ASIP 2025

2025 7th Asia Symposium on Image Processing

June 13-15, 2025 ‖ Tsukuba, Japan

Sponsored By

筑波大学
University of Tsukuba

Supported By

天津市自动化学会

天津市信息传感与智能控制重点实验室

Technically Supported by

IEEE ComSoc
IEEE Communications Society
(Special Interest Group of E-Health)

北京联合大学
BEIJING UNION UNIVERSITY

Table of Content

Welcome Message

On behalf of the Conference Committee, we are pleased to welcome you to 2025 7th Asia Symposium on Image Processing (ASIP 2025) held in Tsukuba, Japan during June 13-15, 2025.

The conference is co-sponsored by University of Tsukuba, Japan and Tianjin University of Technology and Education, China. ASIP 2025 invites authors to submit papers on different aspects of image processing. Key areas of interest include, but are not limited to: target detection and recognition, digital image analysis and computer vision, application of AI in medical image processing,image detection model and classification algorithm, computer-aided design and image processing etc.

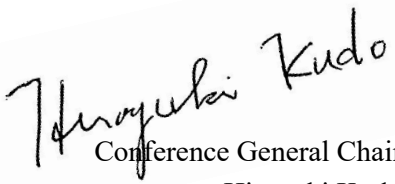
The conference aims to provide an interactive platform for professionals to learn about cutting edge academic and industrial trends, share the latest scientific research and technological achievements, and discuss innovative ideas and methods.

This year, we received more than 30 submissions from China, Malaysia, Thailand, India, Japan, the Philippines, Brunei, Saudi Arabia, Sri Lanka and other countries. More than 20 Technical Program Committee Members participated in the review process. Thanks for their great efforts and excellent work.

We are profoundly grateful to everyone who has helped make this event possible, including the respected authors, the keynote speakers, invited speakers and the peer reviewers. Special thanks also go to the conference committees for their dedication throughout the planning and execution of the conference.

We hope all participants will benefit from this event. Your contributions are essential in advancing the frontiers of knowledge and technology.

Wishing you a successful and inspiring conference experience at ASIP 2025!



Conference General Chair
Hiroyuki Kudo
University of Tsukuba, Japan
June 2025

Organizing Committees

Conference Chair
Hiroyuki Kudo, University of Tsukuba, Japan

Conference Co-Chairs
Hotaka Takizawa, University of Tsukuba, Japan
Mengmeng Zhang, Beijing Union University, China

Program Chairs
Jian Dong, Tianjin University of Education and Technology, China
Liming Zhang, University of Macau, Macau, China (IEEE Member)
Yanqiu Che, Tianjin University of Education and Technology, China

Chapter Chairs
Bok-Min Goi, Universiti Tunku Abdul Rahman, Malaysia (IEEE Member)
Abhishek Kunar, Chandigarh University, India (IEEE Senior Member)
M. Hassaballah, South Valley University, Egypt (IEEE Member)

Publication Chairs
Cheng Chin, Newcastle University, UK (IEEE Senior Member)
Kavita Thakur, Pt. Ravishankar Shukla University, India (IEEE Senior Member)
Aslina Baharum, Sunway University, Malaysia (IEEE Senior Member)

Publicity Chairs
Qiu Chen, Kogakuin University, Japan (IEEE Member)
Gordana Jovanovic Dolecek, Institute National INAOE, Mexico (IEEE Senior Member)
Yew Kee WONG, Hong Kong Chu Hai College, China
Lianly Rompis, Universitas Katolik De La Salle Manado, Indonesia (IEEE Senior Member)
Zhen Wang, Tianjin University of Finance and Economics, China

Special Session Chairs
Lianly Rompis, Universitas Katolik De La Salle Manado, Indonesia
Xuguang Zhang, Hangzhou Dianzi University, China

Local Organizing Chairs
Yasuhiko Igarashi, University of Tsukuba, Japan
Rashed Essam, University of Hyogo, Japan

Finance Chairs
Cheng Zhu, Tianjin University, China
Haytham Ashraf Abdelaal Ali, Sohag University, Egypt

Technical Program Committee Chair

Yen-Wei Chen, Ritsumeikan University, Japan

Technical Program Committee (By Alphabetic order)

Abhishek Kunar, Aryabhatt College of Engineering and Research Centre, India (IEEE Senior Member)
Adnan Fatih Kocamaz, Inonu University, Türkiye
Amir Akramin bin Shafie, International Islamic University Malaysia(IIUM), Malaysia (IEEE Member)
Amlan Basu, University of Strathclyde, UK
Anwaar Ul-Haq, Central Queensland University, Australia
Asim Ali Khan, Sant Longowal Institute of Engineering and Technology, India
Ata Jahangir Moshayedi, Jiangxi University of Science and Technology, China (IEEE Member)
Bao Shi, Inner Mongolia University of Technology, China
BASANT KUMAR, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India (IEEE Senior Member)
David E. Breen, Drexel University, USA (IEEE Senior Member)
Dinesh Bhatia, North Eastern Hill University, India
Djoni H. Setiabudi, Petra Christian University, Indonesia
Guangyan Wang, Tianjin University of Commerce, China
Haigang Wang, Chinese Academy of Sciences, China (IEEE Member)
Honggui Li, Yangzhou University, China
Hongmin Gao, Ho-hai University, China (IEEE Senior Member)
Hua Zheng, Fujian Normal University, China
Jan Kubicek, VSB-Technical University of Ostrava, Czech Republic
Jeffrey F. Calim, Fatima University, Philippines
Jinwei Wang, Nanjing University of Information Science & Technology, China
JungHwan Oh, University of North Texas, USA
Lei Tong, University of Leicester, UK
M. Hassaballah, South Valley University, Egypt (IEEE Member)
Mahesh Kumar Gellaboina, Honeywell International Sdn. Bhd., Malaysia (IEEE Senior Member)
Malik Zawwar Hussain, University of the Punjab, Pakistan
Momina Moetesum, Bahria University, Pakistan
Narendra D Londhe, National Institute of Technology, India (IEEE Senior Member)
Olarik Surinta, Mahasarakham University, Thailand
Omar Farooq, AMU, Aligarh, India (IEEE Senior Member)
Renjith V Ravi, Rashtriya Raksha University, India
Robert S Laramee, University of Nottingham, UK
Seokwon Yeom, Daegu University Gyeongsan, Korea
Shin-Feng Lin, NDHU, Taiwan, China (IEEE Senior Member)
Simi V.R, Manipal Institute of Technology, India
Sirikan Chucherd, Mae Fah Luang University, Thailand
Suraiya Jabin, Jamia Millia Islamia, India
Thumrongrat Amornraksa, King Mongkut’s University of Technology Thonburi, Thailand
Umesh Chandra Pati, National Institute of Technology, India (IEEE Senior Member)

Wadhah Zeyad Tareq, Istinye University,Turkey
Wenbo Wan, Shandong Normal University, China
Wenwu Wang, University of Surrey, UK (IEEE Senior Member)
Xiuxin Wang, Chongqing University of Posts and Telecommunications, China
Yahui Peng, Beijing Jiaotong University, China
Yasmeen M. George, Monash University, Australia
Yijun Yang, Xi'an Jiaotong University, China
Yuji Iwahori, Chubu University, Japan (IEEE Member)
Z.J.M.H. (Zeno) Geradts, University of Amsterdam,Nederland
Zahid Akhtar, State University of New York Polytechnic Institute, USA
Zawwar Husain Malik, University of the Punjab, Pakistan

Onsite Conference Notice

Conference Venue

Tsukuba International Congress Center, Japan

Address: 2-20-3 Takezono, Tsukuba City, Ibaraki Prefecture 305-0032, Japan

More Information about Tsukuba International Congress Center:

https://www.epochal.or.jp/en/

Transportation

Narita Airport

· **Direct Bus (Recommended)**

Service: Airport Liner NATT'S

Duration: Approximately 100–120 minutes

Drop-off: Tsukuba Center (about 10-minute walk to the conference center)

· **Train Route (Transfer required)**

Take the **JR Narita Line** or **Keisei Line** from Narita Airport to Nippori (or Ueno)

Transfer to the **JR Yamanote Line** to Akihabara

Take the **Tsukuba Express** from Akihabara to Tsukuba Station

Walk about 10 minutes to the conference center

Total Duration: Approximately 120–150 minutes

Haneda Airport

· **Train Route**

Take the **Keikyu Line** from Haneda Airport to Shinagawa

Transfer to the **JR Yamanote Line** to Akihabara

Take the **Tsukuba Express** from Akihabara to Tsukuba Station

Walk about 10 minutes to the conference center

Total Duration: Approximately 90 minutes to 120 minutes

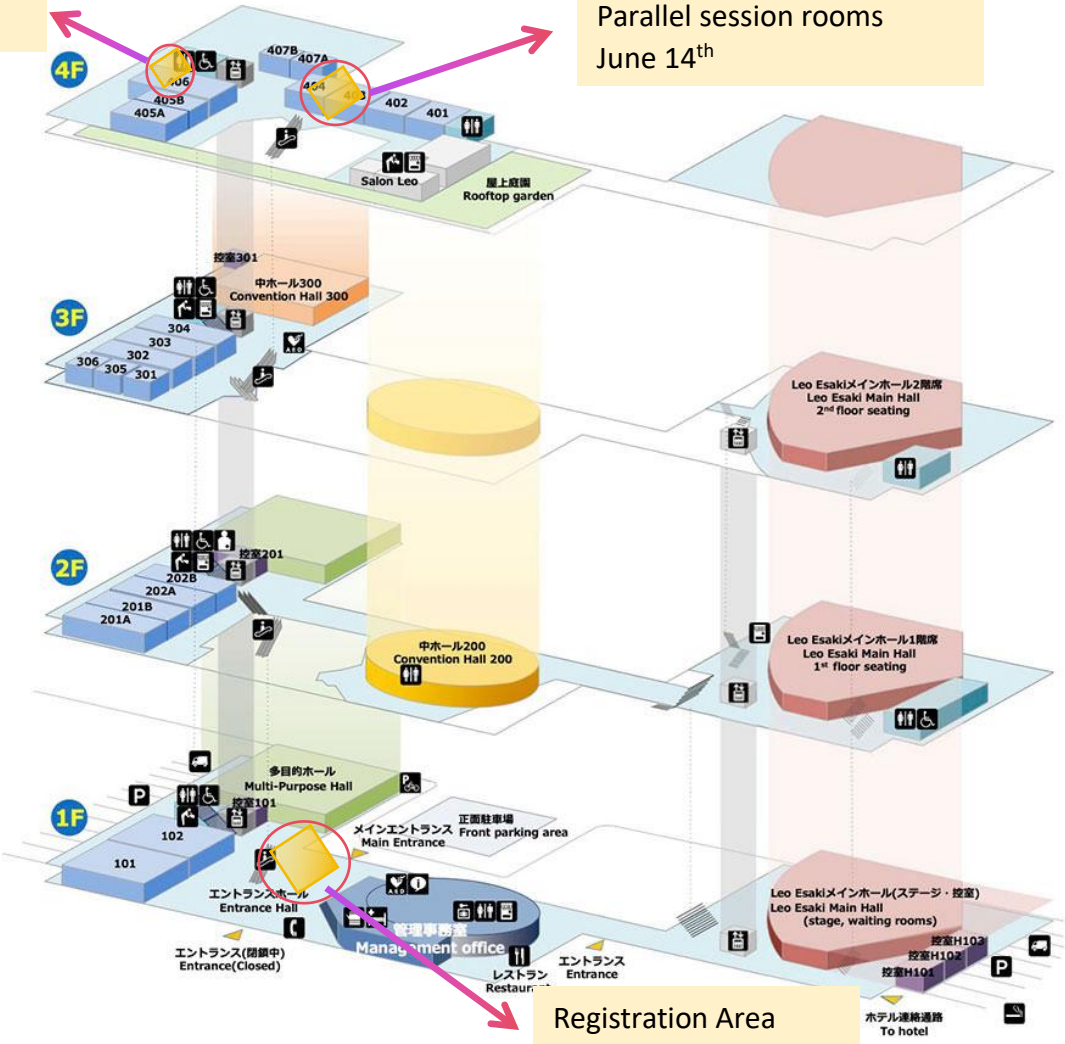
Conference Rooms Information

Tsukuba International Congress Center, Japan	
the Ground Floor	Registration
4th Floor, <406>	Opening Remark
	Keynote Speeches & Invited Speeches
4th Floor, <403>	Invited Speech V & Onsite Sessions 1&3
4th Floor, <404>	Onsite Sessions 2
Common Learning Area	Coffee Break & Group Photo
Lunch	4th Floor, <406>

Conference Venue Guideline

Main Venue
June 14th

Parallel session rooms
June 14th



Registration Area
June 13th

About Onsite Presentation

• Timing: a maximum of 20 minutes total, including speaking time and discussion. Please make sure your presentation is well timed.
• Each speaker is required to meet her / his session chair in the corresponding session rooms 10 minutes before the session starts and copy the slide file (PPT or PDF) to the computer.
• It is suggested that you email a copy of your presentation to your personal in box as a backup. If for some reason the files can't be accessed from your flash drive, you will be able to download them to the computer from your email.
• Please note that each session room will be equipped with a LCD projector, screen, point device, microphone, and a laptop with general presentation software such as Microsoft Power Point and Adobe Reader.

Name Badge

For security purposes, delegates, speakers, exhibitors and staff are required to wear their name badge to all sessions and social functions. Lending your participant card to others is not allowed. Entrance into sessions is restricted to registered delegates only. If you misplace your name badge, please ask the staff at the registration desk to arrange a replacement.

Gentle Reminder

• Please ensure that you take all items of value with you at all times when leaving a room. Do not leave bags or laptops unattended. The conference organizer does not assume any responsibility for the loss of personal belongings of the participants.
• Accommodation is not provided. Delegates are suggested make early reservation.
• Please show the badge and meal coupons when dining.

Emergency

Emergency Number: 119

Online Conference Notice

Platform: Zoom

Download Link: https://zoom.us/download

Sign In and Join

***Join a meeting without signing in.**

A Zoom account is not required if you join a meeting as a participant, but you cannot change the virtual background or edit the profile picture.

***Sign in with a Zoom account.**

All the functions are available.

Time Zone

UTC+9

***You're suggested to set up the time on your computer in advance.**

Online Room Information

Zoom ID: 812 7787 3849

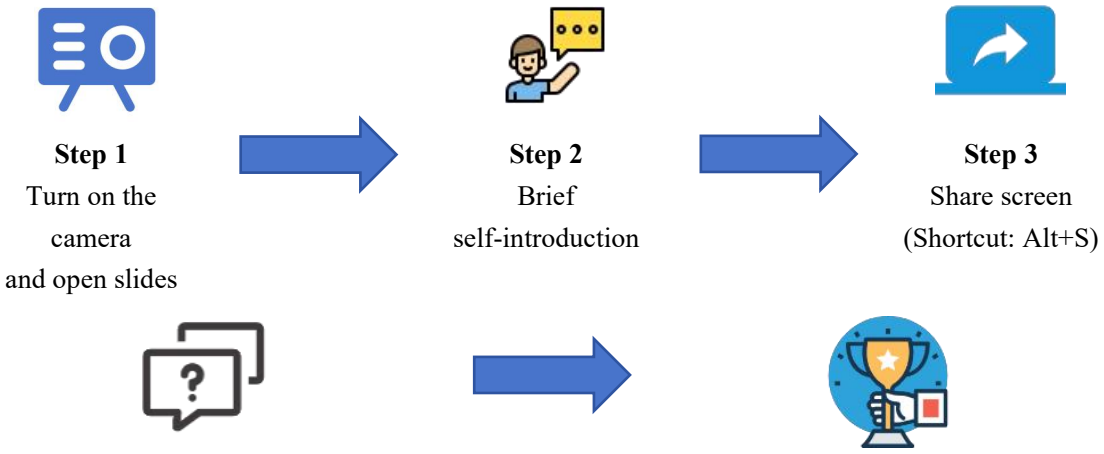
Zoom Link: https://us02web.zoom.us/j/81277873849

You can scan QR code to enter:

1. Prior to the formal conference, presenter shall join the test room to make sure everything is on the right track
2. Note: Please rename your Zoom Screen Name in below format before entering meeting room.

Role	Format	Example
Conference Committee	Position-Name	General Chair-Prof.
Keynote/ Invited Speaker	Position-Name	Keynote Speaker-Prof.
Author	Session Number-Paper ID-Name	S1-AP0001-Name
Delegate	Delegate-Name	LSAP0001-Name

Presentation Process by Zoom Meeting



About Online Presentation

- Every presenter has 20 minutes, including Q & A. Each presentation should have at least 10 minutes.
- The best presentation certificate and all authors' presentation certificates will be sent after conference by email.
- It is suggested that the presenter email a copy of his / her video presentation to the conference email box as a backup in case any technical problem occurs.

Environment & Equipment Needed

- A quiet place; Stable Internet connection; Proper lighting and background
- A computer with internet and camera; Earphone

Conference Recording

- We'll record the whole conference. If you do mind, please inform us in advance. We'll stop to record when it's your turn to do the presentation.
- The whole conference will be recorded. It is suggested that you should dress formally and we appreciate your proper behavior.

* The recording will be used for conference program and paper publication requirements. It cannot be distributed to or shared with anyone else, and it shall not be used for commercial nor illegal purpose.

Daily Schedule

June 13th (Friday)

Onsite Registration & Materials Collection for Onsite Participants

Time: 13:00-17:00

Venue:
The Ground Floor-Tsukuba International Congress Center, Japan

Address:
2-20-3 Takezono, Tsukuba City, Ibaraki Prefecture 305-0032, Japan

Registration Steps:

1. Arrive at the Tsukuba International Congress Center, Japan;
2. Inform the conference staff of your paper ID;
3. Sign your name on the Participants list;
4. Sign your name on Lunch & Dinner requirement list;
5. Check your conference kits: (1 conference program, 1 lunch coupon, 1 dinner coupon, 1 name card, 1 bag, 1 U disk);
6. Finish registration.

Zoom Testing for Online Participants

Zoom ID: 812 7787 3849

Zoom Link: https://us02web.zoom.us/j/81277873849

Online Test (UTC+9)

10:00-10:20	**Prof. Laurent David Cohen,** Universite Paris Dauphine, France
10:20-10:40	**Assoc. Prof. Ts. Dr. Aslina Baharum,** Sunway University, Malaysia
10:40-11:00	**Prof. Narendra D. Londhe,** National Institute of Technology Raipur, India
11:00-11:20	**Assoc. Prof. Xiwen Zhang,** Beijing Language and Culture University, China
11:20-11:40	**Session 4**
11:40-12:00	**Session 5**

You can attend the test of other session if you cannot manage it in your given time.

June 14th (Saturday)

Opening Remark & Keynote Speeches & Invited Speeches

Room / Time	Conference Main Venue: 4th Floor, <406>	
8:55-9:05	**Opening Remark**	**Prof. Hiroyuki Kudo**, University of Tsukuba, Japan
9:05-9:45	**Keynote Speech I**	**Prof. Hiroyuki Kudo**, University of Tsukuba, Japan
9:45-10:25	**Keynote Speech II**	**Prof. Keisuke Kameyama,** University of Tsukuba, Japan
10:25-10:45	**Invited Speech I**	**Prof. Seokwon Yeom,** Daegu University, Korea
10:45-11:00	**Group Photo & Coffee Break**	
11:00-11:20	**Invited Speech II**	**Assoc. Prof. Kazuya Ueki,** Meisei University, Japan
11:20-11:40	**Invited Speech III**	**Assoc. Prof. Muhammad Tariq Mahmood,** Korea University of Technology and Education, Korea
11:40-12:00	**Invited Speech IV**	**Assoc. Prof. Hongyuan Jing,** Beijing Union University, China
12:00-12:45	**Lunch**	

Invite Speech & Onsite Sessions

Room / Time	4th Floor, <403>	4th Floor, <404>
13:00-13:20	**Invited Speech V: Dr. Haytham Ashraf,** Sohag University, Egypt and University of Tsukuba, Japan	
13:20-15:00	**Session 1:** **Target Detection and Recognition** **Chair: Prof. Seokwon Yeom,** Daegu University, Korea **AP0020, AP0005, AP0007,AP0016, AP0006**	**Session 2:** **Digital Image Analysis and Computer Vision** **Chair: Prof. Mengmeng Zhang,** Beijing Union University, China **AP0019, AP0034, AP0003, AP0012, AP0027, AP0031**
15:00-15:10	**Coffee Break**	
15:10-16:30	**Session 3:** **Application of AI in Medical Image Processing** **Chair: Dr. Haytham Ashraf,** Sohag University, Egypt and University of Tsukuba, Japan **AP0004, AP0013, AP0029, AP0030-A**	
17:30-20:00	**Dinner Time - Don-Tei**	

June 15th (Sunday)

Keynote Speech & Invited Speeches & Online Sessions

(UTC+9)

Room / Time	Zoom ID: 812 7787 3849 Zoom Link: https://us02web.zoom.us/j/81277873849	
9:00-9:40	**Keynote Speech III**	**Prof. Laurent David Cohen,** Universite Paris Dauphine, France (IEEE Fellow)
9:40-10:00	**Invited Speech VI**	**Assoc. Prof. Ts. Dr. Aslina Baharum,** Sunway University, Malaysia
10:00-10:20	**Invited Speech VII**	**Assoc. Prof. Xiwen Zhang,** Beijing Language and Culture University, China
10:20-10:40	**Invited Speech VIII**	**Prof. Narendra D. Londhe,** National Institute of Technology Raipur, India
10:40-10:45	**Break Time**	
10:45-13:05	**Session 4:** **Image Detection Model and Classification Algorithm** **Chair: Assoc. Prof. Honggui Li,** Yangzhou University, China **AP0008, AP0022, AP0024, AP0014, AP0032, AP0033, AP0018**	
13:05-15:45	**Session 5:** **Computer-aided Design and Image Processing** **Chair: Prof. Kavita Thakur,** Pt. Ravishankar Shukla University, India **AP0011, AP0017, AP0036, AP0021, AP0023, AP2002, AP0037, AP0035**	

Keynote Speech I

Prof. Hiroyuki Kudo

University of Tsukuba, Japan

Biography

He received the B.Sc. degree from the Department of Electrical Communications, Tohoku University, Japan, in1985, and the Ph.D. degree from the Graduate School of Engineering, Tohoku University, in 1990. In 1992, he joined the University of Tsukuba, Japan. He is currently a Professor with the Institute of Systems and Information Engineering, University of Tsukuba, Japan. His research areas include medical imaging, image processing, and inverse problems. In particular, he is actively working on tomographic image reconstruction for X-ray CT, PET, SPECT, and electron tomography. He received best paper awards more than 10 times from various international and Japanese societies. He received the IEICE (The Institute of Electronics, Information, and Communication Engineers, Japan) Fellow award for his contributions on “cross-sectional image reconstruction methods in medical computed tomography”. In 2018, he obtained Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology for his contributions on “research on design method and image reconstruction method for new CT”. For 2011-2016, he was an Editor-in-Chief of the Journal of Medical Imaging Technology (MIT). From 2020, he is a president of Japanese Society of Medical Imaging Technology (JAMIT).

Speech Info.

Speech Title: Image Reconstruction for Sparse-View CT and Interior CT : From Compressed Sensing to Deep Learning

Abstract: Since 2000, new designs of CT scanners called sparse-view CT and interior CT have been actively investigated in CT community. The sparse-view CT refers to CT in which the number of measured projection data is reduced to accelerate the data acquisition, reduce the patient dose, or due to other physical reasons. The interior CT refers to CT in which x-rays are radiated only to a small region of interest (ROI) such as heart/breast (in the case of cardiac/mammographic imaging) to reduce the patient dose or due to other physical reasons. The key in these new CT scanners lies in how to reconstruct an accurate image from the incomplete projection data arising in each new CT. In the sparse-view CT, the used frameworks to solve this difficult reconstruction problem have been mainly Compressed Sensing (CS) and Deep Learning (DL). In the interior CT, the used frameworks have been mathematically analyzing solution uniqueness, CS, and DL. In this talk, I will overview the progresses of image reconstruction in each new CT. The talk will be constructed as follows. First, I will talk about image reconstruction in the sparse-view CT. After a short explanation on the introductory material, I will explain the image reconstruction using CS and DL including our research activities, where I put a focus on how the shift of used framework from CS to DL led to big advances. Second, I will talk about image reconstruction in the interior CT. After a short explanation on the introductory material and mathematical solution uniqueness in the interior CT, I will explain methodologies of image reconstruction to achieve an accurate reconstruction including our research activities, where I put a focus on how the shift of used techniques from the non-DL approaches to the DL approach led to big advances.

Keynote Speech II

Prof. Keisuke Kameyama

University of Tsukuba, Japan

Biography

Keisuke Kameyama received the B.E., M.E., and PhD degrees from Tokyo Institute of Technology, Tokyo, Japan. From 1992 to 2000 he worked as a Research Associate at Tokyo Institute of Technology. In 2000, he joined the Tsukuba Advanced Research Alliance (TARA),
University of Tsukuba. Currently, he works as a Professor at the Institute of Systems and Information Engineering, University of Tsukuba. He has been working in the field of pattern recognition, signal processing, and neural networks for applications such as image classification, image restoration, media retrieval and biometric authentication. Dr. Kameyama is a member of IEEE, IEICE, JNNS and APNNS.

Speech Info.

Speech Title: siFrom transfer learning to “trans-model curriculum learning” in neural networks

Abstract: The role of neural networks as the core unit for various artificial intelligence technology is becoming ever important. In image processing and related fields, the use of neural networks is widely spread to classification, identification, retrieval, reconstruction, filtering, transformation and generation of media signals in various applications. The drive behind the broad use of neural networks resides in their robust ability to acquire desired mappings through learning. However, the learning in many cases, is time consuming process with much uncertainty. The procedure for obtaining a right function in a right network model is not straightforward; a large-scale network may perform well but require a lot of resource to run, whereas a relatively lightweight network may not learn well enough and could underperform.

Such mismatch of suitable network models in training and deployment have been tackled by way of transfer learning, which tries to crossover different models while maintaining or incrementally improving the networks’ maps. This multi-stage operation may be viewed as a single learning process which involves switching of network models, training sets and evaluation criteria (loss functions) according to a given schedule (curriculum) or operational rules.

In this talk, I will review techniques and research efforts in Knowledge Distillation (KD), Neural Architecture Search (NAS), and Curriculum learning (CL) which are approaches to extend neural network learning to non-static learning conditions. Furthermore, I will discuss about the possible research directions that could be explored in such a general framework for neural network learning.

Keynote Speech III

Prof. Laurent David Cohen

Universite Paris Dauphine, France

(IEEE Fellow)

Biography

Laurent David Cohen was born in 1962. He was student at the Ecole Normale Superieure , rue d'Ulm in Paris, France from 1981 to 1985. He received the Master's and Ph.D. degrees in Applied Mathematics from University of Paris 6, France, in 1983 and 1986, respectively. He got the Habilitation à diriger des Recherches from University Paris 9 Dauphine in 1995.

From 1985 to 1987, he was member at the Computer Graphics and Image Processing group at Schlumberger Palo Alto Research, Palo Alto, California and Schlumberger Montrouge Research, Montrouge, France and remained consultant with Schlumberger afterwards. He began working with INRIA, France in 1988, mainly with the medical image understanding group EPIDAURE.

He obtained in 1990 a position of Research Scholar (Charge then Directeur de Recherche 1st class) with the French National Center for Scientific Research (CNRS) in the Applied Mathematics and Image Processing group at CEREMADE, Universite Paris Dauphine, Paris, France. His research interests and teaching at university are applications of Partial Differential Equations and variational methods to Image Processing and Computer Vision, like deformable models, minimal paths, geodesic curves, surface reconstruction, Image segmentation, registration and restoration.

For many years, he has been editorial member of the Journal of Mathematical Imaging and Vision, Medical Image Analysis and Machine Vision and Applications.

Speech Info.

Speech Title: Shortest Paths and Front Propagation for Image Segmentation. Application to Biomedical images.

Abstract: Minimal paths have been used for long as an interactive tool to find contours or segment tubular and tree structures, like vessels in medical images. The user usually provides start and end points on the image and gets the minimal path as output, as a cost minimizing curve.
These minimal paths correspond to minimal geodesics according to some relevant metric defined on the image domain. Finding geodesic distance and geodesic paths can be solved by the Eikonal equation using the fast and efficient Fast Marching method.
Minimal paths are a way to find the global minimum of a simplified active contour energy.
In the past years, we have extended the minimal path methods with asymmetric Randers Metrics to cover all kinds of active contour energy terms, as well as segmentation by front propagation.
For example a way to penalize the curvature in the framework of geodesic minimal paths was introduced, leading to more natural results in vessel extraction or object segmentation in natural images.
Recently, we introduced new methods combining the efficiency of minimal paths with CNN.
In a first method, CNN are trained to find a set of keypoints and minimal paths are found that link these keypoints.
In another context, CNN are used to generate relevant metrics adapted to a problem.

Invited Speech I

Prof. Seokwon Yeom

Daegu University, Korea

Biography

Seokwon Yeom has been a faculty member of Daegu University since 2007. He has a Ph.D. in Electrical and Computer Engineering from the University of Connecticut in 2006.
He has been a guest editor of Applied Sciences and Drones in MDPI since 2019. He has served as a board member of the Korean Institute of Intelligent Systems since 2016, and a member of the board of directors of the Korean Institute of Convergence Signal Processing since 2014. He has been program chair of several international conferences. He was a vice director of the AI homecare center and a head of the department of IT convergence engineering at Daegu University in 2020-2023, a visiting scholar at the University of Maryland in 2014, and a director of the Gyeongbuk techno-park specialization center in 2013. He has been a keynote or invited speaker at several international conferences.

Speech Info.

Speech Title: Tracks-to-Track Association for Broken Tracks in Search and Rescue Missions with a Drone

Abstract: This invited talk deals with multi-target tracking using a small drone equipped with a thermal infrared camera. Thermal objects are detected using a YOLO detection model trained on a custom dataset. Then, the targets are tracked using a Kalman filter. Track association and fusion select the fittest tracks and fuse them during track formation. Track segment association connects broken track segments over time. In the experiment, three lost hikers in the mountains were captured by a thermal camera mounted on a drone. Robust tracking results are obtained in terms of total trajectory lifetime, average trajectory lifetime, and trajectory purity.

Invited Speech II

Assoc. Prof. Kazuya Ueki

Meisei University, Japan

Biography

He received a B.S. in Information Engineering in 1997, and an M.S. in the Department of Computer and Mathematical Sciences in 1999, both from Tohoku University, Sendai, Japan. In 1999, he joined NEC Soft, Ltd., Tokyo, Japan. He was mainly engaged in research on face recognition. In 2007, he received a Ph.D. from Graduate School of Science and Engineering, Waseda University, Tokyo, Japan. In 2013, he became an assistant professor at Waseda University. He is currently an associate professor in the School of Information Science, Meisei University. His current research interests include pattern recognition, video retrieval, character recognition, and semantic segmentation. He is currently working on the video retrieval evaluation benchmark (TREVID) sponsored by the National Institute of Standards and Technology (NIST), contributing to the development of video retrieval technology. In 2016, 2017, and 2022, his submitted systems achieved the highest performance in the TRECVID AVS task.

Speech Info.

Speech Title: Multimodal Approaches to Video Search: The Power of Vision-Language Models

Abstract: This talk introduces three multimodal approaches to enhance video retrieval performance. First, we investigate how text generation techniques can expand and diversify search queries. Second, we show how image generation techniques help visualize and clarify query intent. Third, we showcase how vision-language models (VLMs) effectively bridge visual and textual modalities, resulting in more accurate and efficient retrieval. By integrating these methodologies, our goal is to achieve significant improvements in large-scale video retrieval systems by generating richer and more semantically aligned representations, ultimately providing higher search accuracy and more relevant results.

Invited Speech III

Assoc. Prof. Muhammad Tariq Mahmood

Korea University of Technology and Education, Korea

Biography

He received the MCS degree in computer science from AJK University of Muzaffarabad, Pakistan, in 2004. After That he worked as a Software Engineer for more than 8 years at Khaksar and Co. Islamabad Pakistan. Then, in 2005, he made a significant shift in his career by leaving software development and by joining various institutes for his higher studies and research. He received the MS degree in intelligent software systems from Blekinge Institute of Technology, Sweden, in 2006 and the PhD degree in information and mechatronics from Gwangju Institute of Science and Technology, Korea, in 2011. Now, he is working as an Associate Professor at School of Computer Science and Engineering, Korea University of Technology and Education, Cheonan, Korea. His research interests include image processing, 3D shape recovery from image focus, computer vision, pattern recognition and machine/deep learning. Currently, he is working on various projects funded by National research foundation (NRF), Korea related to shape from focus/defocus, smart cities and underwater imaging.

Speech Info.

Speech Title: Leaning Depth from Focus via Multi-Scale Recurrent Networks

Abstract: Depth estimation is a fundamental task in computer vision, enabling machines to perceive and interpret the 3D structure of the scene from 2D images. Among passive methods, depth from focus (DFF) is particularly advantageous due to its simple setup and the non-requirement of additional hardware. A deep learning framework for depth-from-focus is proposed that estimates scene depth from images captured at different focus settings by leveraging a recurrent network to iteratively refine depth estimates. Unlike conventional deep learning methods that collapse a 3D focus volume (a volumetric representation encoding the per-pixel focus likelihood across the focal stack) into a 2D depth map via direct regression — which often amplifies errors and loses spatial context — our approach treats depth estimation as an iterative optimization process. First, an encoder backbone extracts multi-scale features from each focal slice, while a separate context encoder processes a reference image to provide global scene cues. A dedicated focus mapping module then fuses these multi-scale features into single-channel focus volumes that capture per-pixel focus likelihood across the focal stack. Building on these focus volumes, a recurrent depth extraction module comprising multiple Gated Recurrent Unit (GRU) layers at different resolutions iteratively refines the depth prediction, with joint supervision on all iterative estimates to produce accurate depth maps. Experiments on both synthetic and real-world datasets demonstrate that our framework produces quantitatively accurate and qualitatively sharp depth maps and generalizes better on unknown real-world datasets than state-of-the-art methods.

Invited Speech IV

Assoc. Prof. Hongyuan Jing

Beijing Union University, China

Biography

Hongyuan Jing is currently an Associate Professor at the School of Artificial Intelligence, Beijing Union University, China. He received his Ph.D. in Engineering from the Communications and Embedded Systems Laboratory at the University of Leicester, UK in 2019. He serves as a technical program committee member and session chair for several international conferences such as ICGIP and ICCCS. His research interests include computer vision, deep learning, object detection, image restoration, and multi-sensor fusion. He has published more than 30 papers in SCI-indexed journals including IEEE TIM and IoT. He has led or participated in multiple research projects funded by the EU FP7 program, China's National Key R&D Program, the Beijing Natural Science Foundation, and industry partners. He won the championship in the CVPR-NTIRE Multi-Scenario Raindrop Removal Challenge. He also serves as a reviewer for journals such as MS and TVC et.al.

Speech Info.

Speech Title: TAE-Net: Top-k Attention Enhancement Network for Single Image Dehazing

Abstract: Single image dehazing is an ill-posed problem. Due to the presence of water mist particles, the refraction of light in the air is uneven, making it difficult to restore a true and clear image. The haze in the image will affect the later advanced computer vision tasks, such as target detection, instance segmentation, etc., and will also affect the safety and reliability of intelligent monitoring and unmanned driving. At present, many works on image dehazing use CA (Channel Attention) and PA (Pixel Attention)[8] to enhance the areas with dense haze and rich details, better process the features, and improve the dehazing effect. However, this CA and PA mechanism will still give a certain weight to the areas that do not need to be paid attention to at all, resulting in the distraction of network attention. Therefore, we propose to use top-k to enhance the attention mechanism, that is, to enhance the important areas in the front part and give a very small value to the unimportant parts. The model can pay full attention to the important parts and enhance the features twice.

Invited Speech V

Asst. Prof. Haytham Ashraf

Sohag University, Egypt and University of Tsukuba, Japan

Biography

Dr. Haytham Ashraf is a JSPS Postdoctoral Fellow at the University of Tsukuba, Japan, and an Assistant Professor at Sohag University, Egypt. He received his B.Sc. degree in Mathematics and Computer Science from Sohag University in 2013, and his M.Sc. in Mathematics from the same university in 2018. In 2020, he joined the University of Tsukuba, where he completed his Ph.D. in Engineering in 2023.

His research focuses on solving inverse problems related to medical imaging, with a particular interest in deep learning for computed tomography (CT) imaging. His doctoral work centered on geometric tomography using parametric level-set methods. He is passionate about developing innovative solutions that contribute to improved healthcare outcomes and broader societal impact.

Currently, his research explores the integration of deep learning techniques into sparse-view and limited-angle CT reconstruction. By combining classical reconstruction approaches with modern AI methods—such as generative models—he aims to achieve high-fidelity image reconstruction from limited data.

He has published widely in international journals and conferences and has received multiple honors, including the prestigious JSPS Fellowship and several best presentation awards.

Speech Info.

Speech Title: Dual-Stage Self-Attention GAN for High-Quality CT Reconstruction from Limited and Incomplete Data

Abstract: Reconstructing high-fidelity computed tomography (CT) images from limited-angle or sparse-view data remains a fundamental challenge due to severe noise, artifacts, and loss of structural detail inherent in under-sampled acquisitions. Traditional analytical methods such as Filtered Back Projection (FBP) fail to deliver reliable diagnostic quality under such conditions, while many deep learning approaches struggle with generalization and artifact suppression.

In this work, we will present a novel Dual-Stage Self-Attention GAN framework designed to robustly reconstruct high-quality CT images from incomplete data. The proposed method integrates two synergistic stages: an initial artifact correction stage employing a U-Net-based generator with residual learning and multi-scale self-attention for capturing global context and suppressing coarse artifacts, followed by a refinement stage that enhances fine structures and texture fidelity through adversarial training with a PatchGAN discriminator.

To ensure both perceptual realism and quantitative accuracy, we leverage a comprehensive loss function combining L1 loss, SSIM, perceptual loss from a pretrained VGG network, and adversarial loss. This hybrid objective guides the network in preserving anatomical integrity while minimizing reconstruction error.

Extensive experiments on simulated and clinical CT datasets demonstrate that our approach outperforms state-of-the-art methods in PSNR, SSIM, and MSE, while qualitatively delivering clearer, artifact-free reconstructions with improved structural detail. This dual-stage framework offers a promising solution for low-dose and limited-angle CT imaging scenarios where data acquisition is constrained.

Invited Speech VI

Asso. Prof. Ts. Dr. Aslina Baharum

Sunway University, MALAYSIA

Biography

Associate Professor Ts. Dr. Aslina Baharum (Dr. Ask) holds the esteemed position of Associate Professor at the School of Engineering and Technology within Sunway Uni-versity. Previously, she has served as a Senior Lecture at the Faculty of Computer and Mathematical Sciences in Universiti Teknologi MARA (UiTM), and as a Senior Lecturer at the Faculty of Computing and Informatics in Universiti Malaysia Sabah (UMS), where she led the User Experience (UX) research group. Completing her academic journey, she also brings valuable industry experiences as a former IT Officer at the Forest Research Insti-tute of Malaysia (FRIM). She had experienced more than 20 years in the IT field.

She earned her PhD in Visual Informatics from UKM, a Master Science degree in IT from UiTM, and Bachelor of Science (Hons.) in E-Commerce from UMS. Dr. Ask is an active member of the Young Scientists Network - Academy of Science Malaysia, a Senior Mem-ber IEEE, and a certified Professional Technologist recognized by MBOT. She has further contributed to the field by serving as an auditor for MBOT/MQA.

She has received medals at research and innovation showcases and has been honored with awards for her teaching, excellence in service, and outstanding contributions as a researcher. Her bibliography showcases her prolific output, including co-authored and co-edited books, over 20 book chapters, technical papers presented at conferences, and more than 60 peer-reviewed and indexed journals publications. She has also taken on editorial roles for several journals and actively participated as a committee member, ses-sion chair, and part of editorial teams while actively participating as a reviewer. Dr. Ask has graced numerous conferences with her wisdom, delivering keynote, invited and ple-nary talks.

Her research interests span a wide spectrum, encompassing UX/UI, HCI/Interaction De-sign, Product & Service Design, Software Engineering & Mobile Development, Infor-mation Visualization & Analytics, Multimedia, ICT, IS and Entre/Technopreneurship. Dr. Ask's expertise extends beyond the academic realm; she imparts her knowledge through workshops and talks on various subjects,

including UI/UX, Entrepreneurship, Vid-eo/Image Editing, E-Commerce/Digital Marketing, STEM, Design Thinking and etc.

Furthermore, she is certified as a Professional Entrepreneurial Educator, Executive En-trepreneurial Leader, and HRDF Professional Trainer, which highlights her strong com-mitment to education and entrepreneurship. Dr. Ask is highly regard in her field, dedicat-ed and consistently pushing the boundaries of knowledge and sharing her wealth of ex-pertise with others.

Speech Info.

Speech Title: Humanizing Machines: Advancing Multimodal Interaction Through Emotion-Aware Image Processing.

Abstract: As artificial intelligence continues to evolve, the boundary between human and machine interaction becomes increasingly dynamic. In this study, we explore the transformative role of emotion-aware image processing in enabling more intuitive and empathetic human-computer interactions. By integrating computer vision, affective computing, and multimodal data sources, including facial expressions, gesture analysis, gaze tracking, and physiological signals, we can better model user context, intent, and emotional state. This study will discuss recent breakthroughs and challenges in building adaptive, emotionally intelligent interfaces across domains such as healthcare, education, and assistive technologies. Drawing on real-world projects and user-centred design frameworks, I will share insights into designing multimodal systems that not only see and hear but also feel. This study aims to foster interdisciplinary collaboration between researchers in image processing, HCI, and cognitive science, toward the goal of humanizing digital interaction for next-generation intelligent systems.

Invited Speech VII

Assoc. Prof. Xiwen Zhang

Beijing Language and Culture University, China

Biography

XiWen Zhang is currently a full professor of Digital Media Department, School of Information Science, Beijing Language and Culture University.

Prof. Zhang worked as an associated professor from 2002 to 2007 at the Human-computer interaction Laboratory, Institute of Software, Chinese Academy of Sciences. From 2005 to 2006 he was a Post doctor advised by Prof. Michael R. Lyu in the Department of Computer Science and Engineering, the Chinese University of Hong Kong. From 2000 to 2002 he was a Post doctor advised by Prof. ShiJie Cai in the Computer Science and Technology department, Nanjing University.

Prof. Zhang's research interests include pattern recognition, computer vision, and human-computer interaction, as well as their applications in digital image, video, and ink. Prof. Zhang has published over 60 refereed journal and conference papers. His SCI papers are published in Pattern Recognition, IEEE Transactions on Systems Man and Cybernetics B, Computer-Aided Design. He has published more than twenty EI papers.

Prof. Zhang received his B.E. in Chemical equipment and machinery from Fushun Petroleum Institute (became Liaoning Shihua University since 2002) in 1995, and his Ph.D. advised by Prof. ZongYing Ou in Mechanical manufacturing and automation from Dalian University of Technology in 2000.

Speech Info.

Speech Title: Intelligently Recognizing and Generating Information from Digital Image

Abstract: Due to pattern recognition and deep learning, various information can be recognized and generated from image. Our work has focused on the proposed hierarchy models, local homogeneity, and adversarial generation.

Various digital images are recognized, such as ones scanned from mechanical paper drawings and paper text, face images, portrait ones with line drawings, and microscopic bone marrow images.

Various information is recognized using the proposed hierarchy models. Graphics and their multi-levels compounded objects are recognized from images scanned from mechanical paper drawings using a hierarchy model of engineering drawings. Faces and their components are recognized from photos using a facial model.

Various information is recognized using the proposed local homogeneity. Karyocytes and their components from microscopic bone marrow images based on regional color features.

Various information is generated from image using cycle-Consistent adversarial networks. Text is separated from grid background using cycle-Consistent adversarial networks. Digital images of Chinese classical upper-class lady paintings are generated from images with line drawings using conditional generative adversarial networks.

Invited Speech VIII

Prof. Narendra D. Londhe

National Institute of Technology Raipur, India

Biography

Dr. Narendra D. Londhe is presently working as Professor in the Department of Electrical Engineering of National Institute of Technology Raipur, Chhattisgarh, India. He completed his B.E. from Amravati University in 2000 followed by M.Tech. and Ph.D. from Indian Institute of Technology Roorkee in the years 2006 and 2011, respectively. He has 15 years of rich experience in academics and research. He has published more than 170 articles in recognized journals, conferences, and books. His main areas of research include medical signal and image processing, biomedical instrumentation, speech signal processing, biometrics, intelligent healthcare, brain–computer interface, artificial intelligence, and pattern recognition. He has been awarded by organizations like Taiwan Society of Ultrasound in Medicine, Ultrasonics Society of India, and NIT Raipur. He is an active member of different recognized societies from his areas of research including senior membership of IEEE.

Speech Info.

Speech Title: Reviewing Progress and Innovations in Devanagari Script-Based P300 Spellers: A Historical Perspective on Advances Over Time

Abstract: The P300 speller is a foundational brain-computer interface (BCI) system that enables communication without muscular movement by leveraging the P300 event-related potential (ERP) from EEG signals. Initially designed for English, the growing emphasis on inclusive neurotechnology led to adaptations for various languages, including the Devanagari script (DS), which supports several major Indian languages. This talk offers a focused review of Devanagari Script-Based P300 Spellers (DSP3S), highlighting key innovations aligned with usability, speed, and accuracy goals. The DSP3S pipeline, comprising stimulus presentation, data acquisition, preprocessing, feature extraction, classification, and character detection, has been substantially refined. Display paradigms have progressed from basic row-column and zigzag layouts to more engaging dual-character and facial-expression paradigms, enhancing user convenience and P300 detection through improved stimulus saliency. Simultaneously, limitations of traditional machine learning approaches in handling noisy, low-signal EEG data led to the adoption of deep learning architectures, which offer superior capacity for modelling non-linear patterns and generalizing across sessions. This shift has enabled significant gains in classification accuracy and robustness. State-of-the-art research has increasingly targeted single-trial detection and compact models to boost information transfer rates (ITRs). In parallel, channel selection has evolved from optimization to attention-based deep learning, improving performance while reducing computational overhead. In addition, spelling correction and predictive text approaches have been integrated to enhance user experience and reduce cognitive load. Notably, recent developments have achieved **94% single-trial classification accuracy** with an average spelling time of just **2.2 seconds per character**, marking a significant step toward real-time communication. Looking ahead, the trajectory of DSP3S research is moving toward achieving robust real-time performance, with a particular focus on deploying energy-efficient, portable hardware solutions that balance responsiveness with practical deployment constraints.

Session 1

Topic: Target Detection and Recognition
Session Chair: Prof. Seokwon Yeom, Daegu University, Korea
Time: 13:20-15:00, June 14th, 2025
Onsite Room: 4th Floor, <403>

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters in this session.**

AP0020

End-to-End Multi-Domain Joint Learning for Robust Person Re-Identification in Large-Scale Surveillance

Yuan-Kai Wang, Tung-Ming Pan, Tian-You Chen and **Yan-Lok Sing**
Fu Jen Catholic University

Abstract-End-to-end person re-identification (re-ID) remains a critical challenge due to domain shifts, inconsistent feature representations, and the complexity of integrating detection, tracking, and retrieval. This study proposes a multi-domain joint learning (MDJL) framework for an end-to-end person re-ID system, incorporating pedestrian detection, multi-object tracking, person re-ID, and keyframe extraction to improve real-world applicability. The framework enhances feature learning by refining domain-guided dropout (DGD) with refined neuron dropout (RND) and optimizing re-ranking through recursive reciprocal expansion (RRE). Unlike conventional methods that rely on manually cropped pedestrian images, the proposed system processes raw surveillance footage, enabling seamless person tracking across multiple cameras. Extensive experiments on multiple datasets validate the model's superior accuracy and generalization. The results demonstrate a significant improvements on cross-camera tracking and person re-ID performance for large-scale surveillance applications.

AP0005

YOLO-Based Hip Fracture Detection with Hyperparameter Tuning

Praechompoo Nilbanjong, Prin Twinprai, Nattaphon Twinprai, Patinya Muangkammuen and Puripong Suthisopapan
Khon Kaen University, Khon Kaen, Thailand

Abstract-Hip fractures are a significant global health issue that can occur across all ages and genders. The condition is a complex disorder that limits human visual perception and causes a certain margin of error during diagnosis through X-ray inspection, with high misdiagnosis rates in worst-case scenarios reaching up to 30%. Despite the advancement in the current artificial intelligence (AI) model for hip fracture detection, small prediction errors still persist. In order to address these issues, the study proposed an AI model for hip fracture detection through an optimization strategy with hyperparameter tuning. The strategy combines grid search and genetic algorithm based on the state-of-the-art YOLO algorithm. Consequently, the results demonstrate that meticulous tuning of 33 hyperparameters significantly affects model performance. The tuning strategy contributes to a notable outcome, the optimal model achieves an average accuracy of 0.982 on a test set comprising 1,144 images from internal and

external clinical sources. A significant improvement over the suboptimal model achieves an average accuracy of 0.943. Furthermore, the proposed method demonstrates the potential of hyperparameter tuning not only to improve diagnostic accuracy but also to enhance clinical decision-making in hip fracture diagnosis.

AP0007

Deep Learning for Tear Meniscus Analysis: A YOLOv8 Approach to Dry Eye Disease Diagnosis

Anand[1], **Bala Murugan MS[1]** and Yoshiro Okazaki[2]
1. Vellore Institute of Technology, Chennai, India
2. Waseda University, Tokyo, Japan

Abstract-Dry Eye Disease (DED) is a prevalent ocular condition characterized by insufficient tear production or excessive tear evaporation, leading to discomfort, visual impairment, and reduced quality of life. Accurate assessment of the tear meniscus height (TMH) is crucial for diagnosing and managing DED. Traditional manual evaluation techniques are often subjective and prone to variability, necessitating the development of automated, objective methods. This study presents a YOLOv8-based deep learning model for real-time detection and measurement of TMH, leveraging object detection techniques rather than traditional segmentation approaches. The dataset consists of 40 ocular images, augmented to enhance generalization. The model was trained using CIoU loss, classification loss, and distribution focal loss (DFL) to improve localization precision. Experimental results indicate that the model achieves a mean Average Precision (mAP) of 86.7%, precision of 74.9%, and recall of 87.5%, demonstrating its efficacy in identifying tear meniscus regions. Loss function analysis further validates the model's convergence and stability. The proposed YOLOv8-based approach offers a fast, objective, and automated alternative to manual assessments, enhancing clinical efficiency in diagnosing DED. Future work will focus on expanding the dataset and integrating multi-modal imaging for improved accuracy.

AP0016

Accuracy Improvement for License Plate Character Recognition Using Learning Method

Yuki Esumi and Tomio Goto
Nagoya Institute of Technology, Nagoya, Japan

Abstract-Increased social awareness of traffic problems has led to an increase in the number of opportunities to see drive recorder footage. In fact, drive recorder footage is used in many situations for analyzing the cause of accidents and resolving troubles. On the other hand, the degradation of drive recorder images due to motion blur and out-of-focus blur still remains an issue. This degradation reduces the visibility of license plates in the drive recorder footage, negatively affecting the evidentiary ability of the drive recorder and severely hindering the resolution of traffic problems. Therefore, this paper aims to improve the accuracy of license plate character recognition by using a method that can recover degraded images that include both motion blur and blur due to out-of-focus images. We propose three methods with an improved training dataset. Experimental results show that the proposed method improves the accuracy of character recognition.

AP0006

Fully Automated Femoral Neck-Shaft Angle Measurement Based on Keypoint Detection

Thanachai Jantana, Nattaphon Twinprai, Prin Twinprai, Patinya Muangkammuen and Puripong Suthisopapan

Khon Kaen University, Khon Kaen, Thailand

Abstract-The femoral neck-shaft angle (FNSA) is a critical parameter to evaluate hip joint biomechanics in orthopedic medicine and plays a vital role in diagnosing pathologies in newborns, children, and adults. Normally, physicians manually measure FNSA by selecting key points along the femoral neck and shaft on radiographs to create reference axes, essential for FNSA calculation via the Picture Archiving and Communication System (PACS). Unfortunately, this manual measurement is time-consuming and varies due to the subjective nature of landmark placement. In order to assist physicians in advancing faster and more accurate evaluations of the FNSA measurement, deep learning-based keypoint detection is introduced to identify the critical femoral landmarks on radiographs. From the results, it is indicated that this proposed model can automatically identify 16 keypoints with an average accuracy of 97.19% and object keypoint similarity of 0.9873. Furthermore, in terms of clinical application, the model achieves an average absolute deviation of 3.6909°, staying within the 5% threshold deemed acceptable by specialists. Moreover, the prediction closely aligns with expert physician measurements. As a result, the findings suggest that the proposed approach can enhance orthopedic diagnoses, particularly FNSA evaluation, as well as improve patient treatment through a reliable, consistent, and efficient solution.

Session 2

Topic: Digital Image Analysis and Computer Vision
Session Chair: Prof. Mengmeng Zhang, Beijing Union University, China
Time: 13:20-15:00, June 14th, 2025
Onsite Room: 4th Floor, <404>

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters in this session.**

AP0019

Self-Attention Guided Low-Light Image Enhancement GAN with adaptive Multi-Objective Loss Optimization

Ruofan Jia, Xiaofeng Li, Likun Wu and **Minhang Zhang**
University of Electronic Science and Technology of China, China

Abstract-Low-light image enhancement in zero-shot area has gained attention as it relies on image priors for brightness adjustment avoiding the need for paired low/normal-light images. However, complex lighting and noise affect prior extraction, leading to noise and artifacts in reconstructed images. To address this issue, we propose a Self-Attention-guided Generative Adversarial Network (SA-GAN) for low-light image enhancement, which will improve the visual quality and details of images. Furthermore, a multi-objective adaptive balancing mechanism is introduced to dynamically optimize the weights of multiple loss functions, achieving optimal enhancement. Extensive experiments on multiple benchmarks demonstrate that the proposed method achieves performance comparable to state-of-the-art approaches.

AP0034

Scalable AI Pipeline for Specified Colour Transfer from Concept Sketches into Static Animation Frames

Bo-Wei Chen, Yi-Chun Yeh, **Shih-Yen Wu**, Ching-Ying Yu, Long-Xuan Huang, Hsiao-Chu Chiou and Yi-Ming Chen
National Central University

Abstract-Every technological innovation brings new possibilities to the animation industry. This study explores the application of Artificial Intelligence (AI) in the 2D animation production process. In this study, we establish an AI Colour filling Pipeline, which includes a variety of steps and model components. For the core part of generative colour fill, we use the U-Net architecture of Pix2Pix as the foundation, combined with the Conditional Generative Adversarial Network (CGAN), to train a model that can generate colour filling results according to the specified colour, which we name 2DColorGAN. The training materials, sourced from Midjourney and Kaggle, are classified into character design drawings (CHD), character sketch drawings (CHS) and character sketch finished drawings (CHSF), with motion consistency, colour consistency and colouring accuracy as the evaluation indicators. The automatic colour filling and colouring process is divided into two stages: data collation and component colouring. The former helps users organize character images, while the latter relies on the segmentation model 2DColorSEG for colouring. The results show that 2DColorSEG has a high degree of stability and accuracy for general components

but performs less effectively on unique components due to limited recognition capabilities. In contrast, 2DColorGAN has higher colour flexibility and can handle components that cannot be covered by colour filling, but it still faces colour stability and overflow issues, which could be mitigated by increasing data diversity. Overall, the combination of automated colour fill and generative colour fill can improve the overall colour filling effect and application flexibility.

AP0003

Extending Heritage Tourism Beyond Reality with Extended Reality Technologies: A Systematic Literature Review

Abby Lian Hendrick[1], **Suriati Khartini Jali[1]**, Mohammad Imran Bandan[1], Phei Chin Lim[1], Yin Chai Wang[1], Ally Dian Hendrick[1], Nurul Farizah Ridzuan[1], Johannes Hamonangan Siregar[2] and Rufman Iman Akbar[2]

1. Universiti Malaysia Sarawak, Kota Samarahan, Malaysia
2. Universitas Pembangunan Jaya, Jakarta, Indonesia

Abstract-This paper is a systematic literature review that outlines the evolution of heritage tourism with the use of extended reality technologies. Through the use of three XR technologies, augmented reality, virtual reality, and mixed reality, travelling becomes a more immersive experience. This review explores the types of extended reality technologies in heritage tourism, what tourists prioritize when using them, and what considerations need to be taken by developers when implementing. The Preferred Reporting Items for Systematic reviews and Meta-Analyses statement was adopted as the methodology. Reports from 2018 to 2025 were gathered from SCOPUS, Web of Science, ScienceDirect, and ProQuest as the source databases. The identification process was according to keywords related to extended reality technologies, heritage, and tourism as a singular domain. 37 reports were deemed eligible after screening for their relevance to the topic, document types, publication status, and written in English. From the conducted analysis, virtual reality was applied more often than augmented reality and mixed, with a steady rise across the years. Tourists prioritized engagement with extended reality technology the most while user collaboration and cognitive-affective factors the least. For the developers, reliability when using extended reality technology was regarded as the biggest concern but information architecture of the system was the least. The findings of this review were limited to English documents for consistency and academic publications according to institutional access. From the analysis, the evolution of heritage tourism relies on humans, the natural environment and technologies as interdependent elements of an ecosystem. The relationship encourages the growth of each involved party socioeconomically and technologically. Future research should investigate applicable methodologies and usability metrics of extended reality technology across diverse forms of heritage.

AP0012

Multi-Frame Fusion-Based Ground Target Localization for Unmanned Aerial Vehicles

Shiyi Xiong, Yali xue and Liangchen Xie

1. Nanjing University of Aeronautics and Astronautics, NanJing, China
2. ByteDance, Beijing, China

Abstract-The ground target localization algorithm, based on homogeneous coordinate transformation, plays a crucial role in unmanned aerial vehicle target reconnaissance and outdoor search and rescue. However, its accuracy is hampered by sensor measurement noise. To address this issue, we propose an improved YOLOv5 algorithm with a detection head and attention mechanism to reduce positioning errors in the image coordinate system. We further employ a Kalman filtering algorithm for multi-frame image data fusion, effectively mitigating errors from

single-frame positioning. Our results show a significant increase in unmanned aerial vehicle localization accuracy compared to other algorithms.

AP0027

Image-Based Data Acquisition System for Wireless Sensor Networks Using Python

Wang Yucheng[1] and Zhang Xiaohua[2]
1. Wenzhou Semir United International School, China
2. Wenzhou Business College, China

Abstract-Wireless sensor networks face challenges such as complex node structures and massive data volumes when acquiring image data, making it difficult to quickly extract the diverse image information required by users. To address this issue, this paper proposes an Image-Based Data Acquisition System for Wireless Sensor Networks Using Python. The hardware of the system includes the selection of image sensors, wireless transceiver chips, and serial communication circuit design, while the software consists of an image data acquisition framework, data compression module, and data caching module. Through the integration of hardware and software, the system can efficiently acquire and process large volumes of image data and optimize data transmission efficiency. Experimental results show that compared to the network data acquisition system based on the Android security container, the proposed system achieves shorter acquisition time, higher compression rates, and lower energy consumption, demonstrating its superior performance in image data acquisition.

AP0031

Enhanced q-logarithm Based Image Watermarking Using Fuzzy Image Filter

Piyanart Chotikawanid and **Thumrongrat Amornraksa**
King Mongkut's University of Technology Thonburi, Thailand

Abstract-Watermarking is important for copyright protection, preventing unauthorized use of content, and establishing ownership, particularly in digital format such as images and videos. Robust image watermarking aims to embed a watermark that can survive various intentional or unintentional attacks, e.g. noise addition, compression, and geometric transformations, while maintaining the perceptual quality of the original image. In this paper, an improved version of a robust image watermarking method based on the q-logarithm transform is proposed. With a new prediction model of original q-logarithm component based on fuzzy image filtering technique, the embedded watermark can be blindly extracted, and withstand more manipulation and attacks. The proposed watermarking method is tested using two categories of images, i.e. photographic and computer-generated images, and its performance is evaluated using Bit Correction Rate (BCR) and Normalized Correlation (NC) metrics. The results at various wPSNRs and common attacks are presented and compared to the two previous watermarking methods [5] and [6] to demonstrate its superior performance.

Session 3

Topic: Application of AI in Medical Image Processing
Session Chair: Dr. Haytham Ashraf, Sohag University, Egypt and University of Tsukuba, Japan
Time: 15:10-16:30, June 14th, 2025
Onsite Room: 4th Floor, <403>

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters in this session.**

AP0004

Gamified eXtended Reality Potential for Kenyah and Kayan Traditional Medicine Knowledge Transfer and Preservation

Ally Dian Hendrick, Suriati Khartini Jali, Nurul Farizah Ridzuan, **Jane Labadin**, Abby Lian Hendrick and Juna Liau
Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

Abstract-The rapid modernization of Sarawak presents a significant threat to the preservation of traditional knowledge within its indigenous communities. Among the most vulnerable aspects is traditional medicinal knowledge, an intangible cultural heritage at risk of fading without immediate intervention. To address this challenge, this study investigates the potential of an extended reality-based application designed to facilitate the dissemination of traditional medicinal knowledge. The application integrates gamified augmented reality and non-immersive virtual reality elements. Its effectiveness is evaluated using the System Usability Scale, Intrinsic Motivation Inventory, Augmented Reality Sickness Questionnaire, observational analysis, and open-ended qualitative feedback. The findings provide insights into the usability of the application, its capacity to enhance user engagement, and critical design considerations for developing extended reality-based tools aimed at preserving and transmitting traditional medicinal knowledge.

AP0013

Optimizing Source Recovery for fMRI by Integrating Biological and Mathematical Priors with Data-Driven Learning

Muhammad Usman Khalid[1], **Malik Muhammad Nauman[2]**, Hatoon S. AlSagri[1] and Muhammad Abid[2]
1. College of Computer and IS IMAM University (IMSIU), Riyadh, Saudi Arabia
2. Faculty of Integrated Techn. Universiti Brunei Darussalam, Bandar Seri Begawan, Brunei

Abstract-Despite its popularity, dictionary learning (DL) for functional magnetic resonance imaging (fMRI) data faces a significant challenge in accurately modeling the sources signals. This problem mainly arises because the conventional loss function tends to overfit by aligning too closely with the training data. To address this, our paper introduces an integration of data-driven learning with biological and mathematical priors and regularizes the model complexity. This novel approach resulted in two algorithms MPDL and MBPDL, which aim to prevent overfitting in DL while retaining the strict objective of minimizing the loss function. Using the proposed synthesis cost function,

which exploits quad-factorization for matrix approximation, these algorithms leverage coordinate descent to solve rank-1 approximations. Both algorithms effectively model temporal smoothness and spatially reduce false positives, enhancing the precision of source signal recovery. Their effectiveness was assessed by comparing them against existing DL techniques, using both experimental and synthetic fMRI datasets. The mean correlation value and F-score was found to be higher for the proposed algorithms than stateof-the-art ODL, ACSD, and ACSDBE algorithms. The MATLAB code for all algorithms used in this study is available at.

AP0029

Retinal Fundus Image Enhancement for CNN-Based Multiclass Eye Disease Classification

Alvin Choo[1], **Stephanie Chua[1]**, Lik Thai Lim[1], Dayang Nurfatimah Awang Iskandar[1], Mohd Hanafi Ahmad Hijazi[2] and Puteri Nor Ellyza Nohuddin[3]

1. Universiti Malaysia Sarawak, Malaysia
2. Universiti Malaysia Sabah
3. Higher Colleges of Technology, Sharjah, United Arab Emirates

Abstract-Retinal fundus imaging plays a vital role in the early detection and monitoring of various eye conditions by providing detailed views of retinal structures. The advancement of Convolutional Neural Networks (CNNs) has driven the development of automated eye disease classification using retinal fundus images. However, variations in image quality, such as low contrast and uneven illumination, can affect model performance. This study investigates the impact of applying image enhancement techniques on the performance of models in multiclass eye disease classification. Specifically, we applied grayscale conversion and Contrast Limited Adaptive Equalization (CLAHE) to different image channels, resulting in five dataset variants: (1) original dataset, (2) grayscale images, (3) RGB channel-enhanced images (with CLAHE applied to each channel separately), (4) green channel-enhanced images, and (5) L* channel from CIELAB color space enhanced images. Two CNN models, DenseNet121 and EfficientNet-B0, were trained and evaluated under identical hyperparameter settings. Experimental results show that L* channel enhancement consistently outperforms others, with DenseNet121 achieving accuracy of 82.63% %, recall of 82.63% and precision of 84.12%, and EfficientNet-B0 achieving accuracy of 81.09%, recall of 81.09%, and precision of 81.99%. These findings highlight the effectiveness of targeted channel-wise enhancement in improving the performance of CNN models and highlight the importance of image enhancement in medical image classification tasks.

AP0030-A

An Integrity Assurance Framework for Medical Images in Telemedicine Applications

Chia-Chen Lin, Yen-Heng Lin, En-Ting Chu and Ku-Yaw Chang
National Chin-Yi University of Technology

Abstract-Over the past few decades, digital images have been integrated into a wide range of applications [1, 2], gaining significant attention for their role as a crucial source of information within the healthcare sector. The increasingly advanced image editing software also poses an indirect threat, as these enhanced tools make it easier for malicious users to illegally tamper with medical images, thereby indirectly facilitating incidents such as medical insurance fraud [3, 4]. As a result, how to prevent theft and tampering of medical data
has become an issue of growing concern.

Session 4

Topic: Image Detection Model and Classification Algorithm
Session Chair: Assoc. Prof. Honggui Li, Yangzhou University, China
Time: 10:45-13:05, June 15th, 2025

Zoom ID: 812 7787 3849
Zoom Link: https://us02web.zoom.us/j/81277873849

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters in this session.**

AP0008
Using YOLOv9 to Classify the Freshness of Loligo Duvauceli
Joaquin S. Santos II, Miguel Angel A. Rioveros and Carlos C. Hortinela IV
Mapúa University, Manila, Philippines

Abstract-This study presents an automated image processing system based on the YOLOv9 algorithm for classifying the freshness of Loligo duvauceli (Indian squid) using color analysis. Traditional methods for evaluating squid freshness are often subjective and invasive, leading to inconsistencies in quality control. To address this, the proposed system employs HSV color filtering, segmentation, and object detection using YOLOv9, integrated with OpenCV, to non-invasively assess freshness indicators from squid images captured under controlled lighting conditions. The system achieved an overall classification accuracy of 95.31%, outperforming earlier deep learning models such as Faster R-CNN and CNN-based methods in both speed and accuracy. This improvement highlights YOLOv9's potential as a real-time, scalable solution for seafood quality assessment. The findings contribute to the advancement of non-destructive quality control technologies in the aquaculture and food safety industries.

AP0022
3D Object Detection for Near-field Applications
Shuta Doura and Tomio Goto
Nagoya Institute of Technology, Nagoya Japan

Abstract-Accurate perception of the surrounding environment is crucial for improving the safety of autonomous driving. This study proposes a 3D object detection model that detects surrounding vehicles to prevent collisions in autonomous driving. We introduce a novel loss function that prioritizes short-range detection accuracy, as it is particularly critical for collision avoidance. This loss function is integrated into a Virtual Sparse Convolution model. Our experiments demonstrate that the proposed method effectively improves the accuracy of short-range object detection tasks. The results of this study are expected to contribute to the safety of autonomous driving systems. This approach prioritizes the accuracy of object detection at close range, which is particularly effective in scenarios that require emergency braking or rapid evasive maneuvers just before stopping. For instance, it enables the autonomous vehicle to respond appropriately to vehicles making unpredictable movements. As a result, safety considerations for

drivers, pedestrians, and other vehicles are enhanced. Additionally, the proposed method allows for accurate identification of surrounding vehicles even in congested traffic situations, where traditional object detection models often struggle. This enables rapid and reliable decision-making regarding vehicle movement.

AP0024

Transfer Learning with Linear Channel Adaptation for Parkinson's Disease Detection in T1-MRI

Rui Yan and Saimei Guo
Hebei University of Science and Technology, China

Abstract-Early detection of Parkinson's disease (PD) via magnetic resonance imaging (MRI) is critical for timely interventions, yet limited annotation datasets challenge the model generalization of deep learning approaches. This study introduces a promising transfer learning framework that adapts pre-trained models on natural images to brain T1-weighted MRI scans. The framework presents a novel linear-fitting technique to align single-channel grayscale medical images with RGB inputs of pre-trained models to preserve the well-trained features in the pre-trained models. Complementary strategies include data augmentation and a class-weighted optimizer are adopted to enhance generalization and address class imbalance. Model performance is evaluated on 1414 subjects from the PPMI dataset via 5-fold cross-validation achieving 88.2% accuracy, 0.84 F1-score. Experimental results demonstrated the superiority of transfer learning over training models from scratch in both convergence speed and detection accuracy.

AP0014

Growth Habit Classification of Bamboo Using YOLOv5 Algorithm

Carla C. Ty, Dustin Jeiondre A. Ventura and Noel B. Linsangan
Mapúa University, Manila, Philippines

Abstract-Bamboo is vital to the Philippines' ecology and economy. However, classifying its species and growth patterns remains challenging due to limited data and reliance on time-consuming, error-prone morphological methods. Furthermore, accurate classification is affected by ecological factors, reducing reliability. This research aims to develop a device to distinguish between running and clumping bamboo growth habits using image data. It utilizes the YOLOv5 algorithm to automatically detect and classify bamboo growth habit patterns, allowing for efficient processing of image-based datasets. The system's performance was assessed using a confusion matrix, achieving an overall accuracy of 86%. Despite the promising results, further improvements are recommended. Enhancing the dataset, optimizing camera positioning for better image capture, and refining the model to adapt to varying environmental conditions could improve classification accuracy and real-world applicability.

AP0032

Object Detection for Grain-Leveling Robots in Complex Granary Environments

Chi Zhang[1], Bin Yang[2], Xiaoguang Zhou[3] and Xiaosong Zhu[1]
1. Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin University of Technology and Education, Tianjin, China
2. Tianjin Office, Hangzhou HollySys Automation Co., Ltd., Tianjin University of Technology and Education, Tianjin, China
3. Beijing University of Posts and Telecommunications, Beijing, China

Abstract-Addressing the challenges of grain-leveling robots detection in complex granary environments characterized by low-light and heavy-dust, an improved YOLO v11 model is proposed, which integrates two specialized modules. The one is the frequency spectrum dynamic aggregation (FSDA) module embedded within the backbone's C3K2 blocks, which applies Fourier transforms and channel attention mechanisms to filter and fuse adaptively frequency domain features, thereby suppressing noise from low-light and dust particulates. The other one is the multi-scale grouped dilated convolution (MSGDC) module in the neck, which employs parallel grouped dilated convolutions with varying dilation rates to extract and merge features at multiple spatial scales, enhancing adaptability to target size variations. We constructed a dataset of images captured by a surveillance camera in a 13 m diameter silo for model training and evaluation. Experimental results demonstrate that the proposed model achieves a precision of 0.83 and a recall of 0.71 (absolute increase of 0.11 and 0.09 over the original model, respectively). The speed is 78 FPS, which meets real-time requirements. Ablation studies further validate the independent and synergistic contributions of the FSDA and MSGDC modules. The results confirm that our method provides robust and reliable grain-leveling robot detection under challenging low-light and heavy-dust conditions.

AP0033

PDM-Based Graph Cut Clustering Method for Streamline Visualization

Zhen Wang[1], **Chenpeng Zhao[2]**, Xionghao Zhao[1], Xiaoxue Li[1] and Zhan Wang[1]
1. Tianjin University of Finance and Economics, Tianjin, China
2. Yanshan University, Qinhuangdao, China

Abstract-For ocean vector field visualization, the streamline visualization system usually generates the streamline set covering the marine environment vector field, then various clustering methods are applied to cluster the streamline sets and select representative streamlines to alleviate clutter and occlusion. In this paper, we propose a new semi-supervised streamline clustering approach to simplify the vector field to visualize the marine environmental data, at the will of the user by selecting several streamlines from each target cluster. By using the schema of the graph cut in spectral clustering methods, streamlines are modeled as vertices of a weighted graph, and the PDM, Hausdorff, MCP distance are applied as a distance metric in weight function defined on all pairs of vertices for measuring similarity between the two vertices. The clustering algorithm is proposed as a minimization of a convex functional of the energy function, and the region-force term is calculated by the affinity between each vertex and the labeled vertices, characterizing the conditional probability of each vertex belonging to each class. Additionally, we apply the MDS algorithm to accelerate computing speed. Experiments are conducted using ocean flow field data from China coast, our proposed method improves the Hausdorff and MCP similarity measures by 0.133 and 0.06 compared to AHC respectively, which validates the effectiveness and superiority of the method

AP0018

Philippine License Plate Classification Using K-Nearest Neighbors

Robne Kyle L. Fonseca, Neil Ade D. Martinez and Rosemarie V. Pellegrino
Mapúa University, Manila, Philippines

Abstract-Vehicles tend to outnumber the available road space, resulting in dense traffic that authorities struggle to manage. Previous studies have applied YOLOv7 for license plate detection and K-Nearest Neighbors (KNN) for text recognition in countries such as Bangladesh. However, limited attention has been given to categorizing Philippine license plates by color schemes and identifying high-ranking government plates under real-time conditions. Earlier

Philippine research mainly relied on convolutional neural networks (CNN) and conventional color segmentation methods, often lacking emphasis on lightweight, real-time systems. The system's goal is to automate the categorization of vehicle classes and identify vehicles assigned to Philippine government officials, helping future researchers gather structured license plate classification data more efficiently. It specifically aims to: (1) create hardware for capturing videos on vehicles on the road in real-time using Raspberry Pi Camera Module v3, (2) utilize YOLOv7 model to detect license plates of vehicles captured by the hardware and classify license plate color schemes with KNN and high-ranking government plates using Tesseract OCR, and (3) evaluate the accuracy of the system using a confusion matrix. The system achieved an overall accuracy of 86.60%. Researchers recommend using Jetson Nano and adopting newer models to improve real-time inference and frame rates.

Session 5

Topic: Computer-aided Design and Image Processing
Session Chair: Prof. Kavita Thakur, Pt. Ravishankar Shukla University, India
Time: 13:05-15:45, June 15th, 2025

Zoom ID: 812 7787 3849
Zoom Link: https://us02web.zoom.us/j/81277873849

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters in this session.**

AP0011

Hanunoo Character Identification using Visual Saliency Models

John Carlo P. Galicia, Mark Wilbur A. Cordero and Jocelyn F. Villaverde
Mapúa University, Manila, Philippines

Abstract-This study focuses on the development and implementation of a prototype device aimed at preserving the Hanunoo script, an indigenous writing system used by the Hanunoo Mangyan people of Mindoro, Philippines. The Hanunoo script holds significant cultural value, as it is used to record myths, folklore, and other oral traditions. However, recognizing and interpreting Hanunoo characters poses challenges due to its complex structure. To address this, the study employs Visual Saliency Model (VSM) algorithms, which are commonly used in computer vision for identifying and highlighting relevant visual elements, to automatically detect and identify 12 distinct characters from the Hanunoo script. The 12 target characters—A-Ba-Ka-Da, I-Bi-Ki-Di, and U-Bu-Ku-Du—are drawn from a dataset consisting of 100 images per character. This approach aims to simplify the process of character recognition and enhance accessibility to the Hanunoo script, enabling easier learning and teaching without requiring fluency in the language. In this case, the system model managed to identify 41 out of 48-character samples correctly, achieving an accuracy of 85.42%for the model. The performance and accuracy of the saliency model in recognizing Hanunoo characters from complex and disorganized images could be further enhanced by incorporating adaptive thresholding and advanced image enhancement techniques in image preprocessing steps and adding more datasets.

AP0017

Metal Artifact Removal in CT Images Based on U-Net Network with Fusion Attention Mechanism

Jian Dong, Qiyu Wang, Minting Wu and **Qiyu Wang**
Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin University of Technology and Education, China

Abstract-Deep learning has achieved good results in the task of Metal Artifact Reduction (MAR) on Computer Tomography (CT) images, but existing methods generally face the problem of poor model interpretability. Therefore, this paper proposes a U-Net network model that integrates channel attention and spatial attention modules. The model adopts the classic encoding-decoding structure, inserts an attention mechanism module between each layer of

convolution operations, and optimizes the feature decoding process with the help of the jump connection mechanism, thereby enhancing the feature extraction capability in the MAR task and improving the quality of CT images. Experimental results show that this method has significant advantages in. MAR for clinical CT images of the chest and abdominal. In terms of visual effects, it can effectively reduce metal artifacts while retaining image detail information. From the perspective of objective evaluation indicators, excellent performance has been achieved in terms of Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The research results not only verify the excellent performance of this method in the CT image MAR task, but also provide new ideas for the optimization and application of deep learning in the field of medical image processing.

AP0036

Computer Art AIGC-Driven Design Path Exploration for Guangxi Ethnic Cultural Virtual Idol Images

LI YuanCheng
Guangxi Vocational Normal University, China

Abstract-With the development of computer art and artificial intelligence generated content (AIGC) technology, virtual idol image design has new opportunities and challenges. This paper focuses on the innovative needs of Guangxi ethnic culture in digital communication and explores the AIGC-driven virtual idol image design path. By systematically reviewing relevant research and analyzing the shortcomings of existing methods in cultural element integration and personalized expression, a trinity innovative path of "cultural knowledge base- model optimization-collaborative design" is proposed. Based on the characteristics of Guangxi ethnic cultural elements, the AIGC generation model is optimized, and the precise design of virtual idol images is realized by combining multimodal interaction technology. Experimental results show that this path effectively improves the cultural fit and visual expressiveness of virtual idol images, providing a new paradigm for the digital communication of ethnic cultures. The research proves that AIGC technology can effectively integrate Guangxi ethnic cultural elements to create virtual idol images with cultural connotations and a modern sense, offering new ideas and methods for the inheritance and communication of ethnic cultures.

AP0021

Investigation of Super-Resolution Technology Based on Image Alignment

Zhang Jiawei and Tomio Goto
Nagoya Institute of Technology, Nagoya Japan

Abstract-In this paper, we propose a super-resolution algorithm based on deep learning. To address the issues of traditional down sampling methods, we align images with different magnifications, remove image pairs unsuitable for training, and modify the loss function to create the dataset. We propose a super-resolution method based on SwinIR and validate its effectiveness by training and testing real-world scene image data.

AP0023

AN INVESTIGATION ON GRAPH CONVOLUTION NETWORKS FOR COLOR ENHANCEMENTS OF V-PCC

Zeliang LI[1], Yu Liu[2], Siu-Kei Au Yeung[1], Shuyuan Zhu[2] and Kevin Hung[1]
1. Hong Kong Metropolitan University
2. University of Electronic Science and Technology of China, Chengdu, China

Abstract-Dynamic Point Cloud (DPC) technology represents realistic 3D scenes in motion and has a wide range of applications. The Moving Picture Experts Group (MPEG) has developed Video-based Point Cloud Compression (V-PCC), achieving remarkable DPC compression performance. However, V-PCC introduces issues such as point reduction and coordinate distortion in the decoded DPC during the compression. In our previous work [14], we proposed an interpolation architecture to address the issue of point reduction by performing interpolation and coordinate adjustment on the decoded DPC. However, the attributes of the interpolated points were not addressed. In this paper, as a continuation of our previous work, we introduce a graph convolution network with an attention module incorporating geometric features to perform the color adjustment. Experimental results demonstrate that the interpolated DPC exhibits significant improvements in color quality, as evidenced by both objective and subjective evaluations. The highest PSNR gains compared to the interpolation architecture were 0.26 dB for the "redandblack" sequence and 0.17 dB for the "soldier" sequence.

AP2002

Enhancing Boosting-Guided Image Captioning with Scene-Graph-Based Vision Transformer Integrated with GAN

Yizhen Wang

Shandong Vocational College of Science and Technology, China

Abstract-Image Captioning (IC), which involves scrutinizing an image and creating a textual explanation based on the identified elements and their interactions, commonly suffers from semantic inconsistency between the text and image, resulting in less human-like captions. As a solution, we introduce an image captioning strategy called Boosting Guided Image Captioning Method with Scene-graph-based Vision Transformers and Generative Adversarial Networks (BICViTGAN). Our method utilizes Adaptive Boosting to integrate scene-graph-based vision transformers (ViTs) with generative adversarial networks (GANs). This approach notably lessens the issue by thoroughly extracting semantic information from the image. Initially, we convert the given image and its linked captions into scene graphs. These are then fed into a ViT integrated with a GAN, generating the initial phrase. Subsequently, we compare this phrase with the original captions and utilize the same manner to train another ViT to produce the second phrase. We continue this process until we have textually represented the entire image and the generated phrases are not discriminated by discriminators. The BICViTGAN model has shown outstanding results on both the MS COCO and Flickr30k datasets, notably outperforming the state-of-the-art existing methods. Thorough ablation studies have been conducted to confirm the effectiveness of each component of the proposed model.

AP0037

Research on Visualization Design of Ecological Environment Image Processing Empowered by AIGC Based on Optimization of LSTM-CNN Model

LI YuanCheng

Guangxi Vocational Normal University, China

Abstract-This study addresses the challenges of complex data image processing and visualization in ecological environment management by proposing a solution for AIGC-empowered image processing visualization design based on the optimization of the LSTM-CNN model. Three AI tools, DeepSeek, Kimi, and Doubao, optimized by the

LSTM-CNN model, are employed to predict the concentrations of ozone, carbon monoxide, nitrogen dioxide, and PM2.5, respectively. Combined with virtual reality technology, ecological environment images of Nanning's air quality are rendered to achieve visual presentation of the data. The experimental results show that: 1) Incorporating image processing visualization design into air quality monitoring data management makes the displayed information more accessible and understandable; 2) The average absolute percentage error of the optimized DeepSeek in predicting ambient air quality is 0.1254%, representing a 15.4% and 39.1% improvement in prediction accuracy compared to AI tools such as Kimi and Doubao, respectively. The prediction accuracy of all optimized AI tools is higher than that before optimization. This research provides an intelligent and visual new approach for ecological environment management, effectively enhancing the efficiency of ecological environment image processing.

AP0035

Facial Shape Image-Based Stunting Classification using CNN with Sliced Feature Extraction

Hanum Khairana Fatmah, Aufaclav Zatu Kusuma Frisky and Mardhani Riasetiawan
Universitas Gadjah Mada, Indonesia

Abstract-We explore the use of Convolutional Neural Networks (CNNs) for stunting classification by comparing two methods: direct classification and sliced feature extraction. The aim is to assess whether slicing facial images and extracting localized features improves classification performance for stunting detection in children. The dataset includes facial images of children aged 24-60 months from East Flores, Indonesia, evaluated using accuracy, precision, recall, and F1-score. The Direct Classification model achieves 68% accuracy but struggles with stunting detection, favoring the Normal class. In contrast, the Sliced Feature Extraction + FCL model achieves 91% accuracy, with balanced precision and recall for both classes, demonstrating better performance by focusing on localized facial features. These findings underline the potential of sliced feature extraction for improving stunting detection. The method outperforms traditional models, offering a promising approach for community-based health screenings. Future work will focus on expanding the dataset, enhancing deep learning techniques, and integrating real-time stunting classification for broader applications.

Notes