CONFERENCE PROGRAM

2026 11th International Conference on Image, Vision and Computing

2026 10th International Conference on Deep Learning Technologies

Kunming, China | July 17-19, 2026

Co-sponsored by

KUNMING UNIVERSITY OF SCIENCE AND TECHNOLOGY

IEEE

Organized by

信息工程与自动化学院
FACULTY OF INFORMATION ENGINEERING AND AUTOMATION

Table of Contents

- Welcome Message 3
- Conference Venue 4
- Online Conference Information 6
- Event at a Glance 8
- Daily Schedule 9

  Friday, July 17, 2026 9

  Saturday, July 18, 2026 10

  Sunday, July 19, 2026 12
- Details of Keynote Speakers 13

  Yaochu Jin 14

  Lap-Pui Chau 15

  Boxin Shi 16
- Details of Invited Speakers 17
- On-site Oral Sessions 24
- On-site Poster Sessions 46
- Online Oral Sessions 50

Welcome Message

On behalf of the Organizing Committee of the 11th International Conference on Image, Vision and Computing (ICIVC 2026) and the 10th International Conference on Deep Learning Technologies (ICDLT 2026), we warmly welcome all distinguished scholars and industry experts. The two conferences will be held concurrently in Kunming, China, from July 17 to 19, 2026, offering a premier global platform for academic exchanges in image, vision, computing and deep learning technologies.

Driven by artificial intelligence and high-performance computing, image, vision and deep learning technologies have become key drivers of the digital economy. They are widely applied in autonomous driving, medical imaging, smart cities and augmented reality, continuously promoting technological innovation, disciplinary development and talent cultivation.

Co-sponsored by Kunming University of Science and Technology and IEEE, hosted by the Faculty of Information Engineering and Automation, Kunming University of Science and Technology, and technically supported by SMC, the joint conferences are also supported by multiple renowned universities across China. Their joint efforts ensure the high standard and international quality of this event.

We sincerely invite global researchers to participate and submit papers. The conference covers deep learning applications in software and hardware, frontier theoretical research, and enterprise-oriented deep learning innovations, as well as advanced image and vision computing studies. The comprehensive program includes keynote speeches, invited talks, and oral, poster, video and online sessions to facilitate idea sharing and in-depth academic discussion.

We extend our sincere gratitude to all institutions, authors, reviewers and volunteers for their dedicated support. We wish all participants an inspiring and fruitful conference experience in Kunming!

ICIVC&ICDLT 2026 Organizing Committee

July 2026

Conference Venue

丽水云泉大酒店
Fontaine Blanche Hotel

云南丽水云泉大酒店

地址：昆明市呈贡区昆明大学城聚贤街 768 号

No. 768, Juxian Street, Kunming University Town, Chenggong District, Kunming City

Emergency Call
Police: 110
Ambulance: 120
Fire: 119

Average Temperature During the Conference

17℃ -26℃

Transportation

From Kunming Changshui International Airport

- By Metro line
  Metro Line 6(KUNMING Airport-JUHUA) → Line 4(JUHUA-WUJIAYING Exit A) → Bus 216 (approximately 1 hour 40 minutes).

- Taxi: approximately 45 minutes, 45-60RMB

From Kunming Railway Station

- By Metro line
  Metro Line 1(KUNMING Railway Station-Municipal Administration Center Station Exit B) →Bus 216 (approximately 1 hour 40 minutes).

- Taxi: approximately 50 minutes

Oral Presentation Tips

- ✧ The duration of a presentation slot is 15 minutes. Please prepare your presentation for 12 minutes & 3 minutes for questions from the audience;
- ✧ An LCD projector & computer will be available in every session room for regular presentations;
- ✧ Presentations MUST be uploaded at the computer at least 15 minutes before the session start.

Poster Presentation Tips

- ✧ We expect that at least one author stands by the poster for (most of the time of) the duration of the poster session. This is essential to present your work to anyone interest into it.
- ✧ Posters should be set-up at least 15 minutes before the session starts and removed at the end of the session. Left behind posters at the end of the session will be disposed of.
- ✧ The size of poster is 1.8 m*0.8m. Posters must be in portrait format (height > width). The paper ID needs to be marked in the top right corner of the poster

Dress Code

- ✧ All participants are kindly requested to dress formally, as casual wear is discouraged.
- ✧ National formal dress is welcome.

Attention Please

- ✧ Please ensure the safety of your belongings in public areas. For personal and property security, delegates are advised to wear their identification badges during the conference and refrain from lending them to unauthorized individuals. The conference cannot be held responsible for the loss of personal items.

扫码观看会议现场图片

Scan the QR code to view conference pictures

扫码发送会议简称添加会议助理微信

Scan the code to add conference secretary on Wechat

Online Conference Information

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

Passcode: A1719	
ROOM A	https://us02web.zoom.us/launch/jc/89024344343
ROOM B	https://us02web.zoom.us/launch/jc/84940237642
ROOM C	https://us02web.zoom.us/launch/jc/83059659555

Time Zone

- ✓ **Beijing Standard Time, UTC/GMT +8**
- ✓ Please make sure that both the clock and the time zone on your computer are set to the correct standard time.

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.

Additional Suggestions

- ✓ A computer with an internet connection (wired connection recommended)
- ✓ USB plug-in headset with a microphone (recommended for optimal audio quality)
- ✓ Webcam (optional): built-in or USB plug-in
- ✓ Stable internet connection
- ✓ Quiet environment
- ✓ Proper lighting

Presentation Tips

- ✓ Each presentation slot is 15 minutes. Please prepare to speak for around 12 minutes, allowing 3 minutes for audience questions.
- ✓ Join the meeting room at least 10 minutes before the session begins.

Zoom Test

- ✓ Prior to the formal conference, presenter shall join the test room to make sure everything is on the right track.
- ✓ For presenters, please rename Zoom Screen Name in “Paper ID - Name” format before entering meeting room.

Conference Recording

- ✓ The entire conference will be recorded. If any participant objects to being recorded during their presentation, they are requested to inform the organizers in advance.
- ✓ The recording will be paused accordingly during their segment. Attendees are expected to dress formally and maintain appropriate decorum throughout the event.
- ✓ The recording is intended solely for conference-related purposes and academic publications. It is strictly prohibited to distribute, use commercially, or utilize the recording for any illegal activities.

Event at a Glance

Date	Time	Event
July 17	10:00-17:00	Sign in
	10:00-17:00	Zoom Testing (For Online Participants)
	18:00-20:00	Dinner
July 18	09:00-09:10	Opening Ceremony
	09:10-09:20	Group Photo
	09:20-10:00	Keynote Speaker 1
	10:00-10:40	Keynote Speaker 2
	10:40-11:00	Coffee Break
	11:00-11:40	Keynote Speaker 3
	11:40-14:00	Lunch
	14:00-16:20	On-site Oral Session 1 & 2 & 3
	14:00-16:00	On-site Poster Session 1
	16:00-16:30	Coffee Break
	16:30-18:30	On-site Oral Session 4 & 5
	16:30-18:30	On-site Poster Session 2
	18:30-20:30	Dinner
July 19	10:00-12:35	Online Oral Session 1 & 2 & 3
	12:35-13:30	Break Time
	13:30-16:05	Online Oral Session 4 & 5 & 6
	16:05-16:30	Break Time
	16:30-19:05	Online Oral Session 7 & 8 & 9

Daily Schedule

Friday, July 17, 2026 | (UTC+8)

Sign-in and Materials Collection	
10:00-17:00	**Sign in and Collect Conference Materials** Locations: Lobby of Yunnan Fontaine Blanche Hotel **签到&领取会议资料** 地点：云南丽水云泉大酒店大堂
Steps: 1. Give your Paper ID to the staff. 2. Sign your name in the attendance list and check meal information. 3. Check your Conference Kit.	
18:00-20:00	**Dinner**

Zoom Testing for Online Participants	
ROOM A: https://us02web.zoom.us/launch/jc/89024344343 **ROOM B: https://us02web.zoom.us/launch/jc/84940237642** **Passcode: A1719**	
10:00-12:00 **14:00-17:00**	**ROOM A** **Test for Online Committees & Online Session Chairs**
10:00-12:00 **14:00-17:00**	**ROOM B** **Test for Online Authors**

Saturday, July 18, 2026 | (UTC+8)

Venue: International Conference Center, Hall 1 (2nd Floor) **地址：国际会议中心一号厅 （二楼）**		
Opening Ceremony		
09:00-09:05	**Opening Remarks** **Ce Zhu,** University of Electronic Science & Technology of China, China	
09:05-09:10	**Welcome Message** **Jinhong Cai,** Kunming University of Science and Technology, China	
09:10-09:20	**Group Photo**	
Keynote Speech		
09:20-10:00	Keynote Speaker 1	**Prof. Yaochu Jin,** Westlake University, China Speech Title: From Brain-like Computing to Brain-like Embodied Intelligence
10:00-10:40	Keynote Speaker 2	**Prof. Chau Lap-Pui,** The Hong Kong Polytechnic University, Hong Kong, China Speech Title: Egocentric Computer Vision
10:40-11:00	**Coffee Break**	
11:00-11:40	Keynote Speaker 3	**Prof. Boxin Shi,** Peking University, China Speech Title: Event-based Photometric Imaging
11:40-14:00	**Lunch** **Yunquan Pavilion, 1st Floor / 云泉阁 （一楼）**	
Author On-Site Presentation Sessions		
14:00-16:20	**On-site Oral Session 1** **Deep Learning and Cross-Domain Applications** Invited Speech-Hongjiao Guan, EC1019, EC3134, EC212, EC318, EC432, EC434, EC3144, EC4190	201, 2nd Floor 二楼 201

14:00-16:05	**On-site Oral Session 2** **3D Vision and Remote Sensing Imaging** Invited Speech-Youneng Bao, EC1008, EC2073, EC2074, EC2076, EC3138, EC3121, EC3140	101, 1st Floor 一楼 101
14:00-16:20	**On-site Oral Session 3** **Image Processing and Video Analysis** Invited Speech-Rui Chen, EC1009, EC1015, EC2042, EC2044, EC2075, EC3142, EC3145, EC4202	102, 1st Floor 一楼 102
14:00-16:00	**On-site Poster Session 1** **Object Detection and Remote Sensing Imaging** EC3104, EC2033, EC2064, EC2089, EC3098, EC3114, EC3147, EC1010, EC3136, EC4168, EC4179, EC4198	202, 2nd Floor 二楼 202
16:00-16:30	**Coffee Break**	
16:30-18:05	**On-site Oral Session 4** **Medical Imaging and Healthcare Applications** Invited Speech, EC2057, EC2081, EC2082, EC2083, EC3146	201, 2nd Floor 二楼 201
16:30-18:30	**On-site Oral Session 5** **Object Detection and Industrial Applications** EC1024, EC2049, EC2056, EC3122, EC3123, EC2043, EC315, EC3106	101, 1st Floor 一楼 101
16:30 -18:30	**On-site Poster Session 2** **Biomedical Imaging and Image Processing** EC313, EC1013, EC2072, EC2080, EC3131, EC3110, EC3152, EC4161, EC4177, EC4187, EC3094, EC4200	202, 2nd Floor 二楼 202
18:30-20:30	**Banquet | 2nd Floor, Yuntu Hall** **晚宴 | 二楼，云图厅**	

Sunday, July 19, 2026 | (UTC+8)

Online Author Presentation Sessions		
Zoom ROOM A: https://us02web.zoom.us/launch/jc/89024344343 Zoom ROOM B: https://us02web.zoom.us/launch/jc/84940237642 Zoom ROOM C: https://us02web.zoom.us/launch/jc/83059659555 Passcode: A1719		
10:00-12:05	**Online Oral Session 1** **Fundamentals of Deep Learning and Trustworthy Artificial Intelligence** Invited Speech-Thong Ming Sen, EC103, EC207, EC431, EC1014, EC440, EC3129, EC4175	ROOM A
10:00-12:35	**Online Oral Session 2** **Object Detection and Tracking** Invited Speech-Xin Nie, EC2071, EC3097, EC3125, EC3105, EC3119, EC3120, EC3153, EC4170, EC317	ROOM B
10:00-12:05	**Online Oral Session 3** **Image Processing and Enhancement** Invited Speech-Guoping Xu, EC2047, EC2084, EC2086, EC4164, EC4184, EC4191, EC2079	ROOM C
12:35-13:30	**Lunch Break**	
13:30-16:05	**Online Oral Session 4** **Medical Imaging and Biometric Recognition** EC1005, EC2054, EC3124, EC314, EC1018, EC2070, EC3111, EC3130, EC2045, Invited Speech-Vesna Zeljkovic	ROOM A
13:30-15:50	**Online Oral Session 5** **Generative AI and AIGC Detection** Invited Speech-Neelendra Badal, EC1012, EC2065, EC2068, EC2085, EC2088, EC319, EC3151, EC4178	ROOM B

13:30-15:35	**Online Oral Session 6** **3D Vision and Reconstruction** Invited Speech-Tianzi Jiang, EC2051, EC2078, EC3148, EC4185, EC4189, EC4195, EC316	ROOM C
16:05-16:30	**Break Time**	
16:30-19:05	**Online Oral Session 7** **Multimodal Learning and Video Understanding** Invited Speech-Ahmad Ali, EC2034, EC3143, EC438, EC1025, EC3128, EC4172, EC2027, EC429, EC439	ROOM A
16:30-19:00	**Online Oral Session 8** **Agricultural Remote Sensing and Environmental Safety** EC1006, EC1007, EC3001, EC3150, EC4169, EC4171, EC4176, EC4182, EC4199, EC2058	ROOM B
16:30-18:35	**Online Oral Session 9** **Industrial Inspection and Intelligent Systems** Invited Speech-Liangxiu Han, EC2048, EC3102, EC442, EC3139, EC4160, EC4173, EC4186	ROOM C

Keynote Speaker

Yaochu Jin

IEEE Fellow
Westlake University, China

Speech Time: 09:20-10:00, July 18
Location: International Conference Center, Hall 1 (2nd Floor)

Yaochu Jin is Chair Professor of AI, Director of the Trustworthy and General AI Lab, Head of Artificial Intelligence Department, School of Engineering, Westlake University, Hangzhou, China. Prior to that, he was Alexander von Humboldt Professor for Artificial Intelligence endowed by the German Federal Ministry of Education and Research, with the Faculty of Technology, Bielefeld University, Germany from 2021 to 2023, and Surrey Distinguished Chair, Professor in Computational Intelligence, Department of Computer Science, University of Surrey, Guildford, U.K. from 2010 to 2021. He was also "Finland Distinguished Professor" with University of Jyväskylä, Finland, and "Changjiang Distinguished Visiting Professor" with the Northeastern University, China from 2015 to 2017. His main research interests include intelligent optimization of complex systems, trustworthy AI, brain-like computing, and brain-like embodied AI.

Prof Jin was the President of the IEEE Computational Intelligence Society and the Editor-in-Chief of the IEEE Transactions on Cognitive and Developmental Systems. He is the recipient of the 2025 IEEE Frank Rosenblatt Award. He is a Member of Academia Europaea and Fellow of IEEE.

Speech Title: From Brain-like Computing to Brain-like Embodied Intelligence

Abstract: This talk begins with an introduction to information process in the human brain and the evolution of neural self-organization in nature. This is followed by an overview of research efforts that endevour to understand principles behind biological neural self-orgainzation in computational systems, including the coupling between the evolution and development of the nervous systems and body plan. Then, recent advances in constructing large spiking neural networks are presented, paying special attention to brain-inspiked models simulating different cognitive pathways in the brain. This talk is concluded by dicussing the need to move towards embodied brain-like computing systems to enable them to autonomously learn, develop and evolve.

Keynote Speaker

Lap-Pui Chau

IEEE Fellow
The Hong Kong Polytechnic University, Hong Kong, China

Speech Time: 10:00-10:40, July 18
Location: International Conference Center, Hall 1 (2nd Floor)

Lap-Pui Chau received a Ph.D. degree from The Hong Kong Polytechnic University in 1997. He is currently a Professor and Global STEM Scholar (under the Global STEM Professorship Scheme) and Director of JC STEM Lab of Machine Learning and Computer Vision in the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University. He was with the School of Electrical and Electronic Engineering, Nanyang Technological University from 1997 to 2022. His current research interests include large language model, perception for autonomous driving, egocentric computer vision, and embodied artificial intelligence. He is an IEEE Fellow.

Speech Title: Egocentric Computer Vision

Abstract: Egocentric computer vision focuses on capturing and interpreting the world from a first-person perspective. This presentation will introduce the potential of this technology for better scene understanding. By recognizing user activities (such as cooking), identifying objects of interest, and analyzing human-object interactions, egocentric vision can significantly enhance activity recognition capabilities. We will also explore the latest advancements and emerging technologies in this field.

Keynote Speaker

Boxin Shi

Peking University, China

Speech Time: 11:00-11:40, July 18
Location: International Conference Center, Hall 1 (2nd Floor)

Boxin Shi is currently a Boya Young Fellow Associate Professor (with tenure) and Research Professor at Peking University, where he leads the Camera Intelligence Lab (http://camera.pku.edu.cn). He is the Associate Director of the Institute of Visual Technology, School of CS, PKU. He is also a BAAI Scholar and Director of the PKU-AI2Robotics Joint Lab of Embodied AI. He received the PhD degree from the University of Tokyo and did postdoc research at MIT Media Lab. His research interests are computational photography and computer vision. His papers were awarded as Best Paper, Runners-Up at CVPR 2024/ICCP 2015, and selected as candidate for Best Paper at ICCV 2015. He received the Okawa Foundation Research Grant in 2021 and led a National Science and Technology Major Project as PI. He has served as associate editors of TAPMI/IJCV.

Speech Title: Event-based Photometric Imaging

Abstract: Compared with conventional frame-based cameras, the neuromorphic vision sensor, such as the event camera and the spike camera have unique advantages, especially in their ability to perceive high-speed moving objects and scenes with high dynamic range. Existing research has actively demonstrated the advantages of event cameras in computer vision tasks such as image deblurring, high dynamic range imaging, and high-speed object detection and recognition. However, the photometric image formation model of event cameras has not been carefully analyzed. This talk will share a series of research progress on modeling and analyzing the photometric image formation model of an event camera that records and responses to high-speed radiance changes, for achieving real-time photometric stereo, normal and reflectance estimation, hyperspectral imaging, light transport analysis, and so on.

Onsite-Invited Speaker

Wali Samad

Quanzhou University of Information Engineering, China

Speech Time: 16:30-16:50, July 18
Location: 201, 2nd Floor

Dr. Wali Samad is an Associate Professor at Quanzhou University of Information Engineering, China. He received his Bachelor of Science in Mathematics and Physics from Forman Christian College, affiliated with the University of the Punjab, Lahore, Pakistan, in 2007, and his Master of Science in Applied Mathematics from The Islamia University of Bahawalpur, Pakistan, in 2009. He subsequently studied Chinese Language at Nankai University, Tianjin, China, before pursuing his Ph.D. in Computational Mathematics at the School of Mathematical Sciences, Nankai University, which he completed in 2018. Dr. Samad conducted postdoctoral research at the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), from 2018 to 2020 and again from 2023 to 2024. Between these appointments, he served as an Assistant Professor in the Department of Mathematics at Namal University, Pakistan. His research interests include medical image analysis, image segmentation, image restoration, level set methods, optimization algorithms, inverse problems, and artificial intelligence. He has published numerous high-impact research articles in leading international journals and conferences and has made significant contributions to the development of advanced mathematical and computational methods for medical image processing.

Rui Chen

Tianjin University, China

Speech Time: 14:00-14:20, July 18
Location: 102, 1st Floor

Rui Chen (Member, IEEE) received the Ph.D. degree in instrument science from Tsinghua University, China, in 2010. He is currently an Associate Professor with the School of Microelectronics, Tianjin University. He has authored or coauthored one book and more than 80 technical articles in refereed journals and proceedings. His current research interests include generative artificial intelligence technology and large model technology. He has accumulated extensive experience in the design, efficient training, and deployment of various deep neural network models and large models. In recent years, his research achievements have

been widely applied to the deployment of large model and agent systems in multiple industry companies.

Youneng Bao

Shenzhen University, China

Speech Time: 14:00-14:20, July 18
Location: 101, 1st Floor

Youneng Bao is an Assistant Professor at the College of Electronics and Information Engineering, Shenzhen University. He received his Ph.D. in Information and Communication Engineering from Harbin Institute of Technology and was a postdoctoral researcher at City University of Hong Kong. His research focuses on intelligent media compression and efficient computing, including learned image and video compression, lightweight neural models, robust coding, and deployment-friendly optimization for ultra-high-definition video, live streaming, VR/AR, and immersive media. He has published papers in leading journals and conferences, including ICCV, AAAI, IEEE T-CSVT, Signal Processing.

Hongjiao Guan

Qilu University of Technology, China

Speech Time: 14:00-14:20, July 18
Location: 201, 2nd Floor

Hongjiao Guan is currently an Associate Professor at Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences). She received her Ph.D. degree in Computer Science and Technology from Harbin Institute of Technology. Her research focuses on machine learning and natural language processing, with current primary interests covering medical information processing, medical large language models, and affective computing. She serves as a committee member of YSSNLP and Affective Computing under the Chinese Information Processing Society (CIPS) of China. She has led and participated six projects, including the Youth Program of the National Natural Science Foundation of China and Youth Project of the Shandong Provincial Natural Science Foundation. She has published 15 SCI-indexed and CCF papers as the first or corresponding author, and held over 10 authorized invention patents as the first inventor.

Online Invited Speaker

Liangxiu Han

Manchester Metropolitan University, UK

Speech Time: 16:30-16:50, July 19
Meeting Room: Room C
Passcode: A1719

Prof. Han is currently a full Professor of Computer Science at the Department of Computing and Mathematics, Faculty of Science and Engineering, Manchester Metropolitan University. Prof. Han is Academic Director for Centre for Digital Data Research, Faculty Lead for AI, Digital and Cyber Physical Systems and Deputy Director of ManMet Crime and Well-Being Big Data Centre. Prof. Han's research areas mainly lie in the development of novel big data analytics/Machine Learning/AI, and development of novel intelligent architectures that facilitates big data analytics (e.g., parallel and distributed computing, Cloud/Service-oriented computing/data intensive computing) as well as applications in different domains (e.g. Precision Agriculture, Health, Smart Cities, Cyber Security, Energy, etc.) As a Principal Investigator (PI) or Co-PI, Prof. Han has a proven track record of successfully leading multi-million-pound projects on both national and international scales (supported by diverse funding sources: UKRI, NIHR, GCRF/Newton, EU, Industry, and Charity) and has extensive research and practical experiences in developing intelligent data driven AI solutions for various application domains (e.g. Health, Food, Smart Cities, Energy, Cyber Security) using various large datasets (e.g. images, numerical values, sensors, geo-spatial data, web pages/texts). Prof. Han has served as an associate editor/a guest editor for a number of reputable international journals and a chair (or Co-Chair) for organisation of a number of international conferences/workshops in the field. She has been invited to give a number of keynotes and talks on different occasions (including international conferences, national and international institutions/organisations). Prof. Han is a member of EPSRC Peer Review College, an independent expert of European Commission for proposal evaluation/mid-term project review, and serves on assessment and panels for UKRI (EPSRC, BBSRC, Innovate UK, MRC etc.) and the British Council.

Vesna Zeljkovic

Lincoln University, USA

Speech Time: 15:45-16:05, July 19
Meeting Room: Room A
Passcode: A1719

Vesna Zeljkovic received BS, MS and Ph.D. degrees in electrical engineering. From 1996 to 2004 she was a research and teaching assistant at the University of Belgrade and University of Novi Sad, receiving Ministry of Science and Technology of Republic of Yugoslavia fellowship for the prominent young researchers. Following up on her interest in global electrical and computer engineering education she gained experience and expertise in global academic programs with the emphasis on electrical and computer engineering education. Dr. Zeljkovic taught engineering courses at undergraduate and graduate level in five countries in United States, Saudi Arabia, China, Malaysia and ex-Yugoslavia. Currently, Dr. Vesna Zeljkovic is a full professor at Lincoln University. Her expertise encompasses signal and image processing developing mathematical models and novel algorithms for the analysis of 2D images that have applications in biomedical engineering, video surveillance, analysis of complex optical spectra, public health, industry application, homeland security and national defense. She has more than 100 scientific papers, a book chapter and four books published in these fields. Dr. Zeljkovic serves as a reviewer for peer reviewed journals and conference proceedings. Dr. Zeljkovic mentored graduate and undergraduate students. She participates and organizes different activities at professional meetings of IEEE and other organizations. She is a member of General IEEE GMEPE/PAHCE management team and served as Associate Editor of the IEEE HPCS for years. Dr. Zeljkovic formed IEEE student branch at New York Institute of Technology, Nanjing campus and served as its Branch Counselor. Dr. Zeljkovic is an IEEE Senior Member.

Neelendra Badal

Kamla Nehru Institute of Technology, India

Speech Time: 13:30-13:50, July 19
Meeting Room: Room B
Passcode: A1719

Dr. Neelendra Badal is a Professor in the Department of Computer Science & Engineering at Kamla Nehru Institute of Technology, (KNIT), at Sultanpur (U.P.), INDIA of Dr. A.P.J. Abdul Kalam Technical University (AKTU), UP Lucknow INDIA(Formerly Uttar Pradesh Technical University, (UPTU), Lucknow (On-Leave). And presently Director of Rajkiya Engineering College, Bijnor. He received B.E. (1997) from Bundelkhand Institute of Technology (BIET), Jhansi (U.P.), INDIA, in Computer Science & Engineering, M.E. (2001) in Communication, Control and Networking from Madhav Institute of Technology and Science (MITS), Gwalior (M.P.), INDIA and PhD (2009) in Computer Science & Engineering from Motilal Nehru National Institute of Technology (MNNIT), Allahabad (U.P.), INDIA. He is Chartered Engineering (CE) from Institution of Engineers (IE), India. He is a Senior Member of IEEE, Member of ACM and Life Member of IE, IETE, ISTE, CSI, India, IoTSocietyofIndia. He has published more than 100 papers in International/National Journals, conferences and seminars. His research interests are Distributed System, Parallel Processing, GIS, Data Warehouse & Data mining, Software engineering, Networking, IoT and Data Analytics.

Tianzi Jiang

Institute of Automation, Chinese Academy of Sciences, China

Speech Time: 13:30-13:50, July 19
Meeting Room: Room C
Passcode: A1719

Tianzi Jiang is Professor and Director of Beijing Key Laboratory of Brainnetome and Brain-Computer Interface at the Institute of Automation, Chinese Academy of Sciences, and Director of International Institute of Brain-Machine Intelligence, Harbin Institute of Technology (Shenzhen). He obtained PhD degree at Zhejiang University and BSc degree at Lanzhou University. His research interests include neuroimaging, Brainnetome, digital twin brain, neuromodulation methods and device, and their clinical applications in brain disorders. He was Chair of Organization of Human Brain Mapping. He was elected a member of the Academy of Europe, a fellow of IEEE, IAPR and AIMBE. He has published over 350 peer-review journal papers. He is the recipient of Glass Brain Award of the Organization of Human Brain Mapping, Hermann von Helmholtz Award of International Neural Network Society, and Natural Science Award of China.

Ahmad Ali

Hainan Bielefeld University of Applied Sciences, China

Speech Time: 16:30-16:50, July 19
Meeting Room: Room A
Passcode: A1719

Dr. Ahmad Ali is currently working as a faculty member at Hainan Bielefeld University of Applied Sciences, China. Prior to this, he completed his postdoctoral research at Shenzhen University, China. His main research interests include intelligent transportation systems, big data analytics, machine learning, cloud computing, and edge computing. His research focuses on spatio-temporal data modeling, traffic flow prediction, multimodal data fusion, and intelligent computing methods for smart city applications. Dr. Ali has published high-quality research articles in well-reputed journals, including Information Sciences, Neural Networks, Future Generation Computer Systems, IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Consumer Electronics, and Computer Science Review. He has also contributed academic services to several CCF and international conferences, including ICPADS 2025, ICIC 2026, and others.

Xin Nie

Wuhan Institute of Technology, China

Speech Time: 10:00-10:20, July 19
Meeting Room: Room B
Passcode: A1719

Dr. Xin Nie is a Professor at the School of Computer Science and Engineering, Wuhan Institute of Technology. His long-standing research agenda spans software engineering, intelligent optimization and evolutionary computation, machine learning and deep learning, and cloud computing. In recent years, he has led and participated in numerous national and provincial research programs, producing a series of original contributions at the intersection of evolvable hardware, evolutionary optimization, embedded machine vision, and edge intelligence for autonomous systems. He has authored more than sixty peer-reviewed articles in mainstream SCIEI-indexed international journals and in proceedings of flagship conferences such as those sponsored by IEEE and ACM, and holds multiple national invention patents and software copyrights. Prof. Nie serves on the editorial boards of several SCIESCI-indexed journals and is a member of the technical program committees of various international conferences, actively contributing to their academic organization. He is a member of the IEEE, the China Computer Federation (CCF), the Chinese Association for Artificial Intelligence (CAAI), and a senior member and committee member of the Chinese Institute of Command and Control (CICC). His research group continues to advance the state of the art in lightweight Transformer architectures, edge-centric autonomous vision, multimodal fusion, and few-shot learning, with demonstrated deployments in intelligent sports analytics, unmanned systems, and industrial defect detection.

Thong Ming Sen

National University of Singapore, Singapore

Speech Time: 10:00-10:20, July 19
Meeting Room: Room A
Passcode: A1719

Dr. Thong Ming Sen is an Advocate and Solicitor of the High Court of Malaya and his area of expertise is in capital markets law, legal tech, and FinTech regulation. As an ex-Legal500 rising star–turned-Web3 aficionado, Dr. Thong was part of a team of researchers who were the first to report on regulatory issues in the blockchain space in Malaysia. Dr. Thong who holds an LLB (Hons) from the University of London, a Master of Laws (LLM) (Distinction), and a Doctor of

Philosophy (PhD) in law from Universiti Malaya is the author of the monograph Blockchain for Financial Governance in Malaysia and Singapore: Transforming Regulatory and Shariah Compliance to Drive Financial Inclusion (Springer, 2025).

Guoping Xu

Wuhan Institute of Technology, China

Speech Time: 10:00-10:20, July 19
Meeting Room: Room C
Passcode: A1719

Dr. Guoping Xu serves as a Lecturer and Master's Supervisor at Wuhan Institute of Technology. He earned his Ph.D. in Information and Communication Engineering from the School of Electronic Information and Communications, Huazhong University of Science and Technology. During his doctoral studies, he received joint training at the University of Pennsylvania (2016–2018) and Harvard University (2019). Following graduation, he undertook two postdoctoral appointments, respectively at Harvard University (2023–2024) and UT Southwestern Medical Center (2024–2026). His core research interests span biomedical imaging, computer vision, and deep learning, with dedicated work on image segmentation, image registration, object tracking, and quantitative disease evaluation. To date, he has authored over 60 peer-reviewed academic papers and holds 8 patents, and his research outputs have garnered more than 1,500 citations.

On-site Oral Session 1

➢ Deep Learning and Cross-Domain Applications

- **Session Chair**: Lin Zhang, Xi'an University of Architecture and Technology, China
- **Time:** 14:00-16:20, July 18
- **Meeting Room:** 201, 2nd Floor
- **Papers:** Invited Speech, EC1019, EC3134, EC212, EC318, EC432, EC434, EC3144, EC4190

Invited Speech
14:00-14:20

Hongjiao Guan
Qilu University of Technology (Shandong Academy of Sciences), China

Research and Practice on Automatic ICD Coding

Abstract: ICD coding serves as the core foundation for DRG - based medical insurance payment and medical record quality control. Currently, manual coding suffers from issues such as low efficiency, insufficient accuracy, and a shortage of professional talents. Existing automated coding solutions also have shortcomings, including a scarcity of Chinese - annotated datasets, difficulty in conforming to real - world clinical processes, and privacy risks due to reliance on closed - source large models. This report first introduces the research work on automatic ICD coding for electronic medical records. It then presents the research, development, and application practice of jointly building a Chinese ICD coding dataset with top - tier hospitals and developing intelligent coding LLM, facilitating the implementation of intelligent medical coding.

EC1019 14:20-14:35	Evaluating BLIP-2 for Fine-Grained Fashion Attribute Annotation: Prompt-Based Label Mapping and Error Taxonomy Author(s): Yang Song and Bo Du Presenter: Yang Song, Dalian Polytechnic University, China

	Abstract: Vision-language models prompted in a VQA style often return free-form text, which conflicts with closed-set requirements in fine-grained fashion attribute annotation. This paper evaluates BLIP-2 under a dimension-wise prompting protocol for three garment groups (outerwear, dresses, tops). Model outputs are recorded as BLIP_RAW and deterministically mapped into group-specific label sets (BLIP_Mapped), with an explicit not clear outcome for unmappable responses. Experiments use a controlled 50-image set with human ground truth from annotation guidelines. Beyond Accuracy and Macro-F1, coverage-aware metrics are reported, including the not clear rate and response-level mappability. BLIP-2 exhibits pronounced coverage limitations, with mappable rates of 41.96% for outerwear, 66.43% for dresses, and 52.04% for tops. Mapping increases schema-aligned coverage by 16.96–40.73 percentage points compared with direct matching on BLIP_RAW. An error taxonomy further distinguishes coverage failures from in-schema misclassification and summarizes recurring RAW drift patterns.
EC3134 14:35-14:50	A Multi-Task Collaborative Modeling Approach for Plant Trait Retrieval by Integrating Multi-Scale Spectral Features and Vegetation Indices Author(s): Xin Zhang, Guanglai Wang, Ye Liang and Jianping Huang Presenter: Xin Zhang, Northeast Forestry University, China Abstract: The precise quantification of plant leaf functional traits is essential for vegetation health monitoring and ecosystem function assessment. To address the limitations of conventional vegetation indices and shallow machine learning methods in resolving spectral overlap between chlorophyll and carotenoids, as well as the negative transfer issue induced by feature conflicts in multi-task learning, this study proposes a Leaf-scale Gated Task-specific Network (LGTN) based on multi-task learning for collaborative leaf trait inversion. The model adopts leaf hyperspectral data as input and utilizes a Multi-scale Dilated Convolution Encoder (MDCE) for in-depth spectral feature extraction, with six vegetation indices incorporated as physical prior knowledge. An asymmetric Multi-gate Mixture-of-Experts (MMoE) architecture is adopted to assign three shared experts for leaf mass per area (LMA), equivalent water thickness (EWT), and total chlorophyll (Chla+b), while four experts are allocated for carotenoids (Ccar). This design realizes the collaborative learning of common features and the effective isolation of task-specific characteristics. Meanwhile, an adaptive multi-task loss weighting strategy is applied to dynamically adjust task gradients. The experimental results demonstrate that the proposed LGTN outperforms single-task models (PLSR, RF, SVR, 1DCNN) and the multi-task 1DCNN. Specifically, the coefficient of determination (R^2) reaches 0.89, 0.86, 0.83, and 0.76 for LMA, EWT, Chla+b, and Ccar, respectively, and the normalized root mean square error (nRMSE) of all traits is less than 20%. Furthermore, ablation experiments verify that the introduction of vegetation indices substantially improves the inversion accuracy of pigment parameters (Chla+b and Ccar), with R^2 increased by 0.05 and 0.11, respectively. This study provides a novel deep learning framework for the joint high-precision inversion of plant leaf functional traits.
EC212 14:50-15:05	Bias2Risk: Dual-Task Evaluation of Bias-Induction Risk in Chinese Large Language Models Author(s): Qianyi Wang, Zhe Li, Mowei Gong, Yuping Qi, Yimin He Presenter: Qianyi Wang, Beijing University of Technology, China Abstract: Bias in Chinese large language models cannot be fully assessed with a single evaluation task. Prior work often examines general safety, refusal behavior, or social bias in isolation, but provides limited support for jointly analyzing safety recognition and open-generation risk in Chinese settings. To address this gap, we propose Bias2Risk, a dual-task risk evaluation framework for Chinese bias-induction scenarios. The framework uses a multiple-choice question (MCQ) task to evaluate a model's ability to recognize safer expressions, and an open-ended generation (OPEN) task to evaluate its tendency to produce risky outputs under bias-inducing contexts. In addition, we

	introduce three levels of induction strength—neutral, mild-induce, and strong-induce—to examine how model behavior changes as contextual pressure increases. Based on 120 seed scenarios, we construct a Chinese evaluation set containing 1,800 MCQ samples and 1,080 OPEN samples, and conduct unified experiments on six open-source instruction-tuned models. The results reveal three consistent patterns: higher MCQ accuracy does not reliably predict lower OPEN risk, stronger induction leads to higher unsafe rates, and risk varies substantially across social dimensions. These findings show that Bias2Risk can serve as a low-cost supplementary tool for pre-deployment risk screening, alignment regression testing, and prompt defense evaluation in Chinese bias-related settings.
EC318 15:05-15:20	Energy-Based Anomaly Detection for Academic Credential Verification on StarkNet Layer 2 Blockchain Author(s): Nhut Minh Hua, Nghia Quoc Phan, Hiep Xuan Huynh Presenter: Nhut Minh Hua, Tra Vinh University, Vietnam Abstract: Fraudulent academic credentials pose a growing threat to educational integrity worldwide. Traditional verification mechanisms, relying on manual processes and centralized databases, are both inefficient and vulnerable to manipulation. This paper proposes a novel framework integrating Energy-Based Models (EBM) with StarkNet Layer 2 blockchain for intelligent, automated, and cost-efficient credential verification. A composite energy function incorporating three complementary components — embedding-level compatibility, structural constraint enforcement, and temporal consistency validation — models the energy landscape of valid credentials, assigning low energy to legitimate records and high energy to anomalous ones. STARK zero-knowledge proofs provide privacy-preserving verification, reducing transaction costs by ∼99.999% compared to Ethereum Layer 1. A four-phase verification pipeline with a threetier decision mechanism (VALID / SUSPICIOUS / INVALID) balances automation with human oversight. Evaluated on 91,736 real StarkNet mainnet transactions (3.1% anomaly rate), the composite EBM achieves the highest unsupervised accuracy (94.84%, AUROC 0.713), while supervised baselines reach F1 up to 0.981, motivating future semi-supervised extensions.
EC432 15:20-15:35	Sequential Enhanced Physics-Informed Neural Networks for Lithium-Ion Battery State-of-Health Prediction Author(s): Rongrong Wang, Qingbin Xi, Chenglong Wang, Yanan Ma, Xuanlin Ma Presenter: Rongrong Wang, Yunnan University, China Abstract: The lithium-ion battery state-of-health (SOH) prediction is important for ensuring battery safety. However, reliable SOH prediction remains challenging due to imprecise physical models and the limitations of purely data-driven analysis. To address this challenge, a sequential-enhanced physics-informed neural network is proposed for SOH prediction. First, a physics-informed neural network (PINN) framework is developed to integrate battery data features under variable operating conditions with physical degradation mechanisms, compensating for the limitations of purely data-driven methods. Then, a gated recurrent unit (GRU) is incorporated into the PINN architecture to enhance its capability of capturing temporal dependencies in battery aging. Extensive simulations based on real-world XJTU, HUST, and MIT datasets are performed to validate the performance of the proposed method. The results demonstrate that the proposed PINN-GRU model achieves a 22% reduction in root mean square error compared to the baseline PINN-MLP on the MIT dataset, with the RMSE decreasing from 0.0087 to 0.0068 and the coefficient of determination improving from 0.808 to 0.903.
EC434 15:35-15:50	Benchmarking Deep Learning Algorithms for Enhancing Stock Price Prediction with Sentiment Fusion Author(s): Ruochi Ma, Haoyu Wang, Dejun Xie Presenter: Dejun Xie, City University of Macau, China

	Abstract: Contemporary social media is an important medium for expressing the public's attitudes toward social hotspots, and in financial markets the emotional tendency of public opinion acts as an investment weathervane for both individual and institutional investors, implying that public-opinion sentiment may significantly affect stock prices. However, most stock-price prediction models in use today do not include public-opinion sentiment as an influencing factor, which can limit modelling accuracy. To verify the importance of public-opinion sentiment for stock-price prediction, this paper proposes a new Cemotion-based stock-price prediction model that incorporates public-opinion sentiment tendency. We first collect all comments for a target stock over a specific period, then use the Cemotion 2.0 model to score the sentiment tendency of each comment and compute a daily sentiment score, and finally input sentiment as a special feature into mainstream machine-learning and neural-network models. After introducing the public-sentiment factor, the LSTM model shows the most significant improvement: a 13.68% reduction in FMAPE, 11.51% reduction in FRMSE, 2.20% improvement in R-squared, 0.18% improvement in MAPE, 9.48% improvement in MSE, 2.83% improvement in MAE, and 4.86% improvement in RMSE. The model combines the strengths of Cemotion sentiment analysis with the ability of LSTM to process serial data to improve the accuracy of closing-price prediction. Combining public sentiment with time-series modelling offers a more comprehensive approach to forecasting and yields valuable insights for investors and analysts.
EC3144 15:50-16:05	Scale Importance and External Knowledge Enhanced Dense Video Captioning Author(s): Zhenyu Liu, Zhiliang Chen, Jingfu Xiao, Jingyao Bi and Wenhui Jiang Presenter: Wenhui Jiang, Jiangxi University of Finance and Economics, China Abstract: Dense Video Captioning (DVC) is a challenging multimodal task that requires temporally localizing multiple events in a video and describing them in natural language. While recent DETR-based DVC frameworks unify localization and captioning end-to-end, we identify two critical limitations that remain underexplored. First, prior DETR-based DVC frameworks share multi-scale visual features for localization and captioning, often leading to imprecise event boundaries. The key reason is that joint optimization for both tasks prevents these features from being explicitly adjusted to amplify those higher temporal resolution scales that are critical for localization, as they carry scene-transition information. Second, in prior DETR-based DVC frameworks, the captioner relies solely on visual features, which lack the syntactic structures and lexical cues essential for generating linguistically descriptive captions. To address these issues, we propose a novel network named Scale Importance and External Knowledge Enhanced Dense Video Captioning (SIEK-DVC). SIEK-DVC introduces a Scale Importance Aware Reweighting Module (SIARM) that explicitly amplifies the higher temporal resolution scales critical for localization, and an External Knowledge Aware Caption Generator (EKACG) that supplies syntactic structures and lexical cues by feeding retrieved textual features at each decoding step. Experiments on ActivityNet Captions show that SIEK-DVC outperforms strong baselines.
EC4190 16:05-16:20	A TCN-Transformer Deep Learning Model for Time-Series Displacement Prediction in Landslide Monitoring Author(s): Xia Zhou, Xiao Wang and Jinyang Shang Presenter: Xia Zhou, Guizhou University, China Abstract: Transmission lines in mountainous regions frequently traverse landslide-prone areas; therefore, the accurate prediction of slope displacement is critical for enhancing the disaster prevention capabilities of power grids. Addressing the "step-like" deformation of the Baijiabao landslide in the Three Gorges Reservoir area—driven by fluctuating reservoir levels and rainfall—traditional Recurrent Neural Networks struggle to dynamically capture the varying importance of historical information. This

	paper proposes a displacement prediction model integrating Temporal Convolutional Networks (TCN) and Transformers. First, Empirical Mode Decomposition (EMD) is employed to decompose the cumulative displacement into trend and periodic components, with the trend component predicted via Brown's exponential smoothing. Second, a candidate feature pool is constructed—comprising rainfall, reservoir levels, historical displacement, and their derived features—with inputs automatically selected based on correlation analysis. Furthermore, a serial TCN-Transformer hybrid network is designed, where the TCN extracts local mutation features and the Transformer captures long-term dependencies. Finally, an ensemble strategy using multiple random seeds is adopted to improve model robustness. Validated against monitoring data from the Baijiabao landslide (2018–2024), the proposed model achieves superior performance in periodic displacement prediction compared to LSTM, GRU, standalone TCN, and standalone Transformer models, yielding a Root Mean Square Error (RMSE) of 2.835 mm and a Coefficient of Determination ( ) of 0.982. This method can be directly integrated into monitoring and early-warning systems for transmission tower foundations, providing technical support for enhancing power grid resilience.

On-site Oral Session 2

3D Vision and Remote Sensing Imaging

- **Session Chair**: Wuzhen Shi, Shenzhen University, China
- **Time**: 14:00-16:05, July 18
- **Meeting Room**: 101, 1st Floor
- **Papers**: Invited Speech, EC1008, EC2073, EC2074, EC2076, EC3138, EC3121, EC3140

Invited Speech
14:00-14:20

Youneng Bao
Shenzhen University, China

Quality Assessment of 3D Gaussian Splatting: Distortions, Benchmarks, and Open Challenges

Abstract: 3D Gaussian Splatting (3DGS) has become a practical scene representation for real-time novel-view rendering, compression, and immersive content delivery. However, its quality assessment still largely follows rendered-view proxy protocols that sample camera poses, render images or videos, and apply inherited image and video quality metrics. While convenient, this practice does not fully capture 3DGS-native degradations that originate from Gaussian primitive distributions, splatting and visibility behavior, and trajectory-dependent artifacts. This survey reviews recent 3DGS quality assessment studies from four perspectives, covering distortion characteristics, subjective benchmarks, objective metric reliability, and emerging directions for native 3DGS evaluation. Across the literature, we find that different benchmarks construct different notions of quality, and that metric effectiveness is highly protocol dependent, varying with distortion sources, stimulus formats, and view and trajectory sampling. Finally, we summarize open challenges and outline practical guidelines for building more comparable and representative 3DGS-QA evaluations.

EC1008 14:20-14:35	A Monocular Depth Perception Method for the Liquid Zoom System Based on the Principle of Parallax Author(s): Zhaoyang Liu, Wanting Shen, Fang Zhang, Zihao Gan, Yanting Luo and Huajie Hong Presenter: Zhaoyang Liu, National University of Defense Technology, China Abstract: Liquid zoom technology has attracted increasing attention due to its flexible zoom capability. However, the application of parallax principle combined with liquid zooming for depth perception is still in its early stages. To address this challenge, this paper proposes a monocular depth perception method based on the parallax principle for a self-developed liquid zoom system. We systematically present the process of achieving depth perception and derive the theoretical framework for depth calculation. Experimental results on depth recovery in real-world scenarios demonstrate that the proposed method exhibits strong depth recovery capability. Moreover, the reconstruction accuracy improves as the focal length of the liquid zoom system increases, with a reconstruction error of 0.3549 mm under short focal length and 0.1649 mm under long focal length. This study enriches the theory and methodology of depth perception based on the parallax principle for liquid zoom systems, and it plays a significant role in advancing the application of liquid zoom technology in depth measurement.
EC2073 14:35-14:50	A Surface Streamline Generation Method for 3D Unstructured Grids Author(s): Cheng Chen, Dan Zhao, Chao Yang and Mengfei Dong Presenter: Cheng Chen, China Aerodynamics Research and Development Center, China Abstract: Surface streamline generation is a critical technique for visualizing and analyzing flow structures in engineering design. However, when dealing with unstructured mesh with complex geometric surfaces and diverse mesh topologies, existing methods suffer from issues such as inaccurate positioning and poor adaptability, making it difficult to stably generate high quality surface streamlines. In this paper, we propose a high precision surface streamline generation method for unstructured meshes. By employing surface constrained tracking mechanism, this method effectively ensures the conformity and continuity of streamlines along complex geometric surface. Experiments on multiple representative unstructured mesh types demonstrate that our approach can generate continuous, smooth, and highly surface aligned streamlines.
EC2074 14:50-15:05	A Virtual Reality-Based Multimodal Spatial Navigation Task with Optimal Path Planning for Rodents Author(s): Bingkun Si, Ting Pang, Junyao Zhu and Yi Yu Presenter: Bingkun Si, Henan Medical University, China Abstract: Spatial navigation in animals involves critical mechanisms of path selection and goal-directed behavior, offering key insights into learning, memory, and decision-making. While existing virtual reality (VR)-based rodent navigation studies provide high environmental controllability and reproducibility, most of them rely on unimodal sensory stimulation, which limits their ability to capture multimodal perception and adaptive behavior in complex environments. To address this limitation, we developed a multimodal navigation task framework on a rodent VR platform by integrating visual, auditory, tactile, and gustatory cues. Furthermore, we introduced the A* algorithm to generate a theoretically optimal path, which served as a benchmark for virtual maze design, training protocol construction, and behavioral evaluation. Behavioral results showed that animals gradually established stable locomotion mapping during the adaptation phase and acquired the ability to utilize unimodal cues in visual and auditory-liquid feedback training. In the final multimodal maze task, animals shifted from broad

	exploration in early trials to a stable optimal-path selection strategy. Notably, they exhibited directional adjustments toward the goal following auditory stimulation and increased locomotor speed near visual landmarks. These findings demonstrated that multimodal sensory cues effectively guided navigation and promoted strategy optimization. Our framework not only enables standardized acquisition of navigation behavior but also provided a robust experimental basis for subsequent behavioral modeling and strategy analysis. Keywords—rodent; virtual reality; spatial navigation; multimodal sensory cues; A* path planning
EC2076 15:05-15:20	Multimodal Deep Learning-Based Single-Step Autofocus Method for Liquid Lenses Author(s): Wanting Shen, Fang Zhang, Zhaoyang Liu, Zihao Gan, Meng Zhang and Huajie Hong Presenter: Wanting Shen, National University of Defense Technology, China Abstract: Compact zoom vision systems have important appli cations in fields such as robotic perception, industrial inspection, and mobile imaging. Owing to their compact size, fast response, and absence of mechanical motion, liquid lenses provide an ideal solution for constructing compact zoom imaging systems. However, liquid lenses exhibit temperature drift and nonlinear re sponse characteristics, making it difficult for traditional autofocus methods based on contrast searching or multi-step iterative ad justment to simultaneously satisfy the requirements for real-time performance and high accuracy. To this end, this paper proposes a multimodal deep learning-based single-step autofocus method for electrowetting liquid lenses and develops an online autofocus control system based on the proposed model. By fusing the visual representation of a defocused image with voltage–temperature state information of the liquid-lens zoom system, the proposed method directly predicts the required voltage increment, thereby enabling end-to-end single-step autofocus control. Furthermore, the model is deployed on a real imaging platform, where voltage prediction and lens driving are completed through a single trigger to achieve instantaneous focusing, thereby demonstrating the effectiveness and stability of the proposed method.
EC3138 15:20-15:35	Research on a High-Precision Attitude Measurement Method for On-Orbit Defocused Star Sensors Based on Image Restoration and Adaptive Filtering Author(s): Haijun Wang, Qun Hao, Yao Meng, Ting Sun, Jiaxin Lei Presenter: Haijun Wang, Changchun University of Science and Technology, China Abstract: The star sensor is a core attitude measurement sensor for high-resolution remote sensing satellites. During on-orbit operation, factors such as temperature variation, launch-induced vibration, and optical system degradation may cause severe on-orbit defocus of the star sensor, resulting in star spot energy dispersion, failure of star spot extraction, and a significant reduction in attitude determination accuracy. To address this issue, this paper establishes a complete ground-processing framework for defocused star sensor images. The proposed method employs Lucy–Richardson deconvolution for defocused star image restoration, combined with star image identification and quaternion-based attitude determination, followed by a sliding average filter with angular velocity compensation to further improve attitude accuracy. Experimental validation is conducted using four sets of on-orbit star image data (Wuhan, Shandong, Kuwait, and Jimo) acquired by the A/B dual star sensors of an on-orbit satellite. The results demonstrate that the restored star spot energy is effectively concentrated within a 6 × 6 pixel region. The attitude accuracy without filtering is better than 3″ (3σ), while the filtered accuracy reaches 0.5″ (3σ), satisfying the requirements for high-precision performance. In addition, the optical axis angle between the dual star sensors remains stable during the 0606–0607 time period. The proposed method can effectively compensate for on-orbit defocus degradation and provide reliable high-precision attitude data for satellites.

EC3121 15:35-15:50	Self-Paced Learning with Cycle-Consistent Labeling for Unsupervised Multi-Spectral Re-Identification Author(s): Wenbin Zhu Presenter: Wenbin Zhu, Nankai University, China Abstract: Supervised Multi-Spectral object Re-IDentification (SMS-ReID) methods require laborious manual trimodal annotations, hindering their application in real-world scenarios. While existing unsupervised ReID methods are confined to single visible modality or visible and near-infrared modalities, failing to effectively optimize trimodal features. To address these limitations, we introduce a new ReID task of Unsupervised MS-ReID (UMS-ReID), which is more challenging than SMS-ReID owing to the shortage in annotations and difficulty in trimodal optimization with noisy labels. To handle these challenges, this paper proposes a novel UMS-ReID framework consisting of Cycle-Consistent Labeling (CCL) and a Confidence-guided Self-paced Learning (CSL) strategy. CCL introduces a cyclic matching strategy to accurately associate cross-modality clusters with triangular consistency, providing cleaner pseudo labels for cross-modality learning. To effectively learn discriminative trimodal features against label noises within global memory banks, CSL dynamically weights training samples based on prediction probabilities to emphasize reliable ones. Different from existing optimization methods that treat training samples with equal weights, this mechanism enables a stable self-paced trimodal feature learning against noisy labels. Extensive experiments on two widely used datasets, i.e., RGBNT201 and RGBNT100, show that our UMS-ReID outperforms state-of-the-art methods by a large margin.
EC3140 15:50-16:05	Selective Multi-Level Knowledge Distillation for Understory Structural Semantic Segmentation Author(s): Zheng Hang Wang, Chong Mo, Ye Liang and Jianping Huang Presenter: Zheng Hang Wang, Northeast Forestry University, China Abstract: Understory semantic segmentation is critical for environmental perception and local scene understanding of UAVs and mobile robots, yet lightweight models often struggle with severe illumination variation, vegetation occlusion, texture similarity, and class imbalance. This paper presents a selective multilevel knowledge distillation framework based on Mask2Former. The proposed method progressively transfers knowledge from high-level backbone features, a key pixel-decoder output, and the mask feature before final prediction, so that both highlevel forest semantics and decoder-stage structural information can be inherited by the lightweight student. On WildScenes, the proposed method improves the student baseline by 2.61% in overall mIoU and by 4.08% in grouped structural-semantic mIoU, while keeping the same parameter count and FLOPs at inference time.

On-site Oral Session 3

- **Image Processing and Video Analysis**
- **Session Chair**: Zhu Meng, Beijing University of Posts and Telecommunications, China
- **Time**: 14:00-16:20, July 18
- **Meeting Room**: 102, 1st Floor
- **Papers**: Invited Speech, EC1009, EC1015, EC2042, EC2044, EC2075, EC3142, EC3145, EC4202

Invited Speech
14:00-14:20

Rui Chen
Tianjin University, China

Multi-Subject Trajectory-Guided Motion Transfer for Text-to-Video Generation

Abstract: Motion transfer has emerged as a promising approach for controllable text-to-video generation, enabling the synthesis of target videos guided by the motion dynamics of a reference video. However, current methods typically rely on simplistic attention mechanisms or basic spatial displacements, causing the features of interacting subjects to remain entangled during complex movements. In this work, we propose DiMo, a novel training-free framework for multi-subject motion transfer. The proposed approach uses semantic object masks to isolate region-aware latent representations from the internal layers of the diffusion model. By computing spatio-temporal gradients across consecutive frames, we extract explicit motion representations, termed Semantic Velocity Flow (SVF). The SVF functions as localized kinematic constraints and adaptive temporal guidance during the generation process. This strategy effectively decouples the motion trajectories of interacting subjects, mitigating representation interference while preserving structural integrity. Extensive experiments demonstrate that DiMo outperforms baseline methods across multiple metrics and human evaluations.

EC1009 14:20-14:35	Saliency-Guided Manifold-Preserving Hashing for Robust Image Copy Detection Author(s): Man Ling, Xianchuan Wang and Xiuming Chen Presenter: Man Ling, Anhui Business and Technology College, China Abstract: This paper presents a robust image hashing algorithm for copy detection that synergistically integrates Graph-Based Visual Saliency (GBVS) with Locality Preserving Projections (LPP). Unlike conventional pipelines that treat saliency detection and feature extraction as independent stages, the proposed framework employs the GBVS saliency map as a spatial attention mask to modulate Gabor-based texture features. This ensures that perceptually significant regions dominate the manifold structure preservation process. The weighted features are then projected via LPP to produce compact hash vectors that respect the local topology of the visual content. Comprehensive experiments on multiple benchmark datasets demonstrate that the proposed method achieves a True Positive Rate (TPR) of 99.84\% at threshold T = 0.98 in robustness evaluation under various content-preserving attacks, an Area Under the Curve (AUC) of 0.994 in discriminability evaluation, and an F1-Score of 0.984 in copy detection benchmarks. Systematic parameter sensitivity analysis and failure boundary evaluation further validate the algorithm's stability. Comparisons with representative traditional and deep learning-based hashing methods confirm that the proposed approach attains a superior balance between robustness, discriminability, and computational efficiency, particularly in resource-constrained scenarios.
EC1015 14:35-14:50	Controllable Neural Image Compression via Analytical Rate-Distortion Modeling and Hierarchical Modulation Author(s): Lizhe Zhang, Quan Zhou, Ruihua Liu and Lang Huyan Presenter: Lizhe Zhang, China Academy of Space Technology (Xi'an), China Abstract: Learned image compression achieves high performance but lacks precise rate controllability, which is essential for real-world deployment. Existing hierarchical models rely on discrete control parameters that do not map directly to target bitrates or adapt to content-specific sensitivities. We propose a rate-controlled enhancement framework with two key components. First, an analytical rate-distortion dynamics model is derived to provide an analytical mapping to estimate the control parameter, enabling single-pass rate control without iterative searching. Second, grounded in Information Bottleneck theory, a hierarchical latent modulation mechanism is introduced to dynamically redistribute bit budgets across multi-scale latent representations. Experimental results show that the proposed method maintains a low error rate (0.3\% at 1.0 BPP) while maintaining competitive rate-distortion performance. This approach is a practical option for systems that need to meet strict bitrate limits without the cost of training multiple models.
EC2042 14:50-15:05	Robust AIGC Image Detection via Unidirectional Consistency Regularization Author(s): Yifan Yang Presenter: Yifan Yang, communication University of China, China Abstract: The malicious proliferation of AI-generated content (AIGC) necessitates robust detection mechanisms capable of withstanding real-world image degradations. While recent large foundation models achieve impressive detection performance, their immense computational cost hinders deployment on edge devices. Conversely, lightweight detectors often suffer catastrophic failure when facing unseen generative models or severe compressions (e.g., heavy JPEG artifacts), revealing a critical vulnerability to shortcut learning on superficial high-frequency noise. To bridge this gap, we propose CoRe-Net (Consistency-Regularized Network), a highly efficient framework for robust AIGC image

	detection. CoRe-Net employs a dual-view Siamese architecture that forces the deep semantic representations of severely degraded views to align with their pristine counterparts during training. By utilizing a stop-gradient mechanism, we prevent trivial feature pollution and guide the network to learn degradation-invariant structural semantics. Extensive experiments demonstrate that CoRe-Net significantly outperforms classical CNNs and recent lightweight vision transformers under extreme degradations (e.g., JPEG Q=10). Crucially, these robustness gains are achieved with zero inference overhead, allowing our 23M-parameter model to strictly adhere to the thermal and battery envelopes of edge devices while dramatically narrowing the robustness gap with massive foundation models.
EC2044 15:05-15:20	TFR: Training-Free Cascaded Refinement for One-Shot Segmentation Author(s): Guoqiang Zhu, Xingpeng Zhuo, Xiuhe Li, Yu Zhang and Yi Zhang Presenter: Guoqiang Zhu, National University of Defense Technology, China Abstract: The Segment Anything Model (SAM) has revolutionized open-world segmentation with its prompt-driven paradigm. However, adapting SAM to segment specific object instances from a single reference image remains a challenge. Existing approaches often struggle to balance customization accuracy with inference efficiency, either relying on computationally expensive fine-tuning or suffering from structural incompleteness due to the limited information in one-shot scenarios. In this paper, we propose a novel Training-Free Refinement (TFR) framework that achieves high-fidelity personalized segmentation without any parameter updates. Our method introduces three key innovations: (1) An Augmented Multi-Scale Feature Bank that decouples semantic features from background noise across different resolutions, enabling robust matching under scale variations; (2) A Structure-Guided Cascaded Refinement module that injects non-parametric semantic heatmaps as spatial constraints to rectify SAM's segmentation bias; (3) An Adaptive Residual Completion mechanism that automatically identifies and recalls missing fine-grained components via relative semantic verification. Extensive experiments on the PerSeg benchmark demonstrate that our method outperforms training-free baselines and achieves performance comparable to fine-tuned counterparts.
EC2075 15:20-15:35	Robust Quaternion Tensor Completion for Color Video Restoration via Truncated Nuclear Norm Regularization Author(s): Runchao Feng, Ying Li and Guyan Ni Presenter: Runchao Feng, National University of Defense Technology, China Abstract: Color video restoration aims to recover high-quality video data from incomplete observations and corrupted entries. In this paper, we propose a robust quaternion tensor completion model for color video recovery, where the red, green, and blue channels are naturally represented in a quaternion framework. To better approximate the intrinsic low-rank structure, we extend the truncated nuclear norm to the domain of third-mode symmetric quaternion tensors. As a nonconvex surrogate of the rank function, it provides a tighter approximation than the conventional nuclear norm. Meanwhile, an ℓ_1-norm regularization term is introduced to characterize sparse noise and outliers. To solve the resulting optimization problem, we develop an efficient two-step algorithm based on the alternating direction method of multipliers (ADMM). Numerical experiments on color video datasets demonstrate that the proposed method achieves superior performance in terms of recovery accuracy compared with approaches from the literature.
EC3142 15:35-15:50	Semantic-aware Multi-level Line Drawing Abstraction based on Line Importance Evaluation Author(s): Xueting Liu, Shilong Deng, Chengze Li and Huisi Wu

	Presenter: Xueting Liu, Saint Francis University, Hong Kong SAR, China Abstract: Line drawing is a commonly used art form characterized by the use of lines, where the sparseness of the lines expresses the richness of the object's details to some extent. Line drawings with different levels of detail are often necessary in various scenarios and design objectives. Traditionally, this task has been performed manually by artists. Some computational methods have been proposed, but they either fail to accommodate arbitrary line drawings or do not ensure that the generated drawings with varying levels of detail maintain high fidelity to the original artwork. This paper introduces a novel multi-level line drawing abstraction method that generates different levels of detail based on a line importance evaluation metric in a semantic-aware manner. Specifically, we train a line importance evaluation network using a paired dataset of line drawings at varying levels of detail to predict the importance of lines in a line drawing. We further propose a semantic-aware line removal algorithm that generates high fidelity line drawings at different levels of detail by removing the lines based on their importance, length, and regions. Our method has been evaluated visually and quantitatively, yielding convincing results in all tested cases.
EC3145 15:50-16:05	FASA: Feature Alignment and Structural Augmentation Framework for Weakly Supervised Semantic Segmentation Author(s): Yilong Ren, Ziyang Wang, Yukun Zhao, Jian Zhang, Ruixi Zhang and Shuai Guo Presenter: Yilong Ren, Zhengzhou University of Aeronautics, China Abstract: Weakly supervised semantic segmentation (WSSS) attempts to obtain pixel-level semantic predictions using only image-level annotations. Although recent end-to-end frameworks based on vision-language pretraining have improved segmentation efficiency, their class activation maps (CAMs) still exhibit incomplete object coverage and background confusion. These limitations become more severe in complex scenes containing ambiguous boundaries and semantic co-occurrence. To address these issues, this paper proposes a lightweight single-stage framework named FASA, short for Feature Alignment and Structural Augmentation. The framework employs a frozen CLIP backbone to generate semantic priors and introduces two decoder-side enhancement modules, namely Semantic-Structure Fusion (SSF) and Cross-Dimensional Attention (CDA). SSF separates fused representations into semantic and structural branches and performs collaborative spatial-channel enhancement to recover low-response object regions. CDA further combines multi-scale spatial aggregation and multi-head channel interaction to suppress noisy activations and strengthen category discrimination. Different from conventional multi-stage pipelines, the proposed architecture jointly optimizes pseudo-label refinement and segmentation prediction in an end-to-end manner. Experiments conducted on PASCAL VOC 2012 and MS COCO 2014 demonstrate that FASA achieves 80.0\% mIoU on VOC validation, 78.6\% on VOC test, and 51.8\% on COCO validation, outperforming recent single-stage WSSS approaches while maintaining a lightweight architecture.
EC4202 16:05-16:20	Motion-Estimation-Based Network for Side Information Generation in Distributed Video Coding Author(s): Chunyu Zhang, Zhen Gong, Huaze He, Qiangyuan Xiao, Haiquan Huang, Hong Mo Presenter: Chunyu Zhang, Kunming University of Science and Technology, China Abstract: Distributed video coding (DVC), with its low- complexity encoder, provides a promising solution for power- constrained wireless video sensors. The rate-distortion perfor- mance of DVC mainly depends on the quality of the side-

	information (SI) generated at the decoder. Traditional motion estimation and compensation methods rely on the assumption of linear motion, which makes them difficult to handle scences with nonlinear or complex motion. To enhance the quality of SI, this study empolys an efficient Bilateral Correlation Motion Modeling (BCMM) network to improve the accuracy of inter-frame motion estimation. In addition, a low-complexity macroblock classifica-tion strategy is designed at the encoder end to improve encoding efficiency. Experiment results on four representative QCIF test sequences demonstrate that, for the small GOP size, the proposed scheme significantly improves SI generation quality, achieving BD-Rate savings from 12.09% to 26.17% and BD-PSNR gains from 0.60 dB to 1.82 dB compared with the DISCOVER baseline.

On-site Oral Session 4

Medical Imaging and Healthcare Applications

- **Session Chair**: Rui Chen, Tianjin University, China
- **Time**: 16:30-18:05, July 18
- **Meeting Room**: 201, 2nd Floor
- **Papers**: Invited Speech, EC2057, EC2081, EC2082, EC2083, EC3146

Invited Speech
16:30-16:50

Wali Samad
Quanzhou University of Information Engineering, China

A Robustly Optimized Level Set Method for Accurate Cardiac Image Segmentation

Abstract: Medical image segmentation plays a vital role in computer-aided diagnosis and clinical decision-making, particularly in cardiac imaging, where accurate delineation of anatomical structures is often hindered by intensity inhomogeneity, low contrast, noise, and blurred boundaries. To address these challenges, this paper proposes an efficient region-based level set segmentation framework integrated with the Alternating Direction Method of Multipliers (ADMM). By embedding ADMM into the level set evolution process, the proposed method improves optimization efficiency, enhances numerical stability, and accelerates convergence while preserving boundary precision. The region-based formulation allows the model to effectively capture local and global intensity information, making it especially suitable for complex cardiac image structures. Extensive experiments on cardiac image datasets demonstrate that the proposed approach achieves highly accurate segmentation, with the segmentation error reduced to a level close to zero and with strong agreement between the automatic and manual contours. These results indicate that the proposed framework offers a reliable and efficient solution for cardiac image segmentation and has strong potential for clinical applications.

EC2057 16:50-17:05	Few-shot Multi-scale Cascaded Network for Robust Neuronal Spike Detection in High-noise Scenarios Author(s): Shuntian Lei, Bingkun Si, Yueqi Zhao, Ting Pang Presenter: Shuntian Lei, Henan Medical University, China Abstract: Brain neuronal spike detection is a critical task in neural signal processing, directly influencing brain–computer interface decoding performance and neural coding mechanism analysis of the brain. Traditional methods using bandpass filtering to separate local field potentials often lack robustness in overlapping waveform scenarios, while existing deep learning models face challenges like limited dynamic detection capability and insufficient training samples. To address these challenges, this paper proposes a multi-scale adaptive cascaded detection framework: 1) the coarse-grained localization stage uses a modified ResNet to extract global temporal features, and 2) the fine-grained calibration stage incorporates a lightweight U-Net to analyze local gradient variations. Experimental results on synthetic datasets (spike overlap rate > 4%) demonstrate the following: 1) the coarse-grained stage achieves 98.00% accuracy with only 5% training samples under standard noise (noise standard deviation/spike amplitude = 0.05), while maintaining 94.62% accuracy under extreme noise interference (0.40× amplitude ratio), indicating strong noise robustness; and 2) the fine-grained stage sustains 95.00% detection accuracy at an IoU threshold of 0.5 under extreme noise conditions, while requiring only 10% training samples to attain a 94.24% recall rate in standard noise scenarios. This framework enables dynamic detection of overlapping spikes with limited samples, demonstrating superior noise robustness and strong generalization capability.
EC2081 17:05-17:20	Attention-Guided Graph Learning with Progressive Adversarial Training for Parkinson's Disease fMRI Classification Author(s): Yueqi Zhao, Junyao Zhu and Yi Yu Presenter: Yueqi Zhao, Henan Medical University, China Abstract: Parkinson's disease (PD) is a common neurodegenerative disorder, and early diagnosis is crucial for delaying disease progression. However, publicly available datasets are often distributed across multiple sources, with imbalanced sample distributions and inconsistent acquisition protocols, which may result in unstable model performance across different data splits and datasets. To address these challenges, this paper proposes a multi-scale feature-based brain- graph classification framework for automatic discrimination between PD patients and healthy controls. The framework integrates functional connectivity modeling, multi-scale node feature extraction, attention-based graph convolution, and progressive adversarial training. For node representation, temporal, frequency-domain, and graph-theoretic features are jointly extracted to form 11-dimensional node vectors. A threelayer attention graph convolutional encoder is employed to learn graph-level representations. In addition to the classification loss, a consistency constraint and an adversarial objective are introduced. During training, the encoder and classifier are first pretrained in a supervised manner, after which the discriminator and encoder are alternately optimized to improve representation stability and robustness while preserving discriminative capability.The experiment result shows that: compared with the MLP baseline, the conventional GCN, and model variants without attention or adversarial training, the full model consistently yields superior performance. Additional analyses on training curves, t-SNE embeddings, further indicate that the learned graph representations are both discriminative and interpretable. These results suggest that the proposed framework provides robust and effective PD classification under multi- source fMRI settings.

EC2082 17:20-17:35	Synthetic Breast X-Ray Data Augmentation and BI-RADS Risk Grading Study Based on Mask-Guided LoRA Diffusion Model Author(s): Chenglong Shi, Jia Lv, Shaoyong Guo and Zongya Zhao Presenter: Chenglong Shi, Henan Medical University, China Abstract: Mammography plays a central role in breast cancer screening, yet the development of learning-based systems is limited by data scarcity, privacy constraints, and the lack of clinically diverse annotated cases. Existing diffusion-based medical image generation methods mainly focus on global synthesis and rarely address local controllability or multi-view consistency in mammography. To address these limitations, we propose a multi-view and BI-RADS prior-guided framework for locally controllable mammography generation. The method constructs pseudo-local masks by integrating cross-view discrepancy, bilateral asymmetry, and BI-RADS-dependent spatial priors, and further introduces a Mask-Guided LoRA mechanism within a Stable Diffusion framework to enable parameter-efficient local generation. In addition, case-level consistency constraints are incorporated to improve coherence across standard four-view mammography studies. The proposed framework is expected to enhance clinical plausibility, local controllability, and data utility for privacy-preserving augmentation and medical AI development.
EC2083 17:35-17:50	Dual-Teacher Personalized Federated Learning with Hyper-Knowledge Distillation for Breast Ultrasound Classification Author(s): Yuelei Hou, Lijie Zhao, Zongya Zhao and Shouying Wang Presenter: Yuelei Hou, Henan Medical University, China Abstract: To address the issues of data silos, cross-center heterogeneity, and the scarcity of segmentation annotations in multi-center breast ultrasound diagnosis, this paper proposes a structure-aware personalized federated learning method. This method constructs a structured hyper-knowledge (Including feature_prototypes and label_prototypes) representation based on the feature prototypes of the lesion region, margin area, and background region, combined with the category soft label prototypes. It uses "dual-teacher distillation"(Including History Teacher and Hyper-Knowledge Teacher) to jointly utilize the historical teacher (preserving local personalization) and the global hyper-knowledge teacher (Help each client obtain a better personalized classification model.). At the feature level, mask-weighted average pooling is used to extract regional prototypes, and MSE loss is adopted to align sample-level regional features with global prototypes. At the label level, it uses KL divergence to align the prediction distribution of the student model with the global label prototype, thereby alleviating category confusion and negative transfer. For centers without segmentation annotations, the framework generates pseudo-masks based on a lightweight segmentation head, reducing the reliance on high-cost pixel-level annotations. The proposed method achieved an average accuracy of 89.73%, F1-score of 87.87%, and AUC of 93.33% across five heterogeneous clients.
EC3146 17:50-18:05	Anatomy of a Drift: Mitigating Autoregressive Collapse in 2.5D Medical Image Synthesis Author(s): Vladimir Valeyev, Aleksandra Vatian, Natalia Gusarova and Ivan Tomilov Presenter: Vladimir Valeyev, ITMO University, Russia Abstract: The synthesis of 3D medical volumes requires balancing high-frequency textural realism with strict anatomical fidelity. While 2.5D autoregressive models offer a computationally efficient alternative to fully 3D architectures, they suffer from compounding exposure bias and spatial drift over extended inference

	horizons. In this paper, we systematically dissect the failure modes of deterministic 2.5D autoregression. We demonstrate that while local spatial priors induce severe anatomical dragging, training the network to self-correct via standard regression objectives forces textural blur. To resolve this, we propose a stabilized framework, termed 2.5D SPADE-AR, that maps Dataset Aggregation (DAgger) to continuous pixel-space generation via a Hallucination Replay Buffer. Furthermore, we explicitly decouple spatial localization from texture generation by injecting Absolute Fourier Positional Encodings. We analyze the mathematical tension between this global coordinate prior and local Markovian context, revealing capacity-dependent vulnerabilities such as out-of-distribution activation collapse. Benchmarking against contemporary 2D, 3D, and discrete-token foundation models demonstrates that for specialized in-domain synthesis, this framework establishes a Pareto-optimal frontier, achieving superior anatomical fidelity and sequential continuity at a fraction of the computational cost.

On-site Oral Session 5

Object Detection and Industrial Applications

- **Session Chair**: Yebo Gu, Kunming University of Science and Technology, China
- **Time**: 16:30-18:30, July 18
- **Meeting Room**: 101, 1st Floor
- **Papers**: EC1024, EC2049, EC2056, EC3122, EC3123, EC2043, EC315, EC3106

EC1024 16:30-16:45	Cascaded Deraining-Enhanced Fruit Detection for All-Weather Orchard Environments Author(s): Miao Wang, Jiale Li, Lihan Wang and Guangcheng Shi Presenter: Miao Wang, Wuhan University of Technology, China Abstract: Robust fruit detection under rainfall is crucial for autonomous orchard monitoring and harvesting. This paper presents a cascaded deraining-enhanced framework combining a pretrained residual decomposition network with a YOLOv8 detector. The deraining module progressively removes rain streaks and reconstructs structural details, mitigating input-level degradation. YOLOv8 then performs accurate fruit localization and classification on restored images, maintaining a lightweight, real-time detection pipeline. Evaluated on UAV-acquired orchard images under light, moderate, and heavy rainfall, the framework consistently improves Precision, Recall, and mAP@0.5 compared with baseline and alternative deraining methods. The approach provides a modular and practical solution for all-weather fruit detection, enhancing robustness in small-scale datasets and challenging environments.
EC2049 16:45-17:00	An Optimized Lightweight YOLOv10 Framework for Handwriting-Based BMI Inference in Forensic Profiling Author(s): Nur Yana Yazid, Raihah Aminuddin, Shafaf Ibrahim and Budi Sunarko Presenter: Raihah binti Aminuddin, MARA University of Technology, Malaysia Abstract: Most forensic investigations rely on fingerprints and DNA; however, these methods become less effective when the available evidence is degraded or incomplete. In such situations, handwriting can support suspect identification, as individual characteristics such as size, shape, and stroke thickness provide distinctive biometric markers that allow for the inference of indirect physiological cues, including body mass index (BMI). This study aims to develop a handwriting-based BMI classification system using the lightweight YOLOv10 architecture to categorize handwriting samples into BMI groups. An initial dataset of 2,329 handwriting images was expanded via augmentation to 5,588 samples to evaluate the YOLOv10s and YOLOv10m variants. The YOLOv10s model was further analyzed using five optimizers: Adam, AdamW, Adamax, Auto, and stochastic gradient descent (SGD) to identify the most effective training configuration. The results show that YOLOv10s performed slightly better than YOLOv10m, achieving mAP@0.5 values of 0.994 and 0.992, respectively. Subsequently, YOLOv10s was additionally trained with different optimizers, with results showing that the model trained using the SGD optimizer achieved 0.985 precision, 0.99 recall, 0.995 mAP@0.5, and 0.938 mAP@0.95. These results validate the feasibility of employing optimized, lightweight deep learning models for real-time, mobile deployable non-invasive forensic profiling.

EC2056 17:00-17:15	Dual-Model Monocular Vision for Fast Silicon Wafer Defect Detection Author(s): Shihang Zhang, Haitian Sun and Zhiyi Zhang Presenter: Shihang Zhang, Northwest A&F University, China Abstract: Silicon wafers are essential for semiconductor manufacturing. This paper presents a low-cost silicon wafer defect detection method based on monocular vision and deep learning. To meet the dual demands of speed and accuracy in industrial settings, we develop a dedicated imaging hardware system and propose a dual-model collaborative framework, involving YOLOv11 rapidly localizes macroscopic defects, while a lightweight ConvNeXt model performs fine-grained classification of subtle micro-cracks. Trained on a self-created dataset of enhanced wafer images, the system achieves real-time inference. Experimental results demonstrate that the proposed approach delivers proper detection accuracy while significantly reducing hardware costs, making it suitable for scalable industrial deployment.
EC3122 17:15-17:30	AMFNet: A Lightweight Adaptive Multi-Scale Detector for Small UAV Detection in Complex Environments Author(s): Huayan Wen, Jingxi Shi, Lei Xie, Ruijia Liu and Xiaowen Zhang Presenter: Huayan Wen, Xihua University, China Abstract: Unmanned aerial vehicles (UAVs) have brought significant convenience but also raised concerns over privacy and public safety. However, reliable detection in real-world scenarios is challenging due to the severe feature degradation of small targets, interference from complex backgrounds, and the demand for real-time performance. To overcome these issues, this paper presents AMFNet, a compact adaptive multi-scale feature learning network for robust UAV detection. AMFNet first employs an efficient context-aware backbone that combines depthwise separable convolution, spatial-aware partial convolution, and pinwheel-shaped convolution to reduce computational redundancy while enhancing small-target representation. It then introduces an adaptive four-scale feature fusion module that preserves fine-grained high-resolution cues and dynamically aggregates neighboring feature levels. A high-resolution detection branch is further added to improve sensitivity to tiny UAVs. Experiments on the DUT Anti-UAV and DET-FLY benchmarks show that AMFNet improves recall and mAP over strong YOLO baselines while maintaining only 2.0M parameters and 9.9G FLOPs. The results demonstrate a favorable accuracy-efficiency trade-off for real-time UAV monitoring.
EC3123 17:30-17:45	An Improved YOLOv11n Detection Method via Multi-Scale Feature Enhancement Author(s): Xia Luo, Xiao Wang, Linhua Xu and Dandan Yang Presenter: XiaLuo, Guizhou University, China Abstract: To address the limitations in power system inspecti on—such as extremely small target scales, complex environmenta l backgrounds, and constrained computational resources of edge devices—this paper proposes an improved object detection metho d based on YOLOv11n. The architectural enhancements include: (1) To mitigate background interference, a Cross Stage Partial S patial Pyramid Pooling (SPPCSPC) module is introduced to restr ucture the spatial pyramid pooling layer. This expansion of the ef fective receptive field, achieved through parallel multi-scale maxpooling and cross-stage residual connections, strengthens the algo rithm's capability to extract salient features from complex backgr ounds. (2) To prevent the misdetection of tiny targets caused by d e-tail loss during downsampling, a C2PSA_HRAMi module is des igned to implement high-resolution feature compensation. By inc orporating an attention mechanism, the model significantly enha nces its perceptual sensitivity to dense, small-scale targets. (3) A weighted Bidirectional Feature Pyramid Net-work

	(BiFPN) is em ployed to replace the traditional feature fusion structure. Throug h cross-scale connections and learnable weights that dynamically adjust feature contributions, the multi-scale feature fusion is reali zed, further optimizing the localization precision for multi-scale t argets. Experimental validation on the TIC powerline dataset de monstrates that the proposed algorithm outperforms the baseline YOLOv11n model. Specifically, the Precision, Recall, mAP50, an d mAP50-95 are improved by 2.6%, 2.7%, 0.7%, and 1.8%, respe ctively, reaching 83.8%, 78.5%, 83.1%, and 63.3%. Given its sup erior performance in handling extremely small targets and compl ex background noise, the underlying archi-tectural design exhibit s significant transferability and application potential for detectin g tiny lesions with blurred boundaries in medical imaging. Keywords—Image object detection; YOLOv11n; Power line component detection; SPPCSPC; Attention mechanism; Multi-scale feature fusion
EC2043 17:45-18:00	Structure-Aware Diffusion Fine-Tuning for Few-Shot Industrial Anomaly Synthesis Author(s): Wenbo Liu, Yi Zhang, Yan Zhao and Dongxin Zuo Presenter: Wenbo Liu, Jiangsu University; National University of Defense Technology, China Abstract: Industrial anomaly detection is fundamentally limited by the scarcity and diversity of defect samples, while collecting pixel-level anomaly masks is costly in real industrial environments. Diffusion models provide a promising solution for synthetic defect generation, but existing controllable pipelines are mainly designed for natural or medical images and struggle to introduce localized high-frequency defects while preserving background consistency under limited supervision. In this work, we propose IndusDiff-FT, a data-efffcient diffusion ffne-tuning framework for controllable industrial anomaly synthesis. The framework adapts a diffusion foundation model to the industrial domain by ffne-tuning Stable Diffusion with a small number of anomaly image–mask pairs. To ensure precise defect generation, we introduce mask-constrained denoising that restricts synthesis within masked regions and prevents background contamination. To further improve structural diversity, we design a random anomaly mask synthesizer that generates diverse defect shapes under morphology-aware constraints. A feature-space consistency based quality selection strategy is also introduced to fflter low-quality samples. Experiments on MVTec AD demonstrate that the generated anomalies consistently improve downstream anomaly detection and segmentation performance while maintaining low training cost.
EC315 18:00-18:15	A Trust-Aware Behavioral Anomaly Detection Framework for Ride-Hailing Systems under Trusted Computing Author(s): Chenggang Lu Presenter: Chenggang Lu, Zhejiang University of Technology, China Abstract: Trusted computing in ride-hailing systems traditionally focuses on platform security, authentication, and system integrity. However, behavioral trustworthiness, especially abnormal driving behaviors reflected in trajectory data, remains insufficiently explored. To address this issue, this paper proposes a trust-aware behavioral anomaly detection framework for ride-hailing systems based on trajectory representation learning. Specifically, ride-hailing trajectories are modeled as multivariate time series containing geographic and motion-related features, including location, velocity, acceleration, and heading direction. A Transformer-based AutoEncoder is employed to learn normal behavioral patterns through a normal-only training protocol, where only normal trajectories are used during model training. To improve anomaly discrimination, a hybrid trust score is further designed by jointly considering trajectory reconstruction error and latent embedding distance to the normal behavior center. This enables more robust trustworthiness assessment under complex real-world conditions. Experiments

	are conducted on both synthetic trajectory datasets and the real-world Porto taxi dataset. Results show that the proposed method achieves strong performance in controlled synthetic scenarios and maintains stable anomaly detection capability on noisy real-world trajectories. The findings demonstrate that behavioral trust assessment can serve as an effective complement to traditional trusted computing mechanisms and provide practical support for trust management in intelligent transportation systems.
EC3106 18:15-18:30	HEAL: A Heterogeneous Ensemble Audit Layer for Confidence-Gated Person Re-Identification Retrieval Author(s): Yuanbo Gao, Yinzhe Ma and Shouxin Wang Presenter: YuanBo Gao, China Satellite Communications Co., Ltd., China Abstract: Deployed person ReID retrieval returns a top-1 candidate per query, yet standard pipelines provide no perquery confidence signal—a setting we term confidence-gated retrieval. Existing approaches fall short in two distinct ways. First, distance-based selective prediction saturates across canonical ReID backbones. Second, single-VLM rerankers produce binary verdicts; when their disagreement is sampled stochastically, it reflects decoding noise rather than epistemic uncertainty. To close both gaps, we propose HEAL, a Heterogeneous Ensemble Audit Layer that wraps any pretrained ReID extractor with two zerotraining components. First, a semantic non-conformity score, computed over a multi-member VLM verdict ensemble, replaces the distance score inside a recall-conditional split-conformal calibration. Second, a reason-similarity Jaccard gate converts the multi-member reason ensemble into a graded per-query signal that is structurally unavailable to any single-VLM pipeline. We adopt a cross-foundation ensemble as our default configuration: it simultaneously achieves above-chance inter-member agreement and training-corpus heterogeneity in a single deployment, providing qualitative robustness against single-vendor failure modes (training-data leakage, vendor-specific reasoning biases, simultaneous deprecation) that high-κ same-vendor ensembles cannot. Empirically, the two components yield regime-disjoint gains—on saturated benchmarks the reason gate drives a recall-conditional coverage–precision Pareto improvement, whereas on degraded benchmarks the semantic score raises accepted precision—in both cases without retraining the underlying retriever. HEAL therefore offers a deployment-ready, retriever-decoupled path to confidence-gated retrieval that operates entirely at inference time.

On-site Poster Session 1

Object Detection and Remote Sensing Imaging

- **Session Chair**: Yi Zhang, National University of Defense Technology, China
- **Time**: 14:00-16:00, July 18
- **Meeting Room**: 202, 2nd Floor

1	EC3104	An Enhanced SCTransNet Infrared Dim and Small Target Detection Method Based on Complementary Feature Extraction Author(s): Xiang Xie, Shanzhu Xiao, Jiping Yao Presenter: Xiang Xie, National University of Defense Technology, China
2	EC2033	Research on a CBP-YOLO-Based Method For Counting Silkworm Eggs Author(s): Liuchuan Liao, Peng Tang, Zhixun Liang, Yunying Shi, Jiaqi Zhao, Peng Chen and Chunting Wan Presenter: Liuchuan Liao, Guangxi Normal University, China
3	EC2064	High Precision SAR Image Sequence Registration via Block Matching and Fitting Author(s): Shize Shang, Kesai Ouyang, Qiang Cheng, Yuhao Yang, Linghao Xia and Pin Li Presenter: Shize Shang, Nanjing Research Institute of Electronics Technology, China
4	EC2089	Task-Oriented Image Restoration for Small Object Detection in UAV Images under Adverse Weather Conditions Author(s): Shiqian Huang, Fuzeng Zhang and Songping Huang Presenter: Shiqian Huang, National University of Defense Technology, China
5	EC3098	Geometry-Supported Local Fusion for Monocular Gaussian Splatting SLAM Author(s): Xinlong Qi, Yan Zhang, Lu Zhang, Xiaoxiao Guo and Hongyong Fu Presenter: Xinlong Qi, University of Chinese Academy of Sciences, China
6	EC3114	MEP-CSNet: Multi Evidence Prior Guided Center Sparse Network for Infrared Small Target Detection Author(s): Maoxuan Wang, Jinnan Gong, Yawei Li, Yongqi Mu, Yiping Liu and Tianjun Shi Presenter: Maoxuan Wang, Harbin Institute of Technology, China
7	EC3147	Design of an Ultra-Lightweight Airborne AI Computing Unit Based on Multi-Source Images Author(s): Haifeng Ma Presenter: Haifeng Ma, Northwest Institute of Mechanical and Electrical Engineering, China
8	EC1010	SLED-YOLO: A Lightweight and Accurate Detector for Small Objects in Autonomous Driving Author(s): Wei Huang, Yunchuan Yang, Lutao Wang and Hu Cheng Presenter: Yunchuan Yang, Wuhan Institute of Technology, China
9	EC3136	Lightweight YOLO-CHINS for Multi-Feature Radar Image Processing and Maritime Small Targe Author(s): Shiji Fan Presenter: Shiji Fan, University of Aveiro, Portugal

10	EC4168	Power Equipment Defect Classification Based on Semantic-Guided and Dual-Attention Chinese CLIP Author(s): Liangyi Kang, Tingting Huang, Haoyu Zhou and Shengmin Hong Presenter: Liangyi Kang, Huaqiao University, China
11	EC4179	Super-Resolution methods for radar ocean wave remote sensing images based on deep learning Author(s): Xiao Wang, Kai Sun, Dawei Li, Xuyao Li, Xuewen Ma and Xianglu Wang Presenter: Xiao Wang, Navigation Department Dalian Naval Academy Dalian, China
12	EC4198	DDA-YOLO11: A Boundary-Refined and Shape-Adaptive Network for Real-Time Steel Surface Defect Detection Author(s): JunHua Fei, Jian Wang, Zhijie Wei and Chengxiang Guo Presenter: JunHua Fei, Kunming University of Science and Technology, China

On-site Poster Session 2

Biomedical Imaging and Image Processing

- **Session Chair**: Wenhui Jiang, Jiangxi University of Finance and Economics, China
- **Time**: 16:30-18:30, July 18
- **Meeting Room**: 202, 2nd Floor

1	EC313	A Responsible and Lightweight Progressive Classification Framework for Longitudinal Glioblastoma MRI Author(s): Zisu Dong, Jianqiang Li Presenter: Zisu Dong, Beijing University of Technology, China
2	EC1013	Image Visually Meaningful Encryption Based on Information Hiding of Text Stroke Region Filling Author(s): Yuqiang Cao, Rui He and Sen Bai Presenter: Yuqiang Cao, Chongqing college of Mobile Communication, China
3	EC2072	Emotion Emerges from Shared Latent Context Author(s): Zhenyu Yang, Gensheng Pei and Xuetao Fu Presenter: Zhenyu Yang, Nanjing University of Science and Technology, China
4	EC2080	Research on Fractal Tree Image Generation Mechanism Based on Ablation Experiment Idea Author(s): Sen Bai, Xiaofan Yang, Yuqiang Cao and Xuejiao Tao Presenter: Yuqiang Cao, Chongqing Institute of Engineering, China
5	EC3131	Adaptive Image Segmentation and Sub-pixel Edge Detection for Silicon Ingot Diameter Measuremen Author(s): Kunyang Fan, Yao Ma, Bangyu Xing, Yongli Liang and Meng Chen Presenter: Kunyang Fan, Sichuan University, China; Advanced Silicon Technology Co., Ltd., China
6	EC3110	Research on 3D Dense Reconstruction Algorithm for Visual SLAM Author(s): Lin Zhang, Xu An, Xi Han, Meiting Xin Presenter: Lin Zhang, Xi'an University of Architecture and Technology, China
7	EC3152	Front-to-Side Keypoint Transformations for Sitting Posture Estimation from a Single Frontal Camera Author(s): Taspol Sawasdee, Worakan Lasudee, Thepbordin Jaiinsom, Nonthapan Wongkanha, Ratchanon Mookkaew, Supachok Butdeekhan, Tanakrit Saiphan, Narongdech Keeratipranon and Somnuek Songvanich Presenter: Taspol Sawasdee, Chulalongkorn University, Thailand
8	EC4161	Channel-based complementary fusion network for RGB-Event sensing and moving-object detection Author(s): Ziqiang Ye, Erjun Xiao, Liyuan Han, Tielin Zhang and Kebin Jia Presenter: Ziqiang Ye, Beijing University of Technology, China
9	EC4177	A Gated Fusion Network with Structural Priors for Anterior Segment OCT Segmentation Author(s): Huiying Wu, Libei Zhao, Jieli Wu and Zhiwen Chen Presenter: Huiying Wu, Central South University, China
10	EC4187	A Classification Model for the Aging Stages of Shredded Tobacco Based on Deep Learning Author(s): Haokai Zhang and Fei Xu

		Presenter: Haokai Zhang, Huazhong University of Science and Technology, China
11	EC3094	Noise-Robust Industrial Anomaly Detection via Clean-Noisy Mixup and Feature Adapter Learning Author(s): Huifang Kong, Haodong Shen, Lei Fan and Xiaopeng Yang Presenter: Haodong Shen, Hefei University of Technology, China
12	EC4200	Latent Flow Matching with Scalable Interpolant Transformers for Few-Shot Chinese Font Generation Author(s): Qing Wang, Xiaoyan Wang, Shenglan Peng Presenter: Qing Wang, Jingdezhen Ceramic University, China

Online Oral Session 1

- **Fundamentals of Deep Learning and Trustworthy Artificial Intelligence**
- **Session Chair**: Meng Ma, Northwestern Polytechnical University, China
- **Time**: 10:00-12:20, July 19
- **Meeting Room**: ROOM A
- **Papers**: Invited Speech, EC103, EC207, EC431, EC1014, EC440, EC3129, EC4175, EC2058

Invited Speech
10:00-10:20

Thong Ming Sen
National University of Singapore, Singapore

From Explainable AI to Verifiable AI: Building Trustworthy and Accountable AI Systems through Distributed Ledger Technologies

Abstract: As artificial intelligence (AI) systems become increasingly embedded in critical digital infrastructures, concerns regarding transparency, accountability, auditability, and trust continue to grow. While explainable AI has emerged as an important approach to enhancing transparency, explainability alone may not be sufficient to ensure verifiable compliance and accountability in complex AI-driven environments. The presentation explores the concept of verifiable AI and examines how distributed ledger technologies such as blockchain can complement existing AI governance mechanisms by providing immutable audit trails, transparent decision records, and verifiable compliance frameworks. Drawing upon developments in AI governance, blockchain regulation, digital identity, and public-sector innovation, the presentation proposes a multidisciplinary framework for strengthening trust in next-generation digital ecosystems while balancing innovation, regulatory compliance, and public accountability.

EC103 10:20-10:35	KRSecurity: Clickjacking Detection Based on Flow Tracking Information during Form Submission Author(s): Senmiao Wang, Huawei Wang, Lei Qiao, Tengfei Tu, Jiangyang Qiao

	Presenter: Tengfei Tu, State Key Laboratory of Networking and Switching Technology Beijing, China Abstract: Today, more than 10 million online shopping websites exist on the web, leading to the huge demand for online payment.The huge demand leads to a signiffcant surge in clickjacking during form submission, a type of cyber attack in which hackers insert malicious code into websitessss to steal ffnancial data. Clickjacking has made millions of users take the risk of payment information leakage, and has made companies receive tens of millions of ffnes because of their failure to protect customer privacy from clickjacking. In this work, we proposed KRSecurity, clickjacking detector based on information ffow tracking during form submission. It constructs the tainted data propagation paths in JavaScript code based on taint tracking and detect clickjacking by checking if the propagation path matches the path of clickjacking. Result of experiment shows that the accuracy of KRSecurity can reach 99%. We evaluated 868 web pages with KRSecurity in Alexa Top 10 million websites with our method and gave detailed analysis.
EC207 10:35-10:50	LibSleuth: Semantic-Aware and Lightweight Third-Party Library Detection for Terminal Device Supply Chain Security Author(s): Ying Yang, Chao Zhou, Can Xu, Kai Wang, Zijun Li Presenter: Zijun Li, Beijing University of Posts and Telecommunications, China Abstract: Third-party libraries introduce software supply chain risks to Android applications. Existing auditing tools are often too heavyweight for large-scale APK analysis and brittle under obfuscation. We propose LibSleuth, a semantic-aware and lightweight Android third-party library detection framework. LibSleuth reduces matching overhead through PageRank-based core class extraction, uses CodeT5-base to represent obfuscated SMALI code, and applies sliding-window multi-round matching for robust identification. On 200 real-world APKs and 155 library samples, LibSleuth achieves 92% accuracy on heavily obfuscated code, outperforming LibScout, LibPecker, LibRadar, ATVHunter, and LibScan on the same Android benchmark while reducing average matching latency to 1.71s per APK.
EC431 10:50-11:05	An Enhanced Framework for Stock Modeling and Prediction Based on Bayesian Methods Author(s): Yiqiong Xue, Xiaodong Liu Presenter: Yiqiong Xue, City University of Hong Kong, HK, China Abstract: In recent years, research in the financial domain has increasingly focused on model interpretability and the dependency structures among different indicators. However, existing approaches still exhibit limitations in jointly handling continuous and discrete data, and often fail to achieve satisfactory time-series forecasting performance for indicators with complex variation patterns. To address these issues, this paper improves upon a previously proposed stock modeling and prediction framework, aiming to further enhance its rationality and applicability. Specifically, this paper employs a static Bayesian network to ensure the rationality and consistency of network structure learning and discretization processes, and to characterize the causal relationships among various stock indicators. Building upon this, Bayesian optimization is introduced to optimize the dynamic Bayesian networks, resulting in the proposed BOL-HMM algorithm, through which the optimization time is significantly reduced while prediction accuracy is preserved. Finally, by integrating static and dynamic Bayesian models, the proposed framework enables inference and prediction for discrete indicators as well as certain continuous indicators that are difficult to model directly, and its effectiveness is analyzed from both interpretability and practical application perspectives.

EC1014 11:05-11:20	Transfer Learning-Based Dense VGG for Spontaneous Smile Recognition Author(s): Zhenzhen Luo, Xiang Zou, Jin Liu, Yong Luo, Feiguo Fu, Yiqi Cheng and Qiangqiang Zhou Presenter: Zhenzhen Luo, Jiangxi Normal University, China Abstract: Smile recognition has always been a hot computer vision topic. To improve the transmission and reuse of deep network information and enhance the generalization of smile recognition models, this research first visualizes the features extracted by different layers of a deep convolution neural network to understand its discrimination and classification mechanisms. And it analyzes unique characteristics of these features. Next, a transfer learning-based Dense-VGG model architecture is proposed. We proposed a Dense-VGG network that adopts dense connection to reuse and fuse features learned by the network to improve efficiency of inter-layer feature transfer to reduce information loss in convolution layers. Second, a large-margin softmax was introduced to enlarge inter-class variance, which helps improve smile recognition accuracy. Then, a transfer learning strategy was used to tweak the network architecture, and the ImageNet-pre-trained weights are employed to initialize the network, thus accelerating convergence. Finally, comparison tests for convergence rates, model training, confusion matrices, and classification probabilities were conducted on the GENKI- 4K and FER2013 datasets of unrestrained spontaneous images for smile recognition to validate the proposed model.
EC440 11:20-11:35	Multi-Criteria JEPA: An Energy-Based Joint-Embedding Predictive Architecture with Criterion-Level Masking for Sparse Tabular Self-Supervision Author(s): Tri Minh Huynh, Van Hung Nang Nguyen, Hanh My Thi Le, Hiep Xuan Huynh, Presenter: Tri Minh Huynh, Kien Giang University, Vietnam Abstract: Joint-Embedding Predictive Architectures (JEPA) have recently been proposed as an energy-based, non-generative paradigm for self-supervised representation learning, in which a scalar energy is associated with every input–target pair and prediction is performed in the abstract embedding space rather than in the raw input space. Promising results have been reported for images, video, and world models; the extension of the JEPA paradigm to sparse tabular signals, however, has remained comparatively unexplored, since the multi-block masking strategy adopted in image-domain instantiations is not directly applicable to tabular data in which structural patches are absent. In this paper, a Multi-Criteria JEPA (E-MC-JEPA) is proposed, in which (i) the criteria of a multi-attribute observation are taken as the natural tabular analogue of image patches, (ii) the JEPA loss is interpreted within the Energy-Based Model framework of LeCun as a latent-variable energy with the criterion-position token as the latent variable, and (iii) a 2-additive Choquet integral is appended as an interpretable readout by which the learnt embeddings are projected to a scalar prediction. The proposed criterion-level masking, embedding-space prediction, and EBM interpretation are evaluated on three sparse tabular benchmarks (Yahoo! 13, OpenTable, ITM-Rec) drawn from the MCRecKit suite. The Sugeno integral is observed to be outperformed by the 2-additive Choquet readout on every dataset; on top-*N* ranking, NDCG@10 improvements of 119%, 107%, and 63% over neural collaborative filtering are delivered by the BPR-augmented variant. Through a four-method criterion-importance convergence diagnostic, the learnt energy is shown to attend to domain-meaningful criteria—Story for movies, Service for restaurants, Ease for educational items—each consistent with prevailing findings in the respective consumer-research literature. The empirical evidence is interpreted as a first demonstration that the energy-based JEPA paradigm extends, with appropriate tabular-domain instantiation, to sparse multi-attribute data.

EC3129 11:35-11:50	A Lightweight Progressive Patch Stem Adaptation for FasterNet in Low-Resolution Image Classification Author(s): Hao Wang, Yufei Guan and Mo Tian Presenter: Yufei Guan, Dalian Neusoft University of Information, China Abstract: Low-resolution image classification is common in edge vision and embedded recognition scenarios, where available spatial information and computational resources are both limited. FasterNet improves the efficiency of lightweight convolutional backbones through partial convolution, but its original one-step stride-4 patch stem can be too aggressive for 32x32 inputs. To address this stem adaptation problem, we propose LRPS-FasterNet, a lightweight progressive patch stem adaptation for FasterNet. Without changing the PConv backbone stages or the classification head, LRPS-FasterNet replaces the original patch stem with a progressive patch stem following 32x32 -> 16x16 -> 8x8, and introduces partial local spatial refinement at the intermediate 16x16 stage. On CIFAR-100, three-seed experiments show that LRPS-FasterNet-T0 improves the average Top-1 accuracy from 58.17% to 60.42% and the average Top-5 accuracy from 81.68% to 83.59%, while increasing parameters from 2.753M to 2.759M and FLOPs from 7.455M to 8.010M. Ablation results indicate that the gain mainly comes from the progressive patch stem rather than from adding convolutional computation after the original stem, while partial local spatial refinement provides consistent additional improvement. Overall, the results indicate that adapting the input stem is an effective and low-cost way to improve FasterNet for low-resolution image classification.
EC4175 11:50-12:05	Cluster-Based Noise Inversion for Structure-Aware Noise-Gradient Estimation in NCSAM Author(s): Jiayu Xu, Zhicheng Wang and Junbiao Pang Presenter: Jiayu Xu, Beijing University of Technology, China Abstract: Instance-dependent visual label noise often follows local feature structure: visually similar images may share similar ambiguity and corruption patterns. Noise-Compensated Sharpness-Aware Minimization (NCSAM) addresses noisy-label learning by estimating a noise-induced update offset and compensating the SAM perturbation. This paper preserves the NCSAM objective and compensation rule, and studies only the simulated noise-gradient estimator inside NCSAM. We propose Cluster-Based Noise Inversion (CBNI), a structure-aware estimator that changes the statistical unit of simulated noise-gradient estimation from individual uncertain samples to feature-local neighborhoods. CBNI clusters mini-batch features, assigns each cluster an uncertainty-dependent inversion rate, and samples temporary counterfactual labels only to estimate the noise-compensation direction. It does not relabel data, select clean samples, or learn a transition matrix. A first-order residual analysis shows that cluster averaging reduces the variance of local inversion-intensity estimation under structured noise, and that the remaining estimator error enters the NCSAM perturbation residual explicitly. Experiments on CIFAR-10 and CIFAR-100 under instance-dependent noise show gains over the original NCSAM in most reported IDN settings, while symmetric-noise results clarify the limitation when corruption lacks local feature structure. Code is available at \url{https://github.com/xujiayuxu/CBNI-NCSAM}.

Online Oral Session 2

➢ Object Detection and Tracking

- **Session Chair**: Xiaomei Qu, Southwest Minzu University, China
- **Time**: 10:00-12:35, July 19
- **Meeting Room**: ROOM B
- **Papers**: Invited Speech, EC2071, EC3097, EC3125, EC3105, EC3119, EC3120, EC3153, EC4170, EC317

Invited Speech
10:00-10:20

Xin Nie
Wuhan Institute of Technology, China

Innovative Applications of Machine Vision Techniques in Industrial Product Inspection

Abstract: Machine vision technology has revolutionized industrial product inspection by enhancing accuracy and reliability. Advanced applications in detecting lithium battery electrode defects, analyzing chip porosity, and inspecting BGA defects have been explored. Integration of machine vision with deep learning and graph convolutional networks offers new strategies for high-precision quality control in industrial settings. The development of specialized algorithms and models has been a key focus in this field. For instance, the application of an innovative X-ray image enhancement algorithm based on MSR and RPCA has significantly improved image clarity for lithium battery inspection. Additionally, the use of the LBPEDNet network has enabled precise detection of pole-piece endpoints in lithium batteries. Furthermore, the ASTM-DGCN model has been employed for industrial robot posture prediction. These advancements address challenges such as poor imaging quality and limited computing resources. They contribute to the progress of industrial automation and provide valuable insights for researchers and practitioners in the field of machine vision technology.

EC2071 10:20-10:35	Region-Aware Gait Recognition with Spatial and Temporal Region Modeling Author(s): Lina Ai, Chin Poo Lee and Kian Ming Lim Presenter: Lina Ai, University of Nottingham, China

	Abstract: Gait Energy Image (GEI) remains a widely used representation in gait recognition due to its compactness, effi- ciency, and privacy-preserving nature. However, by averaging a gait sequence into a static template, GEI inevitably weakens fine- grained temporal cues, which limits its robustness under chal-lenging conditions such as clothing changes and carrying varia- tions. To address this issue, this paper proposes a local region- aware gait recognition framework based on sliding-window GEIs. The proposed method employs an Attention-Guided Residual Backbone to extract shared features, a Spatial Branch to model fine-grained region-level structure, and a Temporal Branch to capture inter-window temporal variation of local body regions. In this way, the framework preserves the compactness of GEI while strengthening discriminative structural and temporal cues. Experimental results on the CASIA-B dataset demonstrate the effectiveness of the proposed method, achieving average rank-1 identification rates of 96.2%, 92.3%, and 74.7% under normal walking (NM), walking with a bag (BG), and walking with a coat (CL), respectively.
EC3097 10:35-10:50	SA-TAL: Scale-Aware Task-Aligned Learning for Efficient Small Object Detection in Traffic Scenes Author(s): Yutong Zhang Ammar Hawbani Liang Zhao Saeed Alsamhi Presenter: Yutong Zhang, Shenyang Aerospace University, China Abstract: Small object detection in traffic scenes is challenging due to limited appearance, background clutter, and scale bias in label assignment. Recent real-time detectors still suffer from insufficient supervision for small objects. We propose a unified framework with Scale-Aware Task-Aligned Learning (SA-TAL), which mitigates scale-induced bias by dynamically adjusting alignment scores during training, improving supervision balance across scales. A high-resolution P2 detection head preserves fine-grained details, while CBAM suppresses background noise and enhances discriminative features. Experiments on KITTI show that our method achieves 0.892 mAP@0.5 at 208 FPS, outperforming YOLOv8 by 6.6%. Cross-domain evaluation on VisDrone confirms consistent improvements in dense small-object scenarios, demonstrating a superior accuracy-efficiency trade-off for real-time intelligent transportation systems. Index Terms—Object Detection, Small Object Detection, YOLOv8, Attention Mechanism, Intelligent Transportation Systems, Real-Time Detection
EC3125 10:50-11:05	META-Tracker: Multi-model Energy Topology Association for Infrared Small Target Tracking Author(s): Longdan Li, Luping Wang and Wuming Wu Presenter: Longdan Li, Sun Yat-sen University, China Abstract: Infrared small target tracking in complex scenarios poses formidable challenges due to the targets' lack of discriminative texture. To address these issues, this paper proposes multi-model energy topology association framework (METATracker) tailored for infrared small targets. First, the adaptive interacting multiple model (AIMM) is proposed to replace the single Kalman filter, effectively predicting abrupt motions. Furthermore, the scale-invariant distance-shape (SIDS) metric is introduced to provide continuous cost gradients, addressing the fragility of the traditional IoU metric on small targets. Finally, local energy topology embedding (LETE) module is designed to extract lightweight thermal radiation signatures, replacing heavy deep re-identification networks that are prone to background noise interference. Experiments demonstrate that the proposed algorithm achieves state-of-the-art performance.
EC3105 11:05-11:20	RDAF-YOLOv11: A Rearrangement-Driven Adaptive Fusion Network for UAV Small-Object Detection Author(s): Ziguan Liu, Wenjun Ma, Huagang Shao, Ruijun Yang

	Presenter: Ziguan Liu, Shanghai Institute of Technology Faculty of Intelligence Technology, China Abstract: Infrared dim small target detection has attracted extensive attention in recent years. U-shaped networks have greatly advanced the development of this field. However, infrared dim small targets suffer from a lack of shape and texture features. Conventional convolutions tend to lose detailed information, accompanied by insufficient feature fusion and unsmooth gradient flow, which degrade detection performance under complex backgrounds. Especially in scenarios with complex backgrounds and low signal-to-noise ratios, missed detection and false alarms are prone to occur. To address the above issues, this paper proposes an enhanced CDF-SCTransNet detection network based on SCTransNet. The specific improvements are as follows: First, a Complementary Feature Extraction (CDF) module is designed, which integrates central difference convolution, deformable convolution, and wavelet transform to extract complementary target features from multiple dimensions, enhancing detailed target features while expanding the receptive field. Second, a dual-layer multi-scale feature fusion architecture is constructed, combining MCA, HFF, SSCA, and CCA modules to achieve deep interaction between shallow and deep features and suppress background clutter. Third, a Progressive Feature Decoding (PFD) module and a composite loss function are introduced to strengthen gradient flow and improve segmentation accuracy and edge prediction. The proposed method achieves mIoU scores of 94.1%, 78.93%, and 71.4% on three public datasets, namely NUDT-SIRST, NUAA-SIRST, and IRSTD-1K, respectively, outperforming current mainstream infrared dim small target detection algorithms, which verifies its superior capability for dim small target detection in complex backgrounds.
EC3119 11:20-11:35	An Improved YOLOv8s for Steel Surface Defect Detection with Multi-Scale Context Compensation and Large-Kernel Spatial Attention Author(s): Lei Song, Xu E, Tianyuan Yang, Jiayin Li and Zhihao Yang Presenter: Lei Song, Bohai University, China Abstract: Steel surface defect detection is challenged by multi-scale defects, strong background interference, and weak-texture targets. Existing YOLO-based improvements often enhance local structures or stack modules, but the deployment rationale across feature stages is not always explicit. Rather than treating WASPP and LSKA as independently stacked modules, this work introduces a hierarchical feature mismatch perspective and aims to match different enhancement mechanisms with the heterogeneous demands of different feature stages. The main novelty lies in establishing a task-driven deployment chain that links feature-stage bottlenecks, enhancement mechanisms, and deployment positions. Under this perspective, the detection bottleneck is decomposed into insufficient multi-scale context before feature fusion and limited spatial selectivity after feature fusion. Accordingly, a weighted atrous spatial pyramid pooling (WASPP) module replaces SPPF at the end of the backbone to compensate deep semantic context, while large-kernel spatial attention (LSKA) is inserted after the P3, P4, and P5 detection-scale features to strengthen defect-related spatial responses. Experiments on the NEU-DET dataset show that the proposed method achieves 81.7% mAP@50 on the validation set and 78.9% on the test set, improving the baseline by 4.4 and 1.7 percentage points, respectively. The final model contains 18.46M parameters and 34.7 GFLOPs, with a measured inference time of 0.9 ms per image under the same input-size setting. The results suggest that matching enhancement mechanisms with feature stages is beneficial for steel surface defect detection.
EC3120 11:35-11:50	An Improved YOLOv11-seg Method for Complex Pipe Segmentation Author(s): Keyi Liao and Xu Zhang

	Presenter: Keyi Liao, Beijing Institute of Technology, China Abstract: This paper addresses the challenges of instance segmentation of complex pipes in industrial scenes, such as slender structures, multiple branches, occlusions, and cluttered backgrounds. A lightweight method is proposed based on a dual improvement strategy. A C3k2Strip module is embedded in the backbone to improve the representation of slender pipes and boundary details, while a ProtoBridge module is introduced in the segmentation head to improve multi-scale feature fusion and contextual modeling. A synthetic dataset with 2,500 images of complex pipes is constructed for training and evaluation. Experimental results show that the proposed method achieves gains of 2.4% in mAP@0.5 and 4.9% in mAP@0.5:0.95 over the baseline and outperforms Mask R-CNN and YOLACT. The method maintains low computational complexity while improving segmentation performance, making it suitable for real-time industrial applications.
EC3153 11:50-12:05	Two-Stage Few-Shot Contrastive Enhancement for Long-Tailed Traffic Participant Detection Author(s): Xinxing Chen, Chin Poo Lee and Kian Ming Lim Presenter: Xinxing Chen, University of Nottingham Ningbo, China Abstract: Traffic participant detection in complex traffic scenes is critical for autonomous driving perception. However, real-world road data often presents a long-tailed distribution, causing baseline detectors to favour head classes and perform poorly on rare traffic participants. To address this problem, this paper proposes a two-stage Few-Shot Contrastive Enhancement (FSCE) framework for long-tailed traffic participant detection. The framework first employs YOLO26 as the base detector to perform preliminary localisation, classification, and proposal generation. Then, with the backbone and detection head of the base detector frozen, a lightweight FSCE branch is introduced to perform discriminative embedding learning, multiprototype memory modelling, and selective tail-class re-ranking on proposal-level Region of Interest (ROI) features. During training, the auxiliary classification loss, prototype classification loss, supervised contrastive loss, and branch consistency loss are jointly optimised, while incorporating tail-class reweighting and prototype warm-up for stable long-tailed learning. During inference, base detector classification priors are fused with the outputs of the auxiliary classification branch and the prototype classification branch, and then selective tail-class gating is applied to correct low-confidence proposals with sufficient tail-class evidence. Experimental results on BDD100K dataset demonstrate that the proposed framework achieves a mean Average Precision at an Intersection over Union threshold of 0.5 (mAP@0.5) of 0.659, improving upon YOLO26 baseline from 0.563 by 0.096, corresponding to a relative gain of approximately 17.1%. For tail classes, the AP@0.5 values of bike, motor, and train increase from 0.493, 0.461, and 0.008 to 0.682, 0.769, and 0.459, respectively. These results show that the proposed framework effectively enhances long-tailed traffic participant detection while maintaining the main architecture of the base detector and stable detection performance on head classes.
EC4170 12:05-12:20	Improvement of YOLOv8n Object Detection Algorithm for UAV Aerial Images Author(s): Heng Lu and Xianchun Zhou Presenter: Heng Lu, Anhui Jianzhu University, China Abstract: UAV aerial images pose persistent challenges: dense small objects, extreme scale variation, severe occlusion, and cluttered backgrounds. Conventional detectors fail to reliably detect fine-grained targets under these conditions. This paper proposes ADS-YOLOv8n, an improved YOLOv8n-based detector for UAV aerial object detection. Three targeted modifications are

	introduced. First, a Lossless Multi-scale Attention Path (LMA-Path) in the neck injects high-resolution P2 features via SPDConv and expands the receptive field via CSPOmniKernel. Second, an Enhanced Multi-Scale Detection Head (EMSD-Head) incorporates a P2-HR branch and DyHead-DCNv3 for spatially adaptive cross-scale feature alignment. Third, C2f_RFCBAMConv replaces standard C2f in the detection head, suppressing background clutter through dual channel-spatial attention. On VisDrone2019, ADS-YOLOv8n achieves mAP50 of 40.18%, a gain of 4.93% over the YOLOv8n baseline, outperforming YOLOv7-tiny by 3.48% and Faster R-CNN by 3.38%.
EC317 12:20-12:35	A Multi-Threat Cooperative Evasion Decision Method for UAVs Based on Improved PPO Author(s): Wei Li Presenter: Wei Li, Naval Submarine Academy, China Abstract: To address the autonomous evasion decision problem of Unmanned Aerial Vehicles (UAVs) facing pursuit by multiple capturing aircraft in dynamic multi-agent environments, this paper proposes a cooperative evasion method based on improved Proximal Policy Optimization (PPO) deep reinforcement learning. The proposed method consists of three components: (1) a threat-aware fixed-dimension state representation that encodes the relative distances, azimuth angles, and velocities of a varying number of threats into a fixed-length state vector; (2) a heading-offset-based continuous action space that enables the target to flexibly select evasion directions; and (3) a reward function based on post-decision proportional evaluation, which provides dense and timely feedback by calculating the proportion of successfully evaded threats after each decision, thereby addressing the slow convergence caused by sparse rewards. Experimental results in various scenarios demonstrate that the proposed method enables UAVs to adaptively select optimal evasion directions, significantly reduces the probability of being captured, and exhibits strong generalization capability.

Online Oral Session 3

➢ Image Processing and Enhancement

- **Session Chair:** Mingtao Liu, Linyi University, China
- **Time**: 10:00-12:05, July 19
- **Meeting Room**: ROOM C
- **Papers**: Invited Speech, EC2047, EC2084, EC2086, EC4164, EC4184, EC4191, EC2079

Invited Speech
10:00-10:20

Guoping Xu
Wuhan Institute of Technology, China

Fundamental Development of Deep Learning in Computer Vision: A Systematic Review

Abstract: This presentation systematically surveys landmark studies and pivotal methodological breakthroughs driving the evolution of deep learning within computer vision. We classify mainstream progress into six core dimensions: data augmentation, neural network architectures, optimization algorithms, feature representation and manipulation, learning paradigms and downstream tasks, and fundamental design principles for high-performance deep neural networks. For each dimension, we elaborate on representative foundational techniques validated with outstanding effectiveness, robustness and generalization across diverse visual tasks. We further conduct critical analysis of open challenges, inherent limitations, and promising emerging research frontiers in contemporary computer vision.

EC2047 10:20-10:35	Optimized Hybrid Transformer Architecture for Zero- Reference Low-Light Image Enhancement Author(s): Yile Yan, Lei Zhang and Ziwen Gong Presenter: Yile Yan, Tianjin University, China Abstract: Low-light image enhancement is an important preprocessing step for computer vision applications. This paper proposes Optimized Hybrid DCENet, a compact CNN-Transformer network for zero-reference low-light enhancement. The proposed model is not a direct modification of the original Zero-DCE network.

	Instead, it only adopts the iterative light-enhancement curve as the final image adjustment rule, while using an independently designed hybrid estimator to predict pixel-wise curve parameters. The network combines local convolutional features, global illumination context, and channel attention to generate enhancement curves. Experiments on the SICE dataset show that the proposed method achieves competitive PSNR and SSIM while reducing the number of trainable parameters by approximately 32% compared with Zero-DCE.
EC2084 10:35-10:50	Full-Reference Image Quality Assessment Based on Multimodal Large Language Model: A Comparative Study on TID2013 Dataset Author(s): Renjie Wu, Xiao Kang, Nan Yang, Junyu Han, Wanru Peng and Yifei Yang Presenter: Renjie Wu, China North Intelligence & Innovation Research Institute, China Abstract: Full-Reference Image Quality Assessment (FR-IQA) based on traditional pixel-level or local feature matching metrics struggles to align with human subjective perception in different image distortion evaluation. To address this limitation, this paper proposes a FR-IQA method driven by the Qwen3-VL multimodal large language model (MLLM), which leverages the global semantic understanding capability of MLLMs for image quality assessment. Comparative experiments are carried out on the TID2013 dataset, which encompasses 24 types of image distortions, to systematically evaluate the performance of Qwen3-VL models with different parameter configurations (2B, 4B, and 8B). The proposed models are further compared with image quality assessment approaches including SSIM, PSNR, FSIM and DIQaM-FR with Spearman's rank-order correlation coefficient (SROCC) and Kendall's rank-order correlation coefficient (KROCC) adopted as quantitative evaluation indicators. Experimental results demonstrate that the proposed MLLM-driven FR-IQA method can effectively perceive various image distortions and achieves remarkable performance (SROCC=0.72; KROCC=0.563). Limitations and future work are discussed in this paper for MLLM assisted FR-IQA performance improvement.
EC2086 10:50-11:05	Surface-aware Blind Color Image Deconvolution with Inverse Filtering Author(s): Shaowen Yan, Yingjun Li, Xuekun Yang, Xue Lu and Lijun Liu Presenter: Shaowen Yan, Beijing Vocational College of Agriculture, China Abstract: We propose a novel regularized inverse filtering framework for blind deconvolution of color images, designed to recover original images without prior knowledge of the point spread function (PSF). By estimating independent inverse filters for each RGB channel, our method effectively leverages multichannel information. Specifically, we integrate a surface-aware regularization term into the STAR-RIF framework to suppress artifacts and preserve sharp edges. Furthermore, to exploit the sparsity of document images, we substitute the l2-norm with an l1-norm in the non-negativity and background constraints, significantly improving restoration performance for text-based content. The resulting convex minimization models are solved via a first-order primal-dual algorithm. Numerical experiments demonstrate that our proposed methods achieve competitive performance compared to existing techniques, showing improvements in peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and visual clarity across various blur types.
EC4164 11:05-11:20	FruParse:Topology-Aware Restoration for Fruit Cell Microscopy Images Author(s): Yibo Liu, Qilei Li and Di Wu Presenter: Yibo Liu, Zhejiang University, China Abstract: Cell wall integrity is critical for fruit cellular image analysis and cell biology research. Conventional image restoration algorithms exhibit poor adaptability to fruit cell wall reconstruction and suffer from several limitations:

	(1) Cell walls of ripe fruits have low distinguishability against microscopic backgrounds; (2) Reconstruction of fractured cell walls requires domain-specific biological knowledge; (3) Textures of restored cell walls deviate greatly from actual morphological features. To address these limitations, we propose FruParse, a structure-aware computational framework dedicated to fruit cell wall restoration and morphological alignment where all modules maintain thin-wall continuity, topology preservation and morphological consistency. Specifically, the framework consists of three core modules: a connectivity predictor that generates dense probability maps for missing cell wall segments, a supervised inpainting module that localizes fractured regions, and a texture alignment module that recovers biologically realistic cell wall morphology. Evaluated on a ripe fruit cell microscopy dataset with unified defect masks, FruParse achieves superior topology-aware restoration performance over both classical and learning-based inpainting methods. Compared with Telea and Navier–Stokes inpainting, our method greatly reduces wall-width error and enhances structural continuity. It also outperforms the supervised U-Net baseline in terms of structural continuity and masked-region Learned Perceptual Image Patch Similarity (LPIPS). Although texture alignment causes a slight drop in global Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), the results demonstrate that our framework prioritizes biologically reasonable cell wall morphology and thin-structure consistency, instead of merely pursuing global pixel-level similarity. Cross-domain experiments further verify that FruParse delivers more stable reconstruction across various microscopy scenarios. Additional comparisons against a topology-augmented U-Net reveal that topology-aware refinement better preserves cell wall continuity and morphology than methods relying purely on topology-related loss functions. The present study provides innovative methodologies and insights for cellular image analysis of ripe plant fruits.
EC4184 11:20-11:35	Spatial-Channel Transformer Based Adaptive Convolutional Dictionary Learning Network for Image Denoising Author(s): Zheng Wang and Xizhan Gao Presenter: Zheng Wang, University of Jinan, China Abstract: Image denoising is a fundamental research task in computer vision with considerable practical and theoretical significance. Inspired by the success of deep neural networks (DNNs), deep unfolding methods integrate traditional image modeling techniques into DNNs, possessing both model-based interpretability and data-driven learning capability for image denoising. However, these methods still suffer from limited dictionary representation, weak prior modeling, and restricted cross-stage feature propagation. To address these limitations, we propose an adaptive convolutional dictionary learning transformer network (ADTNet). The proposed method incorporates a content-adaptive convolutional dictionary mechanism, a spatial-channel hybrid transformer for prior learning, and a selective preservation memory (SPM) module, which enable adaptive feature representation and effective cross-stage information propagation. Extensive experiments on natural and medical image denoising datasets demonstrate the superiority of ADTNet in both quantitative evaluation and visual quality.
EC4191 11:35-11:50	Wavelet-Enhanced CLIP-Based Text-Driven Image Style Transfer Author(s): Hongbin Cai, Guangsheng Ma Presenter: Hongbin Cai, North China Electric Power University, China Abstract: Text-driven image style transfer renders content im ages with textures guided by natural language prompts. Existing CLIP-based methods, while effective, lose fine-grained details during downsampling, particularly for texture-dense styles such as sketch and engraving. We propose a wavelet-enhanced CLIP

	framework that explicitly preserves high-frequency information. Our key contribution is a wavelet-enhanced skip-connection module that caches high-frequency wavelet coefficients at each downsampling stage and reinjects them directly into the de coder, protecting edges and fine lines from compression loss. Experiments on the official CLIPstyler test set demonstrate that our method achieves an average CLIP similarity of 0.3093, outperforming the baseline of 0.3067, with the largest gain on the Sketch style. A 50-image dataset further confirms strong generalization without fine-tuning. Index Terms—Text-driven style transfer, Wavelet transform, CLIP, Frequency decomposition, Skip connection, Haar wavelet
EC2079 11:50-12:05	Tire Pattern Image Processing: Recent Advances, Practical Applications and Future Directions Author(s): Yichen Wu, Shuhao Zhao, Jixiang Du and Chiew Tuan Kiang Presenter: Shuhao Zhao, Xi'an University of Posts and Telecommunications, China Abstract: In recent years, tire pattern image processing has gained significant attention for its applications in traffic safety, industrial quality inspection, and environmental impact assessment. Tire image analysis faces significant challenges, including complex acquisition conditions, pronounced illumination variations, reflections and shadows, frequent contamination, and background interference. Additionally, fine- grained tread pattern differences and the prevalence of small- scale, low-contrast defects further complicate the task.This paper reviews key studies from the past five years and summarizes major research directions, including image enhancement and preprocessing, feature extraction and representation, classification and recognition, similarity detection and matching, as well as damage detetion and wear assessment. It also discusses typical application scenarios and the current status of dataset construction.A comprehensive analysis indicates that deep learning-based methods have substantially improved recognition and localization under complex conditions. However, several critical challenges remain: robustness to diverse backgrounds, limited training samples and class imbalance, effective multimodal fusion, and a lack of standardized industrial-level evaluation protocols and reproducible data resources.Future research is expected to focus on lightweight models and edge deployment, multimodal perception and feature collaboration, open datasets and unified benchmarks, and edge-cloud collaboration with incremental adaptation, thereby supporting the large-scale deployment and practical application of tire image processing technologies.

Online Oral Session 4

- **Medical Imaging and Biometric Recognition**
- **Session Chair**: Dipali Kasat, Sarvajanik College of Engineering & Technology, India
- **Time**: 13:30-16:05, July 19
- **Meeting Room**: ROOM A
- **Papers**: EC1005, EC2054, EC3124, EC314, EC1018, EC2070, EC3111, EC3130, EC2045, Invited Speech

Invited Speech
15:45-16:05

Vesna Zeljkovic
Lincoln University, USA

Image Processing Selected Topics Applied in Biomedical Engineering

Abstract: This talk encompasses several different digital image processing approaches applied in selected biomedical engineering domains. Algorithm for brain structural changes documentation of selected image(s) and a tool assisting doctors in accelerated diagnosis and quantification of the degree of abnormality will be presented. Then, three different computer aided acne detection methods, applied in the domain of dermatology, will be outlined: method A that uses CIELAB Color Space, method B based on a saturation component involving sinus hyperbolics function and method C focused on contrast enhancement. Next, quantification of vaccination effectiveness in viral infectious diseases affected patients will be evaluated by the application of algorithm that numerically quantifies visual symptoms of viral infection and comparatively assesses the degree of disease manifestation in two groups of patients vaccinated versus unvaccinated. After that, statistical analysis of mammogram images as a preprocessing phase in computer assisted breast cancer detection to aid assessment and quantitation of breast tissue malformations indicative of benignancy, malignancy or calcification, will be presented. Finally, the multidimensional aspect of the topography flow parameters, applied in cigarette smoking studies, which reflect the physical and chemical characteristics of the tobacco product, will be outlined using data taken from a population-based study, that aim to better understanding the mechanics of factors that affect the dose of tobacco inhalation.

EC1005 13:30-13:45	Lightweight Facial Emotion Recognition Based on Improved Mini-Xception with Focal Loss and Residual Connections Author(s): Weijia Li, Chuye Zheng, Jing Cai, Jiayi Lv and Xiaoyu Gong Presenter: Weijia Li, Jilin University, China Abstract: Facial emotion recognition is a core research direction at the intersection of computer vision and affective computing, aiming to decode the emotional information embedded in human facial expressions. Current mainstream deep learning models suffer from issues of large parameter sizes and high computational costs, while facial emotion recognition tasks also face challenges including class imbalance and susceptibility to overfitting in lightweight models. To address these problems, this paper proposes a lightweight network, Mini-Xception, which is based on the depthwise separable convolution of Xception. It integrates residual connections, adopts Focal Loss combined with L2 regularization, applies "BN+ReLU" processing, embeds the SE (Squeeze-and-Excitation) attention module, and replaces fully connected layers with global average pooling. Trained and validated on the fused and cleaned FER2013 and CK+ datasets, the model achieves an average accuracy of 82.3% in recognizing 7 categories of emotions.
EC2054 13:45-14:00	Quality-Aware and Enhancement-Assisted Interpretable Hybrid Network for Diabetic Retinopathy Severity Grading Author(s): Bingyu Xiao, Weishu Li and Yan Ma Presenter: Yan Ma, Jilin University, China Abstract: Automated diabetic retinopathy (DR) severity grad- ing from fundus photographs is critical for scalable screening, yet its reliability deteriorates when image quality is poor— a common occurrence in real-world clinical workflows. Most existing grading models treat all inputs uniformly, ignoring quality variation; meanwhile, image enhancement applied in- discriminately can distort features in already adequate images. We propose QEGNet, a lightweight CNN–Transformer hybrid network that jointly predicts a fundus image quality score and a four-class DR severity grade. A quality-gated fusion mechanism selectively activates a feature-space enhancement branch for low- quality inputs while bypassing it for high-quality ones, directly addressing the empirically documented risk of over-enhancement. We evaluate QEGNet on MESSIDOR (1,200 images, 4-class grading). QEGNet achieves a quadratic weighted kappa (QWK) of 0.817±0.009 as a single model and 0.900 as a 3-seed ensemble, surpassing the EfficientNet-B0 baseline (0.840 ± 0.018). QEGNet further provides quality-aware capabilities absent from plain classifiers: quality-stratified analysis reveals +9.5 percentage points improvement on high-quality images, and referable DR detection sensitivity reaches 0.880 versus 0.827—a clinically meaningful advantage. Ablation experiments confirm that both the quality branch and the enhancement branch contribute to overall performance. These results suggest that explicit quality modeling and conditioned enhancement can improve the robust- ness and clinical utility of automated DR grading.
EC3124 14:00-14:15	Lightweight Finger Vein Recognition via Binary-Topographic Feature Fusion for Low-end Microcontrollers Author(s): Guodong Zhao, Xueshuang Li, Chuanxian Xin, Shuang Yang and Panpan Qiu Presenter: Chuanxian Xin, Saint Deem Technology Incorporated Co., Ltd. Hangzhou Branch, China Abstract: During the development of finger vein recognition algorithms, the algorithm determines both the feature size and the processing time, as well as the recognition accuracy. Feature size, processing time, and recognition accuracy are

	three key metrics for evaluating the performance of finger vein recognition algorithms. Currently, the commonly used finger vein recognition algorithm schemes include those based on image texture features, image orientation features, image topographic features, and deep convolutional neural net-works. These finger vein algorithm schemes have achieved good results in terms of recognition accuracy, but they exhibit varying performance in feature size and hardware dependency. It is difficult for these algorithm schemes to achieve a balance among feature size, algorithm time, and recognition accuracy on low-end microcontroller chips. To address these issues, this paper proposes a recognition algorithm that integrates finger vein binary features and simplified topographic concavity-convexity features. This method is characterized by small feature size, low algorithmic complexity, and high recognition accuracy, with low dependence on hardware resources. It can be efficiently deployed directly on low-end microcontroller chips, demonstrating excellent application prospects.
EC314 14:15-14:30	YOLO-GBP: A Multi-Strategy Enhanced YOLOv13 for Cervical Cell Image Detection Author(s): Mengrong Chen, Na Li, Kuangpin Liu, Xudong Jiang, Yushun Jiang Presenter: Mengrong Chen, Yunnan University of Chinese Medicine, China Abstract: Artificial intelligence is advancing cervical cancer detection, but current methods struggle with overlapping cells, blurry boundaries, and unclear nuclei—and incur high computational costs, limiting clinical use. We propose YOLO-GBP: a lightweight, accurate detector built on enhanced YOLOv1. It features three innovations: (i) GhostConv in the backbone reduces redundancy via linear projection, cutting computation without sacrificing accuracy; (ii) Weighted BiFPN in the neck fuses top- down semantics and bottom-up details; (iii) PfAAM replaces Focus to dynamically balance spatial and channel attention, improving detection of small or overlapping cells. On SIPaKMeD, YOLO- GBP achieves 93.9% mAP@0.5 with only 2.45M parameters and 6.2 GFLOPs—outperforming YOLOv8, YOLOv11n, and Cell YOLO. It is an efficient, clinically viable cervical screening tool.
EC1018 14:30-14:45	Fusion-Based Deep Learning Models for Multi-Class Classification of Eye Diseases using Fundus Images Author(s): Ayesha Bintee Rob and Shafiqul Islam Fahim Presenter: Ayesha Bintee Rob, BRAC University, Bangladesh Abstract: Around the world, eye conditions like Cataracts, Diabetic Retinopathy, and age-related Macular Degeneration (AMD) remain the main causes of vision loss. Automated diagnosis from fundus images is essential for faster, more reliable identification, as manual screening is traditionally time-consuming and prone to human error. In this paper, we propose fusion-based frameworks that takes advantage of the distinct strengths of three different backbones: the Swin-Tiny Transformer, DenseNet121, and EfficientNetV2-S. We separately assessed two fusion strategies: a feature-level multi-head self-attention mechanism that prioritizes informative features and decision-level fusion through weighted soft voting and meta-model stacking. The system was trained and evaluated using the AMDNet23 dataset and pre-processing steps like resizing, normalization, and data augmentation were applied to ensure robust generalization. According to our analysis, while the standalone Swin-Tiny model achieved 95\% accuracy, both fusion methodologies significantly improved the performance, reaching up to an accuracy of 97.5\%. This improvement was particularly noticeable in complex classes such as diabetic retinopathy, where the combination of local convolutional features and global transformer context refined classifications that were difficult for individual models to achieve. This framework demonstrates its ability as a dependable clinical tool that can help ophthalmologists perform retinal screenings more quickly and accurately in actual hospital settings.

EC2070 14:45-15:00	AMR-YOLO: Adaptive Multi-scale Refinement for Platelet-Aware Blood Cell Detection in Microscopy Author(s): Yuwei Shi, Chin Poo Lee and Kian Ming Lim Presenter: Yuwei Shi, University of Nottingham Ningbo China, China Abstract: Accurate microscopic blood cell detection is essential for automated hematology, yet platelet detection remains particularly challenging because platelets are extremely small and highly sensitive to staining variation. Standard YOLO backbones rely on early feature downsampling and limited intermediate multiscale representation, which can attenuate platelet-scale cues and reduce robustness under domain shift. To address this issue, this paper proposes Adaptive Multi-scale Refinement YOLO (AMR-YOLO), a topology-preserving backbone enhancement for YOLO-style detectors. The proposed framework introduces an Adaptive Multi-scale Refinement block that combines multi-scale context fusion with residual feature refinement, strengthening intermediate representations without modifying the detection head, anchor setting, or training objective. Extensive experiments on peripheral blood microscopy datasets demonstrate that AMRYOLO consistently improves platelet detection, indicating that targeted backbone-level refinement is effective for platelet-aware blood cell detection.
EC3111 15:00-15:15	FAM-YOLO: Feature-Adaptive Multi-Scale YOLO for Blood Cell Detection Author(s): Yu Wei Shi, Chin Poo Lee and Kian Ming Lim Presenter: Yuwei Shi, University of Nottingham Ningbo China, China Abstract: Accurate blood cell detection in peripheral blood smear images is important for computer-aided hematological analysis. However, automated detection remains challenging because red blood cells, white blood cells, and platelets exhibit large variations in scale, morphology, and visual appearance. In particular, platelets are tiny, weakly contrasted, and easily confused with staining artifacts or background debris. To address these challenges, this paper proposes FAM-YOLO, a Feature- Adaptive Multi-Scale YOLO detector for blood cell detection. FAM-YOLO enhances a YOLOv8-style detector by introducing two lightweight feature refinement modules into the backbone. The Local Adaptive Feature Modulation module strengthens boundary-sensitive and texture-level representations by combining edge-sensitive local responses with low-frequency contextual cues. The Multi-Scale Aggregation module enriches feature maps by aggregating spatial information from multiple receptive fields, improving contextual discrimination for tiny and visually ambiguous objects. Experiments on the BCCD dataset show that FAM-YOLO achieves 0.934 mAP@50 and 0.658 mAP50-95, outperforming the baseline by 2.6 percentage points in mAP50- 95. More importantly, platelet mAP50-95 improves from 0.499 to 0.562, demonstrating the effectiveness of the proposed feature enhancement design for tiny platelet detection.
EC3130 15:15-15:30	Facial Skin Color Quantification of Spleen-Stomach Acupoints via Dlib Localization Author(s): Huashan Ye, Yanjing Li, Jiahao Wang and Helin Zhang Presenter: Yanjing Li, Jiangxi University of Chinese Medicine, China Abstract: Traditional Chinese medicine holds that the internal zang-fu organs are closely correlated with facial acupoints and complexion, and the skin color changes of specific facial regions can reflect the functional state of corresponding organs. Among them, facial acupoints related to the spleen and stomach are important manifestations of digestive and visceral health conditions. To realize objective and quantitative diagnosis of facial complexion, this paper takes spleen-stomach related facial acupoints as the research object. Firstly, the correlation theory of zang-fu organs and facial acupoints is summarized, and the research

	progress of Dlib facial landmark detection combined with acupoint localization is reviewed. On this basis, a landmark-acupoint mapping table is constructed to match typical acupoints corresponding to the spleen and stomach. The 30×30 ROI of target acupoints is extracted by Dlib 68-point localization, and the RGB region is converted into CIE Lab color space. K-Means clustering is adopted to extract dominant color features, and Euclidean distance is used for color matching and classification. The experimental results verify that the proposed method can effectively collect and quantify skin color features of spleen-stomach related facial acupoints, providing an objective technical means for intelligent auxiliary diagnosis of TCM facial diagnosis.
EC2045 15:30-15:45	A Multi-Component Enhanced YOLO-ERGfor Impacted third molar Keypoint Detection Author(s): Yuandong Li, Yue Chen, Qing Li, Haihua Xu and Shenghan Zhou Presenter: Yuandong Li, Beihang University, China Abstract: Aiming at the subjective variability and low efficiency of manual measurement in the clinical diagnosis of impacted third molars, as well as the insufficient accuracy of existing algorithms in keypoint detection from dental panoramic radiographs, this paper proposes a keypoint detection model for impacted third molars based on an improved YOLOv8-Pose, named YOLO-ERG. This model enables accurate and robust detection of four core anatomical keypoints for each impacted third molar and its adjacent second molar through three core improvements: designing the iEMA feature enhancement module, replacing the original detection head with RFAPoseHead, and optimizing the GIoU loss function. Experimental results based on a self-constructed panoramic radiograph dataset show that YOLO-ERG improves mAP@0.5 by 1.4% and mAP@0.5:0.95 by 2.7% compared to the vanilla YOLOv8-Pose, significantly enhancing keypoint detection accuracy and robustness while maintaining a lightweight architecture, thus providing reliable technical support for the automated quantitative diagnosis of impacted third molars.

Online Oral Session 5

- **Generative AI and AIGC Detection**
- **Session Chair**: Olarik Surinta, Mahasarakham University, Thailand
- **Time**: 13:30-15:50, July 19
- **Meeting Room**: ROOM B
- **Papers**: Invited Speech, EC1012, EC2065, EC2068, EC2085, EC2088, EC319, EC3151, EC4178

Invited Speech
13:30-13:50

Neelendra Badal
Kamla Nehru Institute of Technology (KNIT), India

Future of Visual Intelligence in the Era of Generative AI

Abstract: Visual Intelligence has emerged as one of the most advanced transformative domains of Artificial Intelligence, enabling machines to perceive, interpret, and interact with the visual world in increasingly sophisticated manners. The recent rise of Generative AI has accelerated this transformation by introducing powerful models capable of creating, understanding, and reasoning about visual content with unprecedented accuracy and creativity. In the era of Generative AI, the evolving landscape of Visual Intelligence highlighting how foundation models, vision-language architectures, and multimodal learning are redefining traditional computer vision paradigms. The discussion will focus on the transition from task-specific systems to generalized visual intelligence capable of image generation, scene understanding, visual question answering, content synthesis, and autonomous decision-making. The objective of talk will examine emerging applications across healthcare, smart cities, autonomous systems, education, agriculture, and industrial automation. This demonstrating how generative visual models are creating new opportunities for innovation and societal impact. It will also address critical challenges related to explainability, scalability, biasness, privacy, security, computational sustainability, and ethical deployment of AI-driven visual systems. The aim of the talk is to provide the insights into how Generative AI is shaping the next generation of intelligent visual systems and driving the future of digital

transformation to the researchers, academicians, industry professionals, and policymakers by presenting recent advances, future research directions, and global technological trends.	
EC1012 13:50-14:05	Adulteration Detection of Olive Oil using Generative Adversarial Network and Convolutional Neural Network Author(s): Mikhaela Joie Dingrat, Angela Noelle Diala and Noel Linsangan Presenter: Mikhaela Joie Dingrat, Mapúa University, Philippines Abstract: The study proposes an image-based classification system that distinguishes whether an olive oil sample is adulterated. The system, processed by a Raspberry Pi 5, utilizes the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) for image enhancement and the VGG16 Convolutional Neural Network (CNN) for image classification. A Raspberry Pi Camera Module 3 was used for both dataset collection and image acquisition during classification. The dataset comprises 1000 olive oil images, 500 per class, and was split at an 80:20 ratio for training and validation. An evaluation of model accuracy using a confusion matrix found the system to be 72% accurate. The results show the potential of image-based analysis for detecting olive oil adulteration and enhancing overall food safety. Future researchers may introduce changes to the system, whether hardware- or software-oriented, to create a more robust and accurate system that can classify adulterated olive oil under varying conditions.
EC2065 14:05-14:20	Degradation-Consistent Paired Training for Robust AI-Generated Image Detection Author(s): Zongyou Yang, Yinghan Hou and Xiaokun Yang Presenter: Zongyou Yang, University College London, UK Abstract: AI-generated image detectors suffer significant per- formance degradation under real-world image corruptions such as JPEG compression, Gaussian blur, and resolution down- sampling. We observe that state-of-the-art methods, including B-Free [1], treat degradation robustness as a byproduct of data augmentation rather than an explicit training objective. In this work, we propose Degradation-Consistent Paired Training (DCPT), a simple yet effective training strategy that explicitly enforces robustness through paired consistency constraints. For each training image, we construct a clean view and a degraded view, then impose two constraints: a feature consistency loss that minimizes the cosine distance between clean and degraded representations, and a prediction consistency loss based on symmetric KL divergence that aligns output distributions across views. DCPT adds zero additional parameters and zero inference overhead. Experiments on the Synthbuster benchmark (9 genera- tors, 8 degradation conditions) demonstrate that DCPT improves the degraded-condition average accuracy by 9.1 percentage points compared to an identical baseline without paired training, while sacrificing only 0.9% clean accuracy. The improvement is most pronounced under JPEG compression (+15.7% to +17.9%). Ablation further reveals that adding architectural components leads to overfitting on limited training data, confirming that train- ing objective improvement is more effective than architectural augmentation for degradation robustness.
EC2068 14:20-14:35	AIGC drives culture heritage protection and visual innovation:A case study of Hongshan culture Author(s): Qin Bo and Wang Miao Presenter: Miao Wang, Shenyang Jianzhu University, China Abstract: This study explores an innovative approach to AI-enhanced protection inheritance and the visual innovation of cultural heritage, specifically improving the audience awareness, low digital penetration, and insufficient youth awareness in promoting Hongshan Culture. This study discusses the issue of whether mass media is suitable for disseminating academic research.Therefore, taking the

	Hongshan Culture as the subject of study.A research framework has been developed, including "historical evidence-driven symbol extraction, AIGC and effectiveness evaluation." Through semiotic analysis, core cultural symbols were identified using a visual symbol sample library. Subsequently, AI multimodal generation technologies were applied to character design, scene reconstruction, and the production of short drama series. Expert evaluations and user assessments indicate that AI-generated content excels in historical accuracy, visual aesthetics, and cultural communication effectiveness, thereby significantly enhancing production efficiency and dissemination. By integrating AI-driven visual narratives with short-form drama dissemination, this study establishes a novel example for the modernized transformation of traditional cultural expression, thereby expanding audience engagement and offering practical insights into the creative revitalization of intangible cultural heritage.
EC2085 14:35-14:50	Forged Image Model Attribution Based on Hierarchical Fine-Grained Classification Author(s): Yang Hui, Liu Shiyu, Zhang Xiang, Zhang Wenbo and Zhang Wengyong Presenter: Wenyong Zhang, University of Electronic Science and Technology of China, China Abstract: With the rapid development of generative models, model attribution of forged images have become increasingly important. Existing methods are typically designed for singlelevel attribution tasks, such as model-level or architecture-level attribution, and struggle to fully leverage the relational information across different levels, thereby limiting the improvement of model-level classification performance. In this paper, we propose a forged image model attribution method based on hierarchical fine-grained classification. Specifically, after extracting the raw features, the input image is processed by two branches: the architecture-level branch extracts architectural fingerprint features, which not only enable architecture-level discrimination but also serve as conditional information for model-level attribution; guided by this conditional information, the model-level branch integrates its own model-specific fingerprint features to achieve fine-grained discrimination among different model instances within the same architecture. Experimental results demonstrate that the proposed method significantly outperforms traditional singlelevel attribution methods in model-level classification accuracy, confirming the effectiveness of incorporating architecture-level prior information to enhance feature representation capability. This method provides an effective strategy for fine-grained attribution of forged images.
EC2088 14:50-15:05	SareeTry: Culturally-Aware Saree Virtual Try-On Using Keypoint-Guided Overlay and Diffusion-Based Inpainting Author(s): Sindhu R Pai, Dr. Surabhi Narayan, Shreya K, Mrunal Kudtarkar, Sanya Sadashiva Kamath and Vishwas H N Presenter: Sanya Sadashiva Kamath, PES University, India Abstract: In India, saree shopping poses significant challenges due to diverse draping styles, intricate fabric designs, and variation in body shapes. These challenges are further amplified in online retail environments where physical try-on is not feasible. To address this gap, we present an innovative virtual try-on framework designed specifically for sarees, leveraging advanced computer vision and image synthesis techniques to simulate realistic draping experiences. The proposed approach ensures accurate garment alignment, body part visibility, and adaptability across a range of cultural draping styles and user body types. Our framework comprises of five essential modules: human detection, body keypoint estimation, fine-grained garment segmentation, geometric transformation using OpenCV-based approaches, and diffusion-based inpainting for seamless garment-user blending. It demonstrates robustness to flowy user attire, diverse body types

	and complex backgrounds, while maintaining cultural authenticity in the outputs rendered. Additionally, we curated a saree dataset enriched through resolution enhancement and stylistic variation to ensure high visual quality and inclusivity. Our proposed approach provides photorealistic overlays, which significantly enhances the virtual try-on experience. This solution bridges the gap between traditional fashion values and AI-driven innovation, offering a scalable pathway for immersive online saree shopping and virtual retail environments. Index Terms—Virtual try-on, cultural attire simulation, image segmentation, human pose estimation, body shape adaptation, diffusion inpainting, OpenCV.
EC319 15:05-15:20	DSC2F: A Multimodal AI-Generated Image Detection System Using Dual-Stream Convolutional Neural Networks With Fast Fourier Transform Author(s): Anh-Khoa Ngo-Ho, Tri-Vinh Nguyen, Anh-Khoi Ngo-Ho Presenter: Anh-Khoi NGO-HO, Nam Can Tho University (NCTU), Vietnam Abstract: The rapid development of generative models, including Generative Adversarial Networks (GANs) and diffusion-based architectures, has made AI-generated images increasingly difficult to distinguish from authentic photographs, creating significant challenges for digital forensics and media verification. This paper proposes DSC2F (Dual-Stream Convolutional Neural Network with Fast Fourier Transform), a multimodal detection framework that integrates spatial and frequency-domain information for AI-generated image detection. The proposed model consists of two parallel branches: a Spatial Stream that analyzes RGB images enhanced with 30 Steganalysis Rich Model (SRM) high-pass filters, and a Frequency Stream that processes the log-scale amplitude spectrum obtained through the Fast Fourier Transform (FFT). The two 512-dimensional feature representations are fused using a bidirectional Cross-Modal Attention mechanism and classified using a multilayer perceptron with Binary Cross-Entropy loss. Compared with the single-stream ResNet50 baseline of Wang et al., DSC2F incorporates CBAM attention modules and SE blocks while reducing the number of trainable parameters from approximately 25 million to 7.8 million. Both models are trained on the CIFAKE dataset containing 120,000 images and evaluated in a zero-shot setting on a balanced subset of 20,000 images from the ArtiFact benchmark, covering more than 13 unseen image generators. Experimental results show that all models perform near chance level, revealing a substantial generalization gap between single-generator training and real-world multi-generator detection. Nevertheless, DSC2F achieves the most balanced performance, with an overall accuracy of 67.20%, a fake-image recall (AccF) of 72.24%, and an AUC of 73.41%. Although the Wang et al. baseline achieves a higher AUC (77.46%), it suffers from severe class bias (AccF = 21.96%), making DSC2F substantially more reliable in practice. These findings demonstrate that multimodal spatial-frequency fusion improves prediction balance and fake-image recall, while robust real-world deployment requires training on more diverse multi-generator datasets.
EC3151 15:20-15:35	A Shape-Prompted Squeezer Diffusion Model for Low-Light Optical Flow Estimation Author(s): Zhixian Zhang, Haiyan Jin, Fengyuan Zuo, De Mi, Yuanlin Zhang, Bin Wang and Chen Lu Presenter: Zhixian Zhang, University of Macau, China Abstract: Mainstream optical flow estimation methods for low-light conditions primarily focus on enhancing the discriminability of low-level representations, while largely overlooking explicit noise-robust design. Although diffusion models help suppress noise interference, their effectiveness remains limited due to the absence of constraints tailored to specific features, particularly those related to object shape. Recent work on squeezed diffusion models introduces a novel noise

	injection strategy that performs anisotropic scaling along the principal directions of the data distribution, namely the principal components. Motivated by this idea, this paper proposes Diffuser-SPS, a model that exploits the noise sensitivity of diffusion models and adopts a noise-to-flow paradigm to progressively estimate optical flow in dark scenes. More importantly, Diffuser-SPS incorporates a novel Shape-Prompted Squeezer (SPS) module, which extracts salient shape priors to generate a scaling matrix and modulates the generation of enhanced noise along shape-aware representation directions. This process concentrates noise around shape boundaries while injecting less noise into texture regions, thereby achieving shape-oriented noise robustness under low-light conditions. Experimental results on the FCDN and VBOF benchmark datasets demonstrate the superiority of Diffuser SPS.
EC4178 15:35-15:50	A Sanity Check for AI-generated Underwater Image Detection Author(s): Siyi Lv and Yutao Liu Presenter: Yutao Liu, Ocean University of China, China Abstract: The rapid progress of generative models has made AI-generated images increasingly realistic, raising urgent demands for reliable forensic detectors. However, existing AIgenerated image detection studies mainly focus on general natural images, while the underwater domain remains largely unexplored despite its distinctive visual characteristics, such as color attenuation, haze, scattering, and complex lighting. In this paper, we present a sanity check for AI-generated underwater image detection. We first introduce Octopus, a dedicated benchmark for distinguishing real underwater photographs from visually deceptive AI-generated underwater images. The dataset covers diverse categories, including humans, animals, objects, and scenes, and contains high-resolution generated images that are difficult for human observers to identify. We then evaluate 13 representative off-the-shelf AI-generated image detectors on Octopus, revealing that existing methods suffer from notable performance degradation in this challenging domain. To address this problem, we propose AUIDE, an AI-generated underwater image detector with hybrid features. AUIDE combines frequencyguided patchwise noise modeling based on DCT scoring and SRM filtering with global semantic representations extracted by OpenCLIP, enabling it to exploit both low-level forensic traces and high-level visual cues. Extensive experiments show that AUIDE achieves state-of-the-art performance on Octopus, demonstrating the effectiveness of hybrid feature modeling for robust AI-generated underwater image detection.

Online Oral Session 6

➢ 3D Vision and Reconstruction

- **Session Chair:** Jinshui Huang, SouthWest Petroleum University, China
- **Time**: 13:30-15:35, July 19
- **Meeting Room**: ROOM C
- **Papers**: Invited Speech, EC2051, EC2078, EC3148, EC4185, EC4189, EC4195, EC316

Invited Speech
13:30-13:50

Tianzi Jiang
Institute of Automation, Chinese Academy of Sciences, China

Development and Evolution of the Brainnetome Atlas

Abstract: Brain atlas is an indispensable tool for studying the relationship between brain structure and cognitive function. We proposed to create a new brain atlas - the Brainnetome atlas, using brain connectivity profiles. The Brainnetome atlas lays the foundation for research in brain science and brain-inspired intelligence, and opens a new avenue not only for the study of brain science and brain diseases, but also for brain-inspired intelligence. In this lecture, we first introduce the research background and content of the Brainnetome, including the definition and the main research directions of the Brainnetome, the idea of creating the Brainnetome atlas, and the essential differences from existing brain atlas. Then, we will introduce the applications of the Brainnetome atlas in elucidating brain cognitive mechanisms and precise diagnosis and therapy of brain diseases. Then, we will present the studies on development and evolution of the Brainnetome Atlas. Finally, a summary and perspective on future research directions are provided.

EC2051 13:50-14:05	Significance-Aware Pruning and Grouped Intervention Training for Gaussian Point Management Author(s): Shengfang Pan, Yixin Zhou, Huanxin Zhu and Jinhe Su Presenter: Shengfang Pan, School of Computer Engineering, Jimei University, China

	Abstract: Existing point management methods for 3D Gaussian Splatting mainly focus on localized point densification and localized calibration, while paying limited attention to redundant Gaussian removal and optimization responsibility allocation in the middle and late stages of training. This often leads to continuous growth in the number of Gaussians and co-adaptation among them. To address this issue, this paper proposes a Gaussian point management method based on significance-aware pruning and grouped intervention training. The proposed method first performs localized point densification and localized calibration through localized multi-view error correspondence; it then constructs a significance score by jointly considering long-term visibility, opacity contribution, and gradient response, and progressively prunes low-value Gaussians. After the scene structure becomes stable, the remaining unprotected Gaussians are further divided into a normal training group and an intervention group, and the instantaneous imaging participation of part of the Gaussians is weakened through opacity perturbation, so as to promote the reallocation of error explanation responsibility and suppress co-adaptation. Experimental results show that, under 22K iterations, the proposed method maintains reconstruction quality comparable to GS-LPM while reducing the average training time by 34.3% and the number of Gaussians by 12.8%, achieving a better efficiency-quality trade-off.
EC2078 14:05-14:20	3D Reconstruction Model Based on Generative Radiance Fields of Self-Supervised Learning Author(s): Xiaomei Xue, Yachen Wang, Hongbing Xiao, Yu Wang, Peiyuan Guo and Man Bao Presenter: Xiaomei Xue, Beijing Technology & Business University, China Abstract: Generative Radiance Fields (GRAF) is a representative unsupervised framework for volume-rendering-based 3D reconstruction. However, the limited availability of medical datasets and the difficulty of obtaining annotations often constrain the discriminator's capacity to learn features, thereby leading to insufficient reconstruction detail. To address this issue, the discriminator of GRAF is replaced with a self-supervised learning method, which captures richer representations by mining the intrinsic structure of the data. Experimental results demonstrate substantial performance improvements. For knee reconstruction, the proposed model reduces the Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) by 7.52% and 13.79%, respectively, while increasing the Structural Similarity Index (SSIM) by 7.51% and the PSNR by 3.89 dB. For chest reconstruction, FID and KID decrease by 13.19% and 16.98%, respectively, whereas SSIM and PSNR improve by 14.36% and 3.38 dB. These results testify that the self-supervised GRAF model effectively improves the quality and fine-detail fidelity of 3D medical reconstruction.
EC3148 14:20-14:35	Approaches to Sparse Tensorial Radiance Fields: An Empirical Study Author(s): Dhruthan M N, Vishwathma Bhat and Prafullata Kiran Auradkar Presenter: Vishwathma Bhat, PES University, India Abstract: Neural radiance fields have transformed novel-view synthesis, yet their dense evaluation of every basis component at every spatial sample remains a computational bottleneck. Mixture-of-experts routing and structured weight sparsity are well-established acceleration techniques in transformer architectures, but their applicability to radiance-field rendering has not been systematically evaluated. In this work we conduct an empirical study of five distinct sparsity strategies applied to TensoRF, a state-of-the-art tensorial radiance-field model. Our designs span two fundamental paradigms: joint training, where sparsity heads are optimised alongside the radiance-field grid from the first iteration, and post-convergence distillation, where a fully trained dense model is augmented with lightweight routing modules in a brief second

	phase. Index Terms—Neural Radiance Fields, Sparse Activation, Mix- ture of Experts, Model Compression, Tensorial Decomposition
EC4185 14:35-14:50	Multi-Sensor Fusion SLAM Mapping Method Based on ROS for Warehouse Environments Author(s): Jiayu Hu, Chuxin Cao, Cheng Qin and Yue Gao Presenter: Jiayu Hu, Soochow University, China Abstract: Traditional AGVs relying on pre-set physical markers can hardly meet the flexible scheduling requirements of intelligent warehouses, while single-sensor SLAM suffers severe perceptual degradation in complex warehouse environments: 2D LiDAR has 3D blind zones and geometric degradation in long corridors, and vision sensors are extremely sensitive to illumination. This paper proposes a low-cost multi-sensor fusion SLAM mapping system for warehouse environments. A single-line LiDAR and a depth camera are integrated based on the RTAB-Map framework, adopting a fusion strategy where LiDAR provides hard planar geometric constraints and vision delivers semantic and loop closure constraints. Spatiotemporal alignment is realized through camera intrinsic parameter calibration and an approximate time synchronization mechanism, and 3-degree-of-freedom kinematic constraints are introduced to eliminate Z-axis drift. Ablation experiments covering low obstacles and low-light traps are designed. The results show that the closed-loop drift error of the fusion scheme is only 0.1059 m, with the accuracy improved by 45.44% compared with the pure LiDAR scheme. The system can successfully detect low obstacles under low-light conditions, which verifies its robustness and complete perception capability.
EC4189 14:50-15:05	BARD-NeuS: Bilinear Affine Residual Neural Implicit Surface Learning for Multi-view Reconstruction Based on Depth Constraint Author(s): Bang Tang, Sen Wang, Zehua Gu and Wenju Wang Presenter: Wenju Wang, University of Shanghai for Science and Technology, China Abstract: Recently, neural implicit surface reconstruction methods combined with volume rendering, such as NeuS, have made significant progress. However, these methods still face some challenges. Existing approaches have difficulty reconstructing the detailed contours of object surfaces, and due to the lack of geometric supervision, the algorithms exhibit poor robustness in uncontrolled environments. To address these issues, this paper proposes a new neural implicit surface reconstruction method called BARD-NeuS. The algorithm first introduces residual connections and normalization techniques into the multilayer perceptron, enabling it to effectively capture fine details in 3D scenes. Furthermore, the algorithm incorporates depth constraints into the loss function to further enhance its robustness. Experimental results demonstrate that the surfaces reconstructed by BARD-NeuS exhibit richer and more detailed contours compared to existing methods, and the reconstruction quality in uncontrolled environments is also superior to that of existing techniques. We also provide our source code, which is available at: https://github.com/Unicodern/BARD-NeuS.
EC4195 15:05-15:20	Audio2Gauss: Structured Audiovisual Modulation with Explicit 3D Scene Representations Author(s): Xuehao Shi, Lihong Luo and Cuiying Li Presenter: Xuehao Shi, Guangdong University of Technology, China Abstract: Music visualization remains confined to 2D screen-space effects despite decades of advances in photorealistic 3D rendering. We present Audio2Gauss, a learnable framework that maps audio features to 3D Gaussian Splatting (3DGS) parameters, animating static scenes in response to music without paired training data. The key insight is spatial-cluster-conditioned modulation: K-Means clustering on Gaussian positions binds audio-driven parameter offsets to explicit

	3D spatial structure, transforming music response from a screen-space effect into structured scene-space modulation. A lightweight temporal MLP predicts per-cluster parameter deltas from audio context and cluster statistics, trained end-to-end through the 3DGS differentiable rasterizer with three complementary structured losses. Experiments across two scenes and three music genres with four baseline levels demonstrate that learned mapping produces significantly more audio-correlated motion than the best handcrafted rules, with a large effect size and statistically significant improvement, while maintaining comparable beat-alignment quality. An AB blind user study confirms that viewers show a clear and statistically significant preference for Learned over Rule-based mapping, driven by visual appeal and moderated by music genre. Our work presents the first learnable audio-to-3DGS mapping framework and demonstrates that explicit 3D scene structure enables qualitatively new forms of audiovisual modulation---forms that human viewers consistently prefer.
EC316 15:20-15:35	Joint-Embedding Predictive Pretraining for Energy-Calibrated Skin Cancer Screening Author(s): Quyen Van Vo, An Cong Tran, Hiep Xuan Huynh Presenter: Quyen Van Vo, Can Tho University of Medicine and Pharmacy, Vietnam Abstract: Automated skin lesion classification from dermoscopic images offers a promising pathway to earlier cancer detection in resource-limited healthcare settings, yet two fundamental challenges remain inadequately addressed: the scarcity of labelled dermoscopic training data and the lack of well-calibrated uncertainty estimates in standard softmax-based classifiers. A two-stage framework is proposed in this study that combines Image-based Joint-Embedding Predictive Architecture (I-JEPA) self-supervised pretraining with Energy-Based Model (EBM) fine-tuning. In Stage 1, a Vision Transformer encoder is pretrained on the unlabelled portion of the ISIC Archive (77,616 records) through a non-contrastive predictive objective that operates entirely in latent embedding space, avoiding hand-crafted augmentations and negative pairs. In Stage 2, the pretrained encoder is transferred to an EBM head fine-tuned on 22,225 labelled ISIC records under a focal-weighted EBM loss, prioritising sensitivity on malignant minority classes. A clinically motivated seven-class taxonomy is constructed from 32 raw ISIC diagnostic categories based on malignancy risk and treatment pathway. Evaluated through five-fold stratified cross-validation, a macro-AUC of 0.701, Macro-F1 of 0.219, and Macro-Sensitivity of 0.278 are achieved by the proposed I-JEPA+EBM on structured metadata features, consistently outperforming the EBM without I-JEPA pretraining and establishing a quantified baseline for image-level deployment. Clinical false-negative rates are reported per malignant class to inform development priorities: Melanoma FNR 0.854, BCC FNR 0.704, and SCC/AK FNR 0.467 identify the gains required from full dermoscopic image pretraining.

Online Oral Session 7

- **Multimodal Learning and Video Understanding**
- **Session Chair**: Siyuan He, Nanfang College Guangzhou, China
- **Time**: 16:30-19:05, July 19
- **Meeting Room**: ROOM A
- **Papers**: Invited Speech, EC2034, EC3143, EC438, EC1025, EC3128, EC4172, EC2027, EC429, EC439

Invited Speech
16:30-16:50

Ahmad Ali
Hainan Bielefeld University of Applied Sciences, China

Exploiting Attention-Driven Weather-Aware Multimodal Spatio-Temporal Fusion for Urban Traffic Flow Prediction

Abstract: Accurate urban traffic flow prediction is an important task for intelligent transportation systems, smart mobility management, and urban congestion control. However, traffic flow is influenced by complex spatial dependencies, temporal variations, and external factors such as weather conditions. This talk will present an attention-driven weather-aware multimodal spatio-temporal fusion framework for urban traffic flow prediction. The proposed approach integrates traffic flow patterns with weather and contextual information to improve prediction accuracy under dynamic urban conditions. By using attention mechanisms, the model can adaptively capture important spatial, temporal, and weather-related features, leading to more robust and reliable traffic forecasting. The talk will also discuss the importance of multimodal data fusion in intelligent transportation systems and its potential applications in smart city traffic management.

EC2034 16:50-17:05	LAGNet: Language-Aware and Spatial Focusing for Cross-View Object-Level Geo-Localization Author(s): Chenzheng Zhou, Guoan Chen, Yujun Zhang and Shengke Wang Presenter: Chenzheng Zhou, Ocean University of China, China Abstract: Cross-view object geo-localization (CVOGL) offers fine-grained object-level geographic positioning beyond traditional image-level localization. However,

	it faces significant challenges from extreme viewpoint variations and spatial ambiguity among homogeneous targets, particularly in marine environments. We propose a Language-Aware Geo-Localization Network (LAGNet) to address these issues. First, a click-aware spatial feature encoding module incorporates Gaussian heatmap and coordinate priors for spatial robustness. Second, a heatmap-guided sparse cross-attention module (H-SCAN) aggregates multi-scale context with log-bias heatmap priors and dual-pathway top- K denoising to precisely localize targets amidst distractors. Third, a RemoteCLIP-guided semantic fusion module (RCGF) efficiently adapts RemoteCLIP via lightweight adapters, performing semantic cross-attention to bridge visual-spatial and high-order semantic representations. The framework is optimized end-to-end with anchor-adaptive YOLO regression loss. Extensive experiments on the standard CVOGL benchmark demonstrate LAGNet's superiority, achieving 65.76% acc@0.5 on Drone→Satellite (outperforming prior SOTA by 1.32%) and 47.39% on Ground→Satellite (gain of 0.63%), showing remarkable robustness in complex marine scenarios.
EC3143 17:05-17:20	Time-frequency Network for Robust Speaker Recognition Author(s): Jiguo Li, Weixuan Zhang, Xiaobin Liu, Wenbin Zhu and Jing Yuan Presenter: Xiaobin Liu, Nankai University, China Abstract: The wide deployment of speech-based biometric systems usually demands high-performance speaker recognition algorithms. However, most of the prior works for speaker recognition either process the speech in the frequency domain or time domain, which may produce suboptimal results because both time and frequency domains are important for speaker recognition. In this paper, we attempt to analyze the speech signal in both time and frequency domains and propose the timefrequency network (TFN) for speaker recognition by extracting and fusing the features in the two domains. Based on the recent advance of deep neural networks, we propose a convolution neural network to encode the raw speech waveform and the frequency spectrum into domain-specific features, which are then fused and transformed into a classification feature space for speaker recognition. Experimental results on the publicly available datasets TIMIT and LibriSpeech show that our framework is effective to combine the information in the two domains and outperforms state-of-the-art methods for speaker recognition.
EC438 17:20-17:35	V-JEPA-Drive: Uncertainty-Aware Self-Supervised Predictive Learning for Multi-Camera Autonomous Driving Perception Author(s): Phuong Bich Thi Nguyen, Tung Thanh Nguyen, Hiep Xuan Huynh Presenter: Phuong Bich Thi Nguyen, Vinh Long University of Technology Education and Dong Thap University, Vietnam Abstract: Scaling camera-basedautonomous drivingperceptionrequiresbreak Hiep Xuan Huynh∗ hxhiep@ctu.edu.vn Can Tho University CTU-AIMED Leading Research Team Can Tho, Vietnam for Multi-Camera Autonomous Driving Perception. In Proceedings of 10th International Conference on Deep Learning Technologies (ICDLT '26). ACM, ing the dependence on large annotated corpora, yet most self supervised alternatives either waste capacity on pixel-level recon struction or leave calibration entirely to post-hoc procedures. We address both problems at once. V-JEPA-Drive trains a multi-camera BEV encoder to predict the representations of masked scene regions rather than their raw pixels, and it does so under a loss that si multaneously penalises prediction error and encourages calibrated epistemic uncertainty. Six synchronised surround-view cameras are fused into a shared bird's-eye-view grid; an adaptive entropy based masking strategy concentrates the self-supervised task on high-information cells such as intersections and dense traffic clus ters. On nuScenes, V-JEPA-Drive achieves 55.2% NDS and 44.8% mAP for 3D detection and 54.6% mIoU for BEV

	segmentation with full labels, matching fully supervised StreamPETR while using no labels during pre-training. At 40% of annotations, detection perfor mance equals supervised BEVFormer at 100%—a 60% annotation saving. Critically, the epistemic uncertainty output by the model rises by 99% when label budget drops from 100% to 10%, providing a calibrated, cost-free signal for guiding active annotation.
EC1025 17:35-17:50	WAFC: Weather-Adaptive Feature Calibration for Robust Fine-Grained Classification Author(s): Junxiu Zhou and Yangyang Tao Presenter: Junxiu Zhou, Northern Kentucky University, USA Abstract: Robust image classification under adverse weather is a prerequisite for reliable AI deployment in unconstrained environments. While recent research has pivoted heavily toward object detection, we argue that the underlying classification backbone remains the primary point of failure due to weather-induced semantic erosion. We propose a Weather Adaptive Feature Calibration (WAFC) framework, a lightweight and plug-and-play module that dynamically recalibrates feature responses conditioned on degradation-aware statistics. Combined with dual-domain supervision and clean–corrupted feature consistency regularization, WAFC improves robustness while maintaining competitive clean accuracy. Extensive experiments on CIFAR-100 and CIFAR-100-C demonstrate consistent robustness gains with negligible computational overhead.
EC3128 17:50-18:05	An Integrated Smart Library System: Intelligent Syllabus-Driven Book Recommendation and AI-Based Real-Time Seat Occupancy Detection Author(s): Sanket Muttur, Abhishek Yelmamdi, Omkar Hiremath, Prafullata Auradkar and Subhash Reddy B Presenter: Abhishek Yelmamdi, PES University, India Abstract: University libraries face two major problems when they need to connect course materials with appropriate reading materials and find study areas for students during busy times. The document introduces an intelligent library system that resolves these two problems through its two interconnected systems. The first subsystem is a book recommendation engine that analyzes syllabus content using OCR and intelligent text parsing to generate relevant book suggestions. The system determines library connections through fuzzy matching based on the user's academic department and current semester and selected domain. The engineering course testing process achieved approximately 95 percent OCR accuracy and 93 percent overall parsing precision while matching 56.8 percent of cases with 97.6 percent precision. The second subsystem utilizes existing CCTV cameras to create an AI-based computer vision system that detects seat occupancy. The system uses YOLO for human detection and homography technology to create spatial maps that display seat availability through an interactive floor map which needs no extra hardware and does not store video footage. The system provides room booking functionality through its online system. The subsystems improve academic resource access while creating better space use through their ability to measure time and manually check results which makes them suitable for smart campus environments.
EC4172 18:05-18:20	Performance Evaluation of LoRA-Fine-Tuned Vision-Language Models for Elderly Care Behavior Understanding in Edge Deployment Scenarios Author(s): Yueyang Yu, Kexun Yan, Cheng Zhang and Yuhang Yang Presenter: Yueyang Yu, Chengdu Neusoft University, China Abstract: Privacy-sensitive elderly care video monitoring requires vision-language models to perform visual behavior understanding locally on edge devices.

	However, the trade-off between lightweight deployment and task-specific visual perception performance remains insufficiently evaluated. This study constructs a dedicated dataset of 1,500 video-text pairs for caregiving visual interaction scenarios and systematically evaluates Qwen, MiniCPM, and InternVL series models under three settings: non-fine-tuned local models, LoRA-fine-tuned lightweight models, and large-scale API baselines. Using a unified LlamaFactory framework, six lightweight multimodal models are fine-tuned with LoRA and tested on an independent set of 200 samples. The evaluation covers generation quality, semantic fidelity, factual faithfulness, output length, and inference efficiency. Experimental results show that LoRA fine-tuning brings significant but non-uniform effects across lightweight models. Qwen2.5-VL-3B after fine-tuning surpasses the non-fine-tuned 72B model within the same lineage in several metrics, but its hallucination rate also increases, while Qwen3-VL-2B exhibits negative transfer across multiple metrics. The results further indicate that model lineage strongly affects hallucination control, and that longer outputs do not necessarily lead to higher factual coverage. Overall, elderly care behavior understanding presents clear multi-objective trade-offs among generation quality, semantic fidelity, factual faithfulness, output density, and efficiency. This work provides quantitative evidence for multimodal model selection and LoRA fine-tuning practice in resource-constrained and privacy-sensitive edge deployment scenarios.
EC2027 18:20-18:35	StreamScreen: Hierarchical Memory-Augmented Streaming Video Understanding for Mobile Screen Question Answering Author(s): Yang Song Presenter: Yang Song, Beijing OPPO Telecommunications Corp., Ltd. China Abstract: Streaming video question answering (VideoQA) systems that operate over extended temporal horizons face a fundamental challenge: how to efficiently organize and retrieve relevant visual-textual information from continuously growing memory stores. Existing approaches typically compress video streams into flat textual descriptions or store uniformly segmented clips, resulting in significant visual information loss and poor temporal awareness. In this paper, we propose TEFA-VQA (Training-Free Event-Aware Video Question Answering), a novel framework that introduces a completely training-free algorithm for localizing event boundaries in streaming video, constructing event-structured memory representations, and performing targeted retrieval for question answering. Our approach formulates event boundary detection as a statistical change-point detection problem in multimodal feature space, leveraging kernel Maximum Mean Discrepancy (MMD) to identify distributional shifts without requiring any labeled training data. Detected event boundaries are used to construct a hierarchical memory architecture that indexes video content at both the segment and event levels, enabling query-adaptive retrieval that balances semantic relevance with temporal coherence. We evaluate TEFA-VQA on a comprehensive streaming video QA benchmark comprising 2,500 question-answer pairs across five distinct categories. Experimental results demonstrate consistent improvements over the baseline system, with particularly notable gains in retrospective QA (+2.00 points) and robustness to irrelevant queries (+6.28 points), validating the effectiveness of training-free event-aware memory organization for streaming VideoQA.
EC429 18:35-18:50	VL-JEPA World Model with Energy-Based Scoring for Feeding Behavior Detection in Pangasianodon hypophthalmus Author(s): Huy Hoang Le Nguyen, Tan Thien Thai, Khang Hoang Duong, Phu Minh Tran, Hiep Xuan Huynh Presenter: Huy Hoang Le Nguyen, Can Tho University, Vietnam

	Abstract: Automated detection of feeding behavior in intensively farmed striped catfish (Pangasianodon hypophthalmus) is a prerequisite for data-driven feed management in Mekong-Delta aquaculture, where feed represents up to 75% of production cost and feeding is still assessed manually. The task is formulated as a world-model problem: rather than classifying isolated clips, the system learns a compact latent model of how the pond surface evolves over time. CTU-CatraWorld is proposed, a Vision-Language Joint-Embedding Predictive Architecture (VL-JEPA) whose latent dynamics are scored by an Energy-Based Model (EBM) head performing joint state classification and overfeeding-anomaly detection. Beyond the model, we contribute a rigorous, measured empirical study on a 45-video field corpus that isolates the core difficulty of the domain—the adversarial pair of active feeding (S1) versus overfeeding / residual feed (S2⋆), which are visually near-identical but demand opposite dispenser actions. We show that a single pellet-appearance cue cannot separate the pair (macro-F1 = 0.46), that raw optical-flow motion is confounded by hand-held camera egomotion (macro-F1 = 0.49), and that a frozen ImageNet ViT-B/16 backbone—the CTU-CatraWorld encoder prior to self-supervised adaptation— already reaches macro-F1 = 0.82 and AUROC = 0.92 on this pair, extending to macro-F1 = 0.67 over three states. The CTU-CatraWorld energy head trained on these features classifies the pair at macro-F1 = 0.76, and a latent-space JEPA adaptation of the features raises the pair to AUROC 0.98 (linear probe) and 0.96 (energy head)—a measured preview of the full objective. These measured results justify the architectural choices and quantify the separability of the state pairs directly on field video. Completion of the full selfsupervised VL-JEPA + EBM pipeline is identified as future work.
EC439 18:50-19:05	TagJEPA: A Joint-Embedding Predictive Architecture for Energy-Based Tag-Aware Recommendation Author(s): Dao Xuan Thi Nguyen, Khoi Tan Nguyen, Hiep Xuan Huynh Presenter: Dao Xuan Thi Nguyen, University of Economics Ho Chi Minh City (UEH), Vietnam Abstract: Building autonomous intelligent systems that can reason about user preferences without explicit supervision has been recognized as a central problem at the intersection of selfsupervised learning and personalized artificial intelligence. Tagaware recommendation is a demanding testbed for such systems: tagging matrices are typically over 99% sparse, cold-start is the rule rather than the exception, and a single user state is compatible with many plausible items. This work draws on the JointEmbedding Predictive Architecture (JEPA) paradigm of LeCun and its image-based realization I-JEPA to design TagJEPA, which, to the best of the authors' knowledge, is the first JEPA dedicated to tag-aware recommendation. A user's tag history is mapped through a context encoder into a latent representation after random tag-masking; an EMA-updated target encoder maps a candidate item to a target representation; and a predictor, conditioned on a latent variable summarizing the masked tag context, predicts the target from the user representation. The predictive energy is minimized jointly with a VICReg variance– invariance–covariance regularizer so that representation collapse is intrinsically prevented and no negative samples are required. At inference time, items are ranked by a fused score in which JEPA latent compatibility is combined with a tag-cosine signal. On three benchmarks spanning sparse, medium, and dense interaction regimes, TagJEPA achieves competitive or leading cold-start performance among all methods evaluated and consistently outperforms interaction-heavy baselines in low-interaction scenarios. A theoretical bridge is further established between TagJEPA's predictive energy and the empirical energy distance of Szekely–Rizzo, positioning the proposed architecture within ´ both energy-statistics and autonomous-machine-intelligence literatures.

Online Oral Session 8

➢ Agricultural Remote Sensing and Environmental Safety

- **Session Chair:** Hong Mo, Kunming University of Science and Technology, China
- **Time**: 16:30-19:00, July 19
- **Meeting Room**: ROOM B
- **Papers**: EC1006, EC1007, EC3001, EC3150, EC4169, EC4171, EC4176, EC4182, EC4199, EC2058

EC1006 16:30-16:45	Sap Shots: Biomass–Reflectance Dynamics Across Phenological Stages in Rubber (Hevea brasiliensis) Author(s): Roxy Blue De Jesus and Fillmore Masancay Presenter: Fillmore D. Masancay, University of Southeastern Philippines, Philippines Abstract: While optical satellites are useful for monitoring rubber (Hevea brasiliensis) plantations, their performance consistency varies seasonally due to rapid changes in canopy density, particularly within fragmented smallholder systems where inconsistencies can render year-round data unreliable. As such, this study investigates the biomass-reflectance dynamics of a smallholder rubber plantation in Makilala, Cotabato, Philippines, to examine how phenological timing influences the reliability of year-round optical satellite monitoring. Using Sentinel-2 imagery and in-situ biometrics from 50 sample trees, the study analyzed the correlation between Above-Ground Biomass (AGB) and spectral indices (NDVI, NDRE, NDMI) across three distinct phenological windows (Onset, Transition, and Peak Vigor), cross-referenced with NASA POWER meteorological data. Results indicate that monitoring feasibility is strongly conditional on structural maturity; reliable correlations were confined to more mature trees (>75 cm girth), while younger trees failed due to background noise. Significantly, the study refutes the efficacy of a static annual index, identifying a "spectral inversion" during the Transition phase (rs = -0.540) where high-biomass trees exhibit reduced greenness due to delayed refoliation. As a solution based on the findings, an exploratory conceptual "Operational Matrix" is proposed: utilizing NDMI during Onset, NDVI during Transition, and NDRE during Peak Vigor to mitigate saturation. This phenologically synchronized approach suggests the potential for a low-cost, seasonally adaptive conceptual monitoring strategy, subject to further multi-site validation, for the sustainable management of inherently fragmented smallholder rubber plots.
EC1007 16:45-17:00	Automated Cassava Yield Estimation through UAV-derived Vegetation Indices and Machine Learning Author(s): Joyce Kristie Fernandez, Fillmore Masancay, Florelyn Robles and Jbel Raf Espina Presenter: Fillmore D. Masancay, University of Southeastern Philippines, Philippines Abstract: This study aims to develop an automated cassava proximal yield estimation through UAV-Derived Vegetation Indices and Machine Learning techniques. Data were gathered from a 0.5-hectare cassava farm in Tamugan, Davao City using an Autel EVO II drone, which captured high-resolution RGB images at the mid-growth stage. Ground-truth measurements of plant height and

	stem diameter were also recorded to serve as reference data. The UAV images were processed in Agisoft Software to produce orthomosaics, Digital Surface Models (DSMs), and vegetation indices. These indices, along with the field data, were then used to train a Random Forest machine learning model to predict cassava above-ground biomass and yield-related indicators. The results demonstrate that UAV-derived vegetation indices, when combined with machine learning, can effectively monitor crop growth and provide accurate yield predictions. This system has the potential to aid farmers and agricultural stakeholders in harvest planning, resource management, and decision-making, contributing to the growth of digital agriculture and sustainable cassava production in the Davao Region.
EC3001 17:00-17:15	Lightweight Scene Classification Method for Off-Nadir Images on Unmanned Platforms Author(s): Siyan Zhou and Xuebo Zhang Presenter: Lin Zhang, Xi'an University of Architecture and Technology, China Abstract: Addressing the limitation of the ORB-SLAM3 system, which can only generate sparse maps and fails to meet the geometric information requirements of tasks such as robot navigation and obstacle avoidance, this paper proposes an extended dense mapping method. By incorporating an independent dense mapping thread into the original ORB-SLAM3 framework, the proposed method integrates three functionalities: dense mapping, voxel filtering, and octree mapping. First, dense point clouds of the scene are obtained using RGB-D images from keyframes. Second, voxel filtering is introduced to downsample the redundant point clouds, thereby reducing the data volume. Finally, the processed point clouds are efficiently converted into an octree map structure to achieve hierarchical spatial management. Experimental results demonstrate that the proposed method effectively reconstructs the 3D structure of the scene while significantly reducing the memory footprint of the map. This alleviates the storage challenges associated with dense mapping in large-scale scenarios and provides a feasible solution for real-time localization and high-precision mapping of mobile robots in resource-constrained environments.
EC3150 17:15-17:30	Raspberry Pi-Based Image Processing System for the Detection and Treatment Recommendation of Corn Leaf Diseases Author(s): Allana Kyle Deang and Alvin Dave Galasinao Presenter: Alvin Dave S. Galasinao, Mapua University, Philippines Abstract: This paper presents the design and implementation of a Raspberry Pi-based agricultural expert system for the detection and treatment recommendation of corn leaf diseases in the Philippines. The system integrates a ResNet-50 Convolutional Neural Network (CNN) for image-based disease classification and a Mamdani-type Fuzzy Inference System (FIS) for adaptive treatment recommendations based on environmental conditions. The system classifies corn leaves into five categories: Healthy, Northern Corn Leaf Blight (NCLB), Southern Corn Leaf Blight (SCLB), Downy Mildew (DM), and Gray Leaf Spot (GLS). A dataset of 600 annotated corn leaf images was used for model development, consisting of training, validation, and testing subsets. To assess real-world performance, the deployed system was further evaluated using 500 independently collected and expert-validated corn leaf samples, along with 100 additional field validation samples captured under varying environmental conditions. Environmental parameters such as temperature, humidity, and soil moisture were processed by the FIS to generate context-aware treatment recommendations. The hardware prototype consists of a Raspberry Pi 5, camera module, environmental sensors, and an LCD-based graphical user interface. Results showed that the deployed system achieved classification accuracies of

	87% for Healthy leaves, 90% for Downy Mildew, 92% for Gray Leaf Spot, 97% for Northern Corn Leaf Blight, and 92% for Southern Corn Leaf Blight on the 500-sample evaluation dataset. Furthermore, the system achieved an overall classification accuracy of 95% during independent random field testing using 100 corn leaf samples collected under varying environmental conditions. Expert validation confirmed complete agreement between the generated treatment recommendations and established corn disease management guidelines. The findings demonstrate the potential of the proposed system as a portable and intelligent tool for real-time corn disease detection and precision agriculture applications.
EC4169 17:30-17:45	Dynamic Frequency-Domain Spectral Selection for Robust Infrared-Visible Object Detection Author(s): Huilin Yang and Jing Zhang Presenter: Huilin Yang, TIANGONG University, China Abstract: Infrared-visible dual-modal object detection enhances environmental perception by coupling thermal radiation with rich texture details. Nevertheless, traditional spatial-domain fusion paradigms often suffer from static modality weight allocation and limited receptive fields, while neglecting discriminative frequency-domain traits. To overcome these challenges, we propose an end-to-end multi-modal detection framework centered on dynamic frequency-domain spectral selection. At its core, a Frequency-guided Dual-domain Selection Module (FDSM) is developed to adaptively modulate input-dependent spectral components, capturing the intrinsic frequency complementarity of heterogeneous modalities. Concurrently, a Cross-Modality Interactive Fusion (CMIF) module leverages a bidirectional feature gating mechanism to suppress conflicting cross-modal noise while reinforcing task-relevant target clues. Furthermore, a directional strip attention mechanism is seamlessly integrated into the backbone to efficiently expand the receptive field and model long-range context. Benchmark evaluations on the M3FD and LLVIP datasets demonstrate that our method surpasses existing state-of-the-art detectors, establishing an optimal tradeoff between detection accuracy and environmental robustness.
EC4171 17:45-18:00	A Density Inversion Method for High-Energy Flash Radiography Based on Compressed Sensing Theory Author(s): Bo Jin, Cunbo Lu, Hang Li, Ao Sun and Yi Mao Presenter: Bo Jin, Hohai University, China Abstract: Aiming at the density inversion problem of non-axisymmetric objects with limited projection data in high-energy flash radiography, and addressing the drawback that existing total variation (TV) algorithms based on compressed sensing only take into account the local similarity of images while ignoring their non-local similarity, this paper proposes a total variation reconstruction method combined with group sparse regularization (TV-GSR). By embedding the group sparse model into the TV framework, the proposed method simultaneously exploits the local similarity and non-local self-similarity of object images, and fully utilizes the prior sparse information of images. Meanwhile, the four-way symmetry (up, down, left and right) of the object is adopted to reduce the scale of image reconstruction, which improves the reconstruction accuracy and accelerates the reconstruction speed. Simulation results demonstrate that the proposed TV-GSR algorithm achieves higher reconstruction accuracy for both noise-free and noisy images. It performs well on high-energy flash images and computed tomography (CT) images with abundant texture details, and possesses strong generality.
EC4176 18:00-18:15	UAV Object Detection with High-Frequency Augmentation and Adaptive Multi-Scale Fusion

	Author(s): Guang Zhu Presenter: Guang Zhu, China Jiliang University, China Abstract: Aerial images captured by UAVs are often characterized by wide field-of-views and cluttered backgrounds, rendering target objects small, densely packed, and blurry. For these low signal-to-noise ratio (SNR) targets, conventional detectors suffer from severe fine-grained feature loss during downsampling, leading to high miss rates. To address this, we propose HAFD-YOLO, a lightweight detector built upon the YOLO11s baseline. Our architecture introduces four key innovations. First, the neck network is streamlined by eliminating the redundant P5 layer and integrating a fine-grained P2 layer to preserve high-resolution spatial features. Second, a High-Frequency Feature Enhancement Block (HFEB) is embedded to sharpen blurred edges via frequency-domain decomposition. Third, a Modified DySample upsampler is designed with a static constraint factor to mitigate aliasing distortion. Finally, we propose a Refined Adaptive Spatial Feature Fusion (RASFF) mechanism that executes structural pruning on dense convolutional layers, enabling lossless cross-scale feature aggregation. Evaluated on the VisDrone2019 dataset, HAFD-YOLO reduces the parameter count by 40.4% while improving mAP50 and mAP50-95 by 11.1 and 6.1 percentage points, respectively.
EC4182 18:15-18:30	Lightweight Semantic Context Prior FPN for Tiny Object Detection in Aerial Images Author(s): Zhicheng Wang, Jiayu Xu and Junbiao Pang Presenter: Zhicheng Wang, Beijing University of Technology, China Abstract: Tiny object detection in aerial images is difficult because object evidence is severely weakened after repeated convolutional downsampling, while many background textures produce high-frequency responses similar to tiny targets. This paper presents Semantic Context Prior FPN (SCP-FPN), a lightweight extension built on HS-FPN [12]. SCP-FPN injects teacher-derived coarse semantic context into the high-frequency pyramid pathway without adding the segmentation teacher at inference. A public SegFormer-B5 model trained on ADE20K is used only offline to generate trainval pseudo maps, which are mapped to an 8-channel coarse semantic space. During detector training, a compact SCP student branch learns this prior through auxiliary supervision and fuses it with high-frequency features by local residual cross attention. During validation and testing, only the detector and the lightweight SCP branch are retained. On AITOD with Cascade R-CNN and ResNet-50, SCP-FPN improves the HS-FPN baseline from 21.0 to 21.5 AP, from 47.4 to 48.4 AP50, from 15.2 to 15.6 AP75, and from 10.0 to 11.0 APvt. The improvement is achieved with only 1.64% additional FLOPs and 0.33% additional parameters, showing that coarse semantic priors provide a lightweight accuracy-efficiency trade-off for tiny object detection.
EC4199 18:30-18:45	LANDSCAPE ANALYSIS OF HUMAN ACTIVITIES AND VEGETATION LOSS ON WATER QUALITY IN TALOMO–LIPADAS WATERSHEDS, DAVAO CITY Author(s): 1. Allen Jhon Bautista 2. Adrian Chris Cagampang 3. Julius Lance B. Españo 4. Kevin G. Cañada Presenter: Kevin Cañada, University of Southeastern Philippines, Philippines Abstract: This study analyzes the effects of human activities and vegetation loss on water quality in the Talomo and Lipadas Watersheds in Davao City. Geographic Information System (GIS), Land Use and Land Cover (LULC) analysis, Normalized Difference Vegetation Index (NDVI), Spearman correlation, and Holt–Winters forecasting were used to examine data from 2020 to 2025. Results showed that Lipadas had stable and dense vegetation, where NDVI had significant relationships with Dissolved Oxygen, Chloride, pH, and Fecal Coliform. This suggests that

	vegetation helps support water quality and ecological stability. In Talomo, NDVI showed no significant relationships with water quality because urbanization, agriculture, and logging had stronger effects on the watershed. Built-up areas in both watersheds were associated with higher Nitrate, Phosphate, Biochemical Oxygen Demand, and Color, showing the effects of human activities and runoff. Holt–Winters forecasting was reliable for stable parameters such as pH, Temperature, and Dissolved Oxygen, but less reliable for highly changing parameters such as Fecal Coliform, Chloride, Nitrate, and Total Suspended Solids. The study recommends reforestation, riparian protection, waste management, runoff control, and continuous water quality monitoring to help protect the watersheds and support water security in Davao City.
EC2058 18:45-19:00	Lightweight Multi-Scale Attention Network Based on MobileNetV3 for Tomato Leaf Disease Classification Author(s): Xiaoran Zhang, Yunzheng Li and Zhenxue He Presenter: Xiaoran Zhang, Hebei Agricultural University, China Abstract: Automated classification of tomato leaf diseases is essential for reducing crop losses, yet deploying deep models on resource-constrained agricultural devices remains a challenge. Existing lightweight convolutional neural networks typically process features at a single scale, which limits their capacity to capture lesions ranging from small scattered spots to large deformed areas. We propose MSCA-MobileNet, a lightweight network built on MobileNetV3-Small that introduces two complementary modules: (i) Convolutional Block Attention Modules (CBAM) inserted after backbone stages 2, 3, and 4 for hierarchical channel–spatial attention, and (ii) a Cross-layer Multi-scale Feature Fusion (CMFF) module that aligns, upsamples, concatenates, and compresses features from three depths via 1×1 convolutions. On the PlantVillage tomato subset (10 classes, 18,160 images), MSCA-MobileNet achieves 99.78% accuracy with only 1.10 M parameters—a 56% parameter reduction over the TDR-Model baseline. To address robustness gaps in prior work, we systematically evaluate corruption resilience under three corruption types at five severity levels. The full model retains 97.91% accuracy under 20% random occlusion (a 1.80 percentage-point drop vs. 4.04 for the baseline) and reaches 45.04% under Gaussian noise (σ = 0.10), outperforming the CBAM-only variant by 2.57 percentage points. Inference runs at 40.1 FPS on an NVIDIA RTX 3060 laptop GPU, confirming practical suitability for in-field mobile deployment.

Online Oral Session 9

➢ Industrial Inspection and Intelligent Systems

- **Session Chair:** Chen Li, North China University of Technology, China
- **Time**: 16:30-18:35, July 19
- **Meeting Room**: ROOM C
- **Papers**: Invited Speech, EC2048, EC3102, EC442, EC3139, EC4160, EC4173, EC4186

Invited Speech
16:30-16:50

Liangxiu Han
Manchester Metropolitan University, UK

Meeting Societal Challenges: Scalable Big Data-driven, AI-enabled Approaches

Abstract: This talk will present our latest advances in scalable AI and big data learning, spanning fundamental methodological development and real-world application. Through case studies across domains such as health, it will demonstrate how our scalable AI solutions can support better decision-making, enable innovation and deliver meaningful public benefit.

EC2048 16:50-17:05	Performance Enhancement for Industrial Defect Detection via Dynamic Network and Feature Optimization Strategy Author(s): Lei Zhang, Mingrui Kou and Ziwen Gong Presenter: Lei Zhang, Hainan University, China Abstract: Industrial defect detection faces challenges of uneven crosscategory accuracy, insufficient multiscale feature utilization, and high computational cost. This paper proposes a collaborative optimization scheme based on PaDiM, integrating dynamic backbone selection, categoryspecific feature layers, and adaptive dimensionality reduction. On MVTec_AD, dynamic backbone improves Pixel-level AUROC from 0.954 (ResNet18) to 0.967; adding categoryspecific feature layers further raises it to 0.969. Adaptive dimensionality reduction cuts covariance time by 81.9% with minimal accuracy loss. The final combination achieves 0.969 Pixel-level AUROC and 64.7% lower computational cost, striking a balance between precision and efficiency. Extensive experiments on three datasets validate the generalization. Compared with the baseline PaDiM and state-

	of-the-art PatchCore, the proposed method further enhances detection performance and computational efficiency by introducing dynamic backbone selection, category-specific feature layer optimization, and adaptive dimensionality reduction.
EC3102 17:05-17:20	A Prototype Enhanced Text Confidence Guided Fusion Method for Pharmaceutical Package Recognition Author(s): Xutan Liang, Zefan Zuo, Wenqi Chang, Haochen Liu and Fanwei Meng Presenter: Zefan Zuo, Northeastern University at Qinhuangdao, China Abstract: In response to the problems existing in the process of drug identification, such as similar appearances of drug packaging, uneven quality of OCR text, and the tendency of recognition capability to decline in complex scenarios, an image–text fusion recognition method based on prototype enhancement and text confidence guidance is proposed. This method constructs an image branch and a character-level text branch to respectively extract the visual features of drug packaging and the OCR text features; then estimates text quality from three aspects: character confidence, text completeness, and structural rationality, and uses it as a reliability constraint in the fusion process. On this basis, a two-layer fusion mechanism guided by text confidence is designed, which adaptively adjusts the contributions of image and text representations at the feature level, and integrates the prediction results of the image branch, text branch, and fusion branch at the decision level. Furthermore, a category prototype enhancement module is introduced, which improves the discriminative boundaries between similar drug categories through prototype distance constraints. Experimental results show that this study provides a feasible solution for multimodal recognition of drug packaging under complex shooting environments. Keywords—Multimodal fusion; Text-confidence guidance; Pharmaceutical packaging recognition; Image–text fusion
EC442 17:20-17:35	Ultra-Short-Term Prediction of Ship Motions in Irregular Waves Based on RAO and Deep Learning Author(s): Feng Diao, Zhurui Shao, Wenhui Li, Tianyu Liu, Jiang Lyu, Zhen Huang, Jianing Zhang Presenter: Zhurui Shao, Dalian Maritime University, China Abstract: To address the limited coverage of experimental conditions, the difficulty of directly modeling nonstationary responses, and the accumulation of multi-step extrapolation errors in ultra-short-term prediction of ship motions in irregular waves, this study proposes a combined prediction method that uses the response amplitude operator (RAO) frequency-domain transfer relation as a physical reference and combines it with a VMD-LSTM model trained on measured motion-response time histories. Using towing-tank test data of the Yupeng ship model, two representative operating conditions are selected: a following-quartering-wave condition without forward speed and a bow-quartering-wave condition with a larger significant wave height and nonzero forward speed. The RAO is used to establish the frequency-domain relationship between wave excitation and ship motion responses, and irregular-wave motion response characteristics are analyzed accordingly. The prediction performance of LSTM, BiLSTM, Transformer-LSTM, and VMD-LSTM models is then compared. The RAO-based analysis provides a reference for identifying the dominant response frequencies and main energy distribution. The VMD-LSTM model can effectively follow the principal oscillatory trends of roll, pitch, and heave responses and performs well in preserving the amplitude envelope of nonstationary motion responses. The analysis further shows that variations in sea-state intensity, forward speed, and wave heading affect heave prediction errors and local crest trough tracking by altering the response amplitude, encounter relationship, and incident-wave direction.

EC3139 17:35-17:50	Boundary-Aware Semi-Supervised Segmentation for Rail Surface Defects Author(s): Yaru Huang, Zhongyi Shu and Xiaoli Geng Presenter: Zhongyi Shu, Zhongnan University of Economics and Law, China Abstract: Rail surface defect segmentation is important for railway inspection and maintenance, but accurate pixel-level annotation is costly and time-consuming. Semi-supervised segmentation provides a practical solution by exploiting unlabeled images together with limited labeled samples. However, directly applying generic semi-supervised segmentation methods to rail surface defects remains challenging because rail defects are often small, thin, low-contrast, and surrounded by complex rail textures. Confidence-based pseudo labels may therefore contain texture-induced false positives, fragmented defect regions, and inaccurate boundaries. To address these issues, this paper proposes a boundary-aware semi-supervised segmentation method for rail surface defects. Specifically, a structure-aware pseudo-label refinement strategy is designed to suppress isolated texture noise by incorporating connected-component analysis, area constraints, and shape priors. A boundary-aware consistency constraint is introduced to improve defect contour localization, and a local structure propagation strategy is further used to recover low-confidence but spatially related defect regions. Experiments are conducted on the RSDDs under 10%, 20%, and 50% annotation ratios. The proposed method achieves 66.38% Dice and 52.82% mIoU under the 10% annotation setting, and reaches 73.62% Dice and 60.37% mIoU under the 50% annotation setting, outperforming representative semi-supervised baselines. Ablation studies further verify the effectiveness of the proposed structure refinement, boundary consistency, and local propagation modules.
EC4160 17:50-18:05	Efficient Edge Deployment of YOLOv4-Tiny for Real-Time Moving Object Detection on a DSP-NPU Heterogeneous Accelerator Author(s): Junjie Qu, Yuehui Liu, Yongquan Fu and Xiao Hu Presenter: Junjie Qu, National University of Defense Technology, China Abstract: Real-time moving-object detection on embedded platforms is constrained not only by neural-network inference latency but also by serial task scheduling and inter-stage data movement. This paper presents an efficient edge deployment method for YOLOv4-tiny on an FTXXNE DSP-NPU heteroge neous accelerator. Instead of modifying the detector architecture, the proposed design reorganizes the complete vision pipeline into four parallel stages, including image acquisition and display, image resizing, NPU-based inference, and post-processing with landing-distance prediction. Shared on-chip memory is used to reduce inter-core data copying, and ping-pong buffers are introduced to avoid read-write conflicts between adjacent pipeline stages. Experiments on a motion-blurred table-tennis-ball de tection task show that, compared with a single-core sequential implementation, the proposed pipeline reduces the steady-state processing interval from 25.36 ms to 15.02 ms without least squares prediction and from 44.42 ms to 20.11 ms with prediction. The corresponding throughput improvements are 68.8% and 120.9%, respectively. With UINT8 quantization, the additional inference power consumption is only 2.274 W. The results demonstrate that architecture-aware deployment can improve the real-time performance of embedded vision systems without changing the object detector itself.
EC4173 18:05-18:20	P2-Guided RT-DETR for Tiny Industrial Hazard Detection Author(s): Zhaohui Jin and Shuimiao Du Presenter: Zhaohui Jin, Shanghai University, China Abstract: Industrial safety monitoring requires detectors to localize small hazards under occlusion, low contrast, and repeated background structures. RT-DETR can

	lose such evidence after repeated downsampling, whereas shallow features may introduce texture-driven false positives. This paper adapts RT-DETR-R18 by combining a P2 high-resolution path, DySample alignment, EMA filtering, and a Focaler-Inner-MPDIoU-style regression objective. On the internal validation split, the full variant reaches 0.9509 precision, 0.8920 recall, 0.9408 mAP50, and 0.6220 mAP50-95. It improves mAP50-95 by 10.35 points over RT-DETR-R18 and by 2.33 points over pretrained YOLO11m, with 23.06 M parameters and 131.34 GFLOPs. Repeated-seed reruns and FP32/FP16 timing further show that the gain is stable but comes with a clear deployment-cost trade-off. Ablations indicate that P2 feature injection and DySample alignment provide the main strict-localization gain, while the single-split evaluation limits broad generalization claims.
EC4186 18:20-18:35	AGV Path Planning and Obstacle Avoidance Method Based on an Enhanced A* and DWA Fusion Algorithm Author(s): Chuxin Cao, Ruichen Liu, Guanghui Gao and Yue Gao Presenter: Ruichen Liu, Soochow University, China Abstract: To address the issues of numerous inflection points, insufficient obstacle clearance in conventional A*, and fixed weights and weak dynamic avoidance capability inconventional DWA, this paper proposes an AGV path planning and obstacle avoidance method based on the fusion of enhanced A* and adaptive DWA. In the global planning stage, costmap constraints, redundant waypoint pruning, line-of-sight (LOS) connectivity optimization, constrained smoothing and heading compensation are incorporated to improve path safety and trackability. In the local planning stage, the velocity sampling density, obstacle clearance weight and terminal deceleration strategy of DWA are optimized, and a dynamic parameter regulation mechanism is designed according to local costmap density. Residual obstacle clearance and motion stagnation recovery are further introduced to improve robustness. Comparative experiments in ROS and Gazebo demonstrate that the proposed method reduces path length by 5.2% and increases the minimum obstacle clearance by 10.5% in static environments. In dynamic environments, the enhanced A*+DWA method reduces path length by 6.1%, improves the obstacle avoidance success rate from 60.0% to 100.0%, and increases the minimum obstacle clearance by 25.0%.