2026 International Conference on Image and Video Processing (ICIVP 2026)

November 18-20

Kyoto Institute of Technology, Kyoto, Japan

Supported by

Technically Supported by

Indexed by

Compendex on Engineering Village

Scopus

https://icivp.org

Table of Contents

Welcome Message 3
Conference Committees 4
Guideline for Onsite Attendance 6
Guideline for Online Attendance 7
Conference Venue 9
Conference Rooms Directions 10
Simple Program 11
Detailed Program 15
Keynote Speech 1 16
Keynote Speech 2 18
Invited Speech 1 20
Invited Speech 2 22
Invited Speech 3 (Online) 24
Invited Speech 4 (Online) 26
Online Invited Speech 28
Session 1 30
Session 2 36
Online Session 41
Note 44

Welcome Message

On behalf of the organizing committee of ICIVP 2026, it is our great honor and pleasure to extend a warm welcome to all of you attending the 2026 2nd International Conference on Image and Video Processing (ICIVP 2026), taking place from 18th to 20th November 2026 in Kyoto, Japan.

ICIVP 2026 aims to provide researchers and practitioners from academia and industry with a forum to foster interdisciplinary collaboration, address emerging challenges in image and video processing field, and share cutting-edge research and innovations.

We sincerely appreciate the participation of our esteemed keynote speakers, invited speaker, session chairs, authors, and attendees from all over the world. Your expertise and diverse perspectives are the cornerstone of this event.

Over the course of three days, you will have the opportunity to engage in keynote speeches, technical sessions and networking events, all designed to facilitate knowledge exchange and collaboration. We are confident that the diverse perspectives and expertise represented here will lead to meaningful discussions and groundbreaking outcomes.

On behalf of all the conference committees, we would like to thank all the authors as well as the technical program committee members and reviewers. Their high competence, their enthusiasm, their time and expertise knowledge, enabled us to prepare the high-quality final program and helped to make the conference become a successful event.

We wish you a productive, inspiring, and enjoyable conference experience. Welcome to ICIVP 2026!

ICIVP 2026 Organizing Committee
November 2026

Conference Committees

Advisory Chair
Hiroshi Fujita, Gifu University, Japan

General Co-Chairs
Weiwei Du, Kyoto Institute of Technology, Japan (IEEE member)
Muhammad Tariq Mahmood, Korea University of Education and Technology, Korea (IEEE member)

Program Chairs
Huiyu Zhou, University of Leicester, UK
Yi-Peng Liu,Northwestern Polytechnical University, China (IEEE member)

Steering Chairs
Edmund Lai, Auckland University of Technology, New Zealand (IEEE member)
Marco Leo, Institute of Applied Sciences and Intelligent Systems, CNR, Italy (IEEE member)

Publicity Chair
Seokwon Yeom, Daegu University, Korea (IEEE member)

Local Chair
Yijun Duan,Kyoto Institute of Technology, Japan

Finance Chair
Ha Yoon Park, Cheju Halla University, Korea

Technical Committee Chairs
Zhigang Zhu, The City University of New York, USA (IEEE Senior member)
Qiu Chen, Kogakuin University, Japan (IEEE member)

Technical Program Committees
Yanlei Gu, Hiroshima University, Japan
Tetsuya Shimamura, Saitama University, Japan
Edmund Lai, Auckland University of Technology, New Zealand (IEEE member)
Thumrongrat Amornraksa, King Mongkut's University of Technology Thonburi, Thailand

Renjith V Ravi, M.E.A Engineering College, India
Seokwon Yeom, Daegu University, Korea (IEEE member)
Xue-Cheng Tai, Norwegian Research Center, Norway
Pedro Miguel Carvalho, INESC TEC, Portugal (IEEE Senior member)
Sathish Kumar Selvaperumal, Asia Pacific University of Innovation and Technology, Malaysia (IEEE member)
Olarik Surinta, Mahasarakham University, Thailand
Zhigang Zhu, The City University of New York, USA (IEEE Senior member)
Liming ZHANG, University of Macau, China
Yuelong Li, Tiangong University, China
Paul Craig, Xi'an Jiaotong-Liverpool University, China
Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Thailand
Marco Leo, Institute of Applied Sciences and Intelligent Systems, CNR, Italy(IEEE member)
Borko Furht, Florida Atlantic University, USA (IEEE member)
Rahul Shenoy, Samsung Semiconductor, Inc., San Jose, USA
Fanny Spagnolo, University of Calabria, Italy (IEEE Senior member)
Deepak Kumar Jain, Dalian University of Technology, China (IEEE Senior member)
Mas Rina Mustaffa, Universiti Putra Malaysia, Malaysia (IEEE Senior member)
Chutisant Kerdvibulvech, National Institute of Development Administration, Thailand (IEEE member)
Rawesak Tanawongsuwan, Mahidol University, Thailand
Ying Chen, Jiangnan University, China (IEEE Member)
Yasmeen M. George, Monash University, Australia
Yanhui Guo, University of Illinois Springfield, USA
Faridah Yahya, Universiti Kuala Lumpur, Malaysia (IEEE Senior member)
Baha Ihnaini, Wenzhou-Kean University, China
Suraiya Jabin, Jamia Millia Islamia, India
Shun Zhang, Northwestern Polytechnical University, China (IEEE member)
Yoshito Mekada, Chukyo University, Japan (IEEE member)
Yiguang Liu , Sichuan University, China (IEEE Senior member)
Yongqiang Zhao, Northwestern Polytechnical University, China (IEEE member)
Yuan-Kai Wang, Fu Jen Catholic University (IEEE Senior member)
Kharittha Jangsamsi, King Mongkut's University of Technology Thonburi, Thailand
Azim Zaliha Abd Aziz, Universiti Sultan Zainal Abidin (Unisza), Malaysia

Guideline for Onsite Attendance

Important Notes

- Please enter the meeting room at least 15 minutes before your session. Your punctual arrival and active involvement will be highly appreciated.
- Please wear your name tag for all the conference activities. Lending it to others is not allowed. If you have any companying person, please do inform our staff in advance.
- Please keep all your belongings (laptop and camera etc.) at any time. The conference organizer does not assume any responsibility for the loss of personal belongings.
- Please show name tag and meal coupons when dining.
- Due to force majeure including but not limited to earthquake, natural disaster, war and country policy, the organizer reserves the rights to change the conference dates or venue with immediate effect and takes no responsibility.

Oral & Poster Presentation

- Regular oral presentation: 15 minutes (including Q&A).
- Get your presentation PPT or PDF files prepared. Presentations MUST be uploaded at the session room at least 15 minutes before the session starts.
- Laptop (with MS-Office & Adobe Reader), projector & screen, laser pointer will be provided in all oral session rooms.
- Poster Presenters should bring your poster to the conference venue and put it on designated place.

Guideline for Online Attendance

Platform: ZOOM

- Step 1: Download ZOOM from the link: https://zoom.us/download

How to use ZOOM

* 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.

- Rename: Before you enter the conference room, please change your name to Paper ID + Name
- Chat and raise your hand: During the session, if you have any questions, please let us know by clicking “raise your hands” and use “chat” to communicate with conference secretary.
- When you deliver your online speech, please open your camera.
- During the Question section, if you have any questions about keynote speakers or authors, you can also click “raise your hands” or “chat”
- Share Screen: Please open your power point first, and then click “share screen” when it’s your turn to do the presentation.

How to join the conference online

- Find your paper ID and suitable meeting ID on the conference program.
- Open the ZOOM, click the join, paste the meeting ID, then you can join the conference.
- Click the stop share after you finish your presentation

Time Zone

- **Japan Standard Time (GMT+9)**

Device

- A computer with an internet connection (wired connection recommended)
- USB plug-in headset with a microphone (recommended for optimal audio quality)
- Webcam: built-in or USB plug-in

Online Room Information

Online Room Information

Zoom ID: **812 4670 2234**

Zoom Link: **https://us02web.zoom.us/j/81246702234**

* Please rename your Zoom Screen Name in below format before entering meeting room.

Role	Format	Example
Conference Committee	Position-Name	Conference Chair-Prof.
Keynote/ Invited Speaker	Position-Name	Keynote/Invited Speaker-Prof.
Author	Session Number-Paper ID-Name	S1-VP0001-Name
Delegate	Delegate-Name	Delegate-Name

Conference Venue

Venue - 60th Anniversary Memorial Hall, Kyoto Institute of Technology, Kyoto, Japan

Address : Hashigamicho, Matsugasaki, Sakyo-ku, Kyoto, 606-8585, Japan
(〒606-8585京都市左京区松ヶ崎橋上町)



The organizer doesn't provide accommodation, and we suggest you make an early reservation.

Conference Rooms Directions

Venue: 60th Anniversary Memorial Hall， Floor: 2nd



Simple Program

November 18, 2026 (GMT+9)

Onsite Registration

Registration Time: 10:30-16:00
Venue: 60th Anniversary Memorial Hall Floor: 2nd Small Seminar Room
Address: the second floor of the 60th Anniversary Memorial Hall at KIT, 19 Matsugasaki Shodencho, Sakyo Ward, Kyoto, 606-0951
1. Arrive at Kyoto Institute of Technology; **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;** **6. Finish registration.**

Online Test
Time Zone: GMT+9

Online Test Zoom ID: 812 4670 2234 Link: https://us02web.zoom.us/j/81246702234	Duration
Session Chair	10:30-10:40
Online Session (VP0015, VP0028, VP0034, VP0036)	10:40-17:30

Note: If you want to do online zoom test after 17:30, please contact your conference secretary.

November 19, 2026 (GMT+9)

Morning Sessions Venue: 60TH MEMORIAL HALL Floor: 2nd Large Seminar Room		Duration
Opening Remark	Prof. Weiwei Du, Kyoto Institute of Technology, Japan (IEEE member)	9:00-9:05
Keynote Speech 1	Prof. Hiroshi Fujita, Gifu University, Japan Speech Title: The Evolution of AI in Medical Imaging: From Deep Learning to Generative AI, AI Agents, and Beyond	9:05-9:45
Keynote Speech 2	Prof. Jie Yang, Shanghai Jiao Tong University, China Speech Title: Researches on the Defenses and Out-of-Distribution Detection in Trustworthy Deep Learning	9:45-10:25
Group Photo & Coffee Break		**10:25-10:45**
Invited Speech 1	Associate Professor Nian Wang, Advanced Imaging Research Center, UT Southwestern Medical Center, Dallas, Texas, USA Speech Title: High-resolution MRI Predicted Whole Mouse Brain Cell Type Atlas using Multi-modal Fusion Network (MFNet)	10:45-11:10
Invited Speech 2	Associate Professor Yanlei Gu, Hiroshima University, Japan Speech Title: Vision-Based Pedestrian and Cyclist Intention Recognition for Autonomous Driving	11:10-11:35
Invited Speech 3 (Online)	Associate Professor Kazuya Ueki, Meisei University, Japan Speech Title: Lessons from the Front Lines of Ad-hoc Video Search Zoon Link: https://us02web.zoom.us/j/81246702234 Zoom ID: 812 4670 2234	11:35-12:00
Lunch Time		**12:00-13:30**

Afternoon Session Venue: 60th Anniversary Memorial Hall Floor: 2nd Large Seminar Room		Duration
Invited Speech 4 (Online)	Dr. Chiharu Kai, Fujita Health University, Japan Speech Title: How can we develop AI systems that are truly needed in clinical settings? —Examining the case of breast cancer research Zoon Link: https://us02web.zoom.us/j/81246702234 Zoom ID: 812 4670 2234	13:30-13:55
Coffee Time		13:55-14:20
Parallel Session Venue: 60th Anniversary Memorial Hall Floor: 2nd Large Seminar Room		**Duration**
Session 1	Topic: Intelligent Image/Video Enhancement, Generation, and Domain-Specific Analytics Session Chair: Assoc. Prof. Nian Wang, Advanced Imaging Research Center, UT Southwestern Medical Center, Dallas, Texas, USA Presentations: VP0010, VP0022, VP0023, VP0024, VP0033, VP0020, VP0035, VP0038	14:20-16:20
Session 2	Topic: Multimodal Object Detection and Computer Vision Sensing Technologies Session Chair: Assoc. Prof. Yanlei Gu, Hiroshima University, Japan Presentations: VP0008, VP0009, VP0012, VP0013, VP0017, VP0021, VP0026-A	16:20-18:05
Dinner Time		**18:25-19:45**

Note:

(1) One Best Presentation will be selected from each presentation session, and the Certificate for Best Presentation will be awarded at the end of each session by Session Chairs.

(2) Regular each Presentation: about 15 Minutes including 2-3 Minutes for Question and Answer.

November 20, 2026 (GMT+9)

Morning Session		Duration
Online Speeches and online sessions **Zoom ID: 812 4670 2234** **Zoom link: https://us02web.zoom.us/j/81246702234**		
Invited Speaker 5 (Online)	Associate Professor Zichi Wang, Shanghai University, China Speech Title: Steganography in Neural Networks	14:00-14:25
Online Session 1	**Topic:** Machine Vision Algorithms and Industrial Intelligent Inspection Technologies **Session Chair:** Assoc. Prof. Zichi Wang **Presentations:** VP0015, VP0036, VP0028, VP0034	14:25-15:25

Detailed Program

November 19, 2026

9:00-9:05	**Opening Remark**
Venue	**60TH MEMORIAL HALL, Floor: 2nd Large Seminar Room**



Assoc. Prof. Weiwei Du

General Co-Chair

Kyoto Institute of Technology, Japan (IEEE Senior Member)

Weiwei Du received her bachelor's degree from Tianjin Chengjian University, China, her M.S. degree from the University of Aizu, Japan, in 2005, and her Ph.D. degree from Kyushu University, Japan, in 2008.

She joined Kyoto Institute of Technology, Japan, as an Assistant Professor in 2008 and has been an Associate Professor since 2019. In 2013, she was a Visiting Researcher at the University of Chicago, USA. She currently leads the Visual Information Laboratory at Kyoto Institute of Technology.

Her research interests include computer vision, image processing, deep learning, AI-based ophthalmic image analysis, deep learning-based drug design and discovery, generative AI for digital cultural heritage preservation, and intelligent robotic plastering systems for construction automation.

Dr. Du is a Senior Member of IEEE and actively participates in international research collaborations and academic activities. She serves as the General Chair of the International Conference on Image and Video Processing (ICIVP 2026).

Keynote Speech 1

Host	Assoc. Prof. Weiwei Du Kyoto Institute of Technology, Japan	**Time**	9:05-9:45 November 19, 2026
		Venue	2nd Floor, 60th Anniversary Memorial Hall



Prof. Hiroshi Fujita

Gifu University, Japan

Prof. Hiroshi Fujita received his Ph.D. from Nagoya University, Japan, in 1983. From 1983 to 1986, he was a visiting researcher at the Kurt Rossmann Laboratories for Radiologic Image Research at the University of Chicago, USA. He became an Associate Professor in 1991 and a Professor in 1995 in the Faculty of Engineering at Gifu University. In 2002, he was appointed Professor and Chair of Intelligent Image Information at the Graduate School of Medicine, Gifu University. He currently serves as a Research Professor and holds the title of Professor Emeritus at Gifu University. He is an Honorary President of the Society for Medical Image Information (MII), a Fellow of the Institute of Electronics, Information and Communication Engineers (IEICE), and an Honorary Member of the Japan Society for Medical Imaging Technology. He has served as a conference chair, invited speaker, and keynote speaker at numerous national and international scientific conferences. His research interests include computer-aided diagnosis (CAD), AI in medical imaging, medical image analysis, processing, and evaluation, as well as emerging technologies such as generative AI, vision-language models (VLMs), and AI agents for medical applications. He has received numerous awards and honors, including the Medical Imaging Information Society Award and multiple awards from the Radiological Society of North America (RSNA) and the International Society for Optics and Photonics (SPIE). He has authored or co-authored over 1,000 publications, including scientific journal articles, international conference papers, book chapters, and other scholarly publications. He has been repeatedly recognized among the world's top 2% most-cited scientists in both the "Career-long Impact" and "Single-year Impact" categories of the Stanford–Elsevier ranking, including

in the latest 2026 edition.

Speech Contents

Speech Title: The Evolution of AI in Medical Imaging: From Deep Learning to Generative AI, AI Agents, and Beyond

Abstract: Artificial intelligence (AI) in medical imaging has evolved remarkably, from early computer-aided diagnosis (CAD) systems using handcrafted features and conventional machine learning to deep learning approaches that transformed image classification, detection, and segmentation. More recently, foundation models and generative AI have opened new possibilities for medical image understanding and clinical information generation. In particular, vision-language models (VLMs) can integrate images and text to generate imaging findings and support clinical interpretation. The emergence of AI agents, which can plan tasks and use specialized tools, and agentic AI, which emphasizes autonomous workflow execution, feedback, and self-correction, marks another important advance. Looking ahead, generalist medical AI may integrate multimodal patient information for comprehensive clinical decision support. Using clear examples, real-world applications, and visuals, this keynote will review these advances and explore future directions for medical imaging AI, emphasizing reliability, clinical validation, and human oversight as AI evolves toward integrated and increasingly autonomous clinical workflow support.

Keynote Speech 2

Host	Assoc. Prof. Weiwei Du Kyoto Institute of Technology, Japan	**Time**	9:45-10:25 November 19, 2026
		Venue	2nd Floor, 60th Anniversary Memorial Hall



Prof. Jie Yang

Shanghai Jiao Tong University, China

Jie Yang received a bachelor's degree in Automatic Control in Shanghai Jiao Tong University (SJTU), where a master's degree in Pattern Recognition & Intelligent System was achieved three years later. In 1994, he received Ph.D. at Department of Computer Science, University of Hamburg, Germany. Now he is the Professor and Director of Institute of Image Processing and Pattern recognition in Shanghai Jiao Tong University. He is the principal investigator of more than 30 national and ministry scientific research projects in image processing, pattern recognition, data mining, and artificial intelligence. He has published six books, more than five hundreds of articles in national or international academic journals and conferences. Google citation over 24964, H-index 81. Up to now, he has supervised 6 postdoctoral, 48 doctors and 70 masters, awarded six research achievement prizes from ministry of Education, China and Shanghai municipality. Three Ph.D. dissertation he supervised was evaluated as "National Best Ph.D. Dissertation" in 2009, in 2017 and in 2019. He has been chairman and keynote speaker of more than 10 international conferences.

Speech Contents

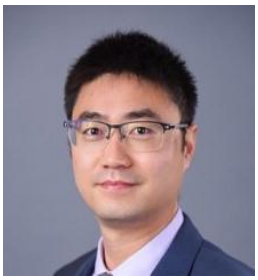
Speech Title: Researches on the Defenses and Out-of-Distribution Detection in Trustworthy Deep Learning

Abstract: The rapid advancement of deep learning has had a transformative effect on the development of technology and society across a multitude of sectors. In safety-critical contexts, the potential for neural network models to produce unreliable outputs in response to "malicious" or "unanticipated" inputs poses a severe risk. This talk delves into the output reliability from neural network models within the domain of trustworthy deep learning. 1) Inputs that involve pixel perturbations, exemplified by adversarial examples,w.r.t the task of adversarial and certified robustness; 2) Inputs that represent distribution shifts, exemplified by Out-of-Distribution (OoD) data, w.r.t the task of out-of-distribution detection. We introduce a novel strategy of model augmentation, adopt a multi-head neural network structure, and pose diversity constraints related to adversarial robustness into the model parameters.We adopt a multi-head neural network structure, use the ensemble of multiple heads in place of the ensemble of multiple neural networks,which significantly reduces the computational load in both training and certification phases. We propose that the non-linearity in InD and OoD data hinders PCA from learning a subspace that fully embodies their diversities. We propose a mode ensemble method that not only enhances detection performance but also significantly reduces the performance variance among independent modes.We propose performing linear dimension reduction on the gradient using a designated subspace that comprises principal components.

Papers about above topics have been published in TPAMI, IJCV, PR, NIPS in recent years.

Invited Speech 1

Host	Prof. Jie Yang Shanghai Jiao Tong University, China	**Time**	10:45-11:10, November 19, 2026
		Venue	2nd Floor, 60th Anniversary Memorial Hall



Associate Professor Nian Wang

Advanced Imaging Research Center, UT Southwestern Medical Center, Dallas, Texas, USA

Dr. Nian Wang is an Associate Professor in the Advanced Imaging Research Center at UT Southwestern Medical Center. My expertise is in developing novel quantitative MRI techniques and analysis methods on the CNS system and musculoskeletal system. His lab has extended the research on the technical development and methodology innovation in medical imaging and Artificial Intelligence (AI) and their translation to MRI, x-ray, and optical imaging. He has published more than 70 papers in reputed journals and been invited to devliver talks in many international conferences.

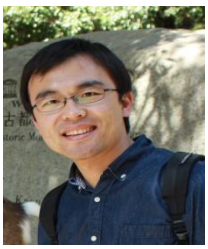
Speech Contents

Speech Title: High-resolution MRI Predicted Whole Mouse Brain Cell Type Atlas using Multi-modal Fusion Network (MFNet)

Abstract: Mapping brain cell types with high spatial precision is fundamental for understanding brain structure and function. Traditional cell atlases, such as the Allen Brain Cell Atlas, rely on labor-intensive techniques like single-cell sequencing and spatial transcriptomics, often at limited spatial resolution. In this study, we present a deep learning framework—Multi-modal Fusion Network (MFNet)—that predicts neuronal cell types across the whole mouse brain by integrating high-resolution diffusion MRI (dMRI) and multi-modal datasets from the BICCN mouse brain atlas. Using the Allen Brain Cell Atlas as ground truth, we trained MFNet to predict dominant neuronal cell neighborhoods and classes, taking advantage of regional specificity and hierarchical cell-type structures. High-resolution (45 μm isotropic) ex vivo dMRI was acquired and processed with multiple diffusion models (DTI, DKI, NODDI, SANDI), then registered to the Allen Mouse Brain Common Coordinate Framework (CCFv3), along with spatial transcriptomics and epigenomics data. MFNet produces a full-scale cell type atlas at 10 μm isotropic resolution, significantly enhancing the spatial detail of the original atlas. Additionally, we examined correlations between dMRI features and gene expression of marker genes. This study offers a novel, efficient method for generating high-resolution brain cell atlases and demonstrates the potential of advanced imaging combined with deep learning to decode brain cellular architecture and its relevance to development and disease.

Invited Speech 2

Host	Prof. Jie Yang Shanghai Jiao Tong University, China	**Time**	11:10-11:35, November 19, 2026
		Venue	2nd Floor, 60th Anniversary Memorial Hall



Associate Professor Yanlei Gu

Hiroshima University, Japan

Dr. Yanlei Gu received his Ph.D. from Nagoya University. He subsequently worked as a researcher at the University of Tokyo and was awarded a Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Overseas Researchers. In 2019, he joined Ritsumeikan University as a Lecturer, and since 2024, he has been an Associate Professor at Hiroshima University. He also serves as a Visiting Researcher at the University of Tokyo. His research interests include Autonomous Driving, Intelligent Transportation Systems (ITS), Machine Learning, and Computer Vision. He has published more than 120 peer-reviewed journal and international conference papers. His research has received several awards, including Best Paper Awards and other honors from IEEE and international conferences. He is a member of the IEEE Intelligent Transportation Systems Society and the IEEE Consumer Technology Society. He has served on the organizing committees of IEEE ICVES 2015 and IEEE ITSC 2017, as well as on the technical program committees of ICIPRob 2021, 2022, 2024, and 2026.

Speech Contents

Speech Title: Vision-Based Pedestrian and Cyclist Intention Recognition for Autonomous Driving

Abstract: Understanding the behavior and intentions of pedestrians and cyclists is essential for safe autonomous driving in complex traffic environments. In particular, accurately recognizing their crossing intentions enables autonomous vehicles to anticipate potential risks and make appropriate driving decisions. Recent advances in computer vision and deep learning have enabled the analysis of road users' appearance, body posture, movement patterns, and surrounding traffic conditions using visual information. However, intention recognition remains challenging due to variations in human behavior, occlusions, environmental conditions, and the uncertainty of human decision-making. This talk presents our research on vision-based intention recognition for pedestrians and cyclists, with a particular focus on predicting their crossing intentions. It discusses the use of visual features, temporal behavior analysis, and contextual information to understand and predict their actions. Key challenges, including early intention prediction, behavior interpretation in different traffic scenarios, and the integration of contextual information, are also examined. Finally, future research directions toward more reliable and context-aware intention recognition for autonomous driving are discussed.

Invited Speech 3 (Online)

Host	Assoc. Prof. Yanlei Gu Hiroshima University, Japan	**Time**	11:35-12:00, November 19, 2026
		Venue	2nd Floor, 60th Anniversary Memorial Hall



Assoc. Prof. Kazuya Ueki

Meisei University, Japan

He received his B.S. degree in Information Engineering in 1997 and his M.S. degree in Computer and Mathematical Sciences in 1999, both from Tohoku University, Sendai, Japan. In 1999, he joined NEC Soft, Ltd., Tokyo, Japan, where he was mainly engaged in research on face recognition. He received his Ph.D. degree from the Graduate School of Science and Engineering, Waseda University, Tokyo, Japan, in 2007. From 2013 to 2017, he served as an Assistant Professor at Waseda University. He is currently an Associate Professor in the School of Information Science, Meisei University. His research interests include information retrieval, video anomaly detection, pattern recognition, and machine learning. He is involved in the video retrieval evaluation benchmark (TRECVID) sponsored by the National Institute of Standards and Technology (NIST), contributing to the development of video retrieval technology. His submitted systems achieved the highest performance in the TRECVID AVS task in 2016, 2017, 2022, and 2025.

Speech Contents

Speech Title: Lessons from the Front Lines of Ad-hoc Video Search

Abstract: The Ad-hoc Video Search (AVS) task in the TRECVID benchmark has been a demanding testbed for video retrieval, where systems must find relevant shots from free-form text queries across large video collections. Over the years, the field has moved from predefined visual concepts to CLIP-style text−video embeddings, and now to vision-language models (VLMs) that interpret user intent and reason about visual content. This talk shares what we learned while competing through these shifts, focusing on three approaches: query expansion using large language models and image generation, caption-driven domain adaptation without manual annotation, and VLM-based semantic verification that re-ranks results through visual question answering. Drawing on experimental findings and hands-on experience, the talk highlights what made a difference in practice and discusses the challenges that remain for future video retrieval.

Invited Speech 4 (Online)

Host	Prof. Hiroshi Fujita Gifu University, Japan	**Time**	13:30-13:55, November 19, 2026
		Venue	2nd Floor, 60th Anniversary Memorial Hall



Dr. Chiharu Kai

Fujita Health University, Japan

She received her M.S. degree in Health Sciences from Kumamoto University in 2019. She joined Konica Minolta, Inc. in 2019, where she was engaged in research on medical image processing. She served as an Assistant Professor at Niigata University of Health and Welfare from 2022 to 2025, where she also received Ph.D. degree in 2025. She is currently an Assistant Professor in the School of Medical Sciences at Fujita Health University. Her research interests include medical image analysis, artificial intelligence (AI), and computer-aided diagnosis (CAD). Her research focuses on developing predictive AI systems for early breast cancer detection using mammograms.

Speech Contents

Speech Title: How can we develop AI systems that are truly needed in clinical settings? —Examining the case of breast cancer research

Abstract: With recent technological advancements, the field of medical AI is expanding rapidly. This talk addresses the core question of how to effectively integrate AI into clinical settings, drawing on fundamental knowledge of breast cancer and research examples. Breast cancer remains a major health challenge where early detection via mammography is crucial, yet generating radiology reports places a heavy burden on radiologists. To tackle this issue, we investigated Vision-Language Models for automatically generating mammogram reports. In addition to model development, this talk focuses on our evaluation framework comparing conventional objective metrics with subjective clinical assessments by a breast clinician. We will share the key insights and lessons learned from analyzing the gap between these evaluation approaches, and discuss the essential factors needed to build AI systems that truly serve clinical needs.

Online Invited Speech

Time	**14:00-14:25, November 20, 2026**
Venue	**https://us02web.zoom.us/j/81246702234** **Zoom ID 812 4670 2234**



Assoc. Prof. Zichi Wang

Shanghai University, China

Zichi Wang received the BS degree in electronics and information engineering from Shanghai University, China, in 2014, and received the MS degree in signal and information processing in 2017, the PhD degree in information and communication engineering from the same university in 2020. He is currently with the School of Communication and Information Engineering, Shanghai University, Shanghai, as an Associate professor. His research interests include steganography, steganalysis, and artificial intelligence security. He has published over 100 papers in these areas. He has served as an invited speaker for several international conferences, including ACMLC 2026 and AIintelli Vision 2027, served as a program chair of international conference AAIP 2027, served as chair of special session for several international conferences, including AAIP 2025, AAIP 2026, and IEEE PRMVAI 2026 and ACMLC 2026, and served as a TPC member for several international conferences, including ICECI 2022, AAIP 2025, CCIE 2026, IEEE PRMVAI 2026, ACMLC 2026 and ICITAC 2027.

Speech Contents

Speech Title: Steganography in Neural Networks

Abstract: Steganography aims to achieve covert communication by hiding secret data into a normal cover and transmitting it over a public channel, which is significant for national information security. Traditional steganography transmits secret data through multimedia such as image, audio and video. Neural network model is a new type of data which grows rapidly in recent years and is widely used. Steganography for neural network models is an emerging research field that needs to be developed. Compared with multimedia, neural network model has complex structure, wide variety and large number of parameters. For this reason, traditional steganographic methods cannot be used for neural network model directly. This speech will discuss the secure steganography for neural network models.

Session 1 (14:20-16:20)

Topic: Intelligent Image/Video Enhancement, Generation, and Domain-Specific Analytics

Session Chair: Assoc. Prof. Nian Wang, Advanced Imaging Research Center, UT Southwestern Medical Center, Dallas, Texas, USA

Time: 14:20-16:20, November 19, 2026
Venue: 2nd Floor,
60th Anniversary Memorial Hall

***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.**

VP0010

Adaptive rANS with Decremental Frequency Updating for Block-Parallel Color Image Compression

Poonlap Lamsrichan
Kasetsart University, Thailand

Abstract- Block Truncation Coding (BTC) remains attractive for resource-constrained image compression due to its low computational complexity and integer-only arithmetic. This paper proposes a novel decremental frequency updating strategy for range Asymmetric Numeral Systems (rANS) integrated into an enhanced BTC pipeline, targeting efficient block-parallel color image compression, with full-precision representative valuecalculation, sentinel-based differential mean encoding, contiguous pixel index assignment, and adaptive symbol frequency encoding. The decremental rANS scheme initializes symbol probabilities from per-block frequency counts and reduces them progressively as encoding proceeds, eliminating the rescaling side information required by prior adaptive rANS methods while preserving full block-parallel decoding capability. For reference, Adaptive Arithmetic Coding (AAC) variants are also evaluated under block and non-block operating modes. Experiments on the 24-image Kodak test set demonstrate that the proposed decremental rANS achieves the best rate-distortion performance among all blockparallel schemes, outperforming static rANS and approaching sequential AAC, making it well-suited for parallel hardware implementation in embedded and resource-constrained systems.

VP0022

Bidirectional Guided Cross-Modulation Adaptive Fusion Transformer for Single Image Super-Resolution

Jan-Ray Liao and Jian-Yi Yao
National Chung Hsing University Taichung, Taiwan

Abstract- Single image super-resolution (SISR) research has increasingly moved toward hybrid architectures that combine the localized textural proficiency of Convolutional Neural Networks (CNNs) with the long-range modeling capabilities of Transformers. However, existing hybrid models face significant challenges, including restricted attention scopes to manage computational costs, distortions and artifacts arising from sparse attention, and feature distribution mismatches between high-frequency local details and low-frequency global structures. To overcome these obstacles, we propose the Bidirectional Guided CrossModulation Adaptive Fusion Transformer (BGCMAFT). Our model introduces three architectural innovations: (1) the Hybrid Attention Transformer Block (HATB), which efficiently expands the receptive field by alternating between dense and sparse attention mechanisms; (2) the High-Frequency Enhancement Block (HFEB), which utilizes asymmetric convolutions and high pass filtering to specifically boost edge information and local textural details; and (3) the Bidirectional Guided Fusion Block (BGFB), which employs dynamic weights and biases to facilitate mutual alignment and guidance between disparate feature spaces, ensuring structural consistency while suppressing artifacts. Extensive quantitative and qualitative experiments demonstrate that our proposed BGCMAFT model consistently outperforms recent state-of-the-art SISR models across standard benchmarks, particularly excelling on texture-rich datasets such as Urban100 and Manga109.

VP0023

A Multidimensional Evaluation of Open-Weight Diffusion Models for Structured Facial Portrait Generation

Yeong Su Lee and Michaela Geierhos
University of the Bundeswehr Munich, Germany

Abstract- This paper presents a systematic, multidimensional evaluation of three representative, open-source diffusion models for structured text-to-image facial portrait generation: FLUX.2-klein-9B, Stable Diffusion 3.5 Large, and SDXL Base. Each model generated 512 × 512- and 1024 × 1024-pixelfacial portraits using its recommended inference configuration, reflecting realistic local deployment scenarios. We evaluated performance using ArcFace distance to quantify the dissimilarity between the original and

generated portraits in terms of identity, CLIPScore to measure semantic prompt fidelity, ImageReward to measure perceptual quality, and inference time and peak GPU memory consumption to measure computational efficiency. Additionally, paired Wilcoxon signed-rank tests and Friedman tests were used to analyze differences dependent on resolution and model. The experimental results indicate that ArcFace distances remain statistically stable across the evaluated model – resolution configurations. However, significant differences were observed in semantic prompt fidelity, perceptual quality, inference speed, and GPU memory consumption. These findings show that evaluating diffusion models requires balancing multiple complementary quality dimensions and computational resources, rather than relying on a single performance metric, providing practical guidance for selecting open-source diffusion models for local facial portrait generation.

VP0024

Heritage-Aware Multimodal Batik Image Generation via Sequential DreamBooth-ControlNet Pipeline with Adaptive Weighted Fusion

Irawati Nurmala Sari[1], Jovan Prabowo Kuncoro[1], Novanto Yudistira[1] and Weiwei Du[2]
1: Brawijaya University, Indonesia
2: Kyoto Institute of Technology, Japan

Abstract- Batik is a UNESCO-recognized Indonesian cultural heritage whose rich motifs: parang, kawung, megamendung, and others, carries centuries of philosophical meaning. As demand grows for batik-inspired digital branding, designers face a challenge that AI tools have not yet solved: how do you generate an image that is both culturally authentic in its motif and structurally faithful to a given image template? These two requirements pull in different directions. DreamBooth fine-tuning learns motif identity from few examples but has no way to enforce spatial layout. ControlNet conditioning preserves image structure precisely but produces textures that look batik-like rather than genuinely motif-specific. Neither approach alone is sufficient for real design workflows. We propose a Sequential Pipeline that treats these as two sequential concerns: Stage 1 uses DreamBooth to embed authentic motif semantics into the latent space, and Stage 2 applies ControlNet with Canny edge conditioning to impose the structural boundaries of the target image. An adaptive Weighted Fusion Layer then blends both latent outputs using a single tunable parameter α. This design gives designers direct control over the motif-structure trade-off without retraining. Evaluated on three canonical Javanese batik motifs using both quantitative metrics and a human opinion survey, our Sequential Pipeline consistently outperforms DreamBooth-only and ControlNet-only baselines across all evaluation dimensions.

VP0033

High-Bit-Depth Medical Image Enhancement: Local Entropy Curve Equalization for 16-Bit Pancreas Computed Tomography

Hafijur Rahman and Tetsuya Shimamura
Saitama University, Japan

Abstract- Contrast enhancement (CE) of pancreas computed tomography (PCT) images plays a vital role in refining feature clarity, thereby facilitating downstream diagnostic and analytical tasks for pancreatic diseases. However, dedicated CE methods for PCT imaging remain underdeveloped due to two major bottlenecks: the absence of publicly available, contrast-oriented PCT datasets and the reliance of existing CE methods on 8-bit natural images, which fail to accommodate high-bit-depth (e.g.,12-, 14-, or 16-bit) medical data. Furthermore, preserving entropy while enhancing localized anatomical features is paramount in medical imaging to avoid structural data loss. To address these limitations, this paper presents a novel 16-bit PCT benchmark dataset (PCT v1.0) dedicated to CE research, consisting of 19 high-quality reference images and 9 distinct types of contrast distortions per reference. In tandem, we propose a localized implementation strategy for Entropy Curve Equalization (ETCE) designed to highlight local features while leveraging ETCE's inherent entropy-preserving capabilities. Experimental evaluations on some 16-bit PCT images against three renowned medical image enhancement and cutting-edge methods demonstrate the localized implementation strategy's potential in the CE of 16-bit PCT images regarding visual assessments and three numerical assessment measures. The proposed dataset and localized implementation strategy of ETCE establish a critical benchmark and methodological foundation to drive future high-bit-depth medical image processing research. The PCT v1.0 dataset is publicly available at: https://doi.org/10.13140/RG.2.2.16452.80000.

VP0020

1000MDE: A 1000-fps Monocular Depth Estimation Video Dataset with Rendering-and-Stereo Hybrid Annotations

Wei Wang[1], Yujing Wu[1], Yuan Li[2], Zihang Wang[1], Ryuji Fuchikami[3] and **Youhua Shi**[1]
1: Waseda University, Japan
2: Changzhou Institute of Technology, China
3: Real-Time Computing Lab, Japan

Abstract- High-frame-rate depth estimation is essential for low-delay 3D perception in fast-motion scenarios, yet existing datasets cannot provide RGB–depth supervision at 1000 fps. This work introduces 1000MDE, an outdoor monocular video depth dataset

containing 45 rendered and 20 hardware-synchronized stereo sequences, totaling 65,000 aligned RGB – depth pairs with dense annotations over valid non-sky regions. The rendering branch generates high-frame-rate sequences through 3D scene reconstruction and trajectory interpolation, while the stereo branch captures real scenes using a synchronized high-speed platform and recovers metric depth. A semantic- and geometryguided recovery pipeline further improves temporal consistency, scale stability, and boundary quality. A high-frame-rate protocol is also introduced to jointly evaluate per-frame accuracy and realtime performance. Experiments demonstrate improved annotation consistency and completeness and reveal a substantial gap between per-frame and real-time performance under 1000-fps input. The dataset is publicly available on the project page.

VP0035

Pedestrian Behavior Recognition Using LiDAR Images

Melvin Erol Sulayman Kisam[1] and **Yanlei Gu**[2]
1: Ritsumeikan University, Japan
2: Hiroshima University, Japan

Abstract- Recognizing pedestrian behavior is important for autonomous driving because it provides activity cues for subsequent intention and trajectory prediction, supporting safer interactions with pedestrians. Existing camera-based approaches capture detailed pedestrian appearance but can become unreliable under poor illumination. LiDAR sensing is less dependent on visible-light illumination; however, range measurements alone may provide insufficient detail to distinguish similar activities. This study investigates whether combining complementary LiDAR channels can improve pedestrian behavior recognition. The proposed method integrates spatially aligned range, signal, reflectivity, and near-infrared images from a single LiDAR sensor. A modified VGG-19 network classifies annotated pedestrian crops into three behaviors: standing, walking, and running. To support this investigation, we construct a controlled dataset comprising 3,600 four-channel samples collected from ten volunteers. The proposed multi-channel method achieves a mean classification accuracy of 85.11% across five reported splits, exceeding signalonly and range-only inputs by 2.00 and 14.41 percentage points, respectively. These findings demonstrate the benefit of combining complementary LiDAR channels for pedestrian behavior recognition under the evaluated conditions.

VP0038

A Dual-Stage Visual Perception Framework for Collaborative Scrub Nurses: Fine-Grained Instrument Recognition and Defect Detection

Junhao Chen[1], Yu-te Wu[2] and Kazushi Ikeda[1]

1: Nara Institute of Science and Technology, Japan
2: National Yang Ming Chiao Tung University, Taiwan

Abstract- In modern healthcare, the increasing volume of orthopedic surgeries, such as Total Hip Arthroplasty (THA), coupled with persistent staffing shortages, has amplified the demand for autonomous robotic scrub nurses. While recent advancements have automated instrument handover, robust visual perception remains a critical bottleneck. Specifically, distinguishing highly similar surgical instruments (e.g., forceps with minor tip variations) and ensuring their physical integrity (e.g., detecting rust and scratches) are imperative to prevent surgical errors and cross-contamination. In this paper, we propose a comprehensive dual-stage visual perception framework. First, we conducted a rigorous comparative analysis using targeted tip-patch datasets, demonstrating that the latest YOLOv11s architecture significantly outperforms the YOLOv8s baseline, achieving an outstanding mAP@50 of 93.6% and mAP@50-95 of 85.3% for fine-grained instrument identification. Second, to overcome the ethical and physical impossibility of acquiring a large-scale dataset of defective instruments, we introduced a highly realistic Synthetic Defect Generation pipeline utilizing AI-based background removal. Integrating this with the PatchCore anomaly detection algorithm, we empirically revealed that preserving the full-scene spatial context is essential for robust defect localization (Image AUROC of 0.839), whereas bounding-box cropping induces severe aspect ratio distortion and catastrophic false positives (Image AUROC of 0.375). This framework establishes a reliable and scalable “Eyes and Brain” module for embodied AI robotic systems, significantly advancing automated surgical assistance and quality assurance.

Session 2 (16:20-18:05)

Topic: Multimodal Object Detection and Computer Vision Sensing Technologies

Session Chair: Assoc. Prof. Yanlei Gu, Hiroshima University, Japan
Time: 16:20-18:05, November 19, 2026
Venue: 2nd Floor,
60th Anniversary Memorial Hall

***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.**

VP0008

Free Region Detection Through Fusion of Pixel Expansion and Multi-Cues for Monocular MAV

Muhamad Wafi bin Abdul Aziz, **Muhammad Faiz bin Ramli**, Syariful Syafiq bin Shamsudin and Siti Nur Mariani binti Mohd Yunos
Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia

Abstract- Vision-based obstacle avoidance using monocular camera for Micro Aerial Vehicles (MAVs) is challenging due to the lack of depth information and sensitivity to motion variation between frames. Previous pixel expansion-based methods are affected by unstable image frame position, leading to free region selection near the obstacles, and rely heavily on distance transform, which reduces robustness in complex and textureless environments. In addition, unsuitable regions such as ceiling and floor may be selected. This paper proposes an enhanced framework that integrates the original pixel expansion method with multi-cue fusion, image alignment, and a macro–micro gridbased selection strategy. Image alignment using ECC reduces frame misalignment, while multiple cues image subtraction, distance transform, optical flow, and edge information are fused to improve reliability. The grid-based rules further ensure that only navigable regions are selected. Experimental results show that the proposed method achieves a FreeRatio of 1.0 and RiskRatio of 0.0 with improved stability across both textured and textureless environments. However, this comes with increased computational time from 0.177 s to 0.281 s (texture) and 0.204 s to 0.267 s (textureless). Overall, the method provides a more robust and reliable solution for safe monocular MAV navigation.

VP0009

VIFIRE: A Three-Layer Vision-Based Framework for Real-Time Indoor Detection of Fire, Smoke, and Dust Using HSV Analysis, CNN, and YOLO

Azim Zaliha Abd Aziz[1], Nurul Nadzirah Adnan[1], Fatma Susilawati Mohamad[1], Jakkarin Weekaew[2], Chaowanan Khundam[2] and Sasithorn Rattanarungrot[2]

1: Universiti Sultan Zainal Abidin, Malaysia

2: Walailak University, Thailand

Abstract- Conventional fire detection systems relying on photoelectric smoke and heat sensors suffer from delayed response and elevated false alarm rates in complex indoor environments, particularly under steam, dust, and variable lighting conditions. Vision-based detection leveraging existing closed-circuit television (CCTV) infrastructure offers a scalable complementary approach; however, single-modality deep learning models remain prone to false activations and have not been systematically validated against Malaysian fire safety regulatory standards. This study proposes and evaluates a three-layer hybrid detection framework integrating pixel-level HSV colour analysis, Convolutional Neural Network (CNN) classification, and YOLOv8 object detection for real-time identification of fire and smoke hazards, with an extended dust detection layer validated on controlled simulated footage. The system incorporates BOMBA OKK-calibrated temporal confirmation timers (heat: 3 s; smoke: 5 s; dust: 4 s) and is benchmarked against OSHA 1994, UBBL 1984, and Fire Services Act 1988 compliance thresholds. Primary training and evaluation were conducted on the D-Fire dataset (~21,000 annotated images), with the BoWFire dataset (226 images) serving as an independent external benchmark for fireclass validation. Experiments span five representative UniSZA building environments: hallways, classrooms, auditoriums, server rooms, and carparks. The proposed system achieves a weightedaverage precision of 92.1%, recall of 89.0%, F1-score of 90.5%, and mAP@0.5 of 91.2%, outperforming all evaluated baselines. External validation on BoWFire yields a fire-class F1-score of 91.9%, confirming cross-dataset generalisability. The overall false alarm rate (FAR) of 4.1% represents an 84.2% reduction relative to conventional photoelectric detectors under steam interference. Real-time operation is confirmed at a mean throughput of 21.6 FPS on mid-range GPU hardware, satisfying UBBL By-law 228 continuous monitoring requirements.

VP0012

Improving Data Efficiency and Stability in Domain Adaptation via Model Ensemble for Instance Segmentation in Viticulture

Tsuyoshi Kuroda[1] and Kota Matsui [2]
1: Yamaha Motor, Co., Ltd., Japan
2: Kyoto University, Japan

Abstract- Limited dataset availability remains a major challenge in applying deep learning to agriculture, where collecting and annotating target-domain data is often costly. Although transfer learning can improve data efficiency, its performance may deteriorate when pretrained source-domain models are poorly matched to the target domain, leading to negative transfer. To address this issue, we propose a source-data-free multi-source transfer-learning method based on model ensembling. The proposed method combines multiple models pretrained on different source domains and adapts them using limited target-domain data, without requiring access to the original source-domain datasets. We evaluated the method on grape-cluster instance segmentation using vineyard RGB image datasets collected under different berry-color and location conditions. The proposed method reduced the amount of target-domain training data required to achieve AP and F1 scores comparable to those obtained by training from scratch with the full dataset, with reductions of 20% or more. While a parameter-matched larger single model achieved higher data efficiency (with a reduction rate of 55% or more), the proposed ensemble exhibited substantially lower performance volatility, providing a better balance between data efficiency and training stability for real-world deployment.

VP0013

RelPos-Co-DETR: Extending Transformer-Based Detection with Discrete Relative Position Classification

Marcel Henkel and Oleksandr Shchetsura
Fraunhofer IOSB, Germany

Abstract- Distinguishing between visually similar military vehicle variants often depends on subtle differences in component placement rather than on the mere presence of a component. Such spatial information is crucial for recognition assistance workflows but cannot be captured by standard object detectors, which predict only bounding boxes and class labels. To address this gap, this work extends Co-DETR, a transformer-based detector, to jointly predict discrete relative positions and associated parent objects for vehicle components. In the proposed RelPos-Co-DETR, two lightweight prediction heads are added: a position head assigning each component a discrete 3D location within a 3 × 3 × 3 grid (e.g. leftfront-bottom), and a parent bounding box head identifying the mounting structure (e.g. wheeled vehicle). A novel annotation schema defines this position discretization and provides semantic interpretations aligned to the vehicle's coordinate frame. Experiments were conducted on a military vehicle dataset with training limited to

only 99 images, resulting in a challenging lowdata regime. Despite these constraints, RelPos-Co-DETR achieves 30.00% Average Precision (AP), representing a decrease of merely 1.8 percentage points compared with the baseline without the new capabilities (31.80% AP). Beyond preserving detection performance, the extension adds substantial spatial reasoning capabilities: it achieves 58.96% position accuracy - far beyond random (3.70%) and majority (15.28%) baselines - and localizes parent bounding boxes with 82.75% mean Intersection over Union (mIoU).

VP0017

Automated Optical Inspection System for ATY: Quantitative Grading of Hairiness and Spatial Density

Chih-Hua Hsieh, Shao-Kuo Tai and Hsien-Chou Liao
Chaoyang University of Technology, Taiwan

Abstract- Air Textured Yarn (ATY) is essential in high-value functional textiles due to its unique loop and hairiness structures that provide a cotton-like feel and superior bulkiness. However, the industry currently relies heavily on manual visual and tactile assessment to grade ATY "hairiness," leading to subjective bias, limited efficiency, and a lack of standardized quantitative metrics. This study proposes an automated optical inspection (AOI) system utilizing a 4K-resolution color line-scan camera and a composite lighting module. For yarn hairiness, a dual-branch morphological Exclusive-OR (XOR) algorithm is employed on the lightness channel to isolate micro-loop features. For woven web density, a local spatial density mapping (LSDM) approach is proposed to overcome the averaging effects inherent in traditional global statistics. Experimental results demonstrate that the system successfully transforms subjective hand-feel into standardized geometric values. The established thresholds provide preliminary quantitative references for local hairiness and density inspection. The proposed framework offers a proof-of-concept technical basis for subsequent calibration and digitalization of textile quality control.

VP0021

Performance Evaluation of YOLO26, YOLO11, RTMDet, and DETR for Sperm Detection using VISEM Dataset

Mohd Zamri Ibrahim
Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia

Abstract- Object detection is crucial for automated sperm analysis in clinical settings. This paper presents a comparative analysis of YOLO26, YOLO11, RTMDet, and DETR on the VISEM dataset. The study evaluates training performance, precision, recall, mean Average Precision (mAP), speed, model size, and parameter count. Results demonstrate

that YOLO11 achieves superior accuracy, with YOLO11s offering the best balance between accuracy and speed, making it the strongest candidate for clinical deployment. YOLO11 models consistently achieve the highest mAP scores, with approximately 10% improvement over DETR in mAP50 and a gap exceeding 32 percentage points in mAP50-95. RTMDet achieves competitive mAP50 scores closely matching YOLO11 but shows a notable drop at the stricter mAP50-95 threshold. In terms of speed, YOLO26n achieves the fastest inference at 2.1ms owing to its NMS-free architecture, while DETR exhibits a significantly slower speed of 12.62ms and a larger model size of 166.6 MB. The findings highlight trade-offs between performance, speed, and model complexity, providing insights for selecting the most suitable detection architecture for sperm analysis applications.

VP0026-A

Ultra-Small Object Detection in YOLOv8 via Super-Resolution Generative Adversarial Networks and Shallow Feature Fusion

Shyang-Jye Chang and Cheng-Wei Peng
National Yunlin University of Science & Technology

Abstract- Accurate detection of ultra-small objects remains a critical challenge in industrial inspection, medical imaging, and remote sensing due to insufficient feature representations in traditional models. This study proposes a novel approach integrating Super-Resolution Generative Adversarial Networks (SRGAN and ESRGAN) with an enhanced YOLOv8 architecture to tackle this limitation. First, we employ PSNR-optimized SRGAN generators to enhance image details and augment training data. To prevent overfitting and ensure data diversity, this data is merged with ESRGAN-upscaled images to form a robust multi-resolution dataset. Furthermore, we introduce YOLOv8-P1, which incorporates an additional shallow feature layer (P1) into the YOLOv8 network to better preserve fine textures and spatial edge information. Evaluated using the mAP@50 metric, our experimental results demonstrate that the YOLOv8-P1 model significantly outperforms the baseline architecture, delivering reliable recognition even in complex backgrounds with indistinct features. This research validates that combining super-resolution augmentation with optimized feature extraction substantially improves micro-object detection capabilities.

Online Session

November 20, 2026, (14:25-15:25, GMT+9)

Topic: Machine Vision Algorithms and Industrial Intelligent Inspection Technologies
Session Chair: Assoc. Prof. Zichi Wang
Time: 14:25-15:25, November 20, 2026
Zoom ID: 812 4670 2234
Link: https://us02web.zoom.us/j/81246702234

***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.**

VP0015

HSQ-VLM: A Novel Spatially-Constrained Quadrant Segmentation VLM Model for Explainability in Diabetic Retinopathy

Shivum Telang
University of Pittsburgh, United States

Abstract- Diabetic Retinopathy (DR) is an aggressive retinal disease and a leading cause of global blindness, yet its clinical management is currently hindered by the black-box nature of diagnostic AI. While deep learning models achieve high classification accuracy, there is a critical lack of explainability methods capable of detailing the exact anatomical landmarks and lesion distributions that lead to a clinical decision for DR. Therefore, we propose HSQ-VLM, a novel quadrant segmentation pipeline on fundus images that utilizes a Landmark Anchored Cartesian Cross-Attention mechanism to unify visual feature extraction with structured clinical reasoning. Unlike traditional methods that rely on arbitrary image partitioning, our pipeline implements 4-quadrant Topological Latent Partitioning (TLP) to dynamically align retinal features with a fovea-centered coordinate system. This allows the Vision-Language Model to generate natural language reports that quantify pathology with anatomical precision. On a dataset of 3,500 high-resolution fundus images, this innovative methodology achieved a lesion detection sensitivity of 99.6% for hemorrhages and 96.4% for microaneurysms, while demonstrating a significant reduction in boundary-ambiguity errors compared to standard segmentation baselines.

VP0036

Optimization of Multi-Class Object Detection for Autonomous Navigation of a

Warehouse Delivery Robot on an Edge Neural Accelerator

Nikita Tursukov, Ruslan Gataullin, Evgeniy Gogolev, Eduard Lazarev and Egor Tikhonov
Saint Petersburg Electrotechnical University «LETI», Russia

Abstract- This paper describes the design, deployment and experimental characterisation of a multi-class object detection subsystem for an autonomous mobile robot operating in warehouse premises. A YOLOv11n detector with 17 logistics-specific classes is compiled for a Hailo-8L neural accelerator attached to a Raspberry Pi 5 and integrated into a ROS 2 navigation stack, where detections are fused with 2D lidar ranges and converted into navigation decisions: an unbounded protective wait for persons and vehicles, and a route penalty with replanning for movable objects. A C++ pre-processing component composed with the camera driver, rate gating, zero-copy inference bindings and on-chip non-maximum suppression make the detector affordable on a four-core host: inference takes 22.05 ms per frame, the capture-to-detection latency is 48.3 ms (median) at a sustained 10 Hz, the detector node uses 9-10 % of one core and the complete stack stays below 60 % of the available CPU. A replay experiment with brightness, blur and noise degradations exposes two safety-relevant failure modes: confident false detections on under-exposed frames and cross-class duplicates from per-class suppression. The paper closes with the modifications required before the subsystem can be relied upon for person safety.

VP0028

Diversity-Aware Representative Coreset Selection for Cross-Domain Generalization in Chest X-Ray Pneumonia Classification

Waqas Ahmed[1] and **Srikanta Pal**[2]
1: University of Gujrat, Pakistan
2: Maynooth University, Ireland

Abstract- Models trained on one dataset often perform poorly when applied to data from a different source, posing a major challenge in medical imaging. This study evaluates DARCS-FL (Diversity-Aware Representative Coreset Selection via Facility Location), a data selection framework that selects diverse and representative training samples to improve classification across different datasets. The method is evaluated for chest X-ray pneumonia classification using seven data-selection strategies and four different training-set sizes, three classifiers, and a fine-tuned DenseNet121 model. In the classical machine learning branch, DARCS-FL reduces the cross-domain performance gap from 43.77% to 16.53%, and in the fine-tuned DenseNet-121 branch from 26.66% to 7.45%, performing substantially better than centroid-based selection. McNemar's test shows a statistically significant difference between the two methods ($\chi^2 = 1807.24$, $p < 10^{-12}$), with DARCS-FL correctly classifying 2,037 cases that the centroid-based model misclassifies, versus

80 cases in the reverse direction. Grad-CAM analysis shows that DARCS-FL focuses on a wider range of anatomical regions and has higher attention entropy than the centroid-based approach (0.856 vs. 0.269; Wilcoxon $p < 10^{-6}$). In addition, feature-space coverage is strongly associated with cross-domain accuracy (Spearman $\rho = 0.857$, $p = 0.014$). Overall, the results show that diversity-aware coreset selection can improve cross-domain robustness without simply increasing the training-set size, providing an effective data-selection strategy for chest X-ray classification under domain shift.

VP0034

KVSwinVesselNet: A Lightweight Framework for Kidney Vessel Segmentation

Ashini Wijesinghe, Shatirah Akib, Muntasir Al-Asfoor and Shahadate Rezvy
Buckinghamshire New University, United Kingdom

Abstract- Accurate segmentation of renal vasculature in Hierarchical Phase-Contrast Tomography (HiP-CT) is challenging because of complex branching patterns, fine vessel structures, and the computational demands of high-resolution volumetric imaging. While three-dimensional models achieve strong segmentation performance, their computational cost can limit scalability, whereas conventional two-dimensional approaches may struggle to preserve vascular continuity. This study proposes KVSwinVesselNet-BoundaryAware, a lightweight 2D topologyaware framework integrating Tiny Vessel Enhancement Blocks, Window Swin Transformer modules, boundary-aware refinement, deep supervision, and topology-aware optimisation. The framework was evaluated on Kidney 1 from the SenNet + Human Organ Atlas dataset using 1,138 training and 1,140 testing slices, with 2D U-Net and Swin-Unet as baseline architectures. The proposed model achieved a Dice score of 0.8603, IoU of 0.7632, recall of 0.9127, clDice of 0.8406, and NSD@1 of 0.8566, using only 4.74 million parameters. Compared with Swin-Unet, Dice improved from 0.8256 to 0.8603 and clDice from 0.7600 to 0.8406, while recall increased from 0.8160 to 0.9127. These results demonstrate that vessel-specific feature enhancement combined with topology-aware learning improves vascular connectivity and fine-vessel preservation while maintaining computational efficiency

Note