

# The 10th International Conference on Image, Vision and Computing (ICIVC 2025)

## Conference Agenda

### WORKSHOP

**2025 9th International Conference on Deep Learning  
Technologies (ICDLT 2025)**

**July 16-18, 2025  
Chengdu, China**

► Co-sponsored by



**四川大学**  
SICHUAN UNIVERSITY



► Organized by



**四川大学 电子信息学院**  
SICHUAN UNIVERSITY College of Electronics and Information Engineering

► Technical sponsored by



► Co-supported by



大数据研究中心  
Data Science Research Center



**中国海洋大学**  
OCEAN UNIVERSITY OF CHINA



**西安科技大学**  
XI'AN UNIVERSITY OF SCIENCE AND TECHNOLOGY



**青岛大学**  
QINGDAO UNIVERSITY



**University  
of Windsor**



**广东省知识产权大数据重点实验室**  
Guangdong Provincial Key Laboratory of Intellectual Property & Big Data

# TABLE OF CONTENTS

03..... Welcome Remarks

04..... Local Information

04..... Attendee Guideline

09..... Conference Agenda

15..... Introduction of Speakers

20 ..... Onsite Oral Session 1

24..... Onsite Oral Session 2

28..... Onsite Poster Session 1

30..... Onsite Oral Session 3-1

33..... Onsite Oral Session 3-2

36..... Onsite Poster Session 2

38..... Onsite Oral Session 3-3

42..... Onsite Oral Session 4

46..... Online Oral Session 1

50..... Online Oral Session 2

54..... Online Oral Session 3-1

58..... Online Oral Session 3-2

62..... Online Oral Session 4

66..... Online Oral Session 5

70..... Note



# WELCOME REMARKS

Dear Colleagues,

On behalf of the organizing committee, it is our great pleasure to welcome you to the 10th International Conference on Image, Vision and Computing (ICIVC 2025), co-sponsored by Sichuan University, China and IEEE, organized by the College of Electronics and Information Engineering of Sichuan University, China. Concurrently, we are excited to host the 9th International Conference on Deep Learning Technologies (ICDLT 2025).

Taking place in the culturally rich city of Chengdu, China from July 16-18, 2025, this conference serves as a premier platform for researchers, academics, and industry experts to converge and exchange groundbreaking ideas in the realms of image processing, computer vision, and deep learning technologies.

Our esteemed keynote speakers and invited speakers bring a wealth of knowledge and experience to the table, promising enlightening discussions on the latest trends and advancements in the field. Their insights will undoubtedly inspire new avenues of research and innovation.

With a diverse program comprising six offline sessions and eight online sessions, ICIVC2025 and ICDLT2025 offer a dynamic environment for fostering collaborations, sharing research findings, and fostering professional relationships.

We urge all participants to actively participate in the sessions, engage in fruitful discussions, and leverage this unique opportunity to broaden your expertise and contribute to the collective knowledge in these rapidly evolving disciplines.

Thank you for being part of ICIVC2025 (Workshop: ICDLT2025). Let us come together to explore the frontiers of image processing, computer vision, and deep learning technologies, shaping the future of these transformative fields.

ICIVC 2025 Organizing Committees  
Chengdu, China



# LOCAL INFORMATION

## Conference Venue



### **Chengdu Xiangyu Hotel (成都祥宇宾馆)**

地址: 四川省成都市武侯区新南路 103 号

Address: No.103, Xinnan Road, Wuhou District, Chengdu, Sichuan, China

Reservation Call: 028-85551111-8122

订房电话: 销售经理 13982229918

# ATTENDEE GUIDELINE

## Transportation

### **From Shungaliu Airport-从双流机场出发**



By Metro line (40 mins)

Take Metro Line 10 (towards Taipingyuan) → Transfer to Line 3 (towards Chengdu Medical College) at Taipingyuan Station → Get off at Moziqiao Station (Exit A). Walk about 500m (7 mins) to the hotel.



By Taxi (40 mins, 40RMB)

### **From Tianfu Airport-从天府机场出发**



By Metro line (1 hour 20mins)

Take Metro Line 18 (towards South Railway Station) → Transfer to Line 1 (towards Weijianian) at South Railway Station → Transfer to Line 3 (towards Chengdu Medical College) at Sichuan Gymnasium Station → Get off at Moziqiao Station (Exit A). Walk about 500m to the hotel.



By Taxi (1 hour 10 mins, 150 RMB)



### From Chengdu East Railway Station-从成都东站出发



By Metro line (40 mins)

Take Metro Line 7 (Inner Loop direction) → Transfer to Line 3 (towards Shuangliu West Station) at Sima qiao Station → Get off at Moziqiao Station (Exit A). Walk about 500m to the hotel.



By Taxi (30 mins, 30 RMB)

### From Chengdu South Railway Station-从成都南站出发



By Metro line (30mins)

Take Metro Line 1 (towards Weijianian) → Transfer to Line 3 (towards Chengdu Medical College) at Sichuan Gymnasium Station → Get off at Moziqiao Station (Exit A). Walk about 500m to the hotel.



By Taxi (20 mins, 25RMB)

### From Chengdu West Railway Station-从成都西站出发



By Metro line (35mins)

Take Metro Line 4 (towards Wannianchang) → Transfer to Line 3 (towards Chengdu Medical College) at Cultural Palace Station → Get off at Moziqiao Station (Exit A). Walk about 500m to the hotel.



By Taxi (45 mins, 50RMB)

## Guideline for Onsite Participation



### Sign In & Material Collecting

**Date: July 16<sup>th</sup>, 2025**

**Time: 10:00-17:00**

**Venue: Lobby of Chengdu Xiangyu Hotel**



### Instructions for Presentation

- Regular Oral Presentation: 15 minutes (including Q&A).
- Get your presentation PPT or PDF files prepared. Please copy your slide file to the desktop before session starts.
- Devices Provided by the Conference Organizer: Laptop (with MS-Office & Adobe Reader), projector & screen, laser pointer.



### Notes and Tips

- Please show the meal ticket to the staff when you enter the dining hall.
- In the consideration of the personal and property security belongings to conference



participants, please take care of your belongings and be sure to take the attendance cards during the conference.

## Guideline for **Poster** Participation

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.
- Please show the meal ticket to the staff when you enter the dining hall.

The size of poster is 180cm\*80cm. Posters must be in portrait format (height > width). This cannot be modified. Please drill round hole at four corners

## Guideline for **Online** Participation

### **ZOOM Download Link**

<https://zoom.us/download> (Oversea)

<https://zoom.com.cn/download> (Author in China)

### **Meeting Rooms**

ZOOM Meeting A: (<https://us02web.zoom.us/j/87471010157>)

ZOOM Meeting B: (<https://us02web.zoom.us/j/87933161872>)

Only one password for all online room. Password: 071618

### **Time Zone**

The conference is arranged based on Beijing Time (GMT+8).

Please carefully check your presentation time and join the conference 10 minutes in advance.

### **Equipment Needed**

- A computer with internet connection and camera
- Headphones
- Stable internet connection
- A quiet place and proper background

### **Test Your Presentation**

**Date: July 16<sup>th</sup>, 2025**

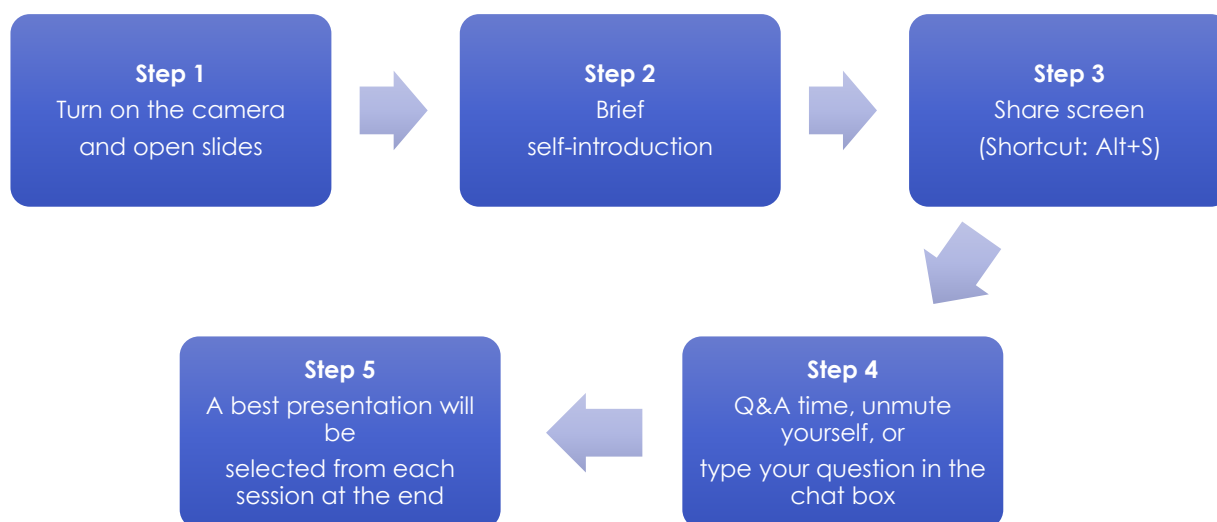
Prior to the formal meeting, online presenters shall join the test room to ensure everything is on the right track. Please check your test Zoom Meeting ID on this program.

## Oral Presentation

- Timing: a maximum of 15 minutes in total, including 2-3 minutes for Q&A. Please make sure your presentation is well timed.
- Please join the meeting room 10 minutes in advance.
- Stay online during Keynote & Invited speeches and your own sessions.
- English Only during the conference.

Rename your screen name before entering the room	Example
<b>Authors: Paper ID-Name</b>	ADxxx-San Zhang
<b>Listener: Listener Number-Name</b>	Listener- San Zhang
<b>Keynote Speaker: Keynote-Name</b>	Keynote- San Zhang
<b>Committee Member: Position-Name</b>	Committee- San Zhang

## Presentation Process



## Conference Recording

The whole conference will be recorded. We appreciate you proper behavior and appearance. The recording will be used for conference program and paper publication requirements. The video recording will be destroyed after the conference and it cannot be distributed to or shared with anyone else, and it shall not be used for commercial nor illegal purpose. It will only be recorded by the staff and presenters have no rights to record.

## Emergency Call

Police: 110

Ambulance: 120

### Other Information



扫码发送会议简称添加会议助理微信

Scan the QR code to add conference secretary on Wechat



扫码观看会议照片

Scan the QR code to view the photos



扫码下载电子版会议日程

Scan the QR code to download the electronic version of  
 the conference agenda



扫码下载会议通知





# CONFERENCE AGENDA

< July 16<sup>th</sup>, 2025, Wednesday (GMT+8)>

Onsite Registration & Materials Collection for Onsite Participants 注册签到		
10:00 – 17:00	Venue: Lobby of Chengdu Xiangyu Hotel 地址：成都祥宇宾馆大堂	
14:00-16:30	Academic Visit to Sichuan University Museum 川大博物馆学术参观 Gathering Place: Xiangyu Hotel (Conference Registration Desk) 集合点：祥宇宾馆 (会议签到处)	
ZOOM Test for Online Participants 线上参会者测试		
14:00 – 16:30	Zoom A: <a href="https://us02web.zoom.us/j/87471010157">https://us02web.zoom.us/j/87471010157</a> Password: 071618	Online Session 1&3-1& 3-2 &Conference Committee
14:00 – 16:30	Zoom B: <a href="https://us02web.zoom.us/j/87933161872">https://us02web.zoom.us/j/87933161872</a> Password: 071618	Online Session 2 & 4 & 5 &Session Chair



## Morning< July 17<sup>th</sup>, 2025, Thursday (GMT+8)>

Venue: Xiangrui Hall- 2F | 会址: 二楼祥瑞厅

Online Room A:

<https://us02web.zoom.us/j/87471010157>

Password: 071618

Host: TBA | 主持人: TBA

### Opening Ceremony

<b>Opening Remarks</b> 09:00-09:05	<b>TBA</b>
<b>Welcome Message</b> 09:05-09:10	<b>TBA</b>
09:10-09:15	<b>Group Photo</b>

### Keynote Speeches

<b>Keynote Speech I</b> 09:15 – 9:55	<b>Prof. Guoping Qiu</b> <b>The University of Nottingham, UK &amp; China Chief Scientist,</b> <b>Everimaging Ltd.</b> <i>Title: High Dynamic Range – The Last Frontier of Digital Imaging</i>
<b>Keynote Speech II</b> 09:55 – 10:35	<b>Prof. Hui Zhang</b> <b>Hunan University, China</b> <i>Title: Multimodal Intelligent Perception Technology and Applications of UAVs in Complex Power Scenarios</i>

**Coffee Break ☕ 10:35—11:00**

<b>Keynote Speech III</b> 11:00 – 11:40	<b>Prof. Yiu-Ming Cheung (FIEEE, FAAAS, FIET, FBCS)</b> <b>Hong Kong Baptist University, Hong Kong, China</b> <i>Title: Imbalanced Data Learning: From Class Imbalance to Long-tailed Data Classification for Visual Recognition</i>
--	--

**Lunch Time 🍴 11:50—13:30**

Venue: Xiangjing Hall - 3F | 午餐: 三楼祥景厅

## Afternoon < July 17th, 2025, Thursday (GMT+8)>

### Keynote Speeches

Online Room A: <https://us02web.zoom.us/j/87471010157> | Password: 071618

Host: TBA | 主持人:

<b>Keynote Speech IV</b> 13:30 – 14:10	<b>Prof. KWONG Tak Wu Sam, Lingnan University Hong Kong, China</b> <b>Fellow of Hong Kong AES, US NAI and IEEE (Online)</b> <i>Title: Deep Learning-Based Video Coding and its Applications</i>
<b>Keynote Speech V</b> 14:10 – 14:50	<b>Prof. Xudong Jiang (IEEE Fellow)</b> <b>Nanyang Technological University, Singapore</b> <i>Title: How Deep CNN Revolutionizes MLP and How Transformer Revolutionizes CNN</i>

### Coffee Break ☕ 14:50—15:20

15:20 – 17:15	<b>Onsite Oral Session 1</b> <b>Image Processing Theory and Application</b>  <b>Session Chair: TBA</b>  Invited Speaker-Longke He, Invited Speaker-Wenhui Jiang, AD361, AD330, AD355, AD374, AD533	<b>Xiangrui Hall</b> <b>2F</b> 二楼祥瑞厅
15:20 – 17:10	<b>Onsite Oral Session 2</b> <b>Deep and Machine Learning Applications</b>  <b>Session Chair: Zhu Meng, Beijing University of Posts and Telecommunications, China</b>  Invited Speaker-Hongping Gan, AD536, AD2012, AD2014, AD3016, AD3024, AD4035	<b>Yutang Chun 1</b> <b>2F</b> 二楼玉堂春 1 号
15:20-17:10	<b>Onsite Poster Session 1</b> <b>Intelligent Image Detection, Recognition and Image Modeling</b>  <b>Session Chair: Tao Zhang, Shanghai Jiao Tong University, China</b>  AD209, AD418, AD385, AD220, AD342, AD344, AD221, AD225, AD409, AD548, AD356, AD333, AD380, AD396, AD398, AD359, AD3022	<b>Yutang Chun 2</b> <b>2F</b> 二楼玉堂春 2 号

### Dinner Time 🍴 18:00 – 20:30

Venue: Manting Fang 2F | 晚宴: 二楼满庭芳

< July 18th, 2025, Friday (GMT+8) >

Onsite Sessions		
10:00-11:50	<b>Onsite Oral Session 3-1</b> <b>Computer Vision Techniques and Application</b>  <b>Session Chair: Yunbo Wang, Central South University, China</b>  Invited Speaker-Yuan Cao, AD217, AD352, AD367, AD336, AD537, AD3020	<b>Xinagqing Hall</b> <b>2F</b> 二楼祥庆厅
10:00-11:50	<b>Onsite Oral Session 3-2</b> <b>Computer Vision Techniques and Application</b>  <b>Session Chair: TBA</b>  Invited Speaker-Hui Liu, AD347, AD390, AD392, AD4029, AD222, AD413	<b>Xiangtai Hall</b> <b>2F</b> 二楼祥泰厅
10:00-11:50	<b>Onsite Poster Session 2</b> <b>AI Based Digital Image Analysis and Processing Technology</b>  <b>Session Chair: TBA</b>  AD351, AD538, AD337, AD211, AD403, AD341, AD226, AD404, AD519, AD526, AD368, AD400, AD532, AD383, AD547, AD1001, AD2015, AD3021	<b>Yutang Chun 1</b> <b>2F</b> 二楼玉堂春 1 号
Lunch Time  12:00—13:30		
13:30- 15:20	<b>Onsite Oral Session 3-3</b> <b>Computer Vision Techniques and Application</b>  <b>Session Chair: TBA</b>  Invited Speaker-Chen Li, AD389, AD387, AD223, AD4032, AD35, AD550	<b>Xianghua Hall</b> <b>3F</b> 三楼祥华厅
13:30- 15:30	<b>Onsite Oral Session 4</b> <b>Multimedia Technology</b>  <b>Session Chair: Yuanzhouhan Cao, Beijing Jiaotong University, China</b>  Invited Speaker-Wuzhen Shi, Invited Speaker-Wenxue Cui, Invited Speaker-Yiyi Liao, AD216, AD546, AD549, AD414	<b>Xiangqing Hall</b> <b>2F</b> 二楼祥庆厅
Coffee Break 		

< July 18th, 2025, Friday (GMT+8) >

Online Sessions		
10:00-11:50	<b>Online Oral Session 1</b> <b>Image Processing Theory and Application</b>  <b>Session Chair: TBA</b>  Invited Speaker-Sinong Quan, AD212, AD360, AD535, AD520, AD417, AD416	<b>ZOOM A</b>  <b>87471010157</b>  <b>Password:071618</b>
10:00-11:50	<b>Online Oral Session 2</b> <b>Computer Graphics and Computational Photography</b>  <b>Session Chair: TBA</b>  Invited Speaker-Zizhao Wu, AD207, AD407, AD412, AD544, AD1002, AD539	<b>ZOOM B</b>  <b>87933161872</b>  <b>Password:071618</b>
<b>Lunch Time</b>  <b>12:00—13:30</b>		
13:30- 15:30	<b>Online Oral Session 3-1</b> <b>Computer Vision Techniques and Application</b>  <b>Session Chair: TBA</b>  AD227, AD354, AD343, AD363, AD366, AD364, AD345, AD346	<b>ZOOM A</b>  <b>87471010157</b>  <b>Password:071618</b>
13:30- 15:30	<b>Online Oral Session 3-2</b> <b>Computer Vision Techniques and Application</b>  <b>Session Chair: TBA</b>  AD353, AD405, AD395, AD384, AD528, AD542, AD530, AD415	<b>ZOOM B</b>  <b>87933161872</b>  <b>Password:071618</b>
<b>Break</b>  <b>15:30-15:45</b>		
15:45- 17:45	<b>Online Oral Session 4</b> <b>Deep and Machine Learning Applications</b>  <b>Session Chair: TBA</b>  AD1007, AD2013, AD523, AD3017, AD3018, AD4027, AD4033, AD379	<b>ZOOM A</b>  <b>87471010157</b>  <b>Password:071618</b>



15:45- 18:00	<p><b>Online Oral Session 5</b>  <b>Innovative Applications and Technological Breakthroughs of Computer Vision and Computational Intelligence in Multiple Fields</b></p> <p><b>Session Chair: TBA</b></p> <p>AD204, AD381, AD522, AD527, AD408, AD219, AD541, AD382, AD543</p>	<p><b>ZOOM B</b>  <b>87933161872</b>  <b>Password:071618</b></p>
--------------	--	--



# Keynote Speaker



**Prof. Guoping Qiu**

*The University of Nottingham, UK & China  
 Chief Scientist, Everimaging Ltd.*

Speech Time: 09:15-9:55, July 17, 2025

Venue: Xiangrui Hall- 2F | 会址: 二楼祥瑞厅

## ***Speech Title: High Dynamic Range – The Last Frontier of Digital Imaging***

**Abstract:** Many years of research and development plus billions of dollars investment in technology have made digital photography device ubiquitous and very sophisticated. Despite huge progress, there are still the occasions, for example when taking a photo of an evening party at a restaurant, where the image quality will still come out poorly. Either the dark shadows are too dark such that no details are visible, or the light areas are so bright such that they are completely saturated, and no details are visible. Even after turning on the high dynamic range (HDR) function in your camera which is now a feature in every smartphone, or adjusting the various control buttons, the situations will not improve much. And yet the photographer on the scene can clearly see every detail both in the dark and in the bright regions. The question is, why? In this talk I will show that this difficulty is caused by the high dynamic range of the light intensities of the scene, and we call this the HDR problem. I will show from first principle that HDR is the cause of many difficulties in digital imaging (photography) and correct some of the misconceptions in many recent literatures on image processing problems such as low-light (or dark) image enhancement, especially those so-called end-to-end blackbox solutions based on deep learning. I will demonstrate both theoretically and in practice that HDR is the last technical obstacle, the last frontier, of digital imaging.

**Biography:** Professor Guoping Qiu has been researching neural networks and their applications in image processing since the 1990s. He spearheaded learning-based super-resolution techniques and developed early neural network solutions for image coding and compression artifact removal, well before deep learning became mainstream in these applications. He also introduced one of the earliest representation learning methods that leveraged unsupervised competitive neural networks for learning image features. He has been at the forefront of high dynamic range (HDR) imaging and pioneered tone-mapping methods that have fundamentally transformed how HDR content is processed and displayed. His group developed some of the best performing practical HDR tone mapping solutions that are widely cited by imaging industrial leaders including smartphone makers, camera manufacturers and imaging software companies. As Chief Scientist at Everimaging ([www.everimaging.com](http://www.everimaging.com)), the company behind the multi-award-winning visual content creation software HDR Darkroom and Fotor with hundreds of millions global users, he is driving advancements in imaging technology research to solve real-world problems. With a distinguished career spanning academia and industry, Professor Qiu has contributed to fundamental research and real-world applications in imaging technology. Currently, he holds the position of Chair Professor of Visual Information Processing at the School of Computer Science, University of Nottingham, UK. Additionally, he is serving as the Vice Provost for Education and Student Experience at the University of Nottingham Ningbo China (UNNC), overseeing the education and student experience of a diverse academic community of over 11,000 students and faculty from over 70 countries and regions. UNNC delivers all its teaching in English and offers undergraduate, Master's, and PhD programs across business, humanities, social sciences, and science and engineering, awarding degrees from the University of Nottingham.



# Keynote Speaker



**Prof. Hui Zhang**  
**Hunan University, China**

Speech Time: 09:55-10:35, July 17, 2025

Venue: Xiangrui Hall- 2F | 会址: 二楼祥瑞厅

## ***Speech Title: Multimodal Intelligent Perception Technology and Applications of UAVs in Complex Power Scenarios***

**Abstract:** To address challenges in drone inspection tasks for complex power scenarios, including infrared thermal fault detection, line vegetation classification, and tower tilt detection, this report proposes intelligent perception technologies based on multimodal information fusion. By integrating visible light images, infrared images, point cloud data, and multispectral data, it overcomes challenges such as environmental complexity, information incompleteness, and sensor perception limitations, significantly enhancing the perception and cognition capabilities of UAV systems in complex environments. The report focuses on the following key aspects: 1) Adaptive image registration and predictive information transfer techniques are proposed to address the spatial alignment of multimodal data, enabling precise localization of power equipment and accurate temperature interpretation; 2) A tree obstacle classification method based on point cloud and multispectral data fusion is designed, leveraging the complementarity of different modalities to accurately identify tree species in power corridors, thereby improving the accuracy and efficiency of inspection tasks; 3) Multimodal information-coordinated tower tilt detection and semantic segmentation technologies are developed, enhancing the intelligence level of power facility inspection in complex environments. Through multimodal data fusion and intelligent processing, this report demonstrates how multimodal sensing technologies can improve the efficiency, accuracy, and safety of drone inspections in complex power scenarios, effectively meeting the demands of national strategic needs.

**Biography:** Hui Zhang, Professor, Ph.D. Supervisor, serves as the Executive Vice Dean of the School of Robotics at Hunan University, Deputy Director of the National Engineering Research Center for Robot Vision Perception and Control Technology, and Council Member and Deputy Secretary-General of the China Society of Image and Graphics. He has been recognized as a Distinguished Professor under the Ministry of Education's "Changjiang Scholars Program" and a Youth Top-notch Talent of the National "Ten Thousand Talents Program." His research focuses on robotic vision inspection, deep learning-based image recognition, and intelligent manufacturing robot technologies and applications. In recent years, he has led more than 20 projects, including a key topic under the Science and Technology Innovation 2030—"New Generation Artificial Intelligence" Major Project, two key projects funded by the National Natural Science Foundation of China, a JW1XX Engineering Key Project, sub-projects under the National Key R&D Program, and sub-projects under the National Science and Technology Support Program. He has published over 70 papers in international and domestic journals such as IEEE Transactions and holds 42 national invention patents and 5 software copyrights. He received the 2018 National Technological Invention Award (Second Prize) as the first contributor, and as the principal investigator, he was awarded the First Prize of Hunan Provincial Science and Technology Progress Award in 2022, the Second Prize in 2019, and the First Prize of the Science and Technology Progress Award of the China General Chamber of Commerce in 2019. Additionally, as a key contributor, he has won 15 provincial and ministerial science and technology progress awards, the Special Prize of the 13th Hunan Provincial Teaching Achievement Award in 2022, and the Second Prize of the National Teaching Achievement Award (Higher Education - Postgraduate) in 2022.



# Keynote Speaker



**Prof. Yiu-Ming Cheung (FIEEE, FAAAS, FIET, FBCS)**  
**Hong Kong Baptist University, Hong Kong, China**

Speech Time: 11:00-11:40, July 17, 2025

Venue: Xiangrui Hall- 2F | 会址: 二楼祥瑞厅

***Speech Title: Imbalanced Data Learning: From Class Imbalance to Long-tailed Data Classification for Visual Recognition***

**Abstract:** Imbalance data refer to the number of samples among classes is extremely imbalanced, which is common in our daily life, e.g. medical diagnosis, and autonomous driving. In general, the problem of learning from imbalanced data is nontrivial and challenging in the field of data engineering and machine learning, which has attracted growing attentions in recent years. In this talk, the imbalance data learning problem is introduced from class imbalance to long-tailed data learning, including their potential applications, and the impacts from a model learning perspective. Then, the latest research progress on imbalance data learning will be reviewed, including some representative methods in the literature. Lastly, the potential research directions in this field will be discussed.

**Biography:** Yiu-ming Cheung is a Chair Professor of the Department of Computer Science in Hong Kong Baptist University (HKBU). He is a Fellow of IEEE, AAAS, IAPR, IET, and BCS. His research interests include Machine Learning and Visual Computing, as well as their applications. He has published over 300 articles in the high-quality conferences and journals. He has been ranked the World's Top 1% Most-cited Scientists in the field of Artificial Intelligence and Image Processing by Stanford University since 2019. He was elected as an IEEE Distinguished Lecturer, and the Changjiang Chair Professor awarded by Ministry of Education of China. He has served in various capacities (e.g., Organizing Committee Chair, Program Committee Chair, Program Committee Area Chair, and Financial Chair) at several top-tier international conferences, including IJCAI'2021, ICPR'2020, ICDM'2017 & 2018, WCCI'2016, WI-IAT'2012, ICDM'2006 & WI-IAT'2006, to name a few. He is currently the Editor-in-Chief of IEEE Transactions on Emerging Topics in Computational Intelligence, besides serving as an Associate Editor for several other prestigious journals. More details can be found at: <https://www.comp.hkbu.edu.hk/~ymc>.



# Keynote Speaker



**Prof. KWONG Tak Wu Sam (Fellow of Hong Kong AES, US NAI and IEEE)**  
**Lingnan University Hong Kong, China**

Speech Time: 13:30-14:10, July 17, 2025

Online Room A: <https://us02web.zoom.us/j/87471010157>

Password: 071618

## ***Speech Title: Deep Learning-Based Video Coding and its Applications***

**Abstract:** In 2016, Cisco released the White paper, VNI Forecast and Methodology 2015-2020, which predicted that by 2020, 82 percent of Internet traffic would come from video applications such as video surveillance and content delivery networks. The report also revealed that in 2015, Internet video surveillance traffic nearly doubled, virtual reality traffic quadrupled, TV grew by 50 percent, and other applications similarly saw significant increases. The report estimated that the annual global traffic would first time exceed the zettabyte (ZB; 1000 exabytes [EB]) threshold in 2016 and will reach 2.3 ZB by 2020, with 1.886 ZB attributed to video data.

Today, AI and machine learning are increasingly being used in video processing to improve video quality, reduce bandwidth requirements, and enhance user experience. For instance, AI algorithms can optimize video encoding parameters based on the content of the video, reducing the bitrate required for a given level of video quality. AI can also be used for video content analysis, enabling automated scene detection, object recognition, and event detection. This has significant applications in video surveillance, where AI algorithms can be used to identify and track individuals or objects of interest in real-time.

Overall, the use of AI in video is a rapidly growing field with immense potential for improving the efficiency and quality of multimedia services. In this talk, I will present the latest research results on machine learning and deep neural network-based video coding, and their applications to the real world, such as saliency detection and underwater imaging.

**Biography:** Sam Kwong received his B.Sc. degree from the State University of New York at Buffalo, M.A.Sc. in electrical engineering from the University of Waterloo in Canada, and Ph.D. from Fernuniversität Hagen, Germany. Before joining Lingnan University, he was the Chair Professor at the City University of Hong Kong and a Diagnostic Engineer with Control Data Canada. He was responsible for designing diagnostic software to detect the manufacturing faults of the VLSI chips in the Cyber 430 machine. He later joined Bell-Northern Research as a Member of the Scientific Staff working on the Integrated Services Digital Network (ISDN) project.

Kwong is currently Chair Professor at the Lingnan University of the Department of Computing and Decision Science. He previously served as Department Head and Professor from 2012 to 2018 at the City University of Hong Kong. Prof Kwong joined CityU as a Department of Electronic Engineering lecturer in 1989. Prof. Kwong is the associate editor of leading IEEE transaction journals, including IEEE Transactions on Evolutionary Computation, IEEE Transactions on Industrial Informatics, and IEEE Transactions on Cybernetics. He was the President of IEEE Systems, Man And Cybernetics Society from 2022-23.





# Keynote Speaker



**Prof. Xudong Jiang (IEEE Fellow)**  
**Nanyang Technological University, Singapore**

Speech Time: 14:10-14:50, July 17, 2025

Venue: Xiangrui Hall- 2F | 会址: 二楼祥瑞厅

***Speech Title: How Deep CNN Revolutionizes MLP and How Transformer Revolutionizes CNN***

**Abstract:** Discovering knowledge from data has many applications in various artificial intelligence (AI) systems. Machine learning from the data is a solution to find right information from the high dimensional data. It is thus not a surprise that learning-based approaches emerge in various AI applications. The powerfulness of machine learning was already proven 30 years ago in the boom of neural networks but its successful application to the real world is just in recent years after the deep convolutional neural networks (CNN) have been developed. This is because the machine learning alone can only solve problems in the training data but the system is designed for the unknown data outside of the training set. This gap can be bridged by regularization: human knowledge guidance or interference to the machine learning. This speech will analyze these concepts and ideas from traditional neural networks to the deep CNN and Transformer. It will answer the questions why the traditional neural networks fail to solve real world problems even after 30 years' intensive research and development and how CNN solves the problems of the traditional neural networks and how Transformer overcomes limitation of CNN and is now very successful in solving various real world AI problems.

**Biography:** Xudong Jiang (Fellow, IEEE) received the bachelor's and master's degrees from University of Electronic Science and Technology of China, and the Ph.D. degree Helmut Schmidt University, Hamburg, Germany. From 1998 to 2004, he was with Institute for Infocomm Research, A\*STAR, Singapore, as a Lead Scientist, and the Head of the Biometrics Laboratory. He joined Nanyang Technological University (NTU), Singapore, as a Faculty Member in 2004, where he served as the Director of the Centre for Information Security from 2005 to 2011. He is currently a professor with the School of EEE, NTU and serves as Director of Centre for Information Sciences and Systems. He has authored over 200 papers with over 60 papers in IEEE journals including 10 papers in T-PAMI and 18 papers in T-IP, and over 30 papers in top conferences such as CVPR/ICCV/ECCV/AAAI/ICLR/NeurIPS. His current research interests include computer vision, machine learning, pattern recognition, image processing, and biometrics. Dr. Jiang served as Associate Editors for IEEE SPL and IEEE T-IP. Currently he is Fellow of IEEE and serves as a Senior Area Editor for IEEE T-IP and the Editor-in-Chief for IET Biometrics.



# Onsite Oral Session 1

**Topic:** Image Processing Theory and Application

**Session Chair:** TBA

July 17th, 2025 | 15:20 – 17:15

**Venue:** Xiangrui Hall - 2nd Floor | 二楼祥瑞厅

Invited Speaker-Longke He, Invited Speaker-Wenhui Jiang, AD361, AD330, AD355,  
AD374, AD533

<p><b>Invited Speaker 15:20-15:40</b></p>	<div data-bbox="569 622 801 855" data-label="Image">  </div> <div data-bbox="842 692 1182 786" data-label="Caption"> <p><b>Longke He</b> Xichang University, China</p> </div> <div data-bbox="497 853 1318 889" data-label="Text"> <p><i>Speech Title: Forest Fire and Smoke Recognition Based on YOLO</i></p> </div> <div data-bbox="349 938 1468 1379" data-label="Text"> <p><b>Abstract:</b> As a major global disaster, forest fire poses a serious threat to the ecological environment and people's lives and property. Therefore, it is crucial to realize rapid and accurate recognition and alarm of forest fires and smokes and develop a forest fire early warning system, in which forest fire recognition is the key technology. YOLO is one of typical models of deep learning, and it has been extensively researched and applied in numerous fields. This paper conducts research on forest fire and smoke recognition based on YOLO, delving into the detailed comparison and analysis of the differences and relations between the performance formulas of target detection and image classification. Through comprehensive simulations involving YOLOv5, YOLOv8, and YOLOv11, we evaluate the performance comparison between object detection and image classification. The paper achieves the conclusion that employing image classification in forest fire and smoke recognition can identify the occurrence of forest fire with an accuracy of about 99%.</p> </div>
<p><b>Invited Speaker 15:40-16:00</b></p>	<div data-bbox="569 1429 801 1662" data-label="Image">  </div> <div data-bbox="842 1482 1220 1630" data-label="Caption"> <p><b>Wenhui Jiang</b> Jiangxi University of Finance and Economics, China</p> </div> <div data-bbox="505 1673 1311 1709" data-label="Text"> <p><i>Speech Title: Combining visual grounding with visual captioning</i></p> </div> <div data-bbox="349 1758 1468 2013" data-label="Text"> <p><b>Abstract:</b> Visual captioning aims to describe the visual content in the images or videos using natural language sentences. To generate sentence descriptions more effectively, existing visual captioning methods usually adopt an encoder-decoder architecture. Recent methods further introduce attention mechanisms, which significantly improve the performance of the model due to its powerful ability to guide the model to focus on relevant visual regions when generating descriptions. Although attention-based models have achieved significant performance, previous researches have shown that attention mechanisms are incapable of correctly</p> </div>

	<p>associating generated words with meaningful visual regions, leading to undesirable behaviours such as object hallucinations. In this talk, we introduce grounded visual captioning, which combines visual grounding and visual captioning. The proposed grounding model encourages the captioner to dynamically focus on informative regions of the objects, therefore improves visual captioning.</p>
Paper ID &Time	Presentation
<p><b>AD361</b> <b>16:00-16:15</b></p>	<p>Diagnosis of Pediatric Hypopigmentary Dermatoses Based on Lightweight HierAttn Network  <b>Authors:</b> Jiaying Chen, Xun Lang, Zhao Zhang, Dongjie Sun and Xieyang Zhang  <b>Presenter:</b> Jiaying Chen, Yunnan University, China</p> <p><b>Abstract:</b> Hypopigmented skin disorders adversely affect both aesthetics and systemic health, undermining quality of life. Early detection is crucial for timely intervention and disease prevention. However, pediatric diagnostic applications of deep learning face significant hurdles due to the scarcity of labeled pediatric datasets. Additionally, the intricate nature of these disorders necessitates the integration of both local and global features, a task that traditional convolutional neural networks (CNNs) struggle with. These networks also require substantial computational resources and lengthy training periods, limiting their practical online application. In response to these challenges, we leverage HierAttn, a lightweight convolutional transformer model, and evaluate its performance on a custom pediatric dermatosis image dataset. Specifically, HierAttn is designed to classify nine types of hypopigmented skin disorders, effectively capturing both local and global features through a multi-stage attention mechanism and a multi-branch deep supervision strategy. Our experiments show that HierAttn offers superior diagnostic accuracy compared to other lightweight models, presenting a viable solution for use in resource-constrained healthcare settings.</p>
<p><b>AD330</b> <b>16:15-16:30</b></p>	<p>Comparative Analysis of Object Detection Algorithms for Bolt Detection: Performance Evaluation of Faster R-CNN, SSD, RetinaNet and YOLOv8n  <b>Authors:</b> Yan Kai Tan, Kar Mun Chin, Yeh Huann Goh, Tsung Heng Chiew, Terence Sy Horng Ting, Ge Ma and Chong Keat How  <b>Presenter:</b> Kar Mun Chin, Tunku Abdul Rahman University of Management and Technology, Malaysia</p> <p><b>Abstract:</b> This paper presents a comparative evaluation of four widely used object detection algorithms for bolt detection in robotic vision applications: Faster Region-based Convolutional Neural Network (Faster R-CNN), RetinaNet, Single Shot MultiBox Detector (SSD), and You Only Look Once version 8 nano (YOLOv8n). In the context of Industry 4.0, automating bolt and nut assembly in dynamic environments where fastener positions vary is increasingly critical. This study provides a comprehensive benchmarking of object detectors specifically for bolt detection while evaluating their performance under a small training dataset to simulate real-world industrial constraints. The robustness of these algorithms is further assessed across diverse conditions, including matched and metal backgrounds, occlusions, and varying camera distances. Results indicate that SSD and YOLOv8n excel in inference and training speed, while Faster R-CNN demonstrates superior accuracy on metal backgrounds. Overall, YOLOv8n achieves the best balance of accuracy and speed across all test datasets, with an inference speed of 0.0360 seconds and a mean average precision (mAP) of 53.488%. These findings provide</p>

	practical insights for selecting object detectors in robotic vision applications for assembly automation.
<b>AD355</b> <b>16:30-16:45</b>	<p>SkyCloud360: Sky and Cloud Segmentation in Equirectangular Images  <b>Authors:</b> Christoph Gerhardt and Wolfgang Broll  <b>Presenter:</b> Christoph Gerhardt, Ilmenau University of Technology, Germany</p> <p><b>Abstract:</b> Image-based sky and cloud segmentation is a promising area of research in computer vision, with applications ranging from climate monitoring and weather forecasting to outdoor scene understanding and autonomous navigation. Existing approaches focus on segmenting sky and cloud regions in natural images using supervised and semi-supervised learning techniques. These methods have demonstrated success in identifying clear skies, thin clouds, and thick clouds under controlled conditions. However, existing methods typically rely on perspective images, which only capture a limited field of view. The emergence of 360° imagery provides an opportunity to overcome these limitations by offering a complete spherical view of the environment. This panoramic perspective is particularly beneficial for applications requiring comprehensive spatial context, such as solar energy forecasting. However, processing equirectangular images introduces unique challenges due to geometric distortions, varying pixel densities across latitudes, and wrap-around boundaries at the edges of the projection. To address these challenges, we present a novel approach for sky and cloud segmentation in equirectangular images. As part of this work, we introduce a dataset comprising 600 high-resolution equirectangular images with dense annotations for sky and cloud segmentation. We evaluate multiple adaptations of the previously proposed SkyCloudNet architecture, including tangent plane projections and equirectangular convolutions, to handle the geometric distortions inherent in 360° imagery. Furthermore, we benchmark state-of-the-art methods for semantic segmentation in 360° images and unsupervised domain adaptation to explore their effectiveness in this domain. Our findings highlight the potential of advanced segmentation techniques to handle the complexities of equirectangular images while providing insights into the limitations and opportunities for future research in panoramic sky and cloud segmentation.</p>
<b>AD374</b> <b>16:45-17:00</b>	<p>High-Precision Human Pose Estimation Algorithm Based on Multi-View LiDAR and Visible Light Sensors  <b>Authors:</b> Yezhao Ju, Haiyang Zhang, Yuanji Li, Le Xin, Changming Zhao and Ziyi Xu  <b>Presenter:</b> Yezhao Ju, Beijing Institute of Technology, China</p> <p><b>Abstract:</b> To address the limitations in multi-person 3D pose estimation algorithms that either lack sufficient three-dimensional information when using visible light sensors or suffer from low resolution with LiDAR sensors, we have developed a system integrating multiple visible light and 3D LiDAR composite sensors. This setup facilitates the creation of a richly detailed, mutually calibrated, and synchronized human pose dataset. We propose an advanced top-down multi-person 3D pose estimation algorithm utilizing this integrated sensor system. By leveraging multi-view fused point clouds and multi-angle visible light data, our approach encompasses modules for human localization, multimodal data fusion, and joint keypoint positioning, achieving enhanced training and inference speeds alongside improved recognition accuracy. Furthermore, our network has been successfully transplanted and accelerated on NVIDIA's Jetson processors as well as Huawei's domestically produced Atlas 200 processor.</p>
<b>AD533</b> <b>17:00-17:15</b>	<p>A Reversible Grayscale Method Based on Bit-Field Multi-Channel Fusion Encoding  <b>Authors:</b> Teng Wang, Wei Pan, Yong Yang and Pascal Lefevre</p>





Presenter: Pascal Lefevre, Xi'an Jiaotong - Liverpool University, China

Abstract: Grayscale image representation is widely adopted to reduce computational complexity, optimize storage, and enhance transmission efficiency. However, restoring the original color information from grayscale images is highly desirable in many applications. Existing methods often suffer from high computational costs, limited interpretability, and dependence on training data. To address these limitations, we propose a novel non-learning-based approach that encodes RGB color information into a 16-bit grayscale image by strategically embedding HSV components, allowing high-fidelity recolorization while remaining lightweight and hardware-friendly. Experimental results validate the effectiveness of our approach in balancing reconstruction accuracy, computational efficiency, and practicality. The proposed method offers a promising alternative to traditional and neural network-based solutions, particularly in resource-constrained environments.



# Onsite Oral Session 2

**Topic:** Deep and Machine Learning Applications

**Session Chair:** Zhu Meng, Beijing University of Posts and Telecommunications, China

July 17th, 2025 | 15:20 - 17:10

**Venue:** Yutang Chun 1 - 2nd Floor | 二楼玉堂春 1 号

Invited Speaker-Hongping Gan, AD536, AD2012, AD2014, AD3016, AD3024, AD4035

**Invited  
Speaker  
15:20-15:40**



**Hongping Gan**  
 Northwestern Polytechnical University, China

*Speech Title: Deep Unfolding Network for Compressive Sensing*

**Abstract:** In the era of big data, compressive sensing (CS) has provided a revolutionary solution for efficient signal acquisition and reconstruction. However, traditional algorithms have long been constrained by bottlenecks such as complex manual parameter tuning, low computational efficiency, and insufficient real-time performance. In recent years, deep unfolding networks (DUN), by integrating the dual advantages of model-driven optimization and data-driven learning, have significantly improved reconstruction speed and accuracy, emerging as a cutting-edge breakthrough in the field of CS. Therefore, this report will analyze state-of-the-art DUN method for CS, explain several classic network architectures, and explore future development trends of this technology.

**Paper ID  
&Time**

**Presentation**

**AD536  
15:40-15:55**

**SCAF-YOLO: Multi-Scale Feature Fusion for Small Object Detection in Remote Sensing Images**

**Authors:** Wei Wang, Ziting Wang, Lina Huo, Qi Zhou, Hongxin Geng and Hanqian Niu

**Presenter:** Ziting Wang, Hebei Normal University, China

**Abstract:** Small object detection in remote sensing images is challenging due to their limited size, indistinct features, and high background similarity. To address these challenges, we propose a method called Spatial Context-Aware Feature Fusion YOLO (SCAF-YOLO). The main contributions of SCAF-YOLO are as follows: First, the Spatial Context Aware Module (SCAM) is adopted to enhance the global context representation of small objects in images. Second, the Adaptive Spatial Feature Fusion (ASFF) detection head is incorporated to replace the original detection structure of YOLOv5, improving multi-scale feature fusion and selection capabilities. Finally, a Multi-Scale Non-Local Attention (MS-GLCA) is adopted. It combines multiscale convolutions and channel attention to enhance feature representation via weighted fusion. This leads to

	more efficient feature weighting. The effectiveness of the proposed method is validated on two public remote sensing datasets, USOD and DOTA. Results show that the SCAF-YOLO model outperforms other state-of-the-art detectors in multi-object detection accuracy.
<b>AD2012</b> <b>15:55-16:10</b>	<p>Meta-Learning Based Cross-Domain Few-Shot Speaker Recognition  <b>Authors:</b> Yixuan Wang, Hongxia Bie  <b>Presenter:</b> Jiahe Yang, Beijing University of Posts and Telecommunications, China</p> <p><b>Abstract:</b> Speaker recognition technology distinguishes identities based on individual voice characteristics and has widespread applications in fields such as voice assistants and smart security. However, in cross-domain scenarios, there are significant differences between speakers of different languages, which poses a major challenge to the inference performance of models in the target language. To address this issue, a few-shot cross-domain speaker recognition method based on meta-learning is proposed. This method constructs N-way K-shot meta-tasks and leverages differences between meta-tasks during training to mitigate the mismatch between the source and target domains. As a result, the model's inference performance in cross-domain scenarios (trained on an English dataset and tested on a Chinese dataset) is significantly improved, achieving a 20% to 40% increase in recognition accuracy. At the same time, it maintains a high accuracy of approximately 98% in non-cross-domain scenarios.</p>
<b>AD2014</b> <b>16:10-16:25</b>	<p>Fine-Grained and Controllable Speech Synthesis Based on Disentangled Latent Representation in StyleGAN  <b>Authors:</b> Peng Wang and Hongxia Bie  <b>Presenter:</b> Jiahe Yang, Beijing University of Posts and Telecommunications, China</p> <p><b>Abstract:</b> In this paper, we implement two fine-grained and controllable speech synthesis tasks - speaker conversion and attribute editing - based on disentangled latent representation in StyleGAN space. An encoder is designed, which aims to map the HuBERT features of real speech to space in StyleGAN, while achieving implicit disentanglement of features and attributes with a small loss of reconstruction quality. Implicit disentanglement means that the features of each dimension are independent of each other and each attribute is linearly separable. In order to train the encoder on general long-term complex data, a bidirectional self-supervised representation learning architecture is constructed to jointly train the encoder and StyleGAN. Speaker conversion is achieved by inputting the disentangled representation into different style blocks of the generator for mixing. The attribute editing direction is found through SVM classification, and the disentangled representation vector is translated along this direction to achieve attribute editing. Fineness is reflected in the ability to achieve smooth and linear transitions in conversion or editing through vector interpolation in the disentangled latent space. Speaker conversion experiments conducted on the LibriSpeech and VCTK datasets demonstrate that the proposed method outperforms FreeVC in terms of content preservation and speaker conversion. Furthermore, to provide a preliminary evaluation of the effectiveness of attribute editing, we analyze variations in fundamental frequency (F0) during gender editing and changes in speech quality during noise editing.</p>
<b>AD3016</b> <b>16:25-16:40</b>	<p>Neural network model of point defect microcavities in two-dimensional silicon-based dielectric pillar photonic crystals  <b>Author:</b> Junhua Chi</p>

	<p><b>Presenter:</b> Jun Hua, Chi, Hangzhou Dianzi University, China</p> <p><b>Abstract:</b> In this study, a neural network-based modeling approach is proposed to address the need for optimal design of point-defect microcavities for two-dimensional silicon-based dielectric column photonic crystals. Photonic crystals have shown great potential for applications in the fields of optics, optoelectronics and photonic integration since their introduction. However, the electron beam etching and ultraviolet etching processes for photonic crystals have long preparation cycles and high maintenance costs. Therefore, the design needs to be optimized before preparation to reduce the cost of trial and error. In this study, two-dimensional photonic crystal energy band data are calculated by PWE algorithm, and a mathematical model of the relationship between photonic crystal energy band data and microcavity size is established by BP neural network. This can effectively overcome the problem of large computational consumption brought by traditional numerical calculation methods.</p>
<p><b>AD3024</b> <b>16:40-16:55</b></p>	<p>Technology Opportunity Analysis Based on Deep Learning and Explainable Artificial Intelligence Model</p> <p><b>Authors:</b> Yingqi Xu, Xian Zhang, Yi Xu</p> <p><b>Presenter:</b> Yingqi Xu, National Science Library(Chengdu), Chinese Academy of Sciences, China</p> <p><b>Abstract:</b> This study introduces an integrated framework combining deep learning and explainable artificial intelligence (XAI) for systematic technology opportunity discovery. Technology Opportunity Analysis (TOA) is operationalized as a data-driven process that identifies emerging technological themes through multi-modal analysis of scientific artifacts. Our methodology employs lithium-ion battery patents filed over the past three years as empirical evidence, with Derwent patent titles being processed through Biterm Topic Modeling (BTM) to address short-text analytical challenges while optimizing input dimensions for subsequent classification tasks. A hybrid architecture incorporating four deep learning classifiers demonstrates patent categorization effectiveness, from which Shapley Additive Explanations (SHAP) analysis reveals critical decision-driving features—specifically those technology themes statistically significant for patent authorization outcomes. Empirical validation confirms the framework's capability in uncovering actionable technological opportunities within the lithium-ion battery sector.</p>
<p><b>AD4035</b> <b>16:55-17:10</b></p>	<p>Gap-Guided Evolutionary Deep Reinforcement Learning for Hierarchical Navigation in Structured Dynamic Environments</p> <p><b>Author:</b> Ketan Anand</p> <p><b>Presenter:</b> Ketan Anand, Georgia Institute of Technology, USA</p> <p><b>Abstract:</b> Multi-agent collision avoidance—where an ego-robot must navigate from a start pose to a goal pose while avoiding both static obstacles and dynamic agents—is a longstanding challenge in motion planning. Recently, reinforcement learning (RL) has been explored for this task, but most RL-based solutions address simplified scenarios, often omitting complex static obstacles and focusing primarily on dynamic agent avoidance. Consequently, these approaches are unsuitable for structured environments such as mazes or indoor layouts with significant static geometry. To address this limitation, we integrate a multi-agent collision avoidance policy—specifically the GPU-accelerated Asynchronous Advantage Actor-Critic for Collision Avoidance (Evolution GA3C-CADRL) within a hierarchical navigation framework. Our method incorporates a local waypoint planner that</p>

leverages global navigation plans and real-time sensor data to generate short-term waypoints, ensuring clear paths devoid of static obstacles between the agent and the next target. This architecture effectively mitigates the original model’s inability to handle static constraints. The full system integrates global planning, perception-informed gap-based waypoint generation, and deep RL for local control. We empirically evaluate our hierarchical planner across five structured environments and demonstrate that the integrated waypoint + RL approach consistently reduces collision rates and improves goal-reaching success compared to the standalone RL model. Results show substantial improvements in both navigation robustness and success rate under realistic, complex scenarios.





# Onsite Poster Session 1

**Topic:** Intelligent Image Detection, Recognition and Image Modeling

**Session Chair:** Tao Zhang, Shanghai Jiao Tong University, China

July 17th, 2025 | 15:20-17:10

**Venue:** Yutang Chun 2 - 2nd Floor | 二楼玉堂春 2 号

AD209, AD418, AD385, AD220, AD342, AD344, AD221, AD225, AD409, AD548, AD356, A  
D333, AD380, AD396, AD398, AD359, AD3022

Paper ID	Presentation
AD209	The New Tai Le Character Recognition System Based on ModelArts Platform <b>Authors:</b> Haoyuan Gan, Yuechao Zhao, Yuming Zhang, Yong Yang, Hao Ge, Ge Peng <b>Presenter:</b> Haoyuan Gan, Dehong Vocational College, China
AD418	Tree-Structure Transformer for Skeleton-based Human Action Recognition <b>Authors:</b> Xiuxiu Li, Huanhuan Guo, Pu Zhang and Wanqing Pei <b>Presenter:</b> guohuanhuan, Xi'an University of Technology, China
AD385	A Knowledge Distillation Based Binarized Separable Convolutional Neural Network for Underwater Acoustic Target Recognition <b>Authors:</b> Hanlin Cui, Xiaohui Chu, Guanqing Li, Xiaofei Dai <b>Presenter:</b> Hanlin Cui, Beijing Institute of Technology, China
AD220	Knowledge Distillation from First-Order Representation for Visual State Space Model <b>Authors:</b> Conghu Li, Wenxuan Wang, Ang Li <b>Presenter:</b> Conghu Li, University of Chinese Academy of Sciences, China
AD342	Compression and Rendering of Time-varying Interplanetary Volumes <b>Authors:</b> Lili Bao, Yanxia Cai, Rui Wang, Yuan Guo, Wei Zheng <b>Presenter:</b> Lili Bao, National Space Science Center, Chinese Academy of Sciences, China
AD344	Exploiting Frequency Correlation for Hyperspectral Image Reconstruction <b>Authors:</b> Muge Yan, Lizhi Wang <b>Presenter:</b> Muge Yan, Beijing Institute of Technology, China
AD221	Evaluating Text-to-Video Alignment: A Hierarchical Benchmark for Video Generation Models <b>Author:</b> Ang Li <b>Presenter:</b> Ang Li, UCAS, China
AD225	A Point Cloud Stitching Approach Based on Image Registration for High-Precision Threaded Surface Modeling in Multi-View 3D Imaging <b>Authors:</b> Wei Luo, Wenzhe Li, Pengfei Sang, Jiang Ma, Zihan Ma, Rui Ni, Wei Zhao and Xiaohai He <b>Presenter:</b> Jiang Ma, Sichuan University, China



<b>AD409</b>	Optimal Mixture Model Distribution Alignment-based 3D-2D Gaussian Splatting Registration for Monocular Endoscopic AR Guidance <b>Authors:</b> Ziang Zhang, Shubo Fan, Hong Song, Jingfan Fan, Tianyu Fu, Danni Ai, Deqiang Xiao, Yuanyuan Wang, Yucong Lin, Long Shao and Jian Yang <b>Presenter:</b> Ziang Zhang, School of Medical Technology Beijing Institute of Technology
<b>AD548</b>	Improved Repeat and Concatenate: A More Effective 2D X-ray to 3D CT Image Translation Model <b>Authors:</b> Haolang Li, Nan Jiang, Chang Liu, Long Lan <b>Presenter:</b> Li HaoLang, National University of Defense Technology, China
<b>AD356</b>	A Real-time Detection Method for Surface Defects in 3D Printing Based on YOLOv12 Algorithm <b>Authors:</b> Xiaoqian Li,Chenglei Zhang, Bo Xu, Jiajia Liu, Liang Yao <b>Presenter:</b> Xiaoqian Li, Linyi University, China
<b>AD333</b>	LViT-GMMs: Semantic Segmentation for Maritime Object Detection <b>Authors:</b> Xiaoting Zhao, Yulong Qiao <b>Presenter:</b> Xiaoting Zhao, Harbin Engineering University, China
<b>AD380</b>	Adaptive Multi-feature Fusion Algorithm for Ship Rust Detection on Coating Surfaces <b>Authors:</b> Bocheng Feng, Zhenqiu Yao and Chuanpu Feng, <b>Presenter:</b> Bocheng Feng, Jiangsu University of Science and Technology, China
<b>AD396</b>	Iterative Similarity Perturbation Point Cloud Registration based on Deformation-Resistant Region Detection <b>Authors:</b> Yifei Yang, Sifan Cao, Long Shao, Jingfan Fan and Jian Yang <b>Presenter:</b> Yifei Yang, Beijing Institute of Technology, China
<b>AD398</b>	Integrating Knowledge for High-Fidelity Remote Sensing Detection of Cross-River Bridges <b>Authors:</b> Liang Chen, Yuhai Qi, Ruiqi Sun, Haohao Zhou, Xiuwei Zhang <b>Presenter:</b> Ruiqi Sun, Information Center of Yellow River Conservancy Commission, China
<b>AD359</b>	A Calibration-Free Method for Large-View Classroom People Counting with Object Detection-Based Structure Matching <b>Authors:</b> Hen Wang, Juan Jiang, Yongdong Li, Guoan Cheng, Guanghao Yuan,Shengke Wang <b>Presenter:</b> Guoan Cheng, Qingdao Harbour Vocational & Technical College, China
<b>AD3022</b>	Application of Deep Learning Techniques in Anomaly Detection and Process Improvement for Electronic Chips <b>Authors:</b> Guoying Wang, Zhen Song, Chunmei Pei, Shuo Wang, Junfeng Shi, Youlan Wu, Gongsheng Zhu <b>Presenter:</b> Guoying Wang, Beijing Polytechnic University,China

# Onsite Oral Session 3-1

**Topic:** Computer Vision Techniques and Application

**Session Chair:** Yunbo Wang, Central South University, China

July 18th, 2025 | 10:00 - 11:50

**Venue:** Xinagqing Hall - 2nd Floor | 二楼祥庆厅

Invited Speaker-Yuan Cao, AD217, AD352, AD367, AD336, AD537, AD3020

<p><b>Invited Speaker</b> <b>10:00-10:20</b></p>	<div data-bbox="512 667 743 898">  </div> <div data-bbox="743 734 1217 835"> <p><b>Yuan Cao</b> <b>Ocean University of China, China</b></p> </div> <div data-bbox="467 936 1315 976"> <p><i>Speech Title: Hashing based Large-scale Multimedia Retrieval</i></p> </div> <div data-bbox="306 1030 515 1064"> <p><b>Abstract: TBA</b></p> </div>
<p><b>Paper ID &amp;Time</b></p>	<p><b>Presentation</b></p>
<p><b>AD217</b> <b>10:20-10:35</b></p>	<p>DSCViTANet: A Hybrid Depthwise Separable Convolution and Vision Transformer for Early Alzheimer's Classification  <b>Authors:</b> Nour-EL Houda Jraibi , Hui Zeng  <b>Presenter:</b> NOUR-ELHOUDA JRAIBI, Southwest University of Science and Technology, China</p> <p><b>Abstract:</b> Alzheimer's disease (AD) is a progressive neurode- generative disorder that deeply affects cognitive abilities, making early classification vital for better patient outcomes. Although several methods have been suggested for early AD diagnosis, many conventional models are limited by low accuracy, high com- putational costs, and a lack of transparency in decision-making. To overcome these limitations, we propose DSCViTANet, a novel hybrid model that combines the</p>

	<p>several deep learning network architectures and these are depthwise separable convolutions (DSC), vision transformers (ViTs), and attention mechanisms. Our main purpose to build this model is not only to achieve high accuracy but also to provide transparency in its decision-making process. We utilized DSC because of its computational costs reduce ability through more efficient convolutions, while ViTs have the ability to enhance feature extraction by capturing long-range dependencies in images. The Attention mechanism improves both performance and interpretability, which is especially important for clinical use. Our model was tested on public dataset, with 98.84% accuracy, 98.81% precision, 98.87% recall, and 98.82% F1 score, surpassing current methods. DSCViTANet also can be use in other's medical image classification task.</p>
<p><b>AD352</b> <b>10:35-10:50</b></p>	<p>GLNet-YOLO: Research on Pedestrian Detection Technology Based on Multimodal Feature Fusion  <b>Authors:</b> Qing Zhao, Yi Zhang, Xiuhe Li, Jinhe Ran, Zhen Zhang, Guoqiang Zhu, Yining Huo, Yu Zhang, Hongyi Yu  <b>Presenter:</b> Qing Zhao, National University of Defense Technology, China</p> <p><b>Abstract:</b> Pedestrian detection holds great significance in various computer vision applications, including intelligent surveillance, autonomous driving, and robot navigation. However, relying solely on single-modal images usually fails to achieve highly accurate detection in complex environments. To overcome this drawback, this study presents GLNet-YOLO, a deep feature fusion framework. The aim is to improve pedestrian detection performance by integrating the features of visible and infrared images. The GLNet-YOLO framework modifies YOLOv11 by adopting a two-branch network architecture. One branch processes visible images, and the other deals with infrared images. To better handle feature fusion, it incorporates the Feature Mixer (FM) module, which is responsible for global feature fusion and enhancement. Additionally, the Dual-Modality Refiner (DMR) module is introduced to separate and interact with local features, thus optimizing the feature fusion procedure.</p> <p>Experimental results in the LLVIP dataset demonstrate that GLNet-YOLO improves mAP@50 by 9.2% for visible images and 0.7% for infrared images compared to the original single-modal YOLOv11. The proposed framework significantly improves the detection accuracy under low light and complex background conditions, improving the robustness of the algorithm.</p>
<p><b>AD367</b> <b>10:50-11:05</b></p>	<p>Binocular Vision-Based Infrastructure Crack Measurement with Morphological Union Enhancement  <b>Authors:</b> Qing Lin Chen, Yuan Fei Gu, Si Yu Tang, Hao Yu Huang  <b>Presenter:</b> Qinglin Chen, East China Jiaotong University, China</p> <p><b>Abstract:</b> Although extensive research has been conducted in the field of infrastructure crack detection, there is still a lack of a universal crack measurement method applicable to complex environments. This study proposes a crack maximum width measurement algorithm based on binocular vision, introducing crack morphological union processing technology, which lays a solid foundation for the complete extraction of crack skeleton morphology and the accurate measurement of crack maximum width. The algorithm has been tested and analyzed on crack datasets of various infrastructures, such as tunnels and building exteriors. The results show that the proposed algorithm exhibits good adaptability and high precision in crack measurement under different environments, especially in the measurement of fine cracks, where it significantly overcomes the limitations of traditional binocular stereo matching methods in terms of feature point dependence and measurement accuracy. This study provides a new approach and feasible framework for crack detection and quantitative analysis of infrastructure in different environments, expanding the research and practice directions in this field.</p>



<p><b>AD336</b> <b>11:05-11:20</b></p>	<p>Automatic Segmentation of Metaplasia in an Endoscopic Decision Support System  <b>Authors:</b> N. Obukhova, A. Motyko, A. Savelev, O. Saveleva, A. Samarin, J. Sigaeva  <b>Presenter:</b> Alexandr Motyko, Saint Petersburg State University, Saint Petersburg Electrotechnical University "LETI", Russia</p> <p><b>Abstract:</b> The goal of the research is to develop a neural network model for the automatic segmentation of images as a part of implementing a decision support system function in a video endoscopic system. The development is based on the use of modern artificial intelligence technologies, specifically neural networks, which enable the automatic segmentation of metaplasia in endoscopic images. Metaplasia, being a precancerous condition, requires special attention from physicians, and the availability of a tool capable of supporting them during examinations is crucial for improving diagnostics and enhancing the level of medical care. As part of the research, a database of images was collected, an original network architecture was proposed, and a neural network model was created, which achieved the target results in metaplasia segmentation. The conducted studies demonstrated that the developed model outperforms modern counterparts in terms of accuracy metrics. The use of the research results accelerates the data analysis process during video endoscopic examinations, reduces the likelihood of medical errors, and represents an important step toward improving clinical outcomes for patients.</p>
<p><b>AD537</b> <b>11:20-11:35</b></p>	<p>Human Eye Optics Simulation and Visual Modeling of Myopia Correction  <b>Author:</b> Wei Wang,Ziyao Li, Lina Huo,Ke Wang,Xueyang Wang, Jiacheng Wang  <b>Presenter:</b> Ziyao Li,Hebei Normal University,China</p> <p><b>Abstract:</b> Visualizing how light forms images in the human eye is essential for understanding refractive mechanisms and guiding clinical applications. We propose a real-time human eye imaging simulation system developed in Unity, which achieves a balance between physical accuracy and interactive performance. Light propagation through ocular media is modeled using Snell's law, and retinal blur is simulated with a geometry-driven diffusion approach. The framework supports the simulation of refractive errors, such as myopia, as well as refractive surgery. Experimental results confirm the system's effectiveness in visualizing image quality variations and postoperative outcomes. These results suggest that the system not only maintains high visual coherence but also holds great promise for ophthalmic education and research.</p>
<p><b>AD3020</b> <b>11:35-11:50</b></p>	<p>Optimization of Blood Glucose Prediction Models for Diabetes Based on Sample Data Volume Variability  <b>Authors:</b> Feiyang Li, lijuan ren, Yaoxi Liu, Yunyu Rao  <b>Presenter:</b> Lijuan Ren, Chengdu University of Information Technology, China</p> <p><b>Abstract:</b> The prediction of blood glucose levels in diabetic patients is challenged by significant variability in individual time-series data volumes. Traditional Long Short-Term Memory (LSTM) models tend to overfit on small samples, limiting their predictive accuracy and generalization. To address this, we propose a segmented training strategy based on sample length, dynamically selecting model structures and training methods. We further introduce DAFNet (Data-Adaptive Forecasting Network), an ensemble model that combines the strengths of LSTM and Gated Recurrent Unit (GRU) networks, adaptively adjusting configurations to suit data volume. Experimental results demonstrate that DAFNet outperforms the standard LSTM model, achieving lower prediction errors. Notably, it reduces MAE and MSE to 6.51 and 80.30, respectively, with consistent improvements in RMSE and <math>R^2</math>, confirming its effectiveness for personalized glucose forecasting.</p>





# Onsite Oral Session 3-2

**Topic:** Computer Vision Techniques and Application

**Session Chair:** TBA

July 18th, 2025 | 10:00 - 11:50

**Venue:** Xiangtai Hall - 2nd Floor | 二楼祥泰厅

Invited Speaker-Hui Liu, AD347, AD390, AD392, AD4029, AD222, AD413

<p><b>Invited Speaker</b>  <b>10:00-10:20</b></p>	<div data-bbox="427 698 657 927">  </div> <div data-bbox="667 770 1353 860"> <p><b>Hui Liu</b>  <b>Kunming University of Science and Technology, China</b></p> </div> <div data-bbox="322 949 1449 1016"> <p><i><b>Speech Title:</b>Difficulty analysis and technology summary and outlook in the process of end-point control of converter steelmaking</i></p> </div> <div data-bbox="322 1070 1449 1438"> <p><b>Abstract:</b> The endpoint determination of converter steelmaking is an important and critical step in the blowing process, and its difficulty lies in achieving accurate real-time measurement of carbon content and temperature in the melt pool. Starting from the production process of converter steelmaking, traditional manual endpoint determination method, contact sensor method, detection theory and method based on spectral radiation, carbon temperature regression method based on flame image processing and recognition, and data-driven endpoint carbon temperature soft measurement modeling method, this paper summarizes the research content and ideas of existing methods, analyzes the content that still needs further improvement, and provides future research and development directions in this field, hoping to play a role in further research and development in this field.</p> </div>
<p><b>Paper ID &amp;Time</b></p>	<p><b>Presentation</b></p>
<p><b>AD347</b>  <b>10:20-10:35</b></p>	<p>Research on Semantic Communication based on Balancing of Task Distortion  <b>Authors:</b> Hong Yang, Honggang Chen, Qizhi Teng, Xiaohai He  <b>Presenter:</b> Hong Yang, Sichuan University,China</p> <p><b>Abstract:</b> With the development of the Artificial Intelligence (AI) and communication technology, how to improve communication efficiency for the semantic communication system in image reconstruction and image classification tasks, meanwhile how to balance these two types of tasks is a hot research topic in the field of semantic communication. This article proposes a semantic communication framework that balances task distortion based on Grad CAM technology, obtains the optimal semantic features that can balance reconstruction quality and classification accuracy, and further compresses the semantic features based on task relevance. Finally, the framework was used in the course design of "Digital Image Communication" and</p>



	<p>verified that the proposed algorithm can balance image reconstruction and image classification tasks while further reducing the amount of semantic data transmitted.</p>
<p><b>AD390</b> <b>10:35-10:50</b></p>	<p><b>FPD: Fringe Photometric Deflectometry When Fringe Meets Photometric Stereo</b>  <b>Authors:</b> Ling Cao, Teng Wang, Yong Yang, Quan Tang, Zi Meng Wang and Wei Pan  <b>Presenter:</b> Wei Pan, OPT Machine Vision, China</p> <p><b>Abstract:</b> We present Fringe Photometric Deflectometry (FPD), a novel approach that synergistically combines Phase Measuring Deflectometry (PMD) and Photometric Stereo (PS) to achieve high-precision, full-field 3D surface reconstruction across both specular and diffuse regions. Leveraging PMD's accurate phase measurement for shiny surfaces and PS's detailed normal estimation for matte areas, we introduce an adaptive fusion mechanism that weights each modality according to local surface reflectance. We provide a rigorous theoretical formulation of the fusion process and validate its robustness through experiments on industrial test specimens. Our results demonstrate that FPD surpasses standalone PMD and PS in reconstruction accuracy and enhances defect detection capabilities, effectively revealing scratches, smudges, and other fine surface anomalies. FPD has been integrated into commercial inspection software, howcasing its practical utility.</p>
<p><b>AD392</b> <b>10:50-11:05</b></p>	<p><b>A robust 3D watermarking method based on statistical features</b>  <b>Authors:</b> Weijun Phoon, Bin Zhang, Liang Xie, Shuhao Wang, Xiaoxi Zhu and Xingjun Wang  <b>Presenter:</b> PHOON WEI JUN , Tsinghua University Shenzhen International Graduate School, China</p> <p><b>Abstract:</b> This paper proposes a robust 3D watermarking method based on statistical features for the copyright protection of 3D models. It initiates with a multi-level coordinate system decomposition of the original model, enabling synchronized watermark embedding across hierarchically structured feature sub-blocks. For each sub-block, it analytically computes the radial distance distribution across all constituent vertices, followed by stability-optimized partitioning regulated according to the specified watermark bit-length. After this, it uses a precision optimization protocol to recalibrate mean radial distances in unstable domains, systematically enhancing discriminative geometric features while maintaining structural integrity. At last, it encodes the resultant amplified geometric signatures into feature sequences, facilitating seamless watermark integration through structured mapping. Many experiment results demonstrate the proposed method obtains good robustness against geometric perturbations. Comparative analyses further underscore its superiority over some state-of-the-art techniques in resisting various attacks.</p>
<p><b>AD4029</b> <b>11:05-11:20</b></p>	<p><b>A Transformer-based Deep Learning Model to Enhance Hope Speech Detection</b>  <b>Authors:</b> Nawal Bint Masood , Saeid Pourroostaei Ardakani, Miao Yu  <b>Presenter:</b> Miao Yu, University of Lincoln, UK</p> <p><b>Abstract:</b> As the volume of information and communication increases on the internet there has been great effort in the reduction of negatively focusing or abusive materials. In as much as negative communication is avoided and controlled, there is need to also look for positive content and promotion of such contents as well. The objective of this research is to study "hope speech" that manifests itself where multilingual and imbalance datasets are available. To that end, we present a machine learning-enabled system that employs the transformer-based model DeBERTa-V3-small to categorise social media texts as hope speech and non-hope speech classes after conducting a rigorous preprocessing and random oversampling to manage imbalance data.</p>

	<p>According to the evaluation results, our proposal outperforms two well-known benchmarks including BERT and Random Forest. It points to the effectiveness of transformer-based approach DeBERTa-V3 in strengthening the positive discourse and brings insightful prospects for the future studies of identifying and promoting hope speech online.</p>
<p><b>AD222</b> <b>11:20-11:35</b></p>	<p>Adaptive Part Shifting for Fine-grained Ship Classification in Remote Sensing Images  <b>Authors:</b> Yan Ma, Xuesong Yang, Wenyu Ma  <b>Presenter:</b> Yan Ma, University of Chinese Academy of Sciences, China</p> <p><b>Abstract:</b> Fine-grained ship classification of remote sensing images faces challenges from high inter-class similarity and intra-class part variance. Traditional methods struggle to capture object-level discriminative features among classes while failing to leverage crucial local information and lacking interpretability in feature representation. This paper proposes an Adaptive Part Shifting (APS) method, which utilizes and iteratively refines spatial focused and semantic discriminative part features for fine-grained ship classification. APS integrates three modules: (i) Part Discovery module to decompose objects into parts in a weakly supervised manner, (ii) Part Weighting module to adaptively enhance the discriminative parts, and (iii) Part Shifting module to refine the discovered parts against background noise or parts homogenization. Experiments on the FGSC-23 and FGSCR-42 datasets demonstrate that APS achieves 91.68% and 99.56% accuracy using ViT_Base, surpassing baseline models by 3.23% and 5.00%. Visualizations further confirm its focus on discriminative regions with human-aligned reasoning, establishing a high-precision and interpretable framework for remote sensing fine-grained ship analysis.</p>
<p><b>AD413</b> <b>11:35-11:50</b></p>	<p>AMF-UNet: A Lightweight Adaptive multi-Mamba Fusion U-shaped Network for Medical Image Segmentation  <b>Authors:</b> Youyang Tao, Hongmin Deng, Tongmeng Yong  <b>Presenter:</b> Youyang Tao, Sichuan University, China</p> <p><b>Abstract:</b> Medical image segmentation remains challenging due to limitations in suppressing background interference, inefficient fusion of local-global features, and excessive computational costs in existing methods. This paper proposes a lightweight adaptive multi-Mamba fusion U-shaped Network, called AMF-UNet, which synergies the local feature extraction capability of convolutional neural networks (CNNs) with the global context modeling of state space models (SSMs). The architecture introduces two novel components: 1) a Mamba fusion module (MFM) utilizing a parallel Mamba architecture that reduces parameters by 99% compared to conventional Mamba designs, i.e. VMamba while maintaining long-range dependency capture, and 2) an adaptive weight decoder module (AWDM) with a dynamic channel-aligned attention mechanism to optimize feature fusion between skip connections and upsampled high-level semantics. Evaluated on the ISIC2017 skin lesion dataset, AMF-UNet achieves good performance with only 0.43M parameters and 2.78 GFLOPs—outperforming 13 state-of-the-art (SOTA) methods including UNet variants, Transformer-based models, and recent Mamba architectures.</p>

# Onsite Poster Session 2

**Topic:** AI Based Digital Image Analysis and Processing Technology

**Session Chair:** TBA

July 18th, 2025 | 10:00-11:50

**Venue:** Yutang Chun 1 - 2nd Floor | 二楼玉堂春 1 号

AD351, AD538, AD337, AD211, AD403, AD341, AD226, AD404, AD519, AD526, AD368, A  
D400, AD532, AD383, AD547, AD1001, AD2015, AD3021

Paper ID	Presentation
AD351	Secret Point Recognition Algorithm via Test-Time Augmentation Based on Large Language Models <b>Authors:</b> ZhenDong Wu, Hu Li, Liang Zhang, JiaoJiao Li <b>Presenter:</b> Liang Zhang, Hangzhou Shiping Information&Technology Co.,Ltd, China
AD538	Masked Face Recognition Method with ArcFace Fusion of Attention and Focal Loss <b>Authors:</b> Lu Huang, Mingwei Chen, Xuan Liu <b>Presenter:</b> Lu Huang, School of Artificial Intelligence and Big Data, Guangdong Business and Technology University, China
AD337	A Multi-Scale Information-Driven Rock Classification Algorithm Based on Enhanced ResNet <b>Authors:</b> Liu Yan, Jia Junyang and Liu Yidong <b>Presenter:</b> Liu Yan, Zhengzhou University of Light Industry, China
AD211	Low-Light Image Enhancement Algorithm Based on Information Fusion Strategy <b>Authors:</b> Boxin Yin, Wencheng Wang, Lei Li, Xiaowei Lan, Lun Li, Xiaojin Wu <b>Presenter:</b> Wencheng Wang, Weifang University, China
AD403	Stokes-S0 Prior-Guided Dual-Branch Network for Polarized Image Enhancement <b>Authors:</b> Tianhe Yu, Yan Wang, Xinran Wei <b>Presenter:</b> Tianhe Yu, Beihang University, China
AD341	SBFS-Net: Smoke Segmentation via Separated Smoke and Background Features <b>Authors:</b> Zifan Mo, Yihui Liang, Kun Zou, Wensheng Li, Fujian Feng and Han Huang <b>Presenter:</b> Zifan Mo, University of Electronic Science and Technology of China, China
AD226	An Efficient Method for Measuring Oil Casing Thread Geometric Parameters Using Point Cloud Data <b>Authors:</b> Xueqiang Wang, Chuanlei Wang, Wei Luo, Pengfei Sang, Haiqing Long, Jing Huang, Jun Shu and Xiaohai He <b>Presenter:</b> Shu Jun, Sichuan University, China
AD404	One-to-Many Fine-grained matching between UAV images and satellite images for UAV Self-Localization

	<b>Authors:</b> Jiaqi Li, Yuli Sun, Yaobing Xiang, Lin Lei <b>Presenter:</b> Jiaqi Li, National University of Defense Technology, China
<b>AD519</b>	Unveiling Histopathological Features of Breast Cancers using Limited Data <b>Authors:</b> Kaiyue Zhou, Bhagya Shree Kottoori, Seeya Awadhut Munj, Guangyu Dong, Suzan Arslanturk and Shengjin Wang <b>Presenter:</b> Kaiyue Zhou, Tsinghua University, China
<b>AD526</b>	Multi-modal Cooperative Distillation for Zero-shot Multi-label Classification <b>Authors:</b> Yiqin Wang, Ying Chen <b>Presenter:</b> Yiqin Wang, Jiangnan University, China
<b>AD368</b>	Artificial Intelligence in CT-Based Diagnosis of Small Pulmonary Nodules: Current Applications and Future Perspectives <b>Authors:</b> Yuanchunsu Tan <b>Presenter:</b> Yuanchunsu Tan, Department of Radiology, The Third People's Hospital of Chengdu, China
<b>AD400</b>	An Improved 3DUNet+ with Inter-slice Difference Awareness for Pulmonary Vessel CT Image Segmentation <b>Authors:</b> Yihong Wang, Tingyan Wang <b>Presenter:</b> Yihong Wang, Hohai University, China
<b>AD532</b>	A Diabetes Screening Algorithm Embedded with Inception Deep Convolution in Swin Transformer <b>Authors:</b> Mingxia Xiao, Shidong Fang, Fei Wang <b>Presenter:</b> Shidong Fang, North Minzu University, China
<b>AD383</b>	From Precomputed Particle Shading to Volumetric Atmospheric Cloud Rendering for Real-Time Gaming: Methods and Advances <b>Authors:</b> Guanyu Chen, Yitian Wang, Anbo Xu <b>Presenter:</b> Guanyu Chen, The University of Sydney, Australia
<b>AD547</b>	Adaptive Far-field Region of Interest Extraction and Its Applications for Long-range Ground Surveillance <b>Authors:</b> Xingxin Li, Yutong Jiang, Hao Li, Jiahe Tian, Yiding Liu, Wentao Wu <b>Presenter:</b> Xingxin Li, China North Vehicle Research Institute, China
<b>AD1001</b>	VMD-Enhanced Temporal Convolutional Networks with Time-Feature Attention for High-resolution PM2.5 Forecasting <b>Authors:</b> Yankun Li, Fang Chen, Dongliang Xiao, Hongling Shi <b>Presenter:</b> Yankun Li, China Agricultural University, China
<b>AD2015</b>	Fourier Detail Injection Implicit Neural Fusion Network for Pansharpening <b>Authors:</b> Ze-Zheng He, Hong-Xia Dou, Yu-Jie Liang <b>Presenter:</b> Ze-Zheng He, Xihua University, China
<b>AD3021</b>	Attention Mechanism Enhanced Joint Multi-region Power Demand Prediction <b>Authors:</b> Shijun Luo, Renjie Xiao, Lu Tang, Chenjun Yang, Jiaxin Liu, Liqiang Zhao <b>Presenter:</b> Jiaxin Liu, Xidian University, China



# Onsite Oral Session 3-3


**Topic:** Computer Vision Techniques and Application

**Session Chair:** TBA

July 18th, 2025 | 13:30 - 15:20

**Venue:** XiangHua Hall - 3nd Floor | 三楼祥华厅

Invited Speaker-Chen Li, AD389, AD387, AD223, AD4032, AD350, AD550

<p><b>Invited Speaker</b> <b>13:30-13:50</b></p>	<div data-bbox="501 667 730 896"></div> <p><b>Chen Li</b> <b>North China University of Technology, China</b></p> <p><i><b>Speech Title:</b> Annotations free survival prediction with WSIs</i></p> <p><b>Abstract:</b> Survival prediction of cancer patients has always been an challenging problem. Tumor microenvironment(TME) Analyzation based on whole-slide-images(WSIs) has provide an effective perspective for survival prediction. However most existing TME analyzation based on cell segmentation or classification relies heavily on labor-intensive cell-level annotations of pathologists. Furthermore, except for each individual cell or local pathological feature, survival prediction also involves local-level pathological features interactions in tumor microenvironments. This requires context-awareness based on histological features to fully infer the patient's survival risk. Therefore, we explored a model based on graph convolutional neural networks(GCNN) to perform survival prediction of cancer patients using WSIs. The model leverages the advantages of graph structures to autonomously learn the histopathological contextual features in WSIs, and therefore can incorporate additional and crucial tumor microenvironment interaction information while avoiding the labor-intensive annotations, which making it an effective supplementary diagnostic tool for oncologists and pathologists.</p>
<p><b>Paper ID &amp;Time</b></p>	<p><b>Presentation</b></p>
<p><b>AD389</b> <b>13:50-14:05</b></p>	<p>Volume Measurement Technology of Dispensing Transparent Adhesives Based on Line Laser Scanning <b>Authors:</b> Ling Cao, Renjie Zhou, Xinhua Wang and Wei Pan <b>Presenter:</b> Ling Cao, Shenzhen University, China</p> <p><b>Abstract:</b> Accurate volume measurement of transparent adhesives is essential for controlling the dispensing process, calibrating adhesive discharge, and enhancing both product quality and manufacturing efficiency. We introduced a complementary spectral background optimization (CSBO) method to enhance three-dimensional (3D) reconstruction accuracy by attenuating transmitted light, thereby reducing internal scattering and reflection</p>



	<p>interference without requiring additional measurement tools or equipment. In addition, we proposed an improved gray-scale center of gravity extraction algorithm by utilizing a unique candidate points mechanism and the novel dynamic weighted multi-factor scoring constraints (DWMF-SC) method for effectively candidate points screening.</p> <p>The proposed method was experimentally validated through various experiments. The experimental results demonstrate that the proposed laser center line extraction method outperforms traditional methods with higher precision and robust anti-interference properties, while the CSBO approach achieved the lowest relative error ratio reaching about 1.40% compared to ground truth values.</p>
<p><b>AD387</b> <b>14:05-14:20</b></p>	<p><b>Synthetic Datasets for Group Activity Recognition</b>  <b>Authors:</b> Gang Zhang, Chong Wang, Yizhong Zhao, Jiankun Zhou and Wenbo Deng  <b>Presenter:</b> Gang Zhang, Shenyang University of Technology, China</p> <p><b>Abstract:</b> Datasets play a critical role in training and evaluating group activity recognition models, which are widely applied in video surveillance, sports analytics, and related domains. While real-world datasets are commonly used, they often suffer from limited data volume, inadequate behavioral diversity, and low annotation accuracy. Synthetic datasets can address annotation challenges but typically lack realistic human-human interactions. To overcome these limitations, we developed a data generator capable of constructing four virtual environments: a plaza, a traffic intersection, an activity center, and a teaching building. These scenarios simulate five types of group activities—walking, waiting, engaging in conversation, queuing, and street crossing, featuring customizable scenes and activity. Each scenario employs three monocular cameras from different angles and one stereo camera to capture video data, with automatic annotations generated for all footage. The generator supports the production of an unlimited amount of labeled crowd video data. We pre-trained a recognition model on 40 synthetic videos and subsequently fine-tuned it using real-world datasets. Experimental results show an improvement in recognition accuracy.</p>
<p><b>AD223</b> <b>14:20-14:35</b></p>	<p><b>FQ-EMCI-Net: A Multi-Head Attention CNN-DQN Approach with Filtered Q-Learning and Equilibrium Monte Carlo Initialization for SP-DLBP</b>  <b>Authors:</b> Zhongyuan Yang, Weiwei Zhai, Xiaowei Xu, Shuo Shi, Yanping Xu, Ruohong Shi  <b>Presenter:</b> Zhongyuan Yang, Ocean University of China, China</p> <p><b>Abstract:</b> The Stochastic Parallel Disassembly Line Balancing Problem (SP-DLBP) presents significant challenges in optimizing task allocation due to large-scale multi-objective optimization and complex production environments. This paper proposes a reinforcement learning-based method that integrates a Multi-Head Attention CNN-DQN algorithm with an Equilibrium Monte Carlo Tree Initialization (EMCI) approach. The method includes a CNN-based feature extraction module for modeling dynamic task distributions and a Multi-Head Attention mechanism to capture task dependencies for improved global optimization. The EMCI method optimizes the exploration strategy, enhancing solution quality and convergence speed. Experimental results on public datasets and real-world applications, such as refrigerator disassembly, demonstrate significant improvements in workstation count, load balancing, and solution stability compared to traditional algorithms. The proposed method proves effective in</p>



	solving high-dimensional, coupled disassembly tasks and offers valuable insights for reinforcement learning applications in complex environments.
<b>AD4032</b> <b>14:35-14:50</b>	<p>SegRNN Optimization and Its Application in SST Forecasting  <b>Authors:</b> Lianzhi Wang, Xianbiao Kang, Guansuo Wang, Haijun Song  <b>Presenter:</b> Lianzhi Wang, Civil Aviation Flight University of China</p> <p><b>Abstract:</b> Sea surface temperature (SST) forecasting plays a crucial role in marine meteorology and climate prediction, yet traditional Segmented Recurrent Neural Networks (SegRNN) exhibit significant limitations in capturing complex spatiotemporal dependencies essential for accurate environmental forecasting. This study presents a comprehensive optimization framework that systematically enhances SegRNN's modeling capabilities through two key innovations: dynamic spatial dependency modeling via adaptive graph attention networks, and multi-scale temporal feature extraction using dilated causal convolution. The framework is evaluated using hourly SST observations from 35 marine meteorological stations across China's eastern coastal waters from January 2021 to June 2024. Experimental results demonstrate substantial performance improvements across all forecasting horizons, with the optimized SegRNN achieving RMSE reductions of up to 15% and correlation coefficient improvements exceeding 0.1 compared to traditional approaches. Case study analysis confirms the model's enhanced ability to capture diurnal temperature cycles and preserve temporal fidelity. This optimization framework establishes a robust foundation for advanced environmental forecasting applications, demonstrating significant potential for operational marine meteorology and climate prediction systems.</p>
<b>AD350</b> <b>14:50-15:05</b>	<p>Visible-Infrared Person Re-Identification with Modality-Specific Expert  <b>Authors:</b> Zishao Qiao, Xiaobin Liu, Xinyu Guo, Jianing Li, Chanho Eom, Jing Yuan  <b>Presenter:</b> Xiaobin Liu, Nankai University, China</p> <p><b>Abstract:</b> Infrared Person Re-Identification (ReID) task requires optimizing feature distance both within and across different modalities, making it more challenging than the previous visible ReID task. Existing methods commonly optimize intra-modality and inter-modality feature distance on a single model simultaneously, leading to a complicated training task and hindering the adequate optimization in the feature space. To handle this issue, this paper proposes a Modality-Specific Expert Network (MSE-Net) for the visible-infrared person ReID task. Specifically, MSE-Net independently trains two modality-specific ReID models for visible and infrared modalities as experts, respectively. As each expert focuses on feature optimization within a single modality, these two experts are equipped with stronger discrimination capacity for each modality, hence could provide accurate modality-specific supervision for the crossmodality joint training of the model. MSE-Net disentangles the feature optimization within and across modalities and thus optimizes the visible-infrared ReID model effectively. Experiments on two widely-used datasets, i.e., SYSU-MM01 and RegDB, demonstrate the superior performance of our MSE-Net over existing state-of-the-art methods. Our code is released at <a href="https://github.com/Qiaozs/MSE-Net">https://github.com/Qiaozs/MSE-Net</a>.</p>
<b>AD550</b> <b>15:05-15:20</b>	<p>AASFNet: An Attention-Aware Spatial-Temporal Fusion Network for Enhanced Pain Intensity Evaluation in Facial Image  <b>Authors:</b> Feng Gao, Linbo Qing, Lindong Li, Wei Zhao, Li Gao  <b>Presenter:</b> Feng Gao, Sichuan University, China</p>



**Abstract:** Video-based pain intensity assessment offers continuous discomfort quantification, overcoming limitations of conventional subjective methods through automated facial expression analysis. However, current deep learning architectures, despite excelling in extracting spatiotemporal pain features from facial expressions, struggle to localize transient pain signals across consecutive frames and specific facial regions, limiting detection accuracy and clinical utility. We propose the Attention-Aware Spatiotemporal Fusion Network (AASFNet). This framework combines a temporal sub-network with feature gating to focus on video frames containing pain-related information, and a spatial sub-network with attention mechanisms to highlight localized pain-related facial regions. Additionally, it integrates facial landmark geometric information to compensate for global spatial feature loss. Feature vectors from both sub-networks are fused via an adaptive feature fusion network with multi-head attention mechanisms to capture global and local dependencies for pain intensity regression analysis. Tested on the UNBC-McMaster Shoulder Pain Expression Archive Database, AASFNet achieves MAE=0.31, MSE=0.52, and PCC=0.89, surpassing state-of-the-art methods. Index Terms—Pain intensity estimation, Attention mechanism, Spatial sub-network, Temporal sub-network, Feature fusion.



# Onsite Oral Session 4

**Topic:** Multimedia Technology

**Session Chair:** Yuanzhouhan Cao, Beijing Jiaotong University, China

July 18th, 2025 | 13:30 – 15:30

**Venue:** Xiangqing Hall - 2nd Floor | 二楼祥庆厅

Invited Speaker-Wuzhen Shi, Invited Speaker-Wenxue Cui, Invited Speaker-Yiyi Liao,  
 AD216, AD546, AD549, AD414

<p><b>Invited Speaker 13:30-13:50</b></p>	<div data-bbox="590 694 821 922">  </div> <div data-bbox="836 763 1190 857"> <p><b>Wuzhen Shi</b>        Shenzhen University, China</p> </div> <div data-bbox="434 927 1369 999"> <p><i>Speech Title: Gradient-Guided Optimization for Large Motion Video Frame Interpolation</i></p> </div> <div data-bbox="359 1050 1445 1415"> <p><b>Abstract:</b> To address the challenges posed by large motions, such as severe pixel displacements, complex occlusions, and illumination variations, this paper proposes a gradient-guided video frame interpolation network tailored for large-motion scenarios. The network generates interpolation results under the guidance of gradient information and employs a cross-scale fusion module to effectively handle multi-scale motion cues. By making full use of motion information across different scales, the network is able to capture both fine local textures and global large-motion patterns, thereby achieving promising results in both large and small motion interpolation tasks. Extensive experiments demonstrate that the proposed method achieves superior interpolation performance on three widely used benchmark datasets.</p> </div>
<p><b>Invited Speaker 13:50-14:10</b></p>	<div data-bbox="590 1451 821 1680">  </div> <div data-bbox="836 1525 1319 1619"> <p><b>Wenxue Cui</b>        Harbin Institute of Technology, China</p> </div> <div data-bbox="391 1697 1414 1769"> <p><i>Speech Title: Message Passing in Deep Unfolding Network for Image Compressive Sensing</i></p> </div> <div data-bbox="359 1821 1445 2004"> <p><b>Abstract:</b> Inspired by certain optimization solvers, the Deep Unfolding Network (DUN) usually inherits a multi-stage structure for image Compressive Sensing (CS) reconstruction. However, in existing DUNs, the message-passing modes within and across stages still face the following issues: 1) the singleness of transmitted information, e.g., the low-dimensional representations or the recovered images. 2) the inefficiency of transmission policy, e.g.,</p> </div>



	<p>simple concatenation or addition between deep features. Therefore, how to construct a more efficient message-passing mechanism has become a focal point in the study of deep unfolding networks. Our research primarily addresses this challenge through two key aspects: 1) Designing more efficient network architectures to extract richer prior information, facilitating the preservation and propagation of messages; 2) Developing more effective message content and transmission strategies to ensure precise information delivery while minimizing information loss. Our recent advances along these lines have yielded state-of-the-art reconstruction performance.</p>
<p><b>Invited Speaker</b> <b>14:10-14:30</b></p>	<div>  <div> <p><b>Yiyi Liao</b> Zhejiang University, China</p> </div> </div> <p><i><b>Speech Title: Towards Efficient Volumetric Video: Representation, Compression and Standardization</b></i></p> <p><b>Abstract:</b> Recent progress in novel view synthesis has made it easier to create high-quality volumetric videos from real-world images, bringing photorealistic immersive media within reach for broad applications. Among these, 3D Gaussian Splatting enables real-time rendering but comes with a substantial memory cost compared to methods like NeRF, highlighting an inherent trade-off between speed and storage. To address this, efficient compression techniques are critical for reducing memory usage while maintaining real-time rendering. In this talk, I will first provide an overview of existing compression approaches for Gaussian Splats. I will then introduce 4D scene representations and corresponding compression techniques for volumetric video built on Gaussian Splatting. Finally, I will present recent exploration efforts within MPEG, the international standardization body, aimed at developing standardized Gaussian Splatting codecs for broad industry adoption.</p>
Paper ID & Time	Presentation
<p><b>AD216</b> <b>14:30-14:45</b></p>	<p>Deep Learning-Based Classification of Planar <sup>99m</sup>Tc Pyrophosphate Scintigraphy for the Diagnosis of Cardiac Amyloidosis  <b>Authors:</b> Jiashu Zhang, Kousuke Imamura, Takayuki Shibutani, Satoru Watanabe, Kenichi Nakajima  <b>Presenter:</b> Jiashu Zhang, Kanazawa University, Japan</p> <p><b>Abstract:</b> Amyloidosis is a rare but life-threatening disease that requires timely diagnosis to prevent severe outcomes. While <sup>99m</sup>Tc-PYP scintigraphy is widely used for the evaluation of cardiac amyloidosis. Although planar imaging has inherent limitations—with additional SPECT imaging often recommended—it remains the primary diagnostic step for the diagnosis, making effective classification of planar imaging critical. Traditional analysis of planar images can be time-consuming and susceptible to diagnostic variability due to these methodological constraints. This study proposes a novel deep learning-based system employing an encoder-analyzer-classifier architecture for automated classification of <sup>99m</sup>Tc-PYP planar images. Unlike existing methods, our approach integrates an attention-based</p>



	<p>analyzer to improve robustness against false positives and organ positional variability. To validate the model, a dataset collected from Kanazawa University Hospital was used, and comparative experiments with ViT-B/16, ResNet-50, and Swin-Tiny were conducted. The proposed model achieved the highest average classification accuracy of 90.10% with an inference time of 18.98 ms, outperforming existing models in both accuracy and computational efficiency. These results demonstrate the model's potential to reduce physician workload and improve diagnostic consistency in clinical settings.</p>
<p><b>AD546</b> <b>14:45-15:00</b></p>	<p>Physical-Model-Guided Dual-Branch Generative Adversarial Network for Thin Cloud Removal  <b>Authors:</b> Mingze Zhu, Tao Zhan, Yuanyuan Zhu  <b>Presenter:</b> Mingze Zhu, Northwest A&amp;F University, China</p> <p><b>Abstract:</b> Remote sensing (RS) imagery is essential for Earth observation applications, yet its utility is often compromised by cloud contamination, which reduces image quality and hinders downstream analysis. Although the semi-transparency of thin clouds permits the possibility of single-image restoration, most existing deep learning approaches rely on paired training data from the same geographic area and offer limited interpretability. To address these challenges, we propose a novel unsupervised framework, termed physical-model-guided dual-branch generative adversarial network (PM-DBGAN), for thin cloud removal using unpaired cloudy and cloud-free images. The core of this model is a dual-branch generator with a feature interaction module that decomposes a cloudy image into transmission map, atmospheric light, and cloud-free radiance. These components are recombined via a physical model to reconstruct the cloudy input and generate a synthetic cloudy image, enforcing cycle consistency at both image and physical levels. By explicitly modeling and inverting the atmospheric degradation process, PM-DBGAN enhances interpretability and versatility, effectively recovering clear surface information under thin cloud cover. Experimental results on two benchmark datasets demonstrate that the proposed method outperforms several state-of-the-art methods, achieving superior restoration quality and generalization capability across diverse RS scenarios.</p>
<p><b>AD549</b> <b>15:00-15:15</b></p>	<p>MetricCol: Metric Depth and Pose Estimation in Colonoscopy via Geometric Consistency and Domain Adaptation  <b>Authors:</b> Yeqi Liu, Deping Yu, Ling Liu,  <b>Presenter:</b> Yeqi Liu, Sichuan University, China</p> <p><b>Abstract:</b> Accurate metric depth and pose estimation are critical for colonoscopic navigation and lesion localization. However, existing methods often struggle with scale ambiguity and domain gaps between synthetic and real datasets. To address these issues, we propose a novel framework consisting of two stages: 1) a fully supervised depth estimation model utilizing synthetic data with anatomical priors to bridge the domain gap between synthetic and real datasets, and 2) a weakly supervised joint learning approach combining camera-aware depth scaling with uncertainty-driven pseudo-labeling to refine metric depth and pose estimation. We validate the framework on both synthetic and real colonoscopy datasets, achieving superior performance in metric depth (RMSE=3.5408) and pose estimation (average ATE=0.6143). Experimental results on both synthetic and real colonoscopy datasets show superior performance, robustness under challenging conditions, and demonstrate the clinical applicability of our method. Code is available at <a href="https://github.com/liuyq055/MetricCol">https://github.com/liuyq055/MetricCol</a>.</p>

<p><b>AD414</b> <b>15:15-15:30</b></p>	<p>A Train-borne video intelligent solution for high-speed railway infrastructure inspection  <b>Authors:</b> Kunzhen Liu, Junbo Liu and Shengchun Wang  <b>Presenter:</b> LIU Kunzhen, City university of Hong Kong</p> <p><b>Abstract:</b> This paper proposes an intelligent monitoring method for high-speed railway environments based on panoramic stitching technology. By installing imaging equipment on comprehensive inspection trains to capture railway environment videos, the method dynamically adjusts the width of the sampling area in each frame to construct a panoramic imaging model, generating four types of panoramic images: "sky-track", "left and right guardrail pillars", and "ground-track". Utilizing deep learning networks to identify abnormal features in the panoramic images and developing corresponding classifiers, the method achieves automatic identification and classification of railway environmental anomalies. Experiments show that this method significantly improves the efficiency of video analysis. It not only compresses video content to shorten browsing time but also enables efficient operation of automatic detection algorithms on compressed panoramic images, thereby advancing the automation process of railway environmental anomaly detection.</p>
--	--



# Online Oral Session 1

**Topic:** Image Processing Theory and Application

**Session Chair:** TBA

July 18th, 2025 | 10:00 - 11:50

**ZOOM A: 87471010157 | Password:071618**

Invited Speaker-Sinong Quan, AD212, AD360, AD535, AD520, AD417, AD416

<p><b>Invited Speaker</b> <b>10:00-10:20</b></p>	<div data-bbox="461 651 676 882">  </div> <p><b>Sinong Quan</b> <b>National University of Defense Technology, China</b></p> <p><i><b>Speech Title:</b>Polarimetric SAR Image Target Detection and Recognition</i></p> <p><b>Abstract:</b> This report introduces the general concepts of radar polarization and target polarization for PolSAR image processing, expounds the basic concepts and physical principles of target polarization scattering, gives the definition, connotation, and mathematical expression of classical polarimetric decomposition, and analyzes the inherent problems in it. On this basis, the concept and general framework of fine polarimetric decomposition are proposed based on physical scattering modeling, and actual samples are given for verification. Finally, using fine polarimetric decomposition as a basic tool, the idea of mathematical programming-based feature design is proposed, and its application and potential in vehicle detection, ship detection, ship recognition and other scenarios are discussed.</p>
<p><b>Paper ID &amp;Time</b></p>	<p><b>Presentation</b></p>
<p><b>AD212</b> <b>10:20-10:35</b></p>	<p>Spectro-textural Integration in Mangrove Delineation: A Case Analysis of Aboitz Cleanergy Park, Davao City, Philippines  <b>Authors:</b> Fillmore D. Masancay, Kenneth S. Castillo, Veah Trisha C. Rivera, &amp; Trixia D. Tesoro  <b>Presenter:</b> Fillmore Masancay, University of Southeastern Philippines, Philippines</p> <p><b>Abstract:</b> Mangroves, located between land and sea, are highly biodiverse ecosystems that provide services, enhancing coastal resilience against climate change. Despite their importance, mangroves face significant degradation due to human activities, highlighting the need for detailed studies on their spatial extent. This study evaluates remote sensing (RS) and geographical information systems (GIS) techniques for delineating mangroves in Aboitz Cleanergy Park, Punta Dumalag, Davao City, Philippines. The Park is selected for its ecological significance as a protected area with diverse coastal habitats, including mangroves. Sentinel-2A images are processed using GIS and the Sentinel Application Platform (SNAP). Mangrove areas are classified using Google Earth Pro imagery and</p>

	<p>ground truthing. Vegetation indices, principal components, and Grey Level Co-Occurrence Matrix (GLCM) textures are analyzed. A Random Forest algorithm achieves robust classification, yielding a kappa value of 0.896. Results reveal a 42.16% increase in mangrove coverage from 2019 to 2024. The study demonstrates the effectiveness of RS in accurately mapping mangroves and supporting sustainable management and conservation efforts.</p>
<p><b>AD360</b> <b>10:35-10:50</b></p>	<p>Study of an SBAS-InSAR Phase Unwrapping Method Integrating ICU and SNAPHU: A Case Study of Dalian City  <b>Authors:</b> JianSheng He and Liang Leng  <b>Presenter:</b> jiansheng He, JiLin Uiniversity,China</p> <p><b>Abstract:</b> To support ground deformation monitoring in Jinzhou District, Dalian, we propose a hybrid SBAS-InSAR approach integrating ICU and SNAPHU for phase unwrapping. A total of 38 ascending Sentinel-1 SAR images from December 2023 to January 2025 were processed. The method dynamically selects unwrapping algorithms based on interferogram coherence: ICU is applied in high-coherence regions above 0.4 to improve speed, achieving five to ten times faster performance compared to SNAPHU; SNAPHU is used in low-coherence areas, which are common during the vegetation-rich months from May to October, to ensure accuracy. Triangular closure analysis confirms reduced phase discontinuities and improved unwrapping consistency. The resulting deformation map reveals localized subsidence of approximately three to five centimeters per year on the southwestern slope of Dahei Mountain, correlating with geological transition zones and potential anthropogenic activity. This framework balances computational efficiency and robustness, demonstrating adaptability for urban subsidence monitoring in monsoonal climates and offering potential for future machine learning-based coherence optimization and multi-band SAR integration.</p>
<p><b>AD535</b> <b>10:50-11:05</b></p>	<p>Monocular 3D Reconstruction Based on Deep Convolutional Neural Networks  <b>Authors:</b> Chenxi Di, Zhen Dai  <b>Presenter:</b> Zhen Dai, Air Traffic Control Products Division, Sun Create Electronics Co., Ltd, China</p> <p><b>Abstract:</b> This research proposes a novel convolutional neural network (CNN) architecture for monocular depth estimation and subsequent 3D reconstruction. Unlike traditional methods that rely heavily on multi-view geometry or expensive depth sensors, this research approach leverages a lightweight encoder decoder network with attention-based feature fusion to predict high-resolution depth maps from a single RGB image. We introduce a multi-scale smoothness loss and a gradient-aware edge preservation loss to sharpen object boundaries and mitigate depth discontinuities. Extensive experiments on the KITTI [1] and NYU Depth V2 [2] datasets demonstrate that proposed method achieves state-of-the-art depth prediction accuracy while maintaining real-time inference speed on a single NVIDIA GTX 1080 GPU. Qualitative 3D point cloud visualizations further validate the effectiveness of this research approach in reconstructing fine geometric details.</p>
<p><b>AD520</b> <b>11:05-11:20</b></p>	<p>Deep Learning Model-based nudity detection with Image Feature Extraction approaches for GLAM Materials  <b>Authors:</b> Vincent Wai-Yip LUM, Hudson Chin-Fung NG  <b>Presenter:</b> Hudson Chin-Fung NG, The Chinese University of Hong Kong Library, China</p>



	<p><b>Abstract:</b> Aiming to enhance the accessibility of GLAM (Galleries, Libraries, Archives, and Museums) materials in cultural heritage collections, this paper explores the application of AI-based methodologies for identifying nudity content present in library catalogs. We specifically discuss methods that incorporate Nudity Detection algorithms to identify sensitive body regions through a case study involving an entertainment magazine from our library. We utilize an automatic content filtering algorithm to blur specific nudity regions, allowing us to unlock sensitive content. Experimental results demonstrate that our approach outperforms traditional methods in terms of accuracy, recall, precision, and F1-scores. A unique score can be computed for each detected body part, indicating the confidence level of nudity detection. This is crucial for our workflow involving human-in-the-loop initiatives, allowing reviewers to focus on low-scored content that may require a higher level of assessment. Moreover, the system can be implemented with limited computational resources, even on computers without GPUs, making it suitable for budgets that are constrained. The proposed workflow further enhances censorship detection accuracy by employing YOLO and Recolorization techniques on our digitized collections, which feature heterogeneous layouts, including a mixture of article text, photos, and illustrations, as well as varying color tones.</p>
<p><b>AD417</b> <b>11:20-11:35</b></p>	<p><b>Hierarchical Multi-task Restoration Network for Old Photo Enhancement</b>  <b>Authors:</b> Chenyang Diwu, Zifei Zhang, Jinglu He, Ting Fan, Weiao Hao  <b>Presenter:</b> Chenyang Diwu, Xi'an University of Posts and Telecommunications, China</p> <p><b>Abstract:</b> With the rapid development of deep learning in the field of image processing, old photo restoration has made significant progress. However, due to the presence of various types of degradation—such as scratches, fading, and structural damage—single-task or single-module models struggle to handle complex degradation effectively. In this paper, we propose a Hierarchical Multi-task Restoration Network (HMR-Net) that integrates three functional modules: degradation awareness, global restoration, and detail enhancement, to improve restoration quality comprehensively. Specifically, a degradation-aware module is designed to generate degradation masks that guide the subsequent restoration process. The restoration module adopts an encoder-decoder architecture enhanced with Transformer attention to model long-range dependencies. Finally, the detail enhancement module focuses on recovering fine textures and edge structures. Experimental results demonstrate that the proposed method performs well on several public datasets for old photo restoration, particularly excelling in detail fidelity and perceptual realism.</p>
<p><b>AD416</b> <b>11:35-11:50</b></p>	<p><b>A Lightweight Perception-Driven Compression Method for Social Media Images</b>  <b>Authors:</b> Qianying Feng, Dan Xu, Ying Zhang  <b>Presenter:</b> Qianying Feng, Xi'an University of Posts and Telecommunications, China</p> <p><b>Abstract:</b> The rapid proliferation of social media content and the increasing heterogeneity of terminal devices have posed significant challenges to traditional image compression methods in terms of perceptual quality, adaptability across platforms, and deployment efficiency. This paper proposes a lightweight image compression framework tailored for social media applications, which integrates semantic-visual attention fusion, platform-aware control, and dynamic structural optimization. The proposed method guides the compression network to allocate more bits to subjectively important regions by combining semantic segmentation and visual saliency, thus improving the quality of perceptual reconstruction. A platform-adaptive module is further introduced to jointly model image</p>





complexity and device-network context, enabling adaptive adjustment of compression rate and structural depth across diverse runtime environments. Additionally, a lightweight encoder-decoder architecture based on depth-wise separable convolutions and conditional path switching is designed to reduce inference latency and computational cost without compromising compression performance. Experimental results on multiple representative social image datasets demonstrate that the proposed method outperforms state-of-the-art baselines in terms of both subjective and objective quality, rate distortion performance, and deployment efficiency, showing strong potential for real-world cross-platform applications.

# Online Oral Session 2

**Topic:** Computer Graphics and Computational Photography

**Session Chair:** TBA

July 18th, 2025 | 10:00 - 11:50

**ZOOM B: 87933161872 | Password:071618**

Invited Speaker-Zizhao Wu, AD207, AD407, AD412, AD544, AD1002, AD539



**Zizhao Wu**

**Hangzhou Dianzi University, China**

***Speech Title:**Data-driven Human Motion Generation*

**Invited  
Speaker  
10:00-10:20**

**Abstract:** In this talk, I will present two of our recent works on data-driven human motion generation, focusing on multi-person motion prediction and sketch-based motion synthesis.

For multi-person motion prediction, we propose a novel Trajectory-Aware Body Interaction Transformer (TBIFormer) via effectively modeling body part interactions. Specifically, we construct a Temporal Body Partition Module that transforms all the pose sequences into a Multi-Person Body-Part sequence to retain spatial and temporal information based on body semantics. Then, we devise a Social Body Interaction Self-Attention (SBI-MSA) module, utilizing the transformed sequence to learn body part dynamics for inter- and intraindividual interactions. Furthermore, different from prior Euclidean distance-based spatial encodings, we present a novel and efficient Trajectory-Aware Relative Position Encoding for SBI-MSA to offer discriminative spatial information and additional interactive clues. Extensive experiments demonstrate that our method greatly outperforms the state-of-the-art methods.

For sketch-based human motion generation, we introduce Sketch-guided human Motion Diffusion (SMD), to address a novel scenario: sketch-to-motion, aiming to generate plausible and natural human motions based on human motion sketches. Specifically, our proposed SMD employs a Dual-branch Time-aware Transformer that utilizes both global semantic and local perspective level attention to condition 2D sketch information for 3D motion generation. Our approach demonstrates proficiency in motion in-betweening and body part editing tasks, seamlessly generating natural motion sequences that harmonize with the provided context. Multiple experiments conducted on the curated sketch-to-motion datasets validate the efficacy of SMD, showcasing the state-of-the-art generation performances.

**Paper ID  
&Time**

**Presentation**

**AD207  
10:20-10:35**

Dental lesion segmentation method based on hypernetwork improved Unet

**Authors:** Tian Ma, Xue Qin, JiaHui Li, HanWen Zhang

**Presenter:** Xue Qin, Xi'an University of Science and Technology, China

**Abstract:** Dental lesion image segmentation is one of the main techniques for oral endoscopy-based

	<p>auxiliary diagnosis of dental diseases. Aiming at the problem that the uneven distribution of pixels and number of samples in the lesion region leads to the inability of the model to accurately capture the tail class lesion regions, a dental lesion segmentation method based on the hypernetwork improved Unet is proposed. Incorporating engineering features in feature extraction provides additional guidance for the model when data are scarce or unevenly distributed. Secondly, a contextual feature mapping block is designed to capture high-level semantic features while generating adaptive weights for different layers of the decoder to improve the segmentation accuracy of tail class lesion regions. Experimental results show that compared with the current state-of-the-art methods, the proposed method achieves better segmentation results on the unbalanced dataset, especially in terms of segmentation accuracy for the tail class of</p>
<b>AD407</b> <b>10:35-10:50</b>	<p>Detailed 3D Modeling and Component Monomerization Extraction of Buildings Using Close-range Photogrammetry  <b>Authors:</b> Shijing Han, Zhao Lu, Shufeng Miao, Qingfang Chang, Ningjing Xu, Riyan Long  <b>Presenter:</b> Riyan Long, Nanning Normal University, China</p> <p><b>Abstract:</b> In urban governance, geospatial planning, and virtual reality applications, it is essential to quickly and accurately capture the characteristics of buildings and their components. This study focuses on buildings within complex environments, using UAVs to gather close-range photogrammetry data for detailed modeling and component monomerization. A model-driven flight path design strategy was proposed during the path planning stage. Based on this method, a detailed 3D model was reconstructed. In the component monomerization phase, a technique was introduced that combines bounding box fitting with surface texture sample fusion for the mesh model. Experimental results demonstrate that the final 3D model achieves higher accuracy, with tile details, window structures, wall textures, and cracks clearly depicted. The highly realistic independent models of building components address issues such as surface unevenness and jagged edges, while significantly improving rendering efficiency.</p>
<b>AD412</b> <b>10:50-11:05</b>	<p>The Surface Defect Detection of Chip Images Based on the Improved FCOS with SENet  <b>Authors:</b> Qing Liu, Yibo Jiang, Kun Chen, Jing Hao, Yuxiang Zhao, Xiangbing Li  <b>Presenter:</b> Yibo Jiang, Tianshui Normal University, China</p> <p><b>Abstract:</b> To efficiently detect surface defects on integrated circuit (IC) chip packaging, an FCOS (Fully Convolutional One-Stage Object Detection) algorithm based on SENet (Squeeze-and-Excitation Network) is proposed. First, SENet introduces channel and spatial attention mechanisms to adaptively adjust feature responses, thereby improving the accuracy and performance of object detection. Second, FCOS adopts a fully convolutional structure and an anchor-free approach for object detection. Finally, the detection precision and robustness are enhanced, enabling detection of targets of any size. In the experiments, the algorithm is validated and compared with popular detection algorithms such as YOLOv8, Faster-RCNN, and SSD. Evaluation metrics demonstrate that the improved FCOS algorithm achieves precision and recall rates of 72.1% and 76.1%, respectively, while also improving the runtime for object detection.</p>
<b>AD544</b> <b>11:05-11:20</b>	<p>Exploration of the Mechanisms Underlying Corneal Decompensation Using Graph Neural Networks  <b>Authors:</b> Chunxu Guo and Bowen Yang  <b>Presenter:</b> Chunxu Guo, Shandong University, China</p> <p><b>Abstract:</b> Objective, materials and methods, results and conclusion should be included This project improves the accuracy of loss-of-substance judgement by capturing graph structural features through graph neural network, and the deep integration of medical and industrial innovations, boosts the intelligent development of medical devices, and responds to the call of Healthy China. Corneal</p>

	<p>endothelial cell loss is highly related to ocular health and intraoperative safety of ocular surgery, meanwhile, due to its non-renewable characteristics and the inevitable damage caused by healthcare workers in clinical operation, the study of its loss mechanism is an important part of the protection of ocular health; so far, ophthalmologists' research on it still remains in the qualitative judgement of the clinical index observation, or relies on the use of time-consuming and highly subjective semi-automatic tools, which requires the operator to use a semi-automatic tool to determine the loss of corneal endothelial cells. semi-automated tools, which require operator interaction. We developed and applied a fully automated graph neural network-based corneal endothelial cell loss analysis system, called the Corneal Endothelial Loss Analysis System, embedded in a non-contact corneal endothelial microscope for segmenting and counting endothelial cells in human corneal images. First, filtering is performed to reduce noise and improve image quality. Second, the watershed algorithm and Venn diagram were used to detect the endothelial cell boundaries, while U-net was used in parallel for image segmentation during the process, and the two were compared to accurately quantify the morphological parameters of human corneal endothelial cells. The segmented images were subjected to image feature extraction, which was used as a feature parameter input into the constructed graph neural network to achieve accurate binary classification prediction of endothelial cell loss. Based on a database consisting of 50 corneal endothelial cell images, the performance of the constructed graph neural network model is tested and cross-validated with ten folds to improve the accuracy as well as robustness of the model.Objective, materials and methods, results and conclusion should be included This project improves the accuracy of loss-of-substance judgement by capturing graph structural features through graph neural network, and the deep integration of medical and industrial innovations, boosts the intelligent development of medical devices, and responds to the call of Healthy China. Corneal endothelial cell loss is highly related to ocular health and intraoperative safety of ocular surgery, meanwhile, due to its non-renewable characteristics and the inevitable damage caused by healthcare workers in clinical operation, the study of its loss mechanism is an important part of the protection of ocular health; so far, ophthalmologists' research on it still remains in the qualitative judgement of the clinical index observation, or relies on the use of time-consuming and highly subjective semi-automatic tools, which requires the operator to use a semi-automatic tool to determine the loss of corneal endothelial cells. semi-automated tools, which require operator interaction. We developed and applied a fully automated graph neural network-based corneal endothelial cell loss analysis system, called the Corneal Endothelial Loss Analysis System, embedded in a non-contact corneal endothelial microscope for segmenting and counting endothelial cells in human corneal images.First, filtering is performed to reduce noise and improve image quality. Second, the watershed algorithm and Venn diagram were used to detect the endothelial cell boundaries, while U-net was used in parallel for image segmentation during the process, and the two were compared to accurately quantify the morphological parameters of human corneal endothelial cells. The segmented images were subjected to image feature extraction, which was used as a feature parameter input into the constructed graph neural network to achieve accurate binary classification prediction of endothelial cell loss. Based on a database consisting of 50 corneal endothelial cell images, the performance of the constructed graph neural network model is tested and cross-validated with ten folds to improve the accuracy as well as robustness of the model.</p>
<b>AD1002</b> <b>11:20-11:35</b>	<p>A segmentation network for optic disc and cup based on channel feature fusion  <b>Authors:</b> Yuxuan Li, Qingyun Huo, Yunyu Wang, Mingtao Liu, Xin Zhang  <b>Presenter:</b> Yuxuan Li, Linyi University, China</p> <p><b>Abstract:</b> Glaucoma is an eye disease that progressively damages vision. Deep learning-based semantic segmentation methods have significantly advanced accurate segmentation of the optic disc and cup in fundus images, enabling quick and precise feature extraction for early glaucoma diagnosis. However, the technique still faces limitations in optic disc and optic cup segmentation, including feature loss during the learning process and insufficient feature expression in the segmented image. Therefore</p>



	<p>this paper proposes a deep learning network called CFF-TransUnet. The network integrates a Channel Feature Recombination (CFR) module in the decoder section, which recombines channel features to enhance image representation while preserving the learned features. Additionally, a re-encoding feature fusion (REFF) module has been proposed to enhance image feature extraction and improve the network's perceptual ability. We evaluate the performance of the DRISHTI-GS, REFUGE, and RIM-ONE-v3 datasets, achieving good results. The experimental results show that the Dice coefficients for the optic cup region are 0.9041, 0.9049, and 0.8692, while the Dice coefficients for the optic disc region are 0.9766, 0.9660, and 0.9702.</p>
<p><b>AD539</b>  <b>11:35-11:50</b></p>	<p>Magnetic Tile Defect Detection with Cross-Scale Visual Feature Fusion: A Cascade Framework of Improved YOLOv11 and SAM Segmentation  <b>Authors:</b> Yaru Huang, Xiaoli Geng, Zhenwei You  <b>Presenter:</b> Jianbin Zhong, Software Engineering Institute of Guangzhou, China</p> <p><b>Abstract:</b> With the widespread application of permanent magnet motors in industries such as automotive and robotics, high-precision visual inspection of magnetic tile surface defects has become a critical component of industrial quality control. Traditional manual inspection is inefficient, costly, and highly subjective, failing to meet the demands of modern industrial production. This study proposes a detection-segmentation cascade framework based on cross-scale visual feature fusion. Specifically, the improved YOLOv11 algorithm expands the receptive field of convolutional kernels using wavelet convolution (WTConv) to accurately capture low-frequency information. The Self-Ensembling Attention Mechanism (SEAM) is employed to significantly enhance the processing capability for occluded and multi-scale features. Detection boxes output by YOLOv11 are converted into input prompts for the Segment Anything Model (SAM), guiding SAM to perform fine-grained segmentation of defect regions. By fusing the semantic features (defect categories) from the detection model with the geometric features (boundary contours, area) from SAM, a visual-geometric joint feature vector is constructed to achieve multi-level analysis of defect location, shape, and severity. Experimental results show that the improved YOLOv11 algorithm achieves a mean average precision (mAP) of 0.913. While maintaining lightness (model size: 5.25MB), it outperforms the best-performing YOLOv8 by 1.67% in mAP@0.5. After object detection, using SAM segmentation to further analyze defect regions enables more accurate acquisition of defect shape, position, and other information. This algorithm provides an efficient and reliable solution for intelligent quality control in the magnetic tile production industry, strongly promoting the advancement of magnetic tile surface defect detection technology.</p>



# Online Oral Session 3-1

**Topic:** Computer Vision Techniques and Application

**Session Chair:** TBA

July 18th, 2025 | 13:30 - 15:30

**ZOOM A: 87471010157 | Password:071618**

AD227, AD354, AD343, AD363, AD366, AD364, AD345, AD346

Paper ID &Time	Presentation
AD227 13:30-13:45	<p>HiLoF-DETR: A Lightweight Framework for SAR Ship Detection with Spatial Frequency Enhancement and Dynamic Alignment  <b>Authors:</b> Yunqi Zhang, Lin Bai, Wenqing Zhou, Danni Xue, Amanda Gozho  <b>Presenter:</b> Yunqi Zhang, Chang'an University, China</p> <p><b>Abstract:</b> Synthetic Aperture Radar (SAR) ship detection faces multiple challenges. The inherent speckle noise and low resolution of SAR images cause the ship features to become blurred. Meanwhile, ship targets in nearshore scenes are often mixed with complex backgrounds such as waves and islands, further interfering with effective detection. The traditional pyramid feature fusion method suffers from multi-scale semantic gaps and spatial misalignment, leading to an increase in the missed detection rate of small targets. The existing models have high computational complexity and are difficult to meet the real-time requirements of edge devices. To this end, this paper proposes an efficient detection framework for High Low Frequency Detection Transformer (HiLoF-DETR): a lightweight MobileNetV4 backbone network is used to achieve fast feature extraction, a High Low Frequency Encoder (HiLoF-Encoder) is designed to suppress noise and enhance details, and a multi-scale alignment mechanism guided by local similarity in FreqFusion is introduced to eliminate feature fusion bias. The experimental results show that HiLoF-DETR achieves AP50 accuracy of 96.7% and 90.5% on the SSDD and HRSID datasets, respectively, while the model parameter size is only 16.8M, achieving a balanced optimization of detection accuracy and computational efficiency.</p>
AD354 13:45-14:00	<p>6D Pose Estimation of Novel Objects: A Survey  <b>Authors:</b> Jiaming Zang, Tianhan Gao, Zichen Zhu, Xinbei Jiang  <b>Presenter:</b> Jiaming Zang, Northeastern University, China</p> <p><b>Abstract:</b> The 6D object pose estimation task, which determines the 3D position and orientation of an object in the camera coordinate system, is critical for applications such as robotic manipulation, augmented reality, and autonomous driving. Traditional approaches are divided into instance-level and category-level methods, while novel object pose estimation eliminates the need for retraining, significantly enhancing real-world applicability. This paper presents a comprehensive survey on 6D pose estimation for novel objects, systematically analyzing the methodologies, benchmark datasets, and evaluation metrics employed in this field. Additionally, we explore emerging trends and highlight promising directions for future research, aiming to inspire further developments in this area.</p>

<p><b>AD343</b> <b>14:00-14:15</b></p>	<p><b>YOLOv8-GAD: A Lightweight Model for Wheat Ear Counting in Field for UAV Edge Computing</b>  <b>Authors:</b> Xu Liu,Yanran Xia,Zuojie Song,Jinbiao Cui,Xia Geng  <b>Presenter:</b> Xu Liu,Shandong Agricultural University ,China</p> <p><b>Abstract:</b> Abstract— Existing deep learning models for wheat ear counting are structurally complex with a large number of parameters, making them difficult to deploy on storage-limited drone edge devices. Moreover, the dense arrangement and mutual occlusion of wheat ear in the field present significant challenges for accurate automated counting. To address these, this paper proposes an innovative lightweight network for detecting and counting wheat ear. It overcomes the deployment challenges of existing models on edge devices due to complexity and designs a drone edge computing solution for real-time counting. This solution incorporates a result caching mechanism to optimize data processing and storage, improving task efficiency and system response. The improved model introduces a gather-and-distribute mechanism for efficient multi-scale feature fusion and integrates an attention-scale sequence fusion module to enhance small target detection. Additionally, the dynamic head detection module optimizes the model's ability to perceive scale, spatial details, and task-specific features. For lightweight optimization, pruning and knowledge distillation techniques reduce computational requirements while maintaining high accuracy. Experimental results show that the improved YOLOv8n model achieves 95.2% mAP and 93.8% detection accuracy in spike detection, an increase of 3.0% and 2.4%, respectively, over the previous version. After lightweight optimization, the model size reduces to 5.6MB, and the FPS reaches 369.2, resulting in a 57.3% reduction in model size and a 116% increase in FPS. These results indicate that the model occupies fewer resources and storage space while ensuring accuracy, making it suitable for deployment on drone edge devices for real-time, accurate wheat ear counting.</p>
<p><b>AD363</b> <b>14:15-14:30</b></p>	<p><b>Vehicle accident detection in video surveillance based on BiFPN-YOLOv8</b>  <b>Authors:</b> Ying Meng, Hongtao Wu and Bingqing Niu  <b>Presenter:</b> Ying Meng, Shanxi Intelligent Transportation Institute Co.ltd, China</p> <p><b>Abstract:</b> This paper addresses the limitations of traditional traffic accident detection methods, which rely on manual monitoring and post-hoc analysis, resulting in inefficiency and lack of real-time performance. To overcome these challenges, we propose a vehicle accident detection method based on an enhanced YOLOv8 object detection framework. Firstly, the BiFPN (Bidirectional Feature Pyramid Network) is introduced to replace the PAN (Path Aggregation Network) in the neck network, thereby improving the feature representation capability. Secondly, the shortest Euclidean distance method is employed to track vehicles detected by the improved YOLOv8, enabling the extraction of motion characteristics such as bounding boxes, centroid displacement, angle, velocity, and acceleration. Finally, by analyzing the relative motion states between vehicles, a comprehensive decision mechanism is designed to identify potential traffic accidents. Experimental results demonstrate that compared with traditional GMM-based and Mask R-CNN-based algorithms, the proposed method achieves significant improvements: an average increase of 0.24 in detection rate (Recall) with 0.03 reduction in false alarm rate, while delivering superior processing speed at 32.3 FPS (5.6× faster than Mask R-CNN). The results fully validate the effectiveness and applicability of this method for vehicle accident detection.</p>



<p><b>AD366</b> <b>14:30-14:45</b></p>	<p>Highway signage breakage detection algorithm based on improved yolov8  <b>Authors:</b> Bingqing Niu, Hongtao Wu and Ying Meng  <b>Presenter:</b> Bingqing Niu, Shanxi Intelligent Transportation Institute Co.ltd, China</p> <p><b>Abstract:</b> With the expansion of the expressway network and the increase in traffic volume, the importance of signboard damage detection in traffic safety management has become increasingly prominent. Traditional detection methods mostly rely on manual inspections or computer vision algorithms based on image processing, which have problems of low efficiency and high cost. In recent years, the development of deep learning technology, especially the application of convolutional neural networks (CNN), has made automated detection possible. This paper proposes an improved YOLOv8 object detection algorithm. The Convolutional Block Attention Module (CBAM) is introduced between the backbone network and the neck network to enhance the focusing ability on key features, and the Lightweight Separable Kernel Attention (LSKA) mechanism is added to the SPPF module to reduce interference from irrelevant backgrounds. In addition, the C2f module is replaced with the lightweight C3Ghost module to reduce the computational load and improve the detection speed. The experimental results show that the mAP50 of the proposed algorithm is improved by 0.7% compared with YOLOv8s, and the detection speed is increased by 6 frames per second, demonstrating excellent performance in terms of accuracy and detection speed, providing strong support for the intelligent management of road traffic facilities.</p>
<p><b>AD364</b> <b>14:45-15:00</b></p>	<p>Vehicle Detection under Complex Weather Conditions Based on an Adaptive Model  <b>Authors:</b> Hongtao Wu, Ying Meng and Bingqing Niu,  <b>Presenter:</b> Hongtao Wu, Shanxi Intelligent Transportation Laboratory Co. ltd, China</p> <p><b>Abstract:</b> Identifying vehicle types in complex climatic conditions on highways is essential for intelligent transportation systems. This study presents an innovative model that integrates a lightweight convolutional neural network, a feature clarification module, and a YOLOv5 detection head to enhance both real-time performance and accuracy in object detection, particularly for vehicle detection in low visibility and adverse weather conditions. The model features a CNN-S module that captures fundamental image features such as edges and textures while reducing the number of parameters to maintain robust extraction capabilities. The CTP module improves feature clarity through processes such as dehazing and white balance adjustment, effectively addressing feature blurriness in challenging weather. The YOLOv5 module, designed as a lightweight single-stage network, processes CTP-enhanced images and employs regression and classification loss optimization for accurate object identification. Utilizing both high-resolution and low-resolution images, the model preprocesses high-resolution input with the CTP module while low-resolution images are processed by the CNN-S module. The combined outputs are then fed into the YOLOv5 module for detection, with final results refined through non-maximum suppression. This approach facilitates end-to-end training and adaptive image processing, enabling efficient vehicle type recognition suitable for all-weather highway monitoring scenarios.</p>
<p><b>AD345</b> <b>15:00-15:15</b></p>	<p>Implicit Diffusion-Based Super-Resolution for Intangible Cultural Heritage Images  <b>Authors:</b> Haoqun Teng, Kun Xiong, Daitao Wang  <b>Presenter:</b> Haoqun Teng, Software Engineering Institute of Guangzhou, China</p> <p><b>Abstract:</b> The digital preservation of Intangible Cultural Heritage (ICH) images faces challenges such as low resolution and loss of details. This study proposes an image super-resolution reconstruction method based on an implicit diffusion model to address these</p>





	<p>issues. The proposed method integrates an implicit diffusion variational autoencoder, multi-scale feature extraction strategy, and multi-objective optimization techniques. We constructed the Chinese Lingnan Intangible Cultural Heritage Image (CLICHI) dataset, which contains 3,568 pairs of high- and low-resolution images covering 16 ICH projects. Experiments conducted on CLICHI and other standard datasets demonstrate that our method outperforms existing state-of-the-art methods in terms of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), and Fréchet Inception Distance (FID). Notably, our method exhibits significant advantages in handling complex textures and fine details. Quantitative analysis shows that, on the CLICHI dataset, our method improves PSNR by 0.28 dB and SSIM by 0.01 compared to the best baseline model. Additionally, ablation experiments validate the effectiveness of key components of the model. This study provides a new solution for the high-quality digital reconstruction of ICH images, which is of great significance for the preservation and inheritance of cultural heritage. Index Terms—Intangible Cultural Heritage, Image Superresolution Reconstruction, Diffusion Model, Deep Learning, Cultural Preservation.</p>
<p><b>AD346</b> <b>15:15-15:30</b></p>	<p><b>EffiHeritageNet: Efficient Semantic Segmentation Method for Intangible Cultural Heritage Scenes</b>  <b>Authors:</b> Tingting Guo, Liangqiong Mai, Yao Fu, Daitao Wang  <b>Presenter:</b> Tingting Guo, Guangdong Teachers College of Foreign Language and Arts, China</p> <p><b>Abstract:</b> The automated analysis of Intangible Cultural Heritage (ICH) scenes is of great significance for cultural preservation and inheritance. However, the complex textures and diverse scales of these scenes present significant challenges for existing computer vision technologies. This paper introduces EffiHeritageNet, an efficient semantic segmentation network specifically designed for ICH scenes. EffiHeritageNet employs a lightweight architecture and depthwise separable convolutions, significantly improving computational efficiency while maintaining accuracy. The introduction of a multi-scale feature extraction mechanism effectively captures both the details and global semantic information of ICH objects. Additionally, this study innovatively combines a Two-Pass connected component analysis algorithm, enabling a seamless transition from semantic segmentation to object detection, providing a comprehensive solution for ICH scene analysis. To support the development and evaluation of the algorithm, we constructed a dataset containing 3,800 high-resolution images of ICH scenes, covering 12 common categories. Experimental results show that EffiHeritageNet-M achieved a real-time inference speed of 427.8 FPS at a resolution of 512×512, while maintaining an average accuracy (mAcc) of 97.72%. EffiHeritageNet-L excelled in accuracy, reaching a mean Intersection over Union (mIoU) of 94.16%. Compared to existing state-of-the-art methods, our approach achieves a better balance between speed and accuracy, making it particularly suitable for real-time analysis and processing of ICH scenes.</p>

# Online Oral Session 3-2

**Topic:** Computer Vision Techniques and Application

**Session Chair:** TBA

July 18th, 2025 | 13:30 - 15:30

**ZOOM B: 87933161872 | Password:071618**

AD353, AD405, AD395, AD384, AD528, AD542, AD530, AD415

Paper ID &Time	Presentation
<b>AD353</b> <b>13:30-13:45</b>	<p>Morphological Correction Method for River Skeleton Lines Based on Sampling Point Offsets</p> <p><b>Authors:</b> Genshan Liang, Liang Leng, Yongxin Sun, Senpeng Wang</p> <p><b>Presenter:</b> Genshan Liang, Jilin University, China</p> <p><b>Abstract:</b> In the computation and extraction of river skeleton lines, limitations in computational capacity and the influence of boundary noise may lead to deviations or jitter in the resulting skeletons. To address this issue, this study employs the Voronoi diagram method for skeleton line computation. The accuracy of the extracted skeleton lines is then systematically evaluated based on the deviation of sampling points along the skeleton. Subsequently, the original skeleton lines are corrected according to the evaluation results. Experimental results based on real-world vector data demonstrate that the corrected skeleton lines effectively reduce the interference caused by boundary noise, improve the smoothness of the generated skeletons, and enhance their robustness. The proposed method provides a high-precision and low-intervention solution for vector-based river skeleton line extraction, offering practical value in hydrological mapping and geographic analysis.</p>
<b>AD405</b> <b>13:45-14:00</b>	<p>Research on ultrasound image feature enhancement based on frequency-domain self-attention</p> <p><b>Authors:</b> Zhe Li, Zhiyuan Wang,Guoping Tan and Ruixin Wang</p> <p><b>Presenter:</b> Zhe Li, the Second Xiangya Hospital,China</p> <p><b>Abstract:</b> Medical ultrasound imaging is extensively utilized in clinical diagnostics due to its non-invasive, real-time capabilities and cost-effectiveness. However, the presence of speckle noise significantly compromises image quality, thereby presenting challenges to the accuracy of deep learning-based classification methods. To mitigate this issue, a feature enhancement technique based on Frequency Domain Self-Attention (FDSA) has been proposed to enhance the performance of convolutional neural networks in the classification of ultrasound images. The FDSA method converts feature maps into the frequency domain utilizing the Fast Fourier Transform, organizes feature channels according to their frequency characteristics, and applies a multi-head self-attention mechanism to adaptively integrate features across various frequency bands. This approach effectively amplifies relevant signals while attenuating noise interference. Experiments were conducted using the POCUS dataset for pneumonia classification and the BLU dataset for breast cancer classification, employing ConvNeXt V2 as the foundational model and comparing it against a range of</p>



	<p>traditional and advanced models. The results show that the ConvNeXt V2_FA model, which incorporates FDSA, has an average f1 score of 0.9428 on the (with a standard deviation of 0.0435) BLU dataset and an F1 score of 0.8060 on the POCUS dataset. Such achievements exceed the performance of the comparison models. These findings indicate that the FDSA method can effectively address the challenges brought by noisy ultrasound images and provide a novel feature enhancement strategy for medical image analysis.</p>
<p><b>AD395</b> <b>14:00-14:15</b></p>	<p>Zero-shot Object Detection with Knowledge Enhancement via Dual-branch Subgraph Reasoning  <b>Authors:</b> Xin Zhou, Jiaming Li, Yufei Kong  <b>Presenter:</b> Jiaming Li, Dalian Maritime University, China</p> <p><b>Abstract:</b> Zero-shot object detection utilizes domain prior knowledge to transfer known category information to unknown categories, improving the generalization ability of the detection system, and alleviating the problem of data scarcity. Existing methods that use semantic embedding as single descriptive knowledge information often struggle to support complex real- world visual scenes, making it difficult for features to capture, and lacking corresponding explicit category relationship information to guide them to the correct category. In this paper, we propose a knowledge enhanced dual branch zero-shot object detection method called DSR-Net which introduces a dual branch alignment structure and encodes semantic information to enable adaptive alignment capability, and a subgraph knowledge propagation module is introduced to explicitly model category relationships by constructing a category relationship graph, enhancing prior knowledge and alleviating the problem of knowledge deficiency. The experiments on the MSCOCO dataset show that the proposed method improves the detection mAP for 1.7% of detection model for unseen classes and effectively enhances the detection performance.</p>
<p><b>AD384</b> <b>14:15-14:30</b></p>	<p>Exploring Emotional Engagement with Responsible AI Constructs: A Video-Based Cognitive Experiment  <b>Authors:</b> Uday Nedunuri, Dr Nicolas Hamelin, Dr Abhijit Das Gupta, and Dr Debasish Guha  <b>Presenter:</b> Uday Nedunuri, SP Jain School of Global Management, India</p> <p><b>Abstract:</b> In the discourse of Artificial Intelligence, we are witnessing a rapid proliferation of AI across various industries—from Energy, healthcare and finance to governance and entertainment—where its capacity to create value is becoming undeniable. At the same time, we are also encountering significant repercussions when AI systems are deployed without a genuine commitment to ethical design, inclusivity, or human oversight (see Greene et al., 2001; Phelps, 2006). These instances underscore the critical need to understand the deeper, often unspoken human responses to AI systems. This lays the foundation for developing AI that is not only technically proficient but also socially and morally responsible. The aim of this study is to explore human emotional reactions to Responsible AI (RAI) principles. Unlike traditional survey-based behavioral assessments, this research captures implicit emotional signals and behavioral micro-reactions through facial expressions during structured questioning. The approach addresses the growing need for deeper understanding of trust, resistance, and acceptance in ethical AI frameworks by deploying affective computing and cognitive psychology principles. The study proposes a novel method for decoding non-verbal, subconscious responses that underpin moral reasoning and cognitive</p>



	load in ethical AI evaluation. The results suggest new pathways for human-centered AI design, regulatory foresight, and emotion-informed AI development
<b>AD528</b> <b>14:30-14:45</b>	<p><b>CETrack: A Feature-Match-Based Framework for Lesion Tracking in CE Videos</b>  <b>Authors:</b> Jiwei Wu, Shu Zhang, Weiwei Zheng and Mengli Xue  <b>Presenter:</b> Jiwei Wu, Fuzhou University, China</p> <p><b>Abstract:</b> This paper addresses the problem of target tracking in capsule endoscopy (CE) videos, which is a technique that is crucial for diagnosing and evaluating Crohn's disease. However, this task is particularly challenging due to the irregular motion trajectories of lesions, significant and random changes in appearance, motion discontinuity, and complex environmental noise interference in the intestinal environment. To tackle these challenges, we propose a novel tracking framework based on context-aware multi-feature joint matching. This framework incorporates a long-term re-detection mechanism that stores historical features of key frames, allowing it to handle random appearance changes of lesions effectively. Additionally, we introduce a dynamic trajectory deletion strategy to accommodate the motion discontinuity. To evaluate the effectiveness of our method, we conduct a benchmark evaluation on the CE video dataset. The results demonstrate that our approach outperforms traditional motion-prediction methods and significantly minimizes identity switches, highlighting its robustness in CE video tracking.</p>
<b>AD542</b> <b>14:45-15:00</b>	<p><b>A Direct Zero-Shot Indoor Scene Recognition Method based on Visual Question Answering</b>  <b>Authors:</b> Chen Wang, Junjie Wei, Yangjun Ou, Xiong Pan  <b>Presenter:</b> Chen Wang, Wuhan Textile University, China</p> <p><b>Abstract:</b> Zero-shot indoor scene recognition has the ability to recognize new indoor images from an unseen scene class, which plays an important role in robot navigation and localization. However, it's difficult to collect sufficient semantic information for indoor scene images, which results in limited knowledge transfer from seen classes to unseen classes. To address this challenge, in this paper, we proposed a Direct Zero-Shot Indoor Scene Recognition (DZSISR) method based on Visual Question Answering (VQA). Specifically, the Large Language-and-Vision Assistant (LLaVA) model has demonstrated great generalization ability in the VQA task. Benefiting from this pre-trained LLaVA model, we can directly input a question as "Directly answer which of the following categories this image belongs to: class name 0, class name 1, ...". Then, we will obtain the corresponding output category answers, which serve as the final zero-shot recognition results. This direct method does not contain any training procedures, so it is efficient and can be extended to downstream visual recognition tasks. Comprehensive experiments on three indoor scene datasets demonstrate the effectiveness of the direct zero-shot indoor scene recognition method.</p>
<b>AD530</b> <b>15:00-15:15</b>	<p><b>Research on an Intelligent Security Door Passenger Flow Statistics System Based on an Improved Deep Learning Human Body Recognition Algorithm</b>  <b>Authors:</b> Jingtao Zhang, Xiaolong Qian, Yong Wang, Junqing Xie  <b>Presenter:</b> Jingtao Zhang, Hangzhou Innovation Institute of Beihang University, China</p> <p><b>Abstract:</b> Accurate and intelligent passenger flow statistics at residential security doors are critically important for enabling smart home systems linkage. Determining the number of occupants within a household relies on precisely counting individuals entering and exiting through the main entrance, which serves as a fundamental prerequisite for effective smart</p>





	<p>home automation. However, due to the limitations in camera installation angles and coverage areas, conventional person detection algorithms such as YOLOv5s often yield high error rates in identifying ingress and egress events, thereby impeding accurate indoor people counting. To address these challenges, this study proposes an intelligent passenger flow counting method based on an enhanced deep learning human body recognition algorithm. A complete front-end and back-end application system was designed and implemented accordingly. Experimental results from deployment on security doors developed by a leading manufacturer validate the proposed method's effectiveness and practical applicability.</p>
<p><b>AD415</b> <b>15:15-15:30</b></p>	<p><b>P2P-Net: A PSO-Vision Framework for Accurate Detection and Multi-Class Classification of Parasitic Eggs in Human and Animal in Microscopy Images</b>  <b>Authors:</b> Muhammad Bilal Zia, Xujuan Zhou, Raj Gururajan, Ka Ching Chan  <b>Presenter:</b> Muhammad Bilal Zia, University of Southern Queensland, Australia</p> <p><b>Abstract:</b> Parasitic infections affecting both humans and animals, particularly gastrointestinal helminths in livestock such as Sheep remain a persistent threat to health systems and agricultural economies around the world. Rapid and accurate identification of parasite eggs on fecal microscopy images is essential for effective diagnosis and intervention. Manual microscopy, still the dominant diagnostic method, remains inherently subjective, time-intensive, and poorly suited to high throughput or remote screening environments. To address these limitations, we present P2P-Net (Pixel-to-Parasite Network), a deep learning framework developed for the high-precision detection and multi-class classification of parasitic eggs. The architecture integrates dual-path feature extraction modules that capture both fine-grained textures and broader contextual structures, feeding into a YOLO-based detection head. Model training is enhanced through dynamic learning rate adjustment via Particle Swarm Optimization (PSO), promoting faster convergence and improved generalization. P2P-Net was initially trained on the Chula-ParasiteEgg-11 dataset encompassing 11 morphologically diverse parasite egg classes, and further evaluated on a curated set of Sheep parasite egg dataset grouped into two classes for evaluation: Egg Class 1 (Haemonchus) and Egg Class 2 (a merged category of Teladorsagia and Trichostrongylus), based on their morphological similarities and the balance of dataset. The proposed method achieved 98.55% precision, 98.54% precision, 98.45% recall and an F1 score of 98.54%, exceeding existing state-of-the-art benchmarks. The improved diagnostic reliability offered by P2P-Net holds direct implications for more effective treatment selection and better infection control in both medical and veterinary applications.</p>

# Online Oral Session 4

**Topic:** Deep and Machine Learning Applications

**Session Chair:** TBA

July 18th, 2025 | 15:45 - 17:45

**ZOOM A: 87471010157 | Password:071618**

AD1007, AD2013, AD523, AD3017, AD3018, AD4027, AD4033, AD379

Paper ID &Time	Presentation
<b>AD1007</b> <b>15:45-16:00</b>	<p>Modeling Carbon Emission Using ARIMA and Neural Network  <b>Author:</b> Yihan Zhou  <b>Presenter:</b> Yihan Zhou, Shanghai World Foreign Language Academy, China</p> <p><b>Abstract:</b> This study models carbon emissions in Jiangsu and Zhejiang provinces from two thousand twenty-three to two thousand thirty-one to assess whether they can achieve carbon peak by two thousand thirty. Key factors influencing emissions—population, gross domestic product, energy consumption, and green finance index—are identified and forecasted using an autoregressive integrated moving average model trained on historical data. A trained backpropagation neural network is then used to predict carbon emissions by combining the forecasted data with historical emissions. Results show that Jiangsu will not reach carbon peak by two thousand thirty, while Zhejiang is expected to achieve it approximately five years earlier. This approach demonstrates the effectiveness of combining time series forecasting and machine learning in carbon emission prediction.</p>
<b>AD2013</b> <b>16:00-16:15</b>	<p>Comparison of Different Models for Natural Gas Forecasting and Assessment  <b>Authors:</b> Xintong Zeng, Xiaomei Wang, Guangxun Zhang  <b>Presenter:</b> Xintong Zeng, Mathematics and Technology Wenzhou-Kean University, China</p> <p><b>Abstract:</b> This paper proposes a trend-based approach for predicting U.S. natural gas consumption, which utilizes centered moving averages and extrapolation techniques. Unlike complex ensemble models, the TrendBaseline model balances simplicity with strong predictive performance. Based on 1997–2023 data, the TrendBaseline model achieves high accuracy for LogCosh and Huber losses at 0.0003 and MAE at 0.0192 for prediction on 2020–2023. It effectively captures long-term trends through filtering short-term noise. The results illustrate the practical utility of combining trend extraction with comprehensive error metrics will offer a robust, efficient methodology for energy forecasting</p>



<p><b>AD523</b> <b>16:15-16:30</b></p>	<p>Reference segmentation network based on feature interaction enhancement  <b>Authors:</b> Yuhang Zhang , Zhiguo Zhang  <b>Presenter:</b> Yuhang Zhang, Shandong University of Science and Technology, China</p> <p><b>Abstract:</b> Reference segmentation is a fundamental task in image understanding, aiming to accurately recognize and classify each part of an image to achieve precise object-level segmentation. Despite the development of numerous segmentation approaches, challenges such as large-scale variations among objects, ambiguous category boundaries, and uneven object distributions persist, making it difficult to enhance the discriminative capability of models for complex categories. To address this problem in the context of few-shot image segmentation, we propose a reference-guided feature interaction enhancement module (RGFIM), along with a feature fusion module, to enhance the model's perception and representation of target regions. By incorporating support images and their corresponding masks, the proposed module guides the model to focus on feature regions that are semantically related to the reference objects. We integrate this module into the DINOv model [1] and introduce a novel evaluation protocol. Extensive experiments conducted on the COCO-20i dataset demonstrate that our approach significantly improves segmentation accuracy and yields promising results.</p>
<p><b>AD3017</b> <b>16:30-16:45</b></p>	<p>Prediction of Air Marshals' Physical Performance Based on Least Squares Support Vector Machine and Prediction Error Correction  <b>Authors:</b> FAN ZHANG, LIXIA ZHANG, MINGJIA LI, YUXI REN  <b>Presenter:</b> Fan Zhang, Nanjing Police University, China</p> <p><b>Abstract:</b> In order to obtain better prediction results of air marshals' physical performance, the air marshals' physical performance prediction model with least squares support vector machine and prediction error correction is proposed. Firstly, the physical performance of air marshals is modeled and predicted by lifting wavelets and least squares support vector machine, then the prediction results of physical performance of air marshals are corrected by prediction error correction, and finally, tested through a case study on air marshals' physical performance prediction, and comparison experiments are carried out with other prediction models of the physical performance of air marshals to validate its superiority. The results show that the model proposed in this paper reduces the prediction error of air marshals' physical performance, and improves the stability of the prediction results of air marshals' physical performance through prediction error correction, and the prediction accuracy is better than other air marshals' physical performance prediction models.</p>
<p><b>AD3018</b> <b>16:45-17:00</b></p>	<p>Research on Public Service Bodies' Performance Evaluation Based on Data Mining  <b>Authors:</b> YIDAN ZHANG, JIAXI DAI, FAN ZHANG, YICHEN NIE  <b>Presenter:</b> Yidan Zhang, Nagoya University, Jaoan</p> <p><b>Abstract:</b> With the goal of improving the effectiveness of public service bodies' performance evaluation under the situation of massive data, the public service bodies' performance evaluation based on data mining technology is studied. The association rule mining method utilizes data mining technology to mine data related to public service bodies' performance, and the association rule mining method utilizes the three indicators of support, confidence and relevance to mine the association rules between the data and establish the public service bodies' performance evaluation system. Based on the</p>

	<p>established public service bodies' performance evaluation system, the fuzzy comprehensive evaluation method is used to realize the public service bodies' performance evaluation by establishing judgment matrix, calculating weights, and constructing fuzzy comprehensive evaluation model. The experimental results show that the researched method can effectively evaluate the public service bodies' performance, and the evaluation time is still less than 10 s when the data volume is as high as 10 GB, which is highly practical.</p>
<b>AD4027</b> <b>17:00-17:15</b>	<p>TransResNet: A Transformer-Guided Dilated Residual Network for MRI Bladder Tumor Segmentation  <b>Authors:</b> Yuanchen Dai, Yongtao Yu  <b>Presenter:</b> Yuanchen Dai, Yunnan Normal University, China</p> <p><b>Abstract:</b> Bladder cancer is a malignant tumor with high morbidity and mortality, seriously threatening patient survival. MRI can clearly visualize bladder walls and tumor regions, making accurate segmentation crucial for diagnosis and treatment planning. We propose TransResNet, an encoder-decoder CNN integrating residual learning, dilated convolutions, and Transformer-based global attention. The encoder uses residual blocks for feature reuse, dilated convolutions for enlarged receptive fields, and Transformer blocks for capturing long-range dependencies. The decoder reconstructs segmentation maps via a bottleneck upsampling path. The segmentation results of the bladder MRI dataset show that the Dice coefficients for the two segmentation tasks are 0.93 (tumor) and 0.92 (bladder), respectively, surpassing current advanced segmentation methods and demonstrating the significant performance of the model in this paper in bladder inner and outer wall and tumor segmentation tasks.</p>
<b>AD4033</b> <b>17:15-17:30</b>	<p>Enhancing Transformer Models for Long-Term Time Series Classification with Multi-Channel Frequency-Domain Filters  <b>Authors:</b> Jiandong Luo, Jianyu Kuang, Tongshou Wei, Meng Yang, Zhongrui Hu, Ye Du  <b>Presenter:</b> Ye Du, Beijing Huadian E-Commerce Technology, China</p> <p><b>Abstract:</b> Time-series classification plays a critical role in numerous domains such as healthcare, activity recognition, and industrial monitoring. Traditional Transformer models, while effective in modeling long-range dependencies, tend to focus disproportionately on high-energy frequency components, potentially overlooking valuable low-frequency information. To address this issue, we propose a frequency-domain filter Transformer model that incorporates a multi-channel frequency decomposition module prior to the Transformer encoder. This module applies parallel frequency-specific filters to segment the input sequence into distinct bands, enabling the model to extract features from both high- and low-frequency components. Each frequency stream is processed by a dedicated Transformer branch, and the results are aggregated to form a comprehensive representation. Extensive experiments on benchmark datasets such as KU-HAR demonstrate that our model significantly outperforms baseline methods, including CNN, CNN-LSTM, and Transformer architectures, achieving superior accuracy and robustness. The proposed approach effectively mitigates the attention bias toward single-band information and enhances the interpretability and performance of Transformer-based time-series classifiers.</p>



<p><b>AD379</b>  <b>17:30-17:45</b></p>	<p>Construction of Three-Dimensional Memristor-Enhanced Polynomial Hyperchaotic Map and Its Application in Image Security Protection  <b>Authors:</b> Yulian Zhang, Yue Guo, Keyuan Zhang, Jiaxin Yang, Wei Feng and Bo Cai  <b>Presenter:</b> Wei Feng, Panzhihua University, China</p> <p><b>Abstract:</b> To address the escalating demands for real-time image security protection, this paper first constructs a 3D memristor-enhanced polynomial hyperchaotic map (3DMEPHM). Our constructed 3D-MEPHM features a concise structure, excellent chaotic performance, and high efficiency in generating chaotic sequences. Subsequently, this paper further presents a meticulously designed real-time image security protection scheme based on pixel bit rearrangement (RISP-MEPHM). RISP-MEPHM achieves differentiated encryption processing for the upper 2 bits and the lower 6 bits of pixels through pixel bit rearrangement. The three rounds of diffusion-scrambling employed are designed to be both secure and efficient. A series of security and efficiency evaluations are conducted to ascertain the superiority of RISP-MEPHM. These evaluations indicate that RISP-MEPHM not only exhibits outstanding security performance but also achieves exceptionally high efficiency, with an average encryption rate up to 81.0811 Mbps. Compared to many recently developed alternatives, RISP-MEPHM stands out in addressing the pressing demands for real-time image security in emerging domains like telemedicine and cloud computing.</p>
---	---



# Online Oral Session 5

**Topic:** Innovative Applications and Technological Breakthroughs of Computer Vision and Computational Intelligence in Multiple Fields

**Session Chair:** TBA

July 18th, 2025 | 15:45 - 18:00

**ZOOM A: 87933161872 | Password:071618**

AD204, AD381, AD522, AD527, AD408, AD219, AD541, AD382, AD543

Paper ID &Time	Presentation
AD204 15:45-16:00	<p>Research on defect detection of wire rope of mine hoist based on feature embedding  <b>Authors:</b> Tian Ma , Jiahao Wang, Libing Zhou, Jiahui Li, Yuancheng Li, Jiayi Yang  <b>Presenter:</b> Jiahui Li, Xi'an University of Science and Technology, China</p> <p><b>Abstract:</b> Aiming at problems such as the scarcity of defect data samples, domain bias in feature extraction, and insufficient expression of normal features in current wire rope defect detection methods, a wire rope defect detection method for mine hoists based on feature embedding is proposed. Firstly, a spatial alignment network is designed. Through self-supervised learning, pixel level and feature level alignment of images is achieved, thus alleviating the problem of insufficient feature learning in small data sets caused by industrial cold start. Then, a feature extraction network integrating an attention mechanism is designed. The embedding vectors of different layers of the pre-trained network are connected to reduce the influence of the deep layers of the pre-trained network on features. Finally, an adaptive module for coupled features is designed. The coupled feature module is composed of patch descriptors, which can adapt to the features of the target data set, strengthen the coupling of normal features, reduce the domain bias problem existing in the processing of the pre-trained network, and avoid overestimating the normality of defect samples. The experimental results show that the AUCROC of image-level defect detection reaches 90.8%, and the AUCROC of pixel-level localization reaches 95.7%, which is 1.7% higher than the latest industrial defect detection methods and can more accurately identify the defects of wire ropes.</p>
AD381 16:00-16:15	<p>Noise Algorithms in Game Terrain Generation  <b>Authors:</b> Shuyue Zhang, Zibo Xu, Zhenyuan Liu, Minghe Liu  <b>Presenter:</b> Zibo Xu, Fuzhou University, China</p> <p><b>Abstract:</b> This article explores game terrain generation technology based on noise algorithms. Specifically, it analyzes the fundamental principles of Perlin noise, Simplex noise, Value noise, and Worley noise, as well as their specific applications and advantages in game terrain generation. Terrain generation involves the creation of landforms through artificial algorithms designed to construct geometric structures, elevation profiles, and geomorphological characteristics. In this context, noise algorithms act as foundational tools for generating pseudo-random yet controlled variations in terrain. They can quickly generate complex terrain by simulating the randomness and continuity of natural phenomena, while</p>



	<p>reducing computational costs. The paper first discusses Perlin noise, explaining its theory and its contributions in terrain generation. Then, it analyzes how Simplex noise optimizes the calculation of interpolation based on Perlin noise. Value noise further enhances computational efficiency by interpolating random height values instead of gradient vectors, but it may generate a monotonous terrain. Worley noise generates unique textures based on Voronoi diagrams and it is suitable for simulating special landforms. However, existing noise algorithms still have their different limitations when generating diverse terrains, such as the balance between terrain complexity and computational efficiency. Future research could further enhance performance by exploring avenues such as GPU parallelization and employing machine learning. This paper offers technical insights for research on noise algorithms in game terrain generation.</p>
<p><b>AD522</b> <b>16:15-16:30</b></p>	<p>A Lightweight Anonymous Authenticated Key Agreement Protocol for V2I with Multi-TA Model  <b>Authors:</b> Kun Li, Shanshan Tu, Jiakai Dou, Dazhong Liu, Zexu Li  <b>Presenter:</b> Kun Li, Beijing University of Technology, China</p> <p><b>Abstract:</b> Vehicular Ad-hoc Network (VANET) is critical for intelligent transportation, enabling vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communication to enhance road safety and efficiency. However, existing protocols face challenges such as security issues, high computational overhead and single-point-of-failure risks in V2I communication. In response to these challenges, we propose a lightweight anonymous authenticated key agreement protocol supported by a multi-Trusted Authority (TA) model, using hash functions and XOR operations to achieve mutual authentication between vehicles and Road Side Unit (RSU), as well as between RSU and TA, while establishing secure session key. Performance evaluation shows it has lower computational overhead and meets many security requirements, outperforming existing schemes in efficiency and security for intelligent transportation systems.</p>
<p><b>AD527</b> <b>16:30-16:45</b></p>	<p>A PUF-Assisted Lightweight Mutual Authentication of Low-Cost RFID Tags for Medical Privacy Preservation  <b>Authors:</b> Dazhong Liu, Shanshan Tu, Kun Li, Jiakai Dou and Zexu Li  <b>Presenter:</b> Dazhong Liu, Beijing University of Technology, China</p> <p><b>Abstract:</b> This paper addresses the security and privacy preservation requirements of low-cost RFID tags in medical systems by proposing a lightweight mutual authentication protocol based on Physical Unclonable Functions (PUF). The protocol achieves a balance between security and practicality through a triple innovative design: first, it utilizes dynamic PUF responses to generate session keys, ensuring forward secrecy; second, it introduces a timestamp-random number dual-verification mechanism to effectively defend against replay attacks; and third, it designs a multi-version state synchronization scheme to resolve the long-standing desynchronization problem in RFID systems. The protocol requires only cyclic shift, XOR, and modulo addition operations, making it suitable for resourceconstrained tags, while leveraging dynamic identity obfuscation and the hardware unclonability of PUF to resist cloning attacks, parameter prediction, and location tracking. Security analysis demonstrates that the protocol ensures communication integrity and privacy while defending against attacks such as replay, desynchronization, and message tampering. Compared to existing schemes, the proposed protocol significantly enhances</p>



	security performance while maintaining lightweight computation, offering a viable solution for sensitive data protection in the Internet of Medical Things (IoMT).
<b>AD408</b> <b>16:45-17:00</b>	<p>Stock Price Prediction and Investment Strategy via Machine Learning Model Fusion  <b>Authors:</b> Bowen Wang, Hanru Xu, Yubing Pan, Chenyongyang Yu  <b>Presenter:</b> Bowen Wang, Nanjing University, China</p> <p><b>Abstract:</b> This study explores stock price prediction using multiple machine learning models and enhances accuracy via model fusion with XGBoost as the meta-model and Linear Regression, Support Vector Machine, and Random Forest as the base models. We trained and compared various models (e.g. linear regression, random forests, and support vector machine) using historical stock data to predict future stock prices based on daily closing prices and various technical indicators. The experimental data presented a clear discrepancy in the performance of the various models. Such unification of the models through composite technology offered us a chance to increase not only the accuracy but also the stability of the forecasting process. Results of this technique gave us the opportunity to design profit-making strategies, which were evaluated through a forward-looking simulation framework to ensure out-of-sample validity. The primary goal of this study is to propose a new method for enhancing financial forecast accuracy and machine learning-driven investment decision making.</p>
<b>AD219</b> <b>17:00-17:15</b>	<p>Leveraging Multimodal Large Language Models for Referring Camouflaged Object Detection  <b>Authors:</b> Xuewei Liu, Ziyu Wei, Jizhong Han  <b>Presenter:</b> Xuewei Liu, Institute of Information Engineering, Chinese Academy of Sciences, China</p> <p><b>Abstract:</b> Referring Camouflaged Object Detection (Ref-COD) aims to identify and segment specified objects hidden within their surroundings according to the reference text or images. In this paper, we leverage the rich visual-text information in Multimodal Large Language Models (MLLM) to benefit RefCOD and propose a referring camouflaged object detection framework based on MLLM (MLLM-RCOD). In MLLM-RCOD, a Reference Image Packer is designed to encode the reference image into textual space, and a Triple-Align Loss is included to align between camouflage images, reference images, and reference text. Furthermore, a Camouflaged Chain of Thought (Cam-CoT) is proposed to guide MLLM in better dealing with the Ref-COD task. Extensive experiments on the Ref-COD benchmark show that our method achieves new state-of-the-art performance.</p>
<b>AD541</b> <b>17:15-17:30</b>	<p>Enhancing Multimodal Sarcasm Detection via Global and Local Prompt Mechanisms  <b>Authors:</b> Zewen Li, Xiaowei Xu and Huili Gong  <b>Presenter:</b> Zewen Li, Ocean University Of China, China</p> <p><b>Abstract:</b> With the rapid development of social media, the volume and complexity of multimodal data have significantly increased, posing new challenges for tasks such as multimodal sarcasm detection. This paper proposes a novel framework, Global and Local Prompt Learning (GLPL), to address these challenges by jointly capturing both global contextual and local fine-grained features in multimodal data. The global prompts align semantic information across modalities to preserve the overall meaning, while the local prompts extract detailed, modality-specific interactions from text-image pairs, enhancing sensitivity to subtle sarcastic cues. To further improve robustness and generalization, the</p>



	<p>global channel incorporates a random prompt loss strategy to avoid over-reliance on specific prompt features, and the local channel introduces a multi-scale information mining mechanism to capture sarcasm cues at different levels of granularity. This paper also conducts comprehensive ablation studies to validate the effectiveness of each module. Experimental results on the benchmark dataset demonstrate that GLPL achieves strong and consistent performance, significantly outperforming most existing approaches in identifying nuanced multimodal sarcasm.</p>
<p><b>AD382</b> <b>17:30-17:45</b></p>	<p>Real-Time Simulation of Destructible Objects: From Rigid Fractures to Soft-Body Deformation  <b>Authors:</b> Zheheng Zeng, Ziyue Chen, Yujie Li  <b>Presenter:</b> Ziyue Chen, Jiangnan University, China</p> <p><b>Abstract:</b> This paper reviews real-time simulation techniques for destructible objects, focusing on three advanced approaches: sparse eigenmode-based fracture prediction, rigid-FEM hybrid frameworks, and Gram-Schmidt-constrained voxel methods for soft-body deformation. By balancing computational efficiency (e.g., GPU parallelism, virtual node algorithms) and physical fidelity (volume preservation, material anisotropy), these methods provide scalable solutions for game development. Case studies (e.g., Battlefield, Red Faction) demonstrate engine integration strategies, while future directions highlight the potential of machine learning and cloud computing for dynamic destruction.</p>
<p><b>AD543</b> <b>17:45-18:00</b></p>	<p>An Improved YOLOv8n Algorithm for Small Object Detection in Road Scenes  <b>Authors:</b> Qikang Song, Fan Yu  <b>Presenter:</b> Qikang Song, Beijing University of Civil Engineering and Architecture, China</p> <p><b>Abstract:</b> To address issues such as weak feature representation and inaccurate localization in small object detection under road scenes, this study proposes an improved YOLOv8n Algorithm for small object detection in road scenes, named SRNE-YOLOv8. Firstly, the feature map downsampled by a factor of 4 is connected to the detection layer, and a small-scale detection head is added to enhance the model's capability for detecting small objects. Secondly, the RFE module is introduced into the backbone network to improve the model's ability to capture contextual information. Then, the NWD-FIoU is constructed to replace the original loss function, improving the geometric matching accuracy between predicted and ground truth boxes and enhancing gradient backpropagation for low-IoU samples. Finally, the EVCBlock module is introduced in the head layer to enhance the model's cross-scale feature transmission ability. Experimental results show that compared with YOLOv8n, the improved SRNE-YOLOv8 model achieves increases of 3.7% and 2.3% in mAP@0.5 and mAP@0.5:0.95 respectively on the BDD100K dataset, and improvements of 2.3%, 6.7%, and 4.9% in Precision, Recall, and mAP@0.5 respectively on the KITTI dataset. The improved SRNE-YOLOv8 model demonstrates good detection accuracy for small objects in road scenes.</p>

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.