The Sixth IEEE International Image Processing, Applications and Systems Conference (IPAS25)

Organized by:

The National Engineering School of Sfax, Sfax University, Tunisia
Lyon 2 University, France
INRIA, Sophia Antipolis, France

Hosted by:

Lyon 2 University, France
9-11 January 2025

Technically sponsored by:

IEEE Tunisia Section
IEEE France Section
IEEE SPS France Chapter
IEEE SPS Tunisia Chapter
Region 08 - Europe, Middle East, Africa

Financially sponsored by:

IEEE Tunisia Section
Association Boussole de la Recherche Scientifique - ABORS Tunisia

Preface

IPAS25 will take place in Lyon, France during January 9-11, 2025. For more convenience and safety to some authors, the conference is managed in hybrid mode. Nevertheless, online presentations are managed in the same days of the conference so that authors attending IPAS25 inLyon can follow them and can interact with authors presenting on-line.

Image processing is driving the intelligence and interactivity of modern systems, sometimes exceeding human capabilities. The International Image Processing Applications and Systems conference unites global experts in image processing and computer vision to share their latest breakthroughs and foster collaboration. The conference features high-quality research papers on all aspects of computer vision applications, promoting innovation across academia and industry.

The 6^{th} IEEE International Conference of Image Processing Applications and Systems has 15 regular sessions composed of 95 participations including 87 regular papers from 38 countries (Algeria, Saudia Arabia, Australia, Bahrain, Belgium, Bangladesh, Canada, China, Cyprus, Czech Republic, Korea, Arab United Emirates, Egypt, France, Finland, Germany, Georgia, Greece, Hungry, India, Italy, Iran, Ireland, Jordan, Lebanon, Libya, Malaysia, Niger, Norway, Netherland, United States, United Kingdom, Vietnam, Pakistan, Romania, Spain, Sweden, Tunisia). The review process has been double blinded. On behalf the program committee of IPAS25, we are grateful to all the authors for their fine contributions and for the high quality of their papers.

The 6^{th} IEEE International Conference of Image Processing Applications and Systems has six invited keynote speakers on different topics of image processing applications. On behalf the organizing committee of IPAS25, we would like to thank all the speakers for their support to the success of the conference.

General Co-Chairs
Serge Miguet
Dorra Sellami
François Brémond

Contents

Committees

General Chairs

Serge Miguet (Lyon University, France)
François Brémond (Inria Sophia Antipolis - France)
Dorra Sellami (National Engineering School of Sfax, Sfax University, Tunisia)

Technical Program Chairs

William Puech, France
Laure Tougne-Rodet, France
Carlos Crispim Junior, France
Ali Wali, Tunisia
Riadh Abdelfattah, Tunisia

Steering Committee Chairs

Sébastien Ambellouis, Lille, France)
Fabio Solary, Genova, Italy
Atika Rivenq, France
Ahmed-Abdelmalik Taleb, France
Alima Damak, Tunisia
Petra Perner, Ibai, Germany
Nicolas Dobigeon, France
Nicolas Gillis, Belgium
Cecilia Zanni Merk, France
Stella Mark Swecker, France
Petra Perner, Germany
Jacques Boonaert, France
Cyril Meurie, France
Chao-Cheng Wu, Taiwan
Pau-Choo Chung, Taiwan

Plenary session Chairs

Hela Boulehmi Chtara, Tunisia
Anthony Fleury, France
Hassan Rabah, France
Abdesslam Benzinou, France

Publication Chairs

Habib Kamoun , FSS, Tunisia
Mouna Zouari, CRNS, Tunisia

Exhibits and Industry Liaison Chairs

Mohamed Louati (Logidas, Tunisia)
Anis Benabdennebi (GFI, Tunisia)

Local Arrangement Chairs

Carlos Crispim Junior, LIRIS, France
Bertrand Kerautret, LIRIS, France
Mihaela Scuturici, France
Jihen Frikha, ENIS, Tunisia)
Majd Belaaj, ENIS, Tunisia)

Organizing Chairs

Amir Gargouri, ENETCOM, Tunisia
Manel Loumi, ENIS, Tunisia
Mondher Frikha, ENETCOM, Tunisia

Plenary, tutorial and special session committee members

Nawres Khlifa (Tunisia)
Mounir Sayadi (ENSIT, Tunisia)
Aicha Bouzid (ENIT, Tunisia)
Ali Douik (ENIM, Tunisia)
Mohamed Atri (ENIM, Tunisia)

Mohamed Ali Mahjoub (ENISO, Tunisia)
Zied Lachiri (ENIT, Tunisia)
Hassene Seddik (ENSIT, Tunisia)
Monia Turki (ENIT, Tunisia)
Kamel Hamrouni (ESPRIT, Tunisia)
Sabeur Abid (ENSIT, Tunisia)

Registration Chairs

Hejer Elloumi, PRIMATEC, Tunisia) Wiem Abbes, SOFRECOM, Tunisia)

Publicity Chairs

Atika Rivenq, Valenciennes, France
Aymen Souissi, Tunisia

Organizing Committee members

Andrea Trucco (University of Genoa, Italy)
Fabio Solari (University of Genoa, Italy)
Mouna Zouari (ENIS, Tunisia)
Manel Loumi (ENIS, Tunisia)
Jihène Frikha (ENIS, Tunisia)
Wiem Abbes (ENIS,Tunisia)
Hejer Elloumi (ENIS,Tunisia)

Technical Program Committee Members

Daniele Caviglia (ITL)
Lorenzo Natale (ITL)
Manuela Chessa (ITL)

Fiora Pirri (ITL)
Gobron Stéphane (SWZ)
Francisco Barranco(ES)
Hela Boulemi Chtara (TN)
Tièche François(CH)
Antonio Rodríguez-Sanchez(AUS)
Alaa Balti (TN)
Nicolas Thome (FR)
Jean-Marc Odobez (SWZ)
Ludovic Macaire (FR)
Pascal Vasseur (FR)
Serge Miguet (FR)
James Ferryman (UK)
Andreas Koschan (USA)
Alexander Loui (CA)
Andrea Cavallaro (UK)
Anoop Cherian (AUS)
Bogdan Raducanu (ES)
Simone Calderara (ITL)
Carlo Regazzonui (ITL)
Li Cheng (CHN)
Christian Micheloni (ITL)
Curtis Padgett (USA)
Shiloh Dockstader (USA)
Riad Hammoud (UK)
Hazim Ekenel (SWZ)
Federico Tombari (DE)

Mahesh Gellaboina (India)
Guillaume Ginolhac (FR)
Pierre-Antoine Thouvenin (UK)
Emilie Chouzenoux (FR)
Cédric Demonceaux (FR)

Chiara Bartolozzi(ITL)
Maurizio Valle(ITL)
Giovanni Maria Farinella(ITL)
Dimitri Ognibene(UK)
Norbert Kruger(DK)
Eris Chinellato(UK)
Hanene Sehli (TN)
Carlos Fernando (FR)
Javier Diaz(ES)
Sabeur Abid (TN)
Dominique gruyer (FR)
Sergio Velastin (ES)
Vincent Fremont (FR)
Olivier Losson (FR)
Remy Boutteau (FR)
Juan Carlos Niebles (USA)
François Fleuret (FR)
Aydin Alatan (UK)
Andrea Prati (ITL)
Andreas Hutter (DE)
Atsushi Shimada (JPN)
Boxin Shi (CHN)
Caifeng Shan (UK)
Chan Fang (CHN)
Anustup Choudhury (USA)
Cong Phuoc Huynh (AUS)
Dimitrios Tzovaras (GR)
Donatello Conte (USA)
Ehsan ZareBorzeshi (KOR)
FahimehRezazadegan (KOR)
Frank Pagel (DE)
Yoann Altmann (UK)
Marcelo Pereyra (UK)
Mauro Dalla-Mura (FR)
Emmanuel Soubies (SWZ)

Paul Honeine (FR)	Olivier Lezoray (FR)
Adrian Basarab (FR)	Jérôme Bobin (FR)
Masoud faraki(AUS)	Massimo Piccardi(SYD)
Kamel Hamrouni (TN)	Kamel Abderrahim (TN)
Mohamed Akil (FR)	Taher Battikh(TN)
Amel Benazza (TN)	Mohamed Hédi Bedoui(TN)
Faouzi Benzarti (TN)	Dhibi(TN)
Afef Elloumi(TN)	Mourad Zaeid(TN)
Kais Ouni (TN)	Issam Jabri (TN)
Mohamed Atri (TN)	Anis Kricha (TN)
Sakly Anis (TN)	Kouraichi Maher (TN)
Mourad Moussa (TN)	Olfa Jemai (TN)
Salwa Said (TN)	Sofiane Hachicha (TN)
Marwa Hmida (TN)	Ines Rahmeny(TN)

Opening & Welcome Ceremony (Thursday 9 Jan 2025 at 9:00-09:30)

Keynote I by Prof. François Brémond
Session chair: Prof. Serge Miguet (Thursday 9 Jan 2025 at 9:30-10:10)

Robust Real-Time Monitoring of Complex Human Activities using Multimodal Video Analytics

9 January 2025
9H30-10H10

François Brémond
INRIA – Sophia Antipolis
francois.Bremond@inria.fr

Summary

This talk addresses the challenge of real-time, continuous monitoring of complex human activities using video analytics. Existing work has focused either on simple activities in uncontrolled environments or on more complex activities within tightly constrained, pre-segmented video clips. This limitation prevents accurate, continuous monitoring in real-world scenarios. We introduce novel methods that overcome this limitation by combining multiple modalities (skeleton tracking, optical flow, gaze, emotion) for robust ADL recognition in untrimmed videos. Results on diverse datasets (Toyota SmartHome, NTU-RGB+D, Charades, Northwestern UCLA) highlight the potential for applications in healthcare and home monitoring.

François Bremond biography

François Brémond is a Research Director at Inria Sophia Antipolis-Méditerranée, where he created the STARS team in 2012. He has pioneered the combination of Artificial Intelligence, Machine Learning and Computer Vision for Video Understanding since 1993, both at Sophia-Antipolis and at USC (University of Southern California), LA. In 1997 he obtained his PhD degree in video understanding and pursued this work at USC on the interpretation of videos taken from UAV (Unmanned Airborne Vehicle). In 2000, recruited as a researcher at Inria, he modeled

human behavior for Scene Understanding: perception, multi-sensor fusion, spatio-temporal reasoning and activity recognition. He is a co-founder of Keeneo, Ekinnox and Neosensys, three companies in intelligent video monitoring and business intelligence. He also co-founded the CoBTek team from Nice University in January 2012 with Prof. P. Robert from Nice Hospital on the study of behavioral disorders for older adults suffering from dementia. He is author or co-author of more than 250 scientific papers published in international journals or conferences in video understanding. He has (co)- supervised 20 PhD theses. More information is available at: http://www-sop.inria.fr/members/Francois.Bremond/

Regular Session S1P: Deep Learning and Novel Imaging Techniques for Medical Diagnosis Session chair: Laure Tougne Rodet (9 Jan 2025 at 10:20-12:20)

A Robust Approach for Classifying Laparoscopic Video Distortions using ResNet-50

9 Jan 2025
10:20-12:20
session S1P

Mohamed Belmokeddem[1], Kamila Khemis[2], Salim Loudjedi[3]
[1]:belmoked@gmail.com, [2]:kamila.khemis@univ-tlemcen.dz,
[3]:ismatsalim.loudjdimouedden@univ-tlemcen.dz

[1,2]:Biomedical Engineering Department, Faculty of Technology, University Abou-bekr Belkaid of Tlemcen, Algeria
[3]:Medicine Department, University Abou-bekr Belkaid of Tlemcen, Algeria

Summary

During minimally invasive surgery, the laparoscopic lens can be contaminated by condensation, smoke, blood, and debris, directly impacting both the surgeon's visibility and the quality of the laparoscopic video. To address this issue, various solutions have been proposed for detecting and classifying different types of distortions in laparoscopic videos. Our method focuses on the automated identification and categorization of five distortion categories in laparoscopic videos, as defined in the Laparoscopic Video Quality database by Khan et al. These distortions are noise ('NO'), smoke ('SM'), uneven illumination ('UI'), defocus blur ('DB'), and motion blur ('MB'). Utilizing Residual Networks (ResNet50), our method achieves high prediction rates in detecting and classifying these distortions: 100% for 'SM', 100% for 'NO', 99.82% for 'UI', and 99.92% for 'DB', and 99.87% for 'MB'. These consistently high scores across all categories highlight the model's robust ability to generalize effectively and deliver precise predictions.

---

9 Jan 2025
10:20-12:20
session S1P

8D Hyperchaotic System and Gold Sequences for Improved Medical Image Cube Encryption

Eyad Mamdouh[1], Mohamed Gabr[2], Wassim Alexan[3], Youstina Megalli[4]
[1]:eyad.gaber@ieee.org, [2]:mohamed.gabr@ieee.org, [3]:wassim.alexan@ieee.org,
[4]:youstina.kamal@guc.edu.eg

German University in Cairo, Egypt

Summary

This paper presents an advanced encryption algorithm specifically designed to enhance the security of volumetric medical image data, crucial for the Internet of Medical Things (IoMT). The algorithm encrypts a stack of 256 images, each with dimensions of 256*256 pixels, through a meticulous multi-stage process. It begins by segmenting an image cube into Red, Green, and Blue channels, which are then encrypted through three phases: XOR operations with keys from an 8D hyperchaotic system, substitution with S-boxes derived from Gold sequences, and a final transformation using Fibonacci Q-matrices. This approach significantly improves security by improving entropy, decreasing cross-correlation, and strengthening resistance to statistical attacks, with various cryptographic seeds used at each stage to enhance the robustness of the encryption. Specific performance evaluation metrics include a pixel cross-correlation of approximately 0, an NPCR of 99.61%, a UACI of 34.45%, an entropy of 7.998, and a key space that exceeds 2^{2870}. Extensive cryptographic assessments confirm the effectiveness of this algorithm, making it a vital tool for securing medical image data within IoMT, ensuring safe transmission and storage in healthcare systems.

---

9 Jan 2025
10:20-12:20
session S1P

Enhancing X-ray Image Classification Through Neural Architecture

Hassen Louati[1], Ali Louati[2], Lilibeth Serencio Azuela[3], Meshal Alharbi[4]
[1]:h.louati@ku.edu.bh, [2]: a.louati@psau.edu.sa, [3]: l.azuela@ku.edu.bh, [4]:
mg.alharbi@psau.edu.sa

[1,3]: Kingdom University Bahrain, Bahrain
[2,4]: Prince Sattam bin Abdulaziz University, Saudi Arabia

Summary

X-ray imaging plays a crucial role in medical diagnostics, especially for detecting anomalies such as respiratory diseases. However, developing highly accurate and efficient deep learning models for X-ray image classification remains a challenge due to the need for optimal model architectures and reduced computational complexity. In this work, we propose a three-stage approach to enhance X-ray image classification performance using Neural Architecture Search (NAS), Transfer Learning,

and Model Compression through filter pruning, with a focus on the ChestX-Ray14 dataset. The first step involves using NAS to automatically identify the optimal convolutional neural network (CNN) architecture for the ChestX-Ray14 dataset. This eliminates the need for extensive manual model adjustments, ensuring a design that fits the task at hand. Next, we apply transfer learning, which uses pre-trained models to tap into features learned from large datasets. This approach helps improve the model's generalization and reduces the dependence on large quantities of labeled X-ray images. Finally, model compression is performed through filter pruning, where evolutionary algorithms are used to trim unnecessary parameters. This step enhances efficiency while maintaining the model's accuracy. Our experiments show that this approach not only boosts classification accuracy on the ChestX-Ray14 dataset but also significantly shrinks the model's size. As a result, the model becomes more suitable for deployment in environments with limited resources, such as mobile devices or edge computing platforms. This method offers a practical solution for improving both the accuracy and efficiency of medical image classification.

OCTA-Based Biomarker Characterization in nAMD

9 Jan 2025
10:20-12:20
session S1P

Ioana Damian[1], Simona Tivadar[2], Adrian Groza[3] Simona Delia Nicoara[4]
[1]:simonamaria05@yahoo.com, [2]:inteioana@yahoo.com, [3]:adrian.groza@cs.utcluj.ro

[1,4]: University of Medicine and Pharmacy, Emergency County Hospital, Romania
[2,3]:Technical University of Cluj-Napoca, Romania,

Summary

We aim to enhance ophthalmologists' decision-making when diagnosing the Neovascular Age-Related Macular Degeneration (nAMD). We developed three tools to analyze Optical Coherence Tomography Angiography images: (1) extracting biomarkers such as mCNV area and vessel density using image processing; (2) generating a 3D visualization of the neovascularization for a better view of the affected regions; and (3) applying an ensemble of three white box machine learning algorithms (decision tree, support vector machines and DL-Learner) for nAMD diagnosis. The learned expressions reached 100% accuracy for the training data and 68% accuracy in testing. The main advantage is that all the learned models white-box, which ensures explainability and transparency, allowing clinicians to better understand the decision-making process.

---

9 Jan 2025
10:20-12:20
session S1P

Revolutionary MRI Imaging for Alzheimer's: Cutting-Edge GANs and Vision Transformer Solutions

Houmem Slimi[1], Imen Cherif [2], Sabeur Abid [3], Mounir Sayadi[4]
[1]:s.houmem@gmail.com, [2]:imcherif2001@yahoo.fr, [3]:abid.saber@uvt.tn,
[4]:mounir.sayadi@ensit.u-tunis.tn

National Engineering School of Tunis, University of Tunis, Research Laboratory SIME, Tunisia

Summary

This study introduces a new approach to diagnosing Alzheimer's disease by combining Generative Adversarial Networks (GANs) with Vision Transformers (ViTs) to tackle the common issue of limited medical imaging data. GANs are used to generate synthetic MRI images of Alzheimer's patients, which are combined with real images to create a richer and more varied dataset. In this approach, the GAN model— comprising a generator and discriminator—learns to produce high-quality synthetic MRI images that, together with real data, significantly expand the training set. These images are then fed into a ViT model, which, thanks to its self-attention mechanisms, excels at identifying and classifying the stages of Alzheimer's disease. Our evaluation shows impressive results, with metrics including accuracy, F1-score, AUC, precision, and recall reaching 98.8%, 98.43%, 99.5%, 98%, and 98%, respectively showing the effectiveness of the expanded dataset in boosting classification performance. This integrated GAN-ViT approach not only enhances diagnostic accuracy for Alzheimer's but also sets the stage for applying similar techniques across other medical imaging domains, providing better tools for understanding disease progression and improving patient outcomes.

---

9 Jan 2025
10:20-12:20
session S1P

Advanced Deep Learning Strategies for Breast Cancer Image Analysis

Houmem Slimi[1], Imen Cherif [2], Sabeur Abid[3], Mounir Sayadi[4]
[1]:s.houmem@gmail.com, [2]:imcherif2001@yahoo.fr, [3]:abid.saber@uvt.tn,
[4]:mounir.sayadi@ensit.u-tunis.tn

National High Engineering School of Tunis, University of Tunis, Research Laboratory SIME, Tunisia

Summary

Breast cancer (BC) is a leading cause of cancer related deaths in women, but early detection significantly improves survival rates. Recently, deep learning neural networks have shown potential for enhancing BC screening, identification, and classification. In this research, we propose a modified deep learning model incorporating customized architecture adjustments and advanced data augmentation techniques

for early BC detection and classification from mammography images. Our model achieves a 40accuracy compared to state-of-the-art transfer learning methods on the MIAS and SA datasets, demonstrating improved robustness and precision. However, while our results are promising, the limited size and diversity of the datasets suggest that further studies are needed to validate the generalizability of the model across broader, more varied data sources.

Identifying Surgical Instruments in Pedagogical Cataract Surgery Videos through an Optimized Aggregation Network

9 Jan 2025
10:20-12:20
session S1P

Sanya Sinha[1], Michal Balazia[2], Francois Bremond[3]

[1]:sanya0904@gmail.com, [2]:michal.balazia@inria.fr, [3]:Francois.Bremond@inria.fr

[1]:Imperial College of London, United Kingdom,
[2]: INRIA, France,
[3]: Inria Sophia Antipolis, France

Summary

Instructional cataract surgery videos are crucial for ophthalmologists and trainees to observe surgical details repeatedly. This paper presents a deep learning model for realtime identification of surgical instruments in these videos, using a custom dataset scraped from open-access sources. Inspired by the architecture of YOLOV9, the model employs a Programmable Gradient Information (PGI) mechanism and a novel Generally-Optimized Efficient Layer Aggregation Network (Go- ELAN) to address the information bottleneck problem, enhancing Minimum Average Precision (mAP) at higher Non-Maximum Suppression Intersection over Union (NMS IoU) scores. The Go- ELAN YOLOV9 model, evaluated against YOLO v5, v7, v8, v9 vanilla, Laptool and DETR, achieves a superior mAP of 73.74 at IoU 0.5 on a dataset of 615 images with 10 instrument classes, demonstrating the effectiveness of the proposed model.

Enhancing Auxilliary Cancer Classification Task for Multi-task Breast Ultrasound Diagnosis Network

9 Jan 2025
10:20-12:20
Session S1P

Quan Nguyen[1], Minh Nguyen[2], Tien Dat Chung[3], Vinh Dinh[4]

[1]:quannd@ptit.edu.vn, [2]:minhdc1712@gmail.com, [3]:18119014@student.hcmute.edu.vn,
[4]:vinh.dinhquang@aivietnam.edu.vn

[1]:Posts and Telecommunications Institute of Technology, Vietnam
[2]:University of Medicine and Pharmacy Ho Chi Minh city, Vietnam
[3]:University of Technology and Education Vietnam, Vietnam
[4]:AI Vietnam, Vietnam

Summary

Breast ultrasound is a non-invasive, economical method that is essential for the diagnosis of cancer. With the advent of deep learning in recent years, numerous CNN-based methods have been thoroughly studied for tasks including tumor localization and cancer classification. While prior single models performed well in both challenges, these approaches had several drawbacks, including long inference times, a need for a GPU, and the need for individual model fine-tuning. Our goal in this work is to create a new end-to-end multi-task architecture that can be used for both segmentation and classification. We obtained exceptional performance and time efficiency with our suggested method, achieving 79.8% and 86.4% in the segmentation chal lenge in DeepLabV3+ architecture.

Keynote II by Fabio Solari Session chair: Prof. François Brémond (Thursday 9 Jan 2025 at 12:20-13:00)

Bioinspired computer vision for effective extended reality applications

9 January 2025
12:20-13:00

Fabio Solari
University of Genoa, Italy

Summary

Bio-inspired computational models of visual perception can be useful tools to design extended reality (XR) systems better, encompassing augmented reality (AR), mixed reality (MR), and virtual reality (VR). The proposed neural models are based mainly on dorsal visual processing stream paradigms, considering also recent deep learning techniques. Besides artificial vision applications, the proposed models can mimic and describe human behavioral data, such as motion and depth perception. By leveraging previous outcomes, we can employ the modeled perception to improve the design of XR environments by providing an enhanced user experience.

Fabio Solari biography

Fabio Solari is Associate Professor of Computer Science at the Department of Informatics, Bioengineering, Robotics and Systems Engineering of the University of Genoa. His research activity concerns the study of visual perception with the aim to design novel bio-inspired artificial vision systems and to develop natural human-computer interaction techniques in virtual and augmented reality. In particular, his research interests are related to: (i) neural algorithms for motion and depth estimation, space-variant visual processing and scene interpretation; (ii) perceptual assessment of virtual/augmented reality systems and the development of systems that allow a natural experience and ecological human-computer interactions

in mixed reality environments. He teaches "Augmented Reality" for the master of Computer Science, "Software Technologies for HCI" for the master of Bioengineering, and "Computer Vision" for the European Master on Advanced Robotics. He is Member of the Board of the Doctoral Course in Bioengineering and Robotics at University of Genoa, and Member of the Faculty Joint Commission for teaching and the right to study. He has participated to eight European projects: Interreg Alcotra CLIP "E-Santé/Silver Economy", PROSOL "Jeune" and PROSOL "Senior"; FP7-ICT, EYESHOTS and SEARISE; FP6-IST-FET, DRIVSCO; FP6-NEST, MC-COOP; FP5-IST-FET, ECOVISION. He is a reviewer for Italian PRIN and FIRB projects, and EU Marie Curie fellowships and ERC. He has a pending International Patent Application (WO2013088390) on augmented reality, and two Italian Patent Applications on virtual (No. 0001423036) and augmented (No. 0001409382) reality. More information is available at http://sites.google.com/site/fabiosolarifs/

Keynote III by Prof. Véronique Eglin
Session chair: Prof. Dorra Sellami (Thursday 9 Jan 2025 at 14:00-14:40)

Tracing the Evolution of Text/Image Multimodal Techniques: From OCR to Advanced Vision-Language Models

9 Jan 2025
14:00-14:40

Véronique Eglin

LIRIS, University Lyon 2, France

Summary

Document analysis has seen significant progress over years, thanks to advances in text and image processing techniques. Early methods focused on OCR for text extraction and simple rule-based systems to combine text and image features, but these were limited in their ability to handle complex documents. The introduction of machine learning in the mid-2000s brought feature-level fusion, where handcrafted features from both text and image were combined, though still with some limitations. The advent of deep learning in the 2010s transforms the landscape, enabling neural networks to process and integrate multimodal data more effectively. Transformers from the LayoutLM family pushed the boundaries further by using spatial and semantic relationships within documents, enabling tasks like table recognition, information extraction and form understanding. More recently, models like Donut and PaLM-E have removed the dependance to OCR by processing documents directly in an end-to-end way. These newer approaches represent major steps forward, but challenges like handling complex layouts, scaling models, and improving interpretability remain.

Véronique Eglin biography

Véronique Eglin is a Professor of Computer Science at INSA Lyon (National Institute of Applied Sciences of Lyon) in France, within the CNU 27 section. She leads the Imagine research team at the LIRIS (Laboratoire d'InfoRmatique en Image et Systèmes d'information) laboratory, which is a joint research unit of CNRS (National Center for Scientific Research) and INSA. She is also involved in teaching computer science to first-year students at INSA and serves as the Deputy Director of the First Cycle (undergraduate) program. Additionally, she is a member of the INSA's Council of Studies. She earned her Habilitation à diriger les recherches (HDR - accreditation to supervise research) in June 2014.

Her research focuses on the analysis of document images, particularly by leveraging principles of human visual perception. Her work includes low-level processing, such as segmentation and structure analysis, as well as content identification through classification. She is also involved in indexing and accessing content within manuscript images, using techniques such as shape signatures and content-based retrieval.

Regular Session S2P: Special Session: Workshop on AI methods of video analysis for human behavior recognition
Session chair: Prof. Serge Miguet (Thursday 9 Jan 2025 at 14:40-16:00)

Detection of Cyberattacks on the Metaverse based on First-order logic

9 Jan 2025
14:40-16:00
Session S2P

Sulaiman Alamro
samro@qu.edu.sa

Qassim University, Saudi Arabia

Summary

There are currently considerable challenges concerning data security and privacy, particularly in relation to modern technologies. This includes the virtual world known as the Metaverse, which consists of a virtual space that integrates various technologies, and therefore susceptible to cyber threats such as malware, phishing, and identity theft. This has led recent studies to propose the development of Metaverse forensic frameworks and the integration of advanced technologies, including machine learning for intrusion detection and security. In this context, the application of first-order logic offers a formal and systematic approach to defining the conditions of cyberattacks, thereby contributing to the development of effective detection mechanisms. In addition, formalizing the rules and patterns of cyber threats has the potential to enhance the overall security posture of the Metaverse and thus the integrity and safety of this virtual environment. The current paper focuses on the primary actions employed by avatars for potential attacks, including Interval Temporal Logic (ITL) and behavior-based detection to detect an avatar's abnormal activities within the Metaverse. The research established that the proposed

framework attained an accuracy of 92.307%, resulting in the experimental results demonstrating the efficacy of ITL, including its superior performance in addressing the threats posed by avatars within the Metaverse domain.

---

9 Jan 2025
14:40-16:00
Session S2P

Cricket Shot Analysis using Conditional Directed Spatio-Temporal Graph networks

Tanu Aneja[1], Brejesh Lall[2], Harsha Malaviya[3]

[1]: eez158111@ee.iitd.ac.in, [2]: brejesh@ee.iitd.ac.in,
[3]: malaviya.harsha25@gmail.com

[1,2]: IIT Delhi, India
[3]: Monta Vista High School, United States

Summary

Recent Use of Conditional Spatio-temporal Directed Graph Convolutional Networks(Cond ST-DGCN) to represent human pose estimation has significantly helped in capturing varying non-local dependencies between limbs for different actions. This can be immensely helpful in Sports analytics where player pose plays key role in shot evaluation and can help in corrective action. In this article, we propose Cond-DGCN[based framework to explore use of spatial-temporal relation of batsman shot sequences (labelled and annotated 2D cricket dataset) for Cricket shot action recognition by conditioning the graph network on batsman 2D poses. We achieve 97% accuracy for shot recognition and further explore visualization of conditional graph connections to establish importance of particular limbs for shots. The proposed framework uses fine-tuned 2D Pose estimator OpenPose(fine-tuned for cricket dataset) which in turn helps in easy adaptation of our solution to internet cricket videos for shot analytics.

---

9 Jan 2025
14:40-16:00
Session S2P

Occlusion-aware Data Augmentation and Reconstruction for Improving Student Engagement in Online Learning Environments

Vincent Pham[1], Sonit Singh[2]

[1]: vincent.pham2@outlook.com, [2]: sonit.singh@unsw.edu.au

University of New South Wales, Sydney, Australia

Summary

The demand for adaptive learning in education is an important application for Computer Vision (CV)-based detection models which extract learners' faces to classify engagement. However, a loss in visual data due to the partial occlusion of learners' faces imposes challenge for practical use-cases. In this paper, we propose an

occlusion-aware framework to improve the robustness of non-occlusion-aware models in engagement detection. Firstly, our framework consists of an occlusion-aware data augmentation pipeline that aims to simulate the types of partially occluded faces in-the-wild. Secondly, we investigate the application of masked auto-encoders (MAE) for occlusion recovery. Extensive experiments have been performed to measure the effectiveness of the proposed occlusion-aware framework. Specifically, on the challenging FER-2013 dataset our occlusion-aware data augmentation achieves a 5.24% improvement in accuracy on baseline models and a 1.54% improvement against models with augmentation methods used in the state-of-the-art. Alongside the application of MAE, our proposed framework improves baseline classification accuracy by 13.3% and 4.59% on the FER-2013, and DAiSEE dataset for engagement detection. Our investigation into occlusion-aware engagement detection using the DAiSEE dataset provides important insight into the limitations of occlusion-awareness in engagement detection.

Automatic recognition of human psychological state on the basis of EEG-based data

9 Jan 2025
14:40-16:00
Session S2P

Kartlos Kachiashvili[1], Joseph Kachiashvili[2], Vakhtang Kvaratskhelia[3]
[1]:k.kachiashvili@gtu.ge, [2]:s.kachiashvili@gtu.ge, [3]:v.kvaratskhelia@gtu.ge

Muskhelishvili Institute of Computational Mathematics, Georgian Technical University, Georgia

Summary

The paper deals with the problem of binary classification of EEG data at restricted computing and time resources. Eleven different criteria of similarity of multivariate time series (MTS) were used for this purpose. Nine of them are widely known, two are introduced in this work. On the basis of the computation results of 32 dimensional EEG signals was established the priority among the considered methods. A method based on the weighted Euclidean distance between the eigenvectors of the covariance matrices, as well as the method based on the trends of MTS and the method obtained by combining the previous two methods gave us better results in terms of classification accuracy than others.

Regular Session S3P: Computer Vision and Deep Learning in Smart Transportation Systems: From Road Defect Detection to Satellite-Based Monitoring Session chair: Prof. François Brémond (Thursday 9 Jan 2025 at 16:10-17:50)

Wide Line Segments Detection in Grey-Level Images via Guided Scale Space Radon Transform

09 Jan 2025
16:10-17:50
Session S3P

Aicha Baya Goumeidane[1], Djemel Ziou[2], Nafaa Nacereddine[3], Nawal Yala[4]

[1]:a.goumeidane@crti.dz, [2]:DjemelZiou@USherbrooke.ca,
[3]:n.nacerddine@crti.dz, [4]:n.yala@crti.dz

[1,3,4]:Research Center in Industrial Technologies (CRTI), Algeria
[2]:University of Sherbrooke, Canada

Summary

Line segment detection is a fundamental procedure in computer vision, pattern recognition, and image analysis appli cations. The paper proposes a novel method for wide line segment detection especially endpoints determination based on the Guided Scale Space Radon Transform and Hessian orientations. The method begins by determining the centerlines of wide lines and then exploit the image Hessian orientations around these lines to define binary region support of the line segments and then detect endpoints. The method shows to be robust against blur and noise on synthetic images where, the evaluation of the outcomes reveals the correctness

of the detection by achieving low errors. In addition, results on real images are very promising.

A Segmentation-Based Approach for Road Surface Defect Detection

09 Jan 2025
16:10-17:50
Session S3P

Norah Alsubaie[1], Ghayda Almalki[2], Sarah Alrumaih[3], Ghada Almutairi[4]
[1]:noraalsobuee11@gmail.com, [2]:GhaydaAlmalki@outlook.com,
[3]:sarah.alrumaih21@gmail.com, [4]:ghadansalmutairi@gmail.com

Princess Norah bint Abdurhaman University, Saudi Arabia

Summary

This research introduces "Jaddah," an innovative AI-based system for the automated detection of road infrastructure defects using advanced computer vision and machine learning techniques. The system addresses the limitations of traditional road inspection methods, which are often slow and prone to human error. Jaddah develops a mobile application that efficiently detects, classifies, and segments road defects at the pixel level. By utilizing a comprehensive dataset of high-resolution images, the model training process is significantly enhanced. The YOLOv8-seg model is implemented to achieve precise defect localization and segmentation, ensuring high accuracy in identifying and categorizing road defects. Performance metrics show an impressive 87% mAP50, demonstrating reliable defect detection. These results contribute to improved infrastructure maintenance, enhanced road safety, and greater operational efficiency.

Patch based image processing for complex environment characterization

09 Jan 2025
16:10-17:50
Session S3P

Corentin Menier[1], Cyril Meurie[2], Timothee Guillemaillle[3], Yassine Ruichek[4], Juliette Marais[5]
[1]:corentin.menier@univ-eiffel.fr, [2]:cyril.meurie@univ-eiffel.fr,
[3]:timothee.guillemaille@univ-eiffel.fr, [4]: yassine.ruichek@utbm.fr,
[5]: juliette.marais@univ-eiffel.fr

[1,2,3,5]:Univ Gustave Eiffel, COSYS-LEOST, F-59650 Villeneuve d'Ascq, France
[4]:UTBM, CIAD UMR 7533, F-90010 Belfort, France

Summary

Environment analysis is a critical part of au tonomous vehicle for transport applications and for passenger safety. The solutions demonstrating the greatest robustness have been integrating multiple sensors used for redundancy and refinement purposes.

Vision applications have proven to offer a high degree of flexibility and performance. One particular instance of this is higlighted in vehicle localisation, which predominantly relies on GNSS-based systems for positioning calculation using propagation time measurements. However, this signal may be degraded through the environment around the vehicle, worst case being urban canyons leading to Non Line Of Sight(NLOS) scenarios or multipaths issues due to reflecting obstacles. Previous work have shown vision-based algorithms can be used to mitigate these effects. One widely studied approach relies on the segmentation of an acquired wide-angle image installed on the roof of the vehicle and oriented toward the sky. Because the sky processing module is binary, the pipeline lack any way to express its uncertainty when applying weighting policies to the detected satellite state, which can be detrimental to the resulting positioning. In this paper, we propose a novel way of analysing wide-angle camera images, also known as fisheye images, dividing the image into patches to output the corresponding situation of each region of interest. Additionally we propose a new class to the previous sky versus non-sky segmentation, designated as mixed class and designed to serve as a fuzzy answer by the deep learning model to improve confidence to other scenarios as well as allow for new analysis policies of satellites signals. The data-driven algorithm is designed and tested on a publicly available dataset, composed of a large number of finely labelled images provided by ISAE-SUPAERO reaching a 94% accuracy.

On the use of Vision for the Weighting of GNSS observation: comparisons

09 Jan 2025
16:10-17:50
Session S3P

Zhiye Cheng[1], Timothee Guillemaillle[2], Corentin Menier[3], Cyril Meurie[4], Juliette Marais[5]

[1]:zhiye.cheng@univ-eiffel.fr, [2]:timothee.guillemaille@univ-eiffel.fr,
[3]:corentin.menier@univ-eiffel.fr, [4]: cyril.meurie@univ-eiffel.fr,
[5]: juliette.marais@univ-eiffel.fr

[1,2,3,4,5]:Univ Gustave Eiffel, COSYS-LEOST, F-59650 Villeneuve d'Ascq, France

Summary

Positioning is a critical function in every intelligent vehicle application. Most applied system is a GNSS (Global Navigation Satellite System)-based receiver, cheap and offering a continuous meter-level accuracy. However accuracy strongly depends on the satellite signal reception state: LOS (Line of sight), that is direct, optimal signal or NLOS (Non LOS), i.e received without direct visibilty and after one or more reflections of the signal. Based on previous work, this paper summarizes different weighting schemes applied to mitigate these local effects on GNSSsignals and enhance position accuracy in land transport environment. A database collected by ISAE is used for application and comparison of the different schemes. One of them not only relies on GNSS signals but also on satellite state identification thanks to the use of a fisheye camera. This state allows us to deweight degraded measurements without excluding them, in order to keep availability. The paper

shows that considering this additional information allows the WLS (Weighting Least Square) to significantly increase accuracy in every types of environments. If state of the art weighting schemes already improve accuracy by 65 and 77% over ordinary least squares in the 2D plan, our weight further improves accuracy by 40% compared to the classical elevation-based weight.

---

09 Jan 2025
16:10-17:50
Session S3P

HPDNET: Hyper Prior Dependent Demosaic Neural Network

Youngil Seo[1], Sungho Jun[2], Dongpan Lim[3], Seongwook Song[4]
[1]:kumi012@gmail.com, [2]:sh1248.jun@samsung.com, [3]:dongpan.lim@samsung.com,
[4]:seongwook.song@samsung.com

Samsung Electronics Yongin, Republic of Korea

Summary

Recent advancements in deep neural networks have shown remarkable improvements in image quality during the demosaicking process, surpassing conventional algorithms. How ever, these deep neural network techniques are often character ized by heavy computational requirements, rendering them un suitable for deployment on resource-constrained platforms. This presents a critical challenge in the field of image demosaicking: while deep learning approaches excel in enhancing image quality, their computational intensity poses a significant hindrance to their wider adoption. Consequently, there is a pressing need for methodologies that can strike an optimal balance between achieving superior image quality and maintaining computational efficiency. In this work, we propose a new deep framework, hyper-prior dependent demosaic neural network, HPDNet that utilizes the significant concepts of the conventional algorithm and the characteristic of image sensor data. We designed the network that exploits three concepts, those are pixel gradient prior attention, phase separation, and multi-level sparse and dense feature extraction. We designed the deep neural network that extracts the optimal gradient prior and the multi-level extracted features are fused and attended by gradient prior. It can fully utilize spatially variant information. Also to make the network deployed in the mobile platform, we devised self-pruned image convolution that adopts image filter characteristic and reduce computations. Experiments show that proposed network outperforms SOTA demosaic networks both in terms of image quality and computation.

Special Session S4P: Revolutionizing Waste Management: AI-Driven Innovations for Efficient Sorting and Recycling
Session chair: Prof. Serge Miguet (Friday 10 Jan 2025 at 09:00-10:30)

Invited Talk on Revolutionizing Waste Management: New trends towards efficient sorting and Recycling

10 Jan 2025
09:00-10:30
Session S4P

Dr. Rémi Cuingnet

remi.cuingnet@veolia.com

Veolia, France

---

10 Jan 2025
09:00-10:30
Session S4P

Solid waste classification towards a Recycling Process: A review

Mouna Zouari[1], Jihen Frikha Elleuch[2], Dorra Sellami[3]
[1]: zouarimouna88@gmail.com, [2]:jihen.frikha@fsegs.usf.tn, [3]: dorra.sellami@enis.tn

[1,2]:National Engineering School of Sfax, Tunisia
[2]: Sfax Digital Research Center, Tunisia

Summary Waste management is becoming an intriguing problem all over the word. This latter impacts all the environment's components. Trying to reduce the huge number of trashes all over the word, recycling them may be an efficient tools especially when the sorting process is done automatically. As this later is one of the principle goal of the circular economy, many researches have concentrated on the development of such automatic solution to save the environment. This paper is dedicated to review the basic IA technics used in the literature in order to classify waste and also the main databases used in order to validate the developed solutions.

---

10 Jan 2025
09:00-10:30
Session S4P

Development of an Embedded System for Plastic Waste Segregation

Wejdene Smari[1], Mouna Zouari[2], Jihen Frikha Elleuch[3], Dorra Sellami[4], Ahmed Fakhfakh[5]
[1]:wejden.smari@enis.tn, [2]:zouarimouna88@gmail.com, [3]:jihen.frikha@fsegs.usf.tn, [4]:dorra.sellami@enis.tn, [5]:ahmed.fakhfakh@enetcom.usf.tn

[1,2,3,4]:National Engineering School of Sfax, Tunisia
[2,5]:Sfax Digital Research Center, Tunisia

Summary This paper presents a novel approach to plastic waste segregation, based on two modalities: spectroscopy and optical images. The research focuses on leveraging an optical camera for efficient waste image acquisition, while the ground truth for plastic type identification (with respect to the seven primary plastic types: PET, HDPE, LDPE, PP, PS, PVC, and Other [cite relevant source here]) is established through spectro-scopic analysis. Various deep learning models were trained and evaluated on a dataset of publicly available and custom-collected images, aiming to achieve human-level accuracy in distinguishing between plastic types. The proposed methodology is based on a set of steps, where the model is first built implicitly on a segregation of different wastes, and then applied via transfer learning for separating the seven plastic types. The system performances demonstrate the feasibility and effectiveness of this combined approach for accurate and efficient plastic waste sorting. The proposed system has been implemented on Raspberry Pi 4 Model B+ as the central processing unit, integrating computer vision and deep learning techniques for efficient and accurate sorting.

Waste Sorting System Using AI: Development of An Embedded Solution for Efficient Waste Management

10 Jan 2025
09-10:30
Session S4P

Amal Remili[1], Mouna Zouari[2], Jihen Frikha Elleuch[3], Dorra Sellami[4], Ahmed Fakhfakh[5]

[1]:amal.remili@enis.tn, [2]:zouarimouna88@gmail.com, [3]:jihen.frikha@fsegs.usf.tn,
[4]: dorra.sellami@enis.tn [5]:ahmed.fakhfakh@enetcom.usf.tn

[1,2,3,4]:National Engineering School of Sfax, Tunisia
[2,5]: Sfax Digital Research Center

Summary This project presents an AI-based system for waste identification and sorting, particularly focusing on plastic, using computer vision techniques. The data, employed for system validation, comprises publicly available datasets and a locally collected dataset specifically generated for this project, devoted to plastic classification and detection. Performance metrics demonstrate that YOLOv8 outperforms other deep learning solutions, in real-time waste detection and is well-suited for deployment on a Raspberry Pi 4 Model B. The system outperforms also existing methods in terms of efficiency, accuracy, and speed, and is a promising solution for addressing the global waste management challenge.

Keynote IV by Prof. Riadh Abdelfattah
Session chair: Prof. Fabio Solari (Friday 10 Jan 2025 at 10:40-11:20)

SAR data analysis for disaster management

10 Jan 2025
10H40-11H20

Riadh Abdelfattah
SUP'COM, University of Carthage, Tunisia
riadh.abdelfattah@supcom.tn

Summary

Synthetic Aperture Radar (SAR) has emerged as a crucial asset in disaster management, offering vital data for the monitoring and response to both natural and anthropogenic disasters. This keynote will explore the significant role of SAR data analysis in enhancing our capabilities to detect, assess, and mitigate the impacts of catastrophic events such as earthquakes, floods, and landslides. We will explore cutting-edge advancements in SAR technology, including interferometry and change detection techniques, which provide timely and precise information essential for decision-makers. Case studies will illustrate the practical applications of these technologies in real-world scenarios, showcasing how SAR can effectively monitor disaster-prone areas and assess damage in the aftermath of events. Additionally, the discussion will address the integration of SAR data with other geospatial data sources, highlighting the challenges and methodologies involved in data interpretation. The incorporation of machine learning and high-performance computing (HPC) will also be examined, emphasizing their transformative potential in processing multidimensional SAR data for improved disaster response outcomes. As we look to the future, this keynote will offer insights into the evolving landscape of SAR applications in disaster management, underscoring the importance of developing resilient systems that can adapt to the increasing frequency and intensity of disasters. By fostering collaboration among researchers, practitioners, and policymakers, we aim to enhance our collective capacity to mitigate the impacts of disasters and support recovery efforts effectively.

Riadh Abdelfattah biography

Dr. Riadh Abdelfattah is Professor at the Higher School of Engineering in Communications (SUP'COM) at the University of Carthage In Tunisia. He was the President, and a Vice- President of the University of Carthage (2017-2020), in charge of research activities, technologic development and environmental partnership. He is also Associate Researcher at the Department ITI (Image Traitement de l'Information) at IMT-Atlantique, the "Institut de Télécom", Brest, France. He was member of the scientific council of AUF (Agence Universitaire de la Francophonie) and member of the Expert Regional Committee (2016- 2020). He received the engineer degree from the Telecommunication Engineering School of Tunis, Tunisia in 1995, the Master Degree (DEA) and the the Ph.D degree in Electrical Engineering from the " Ecole Nationale Ingénieurs de Tunis", in 1995 and 2000 respectively, and "le Diplôme de l'Habilitation Universitaire" from the Higher School of Communications (SUP'COM) at the University of Carthage in Tunisia (2008). Between 2000 and 2002 he was a postdoctoral researcher at the " Ecole Nationale des Télécommunications", Paris, France consecutively at the department TSI and then at the department COMELEC. He is a senior member of the IEEE and he served as a member of the Executive Committee of the IEEE Tunisia Section (2013-2015). He has authored and co-authored more than 80 journal papers, conference papers and book chapters. His main research interests include interferometric radar imagining, multiemporal and multiscale image analysis, desertification, flooding and soil salinity mapping from remote sensed data, applied AI for water resource management and SAR-nanosatellite development.

Regular Session S5P: Deep Learning Applications in Forgery Detection, Gesture & Activity Recognition
Session chair: Prof. Riadh Abdelfattah
(Friday 10 Jan 2025 at 11:20-13:00)

Invited Talk By Prof. Jan Alexandersson on Advancing Artificial Empathic Technology: Building an on-premise GDPR-compliant Trusted Research Environments

10 Jan 2025
11:20-13:00
Session S5P

Jan Alexandersson

jan.alexandersson@dfki.de

Head AAL Competence Center DFKI GmbH, Campus D3 2 Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany

---

10 Jan 2025
11:20-13:00
Session S5P

You're not acting like yourself: Deepfake Detection Based on Facial Behavior

Alexandre Libourel[1], Jean-Luc Dugelay[2]
[1]:libourel@eurecom.fr, [2]:jld@eurecom.fr

[1]:Eurecom
[2]:France

Summary Politicians and government leaders are critical targets for deepfake attacks. A single deepfake involving these individuals can severely damage their careers or, in extreme cases, pose a national security threat. Attackers can leverage vast amounts of publicly available audio and video recordings to train their models, making this threat even more pressing. In response, specialized deepfake detectors have been developed to focus on detecting deepfakes targeting a specific Person of Interest (POI). By learning facial expressions and movements unique to the POI, these detectors can identify inconsistencies in deepfakes where these authentic attributes are absent. However, previous methods relied on Facial Action Units, which offer an incomplete representation of the POI's behavior. In this paper, we propose a novel approach to learning POI-specific movements without requiring deepfake samples during training, making it independent of any deepfake generation methods. Although our technique is speaker-dependent, it provides a robust solution for protecting high-profile individuals who are particularly exposed to deepfake threats.

---

10 Jan 2025
11:20-13:00
Session S5P

A Review On Fusion Of Spiking Neural Networks And Transformers

Oumaima Marsi[1], Sébastien Ambellouis[2], José Mennesson[3], Cyril Meurie[4], Anthony Fleury[5], Charles Tatkeu[6]
[1]: oumaima.marsi@univ-eiffel.fr, [2]: sebastien.ambellouis@univ-eiffel.fr,
[3]:jose.mennesson@imt-nord-europe.fr, [4]:cyril.meurie@univ-eiffel.fr,
[5]:anthony.fleury@imt-nord-europe.fr, [6]:charles.tatkeu@univ-eiffel.fr

[1,2,4,6]:Univ. Gustave Eiffel, COSYS, LEOST, France,
[1,3,5]: IMT Nord Europe, Institut Mines-Telecom, France

Summary This paper provides a comprehensive review of the integration of Spiking Neural Networks (SNNs) and Transformers, combining the energy efficiency of SNNs with the high performance of Transformer architectures. By leveraging the event-driven nature of SNNs and the powerful self-attention mechanism of Transformers, this fusion aims to address the challenges of high energy consumption in deep learning while improving task accuracy, especially for complex datasets. We introduce the core concepts of SNNs and Transformers, reviewing state-of-the-art

methods for their combination, including hybrid architectures like Spikformer, Spikingformer, Spike-driven Transformer and SpikingResformer. The performance of these models is evaluated across both static and neuromorphic datasets, highlighting their advantages and limitations. This review also discusses the challenges of integrating self-attention into spiking architectures and outlines future research directions to further enhance model performance and energy efficiency.

Unsupervised domain adaptation in human detection for various human pose

10 Jan 2025
11:20-13:00
Session S5P

Benoit Lagadec[1], Ekaterina Kostrykina[2]

[1]:benoit.lagadec@esifrance.net, [2]:ekaterina.kostrykina@etu.univ-cotedazur.fr

[1]:European System Integration, France,
[2]:Universite Cote d'Azur, France

Summary In human detection algorithms, some recurrent issues are still challenging. Indeed, in a real application we have a lack of various human poses in our data. For example, in video surveillance it could be a significant issue to forget detecting a human in uncommon pose. We propose a compensation to these less frequent cases. We introduce a detection in rotated view to compensate limited various poses in our dataset. Instead of introducing a generative enhancement for every human body pose, which could be exhausting. Indeed, data annotation has become a difficulty for many applications. Moreover a simple data augmentation is not efficient to generalize our training. We need a strategy to learn this new poses. So, we will introduce a process to compute pseudo labels in a sequence of rotated views. It avoids to use a clustering algorithm like dbscan or K-means to compute a pseudo label. A resume of theoretical results is described in a first section to illustrate our baseline. First of all, we will illustrate how we process pseudo label on various views by two mechanisms which have been called relaxation and scanning. We introduce a sequence of rotations of original view, to process a detection in a compensated view. In second step, we will learn this distortion created by the various views and construct an approach to distillate originals knowledges to a more generalized training. The main improvement of this approach is to propose an alternative solution at dbscan approach which is popular, but which is difficult to use in real applications. We obtain a significant result by improving accuracy in especially various pose datasets.

10 Jan 2025
11:20-13:00
Session S5P

Activity Recognition for Mechanical Systems with Complex Kinematics

Ala'a Alshubbak[1], Daniel Goerges[2]
[1]:shubbak@rptu.de, [2]:daniel.goerges@rptu.de

[1,2]:Institute of Electromobility, University of Kaiserslautern-Landau, Kaiserslautern, Germany
[1]:German Jordanian University, Amman, Jordan

Summary Activity recognition is one of the major tasks in computer science and engineering that aims to recognize and understand the actions of one or more agents from a series of observations. It assists different sectors in ensuring safety in human-machine interaction. This paper investigates and tests an encoder classifier based on an LSTM model with various attention mechanism structures over a synthetic time-series motion dataset generated with the NVIDIA Isaac simulator. The simulator simulates the motion of an excavator as an example of a complex kinematic system in a construction environment, gathering data from position sensors connected to the excavator's main joints. The results show that fewer sensors can be used for certain types of motion classification with a large accuracy range between 66.7% and 83.3%. The encoder LSTM model with scaled dot-product attention gives the most accurate results compared to other attention mechanism types, with around 33.4% when using data from only two sensors and around 22.2% for the four main sensors. These results are in comparison to models that used other attention mechanisms.

Keynote V by Prof. Atika Rivenq Session chair: Prof. Bertrand Kerautret (Friday 10 Jan 2025 at 14:00-14:40)

New trends on signal and image processing for Cooperative Intelligent Transport Systems (C-ITS) : An overview of cooperative perception, IA tools and advanced communication for automated vehicles

5 Dec 2022
14:00-14:40

Atika Rivenq Menhaj
IEMN site de Valenciennes, France
INSA Hauts de France, France

Summary

Cooperative Intelligent Transport Systems (C-ITS) systems are based on information from direct interaction between vehicles or between vehicles and road infrastructure (V2I, V2V). At the European level, new structures have been created to develop specific standards (ETSI) for these systems. They allow road users and traffic managers to share information and use it to coordinate their actions. This cooperative element enabled by digital connectivity between vehicles and between vehicles and infrastructure aims to significantly improve road safety, traffic efficiency and driving comfort, to help the driver make the right decisions and adapt to the traffic situation. These connected systems will improve air quality and reduce pollution rates. Many major projects have been launched by the European Union and in collaboration with the European Telecommunications Standards Institute ETSI (European Telecommunications Standards Institute). These projects have been adopted for the establishment of the bases of a cooperative intelligent transport system (C-ITS). Among these projects, we can cite: Scoop, C-Roads, InterCor, INDID, FENIX and SCALE. All these projects will contribute to the European "Sustainable and Smart Mobility Strategy" and indirectly to the European Green Deal. C-ITS sites communicate using data from cellular networks (4G, 5G, ...) and/or IEEE 802.11p (designed by ITS-G5 in Europe). Cooperative perception will ensure more safety

and traffic management for connected vehicles and also for automated vehicles. This tutorial aims to provide insights into vehicular communications research's state of the art and challenges of this type of communication. Many topics related to C-ITS will be drawn by this tutorial (ITS-G5 characteristics, Hybrid communications, Security, Cryptography, Channel Estimation, AI contribution, Services, Cartography HD).

Atika Rivenq Menhaj biography

Professor Atika Rivenq-Menhaj was born in 1970. She was graduated Engineer from the ENSI-MEV School in 1993, received the M.S. Degree in 1993 and then her Ph.D. Degree in 1996 from the University of Valenciennes (France) She is Professor in electronics at this university with the IEMNDOAE Lab. Her main activity is in signal processing applied to intelligent transport systems (localisation, collision avoidance radar, cruise control associated to GPS). She participated to many national and European projects dedicated to guided-train communication and inter-vehicles communication especially using UWB technology.

Regular Session S6P: Machine Learning for Data-Driven Discovery: From Signal Processing to Image Analysis
Session chair: Prof. Sabeur Abid (Friday 10 Jan 2025 at 14:40-16:00)

Conceptualisation of a Parameterizable Low-Pass Filter for Resolving Measurement Data

10 Jan 2025
14:40-16:00
Session S6P

Michael Schmid[1], Bastian Eisenmann[2], Osman Aksu[3], Misel Radosavac[4], Florian Bierwirth[5], Hans-Georg Herzog[6]

[1]:michael.pe.schmid@bmw.de, [2]:bastian.ba.eisenmann@bmw.de, [3]:osman.aksu@bmw.de
[4]:misel.radosavac@bmw.de, [5]:florian.bierwirth@bmw.de, [6]:hg.herzog@tum.de

BMW AG, Germany

Summary The data acquisition process through measurements translates continuous physical quantities, such as temperature or magnetic flux, into discrete digital data sets. Different noise sources, such as quantization or Gaussian noise, distort the sampled points and inhibit subsequent processing during this process. Filtering the measured data is, thereby, a necessary step. In contrast to conventional filter design, this paper describes a novel discrete low-pass filter with a variable shape. It is configured by two independent parameters, allowing it to be sensitive to time-critical peaks at high levels while providing adequate noise suppression at low measured quantities. The theoretical background of this work's filter approach is explained, and guidelines for parameter selection are suggested. In the utilized test data set, the novel approach results in a 19% better peak preservation while providing the same noise suppression as a conventional low-pass filter.

10 Jan 2025
14:40-16:00
Session S6P

Text Recognition in Persian Documents Using Self-Supervised Learning

Atefeh Babaei[1], Alireza Dehghani[2], Mehran Yazdi Shiraz[3], Ali Jamshidi[4], Mahdi Rasouli[5]

[1]:a.babaei@hafez.shirazu.ac.ir, [2]:alireza.dehghani@ucd.ie, [3]: yazdi@shirazu.ac.ir,
[4]:jamshidi@shirazu.ac.ir,
[5]:mahdi.rasouli@etu.univ-grenoble-alpes.fr

[1,3,4]:Shiraz University, Iran
[2]:CeADAR, Ireland Centre for AI, Ireland
[5]:School of Computer Science and Mathematics Universite Grenoble Alpes Grenoble, France

Summary Text is one of humankind's most significant inventions essential for communication and collaboration in modern society. Extracting text from images, especially for languages with cursive and connected scripts like Persian, presents unique challenges due to issues such as font types, word structures, and the scarcity of labeled data. This paper addresses these challenges by implementing deep learning techniques, specifically self-supervised learning with convolutional and recurrent neural networks (CNNs and RNNs). Our approach successfully mitigates the data scarcity problem, achieving an accuracy of 92.9% on the prepared dataset.

---

10 Jan 2025
14:40-16:00
Session S6P

Data procesing, analysis and Machine Learning of Electrical Atomic Force Microscopy (AFM) in DataCube mode: conduction of a nanomaterial for Lithium-ion battery electrodes

Rosine Coq Germanicus[1], Peter De Wolf[2], Petr Klapetek[3]

[1]: rosine.germanicus@unicaen.fr, [2]: peter.dewolf@bruker.com,
[3]: pklapetek@cmi.gov.cz

[1]: University of Caen, France
[2]: Bruker Nano Surfaces, United States
[3]: Czech Metrology Institute, Czechia

Summary With the increasing integration of functional systems, nanoscale characterization has become crucial not only for material investigation but also for advancing the understanding of material behavior and optimizing performance. This paper details data processing techniques and machine learning methodologies for the analysis and prediction of advanced electrical Atomic Force Microscopy datasets. The datasets are acquired using a hyperspectral DataCube approach, which combines both mapping and spectroscopy acquisitions. In the context of decarbonization and electrification, current challenges are related to energy storage and transport. A

cathode material used in Lithium-ion battery electrodes is investigated for its electrical transport properties. To comprehensively characterize the electrical properties, the conductive mode of Atomic Force Microscopy is employed using the DataCube method, which enables current-voltage spectroscopy at each pixel of the acquisition. The analysis of experimental datasets reveals localized electrical behavior of the material's components as a function of applied voltage. Furthermore, machine learning algorithms are employed to predict current maps and cluster distinct regions based on their electrical characteristics.

Low-Data Classification of Historical Music Manuscripts: A Few-Shot Learning Approach

10 Jan 2025
14:40-16:00
Session S6P

Elona Shatri[1], Daniel Raymond[2], George Fazekas[3]

[1]: e.shatri@qmul.ac.uk, [2]: d.k.raymond@se23.qmul.ac.uk, [3]: george.fazekas@qmul.ac.uk

[1,2]: Queen Mary University of London, United Kingdom
[3]: QMUL, United Kingdom

Summary In this paper, we explore the intersection of technology and cultural preservation by developing a self-supervised learning framework for the classification of musical symbols in historical manuscripts. Optical Music Recognition (OMR) plays a vital role in digitising and preserving musical heritage, but historical documents often lack the labelled data required by traditional methods. We overcome this challenge by training a neural-based feature extractor on unlabelled data, enabling effective classification with minimal samples. Key contributions include optimising crop preprocessing for a self-supervised Convolutional Neural Network and evaluating classification methods, including SVM, multilayer perceptrons, and prototypical networks. Our experiments yield an accuracy of 87.66%, showcasing the potential of AI-driven methods to ensure the survival of historical music for future generations through advanced digital archiving techniques.

Regular Session S7P: Advanced Medical Image Segmentation with Deep Learning and Enhanced Pre-processing
Session chair: Prof. Mondher Frikha
(Friday 10 Jan 2025 at 16:10-18:10)

An inventive network intrusion detection system: Composite deep learning CNN-LSTM model trained on UNSW-NB15 dataset

10 Jan 2025
16:10-18:10
Session S7P

Hanene Sahli[1], Souhir M'rabet[2], Bacely Yorobi[3], Mounir Sayadi[4]
[1]:sahli.hanenne@gmail.com, [2]:souhir.mrabet@gmail.com, [3]:bacely@theschooly.tech, [4]:mounir.sayadi@esstt.rnu.tn

[1,4]:National High Engineering School of Tunis, Tunisia
[2,3]:The Schooly & CO, France

Summary

The creation of efficient network intrusion detection systems (NIDS) has become vital with the rising occurrence of network intrusions. In this research, we introduce an innovative NIDS method that integrates the strengths of convolutional neural network (CNN) and long short-term memory (LSTM) mechanisms to examine the network traffic data characteristics. We employ the UNSW-NB15 dataset, which showcases a varied distribution of patterns, including a notable imbalance between the size of the training and testing sets. Unlike conventional machine learning methods, which frequently face challenges with restricted feature sets and reduced accuracy, our proposed model addresses these shortcomings. Current models applied to this dataset generally necessitate manual feature selection and extraction,

which can be less accurate, labor-intensive and time-consuming. Conversely, our model attains better performance in binary classification by harnessing the benefits of combined CNN and LSTM models. By conducting thorough experiments and evaluations with advanced deep learning models, we showcase the superiority and the efficacy of the proposed approach. The obtained results emphasize the promise of integrating CNN and LSTM in order to improve network intrusion detection.

Medical Image Authentication and Self-Recovery Using Fragile Watermarking in the Frequency Domain

10 Jan 2025
16:10-18:10
Session S7P

Bouarroudj Riadh[1], Harrou Fouzi[2], Zerrouki Nabil[3], Souami Feryel[4], Bellala Fatma Zohra[5], Sun Ying[6]

[1]:rbouarroudj@usthb.dz, [2]:fouzi.harrou.2016@ieee.org, [3]:nzerrouki@cdta.dz, [4]:f.souami@univ-alger.dz, [5]:fzbellala@usthb.dz, [6]:ying.sun@kaust.edu.sa

[1,5]:University of Science and Technology Houari Boumediene, Algeria
[2,6]:King Abdullah University of Science and Technology, Saudi Arabia
[3]:Center for Development of Advanced Technology, Algeria
[4]:University of Algiers 1, Algeria

Summary

The rapid growth in digital image sharing, driven by advancements in internet and communication technologies, has raised concerns about image integrity, especially in sensitive fields like healthcare. This paper presents a fully blind fragile watermarking technique for authenticating and self-recovering color and gray-scale medical images. The approach applies the Discrete Wavelet Transform (DWT) to the cover image, and the resulting sub-bands are divided into 3×3 blocks. One authentication and four recovery watermarks are then embedded into the least significant bits (LSBs) of each block. During the extraction phase, if tampering is detected, the model accurately localizes the altered areas, and a three-level recovery process, including a new inpainting technique, is used to recover the original image. Results based on two publicly available datasets demonstrate that this method delivers high-quality watermarked images, achieves optimal watermark extraction accuracy, and maintains high sensitivity to various attacks. Additionally, it provides precise tamper localization and delivers high-quality recovered images, even with tampering rates as high as 60%.

Hypernetwork Design for Self-Supervised Domain Adaptation in Compressive Sensing MRI Reconstruction

10 Jan 2025
16:10-18:10
Session S7P

Libya Thomas[1], Joseph Zacharias[2]
[1]:d-tve21jan026@cet.ac.in, [2]:josephzacharias@cet.ac.in

[1,2]:College of Engineering, Trivandrum, India

Summary

This paper presents a novel hypernetwork design tailored for self-supervised domain adaptation in compressed sensing MRI (CS-MRI) reconstruction. The proposed hypernetwork generates task-specific parameters for a primary reconstruction network based on regularization weights, dynamically adjusting to distribution shifts between training and test data. We demonstrate the efficiency and adaptability of the hypernetwork through test-time training (TTT), improving reconstruction accuracy even in unseen data distributions. Empirical results using the fastMRI dataset highlight significant improvements in Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

Semi-Supervised Semantic Segmentation using Redesigned Self-Training for White Blood Cells

10 Jan 2025
16:10-18:10
Session S7P

Luu Vinh Quoc [1], Nguyen Quan[2], Le Khanh Duy[3], Dinh Vinh[4], Nguyen Minh Thanh[5], Huy Nguyen Thanh[6], Nguyen Thinh[7]
[1]:vinhluu2608@gmail.com, [2]:quannd@ptit.edu.vn,[3]:duy.le1201@hcmut.edu.vn,
[4]:vinh.dq2@vgu.edu.vn, [5]:minhdc1712@gmail.com

[1,3,7]:Ho Chi Minh University of Technology
[2]:Posts and Telecommunications Institute of Technology, Vietnam,
[4]:Department of Computer Science, Vietnamese German University, Vietnam
[5]:University of Medicine and Pharmacy Ho Chi Minh city, Vietnam,
[6]:Ho Chi Minh University of Education, Vietnam

Summary

Artificial Intelligence (AI) in the medical field, especially in the diagnosis of cancer relating to white blood cells, is held back by two primary obstacles: the scarcity of extensively labeled datasets for white blood cell (WBC) segmentation and outdated segmentation techniques. These challenges stall the development of more accurate and modern techniques to diagnose cancer relating to white blood cells. To overcome the first challenge, there is a need to devise a semi-supervised learning framework to efficiently make use of the huge dataset that is not annotated. Our study tackles this by introducing an innovative self-training approach that integrates FixMatch. Self-training is a semi-supervised framework that leverages the model trained on labeled data to generate pseudo-labels for the unlabeled data and then re-train on

both of them. FixMatch is used to regularize the model against weak and strong perturbations in the input image. We discover that by incorporating FixMatch in the self-training pipeline, the performance improves in many cases. When tested, our method achieved superior results on the UNet and DeepLab-V3 networks with ResNet-50 architecture, attaining accuracies of 90.87%, 89.05%, and 75.13% on the Zheng 1, Zheng 2, and LISC datasets, respectively.

---

10 Jan 2025
16:10-18:10
Session S7P

CycleGAN-Based Approach for Enhanced Segmentation of Lung Disease Areas in Chest X-Ray Images

Hage Chehade Aya[1], Abdallah Nassib[2], Marion Jean-Marie[3], Oueidat Mohamad[4], Chauvet Pierre[5]

[1]:aya.hagechehade@etud.univ-angers.fr, [2]:nassib.abdallah@univ-angers.fr, [3]:marion@uco.fr, [4]:mohoueidat@yahoo.com, [5]:pierre.chauvet@uco.fr

[1]:Aya Hage Chehade (LARIS, University of Angers), France
[2]:LARIS, University of Angers, LaTIM, INSERM UMR 1101, University of Brest, France
[3]: LARIS, University of Angers, Catholic University of the West (UCO)), France
[4]:Faculty of Technology, Lebanese University, Lebanon,
[5]:LARIS, University of Angers, Catholic University of the West (UCO), France

Summary

In this study, we develop a new segmentation approach based on CycleGAN model to generate healthy lung images from pathological chest X-ray images, followed by image subtraction and binarization to produce a mask that includes pathology-affected areas. This approach enables the extraction of radiomic features from regions containing pathologies, enhancing disease classification. Our segmentation approach demonstrated effectiveness, improving AUC by 10.92% over conventional segmentation method for classifying effusion and infiltration using the XGBoost model, and outperforming previous studies. This study underscores the importance of precise pathological mask generation for accurate lung disease classification.

Medical Image Segmentation Techniques: Advances and Challenges

10 Jan 2025
16:10-18:10
Session S7P

Boulehmi Hela[1], Filali Rania Linda[2]

[1]:hela.boulehmi@yahoo.fr, [2]:ranialindafilali@gmail.com

[1,2]:National Engineering School of Tunis, Tunisia

Summary

Medical image segmentation plays a pivotal role in computer-aided diagnosis by facilitating the extraction of essential features necessary for disease detection and treatment strategies. The continuous progress in image processing technologies has led to the development of numerous segmentation methods, encompassing traditional algorithms, machine learning (ML)-driven approaches, and cutting-edge deep learning (DL) techniques. This study undertakes a comparative evaluation of these methods, focusing on their efficiency, accuracy, and suitability across different medical imaging modalities. It also delves into prominent segmentation techniques like thresholding, region-based methods, edge detection, graph cuts, active contour models, and convolutional neural networks (CNNs). Additionally, the paper explores ongoing challenges and prospective advancements aimed at enhancing segmentation efficacy in medical imaging.

C2GMatch: Leveraging Dual-View Cross-Guidance and Co-Guidance Framework for Semi-Supervised Cell Segmentation

10 Jan 2025
16:10-18:10
Session S7P

Truc Nguyen[1], Hong Ong[2], Thien Nguyen[3], Bui Hung[4], Nguyen Vu[5], Minh Nguyen[6], Quan Nguyen[7], Thung Tran[8]

[1]:nguyenthingoctruc102@gmail.com, [2]:hong_ong@epam.com,
[3]:n21dccn080@student.ptithcm.edu.vn, [4]:19127414@student.hcmus.edu.vn,
[5]:nguyenvuhoangwork@gmail.com, [6]:minhdc1712@gmail.com, [7]:quannd@ptit.edu.vn,
[8]:hungnh.isi@lqdtu.edu.vn

[1,5]:AI Vietnam, Vietnam,
[2]:EPAM Vietnam, Vietnam,
[3,7]:Posts and Telecommunications Institute of Technology, Vietnam,
[3,4]:Ho Chi Minh University of science, Vietnam,
[6]:University of Medicine and Pharmacy, Ho Chi Minh city, Vietnam,
[8]:Le Quy Don technical university, Vietnam

Summary

Automated white blood cell segmentation is crucial for diagnosing and monitoring severe conditions like leukemia and lymphoma. However, this task is hindered by

the scarcity and quality of available public datasets. To make better use of these limited datasets, a semi-supervised learning framework is required. In our study, we introduce the C2GMatch framework, which leverages weak-to-strong dual-view guidance to enhance segmentation performance from partially labeled data. Our approach surpasses two out of three leading semi-supervised segmentation methods on three public datasets, including BCCD, LISC, and LiveCells, achieving IOU scores of 54.12% and 68.13% and Dice scores of 69.97% and 82.01% on LISC and LiveCells, respectively. Our framework also shows better segmentation masks compared to previous works such as Pseudo-Seg or Unimatch in qualitative results.

Regular Session S8P: Novel Approaches in Computer Vision: From Image Processing to Visual Understanding
Session chair:
(Saturday 11 Jan 2025 at 9:00-10:20)

Fusion Grad-CAM: A Methodology for Generating Unified Attention Maps in Majority Voting Classifiers

11 Jan 2025
9:00-10:40
Session S8P

Ntampakis Nikolaos[1], Konstantinos Diamantaras[2], Vasileios Argyriou[3], Panagiotis Sarigianndis[4]

[1]:ntampakis@metamind.gr, [2]:kdiamant@ihu.gr, [3]:vasileios.argyriou@kingston.ac.uk, [4]:psarigiannidis@metamind.gr

[1,2]:International Hellenic University Sindos, Greece,
[3]: Kingston University London London, United Kingdom,
[4]:University of Western Macedonia Kozani, Greece,
[1,4]:MetaMind Innovations Kozani, Greece

Summary

In this work, we introduce a novel methodology called Fusion Grad-CAM for generating attention maps in ensemble classification systems that utilize majority voting. Currently, there is no established technique for producing a consolidated attention map in such systems. Traditional approaches require inspecting individual attention maps from each contributing model, which is inefficient. Our proposed method integrates Grad-CAM produced attention maps by averaging or weighted averaging them based on both the predicted class and the confidence level of each model's decision. This approach offers a holistic representation of the ensemble's decision-making process. We validate our methodology using various combinations of well-known

convolutional neural networks (CNNs) pretrained on ImageNet, demonstrating the effectiveness and clarity of the resulting attention maps.

A Fast Frequency-Space Transform-Invariant Template Matching using a "Pseudo-Normalizable" Fourier Transform (PNFT)

11 Jan 2025
9:00-10:40
Session S8P

Nordin Zakaria
nordinzakaria@utp.edu.my

Universiti Teknologi PETRONAS, Malaysia

Summary

Given a template image T , the task in template matching is to search for T in a larger image region S. While simple in concept, much effort in recent years has been channeled on improving its robustness to complex transformation, with attention being largely on the similarity measures. Much lesser attention has been on the efficiency of the search computation itself. This paper attempts to fill the gap by proposing Pseudo- Normalizable Fourier Transform (PNFT), a novel complex transform that enables efficient transform-invariant, vectorized template matching. The concept of normalization is overloaded to denote the transformation of a function over an arbitrarily located patch to that about the origin, enabling the efficient use of image integral to perform template matching. The paper discusses on the 'normalizability' – whether the function can be normalized – and further introduces pseudo-normalization as an approximation. The approach proposed requires no learning and runs fast on lower-end CPU-only platform. Further, it is conceptually simple and can be implemented in approximately 200 lines of Python/Numpy code. The efficacy of the approach is empirically analyzed using a complex example image.

On-the-fly Image Compression Using Overfitted Shallow CNN Autoencoders

11 Jan 2025
9:00-10:40
Session S8P

John El Gallab [1], Abdelwahab Ganna [2], Abdallah Kassem [3], Hend Makram [4], Martina Muawad [5], Mina Wagih Mikhael [6], Tamer ElBatt [7]

[1]:john.elgallab@aucegypt.edu*; [2]:abdelwahabganna@aucegypt.edu; [3]:abdullahkassem@aucegypt.edu; [4]:hendmakram@aucegypt.edu; [5]:martinamichelcx@aucegypt.edu; [6]:minawmikhael@aucegypt.edu; [7]:telbatt@ieee.org,

The American University in Cairo

Summary

In the era of pervasive Internet use, managing large volumes of image data becomes crucial. To mitigate storage and bandwidth costs, image compression plays a pivotal role. Traditional image compression techniques like JPEG and PNG, while widely used, may suffer from limited compression rates. In this work, we propose a novel approach using shallow Convolutional Neural Network (CNN) autoencoders for on-the-fly image compression. Our model aims to achieve high compression rates with improved image quality compared to classical methods. Additionally, it supports decompression on the CPU in real-time, making no assumptions about client-side computational resources. We present a comprehensive methodology, including architecture design, experiments, and performance metrics. Our results demonstrate the effectiveness of the proposed approach, providing a competitive alternative for online content compression and decompression that outperforms JPEG in image quality assessment metrics at 33% compression rate. We also outperform other Learned Image Compression (LIC) techniques in both the decompression time and number of trainable parameters with an improvement in the order of 10-times.

Bridging the Machine-Human Gap in Blurred-Image Classification via Entropy Maximization

11 Jan 2025
9:00-10:40
Session S8P

Emilio Sansano[1], Martinez-Garcia[2], Javier Portilla[3]

[1]:esansano@uji.es, [2]:martigar@uji.es, [3]:javier.portilla@csic.es

[1,2]:UJI, Spain, [3]:Consejo Superior de Investigaciones Científicas (CSIC), Spain

Summary

Recent studies point to an accuracy gap between humans and Artificial Neural Network (ANN) models when classifying blurred images, with humans outperforming ANNs. To bridge this gap, we introduce a spectral channel-based range-constrained entropy merit function, from which we devise a zero-phase, circular symmetric blind deblurring method. We apply it as a pre-processing step for image classification and test it using pre-trained classification models and images blurred by Gaussian kernels. We compare our method to state-of-the-art restoration methods, showing its superiority, effectively bridging the machine-human gap for most models and blur levels. Our results also rank higher than the competitors in no-reference and full-reference image quality metrics. Notwithstanding the limitation to zero-phase blur, this work shows that, for image pre-processing aimed at visual tasks, it may be advantageous to use merit functions based on vision science and information theory, rather than the classical minimisation of the expected error relative to the latent image.

---

11 Jan 2025
9:00-10:40
Session S8P

Optimizing Color Vision Accuracy in Ishihara Tests through Spectral Illumination Optimization

Sonain Jamil[1], Muhammad Turab[2]
[1]:sonainjamil@ieee.org, [2]:turabbajeer202@gmail.com

[1]:University of Eastern Finland (UEF), Joensu, Finland
[2]: Norwegian University of Science and Technology (NTNU), Gjovik, Norway

Summary

In order to design optimal illumination that mini- mizes or maximizes differences between two sample points, we utilized particle swarm optimization (PSO), a popular algorithm for finding optimal solutions. In this study, we explored the application of PSO to minimize or maximize cost functions such as $\Delta E, \Delta RGB$, and Michelson contrast, and to output optimal intensities for designing the emission spectra of the illumination. This approach proved to be more convenient than altering the spectral properties of the object or the observer. The experiment involved capturing sample data from the Ishihara Color Blindness test book using Specim V10E, and measuring the emission spectra of ten LEDs from Ledmotive's spectrally tunable LED light source using the Hamamatsu PMA-12 spectrometer. Subsequently, we pre-processed the captured data and applied PSO to obtain optimal intensities for the ten LEDs. The resulting optimized intensities were then used to compute the spectrally optimal emission spectrum. To assess the performance of these emission spectra, we illuminated the samples with the optimized spectra using the spectrally tunable LED light source. From the illuminated samples, we concluded that for the minimum cost function, the red number on the sample was not distinguishable from the green background, whereas for the maximum cost function, the red number was distinctive.

Keynote VI by Prof. Justin Dauwels
Session chair: Dr. Carlos Crispim-Junior
(Saturday 11 Jan 2025 at 10:30-11:10)

11 Jan 2025
10H30-11H10

AI for applications in psychiatry

Justin Dauwels
TU-Delft, Netherlands

Summary In this talk, we will consider applications of AI in the domain of psychiatry. Specifically, we will give an overview of our research towards automated behavioral analysis for assessing psychiatric symptoms. Schizophrenia (SCZ) and depression (MDD) are two chronic mental disorders that seriously affect the quality of life of millions of people worldwide. We aim to develop machine-learning methods with objective linguistic, speech, facial, and motor behavioral cues to reliably predict the severity of psychopathology or cognitive function, and distinguish diagnosis groups. We collected and analyzed the speech, facial expressions, and body movement recordings of 228 participants (103 SCZ, 50 MDD, and 75 healthy controls) from two separate studies. We created an ensemble machine-learning pipeline and achieved a balanced accuracy of 75.3% for classifying the total score of negative symptoms, 75.6% for the composite score of cognitive deficits, and 73.6% for the total score of general psychiatric symptoms in the mixed sample containing all three diagnostic groups. The proposed system is also able to differentiate between MDD and SCZ with a balanced accuracy of 84.7% and differentiate patients with SCZ or MDD from healthy controls with a balanced accuracy of 82.3%. These results suggest that machine-learning models leveraging audio-visual characteristics can help diagnose, assess, and monitor patients with schizophrenia and depression.

Justin Dauwels biography

Dr. Justin Dauwels is an Associate Professor at the TU Delft (Signals and Systems, Department of Microelectronics), and serves as co-Director of the Safety and Security Institute at the TU Delft. He also is the scientific lead of the Model-Driven Decisions Lab (MoDDL), a first lab for the Knowledge Building program between the Netherlands police and the TU Delft. His research interests are in data analytics with applications to intelligent transportation systems, autonomous systems, and analysis of human behavior and physiology. His academic lab has spawned four startups across a range of industries, ranging from AI for healthcare to autonomous vehicles.

Regular Session S9P: Deep Learning for Autonomous Navigation and Smart City Applications
Session chair: Prof. Serge Miguet (Saturday 11 Jan 2025 at 11:10-12:30)

Multichannel Object Detection with Event Camera

11 Jan 2025
11:10-12:30
session S9P

Rafael Iliasov[1], Alessandro Golkar[2]
[1]r.iliasov@tum.de, [2]:golkar@tum.de

Technical University of Munich (TUM), Germany

Summary

Object detection based on event vision has been a dynamically growing field in computer vision for the last 16 years. In this work, we create multiple channels from a single event camera and propose an event fusion method (EFM) to enhance object detection in event-based vision systems. Each channel uses a different accumulation buffer to collect events from the event camera. We implement YOLOv7 for object detection, followed by a fusion algorithm. Our multichannel approach outperforms single-channel- based object detection by 0.7% in mean Average Precision (mAP) for detection overlapping ground truth with IOU=0.5.

---

11 Jan 2025
11:10-12:30
session S9P

Integrating Object Detection in Bird-View Systems for Enhancing Safety in Construction Machinery

Julian Hund[1], Eric Schoneberg[2], Daniel Goerges[3], Hasan Hashib[4]
[1]:jhund@rptu.de, [2]:eric.schoeneberg@rptu.de, [3]:goerges@eit.uni-kl.de,
[4]:hashib@rhrk.uni-kl.de

University of Kaiserslautern-Landau, Germany

Summary

Working with construction machinery necessitates that workers remain attentive to effectively survey their surroundings. The complex environment, blind spots, and limited visibility while simultaneously tackling complex tasks create hazardous conditions. Camera systems providing a bird view help monitoring the machine's surroundings, but still require active monitoring while working. In this paper, we propose extending bird-view systems with object detection. We investigate and evaluate two different approaches: Detecting people in four fisheye camera images and transforming the obtained bounding boxes into a bird view, and directly detecting people in bird view. Our models, based on the YOLOv5 and YOLOv8 architectures, were trained and validated on custom datasets with 3,302 fisheye images and 1,343 bird-view images. The models achieved mAP@.5-.95 scores of 0.514 (fisheye) and 0.546 (bird view) and run at 25 frames per second on an NVIDIA Jetson Nano edge GPU. Our trained models for the detection utilize the YOLOv5 and YOLOv8 architecture and run inference with 25 frames per second on an NVIDIA Jetson Nano edge GPU. The first approach also addresses objects disappearing in stitching areas - characteristic of bird-view projection errors - by displaying bounding boxes even if they are invisible in bird view.

---

11 Jan 2025
11:10-12:30
session S9P

End-to-end Sketch-Guided Path Planning through Imitation Learning for Autonomous Mobile Robots

Anthony Rizk[1], Charbel Abi Hana[2], Youssef Bakouny[3], Flavia Khatounian[4]
[1]:anthony.rizk1@net.usj.edu.lb, [2]:charbel.abihana@idealworks.com,
[3]:youssef.bakouny@usj.edu.lb, [4]:flavia.khatounian@usj.edu.lb

[1,3,4]:Saint Joseph University of Beirut, Lebanon,
[1,2]:IDEALworks GmbH, Germany

Summary

Path planning is crucial for Autonomous Mobile Robots applications. Traditionally, path planning based on human input and preferences has relied on hard to define reward-based learning or costly techniques requiring additional hardware. This work introduces a more accessible and flexible approach through sketch-guided imitation learning, where nontechnical users can simply draw the desired navigational path

on a provided 2D map, which is then used to teach U-net models path planning behaviors. Additionally, the work draws on metrics from the fields of image generation and robotics to provide a novel evaluation framework. The approach is integrated into an end-to-end robotics stack to demonstrate its usability. The dataset and code are provided upon publication.

Event-based Sensor Fusion and Application on Odometry: A Survey

11 Jan 2025
11:10-12:30
session S9P

Jiaqiang Zhang[1], Xianjia Yu[2], Sier Ha[3], Haizhou Zhang[4], Tomi Westerlund[5]

[1]:jizhan@utu.fi, [2]:xianjia.yu@utu.fi, [3]:sierha@utu.fi, [4]:hazhan@utu.fi, [5]:tovewe@utu.fi

University of Turku, Finland

Summary

Event cameras, inspired by biological vision, are asynchronous sensors that detect changes in brightness, offering notable advantages in environments characterized by high-speed motion, low lighting, or wide dynamic range. These distinctive properties render event cameras particularly effective for sensor fusion in robotics and computer vision, especially in enhancing traditional visual or LiDAR-inertial odometry. Conventional frame-based cameras suffer from limitations such as motion blur and drift, which can be mitigated by the continuous, low-latency data provided by event cameras. Similarly, LiDAR-based odometry encounters challenges related to the loss of geometric information in environments such as corridors. To address these limitations, unlike the existing event camera-related surveys, this paper presents a comprehensive overview of recent advancements in event-based sensor fusion for odometry applications particularly, investigating fusion strategies that incorporate frame-based cameras, inertial measurement units (IMUs), and LiDAR. The survey critically assesses the contributions of these fusion methods to improving odometry performance in complex environments, while highlighting key applications, and discussing the strengths, limitations, and unresolved challenges. Additionally, it offers insights into potential future research directions to advance event-based sensor fusion for next-generation odometry applications.

---

11 Jan 2025
11:10-12:30
session S9P

Advanced Deep Learning Approaches for Automated Recognition of Cuneiform Symbols

Shahad Elshehaby[1], Alavikunhu Panthakkan[2], Hussain Al-Ahmad[3], Mina Al-Saad[4]
[1]:s0000002884@ud.ac.ae, [2]:apanthakkan@ud.ac.ae,
[3]:halahmad@ud.ac.ae, [4]:malsaad@ud.ac.ae

University of Dubai, United Arab Emirates

Summary

This paper presents a thoroughly automated method for identifying and interpreting cuneiform characters via ad vanced deep-learning algorithms. Five distinct deep-learning models were trained on a comprehensive dataset of cuneiform characters and evaluated according to critical performance metrics, including accuracy and precision. Two models demon strated outstanding performance and were subsequently assessed using cuneiform symbols from the Hammurabi law acquisition, notably Hammurabi Law 1. Each model effectively recognized the relevant Akkadian meanings of the symbols and delivered precise English translations. Future work will investigate ensemble and stacking approaches to optimize performance, utilizing hybrid architectures to improve detection accuracy and reliability. This research explores the linguistic relationships between Akkadian, an ancient Mesopotamian language, and Arabic, emphasizing their historical and cultural linkages. This study demonstrates the capability of deep learning to decipher ancient scripts by merging computational linguistics with archaeology, therefore providing significant insights for the comprehension and conservation of human history.

Online Session S1V: Advances in Image Segmentation for Biological and Agricultural Applications Session chair: Prof. Mouna Zouari (Thursday 09 Jan 2025 at 10:20-12:20)

Segment Anything for Dendrites from Electron Microscopy

9 Jan 2025
10:20-12:20
Live stream
Session S1V

Zewen Zhuo[1], Ilya Belevich[2], Ville Leinonen[3], Eija Jokitalo[4], Tarja Malm[5], Alejandra Sierra[6], Jussi Tohka[7]

[1]:zewen.zhuo@uef.fi, [2]:ilya.belevich@helsinki.fi, [3]:ville.leinonen@uef.fi, [4]:eija.jokitalo@helsinki.fi, [5]:tarja.malm@uef.fi, [6]:alejandra.sierralopez@uef.fi, [7]:jussi.tohka@uef.fi

[1,3,5,6,7]:University of Eastern Finland, Finland, [2,4]:University of Helsinki, Finland

Summary

Segmentation of cellular structures in electron microscopy (EM) images is fundamental to analyzing the morphology of neurons and glial cells in the healthy and diseased brain tissue. Current neuronal segmentation applications are based on convolutional neural networks (CNNs) and do not effectively capture global relationships within images. Here, we present DendriteSAM, a vision foundation model based on Segment Anything, for interactive and automatic segmentation of dendrites in EM images. The model is trained on high-resolution EM data from healthy rat hippocampus and is tested on diseased rat and human data. Our evaluation results demonstrate better mask quality compared to the original and other fine-tuned models, leveraging the features learned during training. This study introduces the first implementation of vision foundation models in dendrite segmentation, paving the path for computer-assisted diagnosis of neuronal anomalies.

---

9 Jan 2025
10:20-12:20
Live stream
Session S1V

Uncertainty Driven Sampling to Handle Intra-class Imbalance Part Segmentation in Wheat

Reena Reena[1], Kevin Williams[2], Fiona Corke Aberystwyth[3], Huaizhong Zhang[4], Yonghuai Liu[5], John H. Doonan Aberystwyth[6]

[1]:Reenar@edgehill.ac.uk, [2]:ksw09@aber.ac.uk, [3]:fic5@aber.ac.uk, [4]:zhangh@edgehill.ac.uk, [5]:liuyo@edgehill.ac.uk, [6]:jhd2@aber.ac.uk

[1,4,5]:Edge Hill University, United Kingdom,
[2,3,6]:Aberystwyth University, United Kingdom

Summary

We introduce a novel method to address intra-class imbalance in 3D point cloud segmentation of wheat, focusing on distinguishing between ear and non-ear parts. Variability in plant structure, influenced by factors such as curvature and shape, often leads to data imbalance which complicates segmentation tasks. Our approach utilizes Monte Carlo Dropout to identify and prioritize uncertain samples at the end of each training epoch, employing uncertainty-driven sampling to select samples with the lowest confidence. These samples undergo augmentation through scaling and leaf crossover techniques, enhancing their representation in the training set. Our comparative evaluations demonstrate that this strategy significantly improves the mean Intersection over Union (mIoU) and segmentation accuracy, thereby increasing model robustness for complex 3D plant structures.

---

9 Jan 2025
10:20-12:20
Live stream
Session S1V

AI-based robotic trap for real-time insect detection, monitoring and population prediction

Dimitrios Kapetas[1], Valerio Goglia[2], Eleftheria Maria Pechlivani[3]

[1]:dimikape@iti.gr, [2]:valerio.goglia@agrorobotica.it, [3]:riapechl@iti.gr

[1]:The Centre for Research and Technology Hellas, Information Technologies Institute, Greece
[2]:Agrorobotica Srl, Italy
[3]:Centre for Research and Technology Hellas, Greece

Summary

Timely, accurate, and efficient pest detection is essential in agricultural sectors to mitigate economic losses and environmental impact, mainly in insect spreads. Today, AI-based robotic traps are used for insect monitoring and forecasting population growth, enabling rapid responses in pest management to reduce pesticide use and achieve sustainable goals. The core of this challenge is to build fast and accurate Artificial Intelligence (AI) models that promptly alert farmers for immediate action. In this study, we demonstrate the innovative potential of a robust AI-based trap using a robust dataset created in real settings consisting of 217 images and 6787 total

annotations. The SpyFly robotic trap was used for continuous real-time monitoring and detection of the invasive insect Tuta absoluta. The detection system leverages Deep Learning (DL) models, with the best-performing model (YOLOv8x) achieving high accuracy, reaching a mean Average Precision (mAP50) of 97.6%. Additionally, environmental data such as temperature, humidity, and barometric pressure captured by the trap are used to forecast insect population growth utilizing various Machine Learning (ML) models. The ARIMAX model demonstrates significant accuracy in anticipating pest outbreaks, with a Mean Square Error (MSE) of 36.69 for weekly averaged predictions and 309.67 for daily predictions. This translates to an offset of 6 insects for weekly predictions and 17 insects for daily predictions, indicating the average deviation between the predicted and actual insect counts. This integrated approach supports early intervention and sustainable pest management, highlighting the potential of AI and DL in smart agriculture enhancing productivity and reducing pesticide usage. Dataset is available for download at https://doi.org/10.5281/zenodo.13134110.

Sorghum Plant Counting Using Patch-Wise CNN and Histogram Equalization

9 Jan 2025
10:20-12:20
Live stream
Session S1V

Moctar I. Mossi
mmoctar@gmail.com

Laboratoire d'énergétique, d'électronique, d'electrotechnique et d'informatique industrielle, University Abdou Moumouni, Niamey, Niger

Summary

Due to the increasing impact of climate change on agriculture, efficient early-stage control and yield prediction are becoming increasingly crucial. While satellite data has demonstrated significant efficiency in certain applications, deep learning solutions for plant phenotyping and UAV-based imaging have emerged as promising alternatives. Early-stage plant counting is a key strategy for yield forecasting and mitigating the risk of insufficient production. Manual crop counting, however, is time-consuming, expensive, error-prone, and labor-intensive. Automated, accurate counting offers a significant reduction in workload. This study proposes an automating plant counting method using aerial images of sorghum crops. A histogram equalization is used at the image level then a patch-wise counting is derived from dot annotations of patches using a complexity-optimized regression-based deep convolutional neural network (DCNN). The proposed approach estimates the number of plants within patches and aggregates these estimates for each image. Our findings demonstrate that this approach outperforms existing CNN-based solutions on the same dataset, achieving a lower mean absolute error (MAE) while requiring fewer parameters.

---

9 Jan 2025
10:20-12:20
Live stream
Session S1V

Aquaponic Farming with Advanced Decision Support System: Practical Implementation and Evaluation

Eleftheria Maria Pechlivani[1], Georgios Gkogko[2], Panagiotis Christakakis[3], Dimitrios Kapetas[4], Ioannis Hadjigeorgiou[5], Dimosthenis Ioannidis[6], Dimitrios Tzovaras[7]

[1]:riapechl@iti.gr, [2]:ggogos@iti.gr, [3]:christakakis@iti.gr, [4]:dimikape@iti.gr, [5]:ihadjigeorgiou@sidroco.com, [6]:djoannid@iti.gr, [7]:dimitrios.tzovaras@iti.gr

[1,2,3,4,6,7]: Centre for Research and Technology Hellas, Greece,
[5]: Sidroco Holdings Ltd, Cyprus
[6]: Information Technologies Institute, Greece

Summary

The adoption of Decision Support Systems (DSS) by farmers is challenging, requiring user-friendly interfaces, comprehensive training, reliable support, and affordable pricing. This paper presents an advanced DSS for aquaponics, leveraging Artificial Intelligence (AI), the Internet of Things (IoT), and Blockchain to meet global food production demands while optimizing resource management and ensuring cybersecurity. The DSS simplifies complex technologies through an intuitive interface, minimizing jargon and presenting real-time data in an accessible format, enabling farmers of all skill levels to use it effectively. Key innovations include AI-driven robotic traps with 82% accuracy in insect detection, autonomous robots with 94% precision in 3D spot spraying, and nutrient analyzers achieving over 92% accuracy. The blockchain layer ensures secure data verification and traceability while supporting distributed AI model verification. The system reduces pesticide use by 50%, enhances crop yields, and cuts sample analysis costs by 70%, demonstrating efficiency. Additionally, the DSS integrates hierarchical clustering algorithms for advanced analytics, offering actionable insights into pest and nutrient dynamics. By detailing the implementation of these technologies in aquaponic facilities, the paper highlights how the DSS improves productivity, plant health, and the utilization of biopesticides and biofertilizers. The findings illustrate the scalability and effectiveness of the system, providing insights into its broader application for sustainable agriculture. This work demonstrates the transition of theoretical models into real-world applications, showing the transformative potential of digital agriculture.

Applications of Artificial Intelligence and Cloud Computing in the Oil and Gas Industry- A Review

9 Jan 2025
10:20-12:20
Live stream
Session S1V

Alfitouri Ibrahim Jellah[1], Mohamed. M Milad[2], Moaad. I Ibrahim[3]
[1]:alfitouri.jellah@bwu.edu.ly, [2]:mohamedalshopi132@gmail.com, [3]:moad17730@gmail.com

Bani-Waleed Univesrity, Libya

Summary

This paper reviews the latest machine learning and artificial intelligence technology, showcasing its potential to revolutionize the oil and gas industry. Despite the current limitations, the industry is on the brink of a digital transformation that promises to enhance operations and increase overall production. The future of the oil and gas industry, with the help of AI and cloud computing, is one of optimism and potential. The review showed that there was an increase in production significantly when applying AI technology in particular with EOR. The paper summarized the limitations and future challengers of this technology.

LANet: Enhancing Plant Disease Recognition Through Transfer Learning and Layer Attention

9 Jan 2025
10:20-12:20
Live stream
Session S1V

Xin Lei[1], Yonghuai Liu[2], Ardhendu Behera[3], Wuping Zhang[4], Yang Sun[5]
[1]:24764256@edgehill.ac.uk, [2]:liuyo@edgehill.ac.uk, [3]:beheraa@edgehill.ac.uk,
[4]:zwping@126.com, [5]:bless2016@163.com

[1]: Edge Hill University, China,
[2,3]: Edge Hill University United Kingdom,
[4]: Shanxi Agriculture University China,
[5]: Hebei University, China

Summary

Accurate recognition of plant diseases plays a critical role in maintaining agricultural productivity and food security. This study proposes LANet, an innovative network architecture aimed at improving the precision of identifying and categorizing plant diseases from images. LANet consists of two primary components. Firstly, Layer Attention is incorporated to address the challenge of data loss during downsampling and emphasise plant species recognition. Secondly, we propose a transfer learning method to learn the associative features between the leaf categories and their corresponding diseases. Comparative analyses based on the public PlantVillage dataset reveal the superior performance of our method in disease recognition.

Online Session S2V: Medical Imaging Diagnosis and Deep Learning I
Session chair: Prof. Alima Damak (Thursday 09 Jan 2025 at 16:10-17:50)

LFRanker: An Iterative Method for Identifying Top-Performing Loss Functions in SEM Image Denoising

09 Jan 2025
16:10-17:50
Live stream
Session S2V

Sheikh Shah Mohammad Motiur Rahman[1], Michel Salomon[2], Sounkalo Dembele[3]
[1]:motiur.ion@gmail.com, [2]:michel.salomon@femto-st.fr, [3]:sounkalo.dembele@femto-st.fr

FEMTO-ST Institute, France

Summary

This work focuses on identifying the best performing loss functions for denoising scanning electron microscopy (SEM) images. Recent studies on the impact and effect of loss functions in image denoising are multiplying, particularly in the fields of low-light and microscopic imaging. Most studies have focused on hybrid loss functions, which are a mixture of multiple loss functions, and have shown their significance. However, manual experiments with different loss functions to build a hybrid function are generally very time-consuming. Thus, we propose a framework named LFRanker (Loss Function Ranker) that automatically and iteratively experiments loss function from a given set. LFRanker automatically calculates combinations using set theory from a given set of loss functions. The findings indicated significant variations across the different assessment criteria (such as PSNR, SSIM and keypoints preservation). In this work, we focus on SEM image denoising, which may be subject to reconstruction in the next step. Thus, it is necessary to preserve the maximum keypoints for point-cloud reconstruction during denoising, along with high quality. Based on these criteria, we found a hybrid loss function, combining LMSE, LMAE, LPER and LSIM performs best. Finally, LFRanker demonstrates

the ability to automate manual experimental efforts in identifying appropriate loss functions for denoising.

---

09 Jan 2025
16:10-17:50
Live stream
Session S2V

Processing and Classification Algorithms of Medical Images

Oltiana Seitaj
oseitaj2@os.uniroma3.it

Universita degli studi RomaTre, Italy

Summary

This paper evaluates the impact of hybrid deep learning approaches on lung tumor segmentation by combining traditional image processing techniques with advanced AI-driven models. The study integrates Convolutional Neural Networks (CNNs) with preprocessing methods such as noise reduction, adaptive thresholding, and contrast enhancement to address challenges associated with complex anatomical structures and variability in medical image quality. A novel hybrid framework is proposed, leveraging traditional methods to preprocess data and improve input quality for deep learning models, ultimately enhancing segmentation accuracy and reliability.The effectiveness of the proposed approach is assessed using quantitative performance metrics, including Dice Similarity Coefficient (DSC), Hausdorff Distance, Jaccard Index, Precision, and Recall. Preliminary results indicate significant improvements in tumor boundary detection and reduced false-positive rates compared to existing methods. By streamlining segmentation workflows and enabling near-real-time applications in clinical settings, this research offers a pathway to improved diagnostic accuracy, treatment planning, and workflow efficiency. Future implications include the potential for integration into clinical imaging pipelines, fostering advancements in computer-assisted diagnosis and personalized treatment strategies. This study underscores the value of hybrid methodologies in addressing current limitations and paving the way for more precise and efficient medical image segmentation.

---

09 Jan 2025
16:10-17:50
Live stream
Session S2V

Optimizing Multimodal Transformers for Medical Image Captioning: Enhancing Automated Descriptions via AI Systems

Mithila Arman[1], Md Khurshid Jahan[2], Ahmed Haque Dhrubo[3], Md. Mahfuzur Rhaman[4], Sumaya Binte Zilani Choya[5], Din Mohammad Dohan[6], Md. Ashiq Ul Islam Sajid[7], Md. Golam Rabiul Alam[8]

[1]:mithilaarman30@gmail.com, [2]:khurshid.jahan@northsouth.edu,
[3]:ahmed.dhrubo@northsouth.edu, [4]:mrhaman@gmu.edu,
[5]:sumayachowa1998@gmail.com, [6]:din.mohammad.dohan@g.bracu.ac.bd,
[7]:md.ashiq.ul.islam.sajid@g.bracu.ac.bd, [8]:rabiul.alam@bracu.ac.bd

[1,4]:BRAC University, Bangladesh, Bangladesh
[2,3]:North South University, Bangladesh

Summary

In contemporary diagnostic workflows, medical image captioning has emerged as a pivotal advancement, combining deep learning methodologies and transformer architectures to enhance accuracy and efficiency in medical interpretations. This paper proposes the optimization of multimodal transformers for automated medical image captioning, focusing on integrating Vision Transformers (ViT) and Bidirectional Auto-Regressive Transformers (BART) with novel variations such as Swin Transformers and GPT-2. We use a robust multimodal AI framework to explore how these architectures synergize to generate coherent and diagnostically relevant captions for radiological images. We assess the performance of multiple transformer models by employing the ROCO dataset, containing paired X-ray images and expert-generated reports. Our findings demonstrate that the ViT + BART combination yields the most stable and accurate captions, minimizing training and validation loss. In contrast, DEiT + MBART displayed instability, highlighting the need for further hyperparameter tuning. Through this comparative analysis, we underscore the critical role of transformer-based models in reducing the cognitive load on medical professionals, enhancing diagnostic accuracy, and promoting real-time, automated medical image interpretation.

VMambaDA: Visual State Space Model with Unsupervised Domain Adaptation in Cervical Cancer

09 Jan 2025
16:10-17:50
Live stream
Session S2V

Nguyen Vu[1], Truc Nguyen[2], Bui Duc[3], Kim Ngan Huynh[4], Minh Nguyen[5], Hoang Tran[6], Duy Nguyen[7]

[1]:nguyenvuhoangwork@gmail.com, [2]:nguyenthingoctruc102@gmail.com,
[3]:henry.bmduc@gmail.com, [4]:23C01011@student.hcmus.edu.vn,
[5]:minhdc1712@gmail.com, [6]:tmhoang77@gmail.com, [7]:duynvm@iuh.edu.vn

[1,2,3]:AI Vietnam, Vietnam
[4]:Ho Chi Minh University of Science, Vietnam
[5]:University of Medicine and Pharmacy Ho Chi Minh city, Vietnam
[6]:Truong Vuong Hospital, Vietnam
[7]:Industrial University of Ho Chi Minh City, Vietnam

Summary

Cervical cancer remains a major public health concern in Vietnam, ranking second only to breast cancer among women. Early detection through screening, particularly HPV testing and Pap smears, is critical in reducing cervical cancer mortality. While deep learning has shown great promise in medical image analysis, particularly for detecting cervical cancer, challenges remain, especially when models are applied to

real-world datasets with limited labeled data due to privacy concerns and expensive annotations. This research offers an effective cervical cancer diagnostic approach that combines Sliced Wasserstein Distance (SWD), Maximum Classifier Discrepancy (MCD), and the VMamba model. The VMambaDA model learns domain-invariant features and adjusts to domain shifts to address the problems of accuracy, generalization, and speed in cervical cancer screening. VMambaDA has proven to be more adept at handling the intricacies of medical images than earlier mod- els by exhibiting better classification accuracy and sensitivity through extensive testing on both public and private datasets. This automated method could enhance the early detection of cervical cancer and intervention by pushing the limits of domain adaptation in cytopathology.

Enhancing Fatty Liver Disease Diagnosis Using Deep from Ultrasound Images

09 Jan 2025
16:10-17:50
Live stream
Session S2V

Mohammad Naserameri[1], Kavian Amirmozafarisabet[2], Meisam Namazi[3]
[1]:mohammad.naserameri@mail.concordia.ca, [2]:kavianam@gmail.com,
[3]:meisamnamazi@gmail.com

[1]:Computer Science Department at Concordia University, Canada,
[2,3]:Department Of Computer Science University Of Tehran Tehran, Iran

Summary

The increasing prevalence of fatty liver disease necessitates accurate and efficient diagnostic methods. This study investigates the integration of deep learning techniques to enhance the diagnosis of fatty liver disease using ultrasound images. A dataset was utilized to train a deep learning model. The model achieved an impressive accuracy of 96% on the training dataset for distinguishing patients with fatty liver disease from healthy individuals, while maintaining a commendable accuracy of 90% on a completely new dataset. Furthermore, the model demonstrated a sensitivity of 92% in classifying different levels of liver fat in the training dataset, with an accuracy of 83% on the new dataset. These results underscore the effectiveness of combining deep learning in medical imaging, providing a robust framework for the early detection and classification of fatty liver disease. The findings suggest that this hybrid approach can significantly improve diagnostic accuracy, ultimately contributing to better patient management and outcomes in clinical practice. Further research is warranted to validate these results across diverse populations and enhance model generalizability.

Maskify: Precision Detection and Fit Assessment for Mask Compliance

09 Jan
16:10-17:50
Live stream
Session S2V

Prakrit Pathak
prakrit19072@iiitd.ac.in

IIIT-Delhi, India

Summary

We present Maskify, a method for automatically detecting and assessing face mask usage. In light of the COVID-19 pandemic, wearing face masks has proven to be effective in reducing the spread of the virus. As public health guidelines increasingly require face coverings, reliable recognition technologies are essential for ensuring compliance, particularly in confined spaces. To address this need, we propose a deep learning-based solution, which classifies whether an individual is wearing a mask and evaluates the mask's fit with a "wearable score." This score reflects how well the mask covers the necessary facial areas. The development of Maskify necessitates a comprehensive dataset of both masked and unmasked faces.
Our approach involves leveraging pre-trained models to accurately identify faces within images. Once detected, key points on the lower portion of the face are extracted to facilitate mask classification. For this task, we employ ResNet-based models due to their proven efficacy in image classification. To calculate the wearable score, we segment the masked area and determine the percentage of the face that is properly covered by the mask. Maskify thus offers a robust solution for monitoring mask compliance and effectiveness, contributing to ongoing efforts to mitigate COVID-19 transmission.

Rapid Distributed Fine-tuning of a Segmentation Model Onboard Satellites

09 Jan 2025
16:10-17:50
Session S3P

Meghan Plumridge[1], Rasmus Maråk[2], Pablo Gómez[3],
Gabriele Meoni[4], Chiara Ceccobello[5], Filip Svoboda[6], Nicholas Lane[7]
[1]:map205@cam.ac.uk, [2]:rasmus.marak@ai.se, [3]:pablo.gomez@esa.int,
[4]:gabriele.meoni@esa.int, [5]:chiara.ceccobello@ai.se, [6]:fs437@cam.ac.uk,
[7]:ndl32@cam.ac.uk

[1,6,7]:Cambridge University, United Kingdom
[2,5]:AI Sweden, Sweden
[3]:European Space Agency, Spain
[4]:TU Delft, Netherlands
[7]:Samsung AI, United Kingdom

Summary

Segmentation of Earth observation (EO) satellite data is critical for natural hazard analysis and disaster response. However, processing EO data at ground stations

introduces delays due to data transmission bottlenecks and communication windows. Using segmentation models capable of near-real-time data analysis onboard satellites can therefore improve response times. This study presents a proof-of-concept using MobileSAM, a lightweight, pre-trained segmentation model, onboard Unibap iX10-100 satellite hardware. We demonstrate the segmentation of water bodies from Sentinel-2 satellite imagery and integrate MobileSAM with PASEOS, an open-source Python module that simulates satellite operations. This integration allows us to eval uate MobileSAM's performance under simulated conditions of a satellite constellation. Our research investigates the potential of fine-tuning MobileSAM in a decentralised way onboard multiple satellites in rapid response to a disaster. Our findings show that MobileSAM can be rapidly fine-tuned and benefits from decentralised learning, considering the constraints imposed by the simulated orbital environment. We observe improvements in segmentation performance with minimal training data and fast fine-tuning when satellites frequently communicate model updates. This study contributes to the field of onboard AI by emphasising the benefits of decentralized learning and fine-tuning pre-trained models for rapid response scenarios. Our work builds on recent related research at a critical time; as extreme weather events increase in frequency and magnitude, rapid response with onboard data analysis is essential.

Online Session S3V: Advanced Techniques and Deep Learning for Patter Recognition
Session chair: Prof. Norhene Gargouri
(Friday 10 Jan 2025 at 09:00-10:30)

A Speech Enhancement Algorithm Combining Wavelet Transform and Adaptive Filters

10 Jan 2025
09:00-10:30
Session S3V
Live stream

Aayam Shrestha[1], Seyed Ghorshi[2], Marjan Joorabchi[3], Issa Panahi[4], Fereshteh Fakhar Firouzeh[5]

[1]:ashrestha7@patriots.uttyler.edu, [2]:aghorshi@uttyler.edu,
[3]:marjan.joorabchi@gmail.com, [4]:issa.panahi@utdallas.edu,
[5]:fakhar.behnaz@gmail.com

[1,2,5]:University of Texas at Tyler, United States of America
[3]:Sharif University of Technology, International Campus, Iran
[4]:University of Texas at Dallas, United States of America

Summary Noise as an unwanted interference can significantly degrade speech signals, especially those recorded by many microphones. This interference is modeled as additive noise that originates from a range of sources including White Gaussian Noise (WGN), babble, crowd, large city, and traffic noises. These interference signals can alter the characteristics of speech signals, reducing both their quality and intelligibility. This paper introduces a novel approach designed to reduce noise and enhance the quality and intelligibility of speech signals. The proposed method combines Wavelet Transform with Adaptive Filters, specifically the Wiener filter and RLS filter. The evaluation process involves testing noisy speech signals under realistic conditions with different signal-to-noise ratios (SNRs) and different types of

additive noise. The objective measure is used for evaluation, including the perceptual evaluation of speech quality (PESQ). Results show that combining Wiener or RLS filtering with Wavelet Transform significantly improves noise reduction, outperforming the use of Wavelet Transform alone.

Label Consistent Generalized Adaptive Weighted Recursive Least Squares Dictionary Learning

10 Jan 2025
09:00-10:30
Session S3V
Live stream

Mohadeseh Yousefi[1], Yashar Naderahmadian[2], Soosan Beheshti[3]

[1]:mohadesehyousef@msc.guilan.ac.ir, [2]:naderahmadian@gmail.com,
[3]:soosan@torontomu.ca,

[1,3]:University of Guilan, Iran
[2]:Toronto Metropolitan University, Canada

Summary The Generalized Adaptive Weighted Recursive Least Squares (GAWRLS) dictionary learning method has shown potential for unsupervised dictionary learning. This paper advances GAWRLS by incorporating classification error as an additional cost to enable supervised learning tasks and introduces the Label Consistency for online supervised dictionary learning in classification tasks. The new method is denoted as Label Consistent Generalized Adaptive Weighted Recursive Least Squares Dictionary Learning (LC-GAWRLS). By incorporating both sparse representation error and classification error into the cost function, LC-GAWRLS enables simultaneous learning of the dictionary and classifier parameters. Particularly, to ensure label consistency, the proposed algorithm introduces a correction weight to adaptively regulate the impact of each training data during the model update, enhancing robustness against variations in training data compared to previous dictionary learning methods. Simulation results on real datasets demonstrate that LC-GAWRLS achieves higher classification accuracy compared to existing state-of-the-art supervised dictionary learning methods, particularly in scenarios with limited training samples per class.

Semi-Supervised Generalized Adaptive Weighted Recursive Least Squares Dictionary Learning

10 Jan 2025
09:00-10:30
Session S3V
Live stream

Mohadeseh Yousefi[1], Yashar Naderahmadian[2], Soosan Beheshti[3]
[1]:mohadesehyousef@msc.guilan.ac.ir, [2]:naderahmadian@gmail.com,
[3]:soosan@torontomu.ca,

[1,3]:University of Guilan, Iran
[2]:Toronto Metropolitan University, Canada

Summary

This work presents a novel semi-supervised dictionary learning framework that updates the dictionary by online learning and is efficient in utilizing the training data. The method employs a two-stage process to train the dictionary: initial training with limited labeled data, followed by online refinement using abundant unlabeled data. We introduce an adaptive correction weight to control the influence of new unlabeled data on the dictionary update, based on its consistency with the current model estimate. This approach enables efficient use of the training data set. Moreover, results in faster dictionary convergence and improves data representation accuracy, especially in scenarios with limited training data. Experimental results demonstrate significant enhancement in the classification accuracy of the proposed method compared to the state-of-the-art semi-supervised dictionary learning methods, particularly when dealing with a limited number of training samples.

Overview of Accent Conversion Techniques Based on Deep Learning

10 Jan 2025
09:00-10:30
Session S3V
Live stream

Aayam Shrestha[1], Seyed Ghorshi[2], Issa Panahi[3]
[1]:ashrestha7@patriots.uttyler.edu, [2]:aghorshi@uttyler.edu,
[3]:issa.panahi@utdallas.edu

[1,2]:The University of Texas at Tyler, United States of America
[3]:University of Texas at Dallas, United States of America

Summary

In modern communication, speech technology plays a vital role, with accent being one of the key features that reflects the nuances of spoken technology. However, differences in accent can make communication difficult, especially in global contexts like language learning and human computer interaction. Accent conversion approaches often solve these issues by transforming the accent of speaker to that of the other speakers, while maintaining linguistic information and speaker identity. This technology holds great applications to speech recognition, language learning, and cross-lingual communication. This paper explores the use of advanced deep

learning techniques in accent conversion systems which have improved both the naturalness and intelligibility of converted speech. Additionally, we discuss the methods generally used for performance evaluation of these systems.

---

10 Jan 2025
09:00-10:30
Session S3V
Live stream

How Culturally Aware Are Vision-Language Models?

Olena Burda-Lassen[1], Aman Chadha[2], Shashank Goswami[3], Vinija Jain[4]
[1]:oburdalassen@gmail.com, [2]:hi@aman.ai,
[3]:shashankgoswami80@gmail.com, [4]:hi@vinija.ai

[1,3]:Independent Research Scientist, United States of America
[2,4]:Stanford University, United States of America
[2]:Amazon Inc., United States of America

Summary

An image is often considered worth a thousand words, and certain images can tell rich and insightful stories. Can these stories be told via image captioning? Images from folklore genres, such as mythology, folk dance, cultural signs, and symbols, are vital to every culture. Our research compares the performance of four popular vision-language models (GPT-4V, Gemini Pro Vision, LLaVA, and OpenFlamingo) in identifying culturally specific information in such images and creating accurate and culturally sensitive image captions. We also propose a new evaluation metric, the Cultural Awareness Score (CAS), which measures the degree of cultural awareness in image captions. We provide a dataset MOSAIC-1.5k labeled with ground truth for images containing cultural background and context and a labeled dataset with assigned Cultural Awareness Scores that can be used with unseen data. Creating culturally appropriate image captions is valuable for scientific research and can be beneficial for many practical applications. We envision our work will promote a deeper integration of cultural sensitivity in AI applications worldwide. By making the dataset and Cultural Awareness Score available to the public, we aim to facilitate further research in this area, encouraging the development of more culturally aware AI systems that respect and celebrate global diversity.

Language Prediction of Twitch Streamers using Graph Convolutional Network

10 Jan 2025
09:00-10:30
Session S3V
Live stream

Md. Nahid Hasan[1], Nakshi Saha[2], Md. Anisur Rahman[3]
[1]:nahid1740@cseku.ac.bd, [2]:nakshi1736@cseku.ac.bd,
[3]:anis@cse.ku.ac.bd

Khulna University, Bangladesh

Summary

Social Network is a common example of the graph structure. There are many algorithms available for predicting different elements of social networks. Language is one of the elements. Prediction problems generally stay within 2D information. Graphs provide more structural information like edge relationship edge feature etc. While predicting social network elements, information about graph structure can be very useful.
Graph Convolutional Network gives the opportunity to use graph structure as a piece of new information. Twitch is a popular social media platform where people stream gaming-related content. In our work, we have proposed a Graph Convolutional Network model for predicting the language of Twitch gamers. We also showed that it outperforms other machine learning and deep learning algorithms like Decision trees, Random Forest, Support Vector Machine, Neural Network and others. While other models could only reach 45.8% accuracy, GCN was able to achieve 73.43% by utilizing the graph structure.

A Lightweight and Efficient Convolutional Neural Network for Crowd Counting

10 Jan 2025
09:00-10:30
Session S3V
Live stream

Marios Constantinou[1], Panayiotis Kolios [2], Christos Kyrkou[3]
[1]:constantinou.a.marios.1@ucy.ac.cy, [2]:kolios.panayiotis@ucy.ac.cy,
[3]:ckyrkou@gmail.com

[1,2,3]:KIOS Center of Excellence, Cyprus
[1,3]:University of Cyprus, Cyprus

Summary

The crowd counting task plays a key role in ensuring public safety during large gatherings events. Most prominent works in this area, use large and computationally demanding deep learning model architectures, which require substantial computational power, limiting their usage in a real-world scenario under resource constraints. In this work we consider the trade-off between the model's predicted accuracy and computational speed. We propose an improved version of HR-Net, which is substantially smaller and faster than the original, but preserves its localization and counting performance. Through targeted removal of unnecessary modules and branches, we

demonstrate an increase in frames-per-second by 37.71% on an Nvidia Jetson Orin, and a reduction of GMACs and parameters by 77.41% and 73.07% respectively, while retaining competitive localization and counting performance, specifically for aerial imagery scenarios. Our modifications enable the algorithm to process in real-time higher resolution images, which is crucial when dealing with small objects. Furthermore, because most crowd counting datasets contain random images gathered from the web, and limited aerial images of crowds, we introduce a specialized dataset of high-resolution aerial imagery for sparse and dense crowds in various environments, that contains new drone-captured annotated image data.

Online Session S4V: Medical Imaging Diagnosis & Deep Learning II
Session chair: Prof. Hela Boulehmi
(Friday 10 Jan 2025 at 11:20-13:00)

Exploring Advanced Deep Learning Techniques for Multi-Class Melanoma Cancer Classification

10 Jan 2025
11:20-13:00
Live stream
Session S4V

Haifa Alhasson
hhson@qu.edu.sa

Qassim University, Saudi Arabia

Summary

Melanoma represents one of the most lethal forms of skin cancer, underscoring the importance of early detection for effective treatment and improved survival rates. Traditional diagnostic methods, which predominantly rely on visual inspection and biopsies, are often time-consuming and susceptible to human error. Timely diagnosis significantly enhances the likelihood of recovery and can reduce healthcare costs by minimizing the necessity for surgical, radiographic, or chemical treatments. Recently, deep learning techniques have demonstrated considerable promise in automating and improving the accuracy of medical diagnoses, including melanoma classification. In this study, we evaluate the performance of several state-of-the-art deep learning models—DenseNet, ResNet, VGG-16, VGG-19, Inception v3, and AlexNet—for multi-class melanoma cancer classification. Our objective is to identify the model that offers the best performance in terms of accuracy, sensitivity, and specificity. We conduct a comprehensive comparison using publicly available datasets, such as HAM10000 ("Human Against Machine with 10000 training images"), to ensure robust and generalizable results. This evaluation aims to advance

the field of melanoma diagnosis by identifying the most effective deep-learning approach, thereby facilitating early and accurate detection of this life-threatening disease.

---

10 Jan 2025
11:20-13:00
Live stream
Session S4V

Novel Hybrid Deep Learning Model for Enhancing Skin Cancer Detection

Jamel Smida[1], Mohamed Karim Azizi[2], Sam Chandra Bose[3], Amor Smida[4]
[1]:jsmida@um.edu.sa, [2]:medkarim.azizi@gmail.com, [3]:asam@um.edu.sa,
[4]:a.smida@mu.edu.sa

[1,3]: AlMaarefa University, Saudi Arabia
[2]: ISAMM, Saudi Arabia
[4]: College of applied medical science, Saudi Arabia

Summary

Skin Cancer is the most serious form of cancer, often spreading to other parts of the body over time. Early and accurate identification of skin lesions can improve the treatment options and increase the probability of curing the disease before it progresses. However, automatic detection of skin cancer faces difficulties because of the imbalanced datasets, affected areas covered by hair or moles, and other visual obstructions. In recent years, Deep Learning (DL) techniques have been widely used to address these challenges. Despite advancements, achieving consistently high accuracy remains difficult. To tackle these challenges, the research proposes a novel hybrid DL model for efficient skin cancer detection from dermoscopic images. The proposed network consists of three key modules: feature extraction using the Xception network, feature enhancement through the Squeeze and Excitation network (SENet), and final classification using a transformer module. This combination of DL models effectively identifies SC. The proposed network is evaluated using benign and malignant dermoscopic images collected from the Kaggle website. Necessary preprocessing steps, such as resizing, grayscale conversion, hair removal, and data balancing, are applied. The proposed network is compared with popular DL networks such as MobileNet, Xception, and VGG-16. The experimental outcome demonstrates that the proposed network outperforms these models, achieving a higher accuracy, recall, precision, and F1 score of 96.83%, 97%, 96.67%, and 96.83%, respectively. The metrics value shows that the proposed network is promising and can be helpful for dermatologists in identifying skin cancer with minimal effort.

Redesigned Dual-Task Learning Framework for Diagnosis Mammography Screening with BI-RADS and Density Classification

10 Jan 2025
11:20-13:00
Live stream
Session S4V

Quan Nguyen
quannd@ptit.edu.vn

Posts and Telecommunications Institute of Technology, Vietnam

Summary

Mammography plays a pivotal role in breast cancer diagnosis and monitoring, yet the accuracy of Breast Imaging-Reporting and Data System (BI-RADS) assessments can vary among radiologists, particularly concerning breast density evaluations. Computer-aided diagnosis (CADx) systems have emerged to augment diagnostic precision. In this context, we propose a redesigned Dual-Task Learning (DTL) framework for mammography screening, focusing on BI-RADS and breast density classification. Our approach, notably DTL-Variant M, demonstrates superior performance across multiple metrics. DTL-Variant M showcases substantial enhancements in both BI-RADS and breast density classification tasks compared to other variants, emphasizing its efficiency with ResNeXt-50 backbone. Furthermore, we employ focal loss, a highly effective loss function for imbalanced data, in our approach to tackle the problem of class imbalance and achieve better results for BI-RADS classification, which we consider more significant than density classification when using cross-entropy loss.

A 3D U-Net-Based Network for Segmenting Medical Images of Fetal MRI in Three Dimensions

10 Jan 2025
11:20-13:00
Live stream
Session S4V

Maneesha Singh[1], Sameesha Singh[2], Liu Jiancong[3], Jana Hutter[4]
[1]:maneesha.uk@gmail.com, [2]:sameesha.uk@gmail.com, [3]:liu.jiancong@kcl.ac.uk, [4]:jana.hutter@outlook.com

[1,2]:Centre for Health Innovations, MRIIRS Faridabad, India,
[3,4]:King's College London, United Kingdom

Summary

—This study introduces a U-Net based algorithmic framework designed to segment 3D MRI images of perinatal fetal brains from a cohort of 20 fetuses, with gestational ages ranging from 20 to 36 weeks. Furthermore, an optimization network based on U-Net is also proposed to enhance the segmentation efficiency of the proposed U-Net framework. The experimental results indicate that the resultant Dice Score metric, Hausdorff distance and Jaccard Index all showed notable enhancements when compared to the U-Net pipeline. The final results confirm that the proposed method is highly effective for segmenting fetal brains across various gestational ages, with the optimized model yielding superior segmentation quality as compared to the conventional U-Net model.

---

10 Jan 2025
11:20-13:00
Live stream
Session S4V

A Machine Learning Framework for Skin Cancer Detection using Transfer Learning

Sotirios Spanogianopoulos[1], Mary Linda Panakal Augustine[2], Saeed Sharif[3]
[1]: sotiris.spanogianopoulos@gmail.com, [2]:m.l.augustine@uel.ac.uk, [3]:s.sharif@uel.ac.uk

[1]: University of Portsmouth, United Kingdom,
[2,3]:University of East London, United Kingdom

Summary

This paper uses deep learning algorithms including InceptionV2, InceptionV3, DenseNet, MobileNet, and VGG19 to improve skin cancer detection. This research aims to improve skin cancer diagnosis. This work aims to determine the most effective evolutionary metrics-based technique to recognizing skin cancer, which is comparable to other diseases. Ultimately, our paper aims to create a realistic skin cancer detection system that uses the best deep learning algorithm. This discovery might improve medical diagnostics, leading to earlier diagnosis and improved healthcare outcomes.

Online Session S5V: AI for Autonomous Navigation, Infrastructure Inspection, and Environmental Monitoring Session chair: Prof. Ali Wali (Friday 10 Jan 2025 at 14:40-16:00)

Navigating Beyond GPS Horizon Detection for Autonomous Navigation

10 Jan 2025
14:40-16:00
Session S5V

Alastair J Finlinson[1], Sotiris Moschoyiannis[2]
[1]:a.finlinson@surrey.ac.uk, [2]:s.moschoyiannis@surrey.ac.uk

University of Surrey, United Kingdom

Summary

In a scenario where a GPS signal is unavailable, navigation becomes more challenging. To enable precise navigation, visual features of the surroundings can be interpreted to direct vessels. One such navigation method is the visible horizon line, which matches this to an elevation model. The visible horizon line is constructed from features at the intersection between the sky, land masses, and the waterline. Accurate extraction relies on a robust model operating in adverse weather and equipment conditions. We propose to tackle this new problem with an approach based on semantic boundary prediction with a framework that primarily looks at deep supervision for feature conditioning and differential edge detection to incorporate strong priors early in the training process. We also use semantic segmentation as a support task to provide a strong supervision signal and features to the semantic boundaries prediction task. We conduct experiments on the dataset to evaluate the model in various scenarios. Our results show that our framework predicts high-quality boundaries and segmentation masks over various datasets and domains with the ability to perform in low-data scenarios. We demonstrate that the combination

of the labelling and the edge priors improve over the baseline by 32-44% on the semantic boundary estimation task.

---

10 Jan 2025
14:40-16:00
Session S5V

Optimizing Transmission Line Insulator Defect Detection: Lightweight Edge AI on UAVs

Zhifan Song[1], Yuan Zhang[2], Xinrui Liao[3], Abd Al Rahman M. Abu Ebayyeh[4]

[1]:zhifan.song@lip6.fr, [2]:zhangyuan2024@berkeley.edu, [3]:xinrui.liao22@imperial.ac.uk, [4]:a.abu-ebayyeh@imperial.ac.uk

[1]:Sorbonne University, France
[2]:University of California, Berkeley, United States
[3,4]: Imperial College of London, United Kingdom

Summary

Effective transmission line insulator defect detection is crucial for maintaining power grid reliability and safety. Traditional methods, which often rely on ground-based inspections or cloud-based computing, can be slow and inefficient. This paper introduces EdgeIDD, a lightweight solution for real-time UAV-based insulator defect detection leveraging edge computing. By enabling on-site data processing, EdgeIDD reduces the reliance on high-bandwidth communication with central servers and addresses connectivity challenges in remote areas where most power transmission towers are constructed. Built upon YOLOv10, the latest state-of-the-art model with faster speed compared to precedent iterations by removing the post-processing step, EdgeIDD integrates minimalist design principles, replacing the complete backbone with lightweight vanilla blocks and incorporating our novel C2fP block and utilized distribution shifting convolution for accelerated feature fusion. Our model achieves a remarkable 45.2% reduction in parameters compared to YOLOv10n while maintaining a high mean average precision (mAP) of 97.5% with minimal accuracy loss. With hardware acceleration, EdgeIDD reaches 23.3 FPS on a Raspberry Pi 5 with 7.3x speedup. This advancement in edge computing ensures rapid, reliable insulator defect detection directly from the UAV, even in areas with limited connectivity.

A Computer Vision Approach for Autonomous Cars to Drive Safe at Construction Zone

10 Jan 2025
14:40-16:00
Session S5V

Abu Shad Ahammed[1], Roman Obermaisser[2], Md Shahi Amran Hossain[3]
[1]:abu.ahammed@uni-siegen.de, [2]:roman.obermaisser@uni-siegen.de,
[3]:md.hossain@uni-siegen.de

University of Siegen, Germany

Summary

To build a smarter and safer city, a secure, efficient, and sustainable transportation system is a key requirement. The autonomous driving system (ADS) plays an important role in the development of smart transportation and is considered one of the major challenges facing the automotive sector in recent decades. A car equipped with an autonomous driving system (ADS) comes with various cutting-edge functionalities such as adaptive cruise control, collision alerts, automated parking, and more. A primary area of research within ADAS involves identifying road obstacles in construction zones regardless of the driving environment. This paper presents an innovative and highly accurate road obstacle detection model utilizing computer vision technology that can be activated in construction zones and functions under diverse drift conditions, ultimately contributing to build a safer road transportation system. The model developed with the YOLO framework achieved a mean average precision exceeding 94% and demonstrated an inference time of 1.6 milliseconds on the validation dataset, underscoring the robustness of the methodology applied to mitigate hazards and risks for autonomous vehicles.

NetraAI-The 3rd Eye: AI-Empowered Surveillance

10 Jan 2025
14:40-16:00
Session S5V

Harsh Davec[1], Sparsh Pathak[2], Kamal Sutaria[3]
[1]:harsh.dave1506@gmail.com,
[2]:sparshpathak2002@gmail.com,[3]:kamal.sutaria24554@paruluniversity.ac.in

PIET, Parul University, India

Summary

This research paper explores the development and implementation of NetraAI - The 3rd Eye an AI powered surveillance system aimed at enhancing public safety and security measures. The study investigates the technical architecture, real-time functionalities, and ethical considerations surrounding the deployment of NetraAI. It examines its impact on object detection, people tracking, vehicle recognition, and proactive threat detection in diverse surveillance scenarios. The findings highlight the system's efficacy, ethical deployment, and potential contributions to the field of AI-driven surveillance.

---

10 Jan 2025
14:40-16:00
Session S5V

Multi-Model AI-Driven Satellite Image Segmentation for Comprehensive Urbanization and Environmental Degradation Analysis in Bangladesh

Rezwan-ul-alam[1], Ahmed Faizul Haque Dhrubo[2], , Md. Iftekharul Alam[3], Nafiz Iqbal[4], Raian Ruku[5], Fahim Mahmud[6], Md. Mehedi Hasan[7], Mohammad Abdul Qayum[8]

[1]:rezwan.alam1@northsouth.edu, [2]:ahmed.dhrubo@northsouth.ed,
[3]:iftekharul.alam03@northsouth.edu, [4]:nafiz.iqbal@northsouth.edu,
[5]:raian.ruku@northsouth.edu, [6]:fahim.mahmud03@northsouth.edu,
[7]:mehedi.howlader@northsouth.edu, [8]:mohammad.qayum@northsouth.edu

Dept.of ECE North South University Dhaka, Bangladesh

Summary

The accelerated urbanization in Bangladesh has markedly altered landscapes and ecosystems, posing intricate challenges for policymakers and urban planners. This study underscores the necessity for sophisticated monitoring methodologies, employing AI and deep learning for meticulous image processing and annotation. We curated an extensive dataset by amassing and annotating satellite imagery from diverse regions of Bangladesh, delineating buildings, trees, water sources, tin shades, and farmlands. This dataset underpins the analysis of temporal landscape transformations, urban sprawl, environmental attrition, and land encroachments over the past decade. Our AI framework incorporates four avant-garde deep learning models, ensuring unparalleled accuracy and robust object detection through rigorous cross-validation. Despite data annotation complexities, our methodology exhibits substantial potential in refining AI efficacy for satellite image analysis. Future endeavours will focus on augmenting the dataset, mitigating annotation challenges, and integrating supplementary models to enhance precision. This research yields pivotal insights for urban planning, environmental stewardship, and sustainable development.

---

10 Jan 2025
14:40-16:00
Session S5V

An Effective Method for Checking Consistency in Univariate Data-series with Multiple, Incomplete and Inconsistent Entries

Sameesha Singh[1], Maneesha Singh[2]

[1]:sameesha.uk@gmail.com, [2]:maneesha.uk@gmail.com

[1]:Centre for Health Innovations, MRIIRS Faridabad, India,
[1,2]:Railvision Europe Ltd Castle Donington, United Kingdom

Summary

Uni-variate time series is utilized in numerous scientific applications for a variety of purposes. They serve as the foundation for different statistical measurements and show how data evolves over time when viewed in-dependently as real-time data series. Furthermore, binary data sets are frequently utilized in various fields, including health sciences, where each observation results in either a positive or negative outcome. The primary uses of time series analysis include: (a) Predicting future trends; (b) Determining characteristics like average, range, and standard deviation to comprehend the data's behavior; (c) Examining variations in means and variances among multiple time series using statistical methods like t-tests, F-tests, and Anova; and (d) Contrasting multiple time series to identify point-to-point discrepancies using different measures like root mean square error. This article suggests a straightforward yet successful method for comparing data from various time series that are not synchronized, have gaps in data, and require determining the error of the best match to ensure it falls within an acceptable range. We utilize this method for a challenging railway engineering issue to verify the consistency of laser and inertial sensor measurements within a specified error range.

Real-time detection of road hazards for autonomous vehicle systems

10 Jan 2025
14:40-16:00
Session S5V

Anita Keszler[1], *Maté Andras Szabo*[2], *Laszló Tizedes*[3]

[1]:keszler@sztaki.hu, [2]:szabomandras@sztaki.hun-ren.hu,
[3]:tizedes.laszlo@sztaki.hun-ren.hu

European Union's Centre of Excellence in information technology and automation, Hungry

Summary

Automated detection of road hazards such as speed bumps, potholes, road drainage, and manhole covers has become an important area of research due to its potential to improve road safety in autonomous driving. Various techniques have been introduced to detect such hazards using camera vision and artificial intelligence-based image processing methods. More recently convolutional neural networks have shown remarkable success in this field. However, estimating the distance to the traffic hazards is still challenging. To address this problem and to satisfy the requirement for real-time on-board data processing, the proposed system has the following properties: high-accuracy road hazard detection by analyzing mono-images and videos with a re-trained YOLO neural network; precise distance measurement utilizing a LiDAR; and efficient local data processing using ROS, implemented on an NVIDIA Jetson AGX Xavier. An important contribution of this paper is introducing multiple classes of road hazards when training the network, instead of only focusing on speed bumps and potholes. With this we have improved the neural network's ability to recognize the most important types of obstacles on the road. Furthermore we have analyzed different LiDAR technologies to evaluate and compare their precision and

demonstrate that our method can be successfully applied regardless of the scanning pattern of the LiDAR.

Online Session S6V: AI for Security and Human-Robot Interaction: Authentication, Recognition, and Content Protection
Session chair: Dr. Hanene Sahli (Saturday 11 Jan 2025 at 16:10-18:10)

Finding Identity through Gait

Lavanya Srinivasan
lavanya.srinivasan@uwl.ac.uk

University of West London, United Kingdom

10 Jan 2025
16:10-18:10
Session S6V
Live stream

Summary Human gait identification is to recognize a person from a series of walking images. In contrast to fingerprint or iris-based identification methods, gait identification offers the significant benefit of remote execution. Gait detection is emerging as one of the most promising biometric identification techniques. Traditional methods to identify the identity of gait are: Background subtraction, Gait Energy Image, Gait Entropy, etc., human gait identification is a potential new tool for identifying individuals beyond traditional methods. Developing an automated gait detection system and examining the unique aspects of gait were the objectives of this study. A gait identification method was suggested for this purpose, and the identification outcomes were contrasted using various techniques. This study encompassed the following: predictions and estimation based on joint positions, better predictions with higher confidence scores, and the joint coordinates and joint points are fed into Graph Convolutional Network by eliminating low confidence pose to classify labels. The experimental results show that the Graphical Convolutional Network identifies

gait subjects with higher accuracy. In the future, this work will be extended to identify different gait subjects and the gait in forensic identifications.

Authentication and Verification in Human-Robot Cooperative Robotic Cells using Stereo Vision and Gesture Control

10 Jan 2025
16:10-18:10
Session S5V
Live stream

Gabor Kovacs[1], Tamas Sziranyi[2]

[1]:kovacs.gabor@sztaki.hu, [2]:sziranyi.tamas@sztaki.hu

[1,2]:HUN-REN SZTAKI, Institute for Computer Science and Control, Budapest, Hungary

Summary

The integration of human-machine interaction technologies with industrial automation has become increasingly essential for enhancing productivity and safety in manufacturing environments. In this paper, we propose a novel approach to address these challenges by using stereo vision and gesture control in cooperative robotic cells. Our system enables seamless authentication of operators and real-time verification of task execution, ensuring compliance with established protocols and safety standards. Key features of our system include its gesture-based operation with gesture recognition algorithms, allowing operators to interact with robotic systems intuitively and efficiently. By leveraging stereo vision, our system accurately tracks the operators' movement within the workspace, facilitating precise task execution and object manipulation. We present a detailed description of our system architecture, experimental configuration, and real-world performance assessment. Our results demonstrate the effectiveness and feasibility of our approach in enhancing operational efficiency, ensuring quality, and improving the overall user experience in industrial automation.

Few Skeleton Features For Gate-Based Gender Recognition From Landmark Frames

10 Jan 2025
16:10-18:10
Session S5V
Live stream

Amara Bekhouch[1], Walid Barhoumi[2], Noureddine Doghmane[3]

[1]:a.bekhouch@univ-soukahras.dz, [2]:walid.barhoumi@enicarthage.rnu.tn, [3]:ndoghmane@univ-annaba.org

[1]: University of Souk Ahras, Algeria
[2]:National Engineering School of Carthage, Tunisia
[3]:Badji Mokhtar-Annaba University, Algeria

Summary

Surveillance, forensic, and biometric systems are continually evolving with modern technology. Despite these advancements, their effectiveness in crime deterrence remains debatable, especially given the low-resolution nature of footage from distant cameras. Gait-based gender identification is a crucial application in surveillance. This paper investigates the use of minimal human Surveillance, forensic, and biometric systems are continually evolving with modern technology. Despite these advancements, their effectiveness in crime deterrence remains debatable, especially given the low-resolution nature of footage from distant cameras. Gait-based gender identification is a crucial application in surveillance. This paper investigates the use of minimal human skeleton points to derive discriminative features for gender recognition. We propose a new descriptor using the 7 Hu's moments, applied to two regions: shoulders, hips and hips-knees, extracted from only 5 frames per sequence. Experiments were conducted using three databases (CASIAB and two Kinect databases) and evaluated with SVM and KNN classifiers. Comparative analysis with established research highlights the effectiveness of our approach. Results confirm the reliability of shoulders-hips and hips-knees regions as biometric features for gender recognition across various factors. The highest accuracy achieved was 95.19% using the SVM classifier with the Kinect v2 dataset.

Eye side and orientation detection of iris images using lightweight textural descriptors for embedded system

10 Jan 2025
16:10-18:10
Session S5V
Live stream

Chung Tran[1], Jordi Carrabina[2], David Castells Rufas[3], Minh Son Nguyen[4], Le-Anh Tran[5], Nhan Canh Dang[6]

[1]:chungnguyen.tran@autonoma.cat, [2]:jordi.carrabina@uab.cat,
[3]:david.castells@uab.cat, [4]:sonnm@uit.edu.vn,
[5]:latran@mindintech.com, [6]:cach@ut.edu.vn

[1,2,3]:Autonomous University of Barcelona, Spain
[4]:University of Information Technology, Vietnam
[5]:Mindintech, Inc, Vietnam
[6]:HCMC University of Transport, Vietnam

Summary

Iris recognition is a widely used biometric authentication technique due to its high accuracy and uniqueness. However, these systems are vulnerable to spoofing attacks, which can occur by rotating an iris image or an iris scanner during image acquisition. Additionally, correctly recognizing the eye side significantly reduces computational load and decreases the likelihood of false positives in biometric systems. This paper introduces a novel method for automatically detecting the correct left/right and upright/upside-down orientation of an iris image. The proposed method employs a lightweight feature extraction algorithm utilizing Local Binary Pattern (LBP) and Gray-Level Co-Occurrence Matrix (GLCM) to extract features from the iris image. A Support Vector Machine (SVM) classifier is then used to determine the eye side or orientation of the iris images. LBP captures the local texture details, whereas

GLCM describes the statistical features of the iris images. By combining these textural features, the proposed method improves the ability to classify eye side or orientation. The efficient and precise texture descriptors allow for implementation on embedded systems, including IoT or mobile devices. Experimental results on benchmark datasets demonstrate that the proposed method outperforms existing baseline methods in both performance and speed.

---

10 Jan 2025
16:10-18:10
Session S5V
Live stream

Segmentation of Ship Propeller Cavitation Area Using Image-Based Anomaly Detection Network

Wonje Jeong[1], Soon Yong Park[2], Yong-Jin Shin[3]

[1]:autowmacma@gmail.com, [2]:sypark@knu.ac.kr,
[3]:prop.shin@samsung.com

[1,2]:Kyungpook National University, Korea
[3]:Samsung Heavy Industries, Korea

Summary

Cavitation is the formation of vapor-filled cavities in the liquid. Cavitation in rotating propeller can cause various engineering problems to ships, such as hull vibration, noise, thrust reduction, and propeller corrosion. To address this cavitation problem in designing ship propellers, it needs to analyze the cavitation phenomenon in real propellers. In this study, we introduce a computer vision method to detect and segment cavitation areas in a rotating propeller. The pro-posed method captures the images of a high-speed rotating propeller in a water tunnel and detect the image area of the cavitation. As the first study in the im-age-based automatic cavitation detection, we employ a deep neural network, an anomaly detection method called DDAD (Anomaly Detection with Conditioned Denoising Diffusion Models). In addition, we also use a STN (Spatial Transformer Networks) network to compensate the image shaking caused by the vibration from the high-speed propeller and water flow. After the STN compensation, normal and abnormal images are automatically aligned to find cavitation areas more accurately. The proposed method is the first to use an anomaly detection network for ship propeller cavitation detection. Experimental results show that the proposed method can automatically detect the cavitation area without any annotation supervision.

Adversarial Encoder-Driven Filter for Targeted Image Tagging: A Novel Approach to Visual Content Manipulation

10 Jan 2025
16:10-18:10
Session S5V
Live stream

Cole McKee[1], Dominic Flowers[2], Jesse Wood[3], Ethan Shafer[4]
[1]:cole.mckee@westpoint.edu, [2]:dominic.flowers@westpoint.edu,
[3]:jesse.wood@westpoint.edu, [4]:ethan.shafer@westpoint.edu

United States Military Academy, United States of America

Summary

Computer vision, driven by artificial intelligence, has become pervasive in diverse applications such as self-driving cars and law enforcement. However, the susceptibility of these systems to attacks has raised significant concerns among researchers. This paper addresses the vulnerability of image tagging algorithms, particularly focusing on misclassifications induced by autoencoders. We present experiments conducted on Amazon Rekognition, where we developed a specialized autoencoder to manipulate the latent space, forcing it to align with specific tags. By integrating this manipulated latent space with other images, we demonstrate the ability to increase the confidence of a specific tag on Amazon Rekognition, leading to more false positives of the chosen tag. Our study showcases a practical method to exploit Amazon's Rekognition image tagging algorithm using a black box approach.

From classical techniques to convolution-based models: A review of object detection algorithms

10 Jan 2025
16:10-18:10
Session S5V
Live stream

Fnu Neha[1], Deepshikha Bhati[2], Deepak Shukla[3], Md Amiruzzaman[4]
[1]:neha@kent.edu, [2]:dbhati@kent.edu,
[3]:ds1640@scarletmail.rutgers.edu, [4]:mamiruzzaman@wcupa.edu

[1,2]:Kent State University, United States of America
[3]:Rutgers University, United States of America
[4]:West Chester University, Pennsylvania, United States of America

Summary

Object detection is a fundamental task in computer vision and image understanding, with the goal of identifying and localizing objects of interest within an image while assigning them corresponding class labels. Traditionally, object detection methods relied on handcrafted features and shallow architectures, which often struggled to handle the complexity of visual data. These approaches typically involved constructing ensembles of low-level image features and contextual information. However, their performance plateaued due to the limitations of manually designed features and the inability to fully capture high-level semantics. The advent of deep learning addressed many of these shortcomings. By introducing powerful tools such as Convolutional

Neural Networks (CNNs), deep learning enables models to automatically learn rich, hierarchical features directly from data. These learned features include both semantic and high-level representations that are essential for accurate object detection. This paper provides a comprehensive review of object detection frameworks, beginning with an overview of classical computer vision-based methods. We classify object detection approaches into two main categories: (1) classical computer vision based techniques and (2) CNN-based detectors. We also compare major CNN-based models, analyzing their strengths and limitations. In conclusion, this review highlights the significant progress made in the field of object detection through deep learning and draws attention to key areas where further research is needed to push the boundaries of performance.

Building Damage Assessment in Conflict Zones: A Deep Learning Approach Using Geospatial Sub-Meter Resolution Data

10 Jan 2025
16:10-18:10
Session S5V
Live stream

Matteo Risso[1], Alessia Goffi[2], Beatrice Alessandra Motetti[3], Alessio Burrello[4], Jean Baptiste Bove[5], Enrico Macii[6], Massimo Poncino[7], Daniele Jahier Pagliari[8], Giuseppe Maffeis[9]

[1]:matteo.risso@polito.it, [2]:a.goffi@terraria.com,
[3]:beatrice.motetti@polito.it, [4]:alessio.burrello@polito.it,
[5]:jeanbaptiste.bove@cri.it, [6]:enrico.macii@polito.it,
[7]:massimo.poncino@polito.it, [8]:daniele.jahier@polito.it,
[9]:g.maffeis@terraria.com

[1,3,4,6,7,8]:Politecnico di Torino, Italy
[2,9]:TerrAria s.r.l., Italy
[5]:Croce Rossa Italiana, Italy

Summary

Very High Resolution (VHR) geospatial image anal ysis is crucial for humanitarian assistance in both natural and anthropogenic crises, as it allows to rapidly identify the most critical areas that need support. Nonetheless, manually inspecting large areas is time-consuming and requires domain expertise. Thanks to their accuracy, generalization capabilities, and highly parallelizable workload, Deep Neural Networks (DNNs) provide an excellent way to automate this task. Nevertheless, there is a scarcity of VHR data pertaining to conflict situations, and consequently, of studies on the effectiveness of DNNs in those scenarios. Motivated by this, our work extensively studies the applicability of a collection of state-of-the-art Convolutional Neu ral Networks (CNNs) originally developed for natural disasters damage assessment in a war scenario. To this end, we build an annotated dataset with pre- and post-conflict images of the Ukrainian city of Mariupol. We then explore the transferability of the CNN models in both zero-shot and learning scenarios, demonstrating their potential and limitations. To the best of our knowledge, this is the

first study to use sub-meter resolution imagery to assess building damage in combat zones.

Disentangled-Transformer: An Explainable End-to-End Automatic Speech Recognition Model with Speech Content-Context Separation

10 Jan 2025
16:10-18:10
Session S5V
Live stream

Pu Wang[1], Hugo Van hamme [2]
[1]:pu.wang@esat.kuleuven.be, [2]:hugo.vanhamme@esat.kuleuven.be

Catholic University of Leuven (KU Leuven), Belgium

Summary

End-to-end transformer-based automatic speech recognition (ASR) systems often capture multiple speech traits in their learned representations that are highly entangled, leading to a lack of interpretability. In this study, we propose the explainable Disentangled-Transformer, which disentangles the internal representations into sub-embeddings with explicit content and speaker traits based on varying temporal resolutions. Experimental results show that the proposed Disentangled-Transformer produces a clear speaker identity, separated from the speech content, for speaker diarization while improving ASR performance.

Award Ceremony and Closure (Saturday 11 Jan 2025 at 12:30-13:30)
Jan 2025
12:30-13:30

Index