14:00 | | | | | | | | | | | | |
| 14:30 | | | Oral Presentations (Online) 2 | Oral Presentations (Online) 2 | | | VISIGRAPP Session 2 | VISIGRAPP Session 2 | VISIGRAPP Session 2 | VISIGRAPP Session 2 | VISIGRAPP Session 2 | VISIGRAPP Session 2 |
| 15:00 | | | | | | | VISAPP #32, #44, #153, #279, #413 | VISAPP #45, #121, #270, #399, #408 | VISAPP #182, #261, #299, #342 | VISAPP #95, #125, #222, #370, #405 | GRAPP #75, #143, #256, #334 | IVAPP #120, #194, #455 |
| 15:30 | | | | | | | | | | | | |
| 16:00 | | VISIGRAPP Poster Session 1 | Poster Presentations (Online) 1 | Poster Presentations (Online) 1 | | | | | | | | |
| 16:30 | Coffee-Break | | | | | | | | | | | |
| 17:00 | | | | | Keynote Lecture Fahad Khan | | | | | | | |
| 17:30 | | | | | | | | | | | | |
| 18:00 | | | | | | | | | | | | |

Opening Session
09:00 - 09:15

VISIGRAPP
Room Plenary

Panel
09:15 - 10:15

VISIGRAPP
Room Plenary

Session 1A
10:30 - 12:30
Event and Human Activity Recognition

VISAPP
Room Space 1

Complete Paper #62

S3Aug: Segmentation, Sampling, and Shift for Action Recognition

Taiki Sugiura and Toru Tamaki
Nagoya Institute of Technology, Japan

Keywords: Action Recognition, Data Augmentation, Out-Of-Context, Segmentation, Image Translation.

Abstract: Action recognition is a well-established area of research in computer vision. In this paper, we propose S3Aug, a video data augmentation for action recognition. Unlike conventional video data augmentation methods that involve cutting and pasting regions from two videos, the proposed method generates new videos from a single training video through segmentation and label-to-image transformation. Furthermore, the proposed method modifies certain categories of label images by sampling to generate a variety of videos, and shifts intermediate features to enhance the temporal coherency between frames of the generate videos. Experimental results on the UCF101, HMDB51, and Mimetics datasets demonstrate the effectiveness of the proposed method, particularly for out-of-context videos of the Mimetics dataset.

Complete Paper #178

Mediapi-RGB: Enabling Technological Breakthroughs in French Sign Language (LSF) Research Through an Extensive Video-Text Corpus

Yanis Ouakrim^{1,2}, Hannah Bull¹, Michèle Gouiffès¹, Denis Beautemps², Thomas Hueber² and Annelies Braffort¹

¹ LISN, Univ. Paris-Saclay, CNRS, 91405 Orsay, France

² Univ. Grenoble Alpes, GIPSA-Lab, CNRS, F-38000 Grenoble, France

Keywords: Sign Language Processing, Sign Language Translation, Sign Language Corpora, French Sign Language, LSF.

Abstract: We introduce Mediapi-RGB, a new dataset of French Sign Language (LSF) along with the first LSF-to-French machine translation model. With 86 hours of video, it the largest LSF corpora with translation. The corpus consists of original content in French Sign Language produced by deaf journalists, and has subtitles in written French aligned to the signing. The current release of Mediapi-RGB is available at the Ortolang corpus repository (<https://www.ortolang.fr/workspaces/mediapi-rgb>), and can be used for academic research purposes. The test and validation sets contain 13 and 7 hours of video respectively. The training set contains 66 hours of video that will be released progressively until December 2024. Additionally, the current release contains skeleton keypoints, sign temporal segmentation, spatio-temporal features and subtitles for all the videos in the train, validation

and test sets, as well as a suggested vocabulary of nouns for evaluation purposes. In addition, we present the results obtained on this corpus with the first LSF-to-French translation baseline to give an overview of the possibilities offered by this corpus of unprecedented caliber for LSF. Finally, we suggest potential technological and linguistic applications for this new video-text dataset.

Complete Paper #267

Calisthenics Skills Temporal Video Segmentation

Antonio Finocchiaro, Giovanni Farinella and Antonino Furnari

Department of Mathematics and Computer Science, University of Catania, Italy

Keywords: Pose Recognition, Sports Video Analysis, Temporal Video Segmentation.

Abstract: Calisthenics is a fast-growing bodyweight discipline that consists of different categories, one of which is focused on skills. Skills in calisthenics encompass both static and dynamic elements performed by athletes. The evaluation of static skills is based on their difficulty level and the duration of the hold. Automated tools able to recognize isometric skills from a video by segmenting them to estimate their duration would be desirable to assist athletes in their training and judges during competitions. Although the video understanding literature on action recognition through body pose analysis is rich, no previous work has specifically addressed the problem of calisthenics skill temporal video segmentation. This study aims to provide an initial step towards the implementation of automated tools within the field of Calisthenics. To advance knowledge in this context, we propose a dataset of video footage of static calisthenics skills performed by athletes. Each video is annotated with a temporal segmentation which determines the extent of each skill. We hence report the results of a baseline approach to address the problem of skill temporal segmentation on the proposed dataset. The results highlight the feasibility of the proposed problem, while there is still room for improvement.

Complete Paper #242

AR-VPT: Simple Auto-Regressive Prompts for Adapting Frozen ViTs to Videos

Muhammad Yousuf¹, Syed Wasim², Syed Hasany³ and Muhammad Farhan¹

¹ Habib University, Pakistan

² MBZUAI, U.A.E.

³ Université de Rouen, France

Keywords: Computer Vision, Frozen ViT, Temporal Modelling, Video Recognition, Prompt Tuning.

Abstract: The rapid progress of deep learning in image recognition has driven increasing interest in video recognition. While image recognition has benefited from the abundance of pre-trained models, video recognition remains challenging due to the absence of strong pre-trained models and the computational cost of training from scratch. Transfer learning techniques have been used to leverage pre-trained networks for video recognition by extracting features from individual frames and combining them for decision-making. In this paper, we explore the use of Visual-Prompt Tuning (VPT) for video recognition, a computationally efficient technique previously proposed for image recognition. Our contributions are two-fold: we introduce Auto-Regressive Visual

Prompt Tuning (AR-VPT) method to perform temporal modeling, addressing the weakness of VPT in this aspect. Finally, we achieve significantly improved performance compared to vanilla VPT on three benchmark datasets: UCF-101, Diving-48, and Something-Something-v2. Our proposed method achieves an optimal trade-off between performance and computation cost, making it a promising approach for video recognition tasks.

Complete Paper #247

Benchmarking Sensitivity of Continual Graph Learning for Skeleton-Based Action Recognition

Wei Wei¹, Tom De Schepper¹ and Kevin Mets²

¹ University of Antwerp - imec, IDLab, Department of Computer Science, Sint-Pietersvliet 7, 2000 Antwerp, Belgium

² University of Antwerp - imec, IDLab, Faculty of Applied Engineering, Sint-Pietersvliet 7, 2000 Antwerp, Belgium

Keywords: Continual Graph Learning, Action Recognition, Spatio-Temporal Graph, Sensitivity Analysis, Benchmark.

Abstract: Continual learning (CL) is the research field that aims to build machine learning models that can accumulate knowledge continuously over different tasks without retraining from scratch. Previous studies have shown that pre-training graph neural networks (GNN) may lead to negative transfer (Hu et al., 2020) after fine-tuning, a setting which is closely related to CL. Thus, we focus on studying GNN in the continual graph learning (CGL) setting. We propose the first continual graph learning benchmark for spatio-temporal graphs and use it to benchmark well-known CGL methods in this novel setting. The benchmark is based on the N-UCLA and NTU-RGB+D datasets for skeleton-based action recognition. Beyond benchmarking for standard performance metrics, we study the class and task-order sensitivity of CGL methods, i.e., the impact of learning order on each class/task's performance, and the architectural sensitivity of CGL methods with backbone GNN at various widths and depths. We reveal that task-order robust methods can still be class-order sensitive and observe results that contradict previous empirical observations on architectural sensitivity in CL.

Session 1B
10:30 - 12:30

VISAPP
Room Space 3B

Image Enhancement and Restoration

Complete Paper #115

Simple Base Frame Guided Residual Network for RAW Burst Image Super-Resolution

Anderson Cotrim¹, Gerson Barbosa^{2,3}, Cid Santos³ and Helio Pedrini¹

¹ Institute of Computing, University of Campinas, Campinas, SP, 13083-852, Brazil

² Eldorado Research Institute, Campinas, SP, 13083-898, Brazil

³ São Paulo State University, Guaratinguetá, SP, 12516-410, Brazil

Keywords: Super-Resolution, Deep Learning, RAW Image, Multi-Frame, Burst.

Abstract: Burst super-resolution or multi-frame super-resolution (MFSR) has gained significant attention in recent years, particularly in the context of mobile photography. With modern handheld devices consistently increasing their processing power and the ability to capture multiple images even faster, the development of robust MFSR algorithms has become increasingly feasible. Furthermore, in contrast to extensively studied single-image

super-resolution (SISR), burst super-resolution mitigates the ill-posed nature of reconstructing high-resolution images from low-resolution ones by merging information from multiple shifted frames. This research introduces a novel and effective deep learning approach, SBFburst, designed to tackle this challenging problem. Our network takes multiple noisy RAW images as input and generates a denoised, super-resolved RGB image as output. We demonstrate that significant enhancements can be achieved in this problem by incorporating base frame-guided mechanisms through operations such as feature map concatenation and skip connections. Additionally, we highlight the significance of employing mosaicked convolution to enhance alignment, thus enhancing the overall network performance in super-resolution tasks. These relatively simple improvements underscore the competitiveness of our proposed method when compared to other state-of-the-art approaches.

Complete Paper #165

Training Methods for Regularizing Gradients on Multi-Task Image Restoration Problems

Samuel Willingham^{1,2}, Mårten Sjöström² and Christine Guillemot¹

¹ Inria Rennes, Rennes, France

² Mid Sweden University, Sundsvall, Sweden

Keywords: Inverse Problems, Computer Vision, Image Restoration, Deep Equilibrium Models, Deep Priors.

Abstract: Inverse problems refer to the task of reconstructing a clean signal from a degraded observation. In imaging, this pertains to restoration problems like denoising, super-resolution or in-painting. Because inverse problems are often ill-posed, regularization based on prior information is needed. Plug-and-play (pnp) approaches take a general approach to regularization and plug a deep denoiser into an iterative solver for inverse problems. However, considering the inverse problems at hand in training could improve reconstruction performance at test-time. Deep equilibrium models allow for the training of multi-task priors on the reconstruction error via an estimate of the iterative method's fixed-point (FP). This paper investigates the intersection of pnp and DEQ models for the training of a regularizing gradient (RG) and derives an upper bound for the reconstruction loss of a gradient-descent (GD) procedure. Based on this upper bound, two procedures for the training of RGs are proposed and compared: One optimizes the upper bound directly, the other trains a deep equilibrium GD (DEQGD) procedure and uses the bound for regularization. The resulting regularized RG (RERG) produces consistently good reconstructions across different inverse problems, while the other RGs tend to have some inverse problems on which they provide inferior reconstructions.

Complete Paper #250

Beyond Variational Models and Self-Similarity in Super-Resolution: Unfolding Models and Multi-Head Attention

Ivan Pereira-Sánchez^{1,2}, Eloi Sans¹, Julia Navarro^{1,2} and Joan Duran^{1,2}

¹ Departament de Ciències Matemàtiques i Informàtica, Universitat de les Illes Balears (UIB), Spain

² Institute of Applied Computing and Community Code (IAC3), Spain

Keywords: Unfolding, Multi-Head Attention, Image Super-Resolution, Variational Methods, Nonlocal Regularization, Self-Similarity.

Abstract: Classical variational methods for solving image processing problems are more interpretable and flexible than pure deep learning approaches, but their performance is limited by the use of rigid priors. Deep unfolding networks combine the strengths of both by unfolding the steps of the optimization algorithm used to estimate the minimizer of an energy functional into a deep learning framework. In this paper, we propose an unfolding approach to extend a variational model exploiting self-similarity of natural images in the data fidelity term for single-image super-resolution. The proximal, downsampling and upsampling operators are written in terms of a neural network specifically designed for each purpose. Moreover, we include a new multi-head attention module to replace the nonlocal term in the original formulation. A comprehensive evaluation covering a wide range of sampling factors and noise realizations proves the benefits of the proposed unfolding techniques. The model shows to better preserve image geometry while being robust to noise.

Complete Paper #258

Avoiding Undesirable Solutions of Deep Blind Image Deconvolution

Antonie Brožová^{1,2} and Václav Šmíd²

¹ Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic

² Institute of Information Theory and Automation, Czech Academy of Sciences, Czech Republic

Keywords: Blind Image Deconvolution, Deep Image Prior, No-Blur, Variational Bayes.

Abstract: Blind image deconvolution (BID) is a severely ill-posed optimization problem requiring additional information, typically in the form of regularization. Deep image prior (DIP) promises to model a naturally looking image due to a well-chosen structure of a neural network. The use of DIP in BID results in a significant performance improvement in terms of average PSNR. In this contribution, we offer qualitative analysis of selected DIP-based methods w.r.t. two types of undesired solutions: blurred image (no-blur) and a visually corrupted image (solution with artifacts). We perform a sensitivity study showing which aspects of the DIP-based algorithms help to avoid which undesired mode. We confirm that the no-blur can be avoided using either sharp image prior or tuning of the hyperparameters of the optimizer. The artifact solution is a harder problem since variations that suppress the artifacts often suppress good solutions as well. Switching to the structural similarity index measure from L₂ norm in loss was found to be the most successful approach to mitigate the artifacts.

Complete Paper #265

SWViT-RRDB: Shifted Window Vision Transformer Integrating Residual in Residual Dense Block for Remote Sensing Super-Resolution

Mohamed Ibrahim^{1,2}, Robert Benavente², Daniel Ponsa² and Felipe Lumbreras²

¹ Computer Engineering Department, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt

² Computer Vision Center & Computer Science Department, Universitat Autònoma de Barcelona, Barcelona, Spain

Keywords: Computer Vision, Super-Resolution, Remote Sensing, Deep Learning.

Abstract: Remote sensing applications, impacted by acquisition

season and sensor variety, require high-resolution images. Transformer-based models improve satellite image super-resolution but are less effective than convolutional neural networks (CNNs) at extracting local details, crucial for image clarity. This paper introduces SWViT-RRDB, a new deep learning model for satellite imagery super-resolution. The SWViT-RRDB, combining transformer with convolution and attention blocks, overcomes the limitations of existing models by better representing small objects in satellite images. In this model, a pipeline of residual fusion group (RFG) blocks is used to combine the multi-headed self-attention (MSA) with residual in residual dense block (RRDB). This combines global and local image data for better super-resolution. Additionally, an overlapping cross-attention block (OCAB) is used to enhance fusion and allow interaction between neighboring pixels to maintain long-range pixel dependencies across the image. The SWViT-RRDB model and its larger variants outperform state-of-the-art (SoTA) models on two different satellite datasets in terms of PSNR and SSIM.

Session 1C

10:30 - 12:30

Visual Attention and Image Saliency

VISAPP

Room Space 3C

Complete Paper #209

When Medical Imaging Met Self-Attention: A Love Story That Didn't Quite Work out

Tristan Piater, Niklas Penzel, Gideon Stein and Joachim Denzler

Computer Vision Group, Friedrich Schiller University, Jena, Germany

Keywords: Self-Attention Mechanisms, Feature Analysis, Medical Imaging.

Abstract: A substantial body of research has focused on developing systems that assist medical professionals during labor-intensive early screening processes, many based on convolutional deep-learning architectures. Recently, multiple studies explored the application of so-called self-attention mechanisms in the vision domain. These studies often report empirical improvements over fully convolutional approaches on various datasets and tasks. To evaluate this trend for medical imaging, we extend two widely adopted convolutional architectures with different self-attention variants on two different medical datasets. With this, we aim to specifically evaluate the possible advantages of additional self-attention. We compare our models with similarly sized convolutional and attention-based baselines and evaluate performance gains statistically. Additionally, we investigate how including such layers changes the features learned by these models during the training. Following a hyperparameter search, and contrary to our expectations, we observe no significant improvement in balanced accuracy over fully convolutional models. We also find that important features, such as dermoscopic structures in skin lesion images, are still not learned by employing self-attention. Finally, analyzing local explanations, we confirm biased feature usage. We conclude that merely incorporating attention is insufficient to surpass the performance of existing fully convolutional methods.

Complete Paper #387

Frames Preprocessing Methods for Chromakey Classification in Video

Evgeny Bessonitsyn, Artyom Chebykin, Grigorii Stafeev and Valeria Efimova

ITMO University, Kronverksky Pr. 49, bldg. A, St. Petersburg, 197101, Russia

Keywords: Chromakey, Transformer, Video Analysis, Classification, Deep Learning, Image Processing.

Abstract: Currently, video games, movies, commercials, and television shows are ubiquitous in modern society. However, beneath the surface of their visual variety lies sophisticated technology, which can produce impressive effects. One such technology is chromakey — a method that allows to change the background to any other image or video. Recognizing chromakey technology in video plays a key role in finding fake materials. In this paper, we consider approaches based on deep learning models that allows to recognize chromakey in video based on unnatural artifacts that arise during the transition between frames. The video consists of a sequence of frames, and the the video accuracy can be determined in different ways. If we consider the accuracy frame by frame, our method reaches an F1 score equal to 0:67. If we consider the entire video to be fake in case there is one or more fake segments, then the F1 score equal to 0:76. The proposed methods showed better results on the dataset we collected in comparison with existing methods for chromakey detection.

Complete Paper #49

Iterative Saliency Enhancement over Superpixel Similarity

Leonardo Joao^{1,2} and Alexandre Falcao¹

¹ *Institute of Computing, State University of Campinas, Campinas, 13083-872, São Paulo, Brazil*

² *LIGM, Univ. Gustave-Eiffel, Marne-la-Vallée, F-77454, France*

Keywords: Salient Object Detection, Saliency Enhancement, Deep-Learning, Superpixel-Based Saliency, Iterative Saliency.

Abstract: Saliency Object Detection (SOD) has several applications in image analysis. The methods have evolved from image-intrinsic to object-inspired (deep-learning-based) models. However, when a model fails, there is no alternative to enhance its saliency map. We fill this gap by introducing a hybrid approach, the Iterative Saliency Enhancement over Superpixel Similarity (ISESS), that iteratively generates enhanced saliency maps by executing two operations alternately: object-based superpixel segmentation and superpixel-based saliency estimation - cycling operations never exploited. ISESS estimates seeds for superpixel delineation from a given saliency map and defines superpixel queries in the foreground and background. A new saliency map results from color similarities between queries and superpixels at each iteration. The process repeats, and, after a given number of iterations, the generated saliency maps are combined into one by cellular automata. Finally, the resulting map is merged with the initial one by the maximum between their average values per superpixel. We demonstrate that our hybrid model consistently outperforms three state-of-the-art deep-learning-based methods on five image datasets.

Complete Paper #161

Comprehensive Evaluation of End-to-End Driving Model Explanations for Autonomous Vehicles

Chenkai Zhang, Daisuke Deguchi, Jialei Chen and Hiroshi Murase

Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

Keywords: Autonomous Driving, Convolutional Neural Network, End-to-End Model, Explainability.

Abstract: Deep learning technology has rapidly advanced, leading to the development of End-to-End driving models (E2EDMs) for autonomous vehicles with high prediction accuracy. To comprehend the prediction results of these E2EDMs, one of the most representative explanation methods is attribution-based. There are two kinds of attribution-based explanation methods: pixel-level and object-level. Usually, the heatmaps illustrate the importance of pixels and objects in the prediction results, serving as explanations for E2EDMs. Since there are many attribution-based explanation methods, evaluation methods are proposed to determine which one is better at improving the explainability of E2EDMs. Fidelity measures the explanation's faithfulness to the model's prediction method, which is a bottommost property. However, no evaluation method could measure the fidelity difference between object-level and pixel-level explanations, making the current evaluation incomplete. In addition, without considering fidelity, previous evaluation methods may advertise manipulative explanations that solely seek human satisfaction (persuasibility). Therefore, we propose an evaluation method that further considers fidelity, our method enables a comprehensive evaluation that proves the object-level explanations genuinely outperform pixel-level explanations in fidelity and persuasibility, thus could better improve the explainability of the E2EDMs.

Complete Paper #385

A Learning Paradigm for Interpretable Gradients

Felipe Figueroa¹, Hanwei Zhang¹, Ronan Sicre¹, Yannic Avrithis² and Stephane Ayache¹

¹ *Centrale Marseille, Aix Marseille Univ., CNRS, LIS, Marseille, France*

² *Institute of Advanced Research on Artificial Intelligence (IARAI), Austria*

Keywords: Gradient, Guided Backpropagation, Class Activation Maps, Interpretability.

Abstract: This paper studies interpretability of convolutional networks by means of saliency maps. Most approaches based on Class Activation Maps (CAM) combine information from fully connected layers and gradient through variants of backpropagation. However, it is well understood that gradients are noisy and alternatives like guided backpropagation have been proposed to obtain better visualization at inference. In this work, we present a novel training approach to improve the quality of gradients for interpretability. In particular, we introduce a regularization loss such that the gradient with respect to the input image obtained by standard backpropagation is similar to the gradient obtained by guided backpropagation. We find that the resulting gradient is qualitatively less noisy and improves quantitatively the interpretability properties of different networks, using several interpretability methods.

Session 1D
10:30 - 12:30
Deep Learning for Visual Understanding

VISAPP
Room Space 4A

Complete Paper #124

Pixel-Wise Gradient Uncertainty for Convolutional Neural Networks Applied to Out-of-Distribution Segmentation

Kira Maag and Tobias Riedlinger
Technical University of Berlin, Germany

Keywords: Deep Learning, Semantic Segmentation, Gradient Uncertainty, Out-of-Distribution Detection.

Abstract: In recent years, deep neural networks have defined the state-of-the-art in semantic segmentation where their predictions are constrained to a predefined set of semantic classes. They are to be deployed in applications such as automated driving, although their categorically confined expressive power runs contrary to such open world scenarios. Thus, the detection and segmentation of objects from outside their predefined semantic space, i.e., out-of-distribution (OoD) objects, is of highest interest. Since uncertainty estimation methods like softmax entropy or Bayesian models are sensitive to erroneous predictions, these methods are a natural baseline for OoD detection. Here, we present a method for obtaining uncertainty scores from pixel-wise loss gradients which can be computed efficiently during inference. Our approach is simple to implement for a large class of models, does not require any additional training or auxiliary data and can be readily used on pre-trained segmentation models. Our experiments show the ability of our method to identify wrong pixel classifications and to estimate prediction quality at negligible computational overhead. In particular, we observe superior performance in terms of OoD segmentation to comparable baselines on the SegmentMelfYouCan benchmark, clearly outperforming other methods.

Complete Paper #288

Detecting Anomalies in Textured Images Using Modified Transformer Masked Autoencoder

Afshin Dini and Esa Rahtu
Unit of Computing Sciences, Tampere University, Finland

Keywords: Anomaly Detection, Anomaly Localization, Masked Autoencoders.

Abstract: We present a new method for detecting and locating anomalies in textured-type images using transformer-based autoencoders. In this approach, a rectangular patch of an image is masked by setting its value to gray and then fetched into a pre-trained autoencoder with several blocks of transformer encoders and decoders in order to reconstruct the unknown part. It is shown that the pre-trained model is not able to reconstruct the defective parts properly when they are inside the masked patch. In this regard, the combination of the Structural Similarity Index Measure and absolute error between the reconstructed image and the original one can be used to define a new anomaly map to find and locate anomalies. In the experiment with the textured images of the MVTec dataset, we discover that not only can this approach find anomalous samples properly, but also the anomaly map itself can specify the exact locations of defects correctly at the same time. Moreover, not only is our method computationally efficient, as it utilizes a pre-trained model and does not require any training, but also it has a better performance compared to previous

autoencoders and other reconstruction-based methods. Due to these reasons, one can use this method as a base approach to find and locate irregularities in real-world applications.

Complete Paper #331

Alias-Free GAN for 3D-Aware Image Generation

Attila Szabó¹, Yevgeniy Puzikov², Sahan Ayyaz¹, Sonia Aurelio², Peter Gehler², Reza Shirvany² and Malte Alf¹
¹ Zalando SE, Switzerland
² Zalando SE, Germany

Keywords: GAN, NeRF, 3D-aware, Generative AI.

Abstract: In this work we build a 3D-aware generative model that produces high quality results with fast inference times. A 3D-aware model generates images and offers control over camera parameters to the user, so that an object can be shown from different viewpoints. The model we build combines the best of two worlds in a very direct way: alias-free Generative Adversarial Networks (GAN) and a Neural Radiance Field (NeRF) rendering, followed by image super-resolution. We show that fast and high-quality image synthesis is possible with careful modifications of the well designed architecture of StyleGAN3. Our design overcomes the problem of viewpoint inconsistency and aliasing artefacts that a direct application of lower-resolution NeRF would exhibit. We show experimental evaluation on two standard benchmark datasets, FFHQ and AFHQv2 and achieve the best or competitive performance on both. Our method does not sacrifice speed, we can render images at megapixel resolution at interactive frame rates.

Complete Paper #157

GAF-Net: Video-Based Person Re-Identification via Appearance and Gait Recognitions

Moncef Boujou¹, Rabah Iguernaissi¹, Lionel Nicod², Djamel Merad¹ and Séverine Dubuisson¹
¹ LIS, CNRS, Aix-Marseille University, Marseille, France
² CERGAM, Aix-Marseille University, Marseille, France

Keywords: Deep Learning, Computer Vision, Person Re-Identification, Gait Recognition, Representation Learning.

Abstract: Video-based person re-identification (Re-ID) is a challenging task aiming to match individuals across various cameras based on video sequences. While most existing Re-ID techniques focus solely on appearance information, including gait information, could potentially improve person Re-ID systems. In this study, we propose, GAF-Net, a novel approach that integrates appearance with gait features for re-identifying individuals; the appearance features are extracted from RGB tracklets while the gait features are extracted from skeletal pose estimation. These features are then combined into a single feature allowing the re-identification of individuals. Our numerical experiments on the iLIDS-Vid dataset demonstrate the efficacy of skeletal gait features in enhancing the performance of person Re-ID systems. Moreover, by incorporating the state-of-the-art PiT network within the GAF-Net framework, we improve both rank-1 and rank-5 accuracy by 1 percentage point.

Complete Paper #406

Enhanced Deepfake Detection Using Frequency Domain Upsampling

Mamadou Bah and Mohamed Dahmane
Computer Research Institute of Montreal (CRIM), Canada

Keywords: DeepFake Detection, Media Forensics, Fourier Upsampling.

Abstract: Recent advances in deep learning have dramatically reshaped the image processing landscape, specifically targeting the key challenge of detecting deepfakes in digital media. This study investigates the integration of Fourier or frequency domain upsampling techniques with deep learning models for the purpose of detecting deepfakes. Using the FF++ dataset as a benchmark for evaluating deepfake detection, our research rigorously evaluates the effectiveness of models such as Xception. This evaluation includes an in-depth exploration of various upsampling methods as well as the combination of spatial and frequency domain upsampling. The results reveal clear disparities in performance between the different models and techniques, and our experiments highlight the profound impact of the various upsampling and downsampling approaches on the accuracy of the resulting classification. Remarkably, combining the Xception model with upsampling and downsampling techniques increases detection accuracy by over 2%, while maintaining the constant input size inherent in the Xception architecture.

Session 1A
10:30 - 12:30
Visualization and Exploration

GRAPP
Room Space 4B

Complete Paper #367

Thinking on Your Feet: Enhancing Foveated Rendering in Virtual Reality During User Activity

David Petrescu¹, Paul Warren², Zahra Montazeri¹,
Gabriel Strain¹ and Steve Pettifer¹

¹ University of Manchester, Human-Computer Systems, U.K.

² University of Manchester, Virtual Reality Research (VR2) Facility, Division of Psychology, Communication and Human Neuroscience, U.K.

Keywords: Foveated Rendering, Variable Rate Shading, Psychophysics, Movement, Attention, Virtual Reality.

Abstract: As prices fall, VR technology is experiencing renewed levels of consumer interest. Despite wider access, VR still requires levels of computational ability and bandwidth that often cannot be achieved with consumer-grade equipment. Foveated rendering represents one of the most promising methods for the optimization of VR content while keeping the quality of the user's experience intact. The user's ability to explore and move through the environment with 6DOF separates VR from traditional display technologies. In this work, we explore if the type of movement (Active versus Implied) and attentional task type (Simple Fixations versus Fixation, Discrimination, and Counting) affect the extent to which a dynamic foveated rendering method using Variable Rate Shading (VRS) optimizes a VR scene. Using psychophysics methods we conduct user studies and recover the Maximum Tolerated Diameter (MTD) at which users fail to notice drops in quality. We find that during self-movement, performing a task that requires more attention masks severe shading reductions and that only 31.7% of the headset's FOV is required to be rendered at the native pixel sampling rate.

Complete Paper #213

Viewpoint Selection for Molecular Visualization: Analysis and Applications

Vincent Larroque¹, Maxime Maria¹, Stéphane Mérillou¹
and Matthieu Montes^{2,3}

¹ Laboratoire XLIM, UMR CNRS 7252, Université de Limoges, France

² GBCM - Laboratoire Génomique, Bioinformatique et Chimie Moléculaire, EA7528, Conservatoire National des Arts et Métiers, Hésam Université, France

³ Institut Universitaire de France (IUF), France

Keywords: Molecular Visualization, Viewpoint Selection, Virtual Tours.

Abstract: Molecular systems are often visually complex and abstract. It can be difficult, even for experts, to explore and find meaningful viewpoints. Research in viewpoint selection methods has mainly focused on general real-world objects and only a few works tackled molecular scenes. In this paper, we present a study of 20 state-of-the-art viewpoint selection methods from the general field, applied to molecular visualization. Our goal is to determine if these methods can find two important geometrical configurations for molecules. Additionally, we propose an automatic generation of informative and visually pleasing molecular tours to help the study and the communication around molecules.

Complete Paper #390

Exploring the Effect of Display Type on Co-Located Multiple Player Gameplay Performance, Immersion, Social Presence, and Behavior Patterns

Wenge Xu¹, Ruichen Zheng², Diego Monteiro³,
Vijayakumar Nanjappan⁴, Yihong Wang² and Hai-Ning Liang²

¹ DMT Lab, Birmingham City University, Birmingham, U.K.

² Department of Computing, School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China

³ LII Lab, ESIEA, Laval, France

⁴ Center for Ubiquitous Computing, University of Oulu, Oulu, Finland

Keywords: Immersion, Social Presence, Head-Mounted Displays, Large Displays, Small Tablet Displays, Co-located Social Play.

Abstract: With advances in virtual reality (VR) technology, immersive head-mounted displays (HMDs) have become widely accessible. These devices have made social games and platforms like VRChat popular. Although the literature points to several factors that affect immersion and social presence, there has been no study that has explored the effect of social display setup on immersion and gameplay in multi-player social games. This work aims to shed light on this issue and investigates the effect of social display setup on gameplay performance and experience (i.e., immersion and social presence) in a multi-player competitive social game (i.e., Jenga). We conducted a one-way between-subjects experiment with 24 participants equally distributed in three groups (4 pairs of 2 participants in each group, who were all strangers to each other) according to three social display setups (2 small-screen tablets, 1 shared 40-inch large TV, and 2 VR HMDs). Our results indicate that (1) players gave a lower rating to challenge in the VR-based social setting than in the small-screen tablet display setting, and (2) gameplay behavior patterns are different among these social display setups.

Complete Paper #260

Handing Pedagogical Scenarios Back over to Domain Experts: A Scenario Authoring Model for VR with Pedagogical Objectives

Mathieu Risy, Valérie Gouranton and Bruno Arnaldi
Univ. Rennes, INSA Rennes, Inria, CNRS, IRISA, France

Keywords: Virtual Reality, Scenario, Pedagogy, Authoring, Welding Use Case.

Abstract: Teachers and trainers make pedagogical decisions for their training courses, so why not do the same for Virtual Reality (VR) training courses? Virtual Environments for Training (VETs) are becoming prominent educational tools. However, VET models have yet to propose scenario authoring aligned with pedagogical objectives that can account for the diversity of approaches available to teachers. This paper proposes a scenario authoring model for VET that directly involves domain experts and validates their pedagogical objectives. In addition, it proposes the coexistence of multiple pedagogical scenarios within the same VET, using three types of scenarios. The validity of the model is then discussed using a VR welding application as a use case.

Oral Presentations (Online) 1
10:30 - 12:30
Machine Learning Technologies for Vision

VISAPP
Room Online 1

Complete Paper #111

Important Pixels Sampling for NeRF Training Based on Edge Values and Squared Errors Between the Ground Truth and the Estimated Colors

Kohei Fukuda, Takio Kurita and Hiroaki Aizawa
Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-Hiroshima, Japan

Keywords: Neural Radiance Fields, Novel View Synthesis.

Abstract: Neural Radiance Fields (NeRF) has impacted computer graphics and computer vision by enabling fine 3D representations using neural networks. However, depending on the data (especially on synthetic datasets with single-color backgrounds), the neural network training of NeRF is often unstable, and the rendering results become poor. This paper proposes a method to sample the informative pixels to remedy these shortcomings. The sampling method consists of two phases. In the early stage of learning (up to 1/10 of all iterations), the sampling probability is determined based on the edge strength obtained by edge detection. Also, we use the squared errors between the ground truth and the estimated color of the pixels for sampling. The introduction of these tweaks improves the learning of NeRF. In the experiment, we confirmed the effectiveness of the method. In particular, for small amounts of data, the training process of the neural network for NeRF was accelerated and stabilized.

Complete Paper #67

Attacking the Loop: Adversarial Attacks on Graph-Based Loop Closure Detection

Jonathan Kim^{1,2}, Martin Urschler^{1,3}, Patricia Riddle¹ and Jörg Wicker¹

¹ School of Computer Science, University of Auckland, New Zealand

² Callaghan Innovation, Auckland, New Zealand

³ Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Austria

Keywords: Visual SLAM, Machine Learning, Adversarial Attacks, Graph Neural Networks, Loop Closure Detection.

Abstract: With the advancement in robotics, it is becoming increasingly common for large factories and warehouses to incorporate visual SLAM (vSLAM) enabled automated robots that operate closely next to humans. This makes any adversarial attacks on vSLAM components potentially detrimental to humans working alongside them. Loop Closure Detection (LCD) is a crucial component in vSLAM that minimizes the accumulation of drift in mapping, since even a small drift can accumulate into a significant drift over time. A prior work by Kim *et al.*, SymbioLCD2, unified visual features and semantic objects into a single graph structure for finding loop closure candidates. While this provided a performance improvement over visual feature-based LCD, it also created a single point of vulnerability for potential graph-based adversarial attacks. Unlike previously reported visual-patch based attacks, small graph perturbations are far more challenging to detect, making them a more significant threat. In this paper, we present *Adversarial-LCD*, a novel *black-box evasion* attack framework that employs an eigencentality-based perturbation method and an SVM-RBF surrogate model with a Weisfeiler-Lehman feature extractor for attacking graph-based LCD. Our evaluation shows that the attack performance of *Adversarial-LCD* with the SVM-RBF surrogate model was superior to that of other machine learning surrogate algorithms, including SVM-linear, SVM-polynomial, and Bayesian classifier, demonstrating the effectiveness of our attack framework. Furthermore, we show that our eigencentality-based perturbation method outperforms other algorithms, such as Random-walk and Shortest-path, highlighting the efficiency of *Adversarial-LCD*'s perturbation selection method.

Complete Paper #144

Concept Basis Extraction for Latent Space Interpretation of Image Classifiers

Alexandros Doumanoglou^{1,2}, Dimitrios Zarpalas¹ and Kurt Driessens²

¹ Information Technologies Institute, Centre for Research and Technology Hellas, 1st Km Charilaou - Thessaloniki, Greece

² Department of Advanced Computing Sciences, Maastricht University, 6200 MD Maastricht, The Netherlands

Keywords: Concept Basis, Interpretable Basis, Unsupervised Learning, Explainable AI, Interpretability, Computer Vision, Deep Learning.

Abstract: Previous research has shown that, to a large extent, deep feature representations of image-patches that belong to the same semantic concept, lie in the same direction of an image classifier's feature space. Conventional approaches compute these directions using annotated data, forming an interpretable feature space basis (also referred as concept basis). *Unsupervised Interpretable Basis Extraction (UIBE)* was recently proposed as a novel method that can suggest an interpretable basis without annotations. In this work, we show that the addition of a classifica-

tion loss term to the unsupervised basis search, can lead to bases suggestions that align even more with interpretable concepts. This loss term enforces the basis vectors to point towards directions that maximally influence the classifier's predictions, exploiting concept knowledge encoded by the network. We evaluate our work by deriving a concept basis for three popular convolutional networks, trained on three different datasets. Experiments show that our contributions enhance the interpretability of the learned bases, according to the interpretability metrics, by up-to +45.8% relative improvement. As additional practical contribution, we report hyper-parameters, found by hyper-parameter search in controlled benchmarks, that can serve as a starting point for applications of the proposed method in real-world scenarios that lack annotations.

Complete Paper #192

A Fusion Approach for Enhanced Remote Sensing Image Classification

Vian Ahmed, Khaled Jouini, Amel Tuama and Ouajdi Korbaa

University of Sousse, MARS Research Lab, LR17ES05, ISITCom, 4011 H. Sousse, Tunisia

Keywords: Remote Sensing, Land Cover Mapping, Features Fusion, Convolutional Neural Networks (CNNs), Scale-Invariant Feature Transform (SIFT), Image Classification.

Abstract: Satellite imagery provides a unique and comprehensive view of the Earth's surface, enabling global-scale land cover mapping and environmental monitoring. Despite substantial advancements, satellite imagery analysis remains a highly challenging task due to intrinsic and extrinsic factors, including data volume and variability, atmospheric conditions, sensor characteristics and complex land cover patterns. Early methods in remote sensing image classification leaned on human-engineered descriptors, typified by the widely used Scale-Invariant Feature Transform (SIFT). SIFT and similar approaches had inherent limitations in directly representing entire scenes, driving the use of encoding techniques like the Bag-of-Visual-Words (BoVW). While these encoding methods offer simplicity and efficiency, they are constrained in their representation capabilities. The rise of deep learning, fuelled by abundant data and computing power, revolutionized satellite image analysis, with Convolutional Neural Networks (CNNs) emerging as highly effective tools. Nevertheless, CNNs' extensive need for annotated data limits their scope of application. In this work we investigate the fusion of two distinctive feature extraction methodologies, namely SIFT and CNN, within the framework of Support Vector Machines (SVM). This fusion approach seeks to harness the unique advantages of each feature extraction method while mitigating their individual limitations. SIFT excels at capturing local features critical for identifying specific image characteristics, whereas CNNs enrich representations with global context, spatial relationships and hierarchical features. The integration of SIFT and CNN features helps thus in enhancing resilience to perturbations and generalization across diverse landscapes. An additional advantage is the adaptability of this approach to scenarios with limited labelled data. Experiments on the EuroSAT dataset demonstrate that the proposed fusion approach outperforms SIFT-based and CNN-based models used separately and that it achieves either better or comparable results when compared to existing notable approaches in remote sensing image classification.

Complete Paper #225

Ontology-Driven Deep Learning Model for Multitask Visual Food Analysis

Daniel Ponte¹, Eduardo Aguilar^{1,2}, Mireia Ribera¹ and Petia Radeva^{1,3}

¹ Dept. de Matemàtiques i Informàtica, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, Barcelona, Spain

² Dept. de Ingeniería de Sistemas y Computación, Universidad Católica del Norte, Angamos 0610, Antofagasta, Chile

³ Computer Vision Center, Cerdanyola (Barcelona), Spain

Keywords: Food Ontology, Food Image Analysis, Multitask Learning.

Abstract: The food analysis from images is a challenging task that has gained significant attention due to its multiple applications, especially in the field of health and nutrition. Ontology-driven deep learning techniques have shown promising results in improving model performance. Food ontology can leverage domain-specific information to guide model learning and thus substantially enhance the food analysis. In this paper, we propose a new ontology-driven multi-task learning approach for food recognition. To this end, we deal multi-modal information, text and images, in order to extract from the text the food ontology, which represents prior knowledge about the relationship of food concepts at different semantic levels (e.g. food groups and food names), and apply this information to guide the learning of the multi-task model to perform the task at hand. The proposed method was validated on the public food dataset named MAFood-121, specifically on dishes belonging to Mexican cuisine, outperforming the results obtained in single-label food recognition and multi-label food group recognition. Moreover, the proposed integration of the ontology into the deep learning framework allows providing more consistent results across the tasks.

Complete Paper #268

A Cascade Methodology to Evaluate Black-Box Recognition Systems Based on a Copycat Algorithm

Dinh Nguyen¹, Nhan Le², Van Mai³, Tuong Nguyen⁴, Van Nguyen⁴ and The Nguyen¹

¹ Hong Duc University, Thanh Hoa, Vietnam

² Microsoft, Ha Noi, Vietnam

³ ThinkLABs JSC, Thanh Hoa, Vietnam

⁴ Ministry of Public Security, Vietnam

Keywords: White-Box Attack, Black-Box Attack, Adversarial Image, Deep Learning, Convolutional Neural Network.

Abstract: With the significant advancements of deep learning (DL) and convolutional neural networks (CNNs), many complex systems in the field of computer vision (CV) have been effectively solved with promising performance, even equivalent to human capabilities. Images sophisticatedly perturbed in order to cause accurately trained deep learning systems to misclassify have emerged as a significant challenge and major concern in application domains requiring high reliability. These samples are referred to as adversarial examples. Many studies apply white-box attack methods to create these adversarial images. However, white-box attacks might be impractical in real-world applications. In this paper, a cascade methodology is deployed in which the Copycat algorithm is utilized to replicate the behavior of a black-box model (known as an original model) by using a substitute model. The substitute model is employed to generate white-box perturbations,

which are then used to evaluate the black-box models. The experiments are conducted with benchmark datasets as MNIST and CIFAR10 and a facial recognition system as a real use-case. The results show impressive outcomes, as the majority of the adversarial samples generated can significantly reduce the overall accuracy and reliability of facial recognition systems up to over 80%.

Oral Presentations (Online) 1 **HUCAPP**
10:30 - 12:30 **Room Online 2**
Human Computer Interaction Theory and Applications

Complete Paper #275

I Feel Safe with the Prediction: The Effect of Prediction Accuracy on Trust

Lisa Graichen¹ and Matthias Graichen²

¹ *Department of Psychology and Ergonomics, TU Berlin, Marchstraße 23, Berlin, Germany*

² *Universität der Bundeswehr München, Munich, Germany*

Keywords: XAI, HCXAI, Mental Model, Trust.

Abstract: Mental model building and trust are important topics in the interaction with Artificial Intelligence (AI)-based systems, particularly in domains involving high risk to user safety. The presented paper describes an upcoming study on applying methods from Human-Centered Explainable AI (HCXAI) to investigate the influence of AI accuracy on user trust. For the interaction with an AI-based system, we use an algorithm designed to support drivers at intersections by predicting turning maneuvers. We will investigate how drivers subjectively feel towards different presented and implemented accuracy levels. Participants will be asked to rate their respective levels of trust and acceptance. Moreover, we will investigate, whether trust and acceptance are depending on personality traits.

Complete Paper #276

Virtual Reality and Autism Spectrum Disorder: Emergence of Sensory-Motor and Olfactory Potentialities in an Anthropocentric Epistemological Approach

Cécile Lacôte-Coquereau¹, Paul Richard², Emmanuelle Richard² and Patrice Bourdon¹

¹ *Centre de Recherche en Education de Nantes, Nantes University, Chemin de la Censive du Tertre, Nantes, France*

² *LARIS – SFR MATHSTIC, University of Angers, 62 Av. Notre Dame du Lac, Angers, France*

Keywords: Virtual Reality, Virtual Environment, 3D Interaction, User-Centred Design, Autism Spectrum Disorder.

Abstract: The work presented in this paper is part of an innovative program including researchers from computer science, psychology, and education science. The aim is to propose immersive virtual environments to develop autonomy skills of young adults with autism having specific visual, psycho-sensorial, and cognitive capabilities. Several skills towards a progressive autonomy are thus targeted: interactions and social skills, verbal or alternative communication, perception-action coupling, and joint attention. In this context, a virtual supermarket has been developed, allowing participants to be confronted with shopping activity.

Complete Paper #397

Towards Automated Decision Making in Dating Apps Through Pupillary Responses

Jan Ehlers¹, Sebastian Alfonso² and Arup Mazumder³

¹ *Department of Computer Science, Bauhaus-Universität Weimar, Schwanseestr. 143, Weimar, Germany*

² *Unstructured.io, 7580 Horseshoe Bar Rd., Loomis, CA, U.S.A.*

³ *University of Rhode Island, 45 Upper College Rd, Kingston, RI, U.S.A.*

Keywords: Cognitive Pupillometry, Affective Processing, Arousal, Decision Making, Automation.

Abstract: Decision making is a multi-stage process that involves a series of rational evaluations. Recently, bodily arousal has been identified as a factor that mediates individual decisions, particularly during partner selection. The current study investigates pupil size changes in response to facial images of the opposite sex from controlled eye-tracking data (Experiment 1) and by reading out signals from front-facing smartphone cameras in noisy environments (Experiment 2). The aim is to enable automated decision-making in dating apps using arousal-based information. The rating results showed a tendency towards moderate evaluations when coping with facial attractiveness, while pupil diameter did not clearly discriminate between all four rating categories. However, a ROCKET model was trained on the pupil data from Experiment 1 with a prediction accuracy of 77% for binary classification of clearly preferred and non-preferred images. Ambiguous responses will therefore continue to pose a problem for cognition-aware systems. Capturing pupil diameter from mobile phone cameras resulted in a high proportion of inadequate recordings, probably due to a lack of experimental control. However, an overly systematic approach should run contrary to the intended scenario of lifelike mobile dating app usage.

Complete Paper #94

Applying Cognitive and Cultural Frameworks to mHealth Application Design for Elderly Users

Mia Forbes, Joyram Chakraborty and Johannes Badejo

Department of Computer and Information Science, Towson University, Towson, MD, U.S.A.

Keywords: Cognitive Theories, Conceptual Models, Usability, Perceived Usefulness, Perceived Ease of Use, User-Centered Design, Interaction Design, mHealth.

Abstract: Information and communication technologies and mobile health applications are becoming more pervasive in healthcare. Although these technologies aim to be harmoniously ubiquitous, there are important design factors that user experience practitioners must consider for optimal utilization for elderly users. User's perceived usefulness is the driving factor for technology adoption. Incorporating cognitive and conceptual design principles such as information processing, mental models, external cognition, and emotional interaction based upon sociocultural determinants can improve the usefulness of mobile Health tools amongst elderly users. Data was obtained from a small sample (n=30) aged 75 years or older who use telehealth and mobile health apps. Findings show that users have a low perceived ease of use and usefulness of these applications based on the lack of sociocultural elements incorporated into the interface design. This paper aims to analyze the impact cognitive theories and conceptual frameworks have on mobile health design for elderly users, while bridging the gap of an existing digital divide. A qualitative study was conducted involving empirical research to define a correlational relationship between technology dissonance

and elderly users for the purpose of identifying a thematic analysis.

Complete Paper #14

Modelling Cognitive Workload to Build Multimodal Voice Interaction in the Car

Sylvia Bhattacharya¹ and J. Higgins²

¹ Department of Engineering Technology, Kennesaw State University, Marietta, Georgia, U.S.A.

² UX Research, Google, Inc, Mountain view, California, U.S.A

Keywords: In-Vehicle Information Systems, Cognitive Demands, Optimization, Visual/Auditory Modalities, Tactile Screen Tapping, Voice Commands, Visual Cues, Cognitive Load.

Abstract: The paper discusses the integration of in-vehicle information systems and their impact on driver performance, considering the demands of various types such as visual, auditory, manual, and cognitive. It notes that while there's a lot of research on optimizing visual and manual systems, less attention has been paid to systems that use both visual and auditory cues or a combination of different types. The study has found that simple tasks cause the least cognitive strain when drivers use touchscreens, while complex tasks are easier to manage cognitively when voice commands are used alone or with visual aids. These results are important for designing car interfaces that effectively manage the driver's cognitive load.

Complete Paper #48

Look-over-there: Real-World Co-Located Cross-Referencing Using Augmented Reality

Yuqi Zhou and Voicu Popescu

Purdue University, 610 Purdue Mall, West Lafayette, IN, U.S.A.

Keywords: Collaborative Augmented Reality, Transparent Display, Attention Guidance.

Abstract: This paper presents a method that allows a "guide" to point out an element of the real world, i.e., a reference point, to a "tourist". The guide and tourist stand side-by-side and each hold a tablet whose camera is aimed at the scene. The guide annotates the reference point on their tablet, and it is sent and displayed on the tourist's tablet. Then the device zooms in and guides the tourist towards the target object. A user study shows that the method has significantly shorter reference point localization times compared to a conventional augmented reality interface. Furthermore, the study shows that the method can provide directional guidance through the annotation alone, without any reliance on the visual appearance of the region of the reference point, as needed for challenging scenes with repeated patterns or devoid of visual features.

Tutorial

10:30 - 12:30

Room Space 4C

Context-Aware Applications Using Visual Programming: Case Studies on Mobile Apps and Humanoid Robot Applications

Context-Aware Applications Using Visual Programming: Case Studies on Mobile Apps and Humanoid Robot Applications

Martin Zimmermann

Offenburg University, Germany

Abstract: The main privilege of context-aware applications is to provide tailored services by analyzing the environmental context, such as location, time, weather condition, and seasons, and adapting their functionality according to the changing situations in context data without explicit user interaction. For example, mobile devices can obtain context information in various ways in order to provide more adaptable, flexible and user-friendly services. In case of a tourist app, a tourist would like to see relevant tourist attractions on a map together with distance information, depending on its current location. Human robots, require sensors to gather information about the conditions of the environment to allow the robot to make necessary decisions about its position or certain actions that the situation requires. As a consequence, context-aware applications can sense clues about the situational environment making applications more intelligent, adaptive, and personalized. Visual Programming Languages (VPL) let users develop software programs by combining visual program elements, like sensor and actuator objects, loops or conditional statements rather than by specifying them textually. VPLs are considered to be innovative approaches to address the inherent complexity of developing programs, especially for beginners. The tutorial will cover important technologies which can be used to build user-centered context-based applications, discussing challenges and open problems and will give conclusions and insights for research in the field. Several concrete scenarios will be analyzed and corresponding mobile applications will be developed during the tutorial by the participants using a cloud-based visual programming environment.

Session 2A

14:00 - 16:00

Transfer Learning

VISAPP

Room Space 1

Complete Paper #44

Synthesizing Classifiers from Prior Knowledge

G. Burghouts, K. Schutte, M. Kruijthof, W. Huizinga, F. Ruis and H. Kuijff

TNO, Intelligent Imaging, The Netherlands

Keywords: Classifier Synthesis, Knowledge Representation, Semantics, Attributes, Class Descriptions, Large Language Models, Zero-Shot Learning, Few-Shot Learning.

Abstract: Various good methods have been proposed for either zero-shot or few-shot learning, but these are commonly unsuited for both; whereas in practice one often starts without labels and some might become available later. We propose a method that naturally ties zero- and few-shot learning together. We initiate a zero-shot model from prior knowledge about the classes, by recombining the weights from a classification head via a linear reconstruction that is sparse to avoid overfitting. Our mapping is an explicit transfer of knowledge from known to new classes, hence it can be inspected and visualized, which is impossible with

recently popular implicit prompt learning strategies. Our mapping is used to construct a classifier for the new class, by adapting the neural weights of the classifiers for the known classes. Effectively we synthesize a new classifier. Our method is flexible: we show its efficacy for various knowledge representations and various neural networks (whereas prompt learning is limited to language-vision models). Our synthesized classifier can operate directly on test samples in a zero-shot fashion. We outperform CLIP especially for uncommon image classes, sometimes by margins up to 32%. Because the synthesized classifier consists of a tensor layer, it can be optimized further when a (few) labeled images become available. For few-shot learning, our synthesized classifier provides a kickstart. With one label per class, it outperforms strong baselines that require annotation of attributes or heavy pretraining (CLIP) by 8%, and increases accuracy by 39% relative to conventional classifier initialization. The code is available.

Complete Paper #153

Enabling RAW Image Classification Using Existing RGB Classifiers

Rasmus Munkso¹, Mathias Andersen¹, Lau Nørgaard², Andreas Møgelmo¹ and Thomas Moeslund¹

¹ Visual Analysis and Perception Lab, Aalborg University, Denmark

² Phase One A/S, Denmark

Keywords: RAW, RGB, Transfer Learning, RAW Image Dataset, Classification.

Abstract: Unprocessed RAW data stands out as a highly valuable image format in image editing and computer vision due to it preserving more details, colors, and a wider dynamic range as captured directly from the camera's sensor compared to non-linearly processed RGB images. Despite its advantages, the computer vision community has largely overlooked RAW files, especially in domains where preserving precise details and accurate colors are crucial. This work addresses this oversight by leveraging transfer learning techniques. By exploiting the vast amount of available RGB data, we enhance the usability of a limited RAW image dataset for image classification. Surprisingly, applying transfer learning from an RGB-trained model to a RAW dataset yields impressive performance, reducing the dataset size barrier in RAW research. These results are promising, demonstrating the potential of cross-domain transfer learning between RAW and RGB data and opening doors for further exploration in this area of research.

Complete Paper #32

Classification of Towels in a Robotic Workcell Using Deep Neural Networks

Jens Rossen¹, Patrick Terp¹, Norbert Krüger¹, Laus Bigum² and Tudor Morar²

¹ Maersk Mc-Kinney Møller Institute (MMMI), University of Southern Denmark (SDU), Campusvej, Odense, Denmark

² Inwatec, Odense, Denmark

Keywords: Image Classification, AI, Deep Neural Networks, Towels, Laundry Industry.

Abstract: The industrial laundry industry is becoming increasingly more automated. Inwatec, a company specializing in this field, is developing a new robot (BLIZZ) to automate the process of grasping individual clean towels from a pile, and hand them over to an external folding machine. However, to ensure that towels are folded consistently, information about the type and faces of the

towels is required. This paper presents a proof of concept for a towel type and towel face classification system integrated in BLIZZ. These two classification problems are solved by means of a Deep Neural Network (DNN). The performance of the proposed DNN on each of the two classification problems is presented, along with the performance of it solving both classification problems at the same time. It is concluded that the proposed network achieves classification accuracies of 94.48%, 97.71% and 98.52% on the face classification problem for three different towel types with non-identical faces. On the type classification problem, it achieves an accuracy of 99.10% on the full dataset. Additionally, it is concluded that the system achieves an accuracy of 96.96% when simultaneously classifying the type and face of a towel on the full dataset.

Complete Paper #279

Unsupervised Annotation and Detection of Novel Objects Using Known Objectness

Harsh Jadon, Jagdish Deshmukh, Kamakshya Nayak, Kamalakar Thakare and Debi Dogra ψ

Indian Institute of Technology Bhubaneswar, Odisha, PIN 752050, India

Keywords: Novel Object Detection, Unsupervised Learning, Weakly Annotated Dataset, Clustering.

Abstract: The paper proposes a new approach to detecting and annotating novel objects in images that are not precisely part of a training dataset. The ability to detect novel objects is essential in computer vision, enabling machines to recognise objects that have not been seen before. Current models often fail to detect novel objects as they rely on predefined categories in the training data. Our approach overcomes this limitation by leveraging a large and diverse dataset of objects obtained through web scraping. We extract features using a backbone network and perform clustering to remove redundant data. The resulting dataset is used to retrain the object detection models to obtain results. The method provides deep insights into the effect of clustering and data redundancy removal on performance. Overall, the work contributes to the field of object detection by providing a new approach for detecting novel objects. The method has the potential to be applied to a variety of real-world CV applications.

Complete Paper #413

Evaluating Multiple Combinations of Models and Encoders to Segment Clouds in Satellite Images

Jocsan Ferreira¹, Leandro Silva¹, Mauricio Escarpinati², André Backes³ and João Mari¹

¹ Institute of Exact and Technological Sciences, Federal University of Viçosa, Brazil

² School of Computer Science, Federal University of Uberlândia, Brazil

³ Department of Computing, Federal University of São Carlos, Brazil

Keywords: Cloud Segmentation, Satellite Images, Deep Learning.

Abstract: This work evaluates methods based on deep learning to perform cloud segmentation in satellite images. We compared several semantic segmentation architectures using different encoder structures. In this sense, we fine-tuned three architectures (U-Net, LinkNet, and PSPNet) with four pre-trained encoders (ResNet-50, VGG-16, MobileNet V2, and EfficientNet B2). The performance of the models was evaluated using the Cloud-38 dataset. The training process was carried out until the validation loss stabilized, according to the early stopping criterion, which

provides a comparative analysis of the best models and training strategies to perform cloud segmentation in satellite images. We evaluated the performance using classic evaluation metrics, i.e., pixel accuracy, mean pixel accuracy, mean IoU, and frequency-based IoU. Results demonstrated that the tested models are capable of segmenting clouds with considerable performance, with emphasis on the following values: (i) 96.19% pixel accuracy for LinkNet with VGG-16 encoder, (ii) 92.58% mean pixel accuracy for U-Net with MobileNet V2 encoder, (iii) 87.21% mean IoU for U-Net with VGG-16 encoder, and (iv) 92.89% frequency-based IoU for LinkNet with VGG-16 encoder. In short, the results of this study provide valuable information for developing satellite image analysis solutions in the context of precision agriculture.

Session 2B
14:00 - 16:00
Generative AI

VISAPP
Room Space 3B

Complete Paper #45

StyleHumanCLIP: Text-Guided Garment Manipulation for StyleGAN-Human

Takato Yoshikawa, Yuki Endo and Yoshihiro Kanamori
University of Tsukuba, Japan

Keywords: StyleGAN, Text-Guided Garment Manipulation, Full-Body Human Image.

Abstract: This paper tackles text-guided control of StyleGAN for editing garments in full-body human images. Existing StyleGAN-based methods suffer from handling the rich diversity of garments and body shapes and poses. We propose a framework for text-guided full-body human image synthesis via an attention-based latent code mapper, which enables more disentangled control of StyleGAN than existing mappers. Our latent code mapper adopts an attention mechanism that adaptively manipulates individual latent codes on different StyleGAN layers under text guidance. In addition, we introduce feature-space masking at inference time to avoid unwanted changes caused by text inputs. Our quantitative and qualitative evaluations reveal that our method can control generated images more faithfully to given texts than existing methods.

Complete Paper #399

Image Generation from Hyper Scene Graphs with Trinomial Hyperedges Using Object Attention

Ryosuke Miyake, Tetsu Matsukawa and Einoshin Suzuki
Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan

Keywords: Image Generation, Hyper Scene Graph, Object Attention.

Abstract: Conditional image generation, which aims to generate consistent images with a user's input, is one of the critical problems in computer vision. Text-to-image models have succeeded in generating realistic images for simple situations in which a few objects are present. Yet, they often fail to generate consistent images for texts representing complex situations. Scene-graph-to-image models have the advantage of generating images for complex situations based on the structure of a scene graph. We extended a scene-graph-to-image model to an image generation model from a hyper scene graph with trinomial hyperedges. Our model, termed hsg2im, improved the consistency of the generated images. However, hsg2im has difficulty in generating natural and consistent images for hyper scene graphs with many objects.

The reason is that the graph convolutional network in hsg2im struggles to capture relations of distant objects. In this paper, we propose a novel image generation model which addresses this shortcoming by introducing object attention layers. We also use a layout-to-image model auxiliary to generate higher-resolution images. Experimental validations on COCO-Stuff and Visual Genome datasets show that the proposed model generates more natural and consistent images to user's inputs than the cutting-edge hyper scene-graph-to-image model.

Complete Paper #121

Efficient Posterior Sampling for Diverse Super-Resolution with Hierarchical VAE Prior

Jean Prost¹, Antoine Houdard², Andrés Almansa³ and Nicolas Papadakis⁴

¹ *Univ. Bordeaux, Bordeaux IMB, INP, CNRS, UMR 5251, F-33400 Talence, France*

² *Ubisoft La Forge, F-33000 Bordeaux, France*

³ *Université Paris Cité, CNRS, MAP5, F-75006 Paris, France*

⁴ *Univ. Bordeaux, CNRS, INRIA, Bordeaux INP, IMB, UMR 5251, F-33400 Talence, France*

Keywords: Diverse Image Super-Resolution, Hierarchical Variational Autoencoder, Conditional Generative Model.

Abstract: We investigate the problem of producing diverse solutions to an image super-resolution problem. From a probabilistic perspective, this can be done by sampling from the posterior distribution of an inverse problem, which requires the definition of a prior distribution on the high-resolution images. In this work, we propose to use a pretrained hierarchical variational autoencoder (HVAE) as a prior. We train a lightweight stochastic encoder to encode low-resolution images in the latent space of a pretrained HVAE. At inference, we combine the low-resolution encoder and the pretrained generative model to super-resolve an image. We demonstrate on the task of face super-resolution that our method provides an advantageous trade-off between the computational efficiency of conditional normalizing flows techniques and the sample quality of diffusion based methods.

Complete Paper #270

Generating Videos from Stories Using Conditional GAN

Takahiro Kozaki, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Nagoya 466-8555, Japan

Keywords: Generative AI, Story, Multiple Sentences, Video, GAN, Captioning.

Abstract: In this paper, we propose a method for generating videos that represent stories described in multiple sentences. While research on generating images and videos from single sentences has been advancing, the generation of videos from long stories written in multiple sentences has not been achieved. In this paper, we use adversarial learning to train pairs of multi-sentence stories and videos to generate videos that replicate the flow of the stories. We also introduce caption loss for generating more contextually aligned videos from stories.

Complete Paper #408

Image Augmentation for Object Detection and Segmentation with Diffusion Models

Leon Useinov, Valeria Efimova and Sergey Muravyov
ITMO University, Russia

Keywords: Augmentation, Image Generation, Diffusion Models, Object Detection, Segmentation.

Abstract: Training current state-of-the-art models for object detection and segmentation requires a lot of labeled data, which can be difficult to obtain. It is especially hard, when occurrence of an object of interest in a certain required environment is rare. To solve this problem we present a train-free augmentation technique that is based on a diffusion model, pretrained on a large dataset (more than 1 million images). In order to establish the effectiveness of our method and its modifications, experiments on small datasets (less than 500 training images) with YOLOv8 are conducted. We conclude that none of the proposed versions of the diffusion-based augmentation method are universal, however, each of them may be used to improve an object detection (and segmentation) model performance in certain scenarios. The code is publicly available: github.com/PnthrLeo/diffusion-augmentation.

Session 2C
14:00 - 16:00
Image-Based Modeling and 3D Reconstruction

VISAPP
Room Space 3C

Complete Paper #182

Animating NeRFs from Texture Space: A Framework for Pose-Dependent Rendering of Human Performances

Paul Knoll^{1,2}, Wieland Morgenstern¹, Anna Hilsmann^{1,3} and Peter Eisert^{1,3}

¹ Fraunhofer Heinrich Hertz Institute, HHI, Germany

² Johannes Kepler University Linz, Germany

³ Humboldt University of Berlin, Germany

Keywords: Computer Vision Representations, Neural Radiance Fields, Animation, Rendering.

Abstract: Creating high-quality controllable 3D human models from multi-view RGB videos poses a significant challenge. Neural radiance fields (NeRFs) have demonstrated remarkable quality in reconstructing and free-viewpoint rendering of static as well as dynamic scenes. The extension to a controllable synthesis of dynamic human performances poses an exciting research question. In this paper, we introduce a novel NeRF-based framework for pose-dependent rendering of human performances. In our approach, the radiance field is warped around an SMPL body mesh, thereby creating a new surface-aligned representation. Our representation can be animated through skeletal joint parameters that are provided to the NeRF in addition to the viewpoint for pose dependent appearances. To achieve this, our representation includes the corresponding 2D UV coordinates on the mesh texture map and the distance between the query point and the mesh. To enable efficient learning despite mapping ambiguities and random visual variations, we introduce a novel remapping process that refines the mapped coordinates. Experiments demonstrate that our approach results in high-quality renderings for novel-view and novel-pose synthesis.

Complete Paper #261

ImgAdaPoinTr: Improving Point Cloud Completion via Images and Segmentation

Vaagn Chopuryan¹, Mikhail Kuznetsov^{1,2}, Vasilii Latonov¹ and Natalia Semenova^{1,3}

¹ Sberbank PJSC, 19 Vavilova St., Moscow 117312, Russia

² National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 31 Kashirskoe shosse, 31, Moscow 115409, Russia

³ AIRI, Nizhniy Susalny Lane 5, Moscow 105064, Russia

Keywords: Point Cloud Completion, Transformers, View-Guided, Segmentation.

Abstract: Point cloud completion is an essential task consisting of inferring and filling in missing parts of a 3D point cloud representation. In this paper, we present an ImgAdaPoinTr model, which extends the original Transformer encoder-decoder architecture by accurately incorporating visual information. Besides, we assumed using segmentation of 3D objects as a part of the pipeline due to acquiring an additional increase in performance. We also introduce the novel ImgPCN dataset generated by our rendering tool. The results show that our approach outperforms AdaPoinTr by average 2.9% and 10.3% in terms of Chamfer-Distance L1 and L2 metrics, respectively. The code and dataset are available via the link <https://github.com/ImgAdaPoinTr>.

Complete Paper #299

Evaluation of 3D Point Cloud Distances: A Comparative Study in Multi-Point Cloud Fusion Environments

Ulugbek Alibekov, Vanessa Staderini, Geetha Ramachandran, Philipp Schneider and Doris Antensteiner
AIT Austrian Institute of Technology, Vienna, Austria

Keywords: Point Cloud Registration, Complex Shape Reconstruction, Industrial Inspection, Distance Metrics, 3D Reconstruction.

Abstract: In the domain of 3D shape reconstruction and metrology, the precise alignment and measurement of point clouds is critical, especially within the context of industrial inspection where accuracy requirements are high. This work addresses challenges stemming from intricate object properties, including complex geometries or surfaces, resulting in diverse artefacts, holes, or sparse point clouds. We present a comprehensive evaluation of point cloud measurement metrics on different object shapes and error patterns. We focus on the task of point cloud evaluation of objects to assess their quality. This is achieved through the acquisition of partial point clouds acquired from multiple perspectives. This is followed by a point cloud fusion process including an initial alignment and a point cloud refinement step. We evaluate these point clouds with respect to a reference sampled point cloud and mesh. In this work, we evaluate various point cloud metrics across experimentally relevant scenarios like cloud density variations, different noise levels, and hole sizes on objects with different geometries. We additionally show how the approach can be applied in industrial object evaluation.

Complete Paper #342

Toward Multi-Plane Image Reconstruction from a Casually Captured Focal Stack

Shiori Ueda¹, Hideo Saito¹ and Shohei Mori^{1,2}¹ Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, Japan² Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Styria, Austria

Keywords: Focal Stack Imaging, Casual Photography, Multi-Plane Image, View Synthesis.

Abstract: 3D imaging combining a focal stack and multi-plane images (MPI) facilitates various real-time applications, including view synthesis, 3D scene editing, and augmented and virtual reality. Building upon the foundation of MPI, originally derived from multi-view images, we introduce a novel pipeline for reconstructing MPI by casually capturing a focal stack optically using a handheld camera with a manually modulated focus ring. We hypothesized two distinct strategies for focus ring modulation that users could employ, to sample defocus images along the front-facing axis uniformly. Our quantitative analysis using a synthetic dataset suggests tendencies in possible simulated errors in focus modulations, while qualitative results illustrate visual differences. We further showcase applications utilizing the resultant MPI, including depth rendering, occlusion-aware defocus filtering, and de-fencing.

Session 2D

14:00 - 16:00

Motion, Tracking and Stereo Vision

VISAPP

Room Space 4A

Complete Paper #95

BEVSeg2TP: Surround View Camera Bird's-Eye-View Based Joint Vehicle Segmentation and Ego Vehicle Trajectory Prediction

Sushil Sharma^{1,2}, Arindam Das¹, Ganesh Sistu¹, Mark Halton¹ and Ciarán Eising^{1,2}¹ Department of Electronic & Computer Engineering, University of Limerick, Ireland² SFI CRT Foundations in Data Science, University of Limerick, Ireland

Keywords: Surrounded-View Camera, Encoder-Decoder Transformer, Segmentation, Trajectory Prediction.

Abstract: Trajectory prediction is, naturally, a key task for vehicle autonomy. While the number of traffic rules is limited, the combinations and uncertainties associated with each agent's behaviour in real-world scenarios are nearly impossible to encode. Consequently, there is a growing interest in learning-based trajectory prediction. The proposed method in this paper predicts trajectories by considering perception and trajectory prediction as a unified system. In considering them as unified tasks, we show that there is the potential to improve the performance of perception. To achieve these goals, we present BEVSeg2TP - a surround-view camera bird's-eye-view-based joint vehicle segmentation and ego vehicle trajectory prediction system for autonomous vehicles. The proposed system uses a network trained on multiple camera views. The images are transformed using several deep learning techniques to perform semantic segmentation of objects, including other vehicles, in the scene. The segmentation outputs are fused across the camera views to obtain a comprehensive representation of the surrounding vehicles from the bird's-eye-view perspective. The system further predicts the future trajectory

of the ego vehicle using a spatiotemporal probabilistic network (STPN) to optimize trajectory prediction. This network leverages information from encoder-decoder transformers and joint vehicle segmentation. The predicted trajectories are projected back to the ego vehicle's bird's-eye-view perspective to provide a holistic understanding of the surrounding traffic dynamics, thus achieving safe and effective driving for vehicle autonomy. The present study suggests that transformer-based models that use cross-attention information can improve the accuracy of trajectory prediction for autonomous driving perception systems. Our proposed method outperforms existing state-of-the-art approaches on the publicly available nuScenes dataset. This link is to be followed for the source code: <https://github.com/sharmasushil/BEVSeg2TP/>.

Complete Paper #125

Are Semi-Dense Detector-Free Methods Good at Matching Local Features?

Matthieu Vilain, Rémi Giraud, Hugo Germain and Guillaume Bourmaud

Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400 Talence, France

Keywords: Image Matching, Transformer, Pose Estimation.

Abstract: Semi-dense detector-free approaches (SDF), such as LoFTR, are currently among the most popular image matching methods. While SDF methods are trained to establish correspondences between two images, their performances are almost exclusively evaluated using relative pose estimation metrics. Thus, the link between their ability to establish correspondences and the quality of the resulting estimated pose has thus far received little attention. This paper is a first attempt to study this link. We start with proposing a novel structured attention-based image matching architecture (SAM). It allows us to show a counter-intuitive result on two datasets (MegaDepth and HPatches): on the one hand SAM either outperforms or is on par with SDF methods in terms of pose/homography estimation metrics, but on the other hand SDF approaches are significantly better than SAM in terms of matching accuracy. We then propose to limit the computation of the matching accuracy to textured regions, and show that in this case SAM often surpasses SDF methods. Our findings highlight a strong correlation between the ability to establish accurate correspondences in textured regions and the accuracy of the resulting estimated pose/homography. Our code will be made available.

Complete Paper #222

BASE: Probably a Better Approach to Visual Multi-Object Tracking

Martin Larsen^{1,2}, Sigmund Rolfsjord^{1,2}, Daniel Gusland^{1,2}, Jörgen Ahlberg^{2,3} and Kim Mathiassen^{1,2}¹ Norwegian Defence Research Establishment, Norway² University of Oslo, Norway³ Linköping University, Norway

Keywords: Visual Multi-Object Tracking, Probabilistic Tracking, Distance-Aware, Association-Less Track Management.

Abstract: The field of visual object tracking is dominated by methods that combine simple tracking algorithms and ad hoc schemes. Probabilistic tracking algorithms, which are leading in other fields, are surprisingly absent from the leaderboards. We found that accounting for distance in target kinematics, exploiting detector confidence and modelling non-uniform clutter

characteristics is critical for a probabilistic tracker to work in visual tracking. Previous probabilistic methods fail to address most or all these aspects, which we believe is why they fall so far behind current state-of-the-art (SOTA) methods (there are no probabilistic trackers in the MOT17 top 100). To rekindle progress among probabilistic approaches, we propose a set of pragmatic models addressing these challenges, and demonstrate how they can be incorporated into a probabilistic framework. We present BASE (Bayesian Approximation Single-hypothesis Estimator), a simple, performant and easily extendible visual tracker, achieving state-of-the-art (SOTA) on MOT17 and MOT20, without using Re-Id. Code available at <https://github.com/ffi-no/paper-base-visapp-2024>.

Complete Paper #370

A Multilevel Strategy to Improve People Tracking in a Real-World Scenario

Cristiano B. de Oliveira^{1,2}, Joao Neves³, Rafael Ribeiro⁴ and David Menotti¹

¹ Department of Informatics, Federal University of Paraná, Curitiba, Brazil

² Federal University of Ceará, Quixadá, Brazil

³ University of Beira Interior, NOVA-LINCS, Portugal

⁴ National Institute of Criminalistics, Brazilian Federal Police, Curitiba, Brazil

Keywords: People Tracking, Dataset, Video Surveillance, Palácio Do Planalto.

Abstract: The Palácio do Planalto, office of the President of Brazil, was invaded by protesters on January 8, 2023. Surveillance videos taken from inside the building were subsequently released by the Brazilian Supreme Court for public scrutiny. We used segments of such footage to create the UFPR-Planalto801 dataset for people tracking and re-identification in a real-world scenario. This dataset consists of more than 500,000 images. This paper presents a tracking approach targeting this dataset. The method proposed in this paper relies on the use of known state-of-the-art trackers combined in a multilevel hierarchy to correct the ID association over the trajectories. We evaluated our method using IDF1, MOTA, MOTP and HOTA metrics. The results show improvements for every tracker used in the experiments, with IDF1 score increasing by a margin up to 9.5%.

Complete Paper #405

Self-Mounted Motion Capture System Using Mutual Projection of Asynchronous Cameras

Kazusa Ozaki, Fumihiko Sakaue and Jun Sato

Nagoya Institute of Technology, Nagoya, Japan

Keywords: Motion Capture Systems, Bundle Adjustment, Neural Network Representation, Asynchronous Stereo.

Abstract: In this research, we propose a method for restoring three-dimensional motion from time-series images captured by an asynchronous camera in order to realise a motion capture system using a camera attached to the body surface. For this purpose, we represent the motion trajectory of each marker using a neural network, and estimate the motion trajectory by optimising the neural network from the input images. It is also shown that stable 3D restoration can be achieved using a method called mutual projection, assuming that the cameras are reflecting each other. We show that it is possible to estimate 3D motion from asynchronous cameras with high accuracy.

Session 2A
14:00 - 16:00
Animation and Simulation

GRAPP
Room Space 4B

Complete Paper #75

Real-Time Desertscape Simulation with CUDA

Alexander Nilles, Lars Günther and Stefan Müller

Institute for Computational Visualistics, University of Koblenz, Universitätsstr. 1, Koblenz, Germany

Keywords: Procedural Modeling, Desert, Sand-Dunes, Simulation, Aeolian Erosion, Real-Time, GPU, CUDA.

Abstract: We propose a new GPU-based method capable of simulating dune formation, propagation and aeolian sand transport, based on the Desertscape Simulation model. Our method improves upon the original as well as an existing real-time GPU implementation by introducing bilinear interpolation and removing randomness, which leads to a more robust and noise-free method. We implement our method with CUDA and use atomic adds on floats to remove a previous limitation that required block-based discretization of elevation values. A new sand distribution scheme for sand avalanching is proposed that converges faster than the previous work. We propose and evaluate a new method for reptation, which was previously neglected. Our method further improves on the performance compared with the previous real-time method and can generate results that more closely resemble dune evolution with a bidirectional wind scheme as predicted by an accurate offline method. Our method can generate detailed, physically plausible desert environments very quickly, with possible applications in both computer graphics as well as geomorphology. With some restrictions, our method could even be used during gameplay.

Complete Paper #143

Using the Polynomial Particle-in-Cell Method for Liquid-Fabric Interaction

Robert Dennison and Steve Maddock

The University of Sheffield, U.K.

Keywords: Physically-Based Modeling, Fluid Simulation, Cloth Simulation.

Abstract: Liquid-fabric interaction simulations using particle-in-cell (PIC) based models have been used to simulate a wide variety of phenomena and yield impressive visual results. However, these models suffer from numerical damping due to the data interpolation between the particles and grid. Our paper addresses this by using the polynomial PIC (PolyPIC) model instead of the affine PIC (APIC) model that is used in current state-of-the-art wet cloth models. Theoretically, PolyPIC has lossless energy transfer and so should avoid any problems of undesirable damping and numerical viscosity. Our results show that PolyPIC does enable more dynamic coupled simulations. The use of PolyPIC allows for simulations with reduced numerical dissipation and improved resolution of vortical details over previous work. For smaller scale simulations, there is minimal impact on computational performance when using PolyPIC instead of APIC. However, as simulations involve a larger number of particles and mesh elements, PolyPIC can require up to a 2.5x as long to generate 4.0s of simulation due to a requirement for a decrease in timestep size to remain stable.

Complete Paper #256

Data-Driven Viscosity Solver for Fluid Simulation

Wonjung Park, Hyunsoo Kim and Jinah Park

Korea Advanced Institute of Science and Technology, Daejeon, Korea

Keywords: Physical Simulation, Machine Learning, Viscous Fluid, Convolutional Neural Network.

Abstract: We propose a data-driven viscosity solver based on U-shaped convolutional neural network to predict velocity changes due to viscosity. Our solver takes velocity derivatives, fluid volume, and solid indicator quantities as input. The traditional marker-and-cell (MAC) grid stores velocities at the edges of the grid, causing the dimensions of the velocity field vary from axis to axis. In our work, we suggest a symmetric MAC grid that maintains consistent dimensions across axes without interpolation or symmetry breaking. The proposed grid effectively transfers spatial fluid quantities such as partial derivatives of velocity, enabling networks to generate accurate predictions. Additionally, we introduce a physics-based loss inspired by the variational formulation of viscosity to enhance the network's generalization for a wide range of viscosity coefficients. We demonstrate various fluid simulation results, including 2D and 3D fluid-rigid body scenes and a scene exhibiting the buckling effect.

Tuesday, 27

Complete Paper #334

Transformer-Based Two-level Approach for Music-driven Dance Choreography

Yanbo Cheng and Yingying Wang

Department of Computing and Software, McMaster University, Hamilton, Canada

Keywords: Deep Learning, Character Animation, Motion Synthesis, Motion Stylization, Multimodal Synchronization.

Abstract: Human dance motions are complex, creative, and artistic expressions. Synthesizing high-quality dance motions and synchronizing them to music has always been a challenge in animation research. Three problems in synthesizing dance motions are presented: 1) dance movements are complex non-linear motions that follow high-level structures of the dance genre over a long horizon, yet must maintain a stylistic consistency; 2) even for the same genre, dance movements require diversity, expressiveness, and nuances to appear natural and realistic; 3) spatial-temporal features of dance movements can be influenced by music. In this paper, we address these issues using a novel two-level transformer-based dance generation system that can synthesize dance motions to match the audio input. Our high-level transformer network performs the choreography and generates dance movements with consistent long-term structure, and our low-level implementer infuses diversity and realizes actual dance performances. This two-level approach not only allows us to generate dances that are consistent in structure, but also enables us to effectively add styles learnt from a wide range of dance datasets. When training the choreography model, our approach fully utilizes existing dance datasets, even those without musical accompaniment, and thus differs from previous research that requires dance training data to be accompanied by music. Results in this work demonstrate that our two-level system generates high-quality dance motions that flexibly adapt to varying musical conditions trained on a dataset of dance sequences without accompanying music.

Session 2A
14:00 - 16:00
Visualization Techniques

IVAPP
Room Space 4C

Complete Paper #194

Scale and Time Independent Clustering of Time Series Data

Florian Steinwider^{1,2}, Istvan Szilagyi³, Eva Eggeling² and Torsten Ullrich^{1,2}¹ *Institute of Computer Graphics and Knowledge Visualization, Graz University of Technology, Graz, Austria*² *Fraunhofer Austria Research GmbH, Graz, Austria*³ *Medical University Graz, Graz, Austria*

Keywords: Time Series Analysis, Cluster Analysis, Visualization Toolkits.

Abstract: The analysis of time series, and in particular the identification of similar time series within a large set of time series, is an important part of visual analytics. This paper describes extensions of tree-based index structures to find self-similarities within sets of time series. It also describes filters that extend existing algorithms to better fit real-world, error-prone, incomplete data. The ability of time series clustering to detect common errors in real data is also described. These main contributions are illustrated with real data and the findings and lessons learned are summarised.

Complete Paper #120

Fundamental Limitations of Inverse Projections and Decision Maps

Yu Wang and Alexandru Telea

Department of Information and Computing Science, Utrecht University, 3584 CS Utrecht, The Netherlands

Keywords: Dimensionality Reduction, Inverse Projections, Decision Maps, Intrinsic Dimensionality.

Abstract: Inverse projection techniques and decision maps are recent tools proposed to depict the behavior of a classifier using 2D visualizations. However, which parts of the large, high-dimensional, space such techniques actually visualize, is still unknown. A recent result hinted at the fact that such techniques only depict a two-dimensional manifold from the entire data space. In this paper, we investigate the behavior of inverse projections and decision maps in high dimensions with the help of intrinsic dimensionality estimation methods. We find that the inverse projections are always surface-like no matter what decision map method is used and no matter how high the data dimensionality is, i.e., the intrinsic dimensionality of inverse projections is always two. Thus, decision boundaries displayed by decision maps are the intersection of the backprojected surface and the actual decision surfaces. Our finding reveals a fundamental problem of all existing decision map techniques in constructing an effective visualization of the decision space. Based on our findings, we propose solutions for future work in decision maps to address this problem.

Complete Paper #455

Understanding How Different Visual Aids for Augmented Reality Influence Tool-Patient Alignment in Surgical Tasks: A Preliminary Study

Stefano Stradiotti, Nicolas Emiliani, Emanuela Marcelli and Laura Cercenelli

eDIMES Lab - Laboratory of Bioengineering, Department of Medical and Surgical Sciences, University of Bologna, via Massarenti, 9, Bologna, Italy

Keywords: Augmented Reality, Surgery, Magic Leap 2, Visual Aids, Marker Tracking, 6DoF, Alignment, Accuracy, Image-Guided Surgery, Visualization and Rendering Techniques.

Abstract: This study explores the impact of several visual aids on the accuracy of tool-patient alignment in augmented reality (AR) assisted surgical tasks. AR has gained prominence across surgical specialties, integrating virtual models derived from patient anatomy into the surgical field. This opens avenues for innovative visual aids and feedback which can facilitate surgical operations. To assess the influence of different visual aids on surgeon performance, we conducted a tool-patient alignment test on a 3D-printed frame, involving 12 surgical residents. Each participant inserted 12 toothpicks with a release tool into predefined target positions on the frame simulating patient targets, under AR visualization through a Magic Leap 2 Head-Mounted-Display. As visual aids, four holographic solutions were employed, with two of them offering graphical feedback upon the correct alignment to the target. Linear and angular positioning errors were measured, alongside participant responses to a satisfaction questionnaire. The tests maintained a consistent tracking system for estimating target and tool poses in the real-world, ensuring evaluation stability. Preliminary results indicated statistically significant differences among the proposed visual aids, suggesting the need for further exploration in the realm of their usability in relation to the specific surgical task and the expected overall surgical accuracy.

Oral Presentations (Online) 2
14:00 - 16:00
Features Extraction

VISAPP
Room Online 2

Complete Paper #140

Deep Learning-Based Models for Performing Multi-Instance Multi-Label Event Classification in Gameplay Footage

Etienne Julia, Marcelo Zanchetta do Nascimento, Matheus Faria and Rita Julia

Computer Science Department, Federal University of Uberlândia, Uberlândia, Minas Gerais, Brazil

Keywords: MIML Event Classification, Gameplay Footage, Super Mario, Frame-Based and Chunk-Based Data Representations, Deep Extractor Neural Networks, Fine-Tuned Backbone, Deep Classifier Neural Networks.

Abstract: In dynamic environments, like videos, one of the key pieces of information to improve the performance of autonomous agents are the events, since, in a broad manner, they represent the dynamic changes and interactions that happen in the environment. Video games stand out among the most suitable domains for investigating the effectiveness of machine learning techniques. Among the challenging activities explored in such research, it highlights that which endows the automatic game systems with the ability of identifying, in game footage, the events that other players,

interacting with them, provoke in the game environment. Thus, the main contribution of this work is the implementation of deep learning models to perform MIML game event classification in gameplay footage, which are composed of: a data generator script to automatically produce multi-labeled frames from game footage (where the labels correspond to game events); a pre-processing method to make the frames generated by the script suitable to be used in the training datasets; a fine-tuned MobileNetV2 to perform feature extraction (trained from the pre-processed frames); an algorithm to produce MIML samples from the pre-processed frames (each sample corresponds to a set of frames named chunk); a deep neural network (NN) to perform classification of game events, which is trained from the chunks. In this investigation, Super Mario Bros is used as a case study.

Complete Paper #234

Robust Denoising and DenseNet Classification Framework for Plant Disease Detection

Kevin Zhou¹ and Dimah Dera²

¹ *Electrical and Computer Engineering, The University of Texas Rio Grande Valley, Brownsville, TX 78520 U.S.A.*

² *Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623 U.S.A.*

Keywords: Plant Disease Detection, DenseNet Image Classification, Robust Machine Learning, Denoising Neural Networks.

Abstract: Plant disease is one of many obstacles encountered in the field of agriculture. Machine learning models have been used to classify and detect diseases among plants by analyzing and extracting features from plant images. However, a common problem for many models is that they are trained on clean laboratory images and do not exemplify real conditions where noise can be present. In addition, the emergence of adversarial noise that can mislead models into wrong predictions poses a severe challenge to developing preserved models against noisy environments. In this paper, we propose an end-to-end robust plant disease detection framework that combines a DenseNet-based classification with a vigorous deep learning denoising model. We validate a variety of deep learning denoising models and adopt the Real Image Denoising network (RIDnet). The experiments have shown that the proposed denoising classification framework for plant disease detection is more robust against noisy or corrupted input images compared to a single classification model and can also successfully defend against adversarial noises in images.

Complete Paper #317

Facial Point Graphs for Amyotrophic Lateral Sclerosis Identification

Nicolas Gomes¹, Arissa Yoshida¹, Mateus Roder¹, Guilherme Camargo de Oliveira^{1,2} and João Papa¹

¹ *Department of Computing, São Paulo State University (UNESP), Brazil*

² *School of Engineering, Royal Melbourne Institute of Technology (RMIT), Australia*

Keywords: Neurodegenerative Disease, ALS, Graph Neural Networks, Facial Point Graph.

Abstract: Identifying Amyotrophic Lateral Sclerosis (ALS) in its early stages is essential for establishing the beginning of treatment, enriching the outlook, and enhancing the overall well-being of those affected individuals. However, early diagnosis and detecting the disease's signs is not straightforward. A simpler and cheaper way arises by analyzing the patient's facial

expressions through computational methods. When a patient with ALS engages in specific actions, e.g., opening their mouth, the movement of specific facial muscles differs from that observed in a healthy individual. This paper proposes Facial Point Graphs to learn information from the geometry of facial images to identify ALS automatically. The experimental outcomes in the Toronto Neuroface dataset show the proposed approach outperformed state-of-the-art results, fostering promising developments in the area.

Complete Paper #315

On the Use of Visual Transformer for Image Complexity Assessment

Luigi Celona, Gianluigi Ciocca and Raimondo Schettini

Department of Informatics, Systems and Communication, University of Milano-Bicocca, viale Sarca 336, 20126 Milano, Italy

Keywords: Image Complexity, Feature Extraction, Self-Supervised, Supervised, Transfer Learning, Vision Transformers.

Abstract: Perceiving image complexity is a crucial aspect of human visual understanding, yet explicitly assessing image complexity poses challenges. Historically, this aspect has been understudied due to its inherent subjectivity, stemming from its reliance on human perception, and the semantic dependency of image complexity in the face of diverse real-world images. Different computational models for image complexity estimation have been proposed in the literature. These models leverage a variety of techniques ranging from low-level, hand-crafted features, to advanced machine learning algorithms. This paper explores the use of recent deep-learning approaches based on Visual Transformer to extract robust information for image complexity estimation in a transfer learning paradigm. Specifically, we propose to leverage three visual backbones, CLIP, DINO-v2, and ImageNetViT, as feature extractors, coupled with a Support Vector Regressor with Radial Basis Function kernel as an image complexity estimator. We test our approach on two widely used benchmark datasets (i.e. IC9600 and SAVOIAS) in an intra-dataset and inter-dataset workflow. Our experiments demonstrate the effectiveness of the CLIP-based features for accurate image complexity estimation with results comparable to end-to-end solutions.

Complete Paper #410

Enhancement-Driven Pretraining for Robust Fingerprint Representation Learning

Ekta Gavas¹, Kaustubh Olpadkar² and Anoop Namboodiri¹

¹ *Centre for Visual Information Technology, International Institute of Information Technology, Hyderabad, India*

² *Stony Brook University, U.S.A.*

Keywords: Fingerprint Representation Learning, Fingerprint Verification, Self-Supervised Learning, Deep Learning.

Abstract: Fingerprint recognition stands as a pivotal component of biometric technology, with diverse applications from identity verification to advanced search tools. In this paper, we propose a unique method for deriving robust fingerprint representations by leveraging enhancement-based pre-training. Building on the achievements of U-Net-based fingerprint enhancement, our method employs a specialized encoder to derive representations from fingerprint images in a self-supervised manner. We further refine these representations, aiming to enhance the verification capabilities. Our experimental results, tested on publicly available fingerprint datasets, reveal a marked improvement in verifica-

tion performance against established self-supervised training techniques. Our findings not only highlight the effectiveness of our method but also pave the way for potential advancements. Crucially, our research indicates that it is feasible to extract meaningful fingerprint representations from degraded images without relying on enhanced samples.

Oral Presentations (Online) 2

14:00 - 16:00

Deep Learning for Visual Understanding

VISAPP

Room Online 1

Complete Paper #155

Image Inpainting on the Sketch-Pencil Domain with Vision Transformers

Jose Campana, Luís Decker, Marcos Souza, Helena Maia and Helio Pedrini

Institute of Computing, University of Campinas, Campinas, SP, 13083-852, Brazil

Keywords: Image Inpainting, Sketch-Pencil, Image Processing, Transformers.

Abstract: Image inpainting aims to realistically fill missing regions in images, which requires both structural and textural understanding. Traditionally, methods in the literature have employed Convolutional Neural Networks (CNN), especially Generative Adversarial Networks (GAN), to restore missing regions in a coherent and reliable manner. However, CNNs' limited receptive fields can sometimes result in unreliable outcomes due to their inability to capture the broader context of the image. Transformer-based models, on the other hand, can learn long-range dependencies through self-attention mechanisms. In order to generate more consistent results, some approaches have further incorporated auxiliary information to guide the model's understanding of structural information. In this work, we propose a new method for image inpainting that uses sketch-pencil information to guide the restoration of structural, as well as textural elements. Unlike previous works that employ edges, lines, or segmentation maps, we leverage the sketch-pencil domain and the capabilities of Transformers to learn long-range dependencies to properly match structural and textural information, resulting in more consistent results. Experimental results show the effectiveness of our approach, demonstrating either superior or competitive performance when compared to existing methods, especially in scenarios involving complex images and large missing areas.

Complete Paper #384

Single-Class Instance Segmentation for Vectorization of Line Drawings

Rhythm Vohra, Amanda Dash and Alexandra Branzan Albu

University of Victoria, Canada

Keywords: Segmentation, Visual Attention, Image Vectorization.

Abstract: Images can be represented and stored either in raster or in vector formats. Raster images are most ubiquitous and are defined as matrices of pixel intensities/colours, while vector images consist of a finite set of geometric primitives, such as lines, curves, and polygons. Since geometric shapes are expressed via mathematical equations and defined by a limited number of control points, they can be manipulated in a much easier way than by directly working with pixels; hence, the vector format is much preferred to raster for image editing and understanding purposes.

The conversion of a raster image into its vector correspondent is a non-trivial process, called image vectorization. This paper presents a vectorization method for line drawings, which is much faster and more accurate than the state-of-the-art. We propose a novel segmentation method that processes the input raster image by labeling each pixel as belonging to a particular stroke instance. Our contributions consist of a segmentation model (called Multi-Focus Attention UNet), as well as a loss function that handles well infrequent labels and yields outputs which capture accurately the human drawing style.

Complete Paper #78

Data Quality Aware Approaches for Addressing Model Drift of Semantic Segmentation Models

Samiha Mirza, Vuong Nguyen, Pranav Mantini and Shishir Shah

Quantitative Imaging Lab, Dept. of Computer Science, University of Houston, Texas, U.S.A.

Keywords: Model Drift, Semantic Segmentation, Image Quality Assessment Metrics, Feature Learning, Data Selection, Quality-Aware Models.

Abstract: In the midst of the rapid integration of artificial intelligence (AI) into real world applications, one pressing challenge we confront is the phenomenon of model drift, wherein the performance of AI models gradually degrades over time, compromising their effectiveness in real-world, dynamic environments. Once identified, we need techniques for handling this drift to preserve the model performance and prevent further degradation. This study investigates two prominent quality aware strategies to combat model drift: data quality assessment and data conditioning based on prior model knowledge. The former leverages image quality assessment metrics to meticulously select high-quality training data, improving the model robustness, while the latter makes use of learned feature vectors from existing models to guide the selection of future data, aligning it with the model's prior knowledge. Through comprehensive experimentation, this research aims to shed light on the efficacy of these approaches in enhancing the performance and reliability of semantic segmentation models, thereby contributing to the advancement of computer vision capabilities in real-world scenarios.

Complete Paper #303

Transformer-Based Video Saliency Prediction with High Temporal Dimension Decoding

Morteza Moradi, Simone Palazzo and Concetto Spampinato

PeRCeive Lab, University of Catania, Italy

Keywords: Video Saliency Prediction, Gaze Prediction, Visual Attention, Spatio-Temporal Transformer.

Abstract: In recent years, finding an effective and efficient strategy for exploiting spatial and temporal information has been a hot research topic in video saliency prediction (VSP). With the emergence of spatio-temporal transformers, the weakness of the prior strategies, e.g., 3D convolutional networks and LSTM-based networks, for capturing long-range dependencies has been effectively compensated. While VSP has drawn benefits from spatio-temporal transformers, finding the most effective way for aggregating temporal features is still challenging. To address this concern, we propose a transformer-based video saliency prediction approach with high temporal dimension decoding network (THTD-Net). This strategy accounts for the lack of complex

hierarchical interactions between features that are extracted from the transformer-based spatio-temporal encoder: in particular, it does not require multiple decoders and aims at gradually reducing temporal features' dimensions in the decoder. This decoder-based architecture yields comparable performance to multi-branch and over-complicated models on common benchmarks such as DHF1K, UCF-sports and Hollywood-2.

Complete Paper #430

Decoding Visual Stimuli and Visual Imagery Information from EEG Signals Utilizing Multi-Perspective 3D-CNN Based Hierarchical Deep-Fusion Learning Network

Fatma Emanet and Kazim Sekeroglu

Department of Computer Science, Southeastern Louisiana University, Hammond, LA 70402, U.S.A.

Keywords: Hierarchical Deep Learning, Brain Computer Interface, Fusion Learning, Spatiotemporal Pattern Recognition, Multi-Perspective Learning.

Abstract: Brain-Computer Interface Systems (BCIs) facilitate communication between the brain and machines, enabling applications such as diagnosis, understanding brain function, and cognitive augmentation. This study explores the classification of visual stimuli and visual imagery using electroencephalographic (EEG) data. The proposed method utilizes 3D EEG data generated by transforming 1D EEG data into 2D Spatiotemporal EEG image mappings for feature extraction and classification. Additionally, a multi-perspective 3D CNN-based hierarchical deep fusion learning network is employed to classify multi-dimensional spatiotemporal EEG data, decoding brain activity for visual and visual imagery stimulation. The findings show that the suggested multi-perspective fusion method performs better than a standalone model, indicating promising progress in using BCIs to understand and utilize brain signals for visual and imagined stimulation.

Poster Session 1
16:00 - 17:00

GRAPP
Foyer

Complete Paper #73

ML-Tree and MRL-Tree: Combining Mass-Spring System, Rigid-Body Dynamics and L-Systems to Model Physical Effects on Trees

See Min Lim¹ and Like Gobeawan²

¹ *National University of Singapore, 21 Lower Kent Ridge Road, Singapore*

² *Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way #16-16 Connexis, 138632, Singapore*

Keywords: Physically-Based Modeling, Tree Modeling, Root Modeling, L-Systems, MSS, RBD.

Abstract: We apply physically-based modeling methods to a biological tree model for a hybrid model with a better accuracy. Physical aspects of tree growth, such as wind, tropisms, gravity and soil resistance are modelled. The hybrid model also includes the handling of boundary conditions such as momentum conservation and switching between the L-System, Mass-Spring System (MSS) and Rigid-Body Dynamics (RBD) methods. This paper demonstrates resulting models and their potential applications such as tree stress prediction.

Complete Paper #91

Pure Physics-Based Hand Interaction in VR

Mohammed-Bashir Mahdi¹, Erwan Guillou¹, Alexandre Meyer¹, Arash Habibi² and Saïda Bouakaz¹

¹ LIRIS, CNRS, Univ. Lyon, Université Claude Bernard Lyon 1, 69100 Villeurbanne, France

² ICube, CNRS, Université de Strasbourg, 300 bd Sébastien Brant, 67400 Illkirch, France

Keywords: Virtual Reality, Interaction, Hooke's Spring Law, Physics, Physical Interaction, Hands, Natural Interaction, Visual Feedback.

Abstract: Interaction in Virtual Reality is still mainly done using controllers. However, since the early 2000s, there has been a desire to find a way to interact within the virtual environment using only our hands. Pinch motion detection was introduced to detect a grasping action. Since then, hands' motion capture has been highly developed until being directly integrated in VR headsets. Thus, multiple research projects were done in order to exploit this technology for better grasping techniques. Recent works tend to bring physical hand interaction to VR. However, they introduce physical heuristics to determine if an object is grasped or not, but in reality, motion is purely kinematic. In our paper, we introduce a purely physical method based on Hooke's spring law that eliminates the need for a grasping mode. Additionally, we incorporate visual feedback methods to compensate for the absence of the sense of touch. Consequently, with this approach, we can lift objects, throw them, stack them, and interact with them naturally. We carried out extensive tests with several people who had no previous experience, to validate our technique.

Complete Paper #226

Automatic Registration of 3D Point Cloud Sequences

Natálie Vítová, Jakub Frank and Libor Váša

Department of Computer Science And Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 301 00 Plzeň, Czech Republic

Keywords: Registration, Point Cloud, Dynamic, Series, Kinect, Sensor, Alignment.

Abstract: Surface registration is a well-studied problem in computer graphics and triangle mesh processing. A plethora of approaches exists that align a partial 3D view of a surface to another, which is a central task in 3D scanning, where usually each scan only provides partial information about the shape of the scanned object due to occlusion. In this paper, we address a slightly different problem: a pair of depth cameras is observing a dynamic scene, each providing a sequence of partial scans. The scanning devices are assumed to remain in a constant relative position throughout the process, and therefore there exists a single rigid transformation that aligns the two sequences of partial meshes. Our objective is to find this transformation based on the data alone, i.e. without using any specialized calibration tools. This problem can be approached as a set of static mesh registration problems; however, such an interpretation leads to problems when enforcing a single global solution. We show that an appropriate modification of a previously proposed consensus-based registration algorithm is a more viable solution that exploits information from all the frames simultaneously and naturally leads to a single global solution.

Poster Session 1
16:00 - 17:00

IVAPP
Foyer

Complete Paper #16

Flowstrates++: An Approach to Visualize Multi-Dimensional OD Data

Nicolas Fuchs, Pierre Vanhulst, Raphaël Tuor and Denis Lalanne

Human-IST Institute, Université of Fribourg, Boulevard de Pérolles 90, Fribourg, Switzerland

Keywords: Human-Centered Computing, Visualization Design and Techniques, Evaluation Methods.

Abstract: Is it possible to visualize complex Origin-Destination (OD) data along with relevant spatio-temporal data? In this paper, we tackle this issue by presenting Flowstrates++, an augmented version of Flowstrates which aims to visualize additional time-series datasets linked with OD data. On top of Flowstrates' heatmap, we designed a second heatmap for spatio-temporal data, synchronized on the temporal axis, as well as other dataset comparison features. Two versions of Flowstrates++ have been designed and implemented: Switch, that displays one external dataset at a time, and Combi (for "combined"), that displays two external datasets at the same time. We aimed to assess to which extent both variants spur users into making multidimensional findings. To achieve this goal, we evaluated both variants with ninety participants: ten were pilot users in live remote sessions, and eighty were provided by Prolific.co, a crowd-sourcing platform. In a within-groups study, these participants were asked to take relevant annotations about the data on both variants, and to evaluate them through a survey. We then classified the annotations using a framework whose validity was evaluated with an Intercoder Agreement and Fleiss' Kappa. We found that the Combi variant yielded consistently better results, both in terms of number of produced multidimensional annotations, and in terms of appreciation of the participants. Yet regardless of the variant, our solution allows users to highlight potential correlations between time-series data and temporal OD data.

Complete Paper #141

Dashboard Design: Interactive and Visual Exploration of Spotify Songs

Sarah Clavadetscher, Michael Schlotter, Nadine Christen, Juliane Streitberg and Michael Burch

University of Applied Sciences, Chur, Switzerland

Keywords: Dashboard Design, Visualization Tool, Interactive Information Visualization, Spotify.

Abstract: In this paper we describe an approach to create interactive visualization tools for simple datasets that exist in various application domains which many people are familiar with and interested in, like sports, entertainment, traffic, or health care. Such data problems require a simple but elegant visual solution to support the non-experts in information visualization at their tasks at hand, supported by easy-to-understand interaction techniques. We start our approach with the design phase in which a hand-drawn mockup is created and based on this, an interactive dashboard in Dash, Plotly, and Python is built. The design of the tool is guided by user feedback of 23 participants in qualitative interviews taking into account eight relevant criteria before starting the design of a visualization tool. We illustrate the usefulness of the tool by applying it to a dataset focusing on songs from

the music streaming platform Spotify while we integrate several diagrams in a multiple and coordinated views manner to visually explore a given dataset based on several visual perspectives. With the combination of the many diagrams we can find insights in the mood categories of the songs and several other attributes, hence allowing visual analyses and explorations. Finally, we discuss limitations and scalability issues of the approach.

Complete Paper #444

Simultaneous Optimization of Edge Bundling and Node Layout Using Genetic Algorithm

Junsei Meikari¹ and Ryosuke Saga²

¹ Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Sakai, Japan

² Graduate School of Informatics, Osaka Metropolitan University, Sakai, Japan

Keywords: Visualization, Edge Bundling, Node Layout, Genetic Algorithm.

Abstract: This paper describes an evolutionary visualization method that performs edge bundling during the execution of the genetic algorithm. There are several node layout algorithms and edge bundling, however, there are no methods considering both algorithms simultaneously. This paper proposes an algorithm to optimize the fitness function of GABEB, which is genetic algorithm-based edge bundling, and Zhang's node layout simultaneously. The experiments for the sample graphs show the improved result from the viewpoints of several evaluation criteria.

Poster Session 1
16:00 - 17:00

HUCAPP
Foyer

Abstract #30

Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality

Dawid Ratajczyk, Monika Żyła, Piort Jaworski and Paweł Łupkowski

Adam Mickiewicz University in Poznan, ul. Wieniawskiego 1, Poznań, Poland

Keywords: Uncanny Valley, Virtual Reality, Emotion Recognition, Mind Attribution, Agent Appearance, Humanlike Agents, Eeriness.

Abstract: The study aimed to investigate the joint contributions of appearance and type of mind attributed to agents on the uncanny valley (UV) effect. An experiment was conducted using an ecologically valid virtual reality environment to improve the validity of the results and gain a more comprehensive understanding of the UV effect. The study used a 2 × 2 experimental design, manipulating both the agent's appearance (robot or human) and the identity behind the character (artificial intelligence or a user's mind). The final sample consisted of 116 participants. Self-reported eeriness and likability, as well as, electrodermal and heart activities were measured. It was found that the appearance of the agent was crucial in determining users' perceptions of humanlikeness and feelings of eeriness, but the type of mind attributed to the agent did not significantly affect feelings of eeriness. Additionally, the interaction between appearance and type of mind attributed to the agent influenced behavioral realism, which, in turn, affected likability. People perceive other users with more humanlike avatars and artificial intelligence presented as robots as more realistic and, thus, more likable. We discuss the possibility that low skin

conductance response and feelings of eeriness are attributed to the inability to recognize emotion, which we identified as a possible cause for uncanny ratings. Our experiment suggests that cues related to the ability to experience do not increase feelings of eeriness themselves, but rather a violation of the expectancy regarding these cues increases feelings of eeriness. Our study provides insight into the factors that contribute to human-robot interactions and highlights the importance of appearance in designing effective and likable artificial agents.

Complete Paper #42

Hôsea: A Touch Table for Cognitive and Motor Rehabilitation for the Elderly - A Preliminary Study

Maxime Macé^{1,2,3}, Lise Macé⁴, Emmanuelle Ménétrier⁵, Paul Richard³ and Tassadit Amghar²

¹ KARA Technology, Saint-Barthélemy-d'Anjou, France

² LERIA SFR Mathstic, University of Angers, 2 Bd. Lavoisier, Angers, France

³ LARIS SFR Mathstic, University of Angers, 62 Av. Notre-Dame du Lac, Angers, France

⁴ Department of Psychology, University of Nantes, Nantes, France

⁵ LPPL SFR Confluence, University of Angers, 2 Bd. Lavoisier, Angers, France

Keywords: Touch Table, Elderly, Cognitive Decline, Rehabilitation, Usability.

Abstract: As the population ages, it is becoming increasingly important to offer technological solutions for cognitive-motor rehabilitation. In this context, we have designed Hôsea, a new touch table offering an attractive, accessible and stimulating interface for the elderly. The table integrates various motor and cognitive exercises. The study presented in this article examines the ease of use of the table, as well as the motivation and effort perceived by elderly people. To achieve this, we engaged with 43 elderly individuals in good health, presenting them with a 15-minute session involving three games accessible on the table. At the end of the session, participants were administered standardized scales, evaluating the usability of the table via the F-SUS, the degree of perceived effort with the NASA-TLX, and their motivation with the SIMS. The results suggest that the Hôsea touch table offers a user-friendly and motivating environment. These results motivate further work on a personalized rehabilitation program.

Complete Paper #285

ARTISTA: Redefining Pottery Design with Virtual Reality and Physically Simulated Clay

Gilda Manfredi¹, Gabriele Gilio¹, Nicola Capece¹, Ugo Erra¹ and Vincenzo Baldi²

¹ University of Basilicata, Department of Mathematics, Computer Science and Economics, Potenza 85100, Italy

² SAUTECH S.R.L., Cava de' Tirreni SA 84013, Italy

Keywords: Applied Computing Arts and Humanities, Computing Methodologies Virtual Reality, Computing Methodologies Modeling and Simulation, Human-Centered Computing Virtual Reality.

Abstract: Virtual Reality (VR) applications have gained significant popularity in various fields, including design and manufacturing. This paper introduces a novel VR application that simulates a 3D clay printer. The application utilizes the Oculus Quest 2 Head Mounted Display (HMD), enabling users to create virtual pots by

manoeuvring a virtual extruder that extrudes and deposits strands of clay in layers. Two features define this application: incorporating a virtual coiling technique, offering guided assistance throughout the pottery prototyping process, and integrating physical simulation that provides users with real-time feedback on the final results achievable using a real 3D clay printer. Users can now design and iterate virtually without physically printing, reducing material waste and improving the overall creative process. Moreover, this application saves the information necessary to create the designed pottery piece with a 3D clay printer. We present a preliminary version that enables artisans in VR to craft unique, handmade virtual pottery pieces, fostering creativity and enabling the exploration of many design possibilities.

Complete Paper #350

Assessment of the Detectability of Vulnerable Road Users: An Empirical Study

Wentong Yang¹, Shota Matsubayashi², Kazuhisa Miwa¹, Shinya Kitayama³, Manabu Otsuka³ and Koji Hamada³

¹ Graduate School of Informatics, Nagoya University, Nagoya, 4648601, Japan

² Institutes of Innovation for Future Society (InFuS), Nagoya University, Nagoya, 4648601, Japan

³ Advanced Research and Innovation Center, DENSO Corporation, Aichi, 4700111, Japan

Keywords: Traffic Safety, Advanced Driver Assistance System, Driver Cognition, Collision Risk, Vulnerable Road Users.

Abstract: This study analysed the detectability of vulnerable road users (VRUs) as a function of time to the closest point of approach. We defined four patterns: Gradual Increasing Pattern and High Detectability Pattern as the safe situation, Immediate Increasing Pattern and Low Detectability Pattern as the dangerous situation and investigated empirically drivers' detection patterns. The results showed that motorbikes in the same direction showed the dangerous pattern, and pedestrians in the same direction also showed the dangerous patterns but influenced by the distance of closest point of approach. Bicycles in the same direction showed higher detectability due to their positioning in the driver's field of view for longer time. For the VRUs in the opposite direction, and those in the left and right direction, participants also showed high detectability. The results give implications for advanced driver assistance systems (ADAS) design.

Complete Paper #409

Visual Behavior Based on Information Foraging Theory Toward Designing of Auditory Information

Yuta Kurihara¹, Motoki Shino², Katsuko Nakahira³ and Muneo Kitajima³

¹ Department of Human & Engineered Environmental Studies, The University of Tokyo, Kashiwanoha, Kashiwa, Chiba, Japan

² Department Mechanical Engineering, Tokyo Institute of Technology, Ookayama, Tokyo, Japan

³ Department of Information & Management Systems Engineering, Nagaoka University of Technology, Nagaoka, Niigata, Japan

Keywords: Memory, Audio Guide, Visual Behavior, Information Foraging Theory, Cognitive Model.

Abstract: Knowledge acquisition through appreciation behavior is a commonly experienced phenomenon. Appreciation behavior is characterized by real-time processing of information input through

the five senses. This study focuses on multimodal information processing triggered during appreciation behavior, aiming to enhance knowledge acquisition, i.e., learning, by appropriately designing the provided information. While information during appreciation is presented as visual and auditory stimuli, learning is assumed to occur through the memorization of the content of the auditory information provided. Appreciation behavior is measured as visual behavior and modeled based on the well-established theory of information foraging. According to the information foraging theory, the process leading to information acquisition involves two states: the foraging state, where individuals actively seek information sources, and the acquisition transition state, where attention is directed towards information sources for acquiring information. Based on this theory, the characteristics of visual behavior are extracted for foraging and acquisition transition behaviors. This paper suggested that foraging state can be discerned by setting a threshold for gaze point movement frequency, while the acquisition transition state can be clearly delineated by examining the movement patterns of central and peripheral vision until reaching acquisition.

Complete Paper #428

A Comparative Study: Augmented and Virtual Reality Applications for Improving Comprehension of Abstract Programming Concepts

Omer Cinar, Karen Rafferty, David Cutting and Hui Wang

School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, U.K.

Keywords: Augmented Reality, Booklet, Education, Engineering, Programming, Python, Virtual Reality.

Abstract: This paper presents the development of a Virtual Reality (VR) application to enhance the learning experience of Python collection data types for electrical and electronic engineering students, and the comparison of this VR application with an Augmented Reality (AR) application developed for the same field, and a paper-based content. AR and VR, as emerging technologies, hold promise for transforming educational and cognitive processes, especially in engineering education where students often struggle with abstract programming concepts. The research focuses on using VR to make abstract concepts like Python collections more accessible and understandable through immersive and interactive experiences. The VR application allows students to engage with virtual representations of Python data types, promoting a deeper comprehension. A comparative user study was conducted, involving participants using either the VR or AR application or learning through a printed booklet. The effectiveness of these methods was assessed by pre- and post-comprehension tests. The results indicated a significant improvement in understanding Python collections with the VR application, as evidenced by the substantial difference between its pre- and post-test score means. This finding suggests that VR, more effectively than AR or traditional methods, aids in grasping abstract programming concepts.

Poster Session 1
16:00 - 17:00

VISAPP
Foyer

Complete Paper #27

Synthetic Data-Driven Approach for Missing Nut and Bolt Classification in Flange Joints

Frankly Toro, Hassane Trigui, Yazeed Alnumay, Siddharth Mishra and Sahejad Patel
Aramco, Thuwal, Saudi Arabia

Keywords: Asset Integrity, Flange Joints, Missing Bolt/Nut, Missing Nut, Multi-View CNN, Synthetic Data, Domain Adaptation, Image Classification, Grad-CAM.

Abstract: Inspection of bolted flange joints is a routine procedure typically done manually in process-based industries. However, this is a time-consuming task since there are many flanges in a typical operational facility. We present a computer vision-based tool that can be integrated into other systems to enable automated inspection of these flanges. We propose a multi-view image classification architecture for detecting a missing bolt or nut in a flange joint image. To guide the training process, a synthetic dataset with 60,000 image pairs was created to simulate realistic environmental conditions of flange joints. To demonstrate the effectiveness of our approach, an additional real-world dataset of 1,080 flange joint image pairs was manually collected. The proposed approach achieved remarkable performance in classifying missing bolt instances with an accuracy of 95.28% and 95.14% for missing nut instances.

Complete Paper #28

Finding and Navigating to Humans in Complex Environments for Assistive Tasks

Asfand Yaar¹, Antonino Furnari¹, Marco Rosano¹, Aki Härmä² and Giovanni Farinella¹

¹ *Department of Mathematics and Computer Science, University of Catania, Italy*

² *Department of Advanced Computing Sciences, Maastricht University, The Netherlands*

Keywords: Assistive AI, Autonomous Exploration, Human-Robot Interaction, Embodied Navigation, Habitat Simulator.

Abstract: Finding and reaching humans in unseen environments is a major challenge for intelligent agents and social robots. Effective exploration and navigation strategies are necessary to locate the human performing various activities. In this paper, we propose a problem formulation in which the robot is required to locate and reach humans in unseen environments. To tackle this task, we design an approach that makes use of state-of-the-art components to allow the agent to explore the environment, identify the human's location on the map, and approach them while maintaining a safe distance. To include human models, we utilized Blender to modify the scenes of the Gibson dataset. We conducted experiments using the Habitat simulator, where the proposed approach achieves promising results. The success of our approach is measured by the distance and orientation difference between the robot and the human at the end of the episode. We will release the source code and 3D human models for researchers to benchmark their assistive systems.

Complete Paper #36

Towards Better Morphed Face Images Without Ghosting Artifacts

Clemens Seibold¹, Anna Hilsmann¹ and Peter Eisert^{1,2}

¹ *Fraunhofer HHI, Berlin, Germany*

² *Humboldt University of Berlin, Berlin, Germany*

Keywords: Face Morphing Attacks, Ghosting Artifact Prevention, Morphed Face Images Dataset.

Abstract: Automatic generation of morphed face images often produces ghosting artifacts due to poorly aligned structures in the input images. Manual processing can mitigate these artifacts. However, this is not feasible for the generation of large datasets, which are required for training and evaluating robust morphing attack detectors. In this paper, we propose a method for automatic prevention of ghosting artifacts based on a pixel-wise alignment during morph generation. We evaluate our proposed method on state-of-the-art detectors and show that our morphs are harder to detect, particularly, when combined with style-transfer-based improvement of low-level image characteristics. Furthermore, we show that our approach does not impair the biometric quality, which is essential for high quality morphs.

Complete Paper #47

Teeth Localization and Lesion Segmentation in CBCT Images Using SpatialConfiguration-Net and U-Net

Arnela Hadzic¹, Barbara Kirnbauer², Darko Štern³ and Martin Urschler¹

¹ *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria*

² *Department of Dental Medicine and Oral Health, Medical University of Graz, Graz, Austria*

³ *Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria*

Keywords: Teeth Localization, Lesion Segmentation, SpatialConfiguration-Net, U-Net, CBCT, Class Imbalance.

Abstract: The localization of teeth and segmentation of periapical lesions in cone-beam computed tomography (CBCT) images are crucial tasks for clinical diagnosis and treatment planning, which are often time-consuming and require a high level of expertise. However, automating these tasks is challenging due to variations in shape, size, and orientation of lesions, as well as similar topologies among teeth. Moreover, the small volumes occupied by lesions in CBCT images pose a class imbalance problem that needs to be addressed. In this study, we propose a deep learning-based method utilizing two convolutional neural networks: the SpatialConfiguration-Net (SCN) and a modified version of the U-Net. The SCN accurately predicts the coordinates of all teeth present in an image, enabling precise cropping of teeth volumes that are then fed into the U-Net which detects lesions via segmentation. To address class imbalance, we compare the performance of three reweighting loss functions. After evaluation on 144 CBCT images, our method achieves a 97.3% accuracy for teeth localization, along with a promising sensitivity and specificity of 0.97 and 0.88, respectively, for subsequent lesion detection.

Complete Paper #58

Multi-View 3D Reconstruction for Construction Site Monitoring

Guangan Chen, Michiel Vlamincx, Wilfried Philips and Hiep Luong

Image Processing and Interpretation (IPI), imec Research Group at Ghent University, Department of Telecommunications and Information Processing (TELIN), Ghent University, Belgium

Keywords: Construction Progress Monitoring, Structure-from-Motion, Multi-View Stereo, Point Cloud Registration.

Abstract: Monitoring construction sites is pivotal in effective construction management. Building Information Modeling (BIM) is vital for creating detailed building models and comparing actual construction with planned designs. For this comparison, a 3D model of the building is often generated using images captured by handheld cameras or Unmanned Aerial Vehicles. However, this approach does not provide real-time spatial monitoring of on-site activities within the BIM model. To address this challenge, our study utilizes fixed cameras placed at predetermined locations within an actual construction site. We captured images from these fixed viewpoints and used classical multi-view stereo techniques to create a 3D point cloud representing the as-built building. This point cloud is then aligned with the as-planned BIM model through point cloud registration. In addition, we proposed an algorithm to convert SfM reprojection error into a value with metric units, resulting in a mean SfM reprojection error of 4.17cm. We also created voxel volumes to track and visualize construction activities within BIM coordinate system, enhancing real-time site monitoring and improving construction management.

Complete Paper #60

Pedestrian's Gaze Object Detection in Traffic Scene

Hiroto Murakami¹, Jialei Chen¹, Daisuke Deguchi¹, Takatsugu Hirayama^{1,2}, Yasutomo Kawanishi^{1,3} and Hiroshi Murase¹

¹ Graduate School of Informatics, Nagoya University, Nagoya, Japan

² Faculty of Environmental Science, University of Human Environments, Okazaki, Japan

³ Multimodal Data Recognition Research Team, Guardian Robot Project, Riken, Kyoto, Japan

Keywords: Pedestrian's Gaze Object Detection, Object Detection, Gaze Estimation, Traffic Scene, Dataset.

Abstract: In this paper, we present a new task of detecting an object that a target pedestrian is gazing at in a traffic scene called PEdestrian's Gaze Object (PEGO). We argue that the detection of gaze object can provide important information for pedestrian's behavior prediction and can contribute to the realization of automated vehicles. For this task, we construct a dataset of in-vehicle camera images with annotations of the objects that pedestrians are gazing at. Also, we propose a Transformer-based method called PEGO Transformer to solve the PEGO detection task. The PEGO Transformer directly performs gaze object detection with the utilization of whole-body features without a high-resolution head image and a gaze heatmap which the traditional methods rely on. Experimental results showed that the proposed method could estimate pedestrian's gaze object accurately even if various objects exist in the scene.

Complete Paper #64

Improved Pest Detection in Insect Larvae Rearing with Pseudo-Labeling and Spatio-Temporal Masking

Paweł Majewski¹, Piotr Lampa², Robert Burduk¹ and Jacek Reiner²

¹ Faculty of Information and Communication Technology, Wrocław University of Science and Technology, Poland

² Faculty of Mechanical Engineering, Wrocław University of Science and Technology, Poland

Keywords: Pseudo-Labeling, Spatio-Temporal, Optical Flow, Object Detection, Insect, Monitoring, Tenebrio Molitor.

Abstract: Pest detection is an important application problem as it enables early reaction by the farmer in situations of unacceptable pest infestation. Developing an effective pest detection model is challenging due to the problem of creating a representative dataset, as episodes of pest occurrence under real rearing conditions are rare. Detecting the pest *Alphitobius diaperinus* Panzer in mealworm (*Tenebrio molitor*) rearing, addressed in this work, is particularly difficult due to the relatively small size of detection objects, the high similarity between detection objects and background elements, and the dense scenes. Considering the problems described, an original method for developing pest detection models was proposed. The first step was to develop a basic model by training it on a small subset of manually labelled samples. In the next step, the basic model identified low/moderate pest-infected rearing boxes from many boxes inspected daily. Pseudo-labelling was carried out for these boxes, significantly reducing labelling time, and re-training was performed. A spatio-temporal masking method based on activity maps calculated using the Gunnar-Farneback optical flow technique was also proposed to reduce the numerous false-positive errors. The quantitative results confirmed the positive effect of pseudo-labelling and spatio-temporal masking on the accuracy of pest detection and the ability to recognise episodes of unacceptable pest infestation.

Complete Paper #71

Online Human Activity Recognition Using Efficient Neural Architecture Search with Low Environmental Impact

Nassim Mokhtari, Alexis Nédélec, Marlène Gilles and Pierre De Loor

Lab-STICC (CNRS UMR 6285), ENIB, Centre Européen de Réalité Virtuelle, Brest, France

Keywords: 3D Skeleton Data, Image Encoding, Online Human Activity Recognition, Deep Learning, Neural Architecture Search, FireFly Algorithm, NAS-BENCH-101, Efficiency Estimation, Energy Consumption.

Abstract: Human activity recognition using sensor data can be approached as a problem of classifying time series data. Deep learning models allow for great progress in this domain, but there are still some areas for improvement. In addition, the environmental impact of deep learning is a problem that must be addressed in today's machine learning studies. In this research, we propose to automate deep learning model design for human activity recognition by using an existing training-free Neural Architecture Search method. By this way, we decrease the time consumed by classical NAS approaches (GPU based) by a factor of 470, and the energy consumed by a factor of 170. Finally, We propose a new criterion to estimate the relevance of a deep

learning model based on a balance between both performance and computational cost. This criterion allows to reduce the size of neural architectures by preserving its capacity to recognize human activities.

Complete Paper #74

Towards Rapid Prototyping and Comparability in Active Learning for Deep Object Detection

Tobias Riedlinger¹, Marius Schubert², Karsten Kahl², Hanno Gottschalk¹ and Matthias Rottmann²

¹ *Institute of Mathematics, Technical University Berlin, Germany*

² *School of Mathematics and Natural Sciences, IZMD, University of Wuppertal, Germany*

Keywords: Active Learning, Deep Object Detection, Prototyping.

Abstract: Active learning as a paradigm in deep learning is especially important in applications involving intricate perception tasks such as object detection where labels are difficult and expensive to acquire. Development of active learning methods in such fields is highly computationally expensive and time consuming which obstructs the progression of research and leads to a lack of comparability between methods. In this work, we propose and investigate a sandbox setup for rapid development and transparent evaluation of active learning in deep object detection. Our experiments with commonly used configurations of datasets and detection architectures found in the literature show that results obtained in our sandbox environment are representative of results on standard configurations. The total compute time to obtain results and assess the learning behavior can be reduced by factors of up to 14 compared to Pascal VOC and up to 32 compared to BDD100k. This allows for testing and evaluating data acquisition and labeling strategies in under half a day and contributes to the transparency and development speed in the field of active learning for object detection.

Complete Paper #83

A Study of Real World Information Mapping for Information Sharing Using Edge Devices

Takenori Hara and Hideo Saito

Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, 223-8522, Japan

Keywords: Digital Twin, Information Sharing, Edge Device.

Abstract: We are developing a service platform to achieve use cases such as finding people in trouble and sharing them with AR applications to ask for help from the people around them, finding information about a sale at a store and sharing it with a map service, or alerting the police when a crime occurs, as well as quickly alerting the people around them. Our system detects "objects" and "events" in the real world using edge devices that can share the detected information in cyberspace so that the users can remotely obtain real-world information from distant locations. We believe our system can make life more safe and convenient. In this paper, we report on the findings from our experiments on searching for lost dogs by mapping object recognition results, on mapping bird habitat areas by environmental sound recognition results, and on mapping public facility announcement voices by speech recognition results.

Complete Paper #105

Sumo Action Classification Using Mawashi Keypoints

Yuki Utsuro¹, Hidehiko Shishido² and Yoshinari Kameda²

¹ *Master's Program in Intelligent and Mechanical Interaction Systems, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki, Japan*

² *Center for Computational Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki, Japan*

Keywords: Action Classification, Action Recognition, Extended Skeleton Model, Pose Estimation, Pose Sequences, Computer Vision in Sports, Sumo, Japanese Wrestling.

Abstract: We propose a new method of classification for kimarites in sumo videos based on kinematic pose estimation. Japanese wrestling sumo is a combat sport. Sumo is played by two wrestlers wearing a mawashi, a loincloth fastened around the waist. In a sumo match, two wrestlers grapple with each other. Sumo wrestlers perform actions by grabbing their opponents' mawashi. A kimarite is a sumo winning action that decides the outcome of a sumo match. All the kimarites are defined based on their motions. In an official sumo match, the kimarite of the match is classified by the referee, who oversees the classification just after the match. Classifying kimarites from videos by computer vision is a challenging task. There are two reasons. The first reason is that the definition of kimarites requires us to examine the relationship between the mawashi and the pose. The second reason is the heavy occlusion caused by the close contact between wrestlers. For the precise examination of pose estimation, we introduce a wrestler-specific skeleton model with mawashi keypoints. The relationship between mawashi and body parts is uniformly represented in the pose sequence with this extended skeleton model. As for heavy occlusion, we represent sumo actions as pose sequences to classify the sumo actions. Our method achieves an accuracy of 0.77 in action classification by LSTM. We confirmed that the skeleton model extension by mawashi keypoints improves the accuracy of action classification in sumo through the experiment results.

Complete Paper #189

Automatic Error Correction of GPT-Based Robot Motion Generation by Partial Affordance of Tool

Takahiro Suzuki, Yuta Ando and Manabu Hashimoto

Graduate School of Engineering, Chukyo University, Aichi, Japan

Keywords: Robot Motion Generation, Automatic Error Correction, Part Affordance, Functional Consistency, Robot Motion Template, GPT.

Abstract: In this research, we proposed a technique that, given a simple instruction such as "Please make a cup of coffee" as would commonly be used when one human gives another human an instruction, determines an appropriate robot motion sequence and the tools to be used for that task and generates a motion trajectory for a robot to execute the task. The proposed method uses a large language model (GPT) to determine robot motion sequences and tools to be used. However, GPT may select tools that do not exist in the scene or are not appropriate. To correct this error, our research focuses on function and functional consistency. An everyday object has a role assigned to each region of that object, such as "scoop" or "contain". There are also constraints such as the fact that a ladle must have scoop and grasp functions. The proposed method judges whether the tools in the scene are inconsistent with these constraints, and automatically corrects the tools as necessary. Experiment-

tal results confirmed that the proposed method was able to generate motion sequences a from simple instruction and that the proposed method automatically corrects errors in GPT outputs.

Complete Paper #259

SynthRSF: A Novel Photorealistic Synthetic Dataset for Adverse Weather Condition Denoising

Angelos Kanlis, Vazgken Vanian, Sotiris Karvarsamis, Ioanna Gkika, Konstantinos Konstantoudakis and Dimitrios Zarpalas

Visual Computing Lab (VCL), Information Technologies Institute (ITI), Centre for Research and Technology - Hellas (CERTH), Thessaloniki, Greece

Keywords: Synthetic Dataset, Image Restoration, Adverse Weather Conditions, Semantic Segmentation, Depth Estimation, Benchmarking, Unreal Engine.

Abstract: This paper presents the SynthRSF dataset for training and evaluating single-image rain, snow and haze denoising algorithms, as well as evaluating object detection, semantic segmentation, and depth estimation performance in noisy or denoised images. Our dataset features 26,893 noisy images, each accompanied by its corresponding ground truth image. It further includes 13,800 noisy images accompanied by ground truth, 16-bit depth maps and pixel-accurate annotations for various object instances in each frame. The utility of SynthRSF is assessed by training unified models for rain, snow, and haze removal, achieving good objective metrics and excellent subjective results compared to existing adverse weather condition datasets. Furthermore, we demonstrate its use as a benchmark for the performance of an object detection algorithm in weather-degraded image datasets.

Poster Presentations (Online) 1
16:00 - 17:00

VISIGRAPP
Room Online 1

Poster Presentations (Online) 1
16:00 - 17:00

IVAPP
Room Online 1

Complete Paper #246

Bringing Objects to Life: Supporting Program Comprehension Through Animated 2.5D Object Maps from Program Traces

Christoph Thiede, Willy Scheibel and Jürgen Döllner

Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, Germany

Keywords: Software Visualization, Software Maps, Program Comprehension, Omniscient Debugging.

Abstract: Program comprehension is a key activity in software development. Several visualization approaches such as software maps have been proposed to support programmers in exploring the architecture of software systems. However, for the exploration of program behavior, programmers still rely on traditional code browsing and debugging tools to build a mental model of a system's behavior. We propose a novel approach to visualizing program behavior through animated 2.5D object maps that depict particular objects and their interactions from a program trace. We describe our implementation and evaluate it for different program traces through an experience report and performance measurements. Our results indicate that our approach can benefit

program comprehension tasks, but further research is needed to improve scalability and usability.

Complete Paper #269

Human-Machine Collaboration for the Visual Exploration and Analysis of High-Dimensional Spatial Simulation Ensembles

Mai Dahshan¹, Nicholas Polys², Leanna House³, Karim Youssef² and Ryan Pollyea⁴

¹ School of Computing, University of North Florida, U.S.A.

² Department of Computer Science, Virginia Tech, U.S.A.

³ Department of Statistics, Virginia Tech, U.S.A.

⁴ Department of Geosciences, Virginia Tech, U.S.A.

Keywords: Simulation Ensembles, Spatial Data, Visual Analytics, Large Scale Visualization, Gaussian Process.

Abstract: Continuous improvements in supercomputing have given scientists from various fields the ability to conduct large-scale multi-dimensional numerical simulation ensembles. A simulation ensemble involves running multiple simulations, each with slight variations in model settings, such as input parameters, initial conditions, or boundary values. Exploring and analyzing these ensembles facilitates understanding parameter sensitivity and the correlations between different ensemble members. To capture these relationships, visual analytical tools are used to extract important features from the ensemble. In many cases, however, these visualizations highlight the differences in the ensemble using aggregated or descriptive statistics, ignoring the correlations and local differences between different spatial regions, which could hinder the exploration process. This paper proposes a visual analytical approach, SpatialGLEE, to interactively explore the spatial variability in the simulation ensemble. The proposed approach uses Gaussian Process Regression (GPR) and Semantic Interaction (SI) to help scientists explore the impact of input parameters on the ensemble and find the commonalities and differences across ensemble members and regions of interest (ROI). GPR models the spatial correlation structure in the ensemble. The modeled data is then inputted into the visualization pipeline for analysis and exploration with SI. The effectiveness of SpatialGLEE is demonstrated using a real-life case study.

Poster Presentations (Online) 1
16:00 - 17:00

VISAPP
Room Online 1

Complete Paper #39

Anisotropic Diffusion for Depth Estimation in Shape from Focus Systems

Bilal Ahmad, Ivar Farup and Pål Floor

Department of Computer Science, Norwegian University of Science & Technology, 2815 Gjøvik, Norway

Keywords: Anisotropic Diffusion, Shape from Focus, 3D Reconstruction.

Abstract: Shape from focus is a monocular method that uses the camera's focus as the primary indicator for depth estimation. The initial depth map is usually improved by penalizing the L2 regularizer as a smoothness constraint, which tends to smoothen the structural details due to linear diffusion. In this article, we propose an energy minimization-based framework to improve the initial depth map by utilizing a nonlinear, spatial technique, called anisotropic diffusion as a smoothness constraint, which is

combined with a fidelity term that incorporates the focus values of the initial depth to enhance structural aspects of the observed scene. Experiments are conducted on synthetic and real datasets which demonstrate that the proposed method can significantly improve the depth maps.

Complete Paper #41

Informative Rays Selection for Few-Shot Neural Radiance Fields

Marco Orsingher^{1,2}, Anthony Dell'Eva^{3,2}, Paolo Zani², Paolo Medici² and Massimo Bertozzi¹

¹ University of Parma, Italy

² VisLab, an Ambarella Inc. company, Italy

³ University of Bologna, Italy

Keywords: Neural Radiance Fields, Novel View Synthesis, Few-Shot Learning, 3D Reconstruction.

Abstract: Neural Radiance Fields (NeRF) have recently emerged as a powerful method for image-based 3D reconstruction, but the lengthy per-scene optimization limits their practical usage, especially in resource-constrained settings. Existing approaches solve this issue by reducing the number of input views and regularizing the learned volumetric representation with either complex losses or additional inputs from other modalities. In this paper, we present *KeyNeRF*, a simple yet effective method for training NeRF in few-shot scenarios by focusing on key informative rays. Such rays are first selected at camera level by a view selection algorithm that promotes baseline diversity while guaranteeing scene coverage, then at pixel level by sampling from a probability distribution based on local image entropy. Our approach performs favorably against state-of-the-art methods, while requiring minimal changes to existing NeRF codebases.

Complete Paper #170

PIRO: Permutation-Invariant Relational Network for Multi-Person 3D Pose Estimation

Nicolas Ugrinovic, Adria Ruiz, Antonio Agudo, Alberto Sanfeliu and Francesc Moreno-Noguer

Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain

Keywords: Human Pose Estimation, 3D, Single-View.

Abstract: Recovering multi-person 3D poses from a single RGB image is an ill-conditioned problem due to the inherent 2D-3D depth ambiguity, inter-person occlusions, and body truncation. To tackle these issues, recent works have shown promising results by simultaneously reasoning for different individuals. However, in most cases this is done by only considering pairwise inter-person interactions or between pairs of body parts, thus hindering a holistic scene representation able to capture long-range interactions. Some approaches that jointly process all people in the scene require defining one of the individuals as a reference and a pre-defined person ordering or limiting the number of individuals thus being sensitive to these choices. In this paper, we overcome both these limitations, and we propose an approach for multi-person 3D pose estimation that captures long-range interactions independently of the input order. We build a residual-like permutation-invariant network that successfully refines potentially corrupted initial 3D poses estimated by off-the-shelf detectors. The residual function is learned via a Set Attention (Lee et al., 2019) mechanism. Despite of our model being relatively straightforward, a thorough evaluation demonstrates that our approach is able to boost the

performance of the initially estimated 3D poses by large margins, achieving state-of-the-art results on two standardized benchmarks.

Complete Paper #205

NeRF-Diffusion for 3D-Consistent Face Generation and Editing

Héctor Laria^{1,2}, Kai Wang¹, Joost van de Weijer^{1,2}, Bogdan Raducanu^{1,2} and Yaxing Wang³

¹ Computer Vision Center, Barcelona, Spain

² Universitat Autònoma de Barcelona, Spain

³ Nankai University, China

Keywords: NeRF, Diffusion Models, 3D Generation, Multi-View Consistency, Face Generation.

Abstract: Generating high-fidelity 3D-aware images without 3D supervision is a valuable capability in various applications. Current methods based on NeRF features, SDF information, or triplane features have limited variation after training. To address this, we propose a novel approach that combines pretrained models for shape and content generation. Our method leverages a pretrained Neural Radiance Field as a shape prior and a diffusion model for content generation. By conditioning the diffusion model with 3D features, we enhance its ability to generate novel views with 3D awareness. We introduce a consistency token shared between the NeRF module and the diffusion model to maintain 3D consistency during sampling. Moreover, our framework allows for text editing of 3D-aware image generation, enabling users to modify the style over 3D views while preserving semantic content. Our contributions include incorporating 3D awareness into a text-to-image model, addressing identity consistency in 3D view synthesis, and enabling text editing of 3D-aware image generation. We provide detailed explanations, including the shape prior based on the NeRF model and the content generation process using the diffusion model. We also discuss challenges such as shape consistency and sampling saturation. Experimental results demonstrate the effectiveness and visual quality of our approach.

Poster Presentations (Online) 1
16:00 - 17:00

VISIGRAPP
Room Online 2

Poster Presentations (Online) 1
16:00 - 17:00

VISAPP
Room Online 2

Complete Paper #158

Assessing the Performance of Autoencoders for Particle Density Estimation in Acoustofluidic Medium: A Visual Analysis Approach

Lucas Massa¹, Tiago Vieira¹, Allan Martins² and Bruno Ferreira³

¹ Institute of Computing, Federal University of Alagoas, Lourival Melo Mota Av., Maceió, Brazil

² Department of Electrical Engineering, Federal University of Rio Grande do Norte, Natal, Brazil

³ Edge Innovation Center, Federal University of Alagoas, Maceió, Brazil

Keywords: Particle Density Estimation, Convolutional Autoencoder, Particle Size Estimation, Acoustofluidics.

Abstract: Micro-particle density is important for understanding different cell types, their growth stages, and how they respond to external stimuli. In previous work, a Gaussian curve fitting

method was used to estimate the size of particles, in order to later calculate their density. This approach required a long processing time, making the development of a Point of Care (PoC) device difficult. Current work proposes the application of a convolutional autoencoder (AE) to estimate single particle density, aiming to develop a PoC device that overcomes the limitations presented in the previous study. Thus, we used the AE to bottleneck a set of particle images into a single latent variable to evaluate its ability to represent the particle's diameter. We employed an identical physical apparatus involving a microscope to take pictures of particles in a liquid submitted to ultrasonic waves before the settling process. The AE was initially trained with a set of images for calibration. The acquired parameters were applied to the test set to estimate the velocity at which the particle falls within the ultrasonic chamber. This velocity was later used to infer the particle density. Our results demonstrated that the AE model performed much better, notably exhibiting significantly enhanced computational speed while concurrently achieving comparable error in density estimation.

Complete Paper #160

Image Edge Enhancement for Effective Image Classification

Bu Tianhao¹, Michalis Lazarou² and Tania Stathaki²

¹ *Glory Engineering & Tech Co., LTD, China*

² *Imperial College London, U.K.*

Keywords: Data Augmentation, Image Classification, High Boost Filtering, Edge Enhancement.

Abstract: Image classification has been a popular task due to its feasibility in real-world applications. Training neural networks by feeding them RGB images has demonstrated success over it. Nevertheless, improving the classification accuracy and computational efficiency of this process continues to present challenges that researchers are actively addressing. A widely popular embraced method to improve the classification performance of neural networks is to incorporate data augmentations during the training process. Data augmentations are simple transformations that create slightly modified versions of the training data, and can be very effective in training neural networks to mitigate overfitting and improve their accuracy performance. In this study, we draw inspiration from high-boost image filtering and propose an edge enhancement-based method as means to enhance both accuracy and training speed of neural networks. Specifically, our approach involves extracting high frequency features, such as edges, from images within the available dataset and fusing them with the original images, to generate new, enriched images. Our comprehensive experiments, conducted on two distinct datasets—CIFAR10 and CALTECH101, and three different network architectures—ResNet-18, LeNet-5 and CNN-9—demonstrate the effectiveness of our proposed method.

Complete Paper #333

A Computer Vision Approach to Compute Bubble Flow of Offshore Wells

Rogério Hart and Aura Conci

Institute of Computing, Universidade Federal Fluminense, Niterói, Rio de Janeiro, Brazil

Keywords: Video Analysis, Segmentation, Flow Rate, Neural Network Model, Offshore Substructure.

Abstract: This work presents two approaches for detecting and quantifying the offshore flow of leaks, using video recorded by

a remote-operated vehicle (ROV) through underwater image analysis and considering the premise of no bubble overlap. One is designed using only traditional digital image approaches, such as Mathematical Morphology operators and Canny edge detection, and the second uses segmentation Convolutional Neural Network. Implementation and experimentation details are presented, enabling comparison and reproduction. The results are compared with videos acquired under controlled conditions and in an operational situation, as well as with all previous possible works. Comparison considers the estimation of the average diameter of rising bubbles, velocity of rise, leak flow rate, computational automation, and flexibility in bubble recognition. The results of both techniques are almost the same depending on the video content in the analysis.

Complete Paper #442

Selection of Backbone for Feature Extraction with U-Net in Pancreas Segmentation

Alexandre Araújo, Joao Sousa de Almeida, Anselmo Cardoso de Paiva and Geraldo Braz Junior

Applied Computing Group (NCA), Federal University of Maranhão (UFMA), São Luís, MA, Brazil

Keywords: Image Segmentation, Pancreas Segmentation, U-Net.

Abstract: The survival rate for pancreatic cancer is among the worst, with a mortality rate of 98%. Diagnosis in the early stage of the disease is the main factor that defines the prognosis. Imaging scans, such as Computerized Tomography scans, are the primary tools for early diagnosis. Computer Assisted Diagnosis tools that use these scans usually include in their pipeline the segmentation of the pancreas as one of the initial steps for diagnosis. This paper presents a comparative study of the use of different backbones in combination with the U-Net. This study aims to demonstrate that using pre-trained backbones is a valuable tool for pancreas segmentation and to provide a comparative benchmark for this task. The best result obtained was 85.96% of Dice in the MSD dataset for the pancreas segmentation using backbone efficientnetb7.

Complete Paper #310

Influence of Pixel Perturbation on eXplainable Artificial Intelligence Methods

Juliana Feitosa, Mateus Roder, João Papa and José Brega

Department of Computing, School of Science, Sao Paulo State University (UNESP), Brazil

Keywords: eXplainable Artificial Intelligence, Pixel Perturbation, Artificial Intelligence.

Abstract: The current scenario around Artificial Intelligence (AI) has demanded more and more transparent explanations about the existing models. The use of eXplicable Artificial Intelligence (XAI) has been considered as a solution in the search for explainability. As such, XAI methods can be used to verify the influence of adverse scenarios, such as pixel disturbance on AI models for segmentation. This paper presents the experiments performed with fish images of the Pacu species to determine the influence of pixel perturbation through the following explainable methods: Grad-CAM, Saliency Map, Layer Grad-CAM and CNN Filters. The perturbed pixels were considered the most important for the model during the segmentation process of the input image regions. From the existing pixel perturbation techniques, the images were subjected to three main techniques: white noise, color black noise

and random noise. From the results obtained, it was observed that the Grad-CAM method had different behaviors for each perturbation technique tested, while the CNN Filters method showed more stability in the variation of the image averaging. The Saliency Map was the least sensitive to the three types of perturbation, as it required fewer iterations. Furthermore, of the perturbation techniques tested, Black noise showed the least ability to impact segmentation. Thus, it is concluded that the perturbation methods influence the outcome of the explainable models tested and interfere with these models in different ways. It is suggested that the experiments presented here be replicated on other AI models, on other explainability methods, and with other existing perturbation techniques to gather more evidence about this influence and from that, quantify which combination of XAI method and pixel perturbation is best for a given problem.

develop image and video-based conversation models that merges the representational abilities of a pretrained visual encoder and the generative powers of an LLM, capable of understanding and conversing about images and videos with grounding capabilities.

Complete Paper #318

Convolutional Neural Networks and Image Patches for Lithological Classification of Brazilian Pre-Salt Rocks

Mateus Roder¹, Leandro Passos², Clayton Pereira², João Papa², Altanir Mello Junior³, Marcelo Fagundes de Rezende³, Yaro Silva³ and Alexandre Vidal¹

¹ *Institute of Geosciences, Campinas State University (UNICAMP), Brazil*

² *Department of Computing, São Paulo State University (UNESP), Brazil*

³ *Research Center, Leopoldo Américo Miguez de Mello Research, Development and Innovation Center (Cenpes), Brazil*

Keywords: Lithological Classification, Pre-Salt Rocks, Convolutional Neural Networks.

Abstract: Lithological classification is a process employed to recognize and interpret distinct structures of rocks, providing essential information regarding their petrophysical, morphological, textural, and geological aspects. The process is particularly interesting regarding carbonate sedimentary rocks in the context of petroleum basins since such rocks can store large quantities of natural gas and oil. Thus, their features are intrinsically correlated with the production potential of an oil reservoir. This paper proposes an automatic pipeline for the lithological classification of carbonate rocks into seven distinct classes, comparing nine state-of-the-art deep learning architectures. As far as we know, this is the largest study in the field. Experiments were performed over a private dataset obtained from a Brazilian petroleum company, showing that MobileNetV3large is the more suitable approach for the undertaking.

Keynote Lecture
17:00 - 18:00

VISIGRAPP
Room Plenary

Towards Detailed Understanding of the Visual World

Fahad Khan
MBZUAI, U.A.E.

Abstract: Machine perception that corresponds to the ability to understand the visual world based on the input from sensors, such as cameras is one of the central problems in Artificial Intelligence. To this end, recent years have witnessed tremendous progress in various visual perception tasks having real-world applications in e.g., robotics, autonomous driving and surveillance. In this talk, I will first present our recent results towards understanding state-of-the-art deep learning-based visual recognition networks in terms of their robustness and generalizability. Next, I will discuss moving one step further from instance-level recognition to

Wednesday Sessions: February 28

Wednesday Sessions: February 28 Program Layout

| | Coffee-Break | Foyer | Online 1 | Online 2 | Plenary | Restaurant | Space 1 | Space 3B | Space 3C | Space 4A | Space 4B | Space 4C | Space 4D |
|-------|--------------|----------------------------|---------------------------------|-------------------------------|--------------------------------|------------|------------------------------------|-------------------------------------|-------------------------------|-------------------------------------|----------------------------------|---------------------------|-----------------------|
| 9:00 | | | | | | | VISIGRAPP Session 3 | VISIGRAPP Session 3 | VISIGRAPP Session 3 | VISIGRAPP Session 3 | VISIGRAPP Session 3 | Tutorial Francesco Ragusa | VISIGRAPP Session 3 |
| 9:30 | | | Oral Presentations (Online) 3 | Oral Presentations (Online) 3 | | | VISAPP #23, #51, #365, #407 | VISAPP #173, #255, #272, #368 | VISAPP #239, #264, #293, #302 | VISAPP #171, #220, #326, #424 | GRAPP #98, #301, #320, #451 | | HUCAPP #37, #59, #112 |
| 10:00 | | | | | | | | | | | | | |
| 10:30 | Coffee-Break | | | | | | | | | | | | |
| 11:00 | | | | | Keynote Lecture Gerhard Rigoll | | | | | | | | |
| 11:30 | | | | | | | | | | | | | |
| 12:00 | | | Oral Presentations (Online) 4 | Oral Presentations (Online) 4 | | | VISIGRAPP Session 4 | VISIGRAPP Session 4 | VISIGRAPP Session 4 | VISIGRAPP Session 4 | VISIGRAPP Session 4 | Tutorial Francesco Ragusa | VISIGRAPP Session 4 |
| 12:30 | | | | | | | VISAPP #103, #166, #168 | VISAPP #43, #244, #453 | VISAPP #76, #88, #169 | VISAPP #298, #312, #325 | GRAPP #262, #363, #458 | | IVAPP #237, #436 |
| 13:00 | | | | | | | | | | | | | |
| 13:30 | | | | | | Lunch | | | | | | | |
| 14:00 | | | | | | | | | | | | | |
| 14:30 | | | Oral Presentations (Online) 5 | Oral Presentations (Online) 5 | | | VISIGRAPP Session 5 | VISIGRAPP Session 5 | VISIGRAPP Session 5 | VISIGRAPP Session 5 | VISIGRAPP Session 5 | Tutorial Kai Barthel | |
| 15:00 | | | | | | | VISAPP #56, #123, #201, #359, #366 | VISAPP #154, #200, #327, #369, #423 | VISAPP #174, #188, #248, #445 | VISAPP #128, #191, #308, #322, #323 | HUCAPP #20, #21, #87, #147, #450 | | |
| 15:30 | | | | | | | | | | | | | |
| 16:00 | | | | | | | | | | | | | |
| 16:30 | Coffee-Break | VISIGRAPP Poster Session 2 | Poster Presentations (Online) 2 | | | | | | | | | | |
| 17:00 | | | | | Keynote Lecture Alvitia Ottley | | | | | | | | |
| 17:30 | | | | | | | | | | | | | |
| 18:00 | | | | | | | | | | | | | |

Session 3A
09:00 - 10:30

VISAPP
Room Space 1

Categorization and Scene Understanding

Complete Paper #23

Enabling On-Device Continual Learning with Binary Neural Networks and Latent Replay

Lorenzo Vorabbi^{1,2}, Davide Maltoni², Guido Borghi² and Stefano Santi¹

¹ *Datalogic Labs, Bologna, 40012, Italy*

² *Department of Computer Science and Engineering (DISI), University of Bologna, Cesena, Italy*

Keywords: Binary Neural Networks, On-Device Learning, TinyML, Continual Learning.

Abstract: On-device learning remains a formidable challenge, especially when dealing with resource-constrained devices that have limited computational capabilities. This challenge is primarily rooted in two key issues: first, the memory available on embedded devices is typically insufficient to accommodate the memory-intensive back-propagation algorithm, which often relies on floating-point precision. Second, the development of learning algorithms on models with extreme quantization levels, such as Binary Neural Networks (BNNs), is critical due to the drastic reduction in bit representation. In this study, we propose a solution that combines recent advancements in the field of Continual Learning (CL) and Binary Neural Networks to enable on-device training while maintaining competitive performance. Specifically, our approach leverages binary latent replay (LR) activations and a novel quantization scheme that significantly reduces the number of bits required for gradient computation. The experimental validation demonstrates a significant accuracy improvement in combination with a noticeable reduction in memory requirement, confirming the suitability of our approach in expanding the practical applications of deep learning in real-world scenarios.

Complete Paper #365

Anomaly Detection on Roads Using an LSTM and Normal Maps

Yusuke Nonaka¹, Hideo Saito¹, Hideaki Uchiyama¹, Kyota Higa² and Masahiro Yamaguchi²

¹ *Keio University, Yokohama, Japan*

² *NEC Corporation, Kawasaki, Japan*

Keywords: Deference-in-Level Detection, Unsupervised Learning, Outdoor Navigation, Anomaly Detection.

Abstract: Detecting anomalies on the road is crucial for generating hazard maps within factory premises and facilitating navigation for visually impaired individuals or robots. This paper proposes a method for anomaly detection on road surfaces using normal maps and a Long Short-Term Memory (LSTM). While existing research primarily focuses on detecting anomalies on the road based on variations in height or color information of images, our approach leverages anomaly detection to identify changes in the spatial structure of the walking scenario. The normal (non-anomaly) data consists of time series normal maps depicting previously traversed roads, which are utilized to predict the upcoming road conditions. Subsequently, an anomaly score is computed by comparing the predicted normal map with the normal map at $t + 1$. If the anomaly score exceeds a dynamically set threshold, it indicates the presence of anomalies on the road. The proposed method employs unsupervised learning for anomaly

detection. To assess the effectiveness of the proposed method, we conducted accuracy assessments using a custom dataset, taking into account a qualitative comparison with the results of existing methods. The results confirm that the proposed method effectively detects anomalies on road surfaces through anomaly detection.

Complete Paper #51

Cybersecurity Intrusion Detection with Image Classification Model Using Hilbert Curve

Punyawat Jaroensiripong¹, Karin Sumongkayothin¹, Prarinya Siritanawan² and Kazunori Kotani²

¹ *Department of Computer Engineering, Faculty of Engineering, Mahidol University, Thailand*

² *Japan Advanced Institute of Science and Technology, Japan*

Keywords: Machine Learning (ML), Deep Learning (DL), Cybersecurity, Security Operation Center (SOC), Intrusion Detection System (IDS), Hilbert Curve.

Abstract: Cybersecurity intrusion detection is crucial for protecting an online system from cyber-attacks. Traditional monitoring methods used in the Security Operation Center (SOC) are insufficient to handle the vast volume of traffic data, producing an overwhelming number of false alarms, and eventually resulting in the neglect of intrusion incidents. The recent integration of Machine Learning (ML) and Deep Learning (DL) into SOC monitoring systems has enhanced the intrusion detection capabilities by learning the patterns of network traffic data. Despite many ML methods implemented for intrusion detection, the Convolutional Neural Network (CNN), one of the most high-performing ML algorithms, has not been widely adopted for the intrusion detection systems. This research aims to explore the potentials of CNN implementation with the network data flows. Since the CNN was originally designed for image processing applications, it is necessary to convert the 1-dimensional network data flows into 2-dimensional image data. This research presents a novel approach to convert the network data flow into an image (flow-to-image) by the Hilbert curve mapping algorithm which can preserve the locality of the data. Then, we apply the converted images to the CNN-based intrusion detection system. Eventually, the proposed method and model can outperform the recent methods with 92.43% accuracy and 93.05% F1-score on the CIC-IDS2017 dataset, and 81.78% accuracy and 83.46% F1-score on the NSL-KDD dataset. In addition to the classification capability, the flow-to-image mapping algorithm can also visualize the characteristics of the network attack on the images visually, which can be an alternative monitoring approach for SOC.

Complete Paper #407

SMART-RD: Towards a Risk Assessment Framework for Autonomous Railway Driving

Justin Bescop, Nicolas Goeman, Amel Aissaoui, Benjamin Allaert and Jean-Philippe Vandeborre

IMT Nord Europe, Institut Mines-Télécom, Univ. Lille, Centre for Digital Systems, F-59000 Lille, France

Keywords: Autonomous Driving, Dynamic Risk Assessment, Deep Learning, Railway Dataset.

Abstract: While the automotive industry has made significant contributions to vision-based dynamic risk assessment, progress has been limited in the railway domain. This is mainly due to the lack of data and to the unavailability of security-based annotation for the existing datasets. This paper proposes the

first annotation framework for the railway domain that takes into account the different components that significantly contribute to the vision-based risk estimation in driving scenarios, thus enabling an accurate railway risk assessment. A first baseline based on neural network is performed to prove the consistency of the risk-based annotation. The performances show promising results for vision-based risk assessment according to different levels of risk.

Session 3B **VISAPP**
09:00 - 10:30 **Room Space 3B**
Machine Learning Technologies for Vision

Complete Paper #255

Investigating the Corruption Robustness of Image Classifiers with Random p -norm Corruptions

Georg Siedel^{1,2}, Weijia Shao¹, Silvia Vock¹ and Andrey Morozov²

¹ Federal Institute for Occupational Safety and Health (BAuA), Dresden, Germany

² University of Stuttgart, Germany

Keywords: Image Classification, Corruption Robustness, p -norm.

Abstract: Robustness is a fundamental property of machine learning classifiers required to achieve safety and reliability. In the field of adversarial robustness of image classifiers, robustness is commonly defined as the stability of a model to all input changes within a p -norm distance. However, in the field of random corruption robustness, variations observed in the real world are used, while p -norm corruptions are rarely considered. This study investigates the use of random p -norm corruptions to augment the training and test data of image classifiers. We evaluate the model robustness against imperceptible random p -norm corruptions and propose a novel robustness metric. We empirically investigate whether robustness transfers across different p -norms and derive conclusions on which p -norm corruptions a model should be trained and evaluated. We find that training data augmentation with a combination of p -norm corruptions significantly improves corruption robustness, even on top of state-of-the-art data augmentation schemes.

Complete Paper #368

Beyond the Known: Adversarial Autoencoders in Novelty Detection

Muhammad Asad^{1,2}, Ihsan Ullah^{1,2}, Ganesh Sistu³ and Michael Madden^{1,2}

¹ Machine Learning Research Group, School of Computer Science, University of Galway, Ireland

² Insight SFI Research Centre for Data Analytics, University of Galway, Ireland

³ Valeo Vision Systems, Tuam, Ireland

Keywords: Anomaly Detection, Adversarial Auto-Encoders, Reconstruction, Discriminator, Distribution.

Abstract: In novelty detection, the goal is to decide if a new data point should be categorized as an inlier or an outlier, given a training dataset that primarily captures the inlier distribution. Recent approaches typically use deep encoder and decoder network frameworks to derive a reconstruction error, and employ this error either to determine a novelty score, or as the basis for a one-class classifier. In this research, we use a similar framework

but with a lightweight deep network, and we adopt a probabilistic score with reconstruction error. Our methodology calculates the probability of whether the sample comes from the inlier distribution or not. This work makes two key contributions. The first is that we compute the novelty probability by linearizing the manifold that holds the structure of the inlier distribution. This allows us to interpret how the probability is distributed and can be determined in relation to the local coordinates of the manifold tangent space. The second contribution is that we improve the training protocol for the network. Our results indicate that our approach is effective at learning the target class, and it outperforms recent state-of-the-art methods on several benchmark datasets.

Complete Paper #173

Large Filter Low-Level Processing by Edge TPU

Gerald Krell and Thilo Pionteck

Institute for Information and Communication Technology, Otto von Guericke University Magdeburg, Universitätsplatz 2, Magdeburg, Germany

Keywords: Edge TPU, Tensor Flow, Preprocessing, Low-Level Processing, Deep Learning.

Abstract: Edge TPUs offer high processing power at a low cost and with minimal power consumption. They are particularly suitable for demanding tasks such as classification or segmentation using Deep Learning Frameworks, acting as a neural coprocessor in host computers and mobile devices. The question arises as to whether this potential can be utilized beyond the specific domains for which the frameworks are originally designed. One example pertains to addressing various error classes by utilizing a trained deconvolution filter with a large filter size, requiring computation power that can be efficiently accelerated by the powerful matrix multiplication unit of the TPU. However, the application of the TPU is restricted due to the fact that Edge TPU software is not fully open source. This limits to integration with existing Deep Learning frameworks and the Edge TPU compiler. Nonetheless, we demonstrate a method of estimating and utilizing a convolutional filter of large size on the TPU for this purpose. The deconvolution process is accomplished by utilizing pre-estimated convolutional filters offline to perform low-level preprocessing for various error classes, such as denoising, deblurring, and distortion removal.

Complete Paper #272

Anomaly Detection and Localization for Images of Running Paper Web in Paper Manufacturing

Afshin Dini¹, Marja Mettänen² and Esa Rahtu¹

¹ Unit of Computing Sciences, Tampere University, Tampere, Finland

² Procemex Oy Ltd, Jyväskylä, Finland

Keywords: Anomaly Detection, Anomaly Localization, Running Paper Web Defects, Paper Manufacturing.

Abstract: We introduce a new method based on convolutional autoencoders to detect and locate paper web anomalies that can cause web breaks during the paper production process. In this approach, we pre-process the images, captured by two high-speed cameras located at the opposite sides of the running paper web at a paper machine, in several steps to remove noises and separate the paper web areas from the background. After designing and training a convolutional autoencoder with non-anomalous samples, a novel anomaly score and map are defined to find and locate web irregularities based on an edge detector and a reconstruction error, defined by the combination of absolute error and Structural Similarity Index Measure between the reconstructed and the original images, in each test sample. By

assessing the proposed approach on the images taken from a real paper machine, we discover that this method can detect paper defects properly and, therefore it has the potential to improve machine functionality and even to prevent certain types of web breaks, which reduces the machine downtime, paper losses, maintenance costs, and energy consumption, i.e., increases the performance and efficiency of paper machinery.

Session 3C
09:00 - 10:30
Multimedia Forensics

VISAPP
Room Space 3C

Complete Paper #264

Defying Limits: Super-Resolution Refinement with Diffusion Guidance

Marcelo Santos¹, João Neves², Hugo Proença² and David Menotti¹

¹ Department of Informatics, Federal University of Paraná, Curitiba, Brazil

² Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal

Keywords: Super-Resolution, Face Recognition, Diffusion Models, Diffusion Guidance.

Abstract: Due to the growing number of surveillance cameras and rapid technological advancement, facial recognition algorithms have been widely applied. However, their performance decreases in challenging environments, such as those involving surveillance cameras with low-resolution images. To address this problem, in this paper, we introduce SRDG, a super-resolution approach supported by two state-of-the-art methods: diffusion models and classifier guidance. The diffusion process reconstructs the image, and the classifier refines the image reconstruction based on a set of facial attributes. This combination of models is capable of working with images with a very limited resolution (8×8 and 16×16), being suitable for surveillance scenarios where subjects are typically distant from the camera. The experimental validation of the proposed approach shows that super-resolution images exhibit enhanced details and improved visual quality. More importantly, when using our super-resolution algorithm, the facial discriminability of images is improved compared to state-of-the-art super-resolution approaches, resulting in a significant increase in face recognition accuracy. To the best of our knowledge, this is the first time classifier guidance has been applied to refine super-resolution results of images from surveillance cameras. Source code is available at <https://github.com/marcelowds/SRDG>.

Complete Paper #293

On Feasibility of Transferring Watermarks from Training Data to GAN-Generated Fingerprint Images

Venkata Mannam, Andrey Makrushin and Jana Dittmann

Department of Computer Science, Otto von Guericke University, Magdeburg, Germany

Keywords: Watermarking, Biometrics, Synthetic Fingerprints, Synthetic Data Detection, GAN, Pix2pix.

Abstract: Due to the rise of high-quality synthetic data produced by generative models and a growing mistrust in images published in social media, there is an urgent need for reliable means of synthetic image detection. Passive detection approaches cannot properly handle images created by "unknown" generative models. Embedding watermarks in synthetic images is an active detection

approach which transforms the task from fake detection to watermark extraction. The focus of our study is on watermarking biometric fingerprint images produced by Generative Adversarial Networks (GAN). We propose to watermark images used for training of a GAN model and study the interplay between the watermarking algorithm, GAN architecture, and training hyperparameters to ensure the watermark transfer from training data to GAN-generated fingerprint images. A hybrid watermarking algorithm based on discrete cosine transformation, discrete wavelet transformation, and singular value decomposition is shown to produce transparent logo watermarks which are robust to pix2pix network training. The pix2pix network is applied to reconstruct realistic fingerprints from minutiae. The watermark imperceptibility and robustness to GAN training are validated by peak signal-to-noise ratio and bit error rate respectively. The influence of watermarks on reconstruction success and realism of fingerprints is measured by Verifinger matching scores and NFIQ2 scores respectively.

Complete Paper #302

Towards the Detection of Diffusion Model Deepfakes

Jonas Ricker¹, Simon Damm¹, Thorsten Holz² and Asja Fischer¹

¹ Ruhr University Bochum, Bochum, Germany

² CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

Keywords: Deepfake Detection, Diffusion Models, Generative Adversarial Networks, Frequency Analysis.

Abstract: In the course of the past few years, diffusion models (DMs) have reached an unprecedented level of visual quality. However, relatively little attention has been paid to the detection of DM-generated images, which is critical to prevent adverse impacts on our society. In contrast, generative adversarial networks (GANs), have been extensively studied from a forensic perspective. In this work, we therefore take the natural next step to evaluate whether previous methods can be used to detect images generated by DMs. Our experiments yield two key findings: (1) state-of-the-art GAN detectors are unable to reliably distinguish real from DM-generated images, but (2) re-training them on DM-generated images allows for almost perfect detection, which remarkably even generalizes to GANs. Together with a feature space analysis, our results lead to the hypothesis that DMs produce fewer detectable artifacts and are thus more difficult to detect compared to GANs. One possible reason for this is the absence of grid-like frequency artifacts in DM-generated images, which are a known weakness of GANs. However, we make the interesting observation that diffusion models tend to underestimate high frequencies, which we attribute to the learning objective.

Complete Paper #239

Diffusion-Inspired Dynamic Models for Enhanced Fake Face Detection

Mamadou Bah, Rajjeshwar Ganguly and Mohamed Dahmane

Computer Research Institute of Montreal (CRIM), Canada

Keywords: DeepFake Detection, Dynamic Model, Media Forensics, U-Net.

Abstract: The conventional convolutional U-Net model was originally designed to segment images. Enhanced versions of the architecture with timestep embedding and self-attention

inspired by transformers were proposed in the literature for image classification tasks. In this paper, we investigate a U-Net–encoder architecture for deepfake detection that involves two key features, the use of self-attention blocks to capture both local and global content representation, and the integration of timestep embedding to capture the dynamic perturbation of the input data. The model is trained and evaluated on FF++ dataset, comprising of real and deepfake synthesized videos. Notably, compared to traditional models pretrained on ImageNet, our model demonstrates superior performance. The experimental results highlight the effectiveness of our approach in achieving improved classification results for the challenging task of distinguishing real and deepfake images. The achieved performances suggest that the model aims to leverage both spatial information and dynamic perturbation for improved detection performance.

Session 3D
09:00 - 10:30
Object Detection and Localization

VISAPP
Room Space 4A

Complete Paper #171

End-to-End Chess Recognition

Athanasios Masouris and Jan van Gemert

Computer Vision Lab, Delft University of Technology, Delft, The Netherlands

Keywords: Chess Recognition, Chess Dataset, Computer Vision, Deep Learning.

Abstract: Chess recognition is the task of extracting the chess piece configuration from a chessboard image. Current approaches use a pipeline of separate, independent, modules such as chessboard detection, square localization, and piece classification. Instead, we follow the deep learning philosophy and explore an end-to-end approach to directly predict the configuration from the image, thus avoiding the error accumulation of the sequential approaches and eliminating the need for intermediate annotations. Furthermore, we introduce a new dataset, *Chess Recognition Dataset (ChessReD)*, that consists of 10,800 real photographs and their corresponding annotations. In contrast to existing datasets that are synthetically rendered and have only limited angles, ChessReD has photographs captured from various angles using smartphone cameras; a sensor choice made to ensure real-world applicability. Our approach in chess recognition on the introduced challenging benchmark dataset outperforms related approaches, successfully recognizing the chess pieces' configuration in 15.26% of ChessReD's test images. This accuracy may seem low, but it is $\approx 7x$ better than the current state-of-the-art and reflects the difficulty of the problem. The code and data are available through: <https://github.com/ThanosM97/end-to-end-chess-recognition>.

Complete Paper #424

CSE: Surface Anomaly Detection with Contrastively Selected Embedding

Simon Thomine^{1,2} and Hichem Snoussi¹

¹ University of Technology Troyes, Troyes, France

² AQUILAE, Troyes, France

Keywords: Unsupervised, Anomaly, Pattern, Contrastive, Autoencoder, Feature Extraction.

Abstract: Detecting surface anomalies of industrial materials poses a significant challenge within a myriad of industrial manufacturing processes. In recent times, various methodologies have

emerged, capitalizing on the advantages of employing a network pre-trained on natural images for the extraction of representative features. Subsequently, these features are subjected to processing through a diverse range of techniques including memory banks, normalizing flow, and knowledge distillation, which have exhibited exceptional accuracy. This paper revisits approaches based on pre-trained features by introducing a novel method centered on target-specific embedding. To capture the most representative features of the texture under consideration, we employ a variant of a contrastive training procedure that incorporates both artificially generated defective samples and anomaly-free samples during training. Exploiting the intrinsic properties of surfaces, we derived a meaningful representation from the defect-free samples during training, facilitating a straightforward yet effective calculation of anomaly scores. The experiments conducted on the MVTEC AD and TILDA datasets demonstrate the competitiveness of our approach compared to state-of-the-art methods.

Complete Paper #220

Understanding Marker-Based Normalization for FLIM Networks

Leonardo Joao^{1,2}, Matheus Cerqueira¹, Barbara Benato¹ and Alexandre Falcao¹

¹ Institute of Computing, State University of Campinas, Campinas, 13083-872, São Paulo, Brazil

² LIGM, Univ. Gustave-Eiffel, Marne-la-Vallée, F-77454, France

Keywords: Marker-Based-Normalization, Flim, Z-Score-Normalization, Object-Detection.

Abstract: Successful methods for object detection in multiple image domains are based on convolutional networks. However, such approaches require large annotated image sets for network training. One can build object detectors by exploring a recent methodology, Feature Learning from Image Markers (FLIM), that considerably reduces human effort in data annotation. In FLIM, the encoder's filters are estimated among image patches extracted from scribbles drawn by the user on discriminative regions of a few representative images. The filters are meant to create feature maps in which the object is activated or deactivated. This task depends on a z-score normalization using the scribbles' statistics, named marker-based normalization (MBN). An adaptive decoder (point-wise convolution with activation) finds its parameters for each image and outputs a saliency map for object detection. This encoder-decoder network is trained without backpropagation. This work investigates the effect of MBN on the network's results. We detach the scribble sets for filter estimation and MBN, introduce a bot that draws scribbles with distinct ratios of object-and-background samples, and evaluate the impact of five different ratios on three datasets through six quantitative metrics and feature projection analysis. The experiments suggest that scribble detachment and MBN with object oversampling are beneficial.

Complete Paper #326

AbSynth: Using Abstract Image Synthesis for Synthetic Training

Dominik Penk^{1,2}, Maik Horn², Christoph Strohmeyer², Bernhard Egger¹, Marc Stamminger¹ and Frank Bauer¹

¹ Chair of Visual Computing, Friedrich-Alexander-Universität Erlangen-Nürnberg, Cauerstraße 11, Erlangen, Germany

² Schaeffler Technologies AG & Co. KG, Industriestraße 1-3, Herzogenaurach, Germany

Keywords: Synthetic Training Data, Domain Gap, Deep Learning, Computer Vision.

Abstract: We present a novel pipeline for training neural networks to tackle geometry-induced vision tasks, relying solely on synthetic training images generated from (geometric) CAD models of the objects under consideration. Instead of aiming for photorealistic renderings, our approach maps both synthetic and real-world data onto a common abstract image space reducing the domain gap. We demonstrate that this projection can be decoupled from the downstream task, making our method an easy drop-in solution for a variety of applications. In this paper, we use line images as our chosen abstract image representation due to their ability to capture geometric properties effectively. We introduce an efficient training data synthesis method, that generates images tailored for transformation into a line representation. Additionally, we explore how the use of sparse line images opens up new possibilities for augmenting the dataset, enhancing the overall robustness of the downstream models. Finally, we provide an evaluation of our pipeline and augmentation techniques across a range of vision tasks and state-of-the-art models, showcasing their effectiveness and potential for practical applications.

Session 3A
09:00 - 10:30
Rendering

GRAPP
Room Space 4B

Complete Paper #98

Real-Time Editing of Path-Traced Scenes with Prioritized Re-Rendering

Annalena Ulschmid¹, Bernhard Kerbl¹, Katharina Krösl^{1,2} and Michael Wimmer¹

¹ TU Wien, Austria

² VRVis Forschungs-GmbH, Austria

Keywords: Path Tracing, Scene Editing, Adaptive Rendering, Empirical Studies.

Abstract: With recent developments in GPU ray tracing performance and (AI-accelerated) noise reduction techniques, Monte Carlo Path Tracing at real-time rates becomes a viable solution for interactive 3D scene editing, with growing support in popular software. However, even for minor edits (e.g., adjusting materials or moving small objects), current solutions usually discard previous samples and the image formation process is started from scratch. In this paper, we present two adaptive, priority-based re-rendering techniques with incremental updates, prioritizing the reconstruction of regions with high importance, before gradually moving to less important regions. The suggested methods automatically identify and schedule sampling and accumulation of immediately affected regions. An extensive user study analyzes whether such prioritized renderings are beneficial to interactive scene editing, comparing them with same-time conventional re-rendering. Our evaluation shows that even with simple priority

policies, there is a significant preference for such incremental rendering techniques for interactive editing of small objects over full-screen re-rendering with denoising.

Complete Paper #301

Hierarchical Bitmask Implicit Grids for Efficient Point-in-Volume Queries on the GPU

Julius Ikkala, Tuomas Lauttia, Pekka Jääskeläinen and Markku Mäkitalo

Tampere University, Tampere, Finland

Keywords: Computer Graphics, Real-Time Rendering, Ray Tracing, Data Structures.

Abstract: We propose “Hierarchical Bitmask Implicit Grids”, a novel, memory-efficient spatial index data structure for querying bounding volumes based on a contained point, targeting real-time use cases on GPUs. Like grid structures based on 3D arrays, implicit grids allow for nearly array-like direct lookups of cells without traversal through a spatial tree structure. However, the space complexity of this structure is $O(n)$ with respect to resolution as opposed to $O(n^3)$, which allows for dramatically higher resolutions than would be feasible with a 3D array. We demonstrate the effectiveness of this data structure by applying it to two example use cases: light culling and decal rendering. We measure both cases with ray tracing and multi-view rendering. We show that with tens of thousands of entries, our data structure can be built in 0.1–0.2 milliseconds, being 2.9x faster than the compared state-of-the-art decal method and orders of magnitude faster than dense 3D arrays, while delivering at least similar or even up to doubled rendering performance.

Complete Paper #451

Non-Photorealistic Rendering of 3D Point Clouds Using Segment-Specific Image-Space Effects

Ole Wegen¹, Josafat-Mattias Burmeister¹, Max Reimann², Rico Richter¹ and Jürgen Döllner²

¹ University of Potsdam, Germany

² Hasso-Plattner-Institute, University of Potsdam, Germany

Keywords: 3D Point Clouds, Non-Photorealistic Rendering, Segmentation, Image-Based Artistic Rendering.

Abstract: 3D point clouds are a widely used representation for surfaces and object geometries. However, their visualization can be challenging due to point sparsity and acquisition inaccuracies, leading to visual complexity and ambiguity. Non-photorealistic rendering (NPR) addresses these challenges by using stylization techniques to abstract from certain details or emphasize specific areas of a scene. Although NPR effectively reduces visual complexity, existing approaches often apply uniform styles across entire point clouds, leading to a loss of detail or saliency in certain areas. To address this, we present a novel segment-based NPR approach for point cloud visualization. Utilizing prior point cloud segmentation, our method applies distinct rendering styles to different segments, enhancing scene understanding and directing the viewer’s attention. Our emphasis lies in integrating aesthetic and expressive elements through image-based artistic rendering, such as watercolor or cartoon filtering. To combine the per-segment images into a consistent final image, we propose a user-controllable depth inpainting algorithm. This algorithm estimates depth values for pixels that lacked depth information during point cloud rendering but received coloration during image-based stylization. Our approach supports real-time rendering of large

point clouds, allowing users to interactively explore various artistic styles.

Complete Paper #320

Sparse Spatial Shading in Augmented Reality

Rikard Olajos and Michael Doggett

Faculty of Engineering, Lund University, Sweden

Keywords: Augmented Reality, Colour Correction, Data Structures.

Abstract: In this work, we present a method for acquiring, storing, and using scene data to enable realistic shading of virtual objects in an augmented reality application. Our method allows for sparse sampling of the environment's lighting condition while still delivering a convincing shading to the rendered objects. We use common camera parameters, provided by a head-mounted camera, to get lighting information from the scene and store them in a tree structure, saving both locality and directionality of the data. This makes our approach suitable for implementation in augmented reality applications where the sparse and unpredictable nature of the data samples captured from a head-mounted device can be problematic. The construction of the data structure and the shading of virtual objects happen in real time, and without requiring high-performance hardware. Our model is designed for augmented reality devices with optical see-through displays, and in this work we used Microsoft's HoloLens 2.

Session 3A
09:00 - 10:30
Theories, Models and User Evaluation

HUCAPP
Room Space 4D

Complete Paper #59

Automatic Viewpoint Selection for Interactive Motor Feedback Using Principle Component Analysis

Florian Diller¹, Alexander Wiebel¹ and Gerik Scheuermann²

¹ UX-Vis group, Hochschule Worms University of Applied Sciences, Germany

² BSV group, Universität Leipzig, Germany

Keywords: Automatic Viewpoint Selection, Automatic Perspective, Visual Movement Feedback.

Abstract: We present a novel method to automatically select a viewpoint optimized for the interactive display of a physical exercise which is shown using a human skeleton-like avatar with additional visual motor feedback. Expressive viewpoints are crucial for the users to be able to understand and interactively adapt to the feedback in all its spatial aspects. Selecting camera perspectives for these viewpoints can be challenging when the presentation includes specific visual feedback cues in addition to the instantaneous pose, as many different requirements have to be taken into consideration in this case. The users continuously correcting their movements according to the visual real-time feedback represents a special case of human-computer interaction. Our algorithm employs principal component analysis (PCA) to determine informative viewing directions for the overall pose and specific feedback cues shown at different joints. The final viewpoints are synthesized from the obtained directions in a per-frame manner. To evaluate our method we conducted a user study with 39 participants. They were asked to choose from four exercise videos with motor feedback generated by the presented

method and three competing existing approaches. Additionally, to validate our approach's assumptions, we asked the participants to freely choose a viewpoint, which they considered optimal for the provided motor feedback. The results of the study show that our algorithm was most frequently chosen as being the most informative. Furthermore, our method proved much faster than previous viewpoint selection methods, as it does not require information about upcoming frames. This makes our approach most suitable for real-time and interactive applications.

Complete Paper #37

AR Authoring: How to Reduce Errors from the Start?

Camille Truong-Allié¹, Martin Herbeth² and Alexis Paljic¹

¹ Centre de Robotique, Mines Paris - PSL, Paris, France

² Spectral TMS, Évry-Courcouronnes, France

Keywords: Augmented Reality, Task Assistance, AR Authoring Interface, Knowledge Transmission.

Abstract: Augmented Reality (AR) can be used to efficiently guide users in procedures by overlaying virtual content onto the real world. To facilitate the use of AR for creating procedures, multiple AR authoring tools have been introduced. However, they often assume that authors digitize the procedure perfectly well the first time; this is yet hardly the case. We focus on how AR authoring tools can support authors during the procedure formalization. We introduce three authoring methods. The first one is a video-based method, where a video recording is done before procedure digitization, to improve procedure recency, the second one an in-situ method, where the digitization is made in the procedure environment, to improve context, and the last one is the baseline method, where AR authors digitize from memory. We assess the quality of the procedures resulting from these authoring methods with two simple yet underexplored metrics: the number of errors and the number of versions until the final procedure. We collected feedbacks from AR authors in a field study to validate their significance. We found that participants' performance was better with the video-based method, followed by the in-situ and then the baseline methods. The field study showed the advantages of the different methods depending on the use case and validated the importance of measuring digitization error.

Complete Paper #112

Using Fitts' Law to Compare Sonification Guidance Methods for Target Reaching Without Vision

Coline Fons^{1,2}, Sylvain Huet², Denis Pellerin² and Christian Graff¹

¹ Univ. Grenoble Alpes, Univ. Savoie Mont-Blanc, CNRS, LPNC, 38000 Grenoble, France

² Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

Keywords: Spatial Interaction, Fitts' Law, Sound Guidance, Sonification, Visual Impairment, User Performance Evaluation.

Abstract: Visually impaired people often face challenges in spatial interaction tasks. Sensory Substitution Devices assist them in reaching targets by conveying spatial deviations to the target through sound. Typically, sound guidance systems are evaluated by target reaching times. However, reaching times are influenced by target size and user-target distance, which varies across studies. We propose to explore the potential of Fitts' law

for evaluating such systems. In a preliminary experiment, visually impaired and sighted participants used non-spatialized sonification to reach 3D virtual targets. Movement times were correlated with the Index of Difficulty, confirming that Fitts' law is a valuable model to evaluate target-reaching in 3D non-visual interfaces, even with non-spatialized sonification as feedback. In a second experiment, we compared two non-spatialized sonifications, one dissociating the height and azimuthal direction of the target, and the other combining them into a single 3D angle. Fitts law did allow the comparison of performance in favor of the first sonification. The potential of using Fitts' law to compare performances across studies using different experimental settings deserves exploration in future research. We encourage researchers to provide the full linear regression equations obtained when using Fitts' law, to facilitate standardized performance comparisons across studies.

Oral Presentations (Online) 3 **IVAPP**
09:00 - 10:30 **Room Online 2**
Information Visualization Theory and Applications

Complete Paper #295

Using Retrieval Augmented Generation to Build the Context for Data-Driven Stories

Angelica Lo Duca

Institute of Informatics and Telematics, National Research Council, via G. Moruzzi 1, 56124 Pisa, Italy

Keywords: Data Storytelling, Retrieval Augmented Generation, Data Visualization, Generative AI.

Abstract: Data Storytelling (DS) is building data-driven stories to communicate the result of a data analysis process effectively. However, it may happen that data storytellers lack the competences to build compelling texts to include in the data-driven stories. In this paper, we propose a novel strategy to enhance DS by automatically generating context for data-driven stories, leveraging the capabilities of Generative AI (GenAI). This contextual information provides the background knowledge necessary for the audience to understand the story's message fully. Our approach uses Retrieval Augmented Generation (RAG), which adapts large language models (LLMs), the core concept behind GenAI, to the specific domain required by a data-driven story. We demonstrate the effectiveness of our method through a practical case study on salmon aquaculture, showcasing the ability of GenAI to enrich DS with relevant context. We also describe some possible strategies to evaluate the generated context and ethical issues may raise when using GenAI in DS.

Complete Paper #372

Particle-Wise Higher-Order SPH Field Approximation for DVR

Jonathan Fischer¹, Martin Schulze¹, Paul Rosenthal² and Lars Linsen³

¹ *Department of Computer Science, Chemnitz Technical University, Str. der Nationen 62, 09111 Chemnitz, Germany*

² *Institute for Visual and Analytic Computing, University of Rostock, Albert-Einstein-Str. 22, 18059 Rostock, Germany*

³ *Institute of Computer Science, University of Münster, Einsteinstr. 62, 48149 Münster, Germany*

Keywords: Scientific Visualization, Ray Casting, Higher-Order Approximation, Volume Rendering, Scattered Data, SPH.

Abstract: When employing Direct Volume Rendering (DVR)

for visualizing volumetric scalar fields, classification is generally performed on a piecewise constant or piecewise linear approximation of the field on a viewing ray. Smoothed Particle Hydrodynamics (SPH) data sets define volumetric scalar fields as the sum of individual particle contributions, at highly varying spatial resolution. We present an approach for approximating SPH scalar fields along viewing rays with piecewise polynomial functions of higher order. This is done by approximating each particle contribution individually and then efficiently summing the results, thus generating a higher-order representation of the field with a resolution adapting to the data resolution in the volume.

Oral Presentations (Online) 3 **VISAPP**
09:00 - 10:30 **Room Online 1**
Machine Learning Technologies for Vision

Complete Paper #52

Augmenting Human-Robot Collaboration Task by Human Hand Position Forecasting

Shyngyskhan Abilkassov¹, Michael Gentner^{2,3} and Mirela Popa¹

¹ *Department of Advanced Computing Sciences, Faculty of Science and Engineering, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands*

² *Technical University of Munich, Munich, Germany*

³ *BMW AG, Munich, Germany*

Keywords: Egocentric Vision, Human-Robot Collaboration.

Abstract: Human-Robot collaboration (HRC) plays a critical role in enhancing productivity and safety across various industries. While reactive motion re-planning strategies have proven useful, there is a pressing need for proactive control involving computing human intentions to enable efficient collaboration. This work addresses this challenge by proposing a deep learning-based approach for forecasting human hand trajectories and a heuristic optimization algorithm for proactive robotic task sequencing problem optimization. This work presents a human hand trajectory forecasting deep learning model that achieves state-of-the-art performance on the Ego4D Future Hand Prediction benchmark in all evaluation metrics. In addition, this work presents a problem formulation and a Dynamic Variable Neighborhood Search (DynamicVNS) heuristic optimization algorithm enabling robot to pre-plan their task sequence to avoid human hands. The proposed algorithm exhibits significant computational improvements over the generalized VNS approach. The final framework efficiently incorporates predictions made by the deep learning model into the task sequencer, which is evaluated in an experimental setup for the HRC use-case of the UR10e robot in a visual inspection task. The results indicate the effectiveness and practicality of the proposed approach, showcasing its potential to improve human-robot collaboration in various industrial settings.

Complete Paper #69

Vision-Perceptual Transformer Network for Semantic Scene Understanding

Mohamad Alansari¹, Hamad AlRemeithi^{1,2}, Bilal Hassan^{1,3}, Sara Alansari¹, Jorge Dias^{1,3}, Majid Khonji^{1,3}, Naoufel Werghi^{1,3,4} and Sajid Javed^{1,3}

¹ Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, U.A.E.

² Research and Technology Development Department, Tawauz Technology & Innovation, Abu Dhabi, U.A.E.

³ Center for Autonomous Robotic Systems, Khalifa University, Abu Dhabi, U.A.E.

⁴ Center for Cyber-Physical Systems, Khalifa University, Abu Dhabi, U.A.E.

Keywords: Attention Mechanisms, Computational Resources, Pyramid Vision Transformers, Scene Understanding, Semantic Segmentation.

Abstract: Semantic segmentation, essential in computer vision, involves labeling each image pixel with its semantic class. Transformer-based models, recognized for their exceptional performance, have been pivotal in advancing this field. Our contribution, the Vision-Perceptual Transformer Network (VPTN), ingeniously combines transformer encoders with a feature pyramid-based decoder to deliver precise segmentation maps with minimal computational burden. VPTN's transformative power lies in its integration of the pyramiding technique, enhancing multi-scale variations handling. In direct comparisons with Vision Transformer-based networks and variants, VPTN consistently excels. On average, it achieves 4.2%, 3.41%, and 6.24% higher mean Intersection over Union (mIoU) compared to Dense Prediction (DPT), Data-efficient image Transformer (DeiT), and Swin Transformer networks, while demanding only 15.63%, 3.18%, and 10.05% of their Giga Floating-Point Operations (GFLOPs). Our validation spans five diverse datasets, including Cityscapes, BDD100K, Mapillary Vistas, CamVid, and ADE20K. VPTN secures the position of state-of-the-art (SOTA) on BDD100K and CamVid and consistently outperforms existing deep learning models on other datasets, boasting mIoU scores of 82.6%, 67.29%, 61.2%, 86.3%, and 55.3%, respectively. Impressively, it does so with an average computational complexity just 11.44% of SOTA models. VPTN represents a significant advancement in semantic segmentation, balancing efficiency and performance. It shows promising potential, especially for autonomous driving and natural setting computer vision applications.

Complete Paper #183

Social Distancing Monitoring by Human Detection Through Bird's-Eye View Technique

Gona Rozhbayani¹, Amel Tuama¹ and Fadwa Al-Azzo²

¹ Kirkuk Technical Engineering College, Northern Technical University, Kirkuk 36001, Iraq

² Technical Engineering College Mosul, Northern Technical University, Mosul 41001, Iraq

Keywords: Human Detection, Social Distancing, Deep Learning, YOLOv5.

Abstract: The objective of this study is to offer a YOLOv5 deep learning-based system for social distance monitoring. The YOLOv5 model has been used to detect humans in real-time video frames, and to obtain information on the detected bounding box for the bird's eye view perspective technique. The pairwise distances of the identified bounding box centroid of people are calculated by utilizing euclidean distance. In addition, a threshold

value has been set and applied as an approximation of social distance to pixels for determining social distance violations between people. The effectiveness of this proposed system is tested by experiments on different four video frames. The suggested system's performance showed a high level of efficiency in monitoring social distancing accurately up to 100%.

Complete Paper #356

Visual Perception of Obstacles: Do Humans and Machines Focus on the Same Image Features?

Constantinos Kyriakides¹, Marios Thoma^{1,2}, Zenonas Theodosiou^{1,3}, Harris Partaourides⁴, Loizos Michael^{1,2} and Andreas Lanitis^{1,5}

¹ CYENS Centre of Excellence, Nicosia, Cyprus

² Open University of Cyprus, Nicosia, Cyprus

³ Department of Communication and Internet Studies, Cyprus University of Technology, Limassol, Cyprus

⁴ AI Cyprus Ethical Novelities Ltd, Limassol, Cyprus

⁵ Department of Multimedia and Graphic Arts, Cyprus University of Technology, Limassol, Cyprus

Keywords: Deep Learning Algorithms, Explainability, Eye Tracking, Heatmaps, Obstacle Recognition.

Abstract: Contemporary cities are fractured by a growing number of barriers, such as on-going construction and infrastructure damages, which endanger pedestrian safety. Automated detection and recognition of such barriers from visual data has been of particular concern to the research community in recent years. Deep Learning (DL) algorithms are now the dominant approach in visual data analysis, achieving excellent results in a wide range of applications, including obstacle detection. However, explaining the underlying operations of DL models remains a key challenge in gaining significant understanding on how they arrive at their decisions. The use of heatmaps that highlight the focal points in input images that helped the models reach their predictions has emerged as a form of post-hoc explainability for such models. In an effort to gain insights into the learning process of DL models, we studied the similarities between heatmaps generated by a number of architectures trained to detect obstacles on sidewalks in images collected via smartphones, and eye-tracking heatmaps generated by humans as they detect the corresponding obstacles on the same data. Our findings indicate that the focus points of humans more closely align with those of a Vision Transformer architecture, as opposed to the other network architectures we examined in our experiments.

Tutorial

09:00 - 10:30

Room Space 4C

First Person (Egocentric) Vision: History and Applications

First Person (Egocentric) Vision: History and Applications

Francesco Ragusa
University of Catania, Italy

Abstract: Wearable devices equipped with a camera and computing abilities are attracting the attention of both the market and the society, with commercial devices more and more available and many companies announcing the upcoming release of new devices. The main appeal of wearable devices is due to their mobility and to their ability to enable user-machine interaction through Augmented Reality. Due to these characteristics, wearable devices provide an ideal platform to develop intelligent assistants able to assist humans and augment their abilities, for

which Artificial Intelligence and Computer Vision play a major role. Differently from classic computer vision (the so called "third person vision"), which analyses images collected from a static point of view, first person (egocentric) vision assume that images are collected from the point of view of the user, which gives privileged information on the user's activities and the way they perceive and interact with the world. Indeed, the visual data acquired with wearable cameras usually provides useful information about the users, their intentions, and how they interact with the world. This tutorial will discuss the challenges and opportunities offered by first person (egocentric) vision, covering the historical background and seminal works, presenting the main technological tools and building blocks, and discussing applications.

Keynote Lecture
10:45 - 11:45

VISIGRAPP
Room Plenary

Multi-Modal Human-Machine Interaction: Joint Optimization of Single Modalities and Automatic Learning of Communication Channel Fusion

Gerhard Rigoll

Technical University of Munich, Munich, Germany

Abstract: In multi-modal human-machine communication, users interact with machines using different human communication channels, such as e.g. voice, vision or haptics. It is therefore not amazing that human-machine communication benefited strongly from the extremely dynamic development of advanced machine learning methods during the last decade, since they were the driving factors in most classical pattern recognition areas, such as speech and emotion recognition, or computer vision. In this talk, some of the recent research outcomes from the author's institution will be introduced, including e.g. face recognition from partial and occluded face information, recognition for low resolution face images, or action recognition including gait identification with graph neural networks. The talk will end with the presentation of a multi-modal recognition task for a multi-party speaker activity detection scenario where advanced deep learning methods are not only employed for single modality recognition but especially for the fusion of audio-visual information to solve a real multi-modal complex recognition problem. This approach contributes to the future perspective of human-machine communication, namely to employ advanced machine learning methods to jointly optimize the recognition components for the different modalities as well as automatically learning the strategies for their fusion, to create truly multi-modal interactive systems.

Session 4A
12:00 - 13:00
Segmentation and Grouping

VISAPP
Room Space 1

Complete Paper #166

Semantic State Estimation in Robot Cloth Manipulations Using Domain Adaptation from Human Demonstrations

Georgies Tzelepis¹, Eren Aksoy², Júlia Borràs¹ and Guillem Alenyà¹

¹ Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain

² Halmstad University, Center for Applied Intelligent Systems Research, Halmstad, Sweden

Keywords: Robotic Perception, Garment Manipulation, Semantics, Cloth, Transfer Learning, Domain Adaptation.

Abstract: Deformable object manipulations, such as those involving textiles, present a significant challenge due to their high dimensionality and complexity. In this paper, we propose a solution for estimating semantic states in cloth manipulation tasks. To this end, we introduce a new, large-scale, fully-annotated RGB image dataset of semantic states featuring a diverse range of human demonstrations of various complex cloth manipulations. This effectively transforms the problem of action recognition into a classification task. We then evaluate the generalizability of our approach by employing domain adaptation techniques to transfer knowledge from human demonstrations to two distinct robotic platforms: Kinova and UR robots. Additionally, we further improve performance by utilizing a semantic state graph learned from human manipulation data.

Complete Paper #103

Stereo-Event-Camera-Technique for Insect Monitoring

Regina Pohle-Fröhlich, Colin Gebler and Tobias Bolten

Institute for Pattern Recognition, Niederrhein University of Applied Sciences, Krefeld, Germany

Keywords: Event Camera, Segmentation, Insect Monitoring, Depth Estimation.

Abstract: To investigate the causes of declining insect populations, a monitoring system is needed that automatically records insect activity and additional environmental factors over an extended period of time. For this reason, we use a sensor-based method with two event cameras. In this paper, we describe the system, the view volume that can be recorded with it, and a database used for insect detection. We also present the individual steps of our developed processing pipeline for insect monitoring. For the extraction of insect trajectories, a U-Net based segmentation was tested. For this purpose, the events within a time period of 50 ms were transformed into a frame representation using four different encoding types. The tested histogram encoding achieved the best results with an F1 score for insect segmentation of 0.897 and 0.967 for plant movement and noise parts. The detected trajectories were then transformed into a 4D representation, including depth, and visualized.

Complete Paper #168

Instance Segmentation of Event Camera Streams in Outdoor Monitoring Scenarios

Tobias Bolten¹, Regina Pohle-Fröhlich¹ and Klaus Tönnies²

¹ Institute for Pattern Recognition, Hochschule Niederrhein, Krefeld, Germany

² Department of Simulation and Graphics, University of Magdeburg, Germany

Keywords: Dynamic Vision Sensor, Instance Segmentation, Outdoor Environment.

Abstract: Event cameras are a new type of image sensor. The pixels of these sensors operate independently and asynchronously from each other. The sensor output is a variable rate data stream that spatio-temporally encodes the detection of brightness changes. This type of output and sensor operating paradigm poses processing challenges for computer vision applications, as frame-based methods are not natively applicable. We provide the first systematic evaluation of different state-of-the-art deep learning based instance segmentation approaches in the context of event-based outdoor surveillance. For processing, we consider

transforming the event output stream into representations of different dimensionalities, including point-, voxel-, and frame-based variants. We introduce a new dataset variant that provides annotations at the level of instances per output event, as well as a density-based preprocessing to generate regions of interest (RoI). The achieved instance segmentation results show that the adaptation of existing algorithms for the event-based domain is a promising approach.

Session 4B
12:00 - 13:00
Assistive Computer Vision

VISAPP
Room Space 3B

Complete Paper #43

Enhanced Multimodal Timely Prediction of Pulmonary Fibrosis Progression with Uncertainty Estimation from Chest CT Images and Clinical Metadata

Mohamed Dahmane

Computer Research Institute of Montreal (CRIM), 405 Av. Ogilvy #101, Montreal, Quebec, Canada

Keywords: CT Scans, Uncertainty Estimation, Pulmonary Fibrosis, Multimodal Deep Learning, Clinical Metadata.

Abstract: Pulmonary Fibrosis (PF) is a progressive chronic illness in which the lung tissues become increasingly scarred and damaged, leading to irreversible loss of their capacity to oxygenate vital organs. The specific causes of the illness are often unknown in many cases. Assessment of the severity of the lung disease is critical for physicians to diagnose PF early, control disease decline, and manage damage progression. The Forced Vital Capacity (FVC) of the lungs measured by a spirometer, is a good indicator of the severity of the condition of the lungs. In this work, we investigated an approach for early diagnosis of PF and showcased a multimodal architecture that predicts the FVC of patients at different stages of the disease. We propose an anti-Elu intermediate block and an anti-Relu confidence block to predict the pulmonary fibrosis progression. The uncertainty estimation block proved effective in predicting the FVC using data from initial spirometry measurements, clinical meta-data and CT images. Evaluation of the model on the OSIC pulmonary fibrosis progression dataset showed improved performance compared to state-of-the-art methods, with an average modified Laplace log-likelihood score of -6.8227, on a private test set.

Complete Paper #244

Physical Ergonomics Anticipation with Human Motion Prediction

Mattias Billast¹, Jonas De Bruyne², Klaas Bombeke², Tom De Schepper^{1,3} and Kevin Mets⁴

¹ University of Antwerp - imec, IDLab, Department of Computer Science, Sint-Pietersvliet 7, 2000 Antwerp, Belgium

² Imec-mict-UGent, Department of Communication Sciences, Ghent, Belgium

³ AI & Data Department, Imec, Leuven, Belgium

⁴ University of Antwerp - imec, IDLab, Faculty of Applied Engineering, Sint-Pietersvliet 7, 2000 Antwerp, Belgium

Keywords: Motion Prediction, Ergonomics, VR, AI.

Abstract: Good physical ergonomics is a crucial aspect of performing repetitive tasks sustainably for a long period. We de-

veloped a VR training environment that improves the ergonomics and experience of the user during a task. Through human motion prediction, we can predict the posture of the user accurately up to three seconds ahead of time. Based on this posture, a physical ergonomics score, called REBA (Hignett and McAtamney, 2000), is computed and can warn the user ahead of time to adapt their posture. We used the lightweight STS-GCN model (Sofianos et al., 2021) as it can infer predictions in real-time to give feedback to the users. We show in our experiments that using multi-task learning improves human motion prediction significantly. Our method is generally applicable for various manual tasks as almost all tasks can be simulated in a VR environment.

Complete Paper #453

Automatic Assessment of Skill and Performance in Fencing Footwork

Filip Malawski, Marek Krupa and Ksawery Kapela

Faculty of Computer Science, AGH University of Science and Technology, Krakow, Poland

Keywords: Action Recognition, Pattern Recognition, Sports Support, Fencing, Assistive Computer Vision.

Abstract: Typically human action recognition methods focus on the detection and classification of actions. In this work, we consider qualitative evaluation of sports actions, namely in fencing footwork, including technical skill and physical performance. In cooperation with fencing coaches, we designed, recorded, and labeled an extensive dataset including 28 variants of incorrect executions of fencing footwork actions as well as corresponding correct variants. Moreover, the dataset contains action sequences for action recognition tasks. This is the most extensive fencing action dataset collected to date. We propose and evaluate an expert system, based on pose estimation in video data, for measuring relevant motion parameters and distinguishing between correct and incorrect executions of actions. Additionally, we validate a method for temporal segmentation and classification of actions in sequences. The obtained results indicate that the proposed solution can provide relevant feedback in fencing training.

Session 4C
12:00 - 13:00
Self-supervised Learning

VISAPP
Room Space 3C

Complete Paper #76

Deep Active Learning with Noisy Oracle in Object Detection

Marius Schubert¹, Tobias Riedlinger², Karsten Kahl¹ and Matthias Rottmann¹

¹ School of Mathematics and Natural Sciences, IZMD, University of Wuppertal, Germany

² Institute of Mathematics, Technical University Berlin, Germany

Keywords: Active Learning, Label Noise, Robustness, Label Error Detection, Object Detection.

Abstract: Obtaining annotations for complex computer vision tasks such as object detection is an expensive and timeintense endeavor involving numerous human workers or expert opinions. Reducing the amount of annotations required while maintaining algorithm performance is, therefore, desirable for machine learning practitioners and has been successfully achieved by active learning. However, it is not merely the amount of annotations which influences model performance but also the annotation quality. In

practice, oracles that are queried for new annotations frequently produce significant amounts of noise. Therefore, cleansing procedures are oftentimes necessary to review and correct given labels. This process is subject to the same budget as the initial annotation itself since it requires human workers or even domain experts. Here, we propose a composite active learning framework including a label review module for deep object detection. We show that utilizing part of the annotation budget to correct the noisy annotations partially in the active dataset leads to early improvements in model performance, especially when coupled with uncertainty-based query strategies. The precision of the label error proposals significantly influences the measured effect of the label review. In our experiments we achieve improvements of up to 4.5mAP points by incorporating label reviews at equal annotation budget.

Complete Paper #88

A Comparative Evaluation of Self-Supervised Methods Applied to Rock Images Classification

Van Nguyen^{1,2}, Dominique Fourer², Désiré Sidibé², Jean-François Lecomte¹ and Souhail Youssef¹

¹ IFPEN, Ruel-Malmaison, France

² IBISC - Univ.Évry Paris-Saclay,Évry-Courcouronnes, France

Keywords: Self-Supervised Learning, Representation Learning, Digital Rock Physics Digital Rock Physics (DRP), Image Classification.

Abstract: Digital Rock Physics DRP is a discipline that employs advanced computational techniques to analyze and simulate rock properties at the pore-scale level. Recently, Self-Supervised Learning (SSL) has shown promising outcomes in various application domains, but its potential in DRP applications remains largely unexplored. In this study, we propose to assess several self-supervised representation learning methods designed for automatic rock category recognition. Hence, we demonstrate how different SSL approaches can be specifically adapted for DRP, and comparatively evaluated on a new dataset. Our objective is to leverage unlabeled micro-CT (Computed Tomography) image data to train models that capture intricate rock features and obtain representations that enhance the accuracy of classical machine-learning-based rock images classification. Experimental results on a newly proposed rock images dataset indicate that a model initialized using SSL pretraining outperforms its non-self-supervised learning counterpart. Particularly, we find that MoCo-v2 pretraining provides the most benefit with limited labeled training data compared to other models, including supervised model.

Complete Paper #169

Diverse Data Selection Considering Data Distribution for Unsupervised Continual Learning

Naoto Hayashi, Naoki Okamoto, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi

Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi, Japan

Keywords: Deep Learning, Image Classification, Self-Supervised Learning, Continual Learning.

Abstract: In continual learning, the train data changes during the learning process, making it difficult to solve previously learned tasks as the model adapts to the new task data. Many methods have been proposed to prevent catastrophic forgetting in continual

learning. To overcome this problem, Lifelong Unsupervised Mixup (LUMP) has been proposed, which is capable of learning unlabeled data that can be acquired in the real world. LUMP trains a model by self-supervised learning method, and prevents catastrophic forgetting by using a mixup of a data augmentation method and a replay buffer that stores a part of the data used to train previous tasks. However, LUMP randomly selects data to store in the replay buffer from the train data, which may bias the stored data and cause the model to specialize in some data. Therefore, we propose a method for selecting data to be stored in the replay buffer for unsupervised continuous learning method. The proposed method splits the distribution of train data into multiple clusters using the k-means clustering. Next, one piece of data is selected from each cluster. The data selected by the proposed method preserves the distribution of the original data, making it more useful for self-supervised learning.

Session 4D

12:00 - 13:00

VISAPP

Room Space 4A

Image Formation, Acquisition Devices and Sensors

Complete Paper #298

The Risk of Image Generator-Specific Traces in Synthetic Training Data

Georg Wimmer, Dominik Söllinger and Andreas Uhl

Paris Lodron University Salzburg, Jakob-Haringer-Str. 2, 5020 Salzburg, Austria

Keywords: Image Synthesis, GAN, Diffusion Model, Synthetic Artifacts.

Abstract: Deep learning based methods require large amounts of annotated training data. Using synthetic images to train deep learning models is a faster and cheaper alternative to gathering and manually annotating training data. However, synthetic images have been demonstrated to exhibit a unique model-specific fingerprint that is not present in real images. In this work, we investigate the effect of such model-specific traces on the training of CNN-based classifiers. Two different methods are applied to generate synthetic training data, a conditional GAN-based image-to-image translation method (BicycleGAN) and a conditional diffusion model (Palette). Our results show that CNN-based classifiers can easily be fooled by generator-specific traces contained in synthetic images. As we will show, classifiers can learn to discriminate based on the traces left by the generator, instead of class-specific features.

Complete Paper #312

FuDensityNet: Fusion-Based Density-Enhanced Network for Occlusion Handling

Zainab Ouairdirhi^{1,2}, Otmame Amei¹, Mostapha Zbakh² and Sidi Mahmoudi¹

¹ Computer and Management Engineering Department, FPMS, University of Mons, Belgium

² Communication Networks Department, ENSIAS, Mohammed V University in Rabat, Morocco

Keywords: Object Detection, Occlusion Handling, Voxelization, Density-Aware, Multimodal Fusion.

Abstract: Our research introduces an innovative approach for detecting occlusion levels and identifying objects with varying degrees of occlusion. We integrate 2D and 3D data through advanced network architectures, utilizing voxelized density-based

occlusion assessment for improved visibility of occluded objects. By combining 2D image and 3D point cloud data through carefully designed network components, our method achieves superior detection accuracy in complex scenarios with occlusions. Experimental evaluation demonstrates adaptability across concatenation techniques, resulting in notable Average Precision (AP) improvements. Despite initial testing on a limited dataset, our method shows competitive performance, suggesting potential for further refinement and scalability. This research significantly contributes to advancements in effective occlusion handling for object detection methodologies. The abstract and conclusion highlight the substantial increase in AP achieved through our model.

Complete Paper #325

Multidimensional Compressed Sensing for Spectral Light Field Imaging

Wen Cao, Ehsan Miandji and Jonas Unger

Media and Information Technology, Department of Science and Technology, Linköping University, SE-601 74 Norrköping, Sweden

Keywords: Spectral Light Field, Compressive Sensing.

Abstract: This paper considers a compressive multi-spectral light field camera model that utilizes a one-hot spectral-coded mask and a microlens array to capture spatial, angular, and spectral information using a single monochrome sensor. We propose a model that employs compressed sensing techniques to reconstruct the complete multi-spectral light field from under-sampled measurements. Unlike previous work where a light field is vectorized to a 1D signal, our method employs a 5D basis and a novel 5D measurement model, hence, matching the intrinsic dimensionality of multispectral light fields. We mathematically and empirically show the equivalence of 5D and 1D sensing models, and most importantly that the 5D framework achieves orders of magnitude faster reconstruction while requiring a small fraction of the memory. Moreover, our new multidimensional sensing model opens new research directions for designing efficient visual data acquisition algorithms and hardware.

Session 4A
12:00 - 13:00
Information and Virtual Reality

GRAPP
Room Space 4B

Complete Paper #262

Real-Time 3D Information Visualization on Mobile Devices: Efficient Occlusion Detection for Geospatial Applications

Agata Migalska

Department of Control Systems and Mechatronics, Wrocław University of Science and Technology, Wrocław, Poland

Keywords: Augmented Reality, Geospatial Data Visualization, Occlusion Detection.

Abstract: The rapid advancement of geospatial applications on mobile devices has revolutionized outdoor activities and professions by providing invaluable tools for navigating, exploring, and highlighting geographic features. These applications often face notable challenges, including limited connectivity as well as battery and storage constraints. The objective of this paper is to provide users with a seamless user experience and augmented reality interactions in diverse outdoor scenarios within geospatial applications on mobile devices. Towards this goal, we introduce

visibility map of a point that contains distances from the point to the closest obstacles at multiple angular viewpoints. We propose that every point of interest is associated with a such a visibility map, pre-calculated on a digital world model, in order to render annotated data in real-time without relying on time-consuming depth estimation. We specifically address the aspect of displaying 3D virtual information that is registered at different elevations, justified by the importance of elevation information in domains such as earth science, culture science or construction and by a lack of methods that would allow that. Through our study, we demonstrate the real-time rendering capability and accuracy of our proposed method. We present the findings, discuss potential limitations, and suggest future research directions.

Complete Paper #363

Exploring Foveation Techniques for Virtual Reality Environments

Razeen Hussain, Manuela Chessa and Fabio Solari

Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genoa, Genoa, Italy

Keywords: Foveation, Virtual Reality, Immersive Media, Image Quality, Gaze-Contingency, Visual Perception.

Abstract: Virtual reality technology is constantly advancing leading to the creation of novel experiences for the user. High-resolution displays often are accompanied by higher processing power needs. Foveated rendering is a potential solution to circumvent this issue as it can significantly reduce the computational load by rendering only the area where the user is looking with higher detail. In this work, we compare different foveated rendering algorithms in terms of the quality of the final rendered image. The focus of this work is on evaluating 4K images. These algorithms are also compared based on computational models of human visual processing. Our analysis revealed that the non-linear content-aware algorithm performs best.

Abstract #458

Domain Specific Language for Type Casting in Image Processing

Yamato Kanetaka¹, Haruki Nogami¹, Yuki Naganawa¹ and Norishige Fukushima²

¹ Nagoya Institute of Technology, Japan

² Graduate School of Engineering, Nagoya Institute of Technology, Nagoya, Japan

Keywords: Domain Specific Language, Image Processing, Filtering.

Abstract: High-speed image processing is growing in importance as image resolution is also growing. Type-casting or quantizing data has a variety of benefits: reduction in memory usage and enhancements to parallelism. For instance, casting data from float to byte results in memory usage of only one-fourth, and the parallelism increases 4 times vector operations. However, the accuracy of the output after quantization varies depending on the method of quantization. Furthermore, there are various options for quantizing the stage of multiple steps in image processing and to what extent. Accuracy changes based on these combinations, and there are countless such combinations to consider. Consequently, there is a need to write codes for an innumerable array of possible combinations when using a usual programming language (e.g., C, C++). There is Halide, a domain-specific language (DSL), to solve complications of image processing. However, Halide cannot

dynamically modify types without changing algorithm parts. This results in writing redundant code changing types. Therefore, in this paper, we develop a prototype within a Halide-based language that allows for a decoupling algorithm and type-casting codes.

The proposed DSL, named CastFunc, can solve the previous Halide's limitation. CastFunc is a DSL based on Halide for arbitrarily image processing that allows adjustments to the timing and level of casting and quantizing. CastFunc can describe casting and quantizing without changing the algorithm part, and have the same ability as Halide's Func.

For casting, CastFunc has two casting types: cast type and output type. The cast type is a type for casting and computing values; that is, CastFunc computes all expressions on the cast type. The output type is the final output type, i.e., CastFunc cast values from computed values for output.

For quantizing, CastFunc has the quantization factor. We use a quantization factor of 255 to cast the float kernel into the integer kernel, whose range domain is mapped from [0.0:1.0] to [0:255]. CastFunc generates Halide's Func using interfaces of Halide::Internal that can generate Halide code to define Func programmatically.

Experimental results show the importance of types on image processing by comparing 16-bit integer-based image processing with 32-bit floating-point-based image processing.

Session 4A
12:00 - 13:00
Text Visualization and Storytelling

IVAPP
Room Space 4D

Complete Paper #237

Quantifying Topic Model Influence on Text Layouts Based on Dimensionality Reductions

Daniel Atzberger¹, Tim Cech², Willy Scheibel¹, Jürgen Döllner¹ and Tobias Schreck³

¹ Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, Germany

² Digital Engineering Faculty, University of Potsdam, Germany

³ Graz University of Technology, Austria

Keywords: Topic Modeling, Dimensionality Reductions, Text Spatializations.

Abstract: Text spatializations for text corpora often rely on two-dimensional scatter plots generated from topic models and dimensionality reductions. Topic models are unsupervised learning algorithms that identify clusters, so-called topics, within a corpus, representing the underlying concepts. Furthermore, topic models transform documents into vectors, capturing their association with topics. A subsequent dimensionality reduction creates a two-dimensional scatter plot, illustrating semantic similarity between the documents. A recent study by Atzberger et al. has shown that topic models are beneficial for generating two-dimensional layouts. However, in their study, the hyperparameters of the topic models are fixed, and thus the study does not analyze the impact of the topic models' quality on the resulting layout. Following the methodology of Atzberger et al., we present a comprehensive benchmark comprising (1) text corpora, (2) layout algorithms based on topic models and dimensionality reductions, (3) quality metrics for assessing topic models, and (4) metrics for evaluating two-dimensional layouts' accuracy and cluster separation. Our study involves an exhaustive evaluation of numerous parameter configurations, yielding a dataset that quantifies the quality of each dataset-layout algorithm combination. Through a rigorous analysis of this dataset, we derive practical guidelines for effectively employing topic models in text spatializations. As a main result, we conclude that the quality of a topic model measured by coherence is positively correlated to the layout quality in the case of Latent Semantic Indexing and Non-Negative Matrix Factorization.

Complete Paper #436

A Survey on Storytelling Techniques for Heritage on Nazi Persecution

Niek Meffert¹, Camilla Østergaard¹, Stefan Jänicke¹, Richard Khulusi², Esther Rachow³ and Nicklas Andersen¹

¹ Centre for Visual Data Science, University of Southern Denmark, Odense, Denmark

² Bergen-Belsen Memorial, Lohheide, Germany

³ The Hebrew University of Jerusalem, Jerusalem, Israel

Keywords: Visual Storytelling, Interactive Media, Cultural Heritage, Nazi Persecution, Digital Humanities.

Abstract: This paper explores Visual Storytelling (VS) as a means of conveying historical narratives, with a particular focus on Heritage related to Nazi Persecution (HNP). We refine and augment existing design spaces in information visualization to broaden the scope and emphasize rich media elements while orienting our refined design space towards VS for HNP. We analyze dimensions central for storytelling focusing on cultural heritage, while digging deeper into aspects relevant for HNP like specific types of text (testimonies, diaries, official documents) and person types (victims, survivors, persecutors). The key contribution of our study is the development of a design space uniquely tailored to HNP, which highlights critical elements and trends from existing storytelling projects, and comprehensively examines the unique challenges and opportunities within VS for HNP. Furthermore, we discuss future directions, enriching the evolving domain of VS by equipping heritage professionals and researchers with practical strategies to craft compelling narratives that aim to engage contemporary audiences and to preserve the historical accuracy and ethical integrity of HNP.

Oral Presentations (Online) 4

12:00 - 13:00

Human Computer Interaction Theory and Applications

HUCAPP

Room Online 2

Complete Paper #129

Real-Time Heart Rate Visualization for Individuals with Autism Spectrum Disorder: An Evaluation of Technology Assisted Physical Activity Application to Increase Exercise Intensity

Bo Fu¹, Katrina Orevillo¹, Dennis Lo¹, Andrew Bae¹ and Melissa Bittner²

¹ Computer Engineering and Computer Science, California State University, Long Beach, U.S.A.

² Department of Kinesiology, California State University, Long Beach, U.S.A.

Keywords: Real-Time Heart Rate Visualization, Eye Tracking, Autism Spectrum Disorder, Technology Assisted Exercise.

Abstract: Individuals with autism spectrum disorder (ASD) often experience negative relationships with physical activity and a severe lack of motivation for exercise. While specialized exercise technologies such as internet-enabled machines and mobile applications have provided some solutions for typically developing individuals, there is a lack of research in providing exercise technology that specifically considers the needs of individuals with ASD. This paper presents a real-time heart-rate visualization

application, namely the HeartRunner 2.0 App, which aims to engage individuals with ASD to exercise at higher intensity. When used to supplement exercise sessions, amongst a group of 20 participants with ASD, evaluation results showed that the App helped 83% of participants achieve higher heart rates, 66.6% to maintain heart rates at or above 90 BPM, and 27.7% to re-engage and achieve heart rates at or above 90 BPM after dropping below that threshold. Furthermore, eye tracking analyses indicate that those individuals who achieved higher heart rates have employed a more focused gaze patterns with less distributed fixations in their visual searches, as well as greater efforts in scanning various cues in the given visual scene, suggesting that visual interaction with the App may have contributed to elevated performance in the experiment.

Oral Presentations (Online) 4
12:00 - 13:00
Object Detection and Localization

VISAPP
Room Online 1

Complete Paper #184

Attentive-YOLO: On-Site Water Pipeline Inspection Using Efficient Channel Attention and Reduced ELAN-Based YOLOv7

Atanu Roy and Priyanka Bagade
Indian Institute of Technology, Kanpur, India

Keywords: YOLOv7, Computer Vision, Water Pipeline Inspection, Pipe Robot.

Abstract: The effective and dependable distribution of clean water to communities depends on the timely inspection and repair of water pipes. Traditional inspection techniques frequently require expensive physical labour, resulting in false and delayed defect detections. Current water pipeline inspection methods include radiography testing, eddy current testing, and CCTV inspection. These methods require experts to be present on-site to conduct the tests. Radiographed and CCTV images are usually used for pipeline defect detection on-site, yet real-time automatic detection is lacking. Current approaches, including YOLOv5 models with Retinex-based illumination, achieve acceptable performance but hinder fast inference due to bulky models, which is especially concerning for edge devices. This paper proposes an Attentive-YOLO model based on the state-of-the-art object detection YOLOv7 model with a reduced Efficient Layer Aggregation Network (ELAN). We propose a lightweight attention model in the head and backbone of the YOLOv7 network to improve accuracy while reducing model complexity and size. The paper aims to present an efficient model to be deployed on edge devices such as the Raspberry Pi to be used in Internet of Things (IoT) systems and on-site robotics applications like pipeline inspection robots. Based on the experiments, the proposed model, Attentive-YOLO, achieves an mAP score of 0.962 over 0.93 (1/3rd channel width) compared to the YOLOv7-tiny model, with an almost 20% decrease in model parameters.

Complete Paper #212

High Precision Single Shot Object Detection in Automotive Scenarios

Soumya A¹, C Krishna Mohan¹ and Linga Reddy Cenkeramaddi²

¹ *Department of Computer Science and Engineering, Indian Institute of Technology, Hyderabad, India*

² *Department of Information and Communication Technology, University of Agder, Grimstad, 4879, Norway*

Keywords: Deep Learning, Convolutional Neural Network, Object Detection, Multi-Class Classification, Computer Vision.

Abstract: Object detection in low-light scenarios is a challenging task with numerous real-world applications, ranging from surveillance and autonomous vehicles to augmented reality. However, due to reduced visibility and limited information in the image data, carrying out object detection in low-lighting settings brings distinct challenges. This paper introduces a novel object detection model designed to excel in low-light imaging conditions, prioritizing inference speed and accuracy. The model leverages advanced deep-learning techniques and is optimized for efficient inference on resource-constrained devices. The inclusion of cross-stage partial (CSP) connections is key to its effectiveness, which maintains low computational complexity, resulting in minimal training time. This model adapts seamlessly to low-light conditions through specialized feature extraction modules, making it a valuable resource in challenging visual environments.

Complete Paper #433

Variational Autoencoders for Pedestrian Synthetic Data Augmentation of Existing Datasets: A Preliminary Investigation

Ivan Nikolov

Computer Graphics Group, Department of Architecture, Design and Media Technology, Aalborg University, Aalborg, Denmark

Keywords: Synthetic Data, Variational Autoencoders, Object Detection, Dataset Augmentation, Surveillance.

Abstract: The requirements for more and more data for training deep learning surveillance and object detection models have resulted in slower deployment and more costs connected to dataset gathering, annotation, and testing. One way to help with this is the use of synthetic data giving more varied scenarios and not requiring manual annotation. We present our initial exploratory work in generating synthetic pedestrian augmentations for an existing dataset through the use of variational autoencoders. Our method consists of creating a large number of backgrounds and training a variational autoencoder on a small subset of annotated pedestrians. We then interpolate the latent space of the autoencoder to generate variations of these pedestrians, calculate their positions on the backgrounds, and blend them to create new images. We show that even though we do not achieve as good results as just adding more real images, we can boost the performance and robustness of a YOLOv5 model trained on a mix of real and small amounts of synthetic images. As part of this paper, we also propose the next steps to expand this approach and make it much more useful for a wider array of datasets.

Tutorial

12:00 - 13:00

Room Space 4C

First Person (Egocentric) Vision: History and Applications**First Person (Egocentric) Vision: History and Applications**

Francesco Ragusa
University of Catania, Italy

Abstract: Wearable devices equipped with a camera and computing abilities are attracting the attention of both the market and the society, with commercial devices more and more available and many companies announcing the upcoming release of new devices. The main appeal of wearable devices is due to their mobility and to their ability to enable user-machine interaction through Augmented Reality. Due to these characteristics, wearable devices provide an ideal platform to develop intelligent assistants able to assist humans and augment their abilities, for which Artificial Intelligence and Computer Vision play a major role. Differently from classic computer vision (the so called "third person vision"), which analyses images collected from a static point of view, first person (egocentric) vision assume that images are collected from the point of view of the user, which gives privileged information on the user's activities and the way they perceive and interact with the world. Indeed, the visual data acquired with wearable cameras usually provides useful information about the users, their intentions, and how they interact with the world. This tutorial will discuss the challenges and opportunities offered by first person (egocentric) vision, covering the historical background and seminal works, presenting the main technological tools and building blocks, and discussing applications.

Session 5A

14:15 - 16:00

VISAPP
Room Space 1**Features Extraction**

Complete Paper #56

Calibration-Accuracy Measurement in Railway Overlapping Multi-Camera Systems

Martí Sánchez^{1,2}, Nerea Aranjuelo¹, Jon Iñiguez de Gordo¹, Pablo Alonso¹, Mikel García^{1,2}, Marcos Nieto¹ and Mikel Labayen³

¹ Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Donostia-San Sebastián, Spain

² University of the Basque Country (UPV/EHU), Donostia-San Sebastián, Spain

³ CAF Signalling, Donostia-San Sebastián, Spain

Keywords: Calibration-Accuracy Degree, Radial Grid, Railway Transportation, Multi-Camera Systems.

Abstract: This paper presents a method for assessing calibration quality in overlapping multi-camera systems used in railway transportation. We propose a novel approach that considers the extrinsic and intrinsic parameters of the cameras and extracts features from their images, providing relevant patterns regarding the pose of the cameras to detect cameras' calibration misalignment. Three feature extractors, including traditional image processing techniques and deep learning approaches, are evaluated and compared. The extracted features are used to provide a calibration quality metric, enabling real-time detection of camera calibration degradation. Additionally, we introduce a radial grid design that weights the contribution of pixels based on their distance from the camera's optical center. The results demonstrate the effectiveness

of our method in assessing the calibration degree between camera pairs. The findings highlight the superior performance of the deep learning approaches in analyzing the similarity degree between captured images. Overall, our method lays a solid foundation for the development of an online camera calibration pipeline.

Complete Paper #123

Kore Initial Clustering for Unsupervised Domain Adaptation

Kyungsik Lee, Youngmi Jun, Eunji Kim, Suhyun Kim, Seong Hwang and Jonghyun Choi

Department of Artificial Intelligence, Yonsei University, Seoul, Republic of Korea

Keywords: Unsupervised Domain Adaptation, Unlabeled Target Domain, Pseudo labels, K-means Clustering, Centroids.

Abstract: In unsupervised domain adaptation (UDA) literature, there exists an array of techniques to derive domain adaptive features. Among them, a particularly successful family of approaches of pseudo-labeling the unlabeled target data has shown promising results. Yet, the majority of the existing methods primarily focus on leveraging only the target domain knowledge for pseudo-labeling while insufficiently considering the source domain knowledge. Here, we hypothesize that quality pseudo labels obtained via classical K-means clustering considering both the source and target domains bring simple yet significant benefits. In particular, we propose to assign pseudo labels to the target domain's instances better aligned with the source domain labels by a simple method that modifies K-means clustering by emphasizing the strengthened notion of centroids, namely, Kore Initial Clustering (KIC). The proposed KIC is readily utilizable with a wide array of UDA models, consistently improving the UDA performance on multiple UDA datasets including Office-Home and Office-31, demonstrating the efficacy of pseudo labels in UDA.

Complete Paper #201

Feature Selection Using Quantum Inspired Island Model Genetic Algorithm for Wheat Rust Disease Detection and Severity Estimation

Sourav Samanta, Sanjay Chatterji and Sanjoy Pratihar

Department of Computer Science & Engineering, Indian Institute of Information Technology, Kalyani, West Bengal, India

Keywords: Smart Agriculture, Wheat Leaf Rust Disease Detection, Quantum Inspired Island Model Genetic Algorithm, Color-Glcm Features, Disease Severity Score.

Abstract: In the context of smart agriculture, an early disease detection system is crucial to increase agricultural yield. A disease detection system based on machine learning can be an excellent tool in this regard. Wheat is one of the world's most important crops. Leaf rust is one of the most significant wheat diseases. In this work, we have proposed a method to detect the leaf rust disease-affected areas in wheat leaves to estimate the severity of the disease. The method works on a reduced Color-GLCM (C-GLCM) feature set. The proposed feature selection method employs Quantum Inspired Island Model Genetic Algorithm to select the most compelling features from the C-GLCM set. The proposed feature selection method outperforms the classical feature selection methods. The healthy and diseased leaves are classified using four classifiers: Decision Tree, KNN, Support Vector Machine, and MLP. The MLP classifier achieved the highest accuracy of 99.20% with the proposed feature selection method. Following the detection of the diseased leaf, the k-means

algorithm has been utilized to localize the lesion area. Finally, disease severity scores have been calculated and reported for various sample leaves.

Complete Paper #359

ELSA: Expanded Latent Space Autoencoder for Image Feature Extraction and Classification

Emerson Vilar de Oliveira, Dunfrey Aragão and Luiz Gonçalves

Universidade Federal do Rio Grande do Norte, Av. Salgado Filho, 3000, Campus Universitário, 59.078-970, Natal, Brazil

Keywords: Autoencoder, Stacked Autoencoder, Latent Space, Image Classification, Feature Extraction.

Abstract: In the field of computer vision, image classification has been aiding in the understanding and labeling of images. Machine learning and artificial intelligence algorithms, especially artificial neural networks, are widely used tools for this task. In this work, we present the Expanded Latent space Autoencoder (ELSA). The ELSA network consists of more than one autoencoder in its internal structure, concatenating their latent spaces and constructing an expanded latent space. The expanded latent space aims to extract more information from input data. Thus, this expanded latent space can be used by other networks for general tasks such as prediction and classification. To evaluate these capabilities, we created an image classification network for the FashionM-NIST and MNIST datasets, achieving 99.97 and 99.98 accuracy for the test dataset. The classifier trained with the expanded latent space dataset outperforms some models in public benchmarks.

Complete Paper #366

On Granularity Variation of Air Quality Index Visualization from Sentinel-5

Jordan Cuno¹, Arthur Bezerra², Aura Conci¹ and Luiz Gonçalves²

¹ *Computer Institute, Universidade Federal Fluminense, Niterói, Brazil*

² *Grad. Prog. in Elec. and Comp. Eng., Univ. Fed. do Rio Grande do Norte, Natal, Brazil*

Keywords: Air Quality Index, Sentinel-5 Dataset, Pandemic Dynamics, Data-Driven Prediction.

Abstract: Air quality has been a hot research topic not only because it is directly related to climate change and the greenhouse effect, but most because it has been strongly associated to the transmission of respiratory diseases. Considering that different pollutants affect air quality, a methodology based on satellite data processing is proposed. The objective is to obtain images and measure the main atmospheric pollutants in Brazil. Using satellite systems with spectrometers is an alternative technology that has been recently developed for dealing with such a problem. Sentinel-5 is one of these satellites that works constantly monitoring the earth surface generating a vast amount of data mainly for climate monitoring, and that is used in this research. The main contribution of this research is a computational workflow that uses Sentinel-5 data to generate images of Brazil and its states, in addition to calculating the average value of the main atmospheric pollutants, data that can be used in the prediction of pollution as well as the identification of most polluted regions.

Session 5B
14:15 - 16:00

Deep Learning for Visual Understanding

VISAPP
Room Space 3B

Complete Paper #154

Challenging the Black Box: A Comprehensive Evaluation of Attribution Maps of CNN Applications in Agriculture and Forestry

Lars Nieradzick¹, Henrike Stephani¹, Jördis Sieburg-Rockel², Stephanie Helmling², Andrea Olbrich² and Janis Keuper³

¹ *Image Processing Department, Fraunhofer ITWM, Fraunhofer Platz 1, 67663, Kaiserslautern, Germany*

² *Thünen Institute of Wood Research, Leuschnerstraße 91, 21031, Hamburg, Germany*

³ *Institute for Machine Learning and Analysis (IMLA), Offenburg University, Badstr. 24, 77652, Offenburg, Germany*

Keywords: Explainable AI, Class Activation Maps, Saliency Maps, Attribution Maps, Evaluation.

Abstract: In this study, we explore the explainability of neural networks in agriculture and forestry, specifically in fertilizer treatment classification and wood identification. The opaque nature of these models, often considered 'black boxes', is addressed through an extensive evaluation of state-of-the-art Attribution Maps (AMs), also known as class activation maps (CAMs) or saliency maps. Our comprehensive qualitative and quantitative analysis of these AMs uncovers critical practical limitations. Findings reveal that AMs frequently fail to consistently highlight crucial features and often misalign with the features considered important by domain experts. These discrepancies raise substantial questions about the utility of AMs in understanding the decision-making process of neural networks. Our study provides critical insights into the trustworthiness and practicality of AMs within the agriculture and forestry sectors, thus facilitating a better understanding of neural networks in these application areas.

Complete Paper #200

Scale Learning in Scale-Equivariant Convolutional Networks

Mark Basting¹, Robert-Jan Brintjes¹, Thaddäus Wiedemer^{2,3}, Matthias Kümmerer², Matthias Bethge^{2,4} and Jan van Gemert¹

¹ *Computer Vision Lab, Delft University of Technology, The Netherlands*

² *Bethgelab, University of Tübingen, Geschwister-Scholl-Platz, Tübingen, Germany*

³ *Machine Learning, Max-Planck-Institute for Intelligent Systems, Max-Planck-Ring, Tübingen, Germany*

⁴ *Tübingen AI Center, Maria-von-Linden-Straße, Tübingen, Germany*

Keywords: Convolutional Neural Networks, Scale, Scale-Equivariance, Scale Learning.

Abstract: Objects can take up an arbitrary number of pixels in an image: Objects come in different sizes, and, photographs of these objects may be taken at various distances to the camera. These pixel size variations are problematic for CNNs, causing them to learn separate filters for scaled variants of the same objects which prevents learning across scales. This is addressed by scale-equivariant approaches that share features across a set of pre-determined fixed internal scales. These works, however, give little information about how to best choose the internal scales

when the underlying distribution of sizes, or scale distribution, in the dataset, is unknown. In this work we investigate learning the internal scales distribution in scale-equivariant CNNs, allowing them to adapt to unknown data scale distributions. We show that our method can learn the internal scales on various data scale distributions and can adapt the internal scales in current scale-equivariant approaches.

Complete Paper #327

AV-PEA: Parameter-Efficient Adapter for Audio-Visual Multimodal Learning

Abduljalil Radman and Jorma Laaksonen

Department of Computer Science, Aalto University, Finland

Keywords: Parameter-Efficient, Audio-Visual Adapter, Audio-Visual Fusion, Multimodal Learning.

Abstract: Fine-tuning has emerged as a widely used transfer learning technique for leveraging pre-trained vision transformers in various downstream tasks. However, its success relies on tuning a significant number of trainable parameters, which could lead to significant costs in terms of both model training and storage. When it comes to audio-visual multimodal learning, the challenge also lies in effectively incorporating both audio and visual cues into the transfer learning process, especially when the original model has been trained with unimodal samples only. This paper introduces a novel audio-visual parameter-efficient adapter (AV-PEA) designed to improve multimodal transfer learning for audio-visual tasks. Through the integration of AV-PEA into a frozen vision transformer, like the visual transformer (ViT), the transformer becomes adept at processing audio inputs without prior knowledge of audio pre-training. This also facilitates the exchange of essential audio-visual cues between audio and visual modalities, all while introducing a limited set of trainable parameters into each block of the frozen transformer. The experimental results demonstrate that our AV-PEA consistently achieves superior or comparable performance to state-of-the-art methods in a range of audio-visual tasks, including audio-visual event localization (AVEL), audio-visual question answering (AVQA), audio-visual retrieval (AVR), and audio-visual captioning (AVC). Furthermore, it distinguishes itself from competitors by enabling seamless integration into these tasks while maintaining a consistent number of trainable parameters, typically accounting for less than 3.7% of the total parameters per task.

Complete Paper #369

Improving Low-Light Image Recognition Performance Based on Image-Adaptive Learnable Module

Seitaro Ono¹, Yuka Ogino², Takahiro Toizumi², Atsushi Ito² and Masato Tsukada^{1,2}

¹ *University of Tsukuba, Ibaraki, Japan*

² *NEC Corporation, Kanagawa, Japan*

Keywords: Low-Light Image Enhancement, Image Recognition.

Abstract: In recent years, significant progress has been made in image recognition technology based on deep neural networks. However, improving recognition performance under low-light conditions remains a significant challenge. This study addresses the enhancement of recognition model performance in low-light conditions. We propose an image-adaptive learnable module which apply appropriate image processing on input images and a hyperparameter predictor to forecast optimal parameters used in the module. Our proposed approach allows for the enhancement

of recognition performance under low-light conditions by easily integrating as a front-end filter without the need to retrain existing recognition models designed for low-light conditions. Through experiments, our proposed method demonstrates its contribution to enhancing image recognition performance under low-light conditions.

Complete Paper #423

Fact-Checked Claim Detection in Videos Using a Multimodal Approach

Frédéric Rayar

LIFAT, University of Tours, Tours, France

Keywords: Multimodal Multimedia Analysis, Computational Journalism, Fact-Checking, Verified Claims Retrieval.

Abstract: The recent advances in technology and social networks have led to a phenomenon that is so called, information disorder. Our information environment is polluted by misinformation, disinformation and fake news and at a global scale. Hence, assessing the veracity of information becomes mandatory nowadays. So far, most of the efforts in the last decades have focused on analysing textual content, scraped from blogs and social media, and trying to predict the veracity of information. However, such false information also appears in multimedia content and have a peculiar life cycle throughout time that is worth leveraging. In this paper, we present a multimodal approach for detecting claims that have already been fact-checked, from videos input. Focusing on political discourse in French language, we demonstrate the feasibility of a complete system, offline and explainable.

Session 5C
14:15 - 16:00
3D Deep Learning

VISAPP
Room Space 3C

Complete Paper #174

Hand Mesh and Object Pose Reconstruction Using Cross Model Autoencoder

Chaitanya Bandi and Ulrike Thomas

Robotics and Human Machine Interaction Lab, Technical University of Chemnitz, Chemnitz, Germany

Keywords: Hand, Object, Pose, Reconstruction, Autoencoder.

Abstract: Hands and objects severely occlude each other, making it extremely challenging to estimate the hand-object pose during human-robot interactions. In this work, we propose a framework that jointly estimates 3D hand mesh and 6D object pose in real-time. The framework shares the features of a single network with both the hand pose estimation network and the object pose estimation network. Hand pose estimation is a parametric model that regresses the shape and pose parameters of the hand. The object pose estimation network is a cross-model variational autoencoder network for the direct reconstruction of an object's 6D pose. Our method shows substantial improvement in object pose estimation on two large-scale open-source datasets.

Complete Paper #188

HD-VoxelFlex: Flexible High-Definition Voxel Grid Representation

Igor Vozniak, Pavel Astreika, Philipp Müller, Nils Lipp, Christian Müller and Philipp Slusallek

German Research Center for Artificial Intelligence, Stuhlsatzenhausweg 3 (Campus D3 2), Saarbrücken, Germany

Keywords: Voxel Grid, 3D Convolutions, Voxel Grid Representation, High-Definition Voxel Grid, Reconstruction.

Abstract: Voxel grids are an effective means to represent 3D data, as they accurately preserve spatial relations. However, the inherent sparseness of voxel grid representations leads to significant memory consumption in deep learning architectures, in particular for high-resolution (HD) inputs. As a result, current state-of-the-art approaches to the reconstruction of 3D data tend to avoid voxel grid inputs. In this work, we propose HD-VoxelFlex, a novel 3D CNN architecture that can be flexibly applied to HD voxel grids with only moderate increase in training parameters and memory consumption. HD-VoxelFlex introduces three architectural novelties. First, to improve the models' generalizability, we introduce a random shuffling layer. Second, to reduce information loss, we introduce a novel reducing skip connection layer. Third, to improve modelling of local structure that is crucial for HD inputs, we incorporate a kNN distance mask as input. We combine these novelties with a "bag of tricks" identified in a comprehensive literature review. Based on these novelties we propose six novel building blocks for our encoder-decoder HD-VoxelFlex architecture. In evaluations on the ModelNet10/40 and PCN datasets, HD-VoxelFlex outperforms the state-of-the-art in all point cloud reconstruction metrics. We show that HD-VoxelFlex is able to process high-definition (128 3, 192 3) voxel grid inputs at much lower memory consumption than previous approaches. Furthermore, we show that HD-VoxelFlex, without additional fine-tuning, demonstrates competitive performance in the classification task, proving its generalization ability. As such, our results underline the neglected potential of voxel grid input for deep learning architectures.

baseline experiments to showcase some state-of-the-art deep learning methods for this task. Our findings highlight the need for more advanced models to effectively address this task. This dataset not only catalyzes the development of sophisticated methods for railway applications, but also encourages exploration of novel research directions.

Complete Paper #445

Vehicle Pose Estimation: Exploring Angular Representations

Ivan Orlov^{1,2}, Marco Buzzelli¹ and Raimondo Schettini¹

¹ *Department of Informatics, Systems and Communication, University of Milano-Bicocca, Italy*

² *Aramis Group, France*

Keywords: Vehicle Pose Recognition, Viewpoint Estimation, Car Azimuth Estimation, PASCAL3D+, Angular Regression.

Abstract: This paper addresses the challenge of azimuth estimation in the context of car pose estimation. Our research utilizes the PASCAL3D+ dataset, which offers a diverse range of object categories, including cars, with annotated azimuth estimations for each photograph. We introduce two architectures that approach azimuth estimation as a regression problem, each employing a deep convolutional neural network (DCNN) backbone but diverging in their output definition strategies. The first architecture employs a sin-cos representation of the car's azimuth, while the second utilizes two directional discriminators, distinguishing between front/rear and left/right views of the vehicle. Our comparative analysis reveals that both architectures demonstrate near-identical performance levels on the PASCAL3D+ validation set, achieving a median error of 3.5°, which is a significant advancement in the state of the art. The minimal performance disparity between the two methods highlights their individual strengths while also underscoring the similarity in their practical efficacy. This study not only proposes effective solutions for accurate azimuth estimation but also contributes to the broader understanding of pose estimation challenges in automotive contexts. The code is available at https://github.com/vani-or/car_pose_estimation.

Complete Paper #248

RailCloud-HdF: A Large-Scale Point Cloud Dataset for Railway Scene Semantic Segmentation

Mahdi Abid, Mathis Teixeira, Ankur Mahtani and Thomas Laurent

FCS Railenium, F-59300 Famars, France

Keywords: Point Cloud Dataset, Railway Scenes, Semantic Segmentation, Deep Learning, LiDAR.

Abstract: Semantic scene perception is critical for various applications, including railway systems where safety and efficiency are paramount. Railway applications demand precise knowledge of the environment, making Light Detection and Ranging (LiDAR) a fundamental component of sensor suites. Despite the significance of 3D semantic scene understanding in railway context, there exists no publicly available railborne LiDAR dataset tailored for this purpose. In this work, we present a large-scale point cloud dataset designed to advance research in LiDAR-based semantic scene segmentation for railway applications. Our dataset offers dense point-wise annotations for diverse railway scenes, covering over 267km. To facilitate rigorous evaluation and benchmarking, we propose semantic segmentation of point clouds from a single LiDAR scan as a challenging task. Furthermore, we provide

Session 5D

14:15 - 16:00

Machine Learning Technologies for Vision

VISAPP

Room Space 4A

Complete Paper #128

Multispectral Stereo-Image Fusion for 3D Hyperspectral Scene Reconstruction

Eric Wisotzky^{1,2}, Jost Triller¹, Anna Hilsmann¹ and Peter Eisert^{1,2}

¹ *Computer Vision & Graphics, Fraunhofer HHI, Einsteinufer 37, 10587 Berlin, Germany*

² *Department of Informatics, Humboldt University, Berlin, Germany*

Keywords: Multispectral, Hyperspectral, Stereo-Reconstruction, Optical Flow, Disparity, Spectral Data Fusion, Demosaicing, Multispectral Snapshot Cameras.

Abstract: Spectral imaging enables the analysis of optical material properties that are invisible to the human eye. Different spectral capturing setups, e.g., based on filter-wheel, push-broom, line-scanning, or mosaic cameras, have been introduced in the last years to support a wide range of applications in agriculture, medicine, and industrial surveillance. However, these systems often suffer from different disadvantages, such as lack of real-time

capability, limited spectral coverage or low spatial resolution. To address these drawbacks, we present a novel approach combining two calibrated multispectral real-time capable snapshot cameras, covering different spectral ranges, into a stereo-system. Therefore, a hyperspectral data-cube can be continuously captured. The combined use of different multispectral snapshot cameras enables both 3D reconstruction and spectral analysis. Both captured images are demosaicked avoiding spatial resolution loss. We fuse the spectral data from one camera into the other to receive a spatially and spectrally high resolution video stream. Experiments demonstrate the feasibility of this approach and the system is investigated with regard to its applicability for surgical assistance monitoring.

Complete Paper #191

Deep Learning-Powered Assembly Step Classification for Intricate Machines

Luca Rodiga^{1,2}, Eva Eggeling², Ulrich Krispel² and Torsten Ullrich^{1,2}

¹ Institute of Computer Graphics and Knowledge Visualization, Graz University of Technology, Graz, Austria

² Fraunhofer Austria Research GmbH, Graz, Austria

Keywords: Computer Vision, Deep Learning, Machine Learning.

Abstract: Augmented Reality-based assistance systems can help qualified technicians by providing them with technical details. However, the applicability is limited by the low availability of real data. In this paper, we focus on synthetic renderings of CAD data. Our objective is to investigate different model architectures within the machine-learning component and compare their performance. The training data consists of CAD renderings from different viewpoints distributed over a sphere around the model. Utilizing the advantages of transfer learning and pre-trained backbones we trained different versions of EfficientNet and EfficientNetV2 on these images for every assembly step in two resolutions. The classification performance was evaluated on a smaller test set of synthetic renderings and a dataset of real-world images of the model. The best Top1-accuracy on the real-world dataset is achieved by the medium-sized EfficientNetV2 with 57.74%, while the best Top5-accuracy is provided by EfficientNetV2 Small. Consequently, our approach has a good classification performance indicating the real-world applicability of such a deep learning classifier in the near future.

Complete Paper #308

Recognizing Actions in High-Resolution Low-Framerate Videos: A Feasibility Study in the Construction Sector

Benjamin Vandersmissen, Arian Sabaghi, Phil Reiter and Jose Oramas

University of Antwerp, Imec-IDLab, Antwerp, Belgium

Keywords: Action Recognition, Computer Vision, Deep Learning, Video Processing.

Abstract: Action recognition addresses the automated comprehension of human actions within images or video sequences. Its applications extend across critical areas, mediating between visual perception and intelligent decision-making. However, action recognition encounters multifaceted challenges, including limited annotated data, background clutter, and varying illumination conditions. In the context of the construction sector, distinct challenges arise, requiring specialized approaches. This study investigates

the applicability of established action recognition methodologies in this dynamic setting. We evaluate both sequence-based (YOWO) and frame-based (YOLOv8) approaches, considering the effect that resolution and frame rate have on performance. Additionally, we explore self-supervised learning techniques to enhance recognition accuracy. Our analysis aims to guide the development of more effective and efficient practical action recognition methods.

Complete Paper #322

Large Scale Graph Construction and Label Propagation

Z. Ibrahim¹, A. Bosaghzadeh² and F. Dornaika^{1,3}

¹ University of the Basque Country UPV/EHU, San Sebastian, Spain

² Shahid Rajaee Teacher Training University, Tehran, Iran

³ IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Keywords: Scalable Graph Construction, Semi-Supervised Learning, Topology Imbalance, Large Scale Databases, Reduced Flexible Manifold Embedding.

Abstract: Despite the advances in semi-supervised learning methods, these algorithms face three limitations. The first is the assumption of pre-constructed graphs and the second is their inability to process large databases. The third limitation is that these methods ignore the topological imbalance of the data in a graph. In this paper, we address these limitations and propose a new approach called Weighted Simultaneous Graph Construction and Reduced Flexible Manifold Embedding (W-SGRFME). To overcome the first limitation, we construct the affinity graph using an automatic algorithm within the learning process. The second limitation concerns the ability of the model to handle a large number of unlabeled samples. To this end, the anchors are included in the algorithm as data representatives, and an inductive algorithm is used to estimate the labeling of a large number of unseen samples. To address the topological imbalance of the data samples, we introduced the Renode method to assign weights to the labeled samples. We evaluate the effectiveness of the proposed method through experimental results on two large datasets commonly used in semi-supervised learning: Covtype and MNIST. The results demonstrate the superiority of the W-SGRFME method over two recently proposed models and emphasize its effectiveness in dealing with large datasets.

Complete Paper #323

Image Augmentation Preserving Object Parts Using Superpixels of Variable Granularity

D. Sun¹ and F. Dornaika^{1,2}

¹ University of the Basque Country UPV/EHU, San Sebastian, Spain

² IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Keywords: Data Augmentation, Image Classification, Superpixel, CutMix, Attention, Object-Part.

Abstract: Methods employing regional dropout data augmentation, especially those employing a cut-and-paste approach, have proven highly effective in addressing overfitting challenges arising from limited data. However, existing cutmix-based augmentation strategies face issues related to the loss of contour details and discrepancies between augmented images and their associated labels. In this study, we introduce a novel end-to-end cutmix-based data augmentation method, incorporating the blending of images with discriminative superpixels of diverse granularity. Our experiments for classification tasks reveal outstanding performance across various benchmarks and deep neural network models.

Session 5A
14:15 - 16:00
Agents and Human Interaction

HUCAPP
Room Space 4B

Complete Paper #87

Evaluating Blink Rate as a Dynamic Indicator of Mental Workload in a Flight Simulator

Simon Schwerd and Axel Schulte

Institute of Flight Systems, University of the Bundeswehr Munich, Germany

Keywords: Blink Rate, Mental Workload, Flight Simulator.

Abstract: This study assesses blink rate as a potential indicator for mental workload (MWL) in a dual task scenario in a flight simulator. Prior research indicated that blink rate decreases as mental workload increases across various tasks and domains. In our study, we aimed to determine if these findings are consistent in a dual task environment within a fast jet simulator. Furthermore, we evaluated blink rate fluctuations caused by the dynamic shifts in MWL as tasks are executed, switched, or completed. To investigate this, we executed a flight simulator experiment involving ten participants. They were tasked with two distinct activities: first, classifying air and ground targets, and second, maintaining a specific flight altitude. The results validated that blink rate decreases with increasing task difficulty. However, when a secondary task imposes significant workload, blink rates did not reliably indicate the primary task's difficulty. We also found that the timing of spontaneous blinks was influenced by task completion and switches. Specifically, blink rates surged immediately after decision-making points and during transitions between tasks.

Complete Paper #20

Gaia: A Social Robot to Help Connect Humans and Plants

Christopher Xenophonos, Teresa Clark, Michael Seals,
 Cole Lampman, Iliyas Tursynbek and Mounia Ziat

Bentley University, 175 Forest St, Waltham, MA, 02452, U.S.A.

Keywords: Social Robots, Human-Plant Interaction, Multimodal Design, Arduino.

Abstract: Interacting with plants has been shown to increase both physical and mental health outcomes and seemed obvious for many during such troubling times. Inevitably, plants died as a result of the owner's lack of awareness of their needs. Gaia, a social robot-planter, was created to communicate a plant's needs to its owner in an easy and enjoyable way. The final prototype, a multimodal interface, was designed to join natural language messaging with an emotive digital face and "voice." Creating a social robot that anthropomorphizes the user's plant to effectively communicate the plants' needs, build empathy, and create a stronger emotional bond between the plant and owner, leading to better outcomes for both.

Abstract #21

Exploring the Uncanny Valley Phenomenon for Hand Prosthetics: Stimuli Presentation Factor

Pawel Lupkowski¹, Aleksandra Wasielewska¹ and Marcin Jukiewicz²

¹ *Department of Psychology and Cognitive Science, Adam Mickiewicz University, ul. Wieniawskiego 1 61-712 Poznań, Poland*

² *Adam Mickiewicz University, Poznań, Poland*

Keywords: N/A

Abstract: In this paper we present the results of the study focused on the uncanny valley effect for the hand prosthetics. The motivation for this study comes directly from the author of the uncanny valley hypothesis who illustrated the very idea with the hand prosthetics (Mori et al., 2012, p. 99). Number of studies concerning the UV effect for hand prosthetics is still relatively low (e.g. Poliakoff et al., 2018; Buckingham et al., 2019), thus we aimed to enrich this field.

We were also interested in a more general question related to UV studies. In most cases the stimuli presented to subjects are static pictures (as in the aforementioned ones). However, one may expect that the way you would present the stimuli will affect its evaluation. We have addressed this issue in our study.

We have designed a study with three types of stimuli: human hand (H), prosthetic (P): "Tolka" by the vBionic company and robotic hand (R): "Tolka" without the artificial skin. These stimuli were presented in three experimental conditions: (G1) Photo; (G2) Photo sequence and (G3) Video. For (G2) and (G3) respective stimuli were presented during actions: pointing (a page in a book); grasping (a card lying on the table) and moving (a mug towards a subject). Subjects were asked to evaluate how human-like is the presented hand; how eerie is the presented hand; how much they like the presented hand and to decide whether the hand is natural or artificial. Our research questions were: 1) Will we observe the UV effect for (P) and 2) Will we observe differences in the stimuli evaluations between G1, G2 and G3.

The study was conducted online. 94 participants (49 women), aged 18-72; average 28.42 (SD=9.89) took part. No UV effect for (P) is observed. There are significant differences between G1, G2 and G3 for the (P) stimuli. For G3 prosthetic hand has the lowest human-likeness assessment, lowest like level and the highest eeriness evaluation. Stimuli presentation method does not affect the evaluation of (H) and (R).

The results are important for the UV studies field as they present the evaluation of an actual, commercially available prosthetic hand. What is more, they clearly suggest that the way the stimuli is presented for such studies affect its evaluation - such knowledge should be used for designing future studies.

Several limitations of this study should be also addressed. First of all the study group was relatively small and should be extended for the stronger results. We also believe that future study should also involve real-life interaction with the used stimuli.

References

- [1] Buckingham, G., Parr, J., Wood, G., Day, S., Chadwell, A., Head, J. & Poliakoff, E. (2019). Upper-and lower-limb amputees show reduced levels of eeriness for images of prosthetic hands. *Psychonomic Bulletin & Review*, 26, 1295-1302.
- [2] Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98-100.
- [3] Poliakoff, E., O'Kane, S., Carefoot, O., Kyberd, P., and Gowen, E. (2018). Investigating the uncanny valley for prosthetic hands. *Prosthetics and orthotics international*, 42(1):21-27.

Complete Paper #147

A Digital Inclusion and Technological Barriers: Investigating the Challenges Faced by Formerly Incarcerated Populations in Adopting and Accessing Technology

Johannes Badejo, Joyram Chakraborty and Mia Forbes
*Department of Computer and Information Science, Towson University,
 Towson, MD, U.S.A.*

Keywords: Digital Inclusion, Technology Barriers, Recidivism, Formerly Incarcerated Population, Digital Training.

Abstract: Digital inclusion and technological barriers are two phenomena that directly impact the lives of the formerly incarcerated population (FIP). Adequate access means reducing the high recidivism rates already being handled through the digitization of prison education. The digitization programs offer digital skills and technical skills that can be a handful in helping the ex-offenders secure employment. On the other hand, inadequate access to digital literacy is common among ex-offenders, as most prisons are yet to offer digital education entirely. Championed with the desire to improve the lives of the FIP, this study interviews 71 participants to understand the barriers they face in adapting and accessing technology. The findings indicate that the FIP benefiting from digital inclusion has the upper hand in securing employment and reintegrating well over those hindered by technological barriers. Limited internet connectivity, inadequate financial resources to afford technology devices, limited availability of internet access points (APs), and legal restrictions are majorly reported by the 71 participants. Going by the need to improve the FIP experience, this study found that the population must undergo digital training as part of the re-entry programs.

Complete Paper #450

Performance Evaluation of Visual Analytics Framework for Monitoring Neuromotor Rehabilitation

Maryam Boumrah¹, Samir Garbaya² and Amina Radgui¹

¹ *Centre d'études doctorales Télécoms et Technologies de l'Information (CEDOC-2TI)", INPT, Rabat, Morocco*

² *Laboratoire END-ICAP, INSERM UMR1179, Arts et Métiers Institute of Technology, CNAM, LIFSE, HESAM University, F-75013 Paris, France*

Keywords: Fog Computing, Opentelemetry, Remote Patient Monitoring, Rehabilitation, Stroke, Latency, Scalability.

Abstract: Remote rehabilitation of stroke patients reinforces in-person rehabilitation and enhances the regaining of neuromotor capabilities. However, monitoring stroke patients' rehabilitation from different locations and on a large scale requires a low latency and scalable approach. A real-time visual analytics framework for monitoring in-home rehabilitation of stroke patients based on fog computing is proposed. The objective of this paper is to evaluate the performance of the proposed framework in terms of latency and scalability. OpenTelemetry was used for the evaluation of the proposed framework. OpenTelemetry was chosen over simulation tools for its real-time observability features providing accurate comprehension of the distributed system behaviors in real-world implementation. Five scenarios were setup by progressively escalating the volume of data flow and the number of packets. These scenarios enabled a thorough examination of the framework's ability to handle higher workloads and scalability. The results of end-to-end latency of the proposed system were compared to the Cloud-only implementation. Compared to Cloud-only

implementation, the findings of the evaluation showed that the latency of the proposed system was significantly low. Reflecting the scalability feature, the capacities of handling workload by the proposed system in terms of latency, throughput, processing, and resource utilization were stable across the first four configurations. However, the limitations noticed in the fifth configuration put in evidence the constraints of the experimental setup used in this research. Moreover, the scalability and efficiency of the system can be further enhanced in a distributed deployment in real-world conditions.

Oral Presentations (Online) 5

14:15 - 16:00

Computer Vision Theory and Applications

VISAPP

Room Online 1

Complete Paper #196

Symmetry Completion Test: A Novel Approach for Visual Distortion Mapping and Correction Using Symmetry Constraints

Ye Ling¹, David Frohlich², Tom Williamson³ and Jean-Yves Guillemaut¹

¹ *Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, U.K.*

² *Digital World Research Centre, University of Surrey, Guildford, U.K.*

³ *Guy's and St Thomas' NHS Foundation Trust, London, U.K.*

Keywords: Metamorphopsia, Visual Distortion, Image Warping, Assistive Technology.

Abstract: Metamorphopsia, commonly referred to as distorted vision, is a serious visual impairment which remains uncorrectable by optical glasses or contact lenses. This paper presents a novel approach to digitally map visual distortion based on patient feedback. The approach is based on the use of low-level geometrical constraints (central symmetry) which provide a simple and intuitive mechanism for a patient to provide feedback on their perceived visual distortion. We derive a set of fundamental constraints and show how visual distortion mapping can be framed as an optimisation problem. Critically, a parametric distortion model based on MLS is used to reduce the dimensionality of the problem and enable detailed visual distortion estimation. An extensive evaluation using simulated data demonstrates the accuracy and robustness of the approach. This approach opens up the possibility of correcting for visual distortion by applying the inverse mapping on the input stream to for instance VR see-through devices or screen-based devices.

Complete Paper #278

Deep Discriminative Feature Learning for Document Image Manipulation Detection

Kais Rouis¹, Petra Gomez-Krämer¹, Mickaël Coustaty¹, Saddock Kébairi² and Vincent Poulain d'Andecy²

¹ *L3i Laboratory, La Rochelle University, La Rochelle, France*

² *R&D Department, Yooz, Aimargues, France*

Keywords: Document Forgery, Manipulation Detection, Deep Neural Model, Residual Network.

Abstract: Image authenticity analysis has become a very important task in the last years with one main objective that is tracing the counterfeit content induced by illegal manipulations and forgeries that can be easily practiced using available software tools. In this paper, we propose a reliable residual-based deep neural

network that is able to detect document image manipulations and copy-paste forgeries. We consider the perceptual characteristics of documents including mainly textual regions with homogeneous backgrounds. To capture abstract features, we introduce a shallow architecture using residual blocks and take advantage of shortcut connections. A first layer is implemented to boost the model performance, which is initialized with high-pass filters to forward low-level error feature maps. Manipulation experiments are conducted on a publicly available document dataset. We compare our method with two interesting forensic approaches that incorporate deep neural models along with first layer initialization techniques. We carry out further experiments to handle the forgery detection problem on private administrative document datasets. The experimental results demonstrate the superior performance of our model to detect image manipulations and copy-paste forgeries in a realistic document fraud scenario.

Complete Paper #307

Explainability and Interpretability for Media Forensic Methods: Illustrated on the Example of the Steganalysis Tool Stegdetect

Christian Kraetzer and Mario Hildebrandt

Dept. of Computer Science, Otto-von-Guericke University Magdeburg, Germany

Keywords: Media Forensics, Explainability and Interpretability, Machine Learning, Artificial Intelligence, Quality Assurance and Proficiency Testing in Forensics.

Abstract: For the explainability and interpretability of the outcomes of all forensic investigations, including those in media forensics, the quality assurance and proficiency testing work performed needs to ensure not only the necessary technical competencies of the individual practitioners involved in an examination. It also needs to enable the investigators to have sufficient understanding of machine learning (ML) or 'artificial intelligence' (AI) systems used and are able to ascertain and demonstrate the validity and integrity of evidence in the context of criminal investigations. In this paper, it is illustrated on the example of applying the multi-class steganalysis tool Stegdetect to find steganographic messages hidden in digital images, why the explainability and interpretability of the outcomes of media forensic investigations are a challenge to researchers and forensic practitioners.

Complete Paper #392

A Tool for 3D Representation of the 2D Thermographic Breast Acquisitions

Eudoxia Moura^{1,2}, Gleidson Costa³, Tiago Borchardt³ and Aura Conci²

¹ *Federal Institute of Education, Science and Technology of Rondônia, Ariquemes, Brazil*

² *Institute of Computing, Federal Fluminense University, Brazil*

³ *Department of Informatics, Federal University of Maranhão, Brazil*

Keywords: 3D Breast Models, Thermography Images, Edge Detection.

Abstract: This study enhances the three-dimensional reconstruction of the breast surface from two-dimensional thermal images at five different angles. The process involves the use of a convolutional neural network U-net for image segmentation, extraction of relevant points and curves, geometric transformations to add a dimension to the curves, and modeling with Rational Non-uniform B-splines Surface (NURBS). The work

aims to improve the previously proposed algorithm by making modifications to the rotation points and cutting regions to achieve a more comprehensive breast contour. In a quantitative evaluation of the results, comparing thermal images and Structured Light Scanner images with the Costa's and new algorithm, we observed that, concerning scanning, the Costa's algorithm has an average of 37.07% with a standard deviation of 6.27, while the new algorithm exhibits an average of 14.06% with a standard deviation of 7.28. When compared to thermal images, the Costa's algorithm has an average of 37.47% with a standard deviation of 7.24, whereas the new algorithm shows an average of 14.77% with a standard deviation of 7.37. These results indicate that the new algorithm significantly improves the representation of the breast shape compared to the Costa's algorithm.

Complete Paper #432

Automatic Computation of the Posterior Nipple Line from Mammographies

Quynh Tran¹, Tina Santner² and Antonio Rodríguez-Sánchez¹

¹ *Department of Computer Science, University of Innsbruck, Innsbruck, Austria*

² *Department of Radiology, Medical University of Innsbruck, Innsbruck, Austria*

Keywords: Mammogram, Posterior Nipple Line Detection, Computer Vision.

Abstract: Breast cancer is the most commonly diagnosed cancer in female patients. Detecting early signs of malignancy by undergoing breast screening is therefore of great importance. For a reliable diagnosis, high-quality examined mammograms are essential since poor breast positioning can cause cancers to be missed, which is why mammograms are subject to strict evaluation criteria. One such criterion is the posterior (or pectoralis) nipple line (PNL). We present a method for computing the PNL length, which consisted of the following steps: Pectoral Muscle Detection, Nipple Detection, and final PNL Computation. A multidirectional Gabor filter allowed us to detect the pectoral muscle. For detecting the nipple we made use of the geometric properties of the breast, applied watershed segmentation and Hough Circle Transform. Using both landmarks (pectoral muscle and nipple), the PNL length could be computed. We evaluated 100 mammogram images provided by the Medical University of Innsbruck. The computed PNL length was compared with the real PNL length, which was measured by an expert. Our methodology achieved an absolute mean error of just 6.39 mm.

Oral Presentations (Online) 5

14:15 - 16:00

Computer Graphics Theory and Applications

GRAPP

Room Online 2

Complete Paper #330

Automated Palette Cycling Animations

Ali Javid and David Mould

School of Computer Science, Carleton University, Ottawa, Canada

Keywords: Animation, Image Processing, Palette Cycling, Pixel Art.

Abstract: We propose an automated method for palette cycling, a technique for animation storage and playback. A palette cycling animation uses a fixed map of indices over the entire animation; each frame, the color lookup table accesses by the

indices changes. Historically, palette cycling animations were created manually. Here, we present a method that automatically creates a set of palettes and an index map from an input video. We use optimization, alternating between phases of choosing per-frame palettes and determining an index map given a fixed set of palettes. Our method is highly effective for scenarios such as time-lapse video, where the lighting changes dramatically but there is little overall motion. We show that it can also produce plausible outcomes for videos containing large-scale motion and moving backgrounds; palette rotations with these features are especially difficult to craft by hand. We demonstrate results over a variety of input videos with different levels of complexity, motion, and subject matter.

Complete Paper #347

Adaptation Speed for Exposure Control in Virtual Reality

Claus Madsen and Johan Kristensen

Computer Graphics Group, Aalborg University, Aalborg, Denmark

Keywords: Virtual Reality, Computer Graphics, Exposure Control, Dynamic Range.

Abstract: We address the topic of real-time, view-dependent exposure control in Virtual Reality (VR). For VR to realistically recreate the dynamic range of luminance levels in natural scenes, it is necessary to address exposure control. In this paper we investigate user preference regarding the temporal aspects of exposure adaptation. We design and implement a VR experience that enables users to individually tune how fast they prefer the adaptive exposure control to respond to luminance changes. Our experiments show that 60% of users feel the adaptation significantly improves the experience. Approximately half the of users prefer a fast adaptation over about 1-2 seconds, and the other half prefer a more gradual adaptation over about 10 seconds.

Complete Paper #345

Virtual Reality for Detailed Visualization and Generation of Proximal and Distal Bone Fracture Patterns

J. J. Jiménez-Delgado¹, F. D. Pérez-Cano¹, G. Parra-Cabrera¹ and I. Remolar-Quintana²

¹ *Department of Computer Science, University of Jaén, Jaén, Spain*

² *Institute of New Imaging Technologies, Universitat Jaume I, Castellón, Spain*

Keywords: Bone, Fracture, Pattern, Virtual Reality, Usability, Interaction.

Abstract: This paper articulates the application of virtual reality (VR) in facilitating the study of bone fractures, particularly those affecting the proximal and distal end segments of long bones, areas of the bone in which there are a large number of irregularities. With the increasing prevalence of bone fractures due to aging and increased life expectancy, we present an innovative virtual reality system that meticulously generates and visualizes bone fracture patterns in three dimensions. The system allows users, including healthcare professionals, to interact with 3D bone models and delineate fracture patterns in a realistic environment. It features haptic feedback, simulating the sensation of drawing on a real bone. The tool also provides an exporting functionality for the drawn fractures, enabling the integration into other medical platforms. Experts have evaluated the interaction and representation of the fracture patterns in a highly positive evaluation. The experts assessed the system positively, emphasizing its potential

towards enhancing surgical planning, improving patient outcomes, and fostering educational advancements. Future work aims at improving the degree of realism within the VR environment and refining the precision of closing fracture lines.

Tutorial

14:15 - 17:00

Room Space 4C

Grid-based Layout Algorithms for Visual Sorting, Dimensionality Reduction, and Optimization

Grid-based Layout Algorithms for Visual Sorting, Dimensionality Reduction, and Optimization

Kai U. Barthel

HTW Berlin, Germany

Abstract: Grid-based layout algorithms have a wide range of applications in visual computing and computer graphics. These techniques are invaluable for tasks such as visualizing high-dimensional data, visually sorting images, and solving optimization problems. In this tutorial, we will focus on managing large amounts of textures or images that can overwhelm human perception. We begin with an overview of dimensionality reduction techniques and an examination of their inadequacy for sorting non-point data such as images or textures. The principles and concepts of various image sorting techniques are explained, including Self-Organizing Maps (SOM), Self-Sorting Maps (SSM), and the new innovative Linear Assignment Sorting (LAS). We also present neural networks for learning latent permutations. A major challenge in image sorting is the lack of appropriate metrics for evaluating sorting quality as perceived by humans. Based on extensive user testing, we present the novel Distance Preservation Quality metric, which shows a stronger correlation with user-perceived sorting quality compared to other metrics. Efficiency is another focus of this tutorial. This includes techniques such as integral image filtering and the use of linear assignment solvers for fast sorting. We also offer practical tips and tricks for achieving specific layout shapes and positioning specifications. It's important to note that this tutorial goes beyond static image sorting to include dynamic image graph visualization that adapts to the ever-changing nature of image collections. Furthermore, we show how these techniques can be effectively applied to classical dimension reduction or optimization tasks, such as the Traveling Salesman Problem.

Poster Session 2

16:00 - 17:00

GRAPP

Foyer

Complete Paper #202

Efficiency of 3D Fractal Generation Through Raymarching

Anna Semrau and Dariusz Sawicki

Warsaw University of Technology, Warsaw, Poland

Keywords: Fractals, Raymarching, Signed Distance Functions, Ray Tracing.

Abstract: The use of fractal geometry in computer graphics enables the modeling of natural objects which mathematical description using traditional Euclidean geometry is difficult. However, fractals, due to their properties and specific description, create application problems, especially related to computational and memory efficiency. There are known attempts to solve these problems using graphic hardware and special algorithms. One of the methods that could bring good results is the quite old and

rarely used algorithm of raymarching with SDF (Signed Distance Functions). The aim of the article is to analyze the possibility of using this method to increase the efficiency of fractal modeling. Demonstration application that allows testing various modeling cases has been developed, also taking into account the hardware of modern graphics cards. Research was carried out for 5 different types of fractals (Sierpinski pyramid, Menger sponge, Julia set, Mandelbulb object, fractal tree). The fractal image generation time and memory consumption were considered. For the Menger sponge, different model generation methods were also compared: traditional boundary generation and those using raymarching with SDF. The conducted research has shown that raymarching is a method worth considering. Moreover, the application of raymarching with SDF can bring many measurable benefits.

Complete Paper #357

A Virtual Reality Prototype as a Tool Against Verbal Abuse in Classrooms: A Multidisciplinary Approach

Irene Fondon¹, Mar Elena², Irene Lagares³ and Susana Gaytan⁴

¹ Departamento de Teoría de la Señal y Comunicaciones, Universidad de Sevilla, Camino de los Descubrimientos, Seville, Spain

² Departamento Ingeniería Electrónica, Universidad de Sevilla, Camino de los Descubrimientos, Seville, Spain

³ Departamento Psicología Evolutiva y de la Educación, Universidad de Sevilla, Camilo José Cela, Seville, Spain

⁴ Departamento Fisiología, Universidad de Sevilla, Avenida Reina Mercedes, Seville, Spain

Keywords: Virtual Reality, Bullying, Verbal Abuse, Depression, Anxiety.

Abstract: The exploration of affective response and its vegetative correlate to verbal abuse (VA) is a relevant research area to prevent bullying at schools. Taking advantage of the possibilities that virtual reality offers regarding to immersion in non-real environments inducing feelings in the users, this paper presents a virtual reality application focused on the study of VA in a school context. The versatility of the proposed project is directly related to its applicability. It has been designed under the premises of the psycho-neural effect of VA. The tool is intended to be used under professional and parental supervision, to perform experiments regarding bullying awareness. In this first stage, the authors propose a prototype that will be optimized and upgraded in future versions.

Poster Session 2
16:00 - 17:00

HUCAPP
Foyer

Abstract #102

Are Attitudes Towards Robots Universal? Corpus of Speech About Robots (COSAR) Study of Human Attitudes Towards Social Robots

Aleksandra Wasielewska

Faculty of Psychology and Cognitive Science, Adam Mickiewicz University, ul. Wieniawskiego 1 61-712 Poznan, Poland

Keywords: Attitudes Towards Robots, Social Robots, Human-Robot Interaction, Corpus, Science-Fiction.

Abstract: Studying human attitudes towards social robots constitutes an important part of the human-robot interaction field. It allows for a better understanding of human behaviours, emotions

and concepts related to robots. In this context, social media offers a unique source of data for such studies, as people may express and describe their attitudes in a spontaneous manner.

Corpus of Speech About Robots (COSAR) is a manually annotated corpus that includes people's attitudes toward real social robots, as well as references to science fiction media and fictional robot characters. Data for COSAR was retrieved from YouTube comments on videos presenting 16 different, really existing social robots.

The tagset of the COSAR is based on the source literature, studies on the linguistic data, and the existing questionnaires that measure attitudes toward robots. The structure of the tagset reflects the three-component structure of attitudes (see, e.g. Breckler 1984; Reich-Stiebert et al., 2019): the cognitive component (people's thoughts and beliefs about a robot and cognitive evaluations of a robot), the affective component (feelings or emotions towards/about a robot) and the behavioural component (behavioural intentions or actual behaviour toward a robot). Due to the research on the influence of science fiction media on attitudes towards real robots (see Bruckenberger et al. 2013), an additional category was developed that concerns references to fictional robot characters and science fiction media.

I will present the results of a COSAR study concerning the structure of attitudes towards robots. The cognitive attitudes towards robots prevail in the entire sample. The second most-often component of attitudes is the behavioural one, and there are fewer attitudes that belong to the affective component. I will present the detailed attributes of these attitudes. As a result, we obtain the picture of human attitudes towards really existing robots, which are present in natural language data.

I will also describe and discuss how the observations for the entire sample are reflected on the level of single robots, and I will ask a question of whether observed attitudes may be related to the degree of humanlikeness of robots. I will present possible theoretical implications and applications of the COSAR findings.

References

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, 47(6).

Bruckenberger, U., Weiss, A., Mirnig, N., Strasser, E., Stadler, S., & Tscheligi, M. (2013). The Good, The Bad, The Weird: Audience Evaluation of a "Real" Robot in Relation to Science Fiction and Mass Media (Vol. 8239, p. 310)

Reich-Stiebert, N., Eyssel, F. A., & Hohnemann, C. (2019). Involve the user! Changing attitudes toward robots by user participation in a robot prototyping process. *Computers in Human Behavior*, 91.

Complete Paper #396

From Surveys to TV Series: Study of the Role of Haptic in Social Interaction

Amine Frihi and Bob-Antoine Ménélas

Computer Science Department, UQAC University, 555 Bd Universite, G7H 2B1, Chicoutimi, Canada

Keywords: Social Interaction, Haptic Technology, Communication, Emotions.

Abstract: Here we discuss the role of haptic technology in social interaction. We report the first steps of a project where we want to study the use of a non-visual tactile-auditory device to connect physically distant individuals engaged in non-intimate activities. First, we report the result of a survey to investigate the significance of touch in social interactions. Second, we detail the analysis of touch in some tv shows.

Poster Session 2
16:00 - 17:00

VISAPP
Foyer

Complete Paper #281

Learning 3D Human UV with Loose Clothing from Monocular Video

Meng-Yu Kuo¹, Jingfan Guo¹ and Ryo Kawahara²

¹ University of Minnesota, U.S.A.

² Kyushu Institute of Technology, Japan

Keywords: 3D Human Reconstruction, UV Mapping, Single View.

Abstract: We introduce a novel method for recovering a consistent and dense 3D geometry and appearance of a dressed person from a monocular video. Existing methods mainly focus on tight clothing and recover human geometry as a single representation. Our key idea is to regress the holistic 3D shape and appearance as a canonical displacement and albedo maps in the UV space, while fitting the visual observations across frames. Specifically, we represent the naked body shape by a UV-space SMPL model, and represent the other geometric details, including the clothing, as a shape displacement UV map. We obtain the temporally coherent overall shape by leveraging a differential mask loss and a pose regularization. The surface details in UV space are jointly learned in the course of non-rigid deformation with the differentiable neural rendering. Meanwhile, the skinning deformation in the garment region is updated periodically to adjust its residual non-rigid motion in each frame. We additionally enforce the temporal consistency of surface details by utilizing the optical flow. Experimental results on monocular videos demonstrate the effectiveness of the method. Our UV representation allows for simple and accurate dense 3D correspondence tracking of a person wearing loose clothing. We believe our work would benefit applications including VR/AR content creation.

Complete Paper #305

Deep Learning Model Compression for Resource Efficient Activity Recognition on Edge Devices: A Case Study

Dieter Balemans^{1,2}, Benjamin Vandersmissen³, Jan Steckel², Siegfried Mercelis¹, Phil Reiter³ and José Oramas³

¹ Faculty of Applied Engineering, University of Antwerp - imec - IDLab, Sint-Pietersvliet 7, 2000 Antwerp, Belgium

² Faculty of Applied Engineering, University of Antwerp - CoSysLab, Groenenborgerlaan 171, 2020 Antwerp, Belgium

³ Faculty of Science, University of Antwerp - imec - IDLab, Sint-Pietersvliet 7, 2000 Antwerp, Belgium

Keywords: Activity Recognition, Model Compression, Pruning, Quantization, Edge Computing.

Abstract: This paper presents an approach to adapt an existing activity recognition model for efficient deployment on edge devices. The used model, called YOWO (You Only Watch Once), is a prominent deep learning model. Given its computational complexity, direct deployment on resource-constrained edge devices is challenging. To address this, we propose a two-stage compression methodology consisting of structured channel pruning and quantization. The goal is to significantly reduce the model's size and computational needs while maintaining acceptable task performance. Our experimental results, obtained by deploying the compressed model on Raspberry Pi 4 Model B, confirm that our approach effectively reduces the model's size and

operations while maintaining satisfactory performance. This study paves the way for efficient activity recognition on edge devices.

Complete Paper #382

RecViT: Enhancing Vision Transformer with Top-Down Information Flow

Štefan Pócoš, Iveta Bečková and Igor Farkaš

Faculty of Mathematics, Physics and Informatics, Comenius University Bratislava, Mlynská dolina F1, 842 48 Bratislava, Slovakia

Keywords: Attention, Transformer, Recurrence, Adversarial Examples, Robustness, Heatmap.

Abstract: We propose and analyse a novel neural network architecture — recurrent vision transformer (RecViT). Building upon the popular vision transformer (ViT), we add a biologically inspired top-down connection, letting the network 'reconsider' its initial prediction. Moreover, using a recurrent connection creates space for feeding multiple similar, yet slightly modified or augmented inputs into the network, in a single forward pass. As it has been shown that a top-down connection can increase accuracy in case of convolutional networks, we analyse our architecture, combined with multiple training strategies, in the adversarial examples (AEs) scenario. Our results show that some versions of RecViT indeed exhibit more robust behaviour than the baseline ViT, on the tested datasets yielding $\approx 18\%$ and $\approx 22\%$ absolute improvement in robustness while the accuracy drop was only $\approx 1\%$. We also leverage the fact that transformer networks have certain level of inherent explainability. By visualising attention maps of various input images, we gain some insight into the inner workings of our network. Finally, using annotated segmentation masks, we numerically compare the quality of attention maps on original and adversarial images.

Complete Paper #96

Estimation of the Inference Quality of Machine Learning Models for Cutting Tools Inspection

Kacper Marciniak¹, Paweł Majewski² and Jacek Reiner¹

¹ Faculty of Mechanical Engineering, Wrocław University of Science and Technology, Poland

² Faculty of Information and Communication Technology, Wrocław University of Science and Technology, Poland

Keywords: Machine Vision, Machine Learning, Tool Inspection, Tool Measurement, Inference Quality.

Abstract: The ongoing trend in industry to continuously improve the efficiency of production processes is driving the development of vision-based inspection and measurement systems. With recent significant advances in artificial intelligence, machine learning methods are becoming increasingly applied to these systems. Strict requirements are placed on measurement and control systems regarding accuracy, repeatability, and robustness against variation in working conditions. Machine learning solutions are often unable to meet these requirements - being highly sensitive to the input data variability. Given the depicted difficulties, an original method for estimation of inference quality is proposed. It is based on a feature space analysis and an assessment of the degree of dissimilarity between the input data and the training set described using explicit metrics proposed by the authors. The developed solution has been integrated with an existing system for measuring geometric parameters and determining cutting tool wear, allowing continuous monitoring of the quality of the obtained results and enabling the system operator to take appropriate action in case of a drop below the adopted threshold values.

Complete Paper #118

Diffusion-Based Image Generation for In-Distribution Data Augmentation in Surface Defect Detection

Luigi Capogrosso, Federico Girella, Francesco Taioli, Michele Chiara, Muhammad Aqeel, Franco Fummi, Francesco Setti and Marco Cristani

Department of Engineering for Innovation Medicine, University of Verona, Italy

Keywords: Diffusion Models, Data Augmentation, Surface Defect Detection.

Abstract: In this study, we show that diffusion models can be used in industrial scenarios to improve the data augmentation procedure in the context of surface defect detection. In general, defect detection classifiers are trained on ground-truth data formed by normal samples (negative data) and samples with defects (positive data), where the latter are consistently fewer than normal samples. For these reasons, state-of-the-art data augmentation procedures add synthetic defect data by superimposing artifacts to normal samples. This leads to out-of-distribution augmented data so that the classification system learns what is not a normal sample but does not know what a defect really is. We show that diffusion models overcome this situation, providing more realistic in-distribution defects so that the model can learn the defect's genuine appearance. We propose a novel approach for data augmentation that mixes out-of-distribution with in-distribution samples, which we call *In&Out*. The approach can deal with two data augmentation setups: *i*) when no defects are available (zero-shot data augmentation) and *ii*) when defects are available, which can be in a small number (few-shot) or a large one (full-shot). We focus the experimental part on the most challenging benchmark in the state-of-the-art, i.e., the Kolektor Surface-Defect Dataset 2, defining the new state-of-the-art classification AP score under weak supervision of .782. The code is available at https://github.com/intelligolabs/in_and_out.

Complete Paper #119

Class Weighted Focal Loss for Improving Class Imbalance in Semi-Supervised Object Detection

Shinichi Hoketsu, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi

Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi, Japan

Keywords: Deep Learning, Object Detection, Semi-Supervised Learning, Class Imbalance, In-Vehicle Camera Image.

Abstract: Object detection is a task for acquiring environmental information in automated driving. Object detection is used to detect the position and class of objects in an image. It can be made more accurate by learning with a large amount of supervised data. However, the high cost of annotating the data makes it difficult to create large supervised datasets. Therefore, research using semi-supervised learning for object detection has been attracting attention. Previous studies on semi-supervised learning in object detection tasks have mainly conducted evaluation experiments only on large datasets with many classes, such as MS COCO, and PASCAL VOC. Therefore, the effectiveness of semi-supervised learning for in-vehicle camera data as input has not yet been demonstrated. We examined the effectiveness of semi-supervised learning in object detection when in-vehicle camera data are used as input. We also proposed a class weighted focal loss

that employs a unique weighting method that takes into account the class imbalance problem. Experimental results indicate that semi-supervised learning is also effective when vehicle-mounted camera images are used as input. We also confirmed that the proposed mitigates improves the class imbalance problem and improves accuracy.

Complete Paper #126

Character Identification in Images Extracted from Portuguese Manuscript Historical Documents

Gustavo Lacerda and Raimundo Vasconcelos

Instituto Federal de Brasília, Taguatinga - DF, Brazil

Keywords: OCR, Text Recognition, Historical Manuscripts, Neural Networks.

Abstract: The creation of writing has facilitated the humanity's accumulation and sharing of knowledge, being a vital part of what differentiates humans from other animals and has a high importance for the culture of all peoples. Thus, the first human records (manuscripts), historical documents of organizations and families, began to have new perspectives with the digital age accumulation. These handwritten records remained the primary source for the history of countries, including Brazil before the period of independence, until the Gutenberg movable type printing press dominated the archival world. Thus, over the decades, these handwritten documents, due to their fragility, became difficult to access and manipulate. This has changed, with the possibility of digitization and, consequently, its distribution over the internet. Therefore, this work shows a solution for transcribing historical texts written in Portuguese, bringing accessibility, searchability, sharing and preservation to these records, which achieved a result of 97% of letters recognized in the database used.

Complete Paper #127

Identifying Representative Images for Events Description Using Machine Learning

Marcos Soares de Sousa and Raimundo Vasconcelos

Instituto Federal de Brasília, Taguatinga - DF, Brazil

Keywords: Representative Images, Machine Learning, Events Description.

Abstract: The use of social networks to record events – disasters, demonstrations, parties – has grown a lot and has begun to receive attention in recent years. Existing research focuses primarily on analyzing text-based messages from social media platforms such as Twitter. Images, photos and other media are increasingly used and can provide valuable information to enhance the understanding of an event and can be used as indicators of relevance. This work explores the Twitter social media platform, based on image and text in the case of the demonstrations that took place in Brazil on September 7, 2021, as a result of the Independence celebrations. This work uses machine learning techniques (VGG-16, VGG-19, ResNet50v2 and InceptionResNetv2) for finding relevant Twitter images. The results show that the existence of an image within a social media message can serve as a high probability indicator of relevant content. An extensive experimental evaluation was carried out and demonstrated that high efficiency gains can be obtained compared to state-of-the-art methods.

Complete Paper #139

Association of Grad-CAM, LIME and Multidimensional Fractal Techniques for the Classification of H&E Images

Thales Lopes¹, Guilherme Roberto², Carlos Soares², Thaína Tosta³, Adriano Silva⁴, Adriano Loyola⁵, Sérgio Cardoso⁵, Paulo R. de Faria⁶, Marcelo Z. do Nascimento⁴ and Leandro Neves¹

¹ Department of Computer Science and Statistics, São Paulo State University, São José do Rio Preto-SP, Brazil

² Faculty of Engineering, University of Porto, Porto, Portugal

³ Institute of Science and Technology, Federal University of São Paulo, São José dos Campos-SP, Brazil

⁴ Faculty of Computer Science, Federal University of Uberlândia, Uberlândia-MG, Brazil

⁵ Area of Oral Pathology, School of Dentistry, Federal University of Uberlândia, Uberlândia-MG, Brazil

⁶ Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia, Uberlândia-MG, Brazil

Keywords: Deep Learning, Fractal Features, Explainable Artificial Intelligence, Histological Images.

Abstract: In this work, a method based on the use of explainable artificial intelligence techniques with multiscale and multidimensional fractal techniques is presented in order to investigate histological images stained with Hematoxylin-Eosin. The CNN GoogLeNet neural activation patterns were explored, obtained from the gradient-weighted class activation mapping and locally-interpretable model-agnostic explanation techniques. The feature vectors were generated with multiscale and multidimensional fractal techniques, specifically fractal dimension, lacunarity and percolation. The features were evaluated by ranking each entry, using the ReliefF algorithm. The discriminative power of each solution was defined via classifiers with different heuristics. The best results were obtained from LIME, with a significant increase in accuracy and AUC rates when compared to those provided by GoogLeNet. The details presented here can contribute to the development of models aimed at the classification of histological images.

Complete Paper #146

A Comparative Analysis of the Three-Alternative Forced Choice Method and the Slider-Based Method in Subjective Experiments: A Case Study on Contrast Preference Task

Olga Cherepkova, Seyed Ali Amirshahi and Marius Pedersen

Norwegian University of Science and Industry, Norway

Keywords: 3AFC, Three-Alternative Forced-Choice, Slider, Contrast, Preferences, Response Format.

Abstract: When it comes to collecting subjective data in the field of image quality assessment, different approaches have been proposed. Most datasets in the field ask observers to evaluate the quality of different test and reference images. However, a number of datasets ask observers to make changes to one or more properties of the image to enhance the image to its best possible quality. Among the methods used in the second approach is the Three-Alternative Forced Choice (3AFC) and the slider-based methods. In this paper, we study and compare the two mentioned methods in the case of collecting contrast preferences for natural images. Fifteen observers participated in two experiments under

controlled settings, incorporating 499 unique and 100 repeated images. The reliability of the answers and the differences between the two methods were analyzed. The results revealed a general lack of correlation in contrast preferences between the two methods. The slider-based method generally yielded lower values in contrast preferences compared to 3AFC experiment. In the case of repeated images, the slider-based method showed greater consistency in subjective scores given by each observer. These results suggest that neither method can serve as a direct substitute for the other, as they exhibited low correlation and statistically significant differences in results. The slider-based experiment offered the advantage of significantly shorter completion times, contributing to higher observer satisfaction. In contrast, the 3AFC task provided a more robust interface for collecting preferences. By comparing the results obtained by the two methods, this study provides information on their respective strengths, limitations, and suitability for use in similar preference acquisition tasks.

Complete Paper #149

Comparative Study Between Object Detection Models, for Olive Fruit Fly Identification

Margarida Victoriano¹, Lino Oliveira¹ and Hélder Oliveira^{1,2}

¹ INESC TEC, FEUP Campus, Porto, Portugal

² Faculty of Sciences, University of Porto, Porto, Portugal

Keywords: Image Processing, Object Detection and Classification, Insect Management.

Abstract: Climate change is causing the emergence of new pest species and diseases, threatening economies, public health, and food security. In Europe, olive groves are crucial for producing olive oil and table olives; however, the presence of the olive fruit fly (*Bactrocera Oleae*) poses a significant threat, causing crop losses and financial hardship. Early disease and pest detection methods are crucial for addressing this issue. This work presents a pioneering comparative performance study between two state-of-the-art object detection models, YOLOv5 and YOLOv8, for the detection of the olive fruit fly from trap images, marking the first-ever application of these models in this context. The dataset was obtained by merging two existing datasets: the DIRT dataset, collected in Greece, and the CIMO-IPB dataset, collected in Portugal. To increase its diversity and size, the dataset was augmented, and then both models were fine-tuned. A set of metrics were calculated, to assess both models performance. Early detection techniques like these can be incorporated in electronic traps, to effectively safeguard crops from the adverse impacts caused by climate change, ultimately ensuring food security and sustainable agriculture.

Complete Paper #159

Applying Prompts and Parameter-Efficient Methods to Enhance Single-Stream Vision-Language Transformers

Xuehao Liu, Sarah Jane Delany and Susan McKeever

School of Computer Science, Technological University Dublin, Ireland

Keywords: Image Captioning, Prompts, Parameter-Efficient Tuning, Vision-Language Transformer.

Abstract: Large-Scale transformer models pose challenges due to resource-intensive training, time, and data requirements for fine-tuning on new tasks, mainly due to their extensive parameter count. To address this, zero-shot and few-shot learning, aided by

techniques like prompts and parameter-efficient modules, have emerged. However, these techniques are often tailored for vision-only or language-only tasks, leaving a gap for their effectiveness in multi-modal tasks like image captioning. This paper explores the effectiveness of prompts and parameter-efficient modules in reducing the training effort for image captioning. Rather than extensive fine-tuning, we trained only the prompt and parameter-efficient modules on the pretrained Oscar transformer model using the COCO dataset. We tested five prompt tuning approaches and two parameter-efficient methods. Notably, combining visual prompt tuning (VPT) with Adapter and LoRA led to a 2% Cider score improvement after just one epoch training, with a minimal increase in trainable parameters (5.7%). Our work paves the way towards using single-stream transformer models for a variety of fine-tuned tasks, but with a huge potential reduction in retraining time and processing resources.

Complete Paper #251

Coreset Based Medical Image Anomaly Detection and Segmentation

Ciprian-Mihai Ceașescu¹, Bogdan Alexe^{1,2} and Riccardo Volpi^{3,4}

¹ University of Bucharest, Romania

² Gheorghe Mihoc-Caius Iacob Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, Romania

³ Quaeșta AI, Cluj-Napoca, Romania

⁴ Transylvanian Institute of Neuroscience (TINS), Cluj-Napoca, Romania

Keywords: Binary Classification of Medical Images, Anomaly Detection, Binary Segmentation of Medical Images, PatchCore, Coreset.

Abstract: We address the problem of binary classification of medical images employing an anomaly detection approach that uses only normal images for training. We build our method on top of a state-of-the-art anomaly detection method for visual inspection of industrial natural images, PatchCore, tailored to our tasks. We deal with the distribution shift between natural and medical images either by fine-tuning a pre-trained encoder on a general medical image dataset with ten classes or by training the encoder directly on a set of discriminative medical tasks. We employ our method for binary classification and evaluate it on two datasets: lung cancer from CT scan images and brain tumor from MRI images showing competitive results when compared to the baselines. Conveniently, this approach is able to produce segmentation masks used for localizing the anomalous regions. Additionally, we show how transformer encoders are up to the task allowing for improved F1 and AUC metrics on the anomaly task, also producing a better segmentation.

Complete Paper #332

Error Analysis of Aerial Image-Based Relative Object Position Estimation

Zsombor Páncsics^{1,2}, Nelli Nyisztor^{1,2}, Tekla Tóth¹, Imre Juhász², Gergely Treplán² and Levente Hajder¹

¹ Faculty of Informatics, Eötvös Loránd University Budapest, Hungary

² Robert Bosch Kft. Budapest, Hungary

Keywords: Aerial Imaging, Bird's-Eye View, Error Analysis, Autonomous Navigation, Pose Estimation.

Abstract: This paper presents a thorough analysis of precision and sensitivity in aerial image-based relative object position estimation, exploring factors such as camera tilt, 3D projection

error, marker misalignment, rotation and calibration error. Our unique contribution lies in simulating complex 3D geometries at varying camera altitudes (20-130 m). The simulator has a built-in unique mathematical model offering an extensive set of error parameters to improve reliability of aerial image-based position estimation in practical applications.

Complete Paper #355

An Assistive Technology Based on Object Detection for Automated Task List Generation

Frédéric Rayar

LIFAT, University of Tours, Tours, France

Keywords: Object Detection, YOLOv7 Model, Task Generation, Assistive Technology, Intellectual Disability.

Abstract: People suffering from Intellectual Disability (ID) face challenges to perform sequential tasks in their daily life, impacting their education and employability. To help them, the usage of Assistive Technology (AT) on mobile devices is a promising direction. However, most of them do not take advantage of the important recent advances in Artificial Intelligence. In this paper, we address this lack by presenting a prototype of an embed AT system that leverage computer vision advances to assist a user suffering from ID to perform sequential tasks in a guesthouse rooms' tidying-up activity. It first relies on a state-of-the-art object detector, namely YOLOv7, that we have adapted for a real-time usage on mobile devices. Then, using a "spot the difference" approach, it identifies objects of interest that are either present, absent or displaced, compared to a template image. A list of tasks to be achieved is automatically generated and conveyed to the user using an accessible and ergonomic interface. Early qualitative experiments of this ongoing work lead us to believe that our contribution could improve the life of people suffering from ID, allowing them to improve both their functioning and independence.

Complete Paper #371

Microplankton Discrimination in FlowCAM Images Using Deep Learning

Francisco Bonin-Font¹, Gorka Buenvaron², Mary Kane³ and Idan Tuval³

¹ Systems, Robotics and Vision Group (SRVG), University of the Balearic Islands, ctra de Valldemossa km 7.5, Palma de Mallorca, Balearic Islands, Spain

² Institute for Cross-Disciplinary Physics and Complex Systems (IFISC), University of the Balearic Islands, ctra de Valldemossa km 7.5, Palma de Mallorca, Balearic Islands, Spain

³ Department of Marine Ecology, Mediterranean Institute of Advanced Studies (IMEDEA), Miquel Marqués 21, Esporles, Balearic Islands, Spain

Keywords: Phytoplankton, Zooplankton, Convolutional Neural Networks.

Abstract: Marine plankton are omnipresent throughout the oceans, and the Mediterranean Sea is no exception. Innovation on microscopy technology for observing marine plankton over the last several decades has enabled scientist to obtain large quantities of images. While these new instruments permit generating and recording large amounts of visual information about plankton, they have produced a bottleneck and overwhelmed our abilities to provide meaningful taxonomic information quickly. The development of methods based on Artificial Intelligence or Deep Learning to process these images in efficient, cost-effective manners is an active area of continued research. In this study, Convolutional Neural Networks (CNNs) were trained to analyze

images of natural assemblages of microplankton ($< 100\mu m$) and laboratory monocultures. The CNN configurations and training were focused on differentiating phytoplankton, zooplankton, and zooplankton consuming phytoplankton. Experiments reveal high performance in the discrimination of these different varieties of plankton, in terms of Accuracy, Precision, F1 scores and mean Average Precision.

Complete Paper #376

Depth Estimation Using Weighted-Loss and Transfer Learning

Muhammad Hafeez¹, Michael Madden^{1,2}, Ganesh Sistu³ and Ihsan Ullah^{1,2}

¹ Machine Learning Research Group, School of Computer Science, University of Galway, Ireland

² Insight SFI Research Centre for Data Analytics, University of Galway, Ireland

³ Valeo Vision Systems, Tuam, Ireland

Keywords: Depth Estimation, Transfer Learning, Weighted-Loss Function.

Abstract: Depth estimation from 2D images is a common computer vision task that has applications in many fields including autonomous vehicles, scene understanding and robotics. The accuracy of a supervised depth estimation method mainly relies on the chosen loss function, the model architecture, quality of data and performance metrics. In this study, we propose a simplified and adaptable approach to improve depth estimation accuracy using transfer learning and an optimized loss function. The optimized loss function is a combination of weighted losses to which enhance robustness and generalization: Mean Absolute Error (MAE), Edge Loss and Structural Similarity Index (SSIM). We use a grid search and a random search method to find optimized weights for the losses, which leads to an improved model. We explore multiple encoder-decoder-based models including DenseNet121, DenseNet169, DenseNet201, and EfficientNet for the supervised depth estimation model on NYU Depth Dataset v2. We observe that the EfficientNet model, pre-trained on ImageNet for classification when used as an encoder, with a simple upsampling decoder, gives the best results in terms of RSME, REL and \log_{10} : 0.386, 0.113 and 0.049, respectively. We also perform a qualitative analysis which illustrates that our model produces depth maps that closely resemble ground truth, even in cases where the ground truth is flawed. The results indicate significant improvements in accuracy and robustness, with EfficientNet being the most successful architecture.

Complete Paper #379

Optical Illusion in Which Line Segments Continue to Grow or Shrink by Displaying Two Images Alternately

Kazuhisa Yanaka and Sota Mihara

Kanagawa Institute of Technology, 1030 Shimo-ogino, Atsugi-shi, Japan

Keywords: Optical Illusion, Motion Illusion, Animation, Visual Psychology.

Abstract: A new illusion has been discovered, wherein line segments, when alternately displayed with their tonal inversion or monochromatic images for approximately 120 ms each on a monochromatic background, seem to grow or shrink continuously. For instance, if the first image features black line segments on a white background and the second image shows the inverse brightness, switching between these two images causes the line

segments to give the illusion of continuous expansion. Although a single line segment suffices, aligning multiple line segments parallel to each other enhances the effect of this illusion. This illusion can be achieved using achromatic colors, such as black and white, as well as chromatic colors, such as red, blue, and green. Specifically, when using an image with a black line segment on a red background alongside its brightness-inverted counterpart, the line segments appear to steadily decrease in length. Our hypothesis suggests a comparison between the mechanisms of this illusion and the changes in water volume in a pond.

Complete Paper #425

Evaluating the Usability of a 3D Map Visualizer Augmented Reality Application

Franklin Fracaro¹, Fabiana Peres¹, Claudio Mauricio¹ and João Teixeira²

¹ Centro de Engenharias e Ciências Exatas, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu 85870-650, Brazil

² Departamento de Eletrônica e Sistemas, Universidade Federal de Pernambuco, Recife 50670-901, Brazil

Keywords: Usability Testing, Augmented Reality, User Experience, Human-Computer Interaction.

Abstract: This study focuses on evaluating the usability of a 3D map visualizer augmented reality application through informal, remote, and synchronous usability tests, utilizing the Think Aloud process and the System Usability Scale questionnaire. The aim is to identify potential usability issues in the prototype and develop an action plan for future iterations of the application's development. The challenges of integrating end-users in the application development process is addressed and emphasizes the significance of early usability testing to uncover and address potential issues. The findings from the usability tests provide valuable insights into user experience, highlighting key usability issues and user feedback. A comprehensive action plan for addressing the identified usability issues is presented and implications for future research in the domain of augmented reality applications is discussed.

Poster Presentations (Online) 2
16:00 - 17:00

VISIGRAPP
Room Online 1

Poster Presentations (Online) 2
16:00 - 17:00

IVAPP
Room Online 1

Complete Paper #136

Visualizing Plasma Physics Simulations in Immersive Environments

Nuno Verdelho Trindade¹, Óscar Amaro², David Brás³, Daniel Gonçalves¹, João Madeiras Pereira¹ and Alfredo Ferreira¹

¹ INESC-ID, Instituto Superior Técnico, University of Lisbon, Rua Alves Redol, 9, 1000-029 Lisbon, Portugal

² Group of Lasers and Plasma (GoLP), Instituto Superior Técnico, University of Lisbon, Complexo Interdisciplinar, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

³ Department of Computer Science and Engineering (DEI), Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal

Keywords: Virtual Reality, Immersive Visualization, Scientific Visualization, Plasma Physics, Data Analysis.

Abstract: Plasma physics simulations create complex datasets

for which researchers need state-of-the-art visualization tools to gain insights. These datasets are 3D in nature but are commonly depicted and analyzed using 2D idioms displayed on 2D screens. These offer limited understandability in a domain where spatial awareness is key. Virtual reality (VR) can be used as an alternative to conventional means for analyzing such datasets. This study presents PlasmaVR, a proof-of-concept VR tool for visualizing datasets resulting from plasma physics simulations. It enables immersive multidimensional data visualization of particles, scalar, and vector fields. The study includes user evaluation with domain experts where PlasmaVR was employed to assess the possible benefits of immersive environments in plasma physics visualization. Participants manifested a high level of engagement when using the prototype, considering it more enjoyable than conventional means. The participant's perception of the usefulness of VR in plasma simulations also increased after experiencing the prototype.

Poster Presentations (Online) 2
16:00 - 17:00

HUCAPP
Room Online 1

Complete Paper #187

Enhancing the User Interaction of Online Students: Analysis of an Interaction Concept for a Learner Dashboard

Thorleif Harder and Gilbert Drzyzga

Institute for Interactive Systems, Technische Hochschule Luebeck, Germany

Keywords: Interaction Concept, Learner Dashboard, Navigation Mechanisms, Iterative Optimization, User Experience.

Abstract: This paper describes the systematic development, evaluation, and iterative refinement of an interaction concept for a Learner Dashboard (LD) to enhance intuitive and user-centered engagement for online students. The core of this concept is represented by precisely designed interaction elements such as icons, navigation mechanisms and adaptive features, each iteratively refined through a feedback loop. The concept emphasizes the crucial role of intuitive icons and clear navigation, aligning with established design principles and user expectations. The implementation of text buttons with arrow icons is particularly notable, enhancing affordances and signifiers and significantly improving the clarity of navigation paths and user guidance. Proactive support elements like alerts and confirmation prompts significantly reduced user errors. The findings highlight the essential role of iterative design and user feedback in tuning the interaction design to resonate with the cognitive and emotional needs of online students, enhancing the user experience. The paper argues the evolved interaction design represents a significant step towards a user-centered and intuitive learning environment. Future work should focus on real-world testing to evaluate the technical feasibility and user acceptance of the developed concept in different learning scenarios.

Poster Presentations (Online) 2
16:00 - 17:00

VISAPP
Room Online 1

Complete Paper #198

Robotics and Computer Vision in the Brazilian Electoral Context: A Case Study

Jairton Falcão Filho, Matheus Costa, Jonas Silva, Cecília Santos da Silva, Felipe Mendonca, Jefferson Norberto, Riei Rodrigues and Marcondes Silva Júnior

Informatics Center, Universidade Federal de Pernambuco, Recife, Brazil

Keywords: Electronic Ballot Box, Integrity Test, Robotic Arm, Computer Vision.

Abstract: Since 2000, Brazil has had fully computerized elections using electronic ballot boxes. To prove the functioning and security of electronic ballot boxes transparently, some tests are carried out, including the integrity test. This test is carried out throughout the national territory in locations defined for each Brazilian state, with a reasonable number of people. Here, an automation system for integrity testing is presented, consisting of a robotic arm and software using computer vision for the two versions of electronic ballot boxes used in the 2022 presidential elections. Two days of tests were carried out, simulating the integrity test in the laboratory and a test during the real integrity test of the 2022 election. The system managed to cast between 197 and 232 votes during the tests with an average vote time of between 2 minutes and 15 seconds to 2 minutes and 43 seconds, depending on the version and test day, and errors between 1.48% and 3.72%, including reading and typing errors. The system allows you to reduce the number of people per electronic ballot box and increase the transparency and efficiency of the integrity test.

Complete Paper #211

Nearest Neighbor-Based Data Denoising for Deep Metric Learning

George Galanakis^{1,2}, Xenophon Zabulis² and Antonis Argyros^{1,2}

¹ *Computer Science Department, University of Crete, Greece*

² *Institute of Computer Science, FORTH, Greece*

Keywords: Label Noise, Data Denoising, Deep Metric Learning, KNN, Image Retrieval.

Abstract: The effectiveness of supervised deep metric learning relies on the availability of a correctly annotated dataset, i.e., a dataset where images are associated with correct class labels. The presence of incorrect labels in a dataset disorients the learning process. In this paper, we propose an approach to combat the presence of such label noise in datasets. Our approach operates online, during training and on the batch level. It examines the neighborhood of samples, considers which of them are noisy and eliminates them from the current training step. The neighborhood is established using features obtained from the entire dataset during previous training epochs and therefore is updated as the model learns better data representations. We evaluate our approach using multiple datasets and loss functions, and demonstrate that it performs better or comparably to the competition. At the same time, in contrast to the competition, it does not require knowledge of the noise contamination rate of the employed datasets.

Complete Paper #214

XYZ Unsupervised Network: A Robust Image Dehazing Approach

Percy Maldonado-Quispe and Helio Pedrini

Institute of Computing, University of Campinas, Campinas, SP, 13083-852, Brazil

Keywords: Image Dehazing, Unsupervised Network, Zero-Shot Learning, Disentanglement, Dark Channel Prior, Atmospheric Scattering Model.

Abstract: In this work, we examine a major less-explored topic in image dehazing neural networks, specifically how to remove haze (natural phenomenon) in an unsupervised manner from a given image. By considering a hazy image as the entanglement of many “simpler” layers, such as a hazy-free image layer, transmission map layer, and atmospheric light layer, as shown in the atmospheric scattering model, we propose a method based on the concept of layer disentanglement. Our XYZ approach presents improvements in the SSIM and PSNR metrics, this being the combination of the XHOT, YOLY and ZID methods, in which the advantages of each of them are maintained. The main benefits of the proposed XYZ are twofold. First, since it is an unsupervised approach, no clean photos, including hazy-clear pairs, are used as the ground truth. In other words, it differs from the traditional paradigm of deep model training on a large dataset. The second is to consider haze issues as being composed of several layers.

Complete Paper #386

Strategies for Classifier Selection Based on Genetic Programming for Multimedia Data Recognition

Rafael Martarelli¹, Douglas Rodrigues¹, Clayton Pereira¹, Jurandy Almeida² and João Papa¹

¹ *Department of Computing, São Paulo State University, Bauru, SP, Brazil*

² *Department of Computing, Federal University of São Carlos, Sorocaba, SP, Brazil*

Keywords: Ensemble Pruning, Genetic Programming, Fast-CoViAR, Action Classification.

Abstract: We live in a digital world with an explosion of data in various forms, such as videos, images, signals, and texts, making manual analysis unfeasible. Machine learning techniques can use this huge amount of data to train models as an excellent solution for automating decision-making processes such as fraud detection, product recommendation, and assistance with medical diagnosis, among others. However, training these classifiers is challenging, resulting in discarding low-quality models. Classifier committees and ensemble pruning have been introduced to optimize classification, but traditional functions used to fuse predictions are limited. This paper proposes the use of Genetic Programming (GP) to combine committee members' forecasts in a new fashion, opening new perspectives in data classification. We evaluate the proposed method employing several mathematical functions and fuzzy logic operations in HMDB51 and UCF101 datasets. The results reveal that GP can significantly enhance the performance of classifier committees, outperforming traditional methods in various scenarios. The proposed approach improves accuracy on training and test sets, offering adaptability to different data features and user requirements.

Keynote Lecture
17:00 - 18:00

VISIGRAPP
Room Plenary

The Predictable Side of Unpredictable Humans

Alvitta Ottley

Washington University, U.S.A.

Abstract: Building AI systems that interact naturally with humans presents a tremendous challenge. This is because humans are not simply logic machines; our emotions, experiences, and social contexts all influence our behavior in unpredictable ways. So how can we, as AI developers, navigate this complexity and design systems that respond intelligently to their human counterparts? In this talk, I discuss how we approach this problem in the context of Visual Analytics by embracing the parallels between human reasoning and AI models. For example, this talk will demonstrate how humans, like AI systems, can be modeled as a set of rules and parameters, which facilitates predicting behavioral outcomes *in specific scenarios*. By employing inferential techniques like Bayesian reasoning, we will go beyond actions and discuss how we might model and infer deeper motivations and beliefs driving them. Truly understanding humans is non-trivial, but we can forge a path toward effective human-AI interaction by leveraging the similarities between human thought processes and AI modeling.

Wednesday, 28

Thursday Sessions: February 29

Thursday Sessions: February 29 Program Layout

| | Coffee-Break | Foyer | Online 1 | Online 2 | Plenary | Restaurant | Space 3A | Space 3B | Space 3C | Space 4A | Space 4B | Space 4D |
|-------|--------------|----------------------------|---------------------------------|---------------------------------|-----------------------------------|------------|------------------------------|-------------------------------|------------------------------|------------------------------|-----------------------------|-----------------------------|
| 9:00 | | | | | | | VISIGRAPP Session 6 | VISIGRAPP Session 6 | VISIGRAPP Session 6 | VISIGRAPP Session 6 | VISIGRAPP Session 6 | VISIGRAPP Session 6 |
| 9:30 | | | Oral Presentations (Online) 6 | Oral Presentations (Online) 6 | | | VISAPP #61, #156, #277, #344 | VISAPP #243, #254, #289, #353 | VISAPP #85, #99, #185, #228 | VISAPP #35, #79, #218, #240 | GRAPP #29, #377, #401 | HUCAPP #343, #434, #435 |
| 10:00 | | | | | | | | | | | | |
| 10:30 | Coffee-Break | VISIGRAPP Poster Session 3 | Poster Presentations (Online) 3 | Poster Presentations (Online) 3 | | | | | | | | |
| 11:00 | | | | | | | | | | | | |
| 11:30 | | | | | | | VISIGRAPP Session 7 | VISIGRAPP Session 7 | VISIGRAPP Session 7 | VISIGRAPP Session 7 | VISIGRAPP Session 7 | VISIGRAPP Session 7 |
| 12:00 | | | Oral Presentations (Online) 7 | Oral Presentations (Online) 7 | | | VISAPP #61, #180, #328, #335 | VISAPP #38, #117, #300, #416 | VISAPP #40, #175, #197, #354 | VISAPP #65, #402, #403, #411 | GRAPP #19, #122, #404, #447 | IVAPP #97, #263, #324, #346 |
| 12:30 | | | | | | | | | | | | |
| 13:00 | | | | | | | | | | | | |
| 13:30 | | | | | | Lunch | | | | | | |
| 14:00 | | | | | | | | | | | | |
| 14:30 | | | | | Keynote Lecture Mel Slater | | | | | | | |
| 15:00 | | | | | | | | | | | | |
| 15:30 | | | | | | | VISIGRAPP Session 8 | VISIGRAPP Session 8 | VISIGRAPP Session 8 | VISIGRAPP Session 8 | VISIGRAPP Session 8 | VISIGRAPP Session 8 |
| 16:00 | | | Oral Presentations (Online) 8 | | Tutorial João Marcelo Teixeira | | VISAPP #340, #360, #378 | VISAPP #137, #148, #383, #427 | VISAPP #22, #110, #138, #142 | VISAPP #89, #100, #448 | GRAPP #219, #231, #249 | HUCAPP #18, #438 |
| 16:30 | | | | | | | | | | | | |
| 17:00 | Coffee-Break | | | | | | | | | | | |
| 17:30 | | | | | Industrial Panel | | | | | | | |
| 18:00 | | | | | | | | | | | | |
| 18:30 | | | | | Closing Session & Awards Ceremony | | | | | | | |
| 19:00 | | | | | | | | | | | | |

Session 6A
09:00 - 10:30
Object Detection and Localization

VISAPP
Room Space 3A

Complete Paper #156

Non-Local Context-Aware Attention for Object Detection in Remote Sensing Images

Yassin Terraf¹, El Mercha¹ and Mohammed Erradi²

¹ HENCEFORTH, Rabat, Morocco

² ENSIAS, Mohammed V University, Rabat, Morocco

Keywords: Object Detection, Deep Learning, Attention, Remote Sensing Images.

Abstract: Object detection in remote sensing images has been widely studied due to the valuable insights it provides for different fields. Detecting objects in remote sensing images is a very challenging task due to the diverse range of sizes, orientations, and appearances of objects within the images. Many approaches have been developed to address these challenges, primarily focusing on capturing semantic information while missing out on contextual details that can bring more insights to the analysis. In this work, we propose a Non-Local Context-Aware Attention (NLCAA) approach for object detection in remote sensing images. NLCAA includes semantic and contextual attention modules to capture both semantic and contextual information. Extensive experiments were conducted on two publicly available datasets, namely NWPU VHR and DIOR, to evaluate the performance of the proposed approach. The experimental results demonstrate the effectiveness of the NLCAA approach against various state-of-the-art methods.

Complete Paper #61

Lens Flare-Aware Detector in Autonomous Driving

Shanxing Ma and Jan Aelterman

TELIN-IPI, Ghent University – imec, Sint-Pietersnieuwstraat, Ghent, Belgium

Keywords: Autonomous Driving, Object Detection, Bayes' Theorem, Lens Flare.

Abstract: Autonomous driving has the potential of reducing traffic accidents, and object detection plays a key role. This paper focuses on the study of object detection in the presence of lens flare. We analyze the impact of lens flare on object detection in autonomous driving tasks and propose a lens flare adaptation method based on Bayesian reasoning theory to optimize existing object detection models. This allows us to adjust the detection scores to re-rank the detections of detection models based on the intensity of lens flare and achieve a higher average precision. Furthermore, this method only requires simple modifications based on the detection results of the existing object detection models, making it easier to deploy on existing devices.

Complete Paper #277

World-Map Misalignment Detection for Visual Navigation Systems

Rosario Forte¹, Michele Mazzamuto^{1,2}, Francesco Ragusa^{1,2}, Giovanni Farinella^{1,2} and Antonino Furnari^{1,2}

¹ FPV@IPLAB, DMI - University of Catania, Italy

² Next Vision s.r.l. - Spinoff of the University of Catania, Italy

Keywords: Egocentric Vision, Computer Vision, Synthetic Data, Acquisition Tool.

Abstract: We consider the problem of inferring when the internal map of an indoor navigation system is misaligned with respect to the real world (world-map misalignment), which can lead to misleading directions given to the user. We note that world-map misalignment can be predicted from an RGB image of the environment and the floor segmentation mask obtained from the internal map of the navigation system. Since collecting and labelling large amounts of real data is expensive, we developed a tool to simulate human navigation, which is used to generate automatically labelled synthetic data from 3D models of environments. Thanks to this tool, we generate a dataset considering 15 different environments, which is complemented by a small set of videos acquired in a real-world scenario and manually labelled for validation purposes. We hence benchmark an approach based on different ResNet18 configurations and compare their results on both synthetic and real images. We achieved an F1 score of 92.37% in the synthetic domain and 75.42% on the proposed real dataset using our best approach. While the results are promising, we also note that the proposed problem is challenging, due to the domain shift between synthetic and real data, and the difficulty in acquiring real data. The dataset and the developed tool are publicly available to encourage research on the topic at the following URL: <https://github.com/fpv-iplab/WMM-detection-for-visual-navigation-systems>.

Complete Paper #344

Effects of Model Drift on Ship Detection Models

Namita Agarwal¹, Anh Vo¹, Michela Bertolotto¹, Alan Barnett², Ahmed Khalid² and Merry Globin²

¹ School of Computer Science, University College Dublin, Dublin, Ireland

² Dell Technologies, Cork, Ireland

Keywords: Object Detection, Remote Sensing Images, LEVIR-Ship Dataset, YOLO Models, Model Drift.

Abstract: The rapid and accurate detection of ships within the wide sea area is essential for maritime applications. Many machine learning (ML) based object detection models have been investigated to detect ships in remote sensing imagery in previous research. Despite the availability of large-scale training datasets, the performance of object detection models can decrease significantly when the statistical properties of input images vary according to, for example, weather conditions. This is known as model drift. The occurrence of ML model drift degrades the object detection accuracy and this reduction in accuracy can produce skewed outputs such as, incorrectly classified images or inaccurate semantic tagging, thus making the detection task vulnerable to malicious attacks. The majority of existing approaches that deal with model drift relate to time series. While there is some work on model drift for imagery data and in the context of object detection, the problem has not been extensively investigated for object detection tasks in remote sensing images, especially with large-scale image datasets. In this paper, the

effects of model drift on the detection of ships from satellite imagery data are investigated. Firstly, a YOLOv5 ship detection model is trained and validated using a publicly available dataset. Subsequently, the performance of the model is validated against images subjected to artificial blurriness, which is used in this research as a form of synthetic concept drift. The reduction of the model's performance according to increasing levels of blurriness demonstrates the effect of model drift. Specifically, the average precision of the model dropped by more than 74% when the images were blurred at the maximum level with a 11×11 Gaussian kernel size. More importantly, as the level of blurriness increased, the mean confidence score of the detections decreased up to 20.8% and the number of detections also reduced. Since the confidence scores and the number of detections are independent of ground truth data, such information has the potential to be utilised to detect model drift in future research.

Session 6B
09:00 - 10:30
Image Enhancement and Restoration

VISAPP
Room Space 3B

Complete Paper #243

Efficient and Accurate Hyperspectral Image Demosaicing with Neural Network Architectures

Eric Wisotzky^{1,2}, Lara Wallburg¹, Anna Hilsmann¹, Peter Eisert^{1,2}, Thomas Wittenberg^{3,4} and Stephan Göb^{3,4}

¹ Computer Vision & Graphics, Fraunhofer HHI, Einsteinufer 37, 10587 Berlin, Germany

² Department of Informatics, Humboldt University, Berlin, Germany

³ Fraunhofer IIS, Erlangen, Germany

⁴ Chair of Visual Computing, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Keywords: Sensor Array and Multichannel Signal Processing, Deep Learning, Biomedical Imaging Techniques, Image Analysis, Image Upsampling.

Abstract: Neural network architectures for image demosaicing have become more and more complex. This results in long training periods of such deep networks and the size of the networks is huge. These two factors prevent practical implementation and usage of the networks in real-time platforms, which generally only have limited resources. This study investigates the effectiveness of neural network architectures in hyperspectral image demosaicing. We introduce a range of network models and modifications, and compare them with classical interpolation methods and existing reference network approaches. The aim is to identify robust and efficient performing network architectures. Our evaluation is conducted on two datasets, "SimpleData" and "SimReal-Data," representing different degrees of realism in multispectral filter array (MSFA) data. The results indicate that our networks outperform or match reference models in both datasets demonstrating exceptional performance. Notably, our approach focuses on achieving correct spectral reconstruction rather than just visual appeal, and this emphasis is supported by quantitative and qualitative assessments. Furthermore, our findings suggest that efficient demosaicing solutions, which require fewer parameters, are essential for practical applications. This research contributes valuable insights into hyperspectral imaging and its potential applications in various fields, including medical imaging.

Complete Paper #254

Two Nonlocal Variational Models for Retinex Image Decomposition

Frank Hammond¹, Catalina Sbert² and Joan Duran²

¹ Higher Polytechnic School, University of the Balearic Islands, Spain

² Department of Mathematics and Computer Science & IAC3, University of the Balearic Islands, Cra. de Valldemossa, km. 7.5, E-07122 Palma, Illes Balears, Spain

Keywords: Retinex Theory, Illumination, Reflectance, Image Decomposition, Low-Light Enhancement, Variational Method, Total Variation, Nonlocal Regularization.

Abstract: Retinex theory assumes that an image can be decomposed into illumination and reflectance components. In this work, we introduce two variational models to solve the ill-posed inverse problem of estimating illumination and reflectance from a given observation. Nonlocal regularization exploiting image self-similarities is used to estimate the reflectance, since it is assumed to contain fine details and texture. The difference between the proposed models comes from the selected prior for the illumination. Specifically, Tychonoff regularization, which promotes smooth solutions, and the total variation, which favours piecewise constant solutions, are independently proposed. A comprehensive theoretical analysis of the resulting functionals is presented within appropriate functional spaces, complemented by an experimental validation for thorough examination.

Complete Paper #289

An Image Sharpening Technique Based on Dilated Filters and 2D-DWT Image Fusion

Victor Bogdan¹, Cosmin Bonchiş¹ and Ciprian Orhei²

¹ West University of Timișoara, Bd. V. Pârvan 4, 045B, Timișoara, RO-300223, Romania

² Polytechnic University of Timișoara, Timișoara, RO-300223, România

Keywords: Dilated Filters, Image Sharpening, Unsharp Masking, Image Contrast Enhancement, Image Processing.

Abstract: Image sharpening techniques are pivotal in image processing, serving to accentuate the contrast between darker and lighter regions in images. Building upon prior research that highlights the advantages of dilated kernels in edge detection algorithms, our study introduces a multi-level dilatation wavelet scheme. This novel approach to Unsharp Masking involves processing the input image through a low-pass filter with varying dilatation factors, followed by wavelet fusion. The visual outcomes of this method demonstrate marked improvements in image quality, notably enhancing details without introducing any undesirable cringing effects. Given the absence of a universally accepted index for optimal image sharpness in current literature, we have employed a range of metrics to evaluate the effectiveness of our proposed technique.

Complete Paper #353

Fast and Reliable Inpainting for Real-Time Immersive Video Rendering

Jakub Stankowski and Adrian Dziembowski

Institute of Multimedia Telecommunications, Poznan University of Technology, Poznań, Poland

Keywords: Immersive Video, Inpainting, Virtual View Synthesis, Real-Time Video Processing.

Abstract: In this paper, the authors describe a fast view inpainting algorithm dedicated to practical, real-time immersive video systems. Inpainting is an inherent step of the entire virtual view rendering process, allowing for achieving high Quality of Experience (QoE) for a user of the immersive video system. The authors propose a novel approach for inpainting, based on dividing the inpainting process into two independent, highly parallelizable stages: view analysis and hole filling. In total, four methods of view analysis and two methods of hole filling were developed, implemented, and evaluated, both in terms of computational time and quality of the virtual view. The proposed technique was compared against an efficient state-of-the-art iterative inpainting technique. The results show that the proposal allows for achieving good objective and subjective quality, requiring less than 2 ms for inpainting of a frame of the typical FullHD multiview sequence.

Session 6C

09:00 - 10:30

Medical Image Applications

VISAPP

Room Space 3C

Complete Paper #85

Pair-GAN: A Three-Validated Generative Model from Single Pairs of Biomedical and Ground Truth Images

Clara Brémond-Martin, Huaqian Wu, Cédric Clouchoux and Kévin François-Bouaou

Witsee, 33 Av. des Champs-Élysées, 75008 Paris, France

Keywords: GAN, Single Input, Auto-Encoder, Biomedical, Pair, Segmentation.

Abstract: Generating synthetic pairs of raw and ground truth (GT) image is a strategy to reduce the amount of acquisition and annotation by biomedical experts. Pair image generation strategies, from single-input paired images (SIP), focus on patch-pyramid (PP) or on dual branch generator but, resulting synthetic images are not natural. With few-input images, for raw synthesis, adversarial auto-encoders synthesises more natural images. Here we propose Pair-GAN, a combination of PP containing auto-encoder generators at each level, for the biomedical image synthesis based upon a SIP. PP allows to synthesise using SIP while the AAE generator renders most natural the image content. We use for this work two biomedical datasets containing raw and GT images. Our architecture is evaluated with seven state of the art method updated for SIP: qualitative, similitude and segmentation metrics, Kullback Leibler divergences from synthetic and original feature image representations, computational costs and statistical analyses. Pair-GAN generates most qualitative and natural outputs, similar to original pairs with complex shape not produced by other methods, however with increased memory needs. Future works may use this generative procedure for multimodal biomedical dataset synthesis to help their automatic processing such as classification or segmentation with deep learning tools.

Complete Paper #99

CaRe-CNN: Cascading Refinement CNN for Myocardial Infarct Segmentation with Microvascular Obstructions

Franz Thaler^{1,2}, Matthias Gsell¹, Gernot Plank¹ and Martin Urschler³

¹ *Gottfried Schatz Research Center: Medical Physics and Biophysics, Medical University of Graz, Graz, Austria*

² *Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria*

³ *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria*

Keywords: Machine Learning, Image Segmentation, Myocardial Infarction.

Abstract: Late gadolinium enhanced (LGE) magnetic resonance (MR) imaging is widely established to assess the viability of myocardial tissue of patients after acute myocardial infarction (MI). We propose the Cascading Refinement CNN (CaRe-CNN), which is a fully 3D, end-to-end trained, 3-stage CNN cascade that exploits the hierarchical structure of such labeled cardiac data. Throughout the three stages of the cascade, the label definition changes and CaRe-CNN learns to gradually refine its intermediate predictions accordingly. Furthermore, to obtain more consistent qualitative predictions, we propose a series of post-processing steps that take anatomical constraints into account. Our CaRe-CNN was submitted to the FIMH 2023 MYOSAIQ challenge, where it ranked second out of 18 participating teams. CaRe-CNN showed great improvements most notably when segmenting the difficult but clinically most relevant myocardial infarct tissue (MIT) as well as microvascular obstructions (MVO). When computing the average scores over all labels, our method obtained the best score in eight out of ten metrics. Thus, accurate cardiac segmentation after acute MI via our CaRe-CNN allows generating patient-specific models of the heart serving as an important step towards personalized medicine.

Complete Paper #185

Comparing 3D Shape and Texture Descriptors Towards Tourette's Syndrome Prediction Using Pediatric Magnetic Resonance Imaging

Murilo Costa de Barros¹, Kaue Duarte², Chia-Jui Hsu³, Wang-Tso Lee⁴ and Marco Garcia de Carvalho¹

¹ *Computing Visual Laboratory, School of Technology - UNICAMP, R. Paschoal Marmo, 1888 - Jd. Nova Itália, 13484-332 - Limeira, São Paulo, Brazil*

² *Vascular Imaging Laboratory, Calgary University, 2500 University Dr NW, Calgary, AB T2N 1N4, Canada*

³ *Department of Pediatrics, National Taiwan University Children's Hospital, Taiwan*

⁴ *Department of Pediatrics, National Taiwan University Hospital Hsinchu Branch, Hsinchu, 300001, Taiwan*

Keywords: Tourette Syndrome, GLCM, Texture Feature, Shape Feature, Image Processing, Classification, Machine Learning, MRI.

Abstract: Tourette Syndrome (TS) is a neuropsychiatric disorder characterized by the presence of involuntary motor and vocal tics, with its etiology suggesting a strong and complex genetic basis. The detection of TS is mainly performed clinically, but brain imaging provides additional insights about anatomical structures. Interpreting brain patterns is challenging due to the complexity of the texture and shape of the anatomical regions. This study

compares three-dimensional texture and shape features using Gray-Level Co-occurrence Matrix and Scale-Invariant Heat Kernel Signature. These features are analyzed in the context of TS classification (via Support Vector Machines), focusing on anatomical regions believed to be associated with TS. The evaluation is performed on structural Magnetic Resonance (MR) images of 68 individuals (34 TS patients and 34 healthy subjects). Results show that shape features achieve 92.6% accuracy in brain regions like the right thalamus and accumbens area, while texture features reach 73.5% accuracy in regions such as right putamen and left thalamus. Majority voting ensembles using shape features obtain 96% accuracy, with texture features achieving 79.4%. These findings highlight the influence of subcortical regions in the limbic system, consistent with existing literature on TS.

Complete Paper #228

Investigation of Deep Neural Network Compression Based on Tucker Decomposition for the Classification of Lesions in Cavity Oral

Vitor Fernandes¹, Adriano Silva¹, Danilo Pereira¹, Sérgio Cardoso², Paulo R. de Faria³, Adriano Loyola², Thaína Tosta⁴, Leandro Neves⁵ and Marcelo Z. do Nascimento¹

¹ Faculty of Computer Science, Federal University of Uberlândia, Brazil

² Area of Oral Pathology, School of Dentistry, Federal University of Uberlândia, Brazil

³ Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia, Brazil

⁴ Science and Technology Institute, Federal University of São Paulo, Brazil

⁵ Department of Computer Science and Statistics (DCCE), São Paulo State University, Brazil

Keywords: Oral Epithelial Dysplasia, Convolutional Neural Network, Tensors, Histological Image, Classifier, Tucker Decomposition.

Abstract: Cancer in the oral cavity is one of the most common, making it necessary to investigate lesions that could develop into cancer. Initial stage lesions, called dysplasia, can develop into more severe stages of the disease and are characterized by variations in the shape and size of the nucleus of epithelial tissue cells. Due to advances in the areas of digital image processing and artificial intelligence, diagnostic aid systems (CAD) have become a tool to help reduce the difficulties of analyzing and classifying lesions. This paper presents an investigation of the Tucker decomposition in tensors for different CNN models to classify dysplasia in histological images of the oral cavity. In addition to the Tucker decomposition, this study investigates the normalization of H&E dyes on the optimized CNN models to evaluate the behavior of the architectures in the classification stage of dysplasia lesions. The results show that for some of the optimized models, the use of normalization contributed to the performance of the CNNs for classifying dysplasia lesions. However, when the features obtained from the final layers of the CNNs associated with the machine learning algorithms were analyzed, it was noted that the normalization process affected performance during classification.

Session 6D
09:00 - 10:30
Features Extraction

VISAPP
Room Space 4A

Complete Paper #218

Feature Selection for Unsupervised Anomaly Detection and Localization Using Synthetic Defects

Lars Heckler^{1,2} and Rebecca König¹

¹ MVTec Software GmbH, Germany

² Technical University of Munich, Germany

Keywords: Anomaly Detection, Feature Selection, Visual Inspection, Synthetic Defects.

Abstract: Expressive features are crucial for unsupervised visual Anomaly Detection and Localization. State-of-the-art methods like PatchCore or SimpleNet heavily exploit such features from pretrained extractor networks and model their distribution or utilize them for training further parts of the model. However, the layers commonly used for feature extraction might not represent the optimal choice for reaching maximum performance. Thus, we present the first application-specific feature selection strategy for the task of unsupervised Anomaly Detection and Localization that identifies the most suitable layer of a pretrained feature extractor based on the performance on a synthetic validation set. The proposed selection strategy is applicable to any feature extraction-based AD method and may serve as a competitive baseline for future work by not only outperforming single-layer baselines but also features ensembled from multiple layer outputs.

Complete Paper #240

SIDAR: Synthetic Image Dataset for Alignment & Restoration

Monika Kwiatkowski, Simon Matern and Olaf Hellwich

Computer Vision & Remote Sensing, Technische Universität Berlin, Marchstr. 23, Berlin, Germany

Keywords: Synthetic Dataset, Image Alignment, Homography Estimation, Dense Correspondences, Image Restoration, Shadow Removal, Background Subtraction, Descriptor Learning.

Abstract: In this paper, we present a synthetic dataset generation to create large-scale datasets for various image restoration and registration tasks. Illumination changes, shadows, occlusions, and perspective distortions are added to a given image using a 3D rendering pipeline. Each sequence contains the undistorted image, occlusion masks, and homographies. Although we provide two specific datasets, the data generation itself can be customized and used to generate an arbitrarily large dataset with an arbitrary combination of distortions. The datasets allow end-to-end training of deep learning methods for tasks such as image restoration, background subtraction, image matching, and homography estimation. We evaluate multiple image restoration methods to reconstruct the content from a sequence of distorted images. Additionally, a benchmark is provided that evaluates keypoint detectors and image matching methods. Our evaluations show that even learned image descriptors struggle to identify and match keypoints under varying lighting conditions.

Complete Paper #35

Machine Learning in Industrial Quality Control of Glass Bottle Prints

Maximilian Bundscherer, Thomas Schmitt and Tobias Bocklet

Department of Computer Science, Technische Hochschule Nürnberg Georg Simon Ohm, Nuremberg, Germany

Keywords: Machine Learning, Quality Control, Industrial Manufacturing, Glass Bottle Prints.

Abstract: In industrial manufacturing of glass bottles, quality control of bottle prints is necessary as numerous factors can negatively affect the printing process. Even minor defects in the bottle prints must be detected despite reflections in the glass or manufacturing-related deviations. In cooperation with our medium-sized industrial partner, two ML-based approaches for quality control of these bottle prints were developed and evaluated, which can also be used in this challenging scenario. Our first approach utilized different filters to suppress reflections (e.g. Sobel or Canny) and image quality metrics for image comparison (e.g. MSE or SSIM) as features for different supervised classification models (e.g. SVM or k-Neighbors), which resulted in an accuracy of 84%. The images were aligned based on the ORB algorithm, which allowed us to estimate the rotations of the prints, which may serve as an indicator for anomalies in the manufacturing process. In our second approach, we fine-tuned different pre-trained CNN models (e.g. ResNet or VGG) for binary classification, which resulted in an accuracy of 87%. Utilizing Grad-Cam on our fine-tuned ResNet-34, we were able to localize and visualize frequently defective bottle print regions. This method allowed us to provide insights that could be used to optimize the actual manufacturing process. This paper also describes our general approach and the challenges we encountered in practice with data collection during ongoing production, unsupervised preselection, and labeling.

Complete Paper #79

Privacy Preservation in Image Classification Using Seam Doppelganger

Nishitha Prakash and James Pope

Department of Engineering Mathematics, University of Bristol, Bristol, U.K.

Keywords: Privacy Preservation, Seam Doppelganger, Image Classification, Structural Similarity, Histogram of Gradient.

Abstract: Cloud storage usage continues to increase and many cloud storage sites use advanced machine learning models to classify user's images for various purposes, possibly malicious in nature. This introduces very serious privacy concerns where users want to store and view their images on the cloud storage but do not want the models to be able to accurately classify their images. This is a difficult problem and there are many proposed solutions including the seam doppelganger algorithm. Seam Doppelganger uses the seam carving content-aware resizing approach to modify the image in a way that is still human-understandable and has been shown to reduce model accuracy. However, the approach was not tested with different classifiers, is not able to provide complete restoration, and uses a limited dataset. We propose several modifications to the Seam Doppelganger algorithm to better enhance the privacy of the image while keeping it human-readable and able to be fully restored. We modify the energy function to use a histogram of gradients, comprehensively compare seam selection, and evaluate with several pre-trained (on ImageNet and Kaggle datasets) image classification models. We use the

structural similarity index measure (SSIM) to determine the degree of distortion as a proxy for human understanding. The approach degrades the classification performance by 70% and guarantees 100% restoration of the original image.

Session 6A

09:00 - 10:30

Interactive 3D Graphics and Immersive Systems

GRAPP

Room Space 4B

Complete Paper #377

An Evaluation Research on Dynamic Hit Stop Using Eye Gaze

Rena Tomizawa¹ and Tomokazu Ishikawa^{1,2}

¹ *Toyo University, 1-7-11 Akabanedai Kita-ku, Tokyo, Japan*

² *Prometech CG Research, 3-34-3 Hongo bunkyo-ku, Tokyo, Japan*

Keywords: Hit Stop, Eye Gaze, Game Design.

Abstract: The purpose of this study was to verify whether the response changes when hit stop, one of the elements of GameFeel, is changed according to gaze information. First, several participants play the action game by changing the hit stop duration and answer a questionnaire. Based on the results of this questionnaire, the boundary between pleasant and unpleasant hit stop duration is determined based on the idea of discriminant analysis method. Next, within a comfortable hit stop duration range, we examined whether the response is different when the hit stop duration is changed according to the gazing duration. As a result, designing hit stop duration that corresponds to staring duration is important to improve GameFeel.

Complete Paper #401

HandWindowTeleportation: Locomotion with Hand Gestures for Virtual Reality Games

Hibiki Kirihata¹ and Tomokazu Ishikawa^{1,2}

¹ *Toyo University, 1-7-11 Akabanedai Kita-ku, Tokyo, Japan*

² *Prometech CG Research, 3-34-3 Hongo bunkyo-ku, Tokyo, Japan*

Keywords: Virtual Reality, Point and Teleport, Hand Tracking, Hand Gesture.

Abstract: This study designs and evaluates a novel one-handed gesture-based control method for teleportation within a VR game space. For three different stages, we measured travel time, number of moving maneuvers, and accuracy in achieving checkpoints, and used two questionnaires, NASA-TLX and SUS, to evaluate the subject's workload and the usability of proposed system. It was suggested that the proposed teleportation method using hand gestures is appropriate for games that require agility, such as action games, and is less tiring and has better ease of use than the previous method and a method on existing products.

Complete Paper #29

Trajectory Augmentation for Robust Neural Locomotion Controllers

Dhruv Agrawal^{1,2}, Mathias König¹, Jakob Buhmann², Robert Sumner^{1,2} and Martin Guay²

¹ Department Informatik, ETH Zürich, Zurich, Switzerland

² DisneyResearch|Studios, Zurich, Switzerland

Keywords: Nearest Neighbor Search, Data Augmentation, Neural Networks, Motion Generation.

Abstract: Neural Locomotion Controllers are promising real-time character controllers that can learn directly from motion data. However, the current state of the art models suffer from artifacts such as Pose Blocking and Foot Skating caused by poor generalization to real world control signals. We show that this is due to training on unbalanced biased datasets with poor representation for many important gaits and transitions. To solve this poor data problem, we introduce Trajectory Augmentation (TrajAug), a fully automatic data augmentation technique that generates synthetic motion data by using motion matching to stitch sequences from the original dataset to follow random trajectories. By uniformly sampling these trajectories, we can rebalance the dataset and introduce sharper turns that are commonly used in-game but are hard to capture. TrajAug can be easily integrated into the training of existing neural locomotion controllers without the need for adaptation. We show that TrajAug produces better results than training on only the original dataset or a manually augmented dataset.

Session 6A
09:00 - 10:30
Multimodal Systems and Application

HUCAPP
Room Space 4D

Complete Paper #435

Emotional and Meditative States in Interactive Media Access with a Positive Computing Perspective

Carla Bernardino¹, Sílvia Martins¹, Hugo Ferreira² and Teresa Chambel¹

¹ LASIGE, Faculdade de Ciências da Universidade de Lisboa, Portugal

² IBEB, Faculdade de Ciências da Universidade de Lisboa, Portugal

Keywords: Affective and Positive Computing, Meditation, Mindfulness, Emotions, Wellbeing, Health, Awareness, Serendipity, Media4WellBeing, Interactive Media Access, Video, Music, Images, Visualization, Physiological Sensors, User Experience.

Abstract: Media influences human behavior, emotions, states-of-mind, health and sense of wellbeing, with great potential for a positive impact in our lives, and it is becoming pervasive. Advances and discoveries in the field of neurosciences and informatics have the potential to help people in emotional and wellbeing awareness and regulation, in a positive computing perspective. Our project aims to explore this potential around the design and development of an interactive application that allows the access, exploration and visualization of different media, based on their impact on emotional and meditative states. This paper presents scientific background, design choices, functionalities and a user study of this application. Results were very encouraging in terms of perceived efficacy, usefulness, usability and user experience and will inform future directions.

Complete Paper #343

AI-Supported Diagnostic of Depression Using Clinical Interviews: A Pilot Study

Bakir Hadžić¹, Julia Ohse², Michael Danner³, Nicolina Peperkorn², Parvez Mohammed¹, Youssef Shiban² and Matthias Rättsch¹

¹ ViSiR, Reutlingen University, Reutlingen, Germany

² Private University of Applied Sciences, Göttingen, Germany

³ CVSSP, University of Surrey, Guildford, U.K.

Keywords: Deep Learning, Depression Diagnostics, Mental Health, NLP, BERT, GPT.

Abstract: In the face of rising depression rates, the urgency of early and accurate diagnosis has never been more paramount. Traditional diagnostic methods, while invaluable, can sometimes be limited in access and susceptible to biases, potentially leading to underdiagnoses. This paper explores the innovative potential of AI technology, specifically machine learning, as a diagnostic tool for depression. Drawing from prior research, we note the success of machine learning in discerning depression indicators on social media platforms and through automated interviews. A particular focus is given to the BERT-based NLP transformer model, previously shown to be effective in detecting depression from simulated interview data. Our study assessed this model's capability to identify depression from transcribed, semi-structured clinical interviews within a general population sample. While the BERT model displayed an accuracy of 0.71, it was surpassed by an untrained GPT-3.5 model, which achieved an impressive accuracy of 0.88. These findings emphasise the transformative potential of NLP transformer models in the realm of depression detection. However, given the relatively small dataset ($N = 17$) utilised, we advise a measured interpretation of the results. This paper is designed as a pilot study, and further studies will incorporate bigger datasets.

Complete Paper #434

Exploring Multimodal Interactions with a Robot Assistant in an Assembly Task: A Human-Centered Design Approach

Simona D'Attanasio, Théo Alabert, Clara Francis and Anna Studzinska

Icam School of Engineering, Toulouse Campus, 75 av. de Grande Bretagne, CS 97615, 31076 Toulouse Cedex 3, France

Keywords: Human-Robot Interaction, Multimodality, Human-Centered Design, Bidirectional Interaction, Robot Assistant.

Abstract: The rise of collaborative robots, or cobots, has opened up opportunities for shared operations between humans and robots. However, the transition to true human-robot collaboration faces challenges depending on the context and on the implemented interactions. This article aims to contribute to the evolving field of Human-Robot Interaction by addressing practical challenges in real-world scenarios and proposing a comprehensive approach to bidirectional communication between humans and robots. In particular, our research focuses on an elemental operation during an assembly task, observed in real SMEs (Small and Medium Sized Enterprises). We propose a multimodal bidirectional approach incorporating voice, gesture, visual, haptic, and feedback cues. The study involves a Wizard of Oz series of experiments with test subjects to evaluate user satisfaction, and the overall feeling of interaction, among other

aspects. Preliminary analysis supports hypotheses related to the effectiveness of multimodality, the positive reception of simple interactions, and the impact of feedback on user experience.

Oral Presentations (Online) 6
09:00 - 10:30
Segmentation and Grouping

VISAPP
Room Online 2

Complete Paper #337

BEVFastLine: Single Shot Fast BEV Line Detection for Automated Parking Applications

Praveen Narasappareddygar¹, Venkatesh Karunamoorthy¹, Venkatesh Karunamoorthy¹, Shubham Sonarghare², Ganesh Sistu², Prasad Deshpande² and Prasad Deshpande²

¹ Valeo India Pvt. Ltd., Chennai, India

² Valeo Vision Systems, Tuam, Ireland

Keywords: Road Marking Detection, Autonomous Parking, Multi Camera Line Detection, Birds Eye View.

Abstract: In autonomous parking scenarios, accurate near-field environmental perception is crucial for smooth operations. Parking line detection, unlike the well-understood lane detection, poses unique challenges due to its lack of spatial consistency in orientation, location, and varied appearances in color, pattern, and background surfaces. Consequently, state-of-the-art models for lane detection, which rely on anchors and offsets, are not directly applicable. This paper introduces BEVFastLine, a novel end-to-end line marking detection architecture in Birds Eye View (BEV) space, designed for 360° multi-camera perception applications. BEVFastLine integrates our single-shot line detection methodology with advanced Inverse Perspective Mapping (IPM) techniques, notably our fast splatting technique, to efficiently detect line markings in varied spatial contexts. This approach is suitable for real-time hardware in Level-3 automated vehicles. BEVFastLine accurately localizes parking lines in BEV space with up to 10 cm precision. Our methods, including the 4X faster Fast Splat and single-shot detection, surpass LSS and OFT in accuracy, achieving 80.1% precision, 90% recall, and nearly doubling the performance of BEV-based segmentation and polyline models. This streamlined solution is highly effective in complex, dynamic parking environments, offering high precision localization within 10 meters around the ego vehicle.

Complete Paper #54

Estimation of Package-Boundary Confidence for Object Recognition in Rainbow-SKU Depalletizing Automation

Kento Sekiya¹, Taiki Yano¹, Nobutaka Kimura² and Kiyoto Ito¹

¹ Research & Development Group, Hitachi, Ltd., Kokubunji, Tokyo, Japan

² Research & Development Division, Hitachi America, Ltd., Holland, Michigan, U.S.A.

Keywords: Object Recognition, Robot, Depalletizing, Boundary Confidence.

Abstract: We developed a reliable object recognition method for a rainbow-SKU depalletizing robot. Rainbow SKUs include various types of objects such as boxes, bags, and bottles. The objects' areas need to be estimated in order to automate a depalletizing robot; however, it is difficult to detect the boundaries between

adjacent objects. To solve this problem, we focus on the difference in the shape of the boundaries and propose package-boundary confidence, which assesses whether the recognized boundary correctly corresponds to that of an object unit. This method classifies recognition results into four categories on the basis of the objects' shape and calculates the package-boundary confidence for each category. The results of our experimental evaluation indicate that the proposed method with slight displacement, which is automatic recovery, can achieve a recognition success rate of 99.0%. This is higher than that with a conventional object recognition method. Furthermore, we verified that the proposed method is applicable to a real-world depalletizing robot by combining package-boundary confidence with automatic recovery.

Complete Paper #101

SAMMI: Segment Anything Model for Malaria Identification

Luca Zedda, Andrea Loddo and Cecilia Di Ruberto

Department of Mathematics and Computer Science, University of Cagliari, Italy

Keywords: Computer Vision, Deep Learning, Object Detection, Image Processing, Blood Smear Images, Malaria Parasite Detection.

Abstract: Malaria, a life-threatening disease caused by the Plasmodium parasite, is a pressing global health challenge. Timely detection is critical for effective treatment. This paper introduces a novel computer-aided diagnosis system for detecting Plasmodium parasites in blood smear images, aiming to enhance automation and accessibility in comprehensive screening scenarios. Our approach integrates the Segment Anything Model for precise unsupervised parasite detection. It then employs a deep learning framework, combining Convolutional Neural Networks and Vision Transformer to accurately classify malaria-infected cells. We rigorously evaluate our system using the IML public dataset and compare its performance against various off-the-shelf object detectors. The results underscore the efficacy of our method, demonstrating superior accuracy in detecting and classifying malaria-infected cells. This innovative Computer-aided diagnosis system presents a reliable and near real-time solution for malaria diagnosis, offering significant potential for widespread implementation in healthcare settings. By automating the diagnosis process and ensuring high accuracy, our system can contribute to timely interventions, thereby advancing the fight against malaria globally.

Complete Paper #203

Unsupervised Few-Shot Image Segmentation with Dense Feature Learning and Sparse Clustering

Kuangdai Leng¹, Robert Atwood², Winfried Kockelmann³, Deniza Chekrygina¹ and Jeyan Thiyagalingam¹

¹ Scientific Computing Department, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot, U.K.

² Diamond Light Source, Rutherford Appleton Laboratory, Didcot, U.K.

³ ISIS Neutron and Muon Source, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot, U.K.

Keywords: Unsupervised Learning, Image and Video Segmentation, Representation Learning, Regional Adjacency Graph.

Abstract: Fully unsupervised semantic segmentation of images has been a challenging problem in computer vision. Many deep learning models have been developed for this task, most of which

using representation learning guided by certain unsupervised or self-supervised loss functions towards segmentation. In this paper, we conduct dense or pixel-level representation learning using a fully-convolutional autoencoder; the learned dense features are then reduced onto a sparse graph where segmentation is encouraged from three aspects: non-normalised cut, similarity and continuity. Our method is one- or few-shot, minimally requiring only one image (i.e., the target image). To mitigate overfitting caused by few-shot learning, we compute the reconstruction loss using augmented size-varying patches sampled from the image(s). We also propose a new adjacency-based loss function for continuity, which allows the number of superpixels to be arbitrarily large whereby the creation of the sparse graph can remain fully unsupervised. We conduct quantitative and qualitative experiments using computer vision images and videos, which show that segmentation becomes more accurate and robust using our sparse loss functions and patch reconstruction. For comprehensive application, we use our method to analyse 3D images acquired from X-ray and neutron tomography. These experiments and applications show that our model trained with one or a few images can be highly robust for predicting many unseen images with similar semantic contents; therefore, our method can be useful for the segmentation of videos and 3D images of this kind with lightweight model training in 2D.

Oral Presentations (Online) 6
09:00 - 10:30
Deep Learning for Visual Understanding

VISAPP
Room Online 1

Complete Paper #25

Improving Pseudo-Labeling and Enhancing Robustness for Semi-Supervised Domain Generalization

Adnan Khan, Mai A. Shaaban and Muhammad Haris Khan
Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, U.A.E.

Keywords: Visual Recognition, Domain Generalization, Semi-Supervised Learning, Transfer Learning.

Abstract: Beyond attaining domain generalization (DG), visual recognition models should also be data-efficient during learning by leveraging limited labels. We study the problem of Semi-Supervised Domain Generalization (SSDG) which is crucial for real-world applications like automated healthcare. SSDG requires learning a cross-domain generalizable model when the given training data is only partially labelled. Empirical investigations reveal that the DG methods tend to underperform in SSDG settings, likely because they are unable to exploit the unlabelled data. Semi-supervised learning (SSL) shows improved but still inferior results compared to fully-supervised learning. A key challenge, faced by the best performing SSL-based SSDG methods, is selecting accurate pseudo-labels under multiple domain shifts and reducing overfitting to source domains under limited labels. In this work, we propose new SSDG approach, which utilizes a novel uncertainty-guided pseudo-labelling with model averaging (UPLM). Our uncertainty-guided pseudo-labelling (UPL) uses model uncertainty to improve pseudo-labelling selection, addressing poor model calibration under multi-source unlabelled data. The UPL technique, enhanced by our novel model averaging (MA) strategy, mitigates overfitting to source domains with limited labels. Extensive experiments on key representative DG datasets suggest that our method demonstrates effectiveness against existing methods. Our code and chosen labelled data seeds are available on GitHub: <https://github.com/Adnan-Khan7/UPLM>.

Complete Paper #135

Interpretable Anomaly Analysis for Surveillance Video

Meng Dong

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

Keywords: Interpretable Anomaly Detection.

Abstract: Nowadays, there exist plenty of techniques for surveillance video anomaly detection. However, most works focus on anomaly detection, ignoring the interpreting process for anomaly reasons, especially for real-time anomaly monitoring. The automatic surveillance systems would respond to plenty of alarms based on the types and scores of anomalies and then report to the proper parties. Usually, there exist various types of anomalies, such as abnormal objects, motion, and behaviors, captured by surveillance cameras and defined by application requirements. In this work, we investigate the perspective of reasons for anomalies and propose a general and interpretable anomaly analysis framework formed by three branches: abnormal object categories detection, anomalous motion detection, and strange/violent behaviors recognition, and then related scores are combined to obtain final results. The above three branches could cover various anomaly types in the real world. Besides, the fusion of branches is multivariate based on specific domains or user requirements. They can work together or individually based on specific requirements. In particular, an online non-parametric hierarchical event updating motion model is proposed to explore general motion anomaly. The events with low frequency or have not been seen before could be detected in an unsupervised and continue-updating way. Besides, abnormal human behaviors, such as falling and violence, could be recognized by a spatial-temporal transformer model. Three branches cover different regions but complement each other for joint detection and output interpretable anomaly results. Evaluated on existing datasets, our results are competitive to the online and offline state-of-the-art on several public datasets, demonstrating the proposed method's scene-independent and interpretable abilities even with simple motion update methods. Moreover, the performance of individual anomaly detectors also validates the effectiveness of our proposed method.

Complete Paper #374

Word and Image Embeddings in Pill Recognition

Richárd Rádli, Zsolt Vörösházi and László Czúni

University of Pannonia, 8200 Veszprém, Egyetem u. 10., Hungary

Keywords: Metrics Learning, Pill Recognition, Multi-Modal Learning, Multihead Attention, Multi-Stream Network, Dynamic Margin Triplet Loss.

Abstract: Pill recognition is a key task in healthcare and has a wide range of applications. In this study, we are addressing the challenge to improve the accuracy of pill recognition in a metrics learning framework. A multi-stream visual feature extraction and processing architecture, with multi-head attention layers, is used to estimate the similarity of pills. We are introducing an essential enhancement to the triplet loss function to leverage word embeddings for the injection of textual pill similarity into the visual model. This improvement refines the visual embedding on a finer scale than conventional triplet loss models resulting in higher accuracy of the visual model. Experiments and evaluations are made on a new pill dataset, freely available.

Complete Paper #426

Multimodal Crowd Counting with Pix2Pix GANs

Muhammad Asif Khan¹, Hamid Menouar¹ and Ridha Hamila²

¹ Qatar Mobility Innovations Center, Qatar University, Doha, Qatar

² Department of Electrical Engineering, Qatar University, Doha, Qatar

Keywords: Crowd Counting, CNN, Density Estimation, Multimodal, RGB, Thermal.

Abstract: Most state-of-the-art crowd counting methods use color (RGB) images to learn the density map of the crowd. However, these methods often struggle to achieve higher accuracy in densely crowded scenes with poor illumination. Recently, some studies have reported improvement in the accuracy of crowd counting models using a combination of RGB and thermal images. Although multimodal data can lead to better predictions, multimodal data might not be always available beforehand. In this paper, we propose the use of generative adversarial networks (GANs) to automatically generate thermal infrared (TIR) images from color (RGB) images and use both to train crowd counting models to achieve higher accuracy. We use a Pix2Pix GAN network first to translate RGB images to TIR images. Our experiments on several state-of-the-art crowd counting models and benchmark crowd datasets report significant improvement in accuracy.

Poster Session 3
10:30 - 11:30

IVAPP
Foyer

Abstract #456

Naming Categorical Palette Colors

Martijn Tennekes¹ and Marco Puts²

¹ Department of Methodology and Quality, Statistics Netherlands, CBS-Weg 11, Heerlen, The Netherlands

² Statistics Netherlands, CBS-Weg 11, Heerlen, The Netherlands

Keywords: N/A

Abstract: Colors that are easy to name are also easy to recall. In the context of information visualization, a palette featuring easily recognizable colors, such as blue, red, and purple, is more memorable than one with less distinctive options like olive green, lavender, and maroon red. In this research, we explore the mapping between categorical palette colors and the eleven main color names identified by Boynton and Olson (1987), referred to as the Boynton color names. Our goal is to establish a 1-to-1 relationship, ensuring each palette color is consistently named with one Boynton color and vice versa.

Color naming is culture and language dependent. While Boynton color names have universal translations, language nuances can affect naming. We conduct a user experiment with Dutch participants, assigning Boynton color names to colors drawn from the sRGB color space. The results inform weights for Boynton color centroids, allowing us to derive language-specific nameability scores for categorical color palettes.

This research contributes to the cols4all open-source software tool (Tennekes and Puts, 2023), an R package with a graphical user interface for analyzing and comparing color palettes. We introduce a new property, "nameability," to assess the 1-to-1 mapping with Boynton colors. The analysis tab called "naming" presents results as a table (see video), where palette colors correspond to rows and Boynton color names to columns. Horizontal lines indicate multiple Boynton names for a palette color, and vertical lines signify one Boynton name used for multiple colors.

Complete Paper #398

Navigating the Trade-Off Between Explainability and Privacy

Johanna Schmidt¹, Verena Pietsch², Martin Nocker³, Michael Rader⁴ and Alessio Montuoro⁵

¹ VRVis Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH, Vienna, Austria

² FILL Gesellschaft m.b.H, Gurten, Austria

³ MCI The Entrepreneurial School, Innsbruck, Austria

⁴ Fraunhofer Austria Research GmbH, Wattens, Austria

⁵ Software Competence Center Hagenberg GmbH, Hagenberg, Austria

Keywords: Explainable AI, Privacy, Explainability, Data Visualization, Homomorphic Encryption.

Abstract: Understanding the rationale behind complex AI decisions becomes increasingly vital as AI evolves. Explainable AI technologies are pivotal in demystifying these decisions, offering methods and tools to interpret and communicate the reasoning behind AI-driven outcomes. However, the rise of Explainable AI is juxtaposed with the imperative to protect sensitive data, leading to the integration of encryption techniques in AI development. This paper explores the intricate coexistence of explainability and encryption in AI, presenting a dilemma where the quest for transparency clashes with the imperative to secure sensitive information. The contradiction is particularly evident in methods like homomorphic encryption, which, while ensuring data security, complicates the provision of clear and interpretable explanations for AI decisions. The discussion delves into the conflicting goals of these approaches, surveying the use of privacy-preserving methods in Explainable AI and identifying potential directions for future research. Contributions include a comprehensive survey of privacy considerations in current Explainable AI approaches, an exemplary use case demonstrating visualization techniques for explainability in secure environments, and identifying avenues for future work.

Complete Paper #400

Evaluation of Approximate Reflectional Symmetry

Martin Maňák¹, David Podgorelec² and Ivana Kolingerová³

¹ New Technologies for the Information Society (NTIS), University of West Bohemia, Univerzitní 8, 301 00 Pilsen, Czech Republic

² Faculty of Electrical Engineering and Computer Science, University of Maribor, Koroška cesta 46, SI-2000 Maribor, Slovenia

³ Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 301 00 Pilsen, Czech Republic

Keywords: Symmetry, Approximate, Visualization, Evaluation.

Abstract: When an object can be split by a plane into two symmetrical parts, one being the mirrored image of the other, the object has a reflectional symmetry with respect to that plane. The symmetry is often only approximate and not necessarily global. Many algorithms exist for the detection of symmetries and there are various applications utilizing symmetrical properties. Yet there are not so many ways to measure the amount of approximate reflectional symmetry. In this paper, we introduce a method for the evaluation of approximate symmetry for objects represented as a point cloud. The method consists of three parts - a relative symmetry distance for measuring the amount of approximate

reflectional symmetry, a plot of relative errors, and visualization of errors. This method offers a way how to compare different objects by the amount of symmetry and improves understanding of the symmetrical properties of objects, both quantitatively and visually.

Complete Paper #454

Visual Analysis of Military Diving Incident Reports

G. Walsh, N. Andersen, J. Kusnick, E. Sørensen and S. Jänicke

Centre for Visual Data Science, Department of Mathematics and Computer Science, University of Southern Denmark, Denmark

Keywords: Geospatial Temporal Visualization, Military Diving, Incident Reporting, Incident Analysis.

Abstract: Military diving can be characterized by its specialized tasks, advanced equipment, and ever-changing operational environments, giving rise to remarkably unique operational challenges. Given this complexity, accidents and incidents can occur. Incident reporting and analysis systems exist to collect data, perform trend analysis on safety interventions, and modify unsafe behaviors. By examining existing Military Diving Incident Reporting (MDIR) systems and literature, we reveal that individual European countries have segregated systems, but these lack standardization and interoperability. This paper introduces a novel visualization tool, focusing primarily on rebreather incidents, a critical piece of equipment with a history of incidents in both military and civilian contexts. We compare our proposed system with existing models, highlighting strengths and areas for improvement. This tool aims to illustrate the potential of a broader, more comprehensive system, which would cover not only rebreather incidents but all types of military diving incidents. The paper concludes with insights into the potential of a comprehensive, standardized MDIR system, proposing future extensions and research opportunities to enhance military diving safety and operational effectiveness.

Poster Session 3
10:30 - 11:30

VISAPP
Foyer

Complete Paper #33

Evaluating Learning Potential with Internal States in Deep Neural Networks

Shogo Takasaki and Shuichi Enokida

Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Keywords: Deep Learning, Anticipating Accidents, Long Short-Term Memory, Evaluating Learning Potential.

Abstract: Deploying deep learning models on small-scale computing devices necessitates considering computational resources. However, reducing the model size to accommodate these resources often results in a trade-off with accuracy. The iterative process of training and validating to optimize model size and accuracy can be inefficient. A potential solution to this dilemma is the extrapolation of learning curves, which evaluates a model's potential based on initial learning curves. As a result, it is possible to efficiently search for a network that achieves a balance between accuracy and model size. Nonetheless, we posit that a more effective approach to analyzing the latent potential of training models is to focus on the internal state, rather than merely relying on the validation scores. In this vein, we propose a module

dedicated to scrutinizing the network's internal state, with the goal of automating the optimization of both accuracy and network size. Specifically, this paper delves into analyzing the latent potential of the network by leveraging the internal state of the Long Short-Term Memory (LSTM) in a traffic accident prediction network.

Complete Paper #456

Semantic Image Synthesis for Realistic Image Generation in Robotic Assisted Partial Nephrectomy

Stefano Mazzocchetti¹, Laura Cerenelli¹, Lorenzo Bianchi^{2,3}, Riccardo Schiavina^{2,3} and Emanuela Marcelli¹

¹ *eDIMES Lab, Laboratory of Bioengineering, Department of Medical and Surgical Sciences, University of Bologna, Via Massarenti, 9, 40138 Bologna, Italy*

² *Division of Urology, IRCCS Azienda Ospedaliero, Universitaria di Bologna, Via Massarenti, 9, 40138 Bologna, Italy*

³ *Department of Medical and Surgical Sciences, University of Bologna, Via Massarenti, 9, 40138 Bologna, Italy*

Keywords: Minimally Invasive Surgery, Robotic Surgery, Semantic Image Synthesis, Deep Learning, GAN, Computer Vision.

Abstract: With the continuous evolution of robotic-assisted surgery, the integration of advanced technologies into the field becomes pivotal for improving surgical outcomes. The lack of labelled surgical datasets limits the range of possible applications of deep learning techniques in the surgical field. As a matter of fact, the annotation process to label datasets is time consuming. This paper introduces an approach for realistic image generation in the context of Robotic Assisted Partial Nephrectomy (RAPN) using the Semantic Image Synthesis (SIS) technique. Leveraging descriptive semantic maps, our method aims to bridge the gap between abstract scene representation and visually compelling laparoscopic images. It is shown that our approach can effectively generate photo-realistic Minimally Invasive Surgery (MIS) synthetic images starting from a sparse set of annotated real images. Furthermore, we demonstrate that synthetic data can be used to train a semantic segmentation network that generalizes on real data reducing the annotation time needed.

Complete Paper #162

Modification of DDIM Encoding for Generating Counterfactual Pathology Images of Malignant Lymphoma

Ryoichi Koga¹, Mauricio Kugler¹, Tatsuya Yokota¹, Kouichi Ohshima^{2,3}, Hiroaki Miyoshi^{2,3}, Miharu Nagaishi², Noriaki Hashimoto⁴, Ichiro Takeuchi^{4,5} and Hidekata Hontani¹

¹ *Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya-shi, Aichi, 466-8555, Japan*

² *Kurume University Department of Pathology, 67 Asahi-cho, Kurume-shi, Fukuoka, 830-0011, Japan*

³ *The Japanese Society of Pathology, 1-2-5 Yushima, Bunkyo-ku, Tokyo, 113-0034, Japan*

⁴ *RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan*

⁵ *Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601, Japan*

Keywords: Counterfactual Images, Diffusion Models, Pathological Images, Malignant Lymphoma, Causal Inference.

Abstract: We propose a method that modifies encoding in DDIM (Denosing Diffusion Implicit Model) to improve the quality of counterfactual histopathological images of malignant lymphoma. Counterfactual medical images are widely employed for analyzing the changes in images accompanying disease. For the analysis of pathological images, it is desired to accurately represent the types of individual cells in the tissue. We employ DDIM because it can refer to exogenous variables in causal models and can generate counterfactual images. Here, one problem of DDIM is that it does not always generate accurate images due to approximations in the forward process. In this paper, we propose a method that reduces the errors in the encoded images obtained in the forward process. Since the computation in the backward process of DDIM does not include any approximation, the accurate encoding in the forward process can improve the accuracy of the image generation. Our proposed method improves the accuracy of encoding by explicitly referring to the given original image. Experiments demonstrate that our proposed method accurately reconstructs original images, including microstructures such as cell nuclei, and outperforms the conventional DDIM in several measures of image generation.

Complete Paper #186

Relationship Between Semantic Segmentation Model and Additional Features for 3D Point Clouds Obtained from on-Vehicle LIDAR

Hisato Hashimoto and Shuichi Enokida

Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Keywords: Semantic Segmentation, Point Cloud, Deep Learning.

Abstract: The study delves into semantic segmentation's role in recognizing regions within data, with a focus on images and 3D point clouds. While images from wide-angle cameras are prevalent, they falter in challenging environments like low light. In such cases, LIDAR (Laser Imaging Detection and Ranging), despite its lower resolution, excels. The combination of LIDAR and semantic segmentation proves effective for outdoor environment understanding. However, highly accurate models often demand substantial parameters, leading to computational challenges. Techniques like knowledge distillation and pruning offer solutions, though with possible accuracy trade-offs. This research introduces a strategy to merge feature descriptors, such as reflectance intensity and histograms, into the semantic segmentation model. This process balances accuracy and computational efficiency. The findings suggest that incorporating feature descriptors suits smaller models, while larger models can benefit from optimizing computation and utilizing feature descriptors for recognition tasks. Ultimately, this research contributes to the evolution of resource-efficient semantic segmentation models for autonomous driving and similar fields.

Complete Paper #193

Teacher-Student Models for AI Vision at the Edge: A Car Parking Case Study

Mbasa Molo^{1,2}, Emanuele Carlini², Luca Ciampi²,
Claudio Gennaro² and Lucia Vadicamo²

¹ *Department of Computer Science, University of Pisa, Pisa, Italy*

² *Institute of Information Science and Technologies (ISTI), Italian National Research Council (CNR), Pisa, Italy*

Keywords: Knowledge Distillation, Computer Vision, Object Recognition, Deep Learning, Edge Computing.

Abstract: The surge of the Internet of Things has sparked a multitude of deep learning-based computer vision applications that extract relevant information from the deluge of data coming from Edge devices, such as smart cameras. Nevertheless, this promising approach introduces new obstacles, including the constraints posed by the limited computational resources on these devices and the challenges associated with the generalization capabilities of the AI-based models against novel scenarios never seen during the supervised training, a situation frequently encountered in this context. This work proposes an efficient approach for detecting vehicles in parking lot scenarios monitored by multiple smart cameras that train their underlying AI-based models by exploiting knowledge distillation. Specifically, we consider an architectural scheme comprising a powerful and large detector used as a teacher and several shallow models acting as students, more appropriate for computational-bounded devices and designed to run onboard the smart cameras. The teacher is pre-trained over general-context data and behaves like an oracle, transferring its knowledge to the smaller nodes; on the other hand, the students learn to localize cars in new specific scenarios without using further labeled data, relying solely on the distilled loss coming from the oracle. Preliminary results show that student models trained only with distillation loss increase their performances, sometimes even outperforming the results achieved by the same models supervised with the ground truth.

Complete Paper #199

Most Relevant Viewpoint of an Object: A View-Dependent 3D Saliency Approach

Marie Pelissier-Combesure, Sylvie Chambon and
G eraldine Morin

University of Toulouse, IRIT, Toulouse INP, France

Keywords: View-Dependent Saliency, Best View Selection, 3D Models/2D Images, User Study Evaluation.

Abstract: A viewpoint of a 3D object is the position from which we observe the object. A viewpoint always highlights some 3D parts and discards other parts of an object. Here, we define a good viewpoint as offering a relevant view of the object: a view that best showcases the object and that is the most representative of the object. Best view selection plays an essential role in many computer vision and virtual reality applications. In this paper, given a model and a particular viewpoint, we want to quantify its relevance -not aesthetics. We propose a geometric method for selecting the most relevant viewpoint for a 3D object by combining visibility and view-dependent saliency. Evaluating the quality of an estimated best viewpoint is a challenge. Thus, we propose an evaluation protocol that considers two different and complementary solutions: a user study with more than 200 participants to collect human preferences and an analysis of image dataset picturing objects of interest. This evaluation highlights the correlation between our method and human preferences. A quantitative comparison demonstrates the efficiency of our approach over reference methods.

Complete Paper #221

Combining Total Variation and Nonlocal Variational Models for Low-Light Image Enhancement

Daniel Torres, Catalina Sbert and Joan Duran

Department of Mathematics and Computer Science & IAC3, University of the Balearic Islands, Cra. de Valldemossa, km. 7.5, E-07122 Palma, Illes Balears, Spain

Keywords: Low-Light Image Enhancement, Illumination Estimation, Reflectance Estimation, Retinex, Variational Method, Total Variation, Nonlocal Regularization.

Abstract: Images captured under low-light conditions impose significant limitations on the performance of computer vision applications. Therefore, improving their quality by discounting the effects of the illumination is crucial. In this paper, we present a low-light image enhancement method based on the Retinex theory. Our approach estimates illumination and reflectance in two steps. First, the illumination is obtained as the minimizer of an energy functional involving total variation regularization, which favours piecewise smooth solutions. Next, the reflectance component is computed as the minimizer of an energy functional involving contrast-invariant nonlocal regularization and a fidelity term preserving the largest gradients of the input image.

Complete Paper #229

Oral Dysplasia Classification by Using Fractal Representation Images and Convolutional Neural Networks

Rafael Carvalho¹, Adriano Silva¹, Alessandro Martins², Sérgio Cardoso³, Guilherme Freire⁴, Paulo R. de Faria⁵, Adriano Loyola³, Thaína Tosta⁶, Leandro Neves⁷ and Marcelo Z. do Nascimento¹

¹ Faculty of Computer Science, Federal University of Uberlândia, Brazil

² Federal Institute of Triângulo Mineiro, Brazil

³ Area of Oral Pathology, School of Dentistry, Federal University of Uberlândia, Brazil

⁴ Department of Informatics Engineering, Faculty of Engineering, University of Porto, Portugal

⁵ Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia, Brazil

⁶ Science and Technology Institute, Federal University of São Paulo, Brazil

⁷ Department of Computer Science and Statistics (DCCE), São Paulo State University, Brazil

Keywords: Dysplasia, Fractal Geometry, Reshape, Convolutional Neural Network, Ensemble, Histological Image.

Abstract: Oral cavity lesions can be graded by specialists, a task that is both difficult and subjective. The challenges in defining patterns can lead to inconsistencies in the diagnosis, often due to the color variations on the histological images. The development of computational systems has emerged as an effective approach for aiding specialists in the diagnosis process, with color normalization techniques proving to enhance diagnostic accuracy. There remains an open challenge in understanding the impact of color normalization on the classification of histological tissues representing dysplasia groups. This study presents an approach to classify dysplasia lesions based on ensemble models, fractal representations, and convolutional neural networks (CNN). Additionally, this work evaluates the influence of color normalization in the preprocessing stage. The results obtained with the proposed methodology were analyzed with and without

the preprocessing stage. This approach was applied in a dataset composed of 296 histological images categorized into healthy, mild, moderate, and severe oral epithelial dysplasia tissues. The proposed approaches based on the ensemble were evaluated with the cross-validation technique resulting in accuracy rates ranging from 96.1% to 98.5% with the non-normalized dataset. This approach can be employed as a supplementary tool for clinical applications, aiding specialists in decision-making regarding lesion classification.

Complete Paper #253

Multi-Task Planar Reconstruction with Feature Warping Guidance

Luan Wei¹, Anna Hilsmann¹ and Peter Eisert^{1,2}

¹ Fraunhofer Heinrich Hertz Institute, Berlin, Germany

² Humboldt University, Berlin, Germany

Keywords: Planar Reconstruction, Real-Time, Neural Network, Segmentation, Deep Learning, Scene Understanding.

Abstract: Piece-wise planar 3D reconstruction simultaneously segments plane instances and recovers their 3D plane parameters from an image, which is particularly useful for indoor or man-made environments. Efficient reconstruction of 3D planes coupled with semantic predictions offers advantages for a wide range of applications requiring scene understanding and concurrent spatial mapping. However, most existing planar reconstruction models either neglect semantic predictions or do not run efficiently enough for real-time applications. We introduce SOLOPlanes, a real-time planar reconstruction model based on a modified instance segmentation architecture which simultaneously predicts semantics for each plane instance, along with plane parameters and piece-wise plane instance masks. We achieve an improvement in instance mask segmentation by including multi-view guidance for plane predictions in the training process. This cross-task improvement, training for plane prediction but improving the mask segmentation, is due to the nature of feature sharing in multi-task learning. Our model simultaneously predicts semantics using single images at inference time, while achieving real-time predictions at 43 FPS. Code is available at: <https://github.com/fraunhoferhhi/SOLOPlanes>.

Complete Paper #294

Learning Projection Patterns for Direct-Global Separation

Takaoki Ueda, Ryo Kawahara and Takahiro Okabe

Department of Artificial Intelligence, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan

Keywords: Direct-Global Separation, Projector-Camera System, Projection Patterns, End-to-End Optimization.

Abstract: Separating the direct component such as diffuse reflection and specular reflection and the global component such as inter-reflection and subsurface scattering is important for various computer vision and computer graphics applications. Conventionally, high-frequency patterns designed by physics-based model or signal processing theory are projected from a projector to a scene, but their assumptions do not necessarily hold for real images due to the shallow depth of field of a projector and the limited spatial resolution of a camera. Accordingly, in this paper, we propose a data-driven approach for direct-global separation. Specifically, our proposed method learns not only the separation module but also the imaging module, i.e. the projection patterns at the same time in an end-to-end manner. We conduct a number of

experiments using real images captured with a projector-camera system, and confirm the effectiveness of our method.

Complete Paper #306

Combining Progressive Hierarchical Image Encoding and YOLO to Detect Fish in Their Natural Habitat

Antoni Burguera

Departament de Matemàtiques i Informàtica, Universitat de les Illes Balears, Ctra. Valldemossa Km. 7.5, 07122 Palma, Illes Balears, Spain

Keywords: Computer Vision, Progressive Image Encoding, Deep Learning, Underwater Robotics.

Abstract: This paper explores the advantages of evaluating Progressive Image Encoding (PIE) methods in the context of the specific task for which they will be used. By focusing on a particular task—fish detection in their natural habitat—and a specific PIE algorithm—Progressive Hierarchical Image Encoding (PHIE)—, the paper investigates the performance of You Only Look Once (YOLO) in detecting fish in underwater images using PHIE-encoded images. This is particularly relevant in underwater environments where image transmission is slow. Results provide insights into the advantages and drawbacks of PHIE image encoding and decoding, not from the perspective of general metrics such as reconstructed image quality but from the viewpoint of its impact on a task—fish detection—that depends on the PHIE encoded and decoded images.

Complete Paper #338

Blind Deblurring of THz Time-Domain Images Based on Low-Rank Representation

Marina Ljubenović¹, Mário Figueiredo² and Arianna Traviglia¹

¹ *Center for Cultural Heritage Technology, Istituto Italiano di Tecnologia, Venice, Italy*

² *Instituto de Telecomunicações, Instituto Superior Técnico, Lisbon, Portugal*

Keywords: THz TDS, Deblurring, Denoising, Low-Rank.

Abstract: Terahertz (THz) time-domain imaging holds immense potential for material characterization, capturing three-dimensional data across spatial and temporal dimensions. Despite its capabilities, the technology faces hurdles such as frequency-dependent beam-shape effects and noise. This paper proposes a novel, dual-stage framework for improving THz image resolution beyond the wavelength limit. Our method combats blur at lower frequencies and noise at higher frequencies. The first stage entails selective deblurring of lower-frequency bands, addressing beam-related blurring, while the second stage involves denoising the entire THz hyperspectral cube through dimensionality reduction, exploiting its low-rank structure. The synergy of these advanced techniques—beam shaping, noise removal, and low-rank representation—forms a comprehensive approach to enhance THz time-domain images. We present promising preliminary results, showcasing significant improvements across all frequency bands, which is crucial as samples may display varying features across the THz spectrum. Our ongoing work is extending this methodology to complex scenarios such as analyzing multilayered structures in closed ancient manuscripts. This approach paves the way for broader application and refinement of THz imaging in diverse research fields.

Complete Paper #373

Identification of Honeybees with Paint Codes Using Convolutional Neural Networks

Gabriel Santiago-Plaza¹, Luke Meyers², Andrea Gomez-Jaime³, Rafael Meléndez-Ríos¹, Fanfan Noel⁴, Jose Agosto⁴, Tugrul Giray⁴, Josué Rodríguez-Cordero⁴ and Rémi Mégret¹

¹ *Department of Computer Science, University of Puerto Rico Río Piedras, Puerto Rico*

² *Department of Biology, Seattle University, U.S.A.*

³ *Department of Biology, Universidad de Los Andes, Colombia*

⁴ *Department of Biology, University of Puerto Rico Río Piedras, Puerto Rico*

Keywords: Computer Vision, Deep Learning, Re-Identification, Honey Bee Monitoring, Paint Code Recognition.

Abstract: This paper proposes and evaluates methods for the automatic re-identification of honeybees marked with paint codes. It leverages deep learning models to recognize specific individuals from images, which is a key component for the automation of wild-life video monitoring. Paint code marking is traditionally used for individual re-identification in the field as it is less intrusive compared to alternative tagging approaches and is human-readable. To assess the performance of re-id using paint codes, we built a mostly balanced dataset of 8062 images of honeybees marked with one or two paint dots from 8 different colors, generating 64 distinct codes, repeated twice on distinct individual bees. This dataset was used to perform an extensive comparison of convolutional network re-identification approaches. The first approach uses supervised learning to estimate the paint code directly; the second approach uses contrastive learning to learn an identity feature vector that is then used to query a database of known identities. Best performance reached 85% correct identification for all 64 identities, and up to 97.6% for 8 identities, showing the potential of the technique. Ablation studies with variation in training data and selection of IDs provide guidance for future use of this technique in the field.

Complete Paper #381

SAM-Based Detection of Structural Anomalies in 3D Models for Preserving Cultural Heritage

David Jurado-Rodríguez¹, Alfonso López¹, J. Jiménez¹, Antonio Garrido², Francisco Feito¹ and Juan Jurado¹

¹ *Department of Computer Science, University of Jaén, Spain*

² *Department of Cartographic, Geodetic and Photogrammetric Engineering, University of Jaén, Spain*

Keywords: 3D Modeling, Artificial Intelligence, Cultural Heritage Protection, Drones.

Abstract: The detection of structural defects and anomalies in cultural heritage emerges as an essential component to ensure the integrity and safety of buildings, plan preservation strategies, and promote the sustainability and durability of buildings over time. In the search to enhance the effectiveness and efficiency of structural health monitoring of cultural heritage, this work aims to develop an automated method focused on detecting unwanted materials and geometric anomalies on the 3D surfaces of ancient buildings. In this study, the proposed solution combines an AI-based technique for fast-forward image labeling and a fully automatic detection of target classes in 3D point clouds. As an advantage of our method, the use of spatial and geometric features in the 3D models enables the recognition of target materials in the whole point cloud from

seed, resulting from partial detection in a few images. The results demonstrate the feasibility and utility of detecting self-healing materials, unwanted vegetation, lichens, and encrusted elements in a real-world scenario.

Complete Paper #391

A Generative Model for Guided Thermal Image Super-Resolution

Patricia Suárez¹ and Angel Sappa^{1,2}

¹ *Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ingeniería en Electricidad y Computación, CIDIS, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador*

² *Computer Vision Center, Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain*

Keywords: Thermal Super-Resolution, HSV Color Space, Luminance-Driven Bicubic Image.

Abstract: This paper presents a novel approach for thermal super-resolution based on a fusion prior, low-resolution thermal image and H brightness channel of the corresponding visible spectrum image. The method combines bicubic interpolation of the $\times 8$ scale target image with the brightness component. To enhance the guidance process, the original RGB image is converted to HSV, and the brightness channel is extracted. Bicubic interpolation is then applied to the low-resolution thermal image, resulting in a Bicubic-Brightness channel blend. This luminance-bicubic fusion is used as an input image to help the training process. With this fused image, the cyclic adversarial generative network obtains high-resolution thermal image results. Experimental evaluations show that the proposed approach significantly improves spatial resolution and pixel intensity levels compared to other state-of-the-art techniques, making it a promising method to obtain high-resolution thermal.

Complete Paper #417

Colorectal Image Classification Using Randomized Neural Network Descriptors

Jarbas Sá Junior¹ and André Backes²

¹ *Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Brazil*

² *Department of Computing, Federal University of São Carlos, Brazil*

Keywords: Randomized Neural Network, Colorectal Database, Texture Analysis.

Abstract: Colorectal cancer is among the highest incident cancers in the world. A fundamental procedure to diagnose it is the analysis of histological images acquired from a biopsy. Because of this, computer vision approaches have been proposed to help human specialists in such a task. In order to contribute to this field of research, this paper presents a novel way of analyzing colorectal images by using a very discriminative texture signature based on weights of a randomized neural network. For this, we addressed an important multi-class problem composed of eight types of tissues. The results were promising, surpassing the accuracies of many methods present in the literature. Thus, this performance confirms that the randomized neural network signature is an efficient tool for discriminating histological images from colorectal tissues.

Complete Paper #431

Deformable Pose Network: A Multi-Stage Deformable Convolutional Network for 2D Hand Pose Estimation

Sartaj Salman¹, Ali Zakir¹ and Hiroki Takahashi^{1,2}

¹ *Department of Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan*

² *Artificial Intelligence Exploration/Meta-Networking Research Center, The University of Electro-Communications, Tokyo, Japan*

Keywords: Deformable Convolution, Multi-Stage DC, EfficientNet, 2D HPE.

Abstract: Hand pose estimation undergoes a significant advancement with the evolution of Convolutional Neural Networks (CNNs) in the field of computer vision. However, existing CNNs fail in many scenarios in learning the unknown transformations and geometrical constraints along with the other existing challenges for accurate estimation of hand keypoints. To tackle these issues we proposed a multi-stage deformable convolutional network for accurate 2D hand pose estimation from monocular RGB images while considering the computational complexity. We utilized EfficientNet as a backbone due to its powerful feature extraction capability, and deformable convolution to learn about the geometrical constraints. Our proposed model called Deformable Pose Network (DPN) outperforms in predicting the 2D keypoints in complex scenarios. Our analysis on the Panoptic studio hand dataset shows that our proposed model improves the accuracy by 2.36% and 7.29% as compared to existing methods i.e., OCPM and CPM respectively.

Complete Paper #437

Unsupervised Domain Adaptation for Human Pose Action Recognition

Mattias Billast¹, Tom De Schepper^{1,2} and Kevin Mets³

¹ *University of Antwerp, Imec, IDLab, Department of Computer Science, Sint-Pietersvliet 7, 2000 Antwerp, Belgium*

² *AI & Data Department, Imec, Leuven, Belgium*

³ *University of Antwerp Imec, IDLab, Faculty of Applied Engineering, Sint-Pietersvliet 7, 2000 Antwerp, Belgium*

Keywords: Domain Adaptation, Human Pose, GCN, Action Recognition.

Abstract: Personalized human action recognition is important to give accurate feedback about motion patterns, but there is likely no labeled data available to update the model in a supervised way. Unsupervised domain adaptation can solve this problem by closing the gap between seen data and new unseen users. We test several domain adaptation techniques and compare them with each other on this task. We show that all tested techniques improve on the model without any domain adaptation and are only trained on labeled source data. We add multiple improvements by designing a better feature representation tailored to the new user. These improvements include added contrastive loss and varying the backbone encoder. We would need between 30% and 40% labeled data of the new user to get the same results.

Complete Paper #446

Exploring Unsupervised Domain Adaptation Approaches for Water Parameters Estimation from Satellite Images

Mauren Coelho de Andrade¹, Anderson Souza², Bruno Oliveira², Maria Starling², Camila Amorim² and Jefersson Santos³

¹ Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brazil

² Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

³ Department of Computer Science, University of Sheffield, Sheffield, U.K.

Keywords: Domain Adaptation, Remote Sensing, Imbalanced Data.

Abstract: In this paper, we compare several domain adaptation approaches in classifying water quality in reservoirs using spectral data from satellite images to two optical parameters: turbidity and chlorophyll-a. This assessment adds a new possibility in monitoring these water quality parameters, in addition to the traditional in-situ investigation, which is expensive and time-consuming. The study acquired images from two data sources characterized by different geographic regions (USA and Brazil) and verified the inference quality of the model trained in the source domain on samples from the target domain. The experiments used two classifiers, OSCVM and ANN, for domain adaptation methods based on instances, features, and depth. The results suggest domain adaptation is an efficient alternative when labeled data is scarce. Furthermore, we evaluate the need to handle imbalanced data, a characteristic of real-world problems like the data explored here. Based on promising accuracy results, we show that applying domain adaptation techniques in databases with little data, such as the Brazilian database, and without labeled data, is an efficient and low-cost alternative that can be useful in monitoring reservoirs in different regions.

Complete Paper #449

M&M: Multimodal-Multitask Model Integrating Audiovisual Cues in Cognitive Load Assessment

Long Nguyen-Phuoc^{1,2}, Renald Gaboriau², Dimitri Delacroix² and Laurent Navarro¹

¹ Mines Saint-Étienne, University of Lyon, University Jean Monnet, Inserm, U 1059 Sainbiose, Centre CIS, 42023 Saint-Étienne, France

² MJ Lab, MJ INNOV, 42000 Saint-Etienne, France

Keywords: Cognitive Load Assessment, Multimodal-Multitask Learning, Multihead Attention.

Abstract: This paper introduces the M&M model, a novel multimodal-multitask learning framework, applied to the AVCAffe dataset for cognitive load assessment (CLA). M&M uniquely integrates audiovisual cues through a dual-pathway architecture, featuring specialized streams for audio and video inputs. A key innovation lies in its cross-modality multihead attention mechanism, fusing the different modalities for synchronized multitasking. Another notable feature is the model's three specialized branches, each tailored to a specific cognitive load label, enabling nuanced, task-specific analysis. While it shows modest performance compared to the AVCAffe's single-task baseline, M&M demonstrates a promising framework for integrated multimodal processing. This work paves the way for future enhancements in multimodal-multitask learning systems, emphasizing the fusion of diverse data types for complex task handling.

Poster Presentations (Online) 3
10:30 - 11:30

VISIGRAPP
Room Online 2

Poster Presentations (Online) 3
10:30 - 11:30

VISAPP
Room Online 2

Complete Paper #114

Enhancing Object Detection Accuracy with Variational Autoencoders as a Filter in YOLO

Shubham Dubey¹, J. Satyanarayana² and C. Mohan¹

¹ Computer Science Department, Indian Institute of Technology Hyderabad, Hyderabad, India

² RCI-DRDO, India

Keywords: Object Detection, YOLO, False Positive, Variational Autoencoders.

Abstract: Object detection is an important task in computer vision systems, encompassing a diverse spectrum of applications, including but not limited to autonomous vehicular navigation and surveillance. Despite considerable advancements in object detection models such as YOLO, the issue of false positive detections remain a prevalent concern, thereby causing misclassifications and diminishing the reliability of these systems. This research endeavors to present an innovative methodology designed to augment object detection accuracy by incorporating Variational Autoencoders (VAEs) as a filtration mechanism within the YOLO framework. This integration seeks to rectify the issue of false positive detections, ultimately fostering a marked enhancement in detection precision and strengthening the overall dependability of object detection systems.

Complete Paper #152

Body-Part Enabled Wildlife Detection and Tracking in Video Sequences

Alberto Lee¹, Kofi Appiah¹ and Sze Kwok²

¹ Department of Computer Science, University of York, Deramore Lane, York, YO10 5GH, U.K.

² Phylo-Cognition Laboratory, Division of Natural and Applied Sciences, Data Science Research Center, Duke Kunshan University, Duke Institute for Brain Sciences, Kunshan, 610101, Jiangsu, China

Keywords: Animal Detection, Deep Convolutional Neural Networks, Real-Time Tracking, Data Augmentation.

Abstract: Tracking wild animals through videos presents a non-intrusive and cost-effective way of gathering scientific information key for conservation. State-of-the-art research has shown convolutional neural networks to be highly accurate, however, the application of this field on wild animal tracking has had relatively little interest. This is potentially due to the challenges of varying illumination, noisy backgrounds and camouflaged animals intrinsic to the problem. The aim of this work is to explore and apply state-of-the-art research to detect and track wild animals (specifically bears and primates, including their body parts) in video sequences in real-time. Due to obstructions such as foliage being prevalent in wild animal environments, body part tracking presents a solution to detecting animals when they are obstructed. Two deep convolutional neural networks (YOLOv4 and YOLOv4-Tiny) are trained to detect and track animals in their natural habitat. By using the knowledge that an animal is composed of body parts, the score of weakly predicted bounding is boosted from the relative distance of related body parts. For tracking, the K-Means

algorithm is used to locate the average position of each animal in frame. With the introduction of a body-part confidence boosting, the detection rate can be increased by approximately 2% for a weakly predicted class.

Complete Paper #177

Real-Time Detection and Mapping of Crowd Panic Emergencies

Ilias Lazarou, Anastasios Kesidis and Andreas Tsatsaris
Department of Surveying and Geoinformatics Engineering, University of West Attica, Athens, Greece

Keywords: Panic Detection, Biometrics, Machine Learning, Classification, Real-Time Data.

Abstract: We present a real-time system that uses machine learning and georeferenced biometric data from wearables and smartphones to detect and map crowd panic emergencies. Our system predicts stress levels, tracks stressed individuals, and introduces the CLOT parameter for better noise filtering and response speed. We also introduce the DEI metric to assess panic severity. The system creates dynamic areas showing the evolving panic situation in real-time. By integrating CLOT and DEI, emergency responders gain insights into crowd behaviour, enabling more effective responses to panic-induced crowd movements. This system enhances public safety by swiftly detecting, mapping, and assessing crowd panic emergencies.

Complete Paper #236

Automated Brain Lobe Segmentation and Feature Extraction from Multiple Sclerosis Lesions Using Deep Learning

Nada Haj Messaoud^{1,2}, Rim Ayari², Asma Ben Abdallah² and Mohamed Hedi Bedoui²

¹ *Faculty of Sciences of Monastir (FSM), University of Monastir, Monastir, Tunisia*

² *Medical Technology and Image Processing Laboratory, Faculty of medicine, University of Monastir, Monastir, Tunisia*

Keywords: Brain Lobes Segmentation, Deep Learning, Multiple Sclerosis Lesion, U-Net, Features Extraction.

Abstract: This study focuses on automating the segmentation of brain lobes in MRI images of Multiple Sclerosis (MS) lesions to extract crucial features for predicting disability levels. Extracting significant features from MRI images of MS lesions is indeed a complex task due to the variability in lesion characteristics and the detailed nature of MRI images. Furthermore, all these studies required continuous patient monitoring. Therefore, our contribution lies in proposing an approach for the automatic segmentation of brain lobes and the extraction of lesion features (number, size, location, etc.) to predict disability levels in MS patients. To achieve this, we introduced a model inspired by U-Net to perform the segmentation of different brain lobes, aiming to accurately locate the MS lesions. We utilized two private and public databases and achieved an average mean IoU score of 0.70, which can be considered encouraging. Following the segmentation phase, approximately 7200 features were extracted from the MRI scans of MS patients.

Complete Paper #266

A Mobile-Phone Pose Estimation for Gym-Exercise Form Correction

Matthew Turner¹, Kofi Appiah¹ and Sze Kwok²

¹ *Department of Computer Science, University of York, Deramore Lane, York, YO10 5GH, U.K.*

² *Phylo-Cognition Laboratory, Division of Natural and Applied Sciences, Data Science Research Center, Duke Kunshan University, Duke Institute for Brain Sciences, Kunshan, 610101, Jiangsu, China*

Keywords: Pose Estimation, Deep Learning, Mobile Application, Electronic Gym Instructor.

Abstract: People learn to perform exercises with good pose (or form) via research or instruction from an experienced individual such as a personal trainer, but 33.3% of injuries still occur due to incorrect form. It is known that the presence of a personal trainer causes a significant reduction in the rate that injuries occur. There are many possible reasons for this such as cost, scheduling limitations and desire to train alone. However, given that 91% of UK adults use a smartphone, a mobile APP could take on the role of a personal trainer. This paper presents a solution using machine learning and a novel proposed method of form anomaly detection to offer form corrections from live exercise video while only using the capabilities of a mobile device. Overall, the work in this paper is capable detecting incorrect exercise pose and offer valid corrections based on the detected anomalies. Experiments have been conducted on live video to judge the system performance in real-time.

Complete Paper #282

Curriculum for Crowd Counting: Is It Worthy?

Muhammad Asif Khan¹, Hamid Menouar¹ and Ridha Hamila²

¹ *Qatar Mobility Innovations Center, Qatar University, Doha, Qatar*

² *Department of Electrical Engineering, Qatar University, Doha, Qatar*

Keywords: Crowd Counting, Curriculum Learning, CNN, Density Estimation.

Abstract: Recent advances in deep learning techniques have achieved remarkable performance in several computer vision problems. A notably intuitive technique called Curriculum Learning (CL) has been introduced recently for training deep learning models. Surprisingly, curriculum learning achieves significantly improved results in some tasks but marginal or no improvement in others. Hence, there is still a debate about its adoption as a standard method to train supervised learning models. In this work, we investigate the impact of curriculum learning in crowd counting using the density estimation method. We performed detailed investigations by conducting 112 experiments using six different CL settings using eight different crowd models. Our experiments show that curriculum learning improves the model learning performance and shortens the convergence time.

Poster Presentations (Online) 3
10:30 - 11:30

VISIGRAPP
Room Online 1

Poster Presentations (Online) 3
10:30 - 11:30

GRAPP
Room Online 1

Complete Paper #72

Towards Generating 3D City Models with GAN and Computer Vision Methods

Sarun Poolkrajang and Anand Bhojan

School of Computing, National University of Singapore, Singapore

Keywords: Generative Adversarial Networks, Neural Networks, Computer Vision, City Generation.

Abstract: City generation for video games is a resource and time-consuming task. With the increasing popularity of open-world games, studies on building virtual environments have become increasingly important for the game research community and industry. The game development team must engage in urban planning, designate important locations, create population assets, integrate game design, and assemble these elements into a cohesive-looking city. Based on our limited knowledge and survey, we are the first to propose a holistic approach that integrates all features in generating a city, including the natural features surrounding it. We employ a generative adversarial network architecture to create a realistic layout of an entire city from scratch. Subsequently, we utilize classical computer vision techniques to post-process the layout into separate features. The chosen model is a simple Convolutional GAN, trained on a modest dataset of 2x2 km² snippets from over two thousand cities around the world. Although the method is somewhat constrained by the resolution of the images, the results indicate that it can serve as a solid foundation for building realistic 3D cities.

Poster Presentations (Online) 3
10:30 - 11:30

VISAPP
Room Online 1

Complete Paper #216

Direct 3D Body Measurement Estimation from Sparse Landmarks

David Bojanić¹, Kristijan Bartol², Tomislav Petković¹ and Tomislav Pribanić¹

¹ *Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, Zagreb, Croatia*

² *TU Dresden, 01062 Dresden, Dresden, Germany*

Keywords: Anthropometry, Body Measurements, 3D Landmarks.

Abstract: The current state-of-the-art 3D anthropometry extraction methods are either template-based or landmark-based. Template-based methods fit a statistical human body model to a 3D scan and extract complex features from the template to learn the body measurements. The fitting process is usually an optimization process, sensitive to its hyperparameters. Landmark-based methods use body proportion heuristics to estimate the landmark locations on the body in order to derive the measurements. Length measurements are derived as distances between landmarks, whereas circumference measurements are derived as cross-sections of the body and a plane defined at the desired landmark location. This makes it very susceptible to noise in the 3D scan data. To address these issues, we propose a simple learning method that infers the body measurements directly using the landmarks defined on the body. Our method avoids fitting a body model, extracting complex features, using heuristics, and handling noise in the data. We compare our method on the CAESAR dataset and show that

using a simple method coupled with sparse landmark data can compete with state-of-the-art methods. To take a step towards open-source 3D anthropometry, we make our code available at <https://github.com/DavidBoja/Landmarks2Anthropometry>.

Complete Paper #273

Conic Linear Units: Improved Model Fusion and Rotational-Symmetric Generative Model

Changqing Fu and Laurent Cohen

CEREMADE, University Paris Dauphine, PSL Research University, UMR CNRS 7534, Paris, 75016, France

Keywords: Generative AI, Deep Learning for Visual Understanding, Machine Learning Technologies for Vision.

Abstract: We introduce Conic Linear Unit (CoLU), a natural generalization of commonly used activation functions in neural networks. The common pointwise ReLU activation is a projection onto the positive cone and is permutation symmetric. We propose a nonlinearity that goes beyond this symmetry: CoLU is a skew projection onto a hypercone towards the cone's axis. Due to the nature of this projection, CoLU enforces symmetry in a neural network with width C from the finite-order permutation group $S(C)$ to the infinite-order rotation/reflection group $O(C-1)$, thus producing deep features that are motivated by the HSV color representation. Recent results on merging independent neural networks via permutation modulus can be relaxed and generalized to soft alignment modulo an optimal transport plan (Singh and Jaggi, 2020), which is useful in aligning models of different widths. CoLU aims to further alleviate the apparent deficiency of soft alignment. Our simulation indicates that CoLU outperforms existing generative models including Autoencoder and Latent Diffusion Model on small or large-scale image datasets. Additionally, CoLU does not increase the number of parameters and requires negligible additional computation overhead. The CoLU concept is quite general and can be plugged into various neural network architectures. Ablation studies on extensions to soft projections, general L_p cones, and the non-convex double-cone cases are briefly discussed.

Complete Paper #443

CL-FedFR: Curriculum Learning for Federated Face Recognition

Devilliers Dube¹, Çiğdem Erdem^{2,3} and Ömer Korçak²

¹ *Department of Electrical and Electronics Engineering, Boğaziçi University, Istanbul, Turkey*

² *Department of Computer Engineering, Faculty of Engineering, Marmara University, Istanbul, Turkey*

³ *Department of Electrical and Electronics Engineering, Özyeğin University, Istanbul, Turkey*

Keywords: Curriculum Learning, Deep Learning, Face Recognition, Federated Learning, Privacy.

Abstract: Face recognition (FR) has been significantly enhanced by the advent and continuous improvement of deep learning algorithms and accessibility of large datasets. However, privacy concerns raised by using and distributing face image datasets have emerged as a significant barrier to the deployment of centralized machine learning algorithms. Recently, federated learning (FL) has gained popularity since the private data at edge devices (clients) does not need to be shared to train a model. FL also continues to drive FR research toward decentralization. In this paper, we propose novel data-based and client-based curriculum learning (CL) approaches for federated FR intending

to improve the performance of generic and client-specific personalized models. The data-based curriculum utilizes head pose angles as the difficulty measure and feeds the images from “easy” to “difficult” during training, which resembles the way humans learn. Client-based curriculum chooses “easy clients” based on performance during the initial rounds of training and includes more “difficult clients” at later rounds. To the best of our knowledge, this is the first paper to explore CL for FR in a FL setting. We evaluate the proposed algorithm on MS-Celeb-1M and IJB-C datasets and the results show an improved performance when CL is utilized during training.

Complete Paper #452

RetailKLIP: Finetuning OpenCLIP Backbone Using Metric Learning on a Single GPU for Zero-Shot Retail Product Image Classification

Muktabh Srivastava
ParallelDots, Gurugram, India

Keywords: Packaged Grocery Goods, Image Recognition, Zero Shot, Vision Transformers.

Abstract: Retail product or packaged grocery goods images need to be classified in various computer vision applications like self-checkout stores, supply chain automation and retail execution evaluation. Previous works explore ways to finetune deep models for this purpose. But because of the fact that finetuning a large model or even linear layer for a pretrained backbone requires to run at least a few epochs of gradient descent for every new retail product added in classification range, frequent retrainings are needed in a real world scenario. In this work, we propose finetuning the vision encoder of a CLIP model in a way that its embeddings can be easily used for nearest neighbor based classification, while also getting accuracy close to or exceeding full finetuning. A nearest neighbor based classifier needs no incremental training for new products, thus saving resources and wait time.

Oral Presentations (Online) 7
11:30 - 13:00
Image and Video Understanding

VISAPP
Room Online 1

Complete Paper #304

Parts-Based Implicit 3D Face Modeling

Yajie Gu and Nick Pears

VGL Research Group, Department of Computer Science, University of York, YO10 5GH, U.K.

Keywords: Face Modeling, Deformation Network, Parts Corresponding Implicit Representations, Signed Distance Functions.

Abstract: Previous 3D face analysis has focussed on 3D facial identity, expression and pose disentanglement. However, the independent control of different facial parts and the ability to learn explainable parts-based latent shape embeddings for implicit surfaces remain as open problems. We propose a method for 3D face modeling that learns a continuous parts-based deformation field that maps the various semantic parts of a subject's face to a template. By swapping affine-mapped facial features among different individuals from predefined regions we achieve significant parts-based training data augmentation. Moreover, by sequentially morphing the surface points of these parts, we learn corresponding latent representations, shape deformation fields, and the signed

distance function of a template shape. This gives improved shape controllability and better interpretability of the face latent space, while retaining all of the known advantages of implicit surface modelling. Unlike previous works that generated new faces based on full-identity latent representations, our approach enables independent control of different facial parts, i.e. nose, mouth, eyes and also the remaining surface and yet generates new faces with high reconstruction quality. Evaluations demonstrate both facial expression and parts disentanglement, independent control of those facial parts, as well as state-of-the-art facial parts reconstruction, when evaluated on FaceScape and Headspace datasets.

Complete Paper #362

Occlusion-Robust and Efficient 6D Pose Estimation with Scene-Level Segmentation Refinement and 3D Partial-to-6D Full Point Cloud Transformation

Sukhan Lee¹, Soojin Lee¹ and Yongjun Yang²

¹ Department of Artificial Intelligence, Sungkyunkwan University, Suwon, Republic of Korea

² Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Republic of Korea

Keywords: Object 6D Pose, Panoptic Segmentation, Dual Associative Point Autoencoder, Point Cloud, Occluded Object.

Abstract: Accurate estimation of the 6D pose of objects is essential for 3D scene modeling, visual odometry, and map building, as well as robotic manipulation of objects. Recently, various end-to-end deep networks have been proposed for object 6D pose estimation with their accuracies reaching the level of conventional regimes but with much higher efficiency. Despite progress, the accurate yet efficient 6D pose estimation of highly occluded objects in a cluttered scene remains a challenge. In this study, we present an end-to-end deep network framework for 6D pose estimation with particular emphasis on highly occluded objects in a cluttered scene. The proposed framework integrates an occlusion-robust panoptic segmentation network performing scene-level segmentation refinement and a dual associative point autoencoder (AE) directly reconstructing the 6D full camera and object frame-based point clouds corresponding to a captured 3D partial point cloud through latent space association. We evaluated the proposed deep 6D pose estimation framework based on the standard benchmark dataset, LineMod-Occlusion (LMO), and obtained the top-tier performance in the current leaderboard, validating the effectiveness of the proposed approach in terms of efficiency and accuracy.

Oral Presentations (Online) 7
11:30 - 13:00

Computer Vision Theory and Applications

VISAPP
Room Online 2

Complete Paper #57

Event-Based Semantic-Aided Motion Segmentation

Chenao Jiang, Julien Moreau and Franck Davoine

Université de Technologie de Compiègne, CNRS, Heudiasyc (Heuristics and Diagnosis of Complex Systems), CS 60319 - 60203 Compiègne Cedex, France

Keywords: Event Cameras, Unconventional Vision, Semantic and Motion Segmentation.

Abstract: Event cameras are emerging visual sensors inspired

by biological systems. They capture intensity changes asynchronously with a temporal precision of up to μs , in contrast to traditional frame imaging techniques running at a fixed frequency of tens of Hz. However, effectively utilizing the data generated by these sensors requires the development of new algorithms and processing. In light of event cameras' significant advantages in capturing high-speed motion, researchers have turned their attention to event-based motion segmentation. Building upon (Mitrokhin et al., 2019) framework, we propose leveraging semantic segmentation enable the end-to-end network not only to segment moving objects from background motion, but also to achieve semantic segmentation of distinct moving objects. Remarkably, these capabilities are achieved while maintaining the network's low parameter count of 2.5M. To validate the effectiveness of our approach, we conduct experiments using the EVIMO dataset and the new and more challenging EVIMO2 dataset (Burner et al., 2022). The results demonstrate improvements attained by our method, showcasing its potential in event-based multi-objects motion segmentation.

Complete Paper #104

PanoTherm: Panoramic Thermal Imaging for Object Detection and Tracking

Thomas Kernbauer, Philipp Fleck and Clemens Arth
ICG, Graz University of Technology, Inffeldgasse 16/2, 8010 Graz, Austria

Keywords: Thermal Imaging, Camera Calibration, Object Detection.

Abstract: Visible-light cameras are used traditionally in object detection and tracking. Thermal imaging can equally be used for this purpose, however at the cost of additional calibration efforts, expenses, and limitations concerning the field of view. Still, thermal imaging is advantageous in various scenarios and basically the only plausible technology to apply in harsh outdoor environments, in which the use of standard RGB cameras is prohibitive due to low-light conditions or in complete darkness. While panoramic imaging using visible light cameras is becoming more popular for advanced photography or action recording, limited work has been done on developing panoramic thermal cameras. In this work, we present the first panoramic thermal camera targeting the constant 360° monitoring of the environment. We describe the calibration and stitching process in detail and demonstrate how to use the camera in a vehicular scenario. Finally, we give an example of the detection and tracking of objects and discuss the advantages and disadvantages of thermal imaging for this purpose.

Complete Paper #339

Camera Self-Calibration from Two Views with a Common Direction

Yingna Su^{1,2}, Xinnian Guo^{1,2} and Yang Shen³

¹ *College of Information Engineering, Suqian University, Suqian, China*

² *Suqian Key Laboratory of Visual Inspection and Intelligent Control, Suqian University, Suqian, China*

³ *Industrial Technology Research Institute, Suqian University, Suqian, China*

Keywords: Camera Self-Calibration, Gravity Direction, Homography Constraints, Principal Point Estimation.

Abstract: Camera calibration is crucial for enabling accurate and robust visual perception. This paper addresses the challenge of recovering intrinsic camera parameters from two views of a planar surface, that has received limited attention due to its inherent degeneracy. For cameras equipped with Inertial Measurement

Units (IMUs), such as those in smartphones and drones, the camera's y-axes can be aligned with the gravity direction, reducing the relative orientation to a one-degree-of-freedom (1-DoF). A key insight is the general orthogonality between the ground plane and the gravity direction. Leveraging this ground plane constraint, the paper introduces new homography-based minimal solutions for camera self-calibration with a known gravity direction. We derive 2.5- and 3.5-point camera self-calibration algorithms for points in the ground plane to enable simultaneous estimation of the camera's focal length and principal point. The paper demonstrates the practicality and efficiency of these algorithms and comparisons to existing state-of-the-art methods, confirming their reliability under various levels of noise and different camera configurations.

Session 7A
11:30 - 13:00
Object and Face Recognition

VISAPP
Room Space 3A

Complete Paper #163

EBA-PRNetCC: An Efficient Bridge Attention-Integration PoseResNet for Coordinate Classification in 2D Human Pose Estimation

Ali Zakir¹, Sartaj Salman¹, Gibran Benitez-Garcia¹ and Hiroki Takahashi^{1,2}

¹ *Department of Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan*

² *Artificial Intelligence Exploration/Meta-Networking Research Center, The University of Electro-Communications, Tokyo, Japan*

Keywords: 2D Human Pose Estimation, EBA-PRNetCC, MLP, EBA, COCO Dataset.

Abstract: In the current era, 2D Human Pose Estimation has emerged as an essential component in advanced Computer Vision tasks, particularly for understanding human behaviors. While challenges such as occlusion and unfavorable lighting conditions persist, the advent of deep learning has significantly strengthened the efficacy of 2D HPE. Yet, traditional 2D heatmap methodologies face quantization errors and demand complex post-processing. Addressing this, we introduce the EBA-PRNetCC model, an innovative coordinate classification approach for 2D HPE, emphasizing improved prediction accuracy and optimized model parameters. Our EBA-PRNetCC model employs a modified ResNet34 framework. A key feature is its head, which includes a dual-layer Multi-Layer Perceptron augmented by the Mish activation function. This design not only improves pose estimation precision but also minimizes model parameters. Integrating the Efficient Bridge Attention Net further enriches feature extraction, granting the model deep contextual insights. By enhancing pixel-level discretization, joint localization accuracy is improved. Comprehensive evaluations on the COCO dataset validate our model's superior accuracy and computational efficiency performance compared to prevailing 2D HPE techniques.

Complete Paper #328

Robust Long-Tailed Image Classification via Adversarial Feature Re-Calibration

Jinghao Zhang¹, Zhenhua Feng¹ and Yaochu Jin²

¹ *School of Computer Science and Electronic Engineering, University of Surrey, Guildford GU2 7XH, U.K.*

² *School of Engineering, Westlake University, Hangzhou 310030, China*

Keywords: Deep Learning, Image Classification, Adversarial Training, Long-Tailed Recognition.

Abstract: Long-tailed data distribution is a common issue in many practical learning-based approaches, causing Deep Neural Networks (DNNs) to under-fit minority classes. Although this biased problem has been extensively studied by the research community, the existing approaches mainly focus on the class-wise (inter-class) imbalance problem. In contrast, this paper considers both inter-class and intra-class data imbalance problems for network training. To this end, we present Adversarial Feature Re-calibration (AFR), a method that improves the standard accuracy of a trained deep network by adding adversarial perturbations to the majority samples of each class. To be specific, an adversarial attack model is fine-tuned to perturb the majority samples by injecting the features from their corresponding intra-class long-tailed minority samples. This procedure makes the dataset more evenly distributed from both the inter- and intra-class perspectives, thus encouraging DNNs to learn better representations. The experimental results obtained on CIFAR-100-LT demonstrate the effectiveness and superiority of the proposed AFR method over the state-of-the-art long-tailed learning methods.

Complete Paper #180

Privacy-Preserving Face Recognition in Hybrid Frequency-Color Domain

Dong Han^{1,2}, Yong Li¹ and Joachim Denzler²

¹ Huawei European Research Center, Riesstraße 25, 80992 München, Germany

² Computer Vision Group, Friedrich Schiller University Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany

Keywords: Privacy-Preserving, Face Recognition, Frequency Information, Color Information, Face Embedding Protection.

Abstract: Face recognition technology has been deployed in various real-life applications. The most sophisticated deep learning-based face recognition systems rely on training millions of face images through complex deep neural networks to achieve high accuracy. It is quite common for clients to upload face images to the service provider in order to access the model inference. However, the face image is a type of sensitive biometric attribute tied to the identity information of each user. Directly exposing the raw face image to the service provider poses a threat to the user's privacy. Current privacy-preserving approaches to face recognition focus on either concealing visual information on model input or protecting model output face embedding. The noticeable drop in recognition accuracy is a pitfall for most methods. This paper proposes a hybrid frequency-color fusion approach to reduce the input dimensionality of face recognition in the frequency domain. Moreover, sparse color information is also introduced to alleviate significant accuracy degradation after adding differential privacy noise. Besides, an identity-specific embedding mapping scheme is applied to protect original face embedding by enlarging the distance among identities. Lastly, secure multiparty computation is implemented for safely computing the embedding distance during model inference. The proposed method performs well on multiple widely used verification datasets. Moreover, it has around 2.6% to 4.2% higher accuracy than the state-of-the-art in the 1:N verification scenario.

Complete Paper #335

Exploring the Impact of Knowledge Graphs on Zero-Shot Visual Object State Classification

Filippos Gouidis^{1,2}, Konstantinos Papoutsakis¹, Theodore Patkos³, Antonis Argyros^{2,3} and Dimitris Plexousakis^{2,3}

¹ Department of Management, Science and Technology, Hellenic Mediterranean University, Agios Nikolaos, Greece

² Computer Science Department, University of Crete, Heraklion, Greece

³ Institute of Computer Science, Foundation for Research and Technology Hellas, Heraklion, Greece

Keywords: Visual Object State Classification, Zero-Shot Learning, Knowledge Graphs, Graph Neural Networks.

Abstract: In this work, we explore the potential of Knowledge Graphs (KGs) towards an effective Zero-Shot Learning (ZSL) approach for Object State Classification (OSC) in images. For this problem, the performance of traditional supervised learning methods is hindered mainly by data scarcity, as they attempt to encode the highly varying visual features of a multitude of combinations of object state and object type classes (e.g. open bottle, folded newspaper). The ZSL paradigm does indicate a promising alternative to enable the classification of object state classes by leveraging structured semantic descriptions acquired by external commonsense knowledge sources. We formulate an effective ZS-OSC scheme by employing a Transformer-based Graph Neural Network model and a pre-trained CNN classifier. We also investigate best practices for both the construction and integration of visually-grounded common-sense information based on KGs. An extensive experimental evaluation is reported using 4 related image datasets, 5 different knowledge repositories and 30 KGs that are constructed semi-automatically via querying known object state classes to retrieve contextual information at different node depths. The performance of vision-language models for ZS-OSC is also assessed. Overall, the obtained results suggest performance improvement for ZS-OSC models on all datasets, while both the size of a KG and the sources utilized for their construction are important for task performance.

Session 7B
11:30 - 13:00
Color and Texture Analyses

VISAPP
Room Space 3B

Complete Paper #38

Generative Texture Super-Resolution via Differential Rendering

Milena Bagdasarian¹, Peter Eisert^{1,2} and Anna Hilsmann¹

¹ Fraunhofer Heinrich-Hertz-Institute, Berlin, Germany

² Humboldt University of Berlin, Germany

Keywords: Texture Super-Resolution, Differentiable Rendering, GAN.

Abstract: Image super-resolution is a well-studied field that aims at generating high-resolution images from low-resolution inputs while preserving fine details and realistic features. Despite significant progress on regular images, inferring high-resolution textures of 3D models poses unique challenges. Due to the non-contiguous arrangement of texture patches, intended for wrapping around 3D meshes, applying conventional image super-resolution techniques to texture maps often results in artifacts and seams at texture discontinuities on the mesh. Additionally, obtaining ground truth data for texture super-resolution becomes highly complex due to

the labor intensive process of hand-crafting ground truth textures for each mesh. We propose a generative deep learning network for texture map super-resolution using a differentiable renderer and calibrated reference images. Combining a super-resolution generative adversarial network (GAN) with differentiable rendering, we guide our network towards learning realistic details and seamless texture map super-resolution without a high-resolution ground truth of the texture. Instead, we use high-resolution reference images. Through the differentiable rendering approach, we include model knowledge such as 3D meshes, projection matrices, and calibrated images to bridge the domain gap between 2D image super-resolution and texture map super-resolution. Our results show textures with fine structures and improved detail, which is especially of interest in virtual and augmented reality environments depicting humans.

Complete Paper #117

CAVC: Cosine Attention Video Colorization

Leandro Stival¹, Ricardo Torres^{2,3} and Helio Pedrini¹

¹ *Institute of Computing, University of Campinas, Av. Albert Einstein 1251, Campinas, SP, 13083-852, Brazil*

² *Wageningen Data Competence Center, Wageningen University and Research, Wageningen, The Netherlands*

³ *Norwegian University of Science and Technology, Larsgårdsvegen 2, 6009 Alesund, Norway*

Keywords: Video Colorization, Deep Learning, Cosine Similarity, Attention Mechanism.

Abstract: Video colorization is a challenging task, demanding deep learning models to employ diverse abstractions for a comprehensive grasp of the task, ultimately yielding high-quality results. Currently, in example-based colorization approaches, the use of attention processes and convolutional layers has proven to be the most effective method to produce good results. Following this way, in this paper we propose Cosine Attention Video Colorization (CAVC), which is an approach that uses a single attention head with shared weights to produce a refinement of the monochromatic frame, as well as the cosine similarity between this sample and the other channels present in the image. This entire process acts as a pre-processing of the data from our autoencoder, which performs a feature fusion with the latent space extracted from the referent frame, as well as with its histogram. This architecture was trained on the DAVIS, UVO and LDV datasets and achieved superior results compared to state-of-the-art models in terms of FID metric in all the datasets.

Complete Paper #300

Using Extended Light Sources for Relighting from a Small Number of Images

Toshiki Hirao, Ryo Kawahara and Takahiro Okabe

Department of Artificial Intelligence, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan

Keywords: Relighting, Display-Camera System, Specular Reflection, Extended Light Sources, End-to-End Optimization.

Abstract: Relighting real scenes/objects is useful for applications such as augmented reality and mixed reality. In general, relighting of glossy objects requires a large number of images, because specular reflection components are sensitive to light source positions/directions, and then the linear interpolation with sparse light sources does not work well. In this paper, we make use of not only point light sources but also extended light sources for efficiently capturing specular reflection components and achieve relighting from a small number of images. Specifically, we propose

a CNN-based method that simultaneously learns the illumination module (illumination condition), i.e. the linear combinations of the point light sources and the extended light sources under which a small number of input images are taken and the reconstruction module which recovers the images under arbitrary point light sources from the captured images in an end-to-end manner. We conduct a number of experiments using real images captured with a display-camera system, and confirm the effectiveness of our proposed method.

Complete Paper #416

Optimization and Learning Rate Influence on Breast Cancer Image Classification

Gleidson Barbosa¹, Larissa Moreira^{1,2}, Pedro Moises de Sousa¹, Rodrigo Moreira¹ and André Backes³

¹ *Institute of Exact and Technological Sciences, Federal University of Viçosa, Rio Paranaíba-MG, Brazil*

² *Faculty of Computing (FACOM), Federal University of Uberlândia, Uberlândia-MG, Brazil*

³ *Department of Computing, Federal University of São Carlos, São Carlos-SP, Brazil*

Keywords: Breast Cancer, CNN, Explainable, Influence of Factors, Classification.

Abstract: Breast cancer is a prevalent and challenging pathology, with significant mortality rates, affecting both women and men. Despite advancements in technology, such as Computer-Aided Diagnosis (CAD) and awareness campaigns, timely and accurate diagnosis remains a crucial issue. This study investigates the performance of Convolutional Neural Networks (CNNs) in predicting and supporting breast cancer diagnosis, considering BreakHis and Biglycan datasets. Through a factorial partial method, we measured the impact of optimization and learning rate factors on the prediction model accuracy. By measuring each factor's level of influence on the validation accuracy response variable, this paper brings valuable insights into the relevance analyses and CNN behavior. Furthermore, the study sheds light on the explainability of Artificial Intelligence (AI) through factorial partial performance evaluation design. Among the results, we determine which and how much the hyperparameters tuning influenced the performance of the models. The findings contribute to image-based medical diagnosis field, fostering the integration of computational and machine learning approaches to enhance breast cancer diagnosis and treatment.

Session 7C

11:30 - 13:00

Deep Learning for Visual Understanding

VISAPP

Room Space 3C

Complete Paper #40

Uncertainty-Based Detection of Adversarial Attacks in Semantic Segmentation

Kira Maag¹ and Asja Fischer²

¹ *Technical University of Berlin, Germany*

² *Ruhr University Bochum, Germany*

Keywords: Deep Learning, Semantic Segmentation, Adversarial Attacks, Detection.

Abstract: State-of-the-Art deep neural networks have proven to be highly powerful in a broad range of tasks, including semantic image segmentation. However, these networks are vulnerable against adversarial attacks, i.e., non-perceptible perturbations

added to the input image causing incorrect predictions, which is hazardous in safety-critical applications like automated driving. Adversarial examples and defense strategies are well studied for the image classification task, while there has been limited research in the context of semantic segmentation. First works however show that the segmentation outcome can be severely distorted by adversarial attacks. In this work, we introduce an uncertainty-based approach for the detection of adversarial attacks in semantic segmentation. We observe that uncertainty as for example captured by the entropy of the output distribution behaves differently on clean and perturbed images and leverage this property to distinguish between the two cases. Our method works in a light-weight and post-processing manner, i.e., we do not modify the model or need knowledge of the process used for generating adversarial examples. In a thorough empirical analysis, we demonstrate the ability of our approach to detect perturbed images across multiple types of adversarial attacks.

Complete Paper #175

Multi-View Inversion for 3D-aware Generative Adversarial Networks

Florian Barthel^{1,2}, Anna Hilsmann² and Peter Eisert^{1,2}

¹ Humboldt Universität zu Berlin, Berlin, Germany

² Fraunhofer HHI, Berlin, Germany

Keywords: 3D GAN Inversion, Multi-View Inversion, Multi-Latent Inversion.

Abstract: Current 3D GAN inversion methods for human heads typically use only one single frontal image to reconstruct the whole 3D head model. This leaves out meaningful information when multi-view data or dynamic videos are available. Our method builds on existing state-of-the-art 3D GAN inversion techniques to allow for consistent and simultaneous inversion of multiple views of the same subject. We employ a multi-latent extension to handle inconsistencies present in dynamic face videos to re-synthesize consistent 3D representations from the sequence. As our method uses additional information about the target subject, we observe significant enhancements in both geometric accuracy and image quality, particularly when rendering from wide viewing angles. Moreover, we demonstrate the editability of our inverted 3D renderings, which distinguishes them from NeRF-based scene reconstructions.

Complete Paper #197

Agrinet: A Hyperspectral Image Based Precise Crop Classification Model

Aditi Palit, Himanshu Dolekar and Kalidas Yeturu

Department of Computer Science and Engineering, Indian Institute of Technology Tirupati, India

Keywords: Agriculture, Crop Classification, Hyperspectral Imagery, Convolution Neural Network.

Abstract: Modern smart agriculture utilizes Unmanned Aerial Vehicles (UAVs) with hyperspectral cameras to enhance crop production to address the food security challenges. These cameras provide detailed crop information for type identification, disease detection, and nutrient assessment. However, processing Hyper Spectral Image (HSI) is complex due to challenges such as high inter-class similarity, intra-class variability, and overlapping spectral profiles. Thus, we introduce the Agrinet model, a convolutional neural network architecture, to handle complex hyperspectral image processing. Our novelty lies in the image pre-processing step of selecting suitable bands for better

classification. In tests, Agrinet achieved an impressive accuracy of 99.93% on the LongKou crop dataset, outperforming the existing methods in classification.

Complete Paper #354

Minimalist CNN for Medical Imaging Classification with Small Dataset: Does Size Really Matter and How?

Marie Économidès^{1,2} and Pascal Desbarats²

¹ GE Healthcare, F-78530 Buc, France

² Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

Keywords: Deep Learning, Medical Imaging, MRI, Classification, Convolutional Neural Network.

Abstract: Deep learning has become a key method in computer vision, and has seen an increase in the size of both the networks used and the databases. However, its application in medical imaging faces limitations due to the size of datasets, especially for larger networks. This article aims to answer two questions: How can we design a simple model without compromising classification performance, making training more efficient? And, how much data is needed for our network to learn effectively? The results show that we can find a minimalist CNN adapted to a dataset that gives results comparable to larger architectures. The minimalist CNN does not have a fixed architecture. Its architecture varies according to the dataset and various criteria such as overall performance, training stability, and visual interpretation of network predictions. We hope this work can serve as inspiration for others concerned with these challenges.

Session 7D

11:30 - 13:00

Image Formation, Acquisition Devices and Sensors

VISAPP

Room Space 4A

Complete Paper #65

Investigating Color Illusions from the Perspective of Computational Color Constancy

Oguzhan Ulucan, Diclehan Ulucan and Marc Ebner

Institut für Mathematik und Informatik, Universität Greifswald, Walther-Rathenau-Straße 47, 17489 Greifswald, Germany

Keywords: Computational Color Constancy, Color Assimilation Illusions, Color Illusion Perception.

Abstract: Color constancy and color illusion perception are two phenomena occurring in the human visual system, which can help us reveal unknown mechanisms of human perception. For decades computer vision scientists have developed numerous color constancy methods, which estimate the reflectance of the surface by discounting the illuminant. However, color illusions have not been analyzed in detail in the field of computational color constancy, which we find surprising since the relationship they share is significant and may let us design more robust systems. We argue that any model that can reproduce our sensation on color illusions should also be able to provide pixel-wise estimates of the light source. In other words, we suggest that the analysis of color illusions helps us to improve the performance of the existing global color constancy methods, and enable them to provide pixel-wise estimates for scenes illuminated by multiple light sources. In this study, we share the outcomes of our investigation in which we take several color constancy methods and modify them to reproduce the behavior of the human visual system on color

illusions. Also, we show that parameters purely extracted from illusions are able to improve the performance of color constancy methods. A noteworthy outcome is that our strategy based on the investigation of color illusions outperforms the state-of-the-art methods that are specifically designed to transform global color constancy algorithms into multi-illuminant algorithms.

Complete Paper #402

Analysis of Scattering Media by High-Frequency Polarized Light Projection Using Polarizing Projector

Aigo Ohno, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Nagoya, Japan

Keywords: Scattering Medium, Polarized Light, First-Order Scattering, Multiple Scattering, Specular Reflection, Diffuse Reflection.

Abstract: This paper proposes a special projection method called high-frequency polarized light projection using a polarizing projector to analyze scenes filled with scattering medium, and proposes a method to separate reflected lights and scattered lights by scattering medium in the observed image. In high-frequency polarized light projection, a high-frequency pattern is created by light with different polarization directions, projected onto a scattering medium, and the reflected light is observed. The light scattered by the medium and the reflected light from the object have different polarization properties, and we show that these two types of light can be easily separated.

Complete Paper #403

Probabilistic NeRF for 3D Shape Recovery in Scattered Medium

Yoshiki Ono, Fumihiko Sakaue and Jun Sato
Nagoya Insutitute of Technology, Nagoya, Japan

Keywords: NeRF (Neural Radiance Fields), Ray Tracing, Scattering Medium, Stochastic Gradient Descent.

Abstract: This research proposes a method for analyzing scene information including the characteristics of the medium by representing the space where objects and scattering media such as fog and smoke exist using the NeRF (Neural Radiance Fields) (Mildenhall et al., 2020) representation method of light ray fields. In this study, we focus on the fact that the behavior of rays inside a scattering medium can be expressed probabilistically, and show a method for rendering an image that changes in a probabilistic manner from only a single ray, rather than the entire scattering. By combining this method with a scene representation using the stochastic gradient descent method and a neural network, we show that it is possible to analyze scene information without generating images that directly render light scattering.

Complete Paper #411

Dense Light Field Imaging with Mixed Focus Camera

Masato Hirose, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Nagoya, Japan

Keywords: Light Field Camera, Variable Focal Lens, Coded Imaging.

Abstract: In this study, we propose a method for acquiring a dense light field in a single shot by taking advantage of the sparsity of the 4D light field (LF). Acquiring the LF with one camera is challenging task due to the amount of data. To acquire the LF efficiently, there are various methods like using micro-lens. However, with these methods, images are taken using a single image sensor, which improves directional resolution but reduces positional resolution. In our method, the focal length of the lens is varied, and the exposure is controlled on a pixel-by-pixel level when capturing a single image to obtain a mixed focus image, where each pixel is captured at a different focal length. Furthermore, by analyzing the captured image with an image generator that does not require prior learning, we show how to recover a LF image that is denser than the captured image. With our method, a high-density LF consisting of 5x5 images can be successfully reconstructed only from a single mixed-focus image taken under a simulated environment.

Session 7A
11:30 - 13:00
Geometry

GRAPP
Room Space 4B

Complete Paper #19

Polyline Simplification with Predefined Edge Directions by Mixed Integer Linear Programs

Steffen Goebbels and Jochen Rethmann

Institute for Pattern Recognition, Faculty of Electrical Engineering and Computer Science, Niederrhein University of Applied Sciences, Reinarzstr. 49, 47805 Krefeld, Germany

Keywords: Contour Simplification, Polyline, Mixed Integer Linear Program, 3D City Model.

Abstract: Mixed integer linear programs are presented that simplify polylines such that edges follow only some predefined directions from a given set. Under this constraint, solutions are computed that are closest to the given vertices, or only close to the original data, but with a minimum number of edges, or with a minimum length. The algorithms are applied to 3D building modeling from point clouds. Boundaries of roof facets (roof polygons) are simplified by considering directions of roof plane gradients and intersection lines between roof planes.

Complete Paper #404

Detection of Local Symmetry Polylines of Polygons Based on Sweeping Paradigm

Martin Safko^{1,2}, Luka Lukač³, Borut Žalik³ and Ivana Kolingerová²

¹ Faculty of Mathematics and Physics, Charles University, Ke Karlovu 3, 121 16 Praha 2, Czech Republic

² Department of Computer Science and Engineering, University of West Bohemia, Technická 8, 306 14 Plzeň, Czech Republic

³ Faculty of Electrical Engineering and Computer Science, University of Maribor, Koroška cesta 46, SI-2000 Maribor, Slovenia

Keywords: Symmetry, Polygon, Sweeping, Segmentation, Characterization, Computer Graphics.

Abstract: Symmetry is a fundamental property of many objects of interest allowing us to simplify computation or reduce complexity and is also a contributing factor of beauty for a human brain. In this work, we try to identify polylines satisfying the local reflection symmetry of polygons and show how to use them for shape characterization, segmentation or to find global approximate reflection symmetries. We describe an algorithm based on a sweep-line paradigm to efficiently compute the polylines by scanning through a polygon at various angles.

Complete Paper #447

Pseudo-Curvature of Fractal Curves for Geometric Control of Roughness

Mohamad Janbein, Christian Gentil, Céline Roudet and Clement Poull

Laboratoire d'Informatique de Bourgogne (LIB), Université de Bourgogne, 9 Av. Alain Savary, 21000 Dijon, France

Keywords: Curve, Fractal Geometry, Iterated Function System (IFS), Nowhere Differentiability, Tangent, Curvature.

Abstract: Fractal geometry is a valuable formalism for synthesizing and analyzing irregular curves to simulate non-smooth geometry or roughness. Understanding and controlling these geometries remains challenging because of the complexity of their shapes. This study focuses on the curvature of fractal curves defined from an Iterated Function System (a set of contractive operators). We introduce the Differential Characteristic Function (DCF), a new tool for characterizing and analyzing their differential behavior. We associate a family of DCF to the fixed point of each operator. For each dyadic point of the curve, there exist left and right families of DCF inducing left and right ranges of curvatures: the pseudo-curvatures. A set of illustrations shows the influence of these pseudo-curvatures on the geometry of fractal curves. We propose a first approach for applying our results to roughness generation and control.

Complete Paper #122

A Predictor for Triangle Mesh Compression Working in Tangent Space

Petr Vaněček, Filip Hácha and Libor Váša

Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 301 00 Plzeň, Czech Republic

Keywords: Geometry Compression, Parallelogram Prediction, Tangent Space.

Abstract: Triangle mesh compression has been a popular research topic for decades. Since a plethora of algorithms has been presented, it is becoming increasingly difficult to come up with significant performance improvements. Some of the recent advances in compression efficiency come at the cost of rather steep implementation and/or computational expense, which has profound consequences on their practicality. Ultimately it becomes increasingly difficult to come up with improvements that are reasonably easy to implement and do not harm the computational efficiency of the compression/decompression procedure. In this paper, we analyze a combination of two previously known techniques, namely using the local coordinates for expressing compression residuals and weighted parallelogram prediction, which were not previously investigated together. We report that such approach outperforms industry standard Draco on a large set of test meshes in terms of rate/distortion ratio, while retaining beneficial properties such as simplicity and computational efficiency.

Session 7A
11:30 - 13:00
Visualization Short Papers

IVAPP
Room Space 4D

Complete Paper #97

Increasing User Engagement with a Tracking App Through Data Visualizations

Daniela Nickmann and Victor Oliveira

St. Pölten University of Applied Sciences, St. Pölten, Austria

Keywords: Infovis, Engagement, Mobile Applications, Food Waste.

Abstract: According to the United Nations, 17 percent of global food production is wasted, causing economic losses and significant environmental impact. Digital solutions like food storage management apps can raise awareness and combat this issue. However, their effectiveness relies on consistent user engagement. Therefore, this paper proposes and evaluates data visualization designs to enhance user engagement in mobile applications for tracking food waste. The study involves three steps: discussing the domain situation supported by relevant literature, outlining the process of creating two sets of four data visualization designs and conducting quantitative user surveys to validate the designs. The first experiment assesses user experience, while the second determines user engagement. Results indicate a preference for design approach two (Chart Set B), which also provides more accuracy and higher user engagement when the design aligns with users' sustainability interests. These findings emphasize the potential of engaging data visualizations to curb food waste and contribute to a more sustainable future.

Complete Paper #263

Visualization of Swedish News Articles: A Design Study

Kostiantyn Kucher¹, Nellie Engström¹, Wilma Axelsson¹,
Berkant Savas^{1,2} and Andreas Kerren^{1,3}

¹ Department of Science and Technology, Linköping University,
Norrköping, Sweden

² iMatrics AB, Linköping, Sweden

³ Department of Computer Science and Media Technology, Linnaeus
University, Växjö, Sweden

Keywords: Information Visualization, Text Visualization, Natural Language Processing, News, Editorial Media, Swedish Language, Journalism.

Abstract: The amount of available text data has increased rapidly in the past years, making it difficult for many users to find relevant information. To solve this, natural language processing (NLP) and text visualization methods have been developed, however, they typically focus on English texts only, while the support for low-resource languages is limited. The aim of this design study was to implement a visualization prototype for exploring a large number of Swedish news articles (made available by industrial collaborators), including the temporal and relational data aspects. Sketches of three visual representations were designed and evaluated through user tests involving both our collaborators and end-users (journalists). Next, an NLP pipeline was designed in order to support dynamic and hierarchical topic modeling. The final part of the study resulted in an interactive visualization prototype that uses a variation of area charts to represent topic evolution. The prototype was evaluated through an internal case study and user tests with two groups of participants with the background in journalism and NLP. The evaluation results reveal the participants' preference for the representation focusing on top topics rather than the topic hierarchy, while suggestions for future work relevant for Swedish text data visualization are also provided.

Complete Paper #324

Visualizing Group Structure in Compound Graphs: The Current State, Lessons Learned, and Outstanding Opportunities

Henry Ehlers¹, Diana Marin², Hsiang-Yun Wu³ and
Renata Raidou¹

¹ Visualization Group, Institute of Visual Computing and Human-Centered Technology, TU Wien, Favoritenstr. 9-11 / E193-02, A-1040 Vienna, Austria

² Rendering and Modeling Group, Institute of Visual Computing and Human-Centered Technology, TU Wien, Favoritenstr. 9-11 / E193-02, A-1040 Vienna, Austria

³ St. Pölten University of Applied Sciences, Department of Media and Digital Technologies, Campus-Platz 1, St. Pölten, A-3100, Austria

Keywords: Compound Graph Visualization, Literature Survey, Group Structure Visualization.

Abstract: Compound graphs are common across domains, from social science to biochemical pathway studies, and their visualization is important to both their exploration and analysis. However, effectively visualizing a compound graph's topology and group structure requires careful consideration, as evident by the many different approaches to this particular problem. To better understand the current advancements in compound graph visualization, we have consolidated and streamlined existing surveys' taxonomies. More specifically, we aim to disentangle the visual relationship between graph topology and group structure from the visual encoding used to visualize its group structure in

order to identify interesting gaps in the literature. In so doing, we are able to enumerate a number of lessons learned and gain a better understanding of the outstanding research opportunities and practical implications across domains.

Complete Paper #346

A Review on Data Terminology in Visual Analytics Tools

Johanna Schmidt and Milena Vuckovic

VRVis Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH,
Vienna, Austria

Keywords: Data Visualization, Data Analysis, Data Engineering.

Abstract: Recent advances in visualization research and related technologies gave rise to several Visual Analytics tools capable of supporting many aspects of a typical data analytics pipeline. More specifically, these tools are showing a promise of a feature-rich environment offering multiple built-in options related to data loading and data management, which are essential initial steps for any data exploratory challenge. In this paper, we review the terms and terminology used to describe data, data parts, and data handling tasks in eighteen commonly used Visual Analytics applications. Throughout this review, we have observed a general lack of standardization of terminology used to describe all related features. Such lack of standardization may affect the overall application potential and increase the complexity when combining different tools, thus creating a user dependency on a specific solution and impeding knowledge exchange.

Keynote Lecture
14:15 - 15:15

VISIGRAPP
Room Plenary

Virtual Reality in Mental Health: A Self-Counselling Approach

Mel Slater

University of Barcelona, Spain

Abstract: Extensive research into virtual reality and its applications started in the 1990s. To date there have been over 1,700,000 scientific publications and patents that mention the terms "virtual reality", and about 5% of these specifically include the term "mental health". Early work concentrated on specific phobias such as fear of heights and flying, and then expanded into social phobia and general anxiety disorders, and more complex syndromes such as depression. It has most commonly been used in research in the context of exposure and cognitive behavioural therapy. The evidence suggests that the results are at least as good as conventional in vivo treatment. VR has also been used in the study and treatment of psychotic illnesses such as paranoia. In this talk I will review research in this field and move on to discuss a particular paradigm that makes use of VR for self-counselling, including its role in helping people to overcome obesity.

Session 8A
15:30 - 17:00
Color and Texture Analyses

VISAPP
Room Space 3A

Complete Paper #360

Conditional Vector Graphics Generation for Music Cover Images

Ivan Jarsky, Valeria Efimova, Ilya Bizyaev and Andrey Filchenkov

ITMO University, Kronverksky Pr. 49, St. Petersburg, Russia

Keywords: GAN, Image Generation, Vector Graphics.

Abstract: Generative Adversarial Networks (GAN) have motivated a rapid growth of the domain of computer image synthesis. As almost all the existing image synthesis algorithms consider an image as a pixel matrix, the high-resolution image synthesis is complicated. A good alternative can be vector images. However, they belong to the highly sophisticated parametric space, which is a restriction for solving the task of synthesizing vector graphics by GANs. In this paper, we consider a specific application domain that softens this restriction dramatically allowing the usage of vector image synthesis. Music cover images should meet the requirements of Internet streaming services and printing standards, which imply high resolution of graphic materials without any additional requirements on the content of such images. Existing music cover image generation services do not analyze tracks themselves; however, some services mostly consider only genre tags. To generate music covers as vector images that reflect the music and consist of simple geometric objects, we suggest a GAN-based algorithm called CoverGAN. The assessment of resulting images is based on their correspondence to the music compared with AttnGAN and DALL-E text-to-image generation according to title or lyrics. Moreover, the significance of the patterns found by CoverGAN has been evaluated in terms of the correspondence of the generated cover images to the musical tracks. Listeners evaluate the music covers generated by the proposed algorithm as quite satisfactory and corresponding to the tracks. Music cover images generation code and demo are available at <https://github.com/lzhanVarsky/CoverGAN>.

Complete Paper #340

Neural Style Transfer for Vector Graphics

Ivan Jarsky¹, Valeria Efimova¹, Artyom Chebykin², Viacheslav Shalamov¹ and Andrey Filchenkov¹

¹ *ITMO University, Kronverksky Pr. 49, St. Petersburg, Russia*

² *SUAI, Bolshaya Morskaya Street 67A, St. Petersburg, Russia*

Keywords: Vector Graphics, Computer Vision, Neural Style Transfer, DiffVG.

Abstract: Neural style transfer draws researchers' attention, but the interest focuses on bitmap images. Various models have been developed for bitmap image generation both online and offline with arbitrary and pre-trained styles. However, the style transfer between vector images has not almost been considered. Our research shows that applying standard content and style losses insignificantly changes the vector image drawing style because the structure of vector primitives differs a lot from pixels. To handle this problem, we introduce new loss functions. We also develop a new method based on differentiable rasterization that uses these loss functions and can change the color and shape parameters of the content image corresponding to the drawing of the style image. Qualitative experiments demonstrate the effectiveness of the

proposed VectorNST method compared with the state-of-the-art neural style transfer approaches for bitmap images and the only existing approach for stylizing vector images, DiffVG. Although the proposed model does not achieve the quality and smoothness of style transfer between bitmap images, we consider our work an important early step in this area. VectorNST code and demo service are available at <https://github.com/lzhanVarsky/VectorNST>.

Complete Paper #378

Reliability and Stability of Mean Opinion Score for Image Aesthetic Quality Assessment Obtained Through Crowdsourcing

Egor Ershov¹, Artyom Panshin¹, Ivan Ermakov¹, Nikola Banić², Alex Savchik³ and Simone Bianco⁴

¹ *Institute for Information Transmission Problems, Russian Academy of Sciences, 119991 Moscow, Russia*

² *Gideon Brothers, 10000 Zagreb, Croatia*

³ *ACMetric, Netherlands*

⁴ *University of Milano-Bicocca, 20126 Milan, Italy*

Keywords: Image Quality, Pairwise Comparison, Statistics, Stability, Aesthetics, Computational Aesthetics, Crowdsourcing.

Abstract: Image quality assessment (IQA) is widely used to evaluate the results of image processing methods. While in recent years the development of objective IQA metrics has seen much progress, there are still many tasks where subjective IQA is significantly more preferred. Using subjective IQA has become even more attractive ever since crowdsourcing platforms such as Amazon Mechanical Turk and Toloka have become available. However, for some specific image processing tasks, there are still some questions related to subjective IQA that have not been solved in a satisfactory way. An example of such a task is the evaluation of image rendering styles where, unlike in the case of distortions, none of the evaluated styles is to be objectively regarded as a priori better or worse. The questions that have not been properly answered up until now are whether the scores for such a task obtained through crowdsourced subjective IQA are reliable and whether they remain stable, i.e., similar if the evaluation is repeated over time. To answer these questions, in this paper first several images and styles are selected and defined, they are then evaluated by using crowdsourced subjective IQA on the Toloka platform, and the obtained scores are numerically analyzed. Experimental results confirm the reliability and stability of the crowdsourced subjective IQA for the problem in question. The experimental data is available at <https://zenodo.org/records/10458531>.

Session 8B
15:30 - 17:00
3D Deep Learning

VISAPP
Room Space 3B

Complete Paper #427

Fooling Neural Networks for Motion Forecasting via Adversarial Attacks

Edgar Medina and Leyong Loh

QualityMinds GmbH, Germany

Keywords: Adversarial Attacks, Human Motion Prediction, 3D Deep Learning, Motion Analyses.

Abstract: Human motion prediction is still an open problem, which is extremely important for autonomous driving and safety

applications. Although there are great advances in this area, the widely studied topic of adversarial attacks has not been applied to multi-regression models such as GCNs and MLP-based architectures in human motion prediction. This work intends to reduce this gap using extensive quantitative and qualitative experiments in state-of-the-art architectures similar to the initial stages of adversarial attacks in image classification. The results suggest that models are susceptible to attacks even on low levels of perturbation. We also show experiments with 3D transformations that affect the model performance, in particular, we show that most models are sensitive to simple rotations and translations which do not alter joint distances. We conclude that similar to earlier CNN models, motion forecasting tasks are susceptible to small perturbations and simple 3D transformations.

Complete Paper #137

Analysis of Point Cloud Domain Gap Effects for 3D Object Detection Evaluation

Aitor Iglesias^{1,2}, Mikel García^{1,2}, Nerea Aranjuelo¹, Ignacio Arganda-Carreras^{2,3,4,5} and Marcos Nieto¹

¹ *Fundación Vicomtech, Connected and Cooperative Automated Systems, Spain*

² *University of the Basque Country (UPV/EHU), Donostia - San Sebastian, Spain*

³ *IKERBASQUE, Basque Foundation for Science, Bilbao, Spain*

⁴ *Donostia International Physics Center (DIPC), Donostia - San Sebastian, Spain*

⁵ *Biofísica Institute, Leioa, Spain*

Keywords: Point Clouds, Deep Learning, Domain Gap, Object Detection, Simulation.

Abstract: The development of autonomous driving systems heavily relies on high-quality LiDAR data, which is essential for robust object detection and scene understanding. Nevertheless, obtaining a substantial amount of such data for effective training and evaluation of autonomous driving algorithms is a major challenge. To overcome this limitation, recent studies are taking advantage of advancements in realistic simulation engines, such as CARLA, which have provided a breakthrough in generating synthetic LiDAR data that closely resembles real-world scenarios. However, these data are far from being identical to real data. In this study, we address the domain gap between real LiDAR data and synthetic data. We train deep-learning models for object detection using real data. Then, those models are rigorously evaluated using synthetic data generated in CARLA. By quantifying the discrepancies between the model's performance on real and synthetic data, the present study shows that there is indeed a domain gap between the two types of data and does not affect equal to different model architectures. Finally, we propose a method for synthetic data processing to reduce this domain gap. This research contributes to enhancing the use of synthetic data for autonomous driving systems.

Complete Paper #148

Incorporating Temporal Information into 3D Hand Pose Estimation Using Scene Flow

Niklas Hermes^{1,2}, Alexander Bigalke¹ and Mattias Heinrich¹

¹ *Institute of Medical Informatics, Lübeck University, Germany*

² *Gestigon GmbH, Lübeck, Germany*

Keywords: Deep Learning, Hand Pose Estimation, 3D Point Clouds, Tracking, Temporal Information, Scene Flow.

Abstract: In this paper we present a novel approach that uses 3D point cloud sequences to integrate temporal information and spatial constraints into existing 3D hand pose estimation methods in order to establish an improved prediction of 3D hand poses. We utilize scene flow to match correspondences between two point sets and present a method that optimizes and harnesses existing scene flow networks for the application of 3D hand pose estimation. For increased generalizability, we propose a module that learns to recognize spatial hand pose associations to transform existing poses into a low-dimensional pose space. In a comprehensive evaluation on the public dataset NYU, we show the benefits of our individual modules and provide insights into the generalization capabilities and the behaviour of our method with noisy data. Furthermore, we demonstrate that our method reduces the error of existing state-of-the-art 3D hand pose estimation methods by up to 7.6%. With a speed of over 40 fps our method is real-time capable and can be integrated into existing 3D hand pose estimation methods with little computational overhead.

Complete Paper #383

Detecting Anomalous 3D Point Clouds Using Pre-Trained Feature Extractors

Dario Mantegazza and Alessandro Giusti

Dalle Molle Institute for Artificial Intelligence (IDSIA), USI-SUPSI, Lugano, Switzerland

Keywords: Anomaly Detection, 3D Point Clouds, Deep Learning.

Abstract: In this paper we explore the status of the research effort for the task of 3D visual anomaly detection; in particular, we investigate whether it is possible to find anomalies on 3D point clouds using off-the-shelf feature extractors, similar to what is already feasible on images, without the requirement of an ad-hoc one. Our work uses a model composed of two parts: a feature extraction module and an anomaly detection head. The latter is fixed and works on the embeddings from the feature extraction module. Using the MVTEC-3D dataset, we contribute a comparison between a 3D point cloud features extractor, a 2D image features extractor, a combination of the two, and three baselines. We also compare our work with other models on the dataset's DETECTION-AUROC benchmark. The experiment results demonstrate that, while our proposed approach surpasses the baselines and some other approaches, our best-performing model cannot beat purposely developed ones. We conclude that a combination of dataset size and 3D data complexity is the culprit to a lack of off-the-shelf feature extractors for solving complex 3D vision tasks.

Session 8C
15:30 - 17:00
Domain Adaptation

VISAPP
Room Space 3C

Complete Paper #110

Reducing Bias in Pre-Trained Models by Tuning While Penalizing Change

Niklas Penzel, Gideon Stein and Joachim Denzler

Computer Vision Group, Friedrich Schiller University, Jena, Germany

Keywords: Debiasing, Change Penalization, Early Stopping, Fine-Tuning, Domain Adaptation.

Abstract: Deep models trained on large amounts of data often incorporate implicit biases present during training time. If later such a bias is discovered during inference or deployment, it is

often necessary to acquire new data and retrain the model. This behavior is especially problematic in critical areas such as autonomous driving or medical decision-making. In these scenarios, new data is often expensive and hard to come by. In this work, we present a method based on change penalization that takes a pre-trained model and adapts the weights to mitigate a previously detected bias. We achieve this by tuning a zero-initialized copy of a frozen pre-trained network. Our method needs very few, in extreme cases only a single, examples that contradict the bias to increase performance. Additionally, we propose an early stopping criterion to modify baselines and reduce overfitting. We evaluate our approach on a well-known bias in skin lesion classification and three other datasets from the domain shift literature. We find that our approach works especially well with very few images. Simple fine-tuning combined with our early stopping also leads to performance benefits for a larger number of tuning samples.

Complete Paper #138

Pre-Training and Fine-Tuning Attention Based Encoder Decoder Improves Sea Surface Height Multi-Variate Inpainting

Théo Archambault^{1,2}, Arthur Filoche^{1,3}, Anastase Charantonis^{4,5} and Dominique Béréziat¹

¹ Sorbonne Université, CNRS, LIP6, Paris, France

² Sorbonne Université, LOCEAN, Paris, France

³ UWA, Perth, Australia

⁴ ENSIIE, LaMME, Evry, France

⁵ Inria, Paris, France

Keywords: Image Inverse Problems, Deep Neural Network, Spatiotemporal Inpainting, Multi-Variate Observations, Transfer Learning, Satellite Remote Sensing.

Abstract: The ocean is observed through satellites measuring physical data of various natures. Among them, Sea Surface Height (SSH) and Sea Surface Temperature (SST) are physically linked data involving different remote sensing technologies and therefore different image inverse problems. In this work, we propose to use an Attention-based Encoder-Decoder to perform the inpainting of the SSH, using the SST as contextual information. We propose to pre-train this neural network on a realistic twin experiment of the observing system and to fine-tune it in an unsupervised manner on real-world observations. We show the interest of this strategy by comparing it to existing methods. Our training methodology achieves state-of-the-art performances, and we report a decrease of 25% in error compared to the most widely used interpolations product.

Complete Paper #22

Learning End-to-End Deep Learning Based Image Signal Processing Pipeline Using a Few-Shot Domain Adaptation

Georgy Perevozchikov^{1,2} and Egor Ershov²

¹ Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia

² Institute for Information Transmission Problems, Moscow, Russia

Keywords: Computational Photography, Image Signal Processing Pipeline, Domain Adaptation, Image Processing.

Abstract: Nowadays the quality of mobile phone cameras plays one of the most important roles in modern smartphones, as a result, more attention is being paid to the camera Image Signal

Processing (ISP) pipeline. The current goal of the scientific community is to develop a neural-based end-to-end pipeline to remove the expensive and exhausting process of classical ISP tuning for each next device. The main drawback of the neural-based approach is the necessity of preparing large-scale datasets each time a new smartphone is designed. In this paper, we address this problem and propose a new method for few-shot domain adaptation of the existing neural ISP to a new domain. We show that it is sufficient to have 10 labeled images of the target domain to achieve state-of-the-art performance on the real camera benchmark datasets. We also provide a comparative analysis of our proposed approach with other existing ISP domain adaptation methods and show that our approach allows us to achieve better results. Our proposed method exhibits notably comparable performance, with only a marginal 2% drop in performance compared to the learned from scratch in the whole dataset baseline. We believe that this solution will significantly reduce the cost of neural-based ISP production for each new device.

Complete Paper #142

How Quality Affects Deep Neural Networks in Fine-Grained Image Classification

Joseph Smith¹, Zheming Zuo¹, Jonathan Stonehouse² and Boguslaw Obara¹

¹ School of Computing, Newcastle University, Newcastle upon Tyne, U.K.

² Procter and Gamble, Reading, U.K.

Keywords: Image Quality Assessment, Neural Networks, Fine-Grained Image Classification, Mobile Imaging.

Abstract: In this paper, we propose a No-Reference Image Quality Assessment (NRIQA) guided cut-off point selection (CPS) strategy to enhance the performance of a fine-grained classification system. Scores given by existing NRIQA methods on the same image may vary and not be as independent of natural image augmentations as expected, which weakens their connection and explainability to fine-grained image classification. Taking the three most commonly adopted image augmentation configurations – cropping, rotating, and blurring – as the entry point, we formulate a two-step mechanism for selecting the most discriminative subset from a given image dataset by considering both the confidence of model predictions and the density distribution of image qualities over several NRIQA methods. Concretely, the cut-off points yielded by those methods are aggregated via majority voting to inform the process of image subset selection. The efficacy and efficiency of such a mechanism have been confirmed by comparing the models being trained on high-quality images against a combination of high- and low-quality ones, with a range of 0.7% to 4.2% improvement on a commercial product dataset in terms of mean accuracy through four deep neural classifiers. The robustness of the mechanism has been proven by the observations that all the selected high-quality images can work jointly with 70% low-quality images with 1.3% of classification precision sacrificed when using ResNet34 in an ablation study.

Session 8D
15:30 - 17:00
Segmentation and Grouping

VISAPP
Room Space 4A

Complete Paper #100

Efficiency Optimization Strategies for Point Transformer Networks

Jannis Unkrig and Markus Friedrich

Department of Computer Science and Mathematics, Munich University of Applied Sciences, Munich, Germany

Keywords: 3D Point Cloud Processing, 3D Computer Vision, Deep Learning, Transformer Architecture.

Abstract: The Point Transformer, and especially its successor Point Transformer V2, are among the state-of-the-art architectures for point cloud processing in terms of accuracy. However, like many other point cloud processing architectures, they suffer from the inherently irregular structure of point clouds, which makes efficient processing computationally expensive. Common workarounds include reducing the point cloud density, or cropping out partitions, processing them sequentially, and then stitching them back together. However, those approaches inherently limit the architecture by either providing less detail or less context. This work provides strategies that directly address efficiency bottlenecks in the Point Transformer architecture, and therefore allows processing larger point clouds in a single feed-forward operation. Specifically, we propose using uniform point cloud sizes in all stages of the architecture, a k-D tree-based k-nearest neighbor search algorithm that is not only efficient on large point clouds, but also generates intermediate results that can be reused for downsampling, and a technique for normalizing local densities which improves overall accuracy. Furthermore, our architecture is simpler to implement and does not require custom CUDA kernels to run efficiently.

Complete Paper #448

FingerSeg: Highly-Efficient Dual-Resolution Architecture for Precise Finger-Level Semantic Segmentation

Gibran Benitez-Garcia¹ and Hiroki Takahashi^{1,2,3}

¹ *Graduate School of Informatics and Engineering, The University of Electro-Communications, Japan*

² *Artificial Intelligence eXploration Research Center (AIX), The University of Electro-Communications, Japan*

³ *Meta-Networking Research Center (MEET), The University of Electro-Communications, Japan*

Keywords: Semantic Segmentation, Finger Segmentation, DDRNet, Real-Time CNN, IPN-Hand Dataset.

Abstract: Semantic segmentation at the finger level poses unique challenges, including the limited pixel representation of some classes and the complex interdependency of the hand anatomy. In this paper, we propose FingerSeg, a novel architecture inspired by Deep Dual-Resolution Networks, specifically adapted to address the nuances of finger-level hand semantic segmentation. To this end, we introduce three modules: Enhanced Bilateral Fusion (EBF), which refines low- and high-resolution feature fusion via attention mechanisms; Multi-Attention Module (MAM), designed to augment high-level features with a composite of channel, spatial, orientational, and categorical attention; and Asymmetric Dilated Up-sampling (ADU), which combines standard and asymmetric atrous convolutions to capture rich contextual information for

pixel-level classification. To properly evaluate our proposal, we introduce IPN-finger, a subset of the IPN-Hand dataset, manually annotated pixel-wise for 13 finger-related classes. Our extensive empirical analysis, including evaluations of the synthetic RHD dataset against current state-of-the-art methods, demonstrates that our proposal achieves top results. FingerSeg reaches 73.8 and 71.1 mIoU on the IPN-Finger and RHD datasets, respectively, while maintaining an efficient computational cost of about 7 GFLOPs and 6 million parameters at VGA resolution. The dataset, source code, and a demo of FingerSeg will be available upon the publication of this paper.

Complete Paper #89

Automated Generation of Instance Segmentation Labels for Traffic Surveillance Models

D. Scholte¹, T. T. G. Urselmann², M. H. Zwemer^{2,1}, E. Bondarev² and P. H. N. de With²

¹ *ViNotion BV, Eindhoven, The Netherlands*

² *Department of Electrical Engineering, Eindhoven University, Eindhoven, The Netherlands*

Keywords: Instance Segmentation, Object Detection, Real-Time Processing, Computer Vision, Traffic Surveillance.

Abstract: This paper focuses on instance segmentation and object detection for real-time traffic surveillance applications. Although instance segmentation is currently a hot topic in literature, no suitable dataset for traffic surveillance applications is publicly available and limited work is available with real-time performance. A custom proprietary dataset is available for training, but it contains only bounding-box annotations and lacks segmentation annotations. The paper explores methods for automated generation of instance segmentation labels for custom datasets that can be utilized to finetune state-of-the-art segmentation models to specific application domains. Real-time performance is obtained by adopting the recent YOLACT instance segmentation with the YOLOv7 backbone. Nevertheless, it requires modification of the loss function and an implementation of ground-truth matching to overcome handling imperfect instance labels in custom datasets. Experiments show that it is possible to achieve a high instance segmentation performance using a semi-automatically generated dataset, especially when using the Segment Anything Model for generating the labels.

Session 8A
15:30 - 17:00
Modeling and Natural Environments

GRAPP
Room Space 4B

Complete Paper #219

Between Gaming and Microclimate Simulations: Temperature Estimation of an Urban Area

Eva Strauss and Dimitri Bulatov

Fraunhofer Institute for Optronics, System Technologies and Image Exploitation (IOSB), Gutleuthausstrasse 1, 76275 Ettlingen, Germany

Keywords: Urban Heat Island, Thermal Simulation, Digital Twin, 3D City Model, Thermal Remote Sensing.

Abstract: With the rising awareness and interest from researchers, local authorities, and industry in the urban heat island effect, thermal remote sensing data is needed as it allows for identification, tracking, or analysis of land surface temperatures. Yet, the accessibility of appropriate thermal data in both the spatial and temporal domain states an inhibiting factor. Whilst

thermal satellite data suffers from both low spatial and temporal resolution, airborne imagery might enable adequate resolutions, however, is not acquired without time and cost consumption. One way to overcome this drawback is the generation of synthetic data, which comprises the simulation of surface temperatures. These rather simplified simulations are either quite fast, as desired in gaming applications, however, highly inaccurate, or rather complex, holistic, time-consuming and computationally intensive, like applied in urban microclimate considerations. In this paper, we present an in-between approach towards the estimation of urban surface temperatures that aims to fill this gap between holistic microclimate simulations and climate maps.

Complete Paper #231

TreeSpecies-PC2DT: Automated Tree Species Modeling from Point Clouds to Digital Twins

Like Gobeawan¹, Xuan Liu¹, Chi Lim¹, Venugopalan Raghavan¹, Joyjit Chatteraj¹, Jan Schindler² and Feng Yang¹

¹ Institute of High Performance Computing, Agency for Science, Technology and Research, 1 Fusionopolis Way #16-16 Connexis, Singapore, Singapore

² Manaaki Whenua - Landcare Research, Wellington, New Zealand

Keywords: Digital Twin, Tree Species Model, Point Cloud Data, Tree Branch Reconstruction, Procedural Modeling, Optimisation.

Abstract: 3D digital twin trees for a city-scale have been limited to low-resolution, static shape models due to challenges in automation/scalability, cost performance, tree growth dynamics, species complexities and compatibilities with simulations and virtual city platforms. To address those challenges for high-resolution tree models, we propose an automated workflow of generating large-scale, lightweight, dynamic digital-twin tree species models from point cloud data. Species digital twins are modelled as detailed hierarchical branch structures by solving for all species profile parameters through stages of branch reconstruction from point cloud data, species profiling by machine learning, tropism transfer, optimisation and species growth modelling based on botany and limited field survey. We show that the generated high-resolution tree models can be lightweight while representing their true species characteristics and dynamic botanical architecture (branching patterns and growth processes).

Complete Paper #249

Parameter-Free Connectivity for Point Clouds

Diana Marin, Stefan Ohrhallinger and Michael Wimmer
Institute of Visual Computing & Human-Centered Technology, TU Wien, Austria

Keywords: Proximity Graphs, Point Clouds, Connectivity.

Abstract: Determining connectivity in unstructured point clouds is a long-standing problem that has still not been addressed satisfactorily. In this paper, we analyze an alternative to the often-used k-nearest neighborhood (kNN) graph - the Spheres of Influence Graph (SIG). We show that the edges that are neighboring each vertex are spatially bounded, which allows for fast computation of SIG. Our approach shows a better encoding of the ground truth connectivity compared to the kNN for a wide range of k, and additionally, it is parameter-free. Our result for this fundamental task offers potential for many applications relying on kNN, e.g., parameter-free normal estimation, and consequently, surface reconstruction, motion planning, simulations, and many more.

Session 8A
15:30 - 17:00
User Performance Evaluation

HUCAPP
Room Space 4D

Complete Paper #18

Sense of Presence, Realism, and Simulation Sickness in Operational Tasks: A Comparative Analysis of Virtual and Mixed Reality

Giorgio Ballestin and Heike Diepeveen

Department of Maritime Operations - Human Factors, MARIN (Maritime Research Institute Netherlands), Haagsteeg, Wageningen, Netherlands

Keywords: Human Computer Interaction, Empirical Studies in HCI, Mixed Reality, Sense of Presence, Simulation Sickness.

Abstract: Virtual Reality (VR) training is often used to replicate real-world situations, which brings many challenges. Interaction in VR is notoriously more difficult with respect to real world interaction, and VR wearables can sometimes decrease the quality of the interaction, leading to less effective simulations. While reliability and stability are important considerations, it is also crucial sometimes to have a high level of physical realism in order to develop proper muscle memory. This study aims to investigate whether Video-See-Through Augmented Reality (VST AR) can enhance the interaction during an operational task, and its impact on the immersivity and sense of presence, when compared to a VR based approach. Our results show that VR seems to perform better than AR in terms of sense of presence, with a much lower impact on sickness symptoms, suggesting that it is important to strive for a balance between reliability and realism in order to create an immersive and effective training environment.

Complete Paper #438

Perception of a Spatial Implausibility Caused by Seamless Covert Teleportation

Mathieu Lutfallah, Dylan Pires, Valentina Gorobets and Andreas Kunz

Innovation Center Virtual Reality, ETH Zurich, Switzerland

Keywords: Locomotion, Walking, Redirection, Human Perception.

Abstract: This paper investigates human perception of spatial implausibility through seamless teleportation in circular and hexagonal closed-loop corridors. Spatial implausibility refers to virtual spaces that deviate from Newtonian Euclidean geometry rules, while seamless teleportation involves changing the user's position in the virtual world without their awareness. In this work, the impression of implausibility is generated by subtle teleportation of the user within these corridors, thereby allowing them to unconsciously skip certain sections of them. Different levels of this "implausibility" are presented by varying the percentage of skipping within these corridors, specifically 0%, 15%, and 30% of the corridor's overall length. These implausibilities are assessed through a within-subject study on naive participants to determine their perception of spatial implausibility. Our findings indicate no significant difference in the detection rates between the two corridor shapes. Interestingly, most participants interpreted the manipulation as a change in the environment's shape or size while only few could perceive the teleportation and the skip. This paves the way for future research to leverage this technique for subtle spatial manipulation.

Oral Presentations (Online) 8
15:30 - 17:00
Deep Learning for Visual Understanding

VISAPP
Room Online 1

Complete Paper #77

Attention-Based Shape and Gait Representations Learning for Video-Based Cloth-Changing Person Re-Identification

Vuong Nguyen, Samiha Mirza, Pranav Mantini and Shishir Shah

Quantitative Imaging Lab, Dept. of Computer Science, University of Houston, Houston, Texas, U.S.A.

Keywords: Video-Based Person Re-Identification, Cloth-Changing Person Re-Identification, Gait Recognition, Graph Attention Networks, Spatial-Temporal Graph Learning.

Abstract: Current state-of-the-art Video-based Person Re-Identification (Re-ID) primarily relies on appearance features extracted by deep learning models. These methods are not applicable for long-term analysis in real-world scenarios where persons have changed clothes, making appearance information unreliable. In this work, we deal with the practical problem of Video-based Cloth-Changing Person Re-ID (VCCRe-ID) by proposing "Attention-based Shape and Gait Representations Learning" (ASGL) for VCCRe-ID. Our ASGL framework improves Re-ID performance under clothing variations by learning clothing-invariant gait cues using a Spatial-Temporal Graph Attention Network (ST-GAT). Given the 3D-skeleton-based spatial-temporal graph, our proposed ST-GAT comprises multi-head attention modules, which are able to enhance the robustness of gait embeddings under viewpoint changes and occlusions. The ST-GAT amplifies the important motion ranges and reduces the influence of noisy poses. Then, the multi-head learning module effectively reserves beneficial local temporal dynamics of movement. We also boost discriminative power of person representations by learning body shape cues using a GAT. Experiments on two large-scale VCCRe-ID datasets demonstrate that our proposed framework outperforms state-of-the-art methods by 12.2% in rank-1 accuracy and 7.0% in mAP.

Complete Paper #82

Multi-Task Learning Based on Log Dynamic Loss Weighting for Sex Classification and Age Estimation on Panoramic Radiographs

Igor Prado, David Lima, Julian Liang, Ana Hougaz, Bernardo Peters and Luciano Oliveira

Intelligent Vision Research Lab, Federal University of Bahia, Brazil

Keywords: Multi-Task Learning, Panoramic Radiographs, Sex Classification, Age Estimation.

Abstract: This paper introduces a multi-task learning (MTL) approach for simultaneous sex classification and age estimation in panoramic radiographs, aligning with the tasks pertinent to forensic dentistry. For that, we dynamically optimize the logarithm of the task-specific weights during the loss training. Our results demonstrate the superior performance of our proposed MTL network compared to the individual task-based networks, particularly evident across a diverse data set comprising 7,666 images, spanning ages from 1 to 90 years and encompassing significant sex variability. Our network achieved an F1-score of $90.37\% \pm 0.54$ and a mean absolute error of 5.66 ± 0.22 through a cross-validation assessment procedure, which resulted in a

gain of 1.69 percentage points and 1.15 years with respect to the individual sex classification and age estimation procedures. To the best of our knowledge, it is the first successful MTL-based network for these two tasks.

Complete Paper #151

Efficient Parameter Mining and Freezing for Continual Object Detection

Angelo Menezes¹, Augusto Peterlevitz², Mateus Chinelatto² and André de Carvalho¹

¹ *Institute of Mathematics and Computer Sciences, University of São Paulo, São Carlos, Brazil*

² *Computer Vision Department, Eldorado Research Institute, Campinas, Brazil*

Keywords: Object Detection, Continual Learning, Continual Object Detection, Replay, Parameter Mining.

Abstract: Continual Object Detection is essential for enabling intelligent agents to interact proactively with humans in real-world settings. While parameter-isolation strategies have been extensively explored in the context of continual learning for classification, they have yet to be fully harnessed for incremental object detection scenarios. Drawing inspiration from prior research that focused on mining individual neuron responses and integrating insights from recent developments in neural pruning, we proposed efficient ways to identify which layers are the most important for a network to maintain the performance of a detector across sequential updates. The presented findings highlight the substantial advantages of layer-level parameter isolation in facilitating incremental learning within object detection models, offering promising avenues for future research and application in real-world scenarios.

Complete Paper #309

Region-Transformer: Self-Attention Region Based Class-Agnostic Point Cloud Segmentation

Dipesh Gyawali, Jian Zhang and Bijaya Karki

School of Electrical Engineering and Computer Science, Louisiana State University, Baton Rouge, LA 70803, U.S.A.

Keywords: 3D Vision, Class-Agnostic Segmentation, Self-Attention, Point Cloud, Region-Growth.

Abstract: Point cloud segmentation, which helps us understand the environment of specific structures and objects, can be performed in class-specific and class-agnostic ways. We propose a novel region-based transformer model called Region-Transformer for performing class-agnostic point cloud segmentation. The model utilizes a region-growth approach and self-attention mechanism to iteratively expand or contract a region by adding or removing points. It is trained on simulated point clouds with instance labels only, avoiding semantic labels. Attention-based networks have succeeded in many previous methods of performing point cloud segmentation. However, a region-growth approach with attention-based networks has yet to be used to explore its performance gain. To our knowledge, we are the first to use a self-attention mechanism in a region-growth approach. With the introduction of self-attention to region-growth that can utilize local contextual information of neighborhood points, our experiments demonstrate that the Region-Transformer model outperforms previous class-agnostic and class-specific methods on indoor datasets regarding clustering metrics. The model generalizes well to large-scale scenes. Key advantages include capturing long-range dependencies through self-attention, avoiding the need

for semantic labels during training, and applicability to a variable number of objects. The Region-Transformer model represents a promising approach for flexible point cloud segmentation with applications in robotics, digital twinning, and autonomous vehicles.

| | |
|--|---------------------|
| Tutorial | Room Plenary |
| 15:30 - 17:00 | |
| Efficient 3D Facial Animation using Low Cost Motion Capture Solutions | |

Efficient 3D Facial Animation using Low Cost Motion Capture Solutions

João Marcelo N. Teixeira
Voxar Labs, Brazil

Abstract: A large number of 3D facial animation techniques have emerged with different goals, to make the representation of facial expressions more realistic, decrease computing time, the need for specialized equipment, etc. With new techniques come new definitions, concepts, and terms that correlate with methods that have existed for decades or even more recent ones. Parameterization, interpolation, blendshapes, motion-capture and others are concepts that often appear in the literature, in a generic way, as techniques, but which actually have different levels in information hierarchy. The first part of this tutorial aims to clearly classify the different techniques and concepts of the 3D facial animation literature, locating them in each step of the 3D facial animation pipeline, through a parametric analysis. The second part of the proposed tutorial will focus on Facepipe, an efficient and low cost alternative for capturing face motion information and applying it on 3D facial animation software. The proposed solution will be compared to existing ones regarding cost, features and utilization. At last, a hands-on step-by-step use of Facepipe will be performed to demonstrate the possibilities it brings to the 3D animation community.

Thursday, 29

| | |
|-------------------------|---------------------|
| Industrial Panel | VISIGRAPP |
| 17:15 - 18:15 | Room Plenary |

| | |
|--|---------------------|
| Closing Session & Awards Ceremony | VISIGRAPP |
| 18:15 - 18:45 | Room Plenary |

Author Index

- Østergaard, C. 77
- A, S. 78
- Abid, M. 82
- Abilkassov, S. 71
- Aelterman, J. 99
- Agarwal, N. 99
- Agosto, J. 111
- Agrawal, D. 104
- Agudo, A. 59
- Aguilar, E. 40
- Ahlberg, J. 46
- Ahmad, B. 58
- Ahmed, V. 40
- Aissaoui, A. 65
- Aizawa, H. 39
- Aksoy, E. 73
- Al-Azzo, F. 72
- Alabert, T. 104
- Alansari, M. 72
- Alansari, S. 72
- Alenyà, G. 73
- Alexe, B. 92
- Alf, M. 37
- Alfonso, S. 41
- Alibekov, U. 45
- Allaert, B. 65
- Almansa, A. 44
- Almeida, J. 95
- Alnumay, Y. 55
- Alon, P. 79
- AlRemeithi, H. 72
- Amaro, O. 93
- Amel, O. 75
- Amghar, T. 53
- Amirshahi, S. 91
- Amorim, C. 113
- Andersen, M. 43
- Andersen, N. 77, 108
- Ando, Y. 57
- Antensteiner, D. 45
- Appiah, K. 113, 114
- Aqeel, M. 90
- Aragão, D. 80
- Aranjuelo, N. 79, 125
- Araújo, A. 60
- Archambault, T. 126
- Arganda-Carreras, I. 125
- Argyros, A. 94, 118
- Arnaldi, B. 39
- Arth, C. 117
- Asad, M. 66
- Astreika, P. 82
- Atwood, R. 105
- Atzberger, D. 77
- Aurelio, S. 37
- Avrithis, Y. 36
- Axelsson, W. 123
- Ayache, S. 36
- Ayari, R. 114
- Ayvaz, S. 37
- Backes, A. 43, 112, 119
- Badejo, J. 41, 85
- Bae, A. 77
- Bagade, P. 78
- Bagdasarian, M. 118
- Bah, M. 38, 67
- Baldi, V. 53
- Balemans, D. 89
- Ballestin, G. 128
- Bandi, C. 81
- Banić, N. 124
- Barbosa, G. 34, 119
- Barnett, A. 99
- Barthel, F. 120
- Barthel, K. 87
- Bartol, K. 115
- Basting, M. 80
- Bauer, F. 69
- Beautemps, D. 33
- Bečková, I. 89
- Bedoui, M. 114
- Ben Abdallah, A. 114
- Benato, B. 68
- Benavente, R. 35
- Benitez-Garcia, G. 117, 127
- Béréziat, D. 126
- Bernardino, C. 104
- Bertolotto, M. 99
- Bertozzi, M. 59
- Bescop, J. 65
- Bessonitsyn, E. 36
- Bethge, M. 80
- Bezerra, A. 80
- Bhattacharya, S. 42
- Bhojan, A. 115
- Bianchi, L. 108
- Bianco, S. 124
- Bigalke, A. 125
- Bigum, L. 43
- Billast, M. 74, 112
- Bittner, M. 77
- Bizyaev, I. 124
- Bocklet, T. 103
- Bogdan, V. 100
- Bojanić, D. 115
- Bolten, T. 73
- Bombeke, K. 74
- Bonchiş, C. 100
- Bondarev, E. 127
- Bonin-Font, F. 92
- Borchardt, T. 86
- Borghi, G. 65
- Borrás, J. 73
- Bosaghzadeh, A. 83
- Bouakaz, S. 52
- Boujou, M. 37
- Boumrah, M. 85
- Bourdon, P. 41
- Bourmaud, G. 46
- Braffort, A. 33
- Branzan Albu, A. 50
- Brás, D. 93
- Braz Junior, G. 60
- Brega, J. 60
- Brémond-Martin, C. 101
- Brožová, A. 35
- Bruinţjes, R. 80
- Buenvaron, G. 92
- Buhmann, J. 104
- Bulatov, D. 127
- Bull, H. 33
- Bundscherer, M. 103
- Burch, M. 52
- Burduk, R. 56
- Burghouts, G. 42
- Burguera, A. 111
- Burmeister, J. 69
- Buzzelli, M. 82
- Camargo de Oliveira, G. 49
- Campana, J. 50
- Cao, W. 76
- Capece, N. 53
- Capogrosso, L. 90
- Cardoso de Paiva, A. 60
- Cardoso, S. 91, 102, 110
- Carlini, E. 109
- Carvalho, R. 110
- Ceaşescu, C. 92
- Cech, T. 77
- Celona, L. 50
- Cenkeramaddi, L. 78
- Cercenelli, L. 49, 108
- Cerqueira, M. 68
- Chakraborty, J. 41, 85
- Chambel, T. 104
- Chambon, S. 109
- Charantonis, A. 126
- Chatterji, S. 79
- Chattoraj, J. 128
- Chebykin, A. 36, 124
- Chekrygina, D. 105
- Chen, G. 56
- Chen, J. 36, 56
- Cheng, Y. 48
- Cherepkova, O. 91
- Chessa, M. 76
- Chiara, M. 90
- Chinelatto, M. 129
- Choi, J. 79
- Chopuryan, V. 45
- Christen, N. 52
- Ciampi, L. 109
- Cinar, O. 54
- Ciocca, G. 50
- Clark, T. 84
- Clavadetscher, S. 52
- Clouchoux, C. 101
- Coelho de Andrade, M. 113
- Cohen, L. 115
- Conci, A. 60, 80, 86
- Costa de Barros, M. 101
- Costa, G. 86
- Costa, M. 94
- Cotrim, A. 34
- Coustaty, M. 85
- Cristani, M. 90
- Cuno, J. 80
- Cutting, D. 54
- Czúni, L. 106
- D'Atanasio, S. 104
- Dahmane, M. 38, 67, 74
- Dahshan, M. 58
- Damm, S. 67
- Danner, M. 104
- Das, A. 46
- Dash, A. 50
- Davoine, F. 116
- De Bruyne, J. 74
- de Carvalho, A. 129
- De Loor, P. 56
- de Oliveira, C. 47
- De Schepper, T. 34, 74, 112
- de With, P. 127
- Decker, L. 50
- Deguchi, D. 36, 56
- Delacroix, D. 113
- Delany, S. 91
- Dell'Eva, A. 59
- Dennison, R. 47
- Denzler, J. 35, 118, 125
- Dera, D. 49
- Desbarats, P. 120
- Deshmukh, J. 43
- Deshpande, P. 105
- Di Ruberto, C. 105
- Dias, J. 72
- Diepeveen, H. 128
- Diller, F. 70
- Dini, A. 37, 66
- Dittmann, J. 67
- Doggett, M. 70
- Dogra ψ , D. 43
- Dolekar, H. 120
- Döllner, J. 58, 69, 77
- Dong, M. 106
- Dornaika, F. 83
- Doumanoglou, A. 39
- Driessens, K. 39
- Drzyzga, G. 94
- Duarte, K. 101
- Dube, D. 115
- Dubey, S. 113
- Dubuisson, S. 37
- Duran, J. 34, 100, 110
- Dziembowski, A. 101
- Ebner, M. 120
- Économidès, M. 120
- Efimova, V. 36, 45, 124
- Eggeling, E. 48, 83
- Egger, B. 69
- Ehlers, H. 123
- Ehlers, J. 41
- Eisert, P. 45, 55, 82, 100, 110, 118, 120
- Eising, C. 46
- Elena, M. 88
- Emanet, F. 51
- Emiliani, N. 49
- Endo, Y. 44

| | | | | | |
|-------------------------|------------|----------------------|--------------------------------|----------------------|---------------|
| Engström, N. | 123 | Gorobets, V. | 128 | Javid, A. | 86 |
| Enokida, S. | 108, 109 | Gottschalk, H. | 57 | Jaworski, P. | 53 |
| Erdem, Ç. | 115 | Gouidis, F. | 118 | Jiang, C. | 116 |
| Ermakov, I. | 124 | Gouiffès, M. | 33 | Jiménez, J. | 111 |
| Erra, U. | 53 | Gouranton, V. | 39 | Jiménez-Delgado, J. | 87 |
| Erradi, M. | 99 | Graff, C. | 70 | Jin, Y. | 117 |
| Ershov, E. | 124, 126 | Graichen, L. | 41 | Joao, L. | 36, 68 |
| Escarpinati, M. | 43 | Graichen, M. | 41 | Jouini, K. | 40 |
| | | Gsell, M. | 101 | Juhász, I. | 92 |
| Fagundes de Rezende, M. | 61 | Gu, Y. | 116 | Jukiewicz, M. | 84 |
| Falcão Filho, J. | 94 | Guay, M. | 104 | Julia, E. | 49 |
| Falcao, A. | 36, 68 | Guillemaut, J. | 85 | Julia, R. | 49 |
| Farhan, M. | 33 | Guillemot, C. | 34 | Jun, Y. | 79 |
| Faria, M. | 49 | Guillou, E. | 52 | Jurado, J. | 111 |
| Farinella, G. | 33, 55, 99 | Günther, L. | 47 | Jurado-Rodríguez, D. | 111 |
| Farkaš, I. | 89 | Guo, J. | 89 | | |
| Farup, I. | 58 | Guo, X. | 117 | Kahl, K. | 57, 74 |
| Feito, F. | 111 | Gusland, D. | 46 | Kameda, Y. | 57 |
| Feitosa, J. | 60 | Gyawali, D. | 129 | Kanamori, Y. | 44 |
| Feng, Z. | 117 | | | Kane, M. | 92 |
| Fernandes, V. | 102 | Habibi, A. | 52 | Kanetaka, Y. | 76 |
| Ferreira, A. | 93 | Hácha, F. | 122 | Kanlis, A. | 58 |
| Ferreira, B. | 59 | Hadžić, A. | 55 | Kapela, K. | 74 |
| Ferreira, H. | 104 | Hadžić, B. | 104 | Karki, B. | 129 |
| Ferreira, J. | 43 | Hafeez, M. | 93 | Karunamoorthy, V. | 105 |
| Figueiredo, M. | 111 | Haj Messaoud, N. | 114 | Karvarsamis, S. | 58 |
| Figueroa, F. | 36 | Hajder, L. | 92 | Kawahara, R. | 89, 110, 119 |
| Filchenkov, A. | 124 | Halton, M. | 46 | Kawanishi, Y. | 56 |
| Filoche, A. | 126 | Hamada, K. | 54 | Kébairi, S. | 85 |
| Finocchiaro, A. | 33 | Hamila, R. | 107, 114 | Kerbl, B. | 69 |
| Fischer, A. | 67, 119 | Hammond, F. | 100 | Kernbauer, T. | 117 |
| Fischer, J. | 71 | Han, D. | 118 | Keren, A. | 123 |
| Fleck, P. | 117 | Hara, T. | 57 | Kesidis, A. | 114 |
| Floor, P. | 58 | Harder, T. | 94 | Keuper, J. | 80 |
| Fondon, I. | 88 | Härmä, A. | 55 | Khalid, A. | 99 |
| Fons, C. | 70 | Hart, R. | 60 | Khan, A. | 106 |
| Forbes, M. | 41, 85 | Hasany, S. | 33 | Khan, M. | 106, 107, 114 |
| Forte, R. | 99 | Hashimoto, H. | 109 | Khonji, M. | 72 |
| Fourer, D. | 75 | Hashimoto, M. | 57 | Khulusi, R. | 77 |
| Fracaro, F. | 93 | Hashimoto, N. | 108 | Kim, E. | 79 |
| Francis, C. | 104 | Hassan, B. | 72 | Kim, H. | 48 |
| François-Bouaou, K. | 101 | Hayashi, N. | 75 | Kim, J. | 39 |
| Frank, J. | 52 | Heckler, L. | 102 | Kim, S. | 79 |
| Freire, G. | 110 | Heinrich, M. | 125 | Kimura, N. | 105 |
| Friedrich, M. | 127 | Hellwich, O. | 102 | Kirihata, H. | 103 |
| Frihi, A. | 88 | Helmling, S. | 80 | Kirnbauer, B. | 55 |
| Frohlich, D. | 85 | Herbeth, M. | 70 | Kitajima, M. | 54 |
| Fu, B. | 77 | Hermes, N. | 125 | Kitayama, S. | 54 |
| Fu, C. | 115 | Higa, K. | 65 | Knoll, P. | 45 |
| Fuchs, N. | 52 | Higgins, J. | 42 | Kockelmann, W. | 105 |
| Fujiyoshi, H. | 75, 90 | Hildebrandt, M. | 86 | Koga, R. | 108 |
| Fukuda, K. | 39 | Hilsmann, A. | 45, 55, 82, 100, 110, 118, 120 | Kolingerová, I. | 107, 122 |
| Fukushima, N. | 76 | Hirakawa, T. | 75, 90 | König, M. | 104 |
| Fummi, F. | 90 | Hirao, T. | 119 | König, R. | 102 |
| Furnari, A. | 33, 55, 99 | Hirayama, T. | 56 | Konstantoudakis, K. | 58 |
| | | Hirose, M. | 121 | Korbaa, O. | 40 |
| Gaboriau, R. | 113 | Hoketsu, S. | 90 | Korçak, Ö. | 115 |
| Galanakis, G. | 94 | Holz, T. | 67 | Kotani, K. | 65 |
| Ganguly, R. | 67 | Hontani, H. | 108 | Kozaki, T. | 44 |
| Garbaya, S. | 85 | Horn, M. | 69 | Kraetzer, C. | 86 |
| Garcia de Carvalho, M. | 101 | Houdard, A. | 44 | Krell, G. | 66 |
| García, M. | 79, 125 | Hougaz, A. | 129 | Krishna Mohan, C. | 78 |
| Garrido, A. | 111 | House, L. | 58 | Krispel, U. | 83 |
| Gavas, E. | 50 | Hsu, C. | 101 | Kristensen, J. | 87 |
| Gaytan, S. | 88 | Hueber, T. | 33 | Krösl, K. | 69 |
| Gebler, C. | 73 | Huet, S. | 70 | Krüger, N. | 43 |
| Gehler, P. | 37 | Huizinga, W. | 42 | Kruihof, M. | 42 |
| Gennaro, C. | 109 | Hussain, R. | 76 | Krupa, M. | 74 |
| Gentil, C. | 122 | Hwang, S. | 79 | Kucher, K. | 123 |
| Gentner, M. | 71 | | | Kugler, M. | 108 |
| Germain, H. | 46 | Ibrahim, M. | 35 | Kuijff, H. | 42 |
| Gilio, G. | 53 | Ibrahim, Z. | 83 | Kümmerer, M. | 80 |
| Gilles, M. | 56 | Iglesias, A. | 125 | Kunz, A. | 128 |
| Giraud, R. | 46 | Iguernaissi, R. | 37 | Kuo, M. | 89 |
| Giray, T. | 111 | Ikkala, J. | 69 | Kurihara, Y. | 54 |
| Girella, F. | 90 | Iñiguez de Gordo, J. | 79 | Kurita, T. | 39 |
| Giusti, A. | 125 | Ishikawa, T. | 103 | Kusnick, J. | 108 |
| Gkika, I. | 58 | Ito, A. | 81 | Kuznetsov, M. | 45 |
| Globin, M. | 99 | Ito, K. | 105 | Kwiatkowski, M. | 102 |
| Göb, S. | 100 | | | Kwok, S. | 113, 114 |
| Gobeawan, L. | 51, 128 | Jääskeläinen, P. | 69 | Kyriakides, C. | 72 |
| Goebbels, S. | 121 | Jadon, H. | 43 | | |
| Goeman, N. | 65 | Janbein, M. | 122 | Laaksonen, J. | 81 |
| Gomes, N. | 49 | Jänicke, S. | 77, 108 | Labayan, M. | 79 |
| Gomez-Jaime, A. | 111 | Jaroensiripong, P. | 65 | Lacerda, G. | 90 |
| Gomez-Krämer, P. | 85 | Jarsky, I. | 124 | Lacôte-Coquereau, C. | 41 |
| Gonçalves, D. | 93 | Javed, S. | 72 | Lagares, I. | 88 |
| Gonçalves, L. | 80 | | | Lalanne, D. | 52 |

| | |
|--------------------------|-----------------|
| Lampa, P. | 56 |
| Lampman, C. | 84 |
| Lanitis, A. | 72 |
| Laria, H. | 59 |
| Larroque, V. | 38 |
| Larsen, M. | 46 |
| Latonov, V. | 45 |
| Laurent, T. | 82 |
| Lauttia, T. | 69 |
| Lazarou, I. | 114 |
| Lazarou, M. | 60 |
| Le, N. | 40 |
| Lecomte, J. | 75 |
| Lee, A. | 113 |
| Lee, K. | 79 |
| Lee, S. | 116 |
| Lee, W. | 101 |
| Leng, K. | 105 |
| Li, Y. | 118 |
| Liang, H. | 38 |
| Liang, J. | 129 |
| Lim, C. | 128 |
| Lim, S. | 51 |
| Lima, D. | 129 |
| Ling, Y. | 85 |
| Linsen, L. | 71 |
| Lipp, N. | 82 |
| Liu, X. | 91, 128 |
| Ljubenović, M. | 111 |
| Lo Duca, A. | 71 |
| Lo, D. | 77 |
| Loddo, A. | 105 |
| Loh, L. | 124 |
| Lopes, T. | 91 |
| López, A. | 111 |
| Loyola, A. | 91, 102, 110 |
| Lukač, L. | 122 |
| Lumbreras, F. | 35 |
| Luong, H. | 56 |
| Łupkowski, P. | 53 |
| Lupkowski, P. | 84 |
| Lutfallah, M. | 128 |
| Møgelmoose, A. | 43 |
| Ma, S. | 99 |
| Maag, K. | 37, 119 |
| Macé, L. | 53 |
| Macé, M. | 53 |
| Madden, M. | 66, 93 |
| Maddock, S. | 47 |
| Madeiras Pereira, J. | 93 |
| Madsen, C. | 87 |
| Mahdi, M. | 52 |
| Mahmoudi, S. | 75 |
| Mahtani, A. | 82 |
| Mai, V. | 40 |
| Maia, H. | 50 |
| Majewski, P. | 56, 89 |
| Mäkitalo, M. | 69 |
| Makrushin, A. | 67 |
| Malawski, F. | 74 |
| Maldonado-Quispe, P. | 95 |
| Maltoni, D. | 65 |
| Maňák, M. | 107 |
| Manfredi, G. | 53 |
| Mannam, V. | 67 |
| Mantegazza, D. | 125 |
| Mantini, P. | 51, 129 |
| Marcelli, E. | 49, 108 |
| Marciniak, K. | 89 |
| Mari, J. | 43 |
| Maria, M. | 38 |
| Marin, D. | 123, 128 |
| Martarelli, R. | 95 |
| Martins, A. | 59, 110 |
| Martins, S. | 104 |
| Masouris, A. | 68 |
| Massa, L. | 59 |
| Matern, S. | 102 |
| Mathiassen, K. | 46 |
| Matsubayashi, S. | 54 |
| Matsukawa, T. | 44 |
| Mauricio, C. | 93 |
| Mazumder, A. | 41 |
| Mazzamuto, M. | 99 |
| Mazzocchetti, S. | 108 |
| McKeever, S. | 91 |
| Medici, P. | 59 |
| Medina, E. | 124 |
| Meffert, N. | 77 |
| Mégret, R. | 111 |
| Meikari, J. | 53 |
| Meléndez-Ríos, R. | 111 |
| Mello Junior, A. | 61 |
| Mendonça, F. | 94 |
| Ménélas, B. | 88 |
| Ménétrier, E. | 53 |
| Menezes, A. | 129 |
| Menotti, D. | 47, 67 |
| Menouar, H. | 107, 114 |
| Merad, D. | 37 |
| Mercelis, S. | 89 |
| Mercha, E. | 99 |
| Mérillou, S. | 38 |
| Mets, K. | 34, 74, 112 |
| Mettänen, M. | 66 |
| Meyer, A. | 52 |
| Meyers, L. | 111 |
| Miandji, E. | 76 |
| Michael, L. | 72 |
| Migalska, A. | 76 |
| Mihara, S. | 93 |
| Mirza, S. | 51, 129 |
| Mishra, S. | 55 |
| Miwa, K. | 54 |
| Miyake, R. | 44 |
| Miyoshi, H. | 108 |
| Moeslund, T. | 43 |
| Mohammed, P. | 104 |
| Mohan, C. | 113 |
| Moises de Sousa, P. | 119 |
| Mokhtari, N. | 56 |
| Molo, M. | 109 |
| Montazeri, Z. | 38 |
| Monteiro, D. | 38 |
| Montes, M. | 38 |
| Montuoro, A. | 107 |
| Moradi, M. | 51 |
| Morar, T. | 43 |
| Moreau, J. | 116 |
| Moreira, L. | 119 |
| Moreira, R. | 119 |
| Moreno-Noguer, F. | 59 |
| Morgenstern, W. | 45 |
| Mori, S. | 46 |
| Morin, G. | 109 |
| Morozov, A. | 66 |
| Mould, D. | 86 |
| Moura, E. | 86 |
| Müller, C. | 82 |
| Müller, P. | 82 |
| Müller, S. | 47 |
| Munksø, R. | 43 |
| Murakami, H. | 56 |
| Murase, H. | 36, 56 |
| Muravyov, S. | 45 |
| Nørgaard, L. | 43 |
| Nagaishi, M. | 108 |
| Naganawa, Y. | 76 |
| Nakahira, K. | 54 |
| Namboodiri, A. | 50 |
| Nanjappan, V. | 38 |
| Narasappareddygar, P. | 105 |
| Navarro, J. | 34 |
| Navarro, L. | 113 |
| Nayak, K. | 43 |
| Nédélec, A. | 56 |
| Neves, J. | 47, 67 |
| Neves, L. | 91, 102, 110 |
| Nguyen, D. | 40 |
| Nguyen, T. | 40 |
| Nguyen, V. | 40, 51, 75, 129 |
| Nguyen-Phuoc, L. | 113 |
| Nickmann, D. | 122 |
| Nicod, L. | 37 |
| Nieradzki, L. | 80 |
| Nieto, M. | 79, 125 |
| Nikolov, I. | 78 |
| Nilles, A. | 47 |
| Nocker, M. | 107 |
| Noel, F. | 111 |
| Nogami, H. | 76 |
| Nonaka, Y. | 65 |
| Norberto, J. | 94 |
| Nyisztor, N. | 92 |
| Obara, B. | 126 |
| Ogino, Y. | 81 |
| Ohno, A. | 121 |
| Ohrhallinger, S. | 128 |
| Ohse, J. | 104 |
| Ohshima, K. | 108 |
| Okabe, T. | 110, 119 |
| Okamoto, N. | 75 |
| Olajos, R. | 70 |
| Olbrich, A. | 80 |
| Oliveira, B. | 113 |
| Oliveira, H. | 91 |
| Oliveira, L. | 91, 129 |
| Oliveira, V. | 122 |
| Olpadkar, K. | 50 |
| Ono, S. | 81 |
| Ono, Y. | 121 |
| Oramas, J. | 83, 89 |
| Orevillo, K. | 77 |
| Orhei, C. | 100 |
| Orlov, I. | 82 |
| Orsingher, M. | 59 |
| Otsuka, M. | 54 |
| Ouakrim, Y. | 33 |
| Ouardirhi, Z. | 75 |
| Ozaki, K. | 47 |
| Palazzo, S. | 51 |
| Palit, A. | 120 |
| Paljic, A. | 70 |
| Páncsics, Z. | 92 |
| Panshin, A. | 124 |
| Papa, J. | 49, 60, 61, 95 |
| Papadakis, N. | 44 |
| Papoutsakis, K. | 118 |
| Park, J. | 48 |
| Park, W. | 48 |
| Parra-Cabrera, G. | 87 |
| Partaourides, H. | 72 |
| Passos, L. | 61 |
| Patel, S. | 55 |
| Patkos, T. | 118 |
| Pears, N. | 116 |
| Pedersen, M. | 91 |
| Pedriani, H. | 34, 50, 95, 119 |
| Pelissier-Combescure, M. | 109 |
| Pellerin, D. | 70 |
| Penk, D. | 69 |
| Penzel, N. | 35, 125 |
| Peperkorn, N. | 104 |
| Pereira, C. | 61, 95 |
| Pereira, D. | 102 |
| Pereira-Sánchez, I. | 34 |
| Peres, F. | 93 |
| Perevozchikov, G. | 126 |
| Pérez-Cano, F. | 87 |
| Peterlevitz, A. | 129 |
| Peters, B. | 129 |
| Petković, T. | 115 |
| Petrescu, D. | 38 |
| Pettifer, S. | 38 |
| Philips, W. | 56 |
| Piater, T. | 35 |
| Pietsch, V. | 107 |
| Pionteck, T. | 66 |
| Pires, D. | 128 |
| Plank, G. | 101 |
| Plexousakis, D. | 118 |
| Pócoş, Ş. | 89 |
| Podgorelec, D. | 107 |
| Pohle-Fröhlich, R. | 73 |
| Pollyea, R. | 58 |
| Polys, N. | 58 |
| Ponsa, D. | 35 |
| Ponte, D. | 40 |
| Poolkrajang, S. | 115 |
| Popa, M. | 71 |
| Pope, J. | 103 |
| Popescu, V. | 42 |
| Poulain d'Andecy, V. | 85 |
| Poull, C. | 122 |
| Prado, I. | 129 |
| Prakash, N. | 103 |
| Pratihar, S. | 79 |
| Pribanić, T. | 115 |
| Proença, H. | 67 |
| Prost, J. | 44 |
| Puts, M. | 107 |

| | | | | | |
|----------------------------|--------------|---------------------------|-----------------|----------------------------|--------------|
| Puzikov, Y. | 37 | Schreck, T. | 77 | Terp, P. | 43 |
| R. de Faria, P. | 91, 102, 110 | Schubert, M. | 57, 74 | Terraf, Y. | 99 |
| Rachow, E. | 77 | Schulte, A. | 84 | Thakare, K. | 43 |
| Rader, M. | 107 | Schulze, M. | 71 | Thaler, F. | 101 |
| Radeva, P. | 40 | Schutte, K. | 42 | Theodosiou, Z. | 72 |
| Radgui, A. | 85 | Schwerd, S. | 84 | Thiede, C. | 58 |
| Rádlí, R. | 106 | Seals, M. | 84 | Thiyagalingam, J. | 105 |
| Radman, A. | 81 | Seibold, C. | 55 | Thoma, M. | 72 |
| Raducanu, B. | 59 | Sekeroglu, K. | 51 | Thomas, U. | 81 |
| Rafferty, K. | 54 | Sekiya, K. | 105 | Thomine, S. | 68 |
| Raghavan, V. | 128 | Semenova, N. | 45 | Tianhao, B. | 60 |
| Ragusa, F. | 72, 79, 99 | Semrau, A. | 87 | Toizumi, T. | 81 |
| Rahtu, E. | 37, 66 | Setti, F. | 90 | Tomizawa, R. | 103 |
| Raidou, R. | 123 | Shaaban, M. | 106 | Tönnies, K. | 73 |
| Ramachandran, G. | 45 | Shah, S. | 51, 129 | Toro, F. | 55 |
| Ratajczyk, D. | 53 | Shalamov, V. | 124 | Torres, D. | 110 |
| Rätsch, M. | 104 | Shao, W. | 66 | Torres, R. | 119 |
| Rayar, F. | 81, 92 | Sharma, S. | 46 | Tosta, T. | 91, 102, 110 |
| Reimann, M. | 69 | Shen, Y. | 117 | Tóth, T. | 92 |
| Reiner, J. | 56, 89 | Shiban, Y. | 104 | Tran, Q. | 86 |
| Reiter, P. | 83, 89 | Shino, M. | 54 | Traviglia, A. | 111 |
| Remolar-Quintana, I. | 87 | Shirvany, R. | 37 | Treplán, G. | 92 |
| Rethmann, J. | 121 | Shishido, H. | 57 | Triguí, H. | 55 |
| Ribeiro, R. | 47 | Sicre, R. | 36 | Triller, J. | 82 |
| Ribera, M. | 40 | Sidibé, D. | 75 | Truong-Allié, C. | 70 |
| Richard, E. | 41 | Sieburg-Rockel, J. | 80 | Tsatsaris, A. | 114 |
| Richard, P. | 41, 53 | Siedel, G. | 66 | Tsukada, M. | 81 |
| Richter, R. | 69 | Silva Júnior, M. | 94 | Tuama, A. | 40, 72 |
| Ricker, J. | 67 | Silva, A. | 91, 102, 110 | Tuor, R. | 52 |
| Riddle, P. | 39 | Silva, J. | 94 | Turner, M. | 114 |
| Riedlinger, T. | 37, 57, 74 | Silva, L. | 43 | Tursynbek, I. | 84 |
| Risy, M. | 39 | Silva, Y. | 61 | Tuval, I. | 92 |
| Roberto, G. | 91 | Siritanawan, P. | 65 | Tzelepis, G. | 73 |
| Roder, M. | 49, 60, 61 | Sistu, G. | 46, 66, 93, 105 | Uchiyama, H. | 65 |
| Rodiga, L. | 83 | Sjöström, M. | 34 | Ueda, S. | 46 |
| Rodrigues, D. | 95 | Slusallek, P. | 82 | Ueda, T. | 110 |
| Rodrigues, R. | 94 | Šmidl, V. | 35 | Ugrinovic, N. | 59 |
| Rodríguez-Cordero, J. | 111 | Smith, J. | 126 | Uhl, A. | 75 |
| Rodríguez-Sánchez, A. | 86 | Snoussi, H. | 68 | Ullah, I. | 66, 93 |
| Rolfjord, S. | 46 | Soares de Sousa, M. | 90 | Ullrich, T. | 48, 83 |
| Rosano, M. | 55 | Soares, F. | 91 | Ulschmid, A. | 69 |
| Rosenthal, P. | 71 | Solari, C. | 76 | Ulucan, D. | 120 |
| Rossen, J. | 43 | Söllinger, D. | 75 | Ulucan, O. | 120 |
| Rottmann, M. | 57, 74 | Sonarghare, S. | 105 | Unger, J. | 76 |
| Roudet, C. | 122 | Sousa de Almeida, J. | 60 | Unkrig, J. | 127 |
| Rouis, K. | 85 | Souza, A. | 113 | Urschler, M. | 39, 55, 101 |
| Roy, A. | 78 | Souza, M. | 50 | Urselmann, T. | 127 |
| Rozhbayani, G. | 72 | Spampinato, C. | 51 | Useinov, L. | 45 |
| Ruis, F. | 42 | Srivastava, M. | 116 | Utsuro, Y. | 57 |
| Ruiz, A. | 59 | Staderini, V. | 45 | Vadicamo, L. | 109 |
| Sørensen, E. | 108 | Stafeev, G. | 36 | van de Weijer, J. | 59 |
| Sá Junior, J. | 112 | Stamminger, M. | 69 | van Gemert, J. | 68, 80 |
| Sabaghi, A. | 83 | Stankowski, J. | 101 | Vandeborren, J. | 65 |
| Safko, M. | 122 | Starling, M. | 113 | Vandersmissen, B. | 83, 89 |
| Saga, R. | 53 | Stathaki, T. | 60 | Vaněček, P. | 122 |
| Saito, H. | 46, 57, 65 | Steckel, J. | 89 | Vanhulst, P. | 52 |
| Sakaue, F. | 44, 47, 121 | Stein, G. | 35, 125 | Vanian, V. | 58 |
| Salman, S. | 112, 117 | Steinwigger, F. | 48 | Váša, L. | 52, 122 |
| Samanta, S. | 79 | Stephani, H. | 80 | Vasconcelos, R. | 90 |
| Sánchez, M. | 79 | Štern, D. | 55 | Verdelho Trindade, N. | 93 |
| Sanfeliu, A. | 59 | Stival, L. | 119 | Victoriano, M. | 91 |
| Sans, E. | 34 | Stonehouse, J. | 126 | Vidal, A. | 61 |
| Santi, S. | 65 | Stradiotti, S. | 49 | Vieira, T. | 59 |
| Santiago-Plaza, G. | 111 | Strain, G. | 38 | Vilain, M. | 46 |
| Santner, T. | 86 | Strauss, E. | 127 | Vilar de Oliveira, E. | 80 |
| Santos da Silva, C. | 94 | Streitberg, J. | 52 | Vítová, N. | 52 |
| Santos, C. | 34 | Strohmeier, C. | 69 | Vlaminck, M. | 56 |
| Santos, J. | 113 | Studzinska, A. | 104 | Vo, A. | 99 |
| Santos, M. | 67 | Su, Y. | 117 | Vock, S. | 66 |
| Sappa, A. | 112 | Suárez, P. | 112 | Vohra, R. | 50 |
| Sato, J. | 44, 47, 121 | Sugiura, T. | 33 | Volpi, R. | 92 |
| Satyanarayana, J. | 113 | Sumner, R. | 104 | Vorabbi, L. | 65 |
| Savas, B. | 123 | Sumongkayothin, K. | 65 | Vörösházi, Z. | 106 |
| Savchik, A. | 124 | Sun, D. | 83 | Vozniak, I. | 82 |
| Sawicki, D. | 87 | Suzuki, E. | 44 | Vuckovic, M. | 123 |
| Sbert, C. | 100, 110 | Suzuki, T. | 57 | Wallburg, L. | 100 |
| Scheibel, W. | 58, 77 | Szabó, A. | 37 | Walsh, G. | 108 |
| Schettini, R. | 50, 82 | Szilagyi, I. | 48 | Wang, H. | 54 |
| Scheuermann, G. | 70 | Taioli, F. | 90 | Wang, K. | 59 |
| Schiavina, R. | 108 | Takahashi, H. | 112, 117, 127 | Wang, Y. | 38, 48, 59 |
| Schindler, J. | 128 | Takasaki, S. | 108 | Warren, P. | 38 |
| Schlotter, M. | 52 | Takeuchi, I. | 108 | Wasielawska, A. | 84, 88 |
| Schmidt, J. | 107, 123 | Tamaki, T. | 33 | Wasim, S. | 33 |
| Schmitt, T. | 103 | Teixeira, J. | 93, 130 | Wegen, O. | 69 |
| Schneider, P. | 45 | Teixeira, M. | 82 | | |
| Scholte, D. | 127 | Telea, A. | 48 | | |
| | | Tennekes, M. | 107 | | |

| | | | | | |
|----------------------|----------|---------------------------|--------------|----------------------------------|----------|
| Wei, L. | 110 | Yamaguchi, M. | 65 | Zakir, A. | 112, 117 |
| Wei, W. | 34 | Yamashita, T. | 75, 90 | Zalik, B. | 122 |
| Wergli, N. | 72 | Yanaka, K. | 93 | Zanchetta do Nascimento, M. | 49 |
| Wicker, J. | 39 | Yang, F. | 128 | Zani, P. | 59 |
| Wiebel, A. | 70 | Yang, W. | 54 | Zarpalas, D. | 39, 58 |
| Wiedemer, T. | 80 | Yang, Y. | 116 | Zbakh, M. | 75 |
| Williamson, T. | 85 | Yano, T. | 105 | Zedda, L. | 105 |
| Willingham, S. | 34 | Yeturu, K. | 120 | Zhang, C. | 36 |
| Wimmer, G. | 75 | Yokota, T. | 108 | Zhang, H. | 36 |
| Wimmer, M. | 69, 128 | Yoshida, A. | 49 | Zhang, J. | 117, 129 |
| Wisotzky, E. | 82, 100 | Yoshikawa, T. | 44 | Zheng, R. | 38 |
| Wittenberg, T. | 100 | Youssef, K. | 58 | Zhou, K. | 49 |
| Wu, H. | 101, 123 | Youssef, S. | 75 | Zhou, Y. | 42 |
| Xenophontos, C. | 84 | Yousuf, M. | 33 | Ziat, M. | 84 |
| Xu, W. | 38 | Z. do Nascimento, M. | 91, 102, 110 | Zimmermann, M. | 42 |
| Yaar, A. | 55 | Zabulis, X. | 94 | Zuo, Z. | 126 |
| | | | | Zwemer, M. | 127 |
| | | | | Žyła, M. | 53 |

Final Program and Book of Abstracts of VISIGRAPP 2024

19th International Joint Conference on Computer Vision, Imaging and Computer Graphics
Theory and Applications

<https://visigrapp.scitevents.org>

EVENT MANAGEMENT SYSTEM:



LOGISTICS:



acm In-Cooperation



ENDORSED BY:



IN COOPERATION WITH:



PROCEEDINGS WILL BE SUBMITTED FOR EVALUATION FOR INDEXING BY:



Copyright © 2024 by SCITEPRESS - Science and Technology Publications, Lda. All Rights Reserved