# 2025 7th International Conference on Image, Video and Signal Processing (IVSP 2025)

## March 4-6, 2025

## Meiji University Ikuta campus, Kawasaki, Japan

**Sponsored by**

明治大学 MEIJI UNIVERSITY

SINGAPORE INSTITUTE OF ELECTRONICS

acm Singapore Chapter

**Technically Supported by**

Saitama University

UNIVERSITY OF AUCKLAND Waipapa Taumata Rau NEW ZEALAND

1865 RIDER UNIVERSITY

UTAR UNIVERSITI TUNKU ABDUL RAHMAN

UPEC UNIVERSITÉ PARIS-EST CRÉTEIL VAL DE MARNE

THE UNIVERSITY of TULSA

輔仁大學理工學院 College of Science and Engineering Fu Jen Catholic University

**Indexed by**

Compendex on Engineering Village

Scopus

https://ivsp.net/

# Table of Contents

# Welcome Message

On behalf of the organizing committee, it is our great honor and pleasure to welcome you to the 2025 7th International Conference on Image, Video, and Signal Processing (IVSP 2025), taking place from 4th to 6th March 2025 at the Meiji University Ikuta Campus in Kawasaki, Japan.

IVSP 2025 brings together leading experts, researchers, and practitioners from around the globe to share cutting-edge advancements, innovative ideas, and collaborative insights in the fields of image, video, and signal processing. This conference serves as a premier platform for fostering interdisciplinary discussions, exploring emerging technologies, and addressing the challenges and opportunities in these rapidly evolving domains.

We are delighted to host this event at the prestigious Meiji University, a hub of academic excellence and innovation. The vibrant city of Kawasaki, with its rich cultural heritage and technological advancements, provides an inspiring backdrop for this gathering of minds.

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

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

We wish you a productive, inspiring, and enjoyable conference experience. Welcome to IVSP 2025!

IVSP 2025 Organizing Committee
February 2025

# Conference Committees

**Advisory Chair**
Wenwu Wang, University of Surrey, UK

**Conference Chairs**
Xudong Jiang, Nanyang Technological University, Singapore
Tae-Kyun Kim, Korea Advanced Institute of Science and Technology, Korea

**Program Co-Chairs**
Kiyoshi Hoshino, Meiji University, Professor Emeritus of University of Tsukuba, Japan
Kenneth K. M. Lam, The Hong Kong Polytechnic University, China

**Steering Co-Chairs**
Ke-Lin Du, Concordia University, Canada
Li-Wei Kang, NTNU

**Chapter Co-Chairs**
Yan Chai Hum, Universiti Tunku Abdul Rahman Sungai Long Campus, Malaysia
Qiuyu Zhu, Shanghai University, China
Tetsuya Shimamura, Saitama University, Japan

**Publicity Co-Chairs**
Naoki Igo, Tokyo Information Design Professional University, Japan
Hideaki Kimata, Kogakuin University, Japan
Tien-Ying Kuo, NTUT

**Technical Committee Chair**
Alice Othmani, University Paris-Est Créteil, France

**Technical Program Committees**
Tsutomu Kinoshita, Tohoku Gakuin University, Japan
Ling Xiao, The University of Tokyo, Japan
Rebeka Sultana, Tokyo University of Agriculture and Technology, Japan
Min-Shiang Hwang, Asia University
Marek R. Ogiela, AGH University of Krakow, Poland
Chuan Qin, University of Shanghai for Science and Technology, China
Suraiya Jabin, Jamia Millia Islamia, India

Xinwei Luo, Southeast University, China
Kavita Thakur, Pt. Ravi Shankar Shukla University, Raipur (C.G.), India
Yuan-Kai Wang, Fu Jen Catholic University
Francesco Zirilli, Universita di Roma La Sapienza, Italy
Deyun Wei, Xidian University, China
Xiaochen Yuan, Macao Polytechnic University, China
Zhi-Fang Yang, NTPU
Sheak Rashed Haider Noori, Daffodil International University, Bangladesh
Ming-Han Tsai, NYCU
Md Liakat Ali, Rider University, USA
Waleed Abdulla, The University of Auckland, New Zealand
Zahid Akthar, State University of New York Polytechnic Institute, USA
Wanwan Li, University of Tulsa, USA
Yi Wang, The Hong Kong Polytechnic University, China
Liming Zhang, University of Macau, China
Siriporn Dachasilaruk, Naresuan University, Thailand
Shervin R. Arashloo , Bilkent University,Turkey
Qiang Li, Northwestern Polytechnical University, China
Tengku Mohd Afendi Zulcaffle, Universiti Malaysia Sarawak, Malaysia
Roziana Binti Ramli, Northumbria University, UK
Punnarai Siricharoen, Chulalongkorn University,Thailand
Gianluca Zaza, University of Bari Aldo Moro, Italy
Riad I. Hammoud, DynaVox Technologies, USA
Ahmad Al Badawi, Applied Cryptography, USA
Fangli Ying, East China University of Science and Technology, China
Costantino Grana, University of Modena and Reggio Emilia, Italy
Patrice Jean Delmas, The University of Auckland, New Zealand
Caifeng Shan, Philips Research, The Netherlands
Hamimah binti Ujir, Universiti Malaysia Sarawak, Malaysia
Evgin Goceri, Akdeniz University, Turkey
Jan-Ray Liao, National Chung Hsing University
I-Cheng Chang, National Dong Hwa University
Nabilah Ibrahim, Universiti Tun Hussein Onn Malaysia, Malaysia
Dongyun Lin, Xiamen University, China
Po-Chyi Su, National Central University
Mohamed Maher Ben Ismail, King Saud University, Saudi Arabia
Jiunn-Lin Wu, National Chung Hsing University
Chao-Lung Chou, Feng Chia University
Suk-Ho Lee, Dongseo University, Korea

Chih-Chang Yu, Chung Yuan Christian University
Ran-Zan Wang, Yuan Ze University
Edward T.-H. Chu, National Yunlin University of Science and Technology
Yung-Chen Chou, Asia University
Jae Youn Hwang, Daegu Gyeongbuk Institute of Science&Technology, Korea
Shinfeng Lin, National Dong Hwa University
Sukanya Phongsuphap, Mahidol University, Thailand
Mohd Shafry Mohd Rahim, University Technology Malaysia, Malaysia
Tien Tsin Wong, Monash University, Australia
Huan Chen, National Chung Hsing University
Yung Cheng Hsu, National Central University
Guo-siang Lin, National Chin-Yi University of Technology
Bo-Wei Chen, National Sun Yat-sen University
Chih-Yuan Hsu, Southern Taiwan University of Science and Technology
Yung Gi Wu, Chang Jung Christian University

# Guideline for Onsite Attendance

## Important Notes

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

## Oral & Poster Presentation

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

# Guideline for Online Attendance

## Platform: ZOOM

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

## How to use ZOOM

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

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

## How to join the conference online

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

## Time Zone

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

## Device

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

## Online Room Information

Online Room Information

Zoom ID: **849 2479 0461**

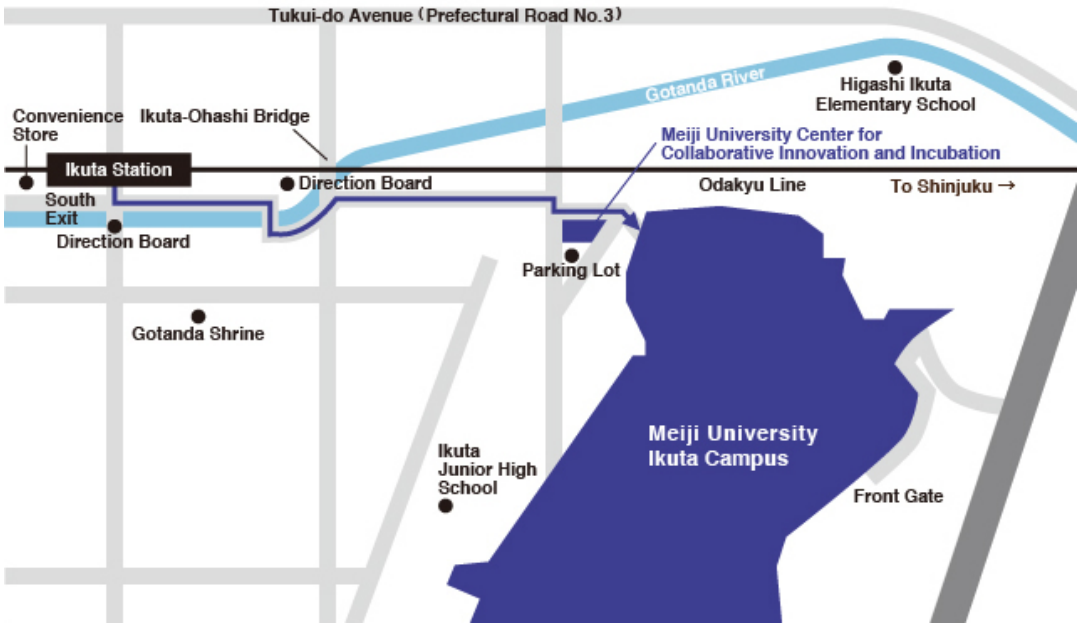
Zoom Link: **https://us02web.zoom.us/j/84924790461**

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

| Role | Format | Example |
|---|---|---|
| Conference Committee | Position-Name | Conference Chair-Prof. |
| Keynote/ Invited Speaker | Position-Name | Keynote/Invited Speaker-Prof. |
| Author | Session Number-Paper ID-Name | S1-MP0001-Name |
| Delegate | Delegate-Name | Delegate-Name |

# Conference Venue

**Central School Building(中央校舎), Meiji University Ikuta Campus**
**https://www.meiji.ac.jp/cip/english/about/campus/ikuta.html**
**Address: 1 Chome-1-1 Higashimita, Tama Ward, Kawasaki, Kanagawa 214-8571**

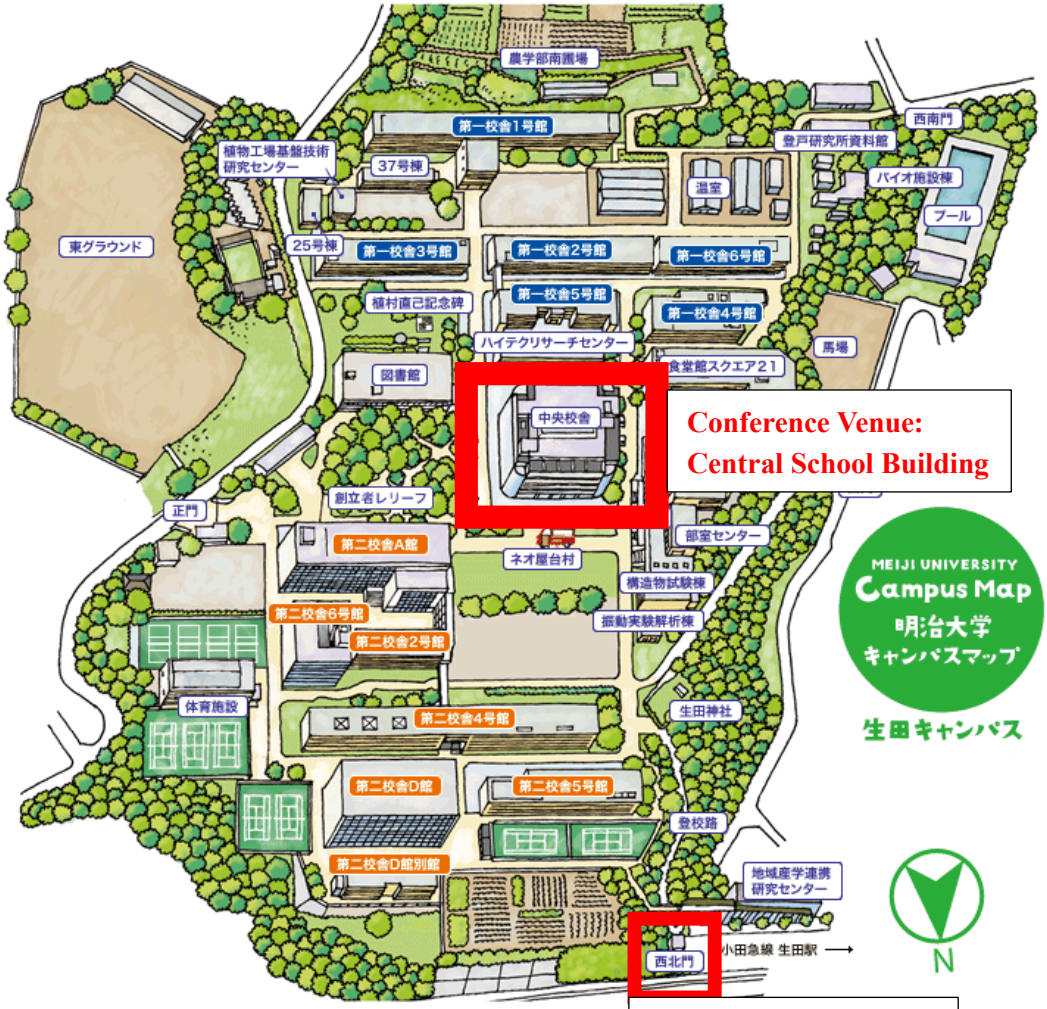



The Ikuta Campus is a vast area situated in the Tama Hills. The campus has several advanced research facilities such as The High-Tech Research Center and Ikuta Structural Test Building to greenhouses and fields. Students from the School of Science and Technology and from the School of Agriculture, as well as graduate students, study in this rich natural environment. For students busy in experiments and research, facilities for daily life such as a cafeteria, shops and ATMs are available on campus. Also, our library is open on Sundays. Thus, the campus is planned to function for the convenience of students.

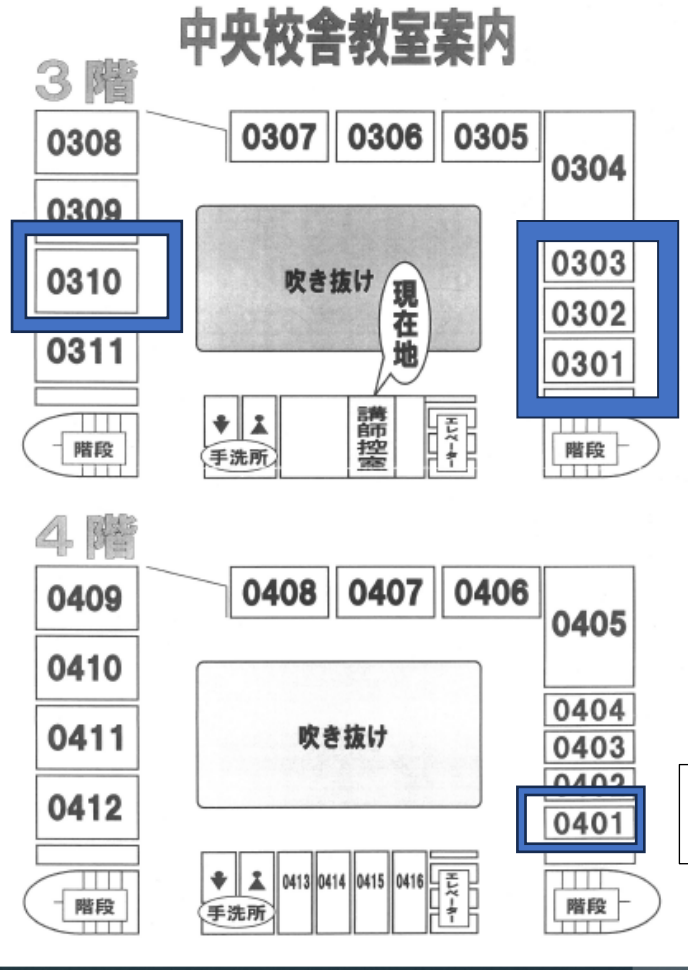
Directions:
10 minutes on foot from Odakyu Line (both semi-express, and local trains), Ikuta Station
URL: http://www.meiji.ac.jp/koho/campus_guide/ikuta/campus.html

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

# Conference Rooms Directions



Room for Keynote Speeches & Invited Speeches on March 5




Rooms for Afternoon Sessions on March 5

Room for registration on March 4

# Simple Program

## March 4, 2025 (Tuesday)

**Onsite Registration**

| **Registration Time: 10:00-16:00** |
|---|
| **Venue: Central School Building(中央校舍), Meiji University Ikuta Campus**<br>**Conference Room: Room401** |
| **Address: 1 Chome-1-1 Higashimita, Tama Ward, Kawasaki, Kanagawa 214-8571** |
| **1. Arrive at Conference Room (401), Central School Building(中央校舍), Meiji University Ikuta Campus;**<br>**2. Inform the conference staff of your paper ID;**<br>**3. Sign your name on the Participants list;**<br>**4. Sign your name on Lunch & Dinner requirement list;**<br>**5. Check your conference kits;**<br>**6. Finish registration.** |

**Online Test**
**Time Zone: GMT+9**

| **Online Test**<br>**Zoom ID:** 849 2479 0461<br>**Link: https://us02web.zoom.us/j/84924790461** | **Duration** |
|---|---|
| Prof. Tae-Kyun Kim | 10:00-10:10 |
| Prof. Maxim Bakaev | 10:10-10:20 |
| Asst. Prof. Xiangyu Yue | 10:20-10:30 |
| Session 7 (MP6001, MP0047, MP0007, MP6002, MP0057, MP0010, MP6003, MP5007) | 10:30-11:10 |

Note: If you want to do online zoom test after 11:10, please contact your conference secretary.

# March 5, 2025 (Wednesday)

| Morning Sessions<br>Venue: Central School Building(中央校舍) Room: 310 | | Duration |
|---|---|---|
| Opening Remark | Prof. Kiyoshi Hoshino, Meiji University,<br>(Professor Emeritus) University of Tsukuba, Japan | 9:55-10:05 |
| Keynote Speech 1 | Prof. Xudong Jiang (IEEE Fellow), Nanyang Technological University, Singapore<br>Speech Title: Critical Foundations and Techniques of Machine Learning That Drive AI | 10:05-10:35 |
| Keynote Speech 2 | Prof. Kenneth K. M. Lam, The Hong Kong Polytechnic University, China<br>Speech Title: Recent Research on Facial Expression Recognition | 10:35-11:05 |
| **Group Photo & Coffee Break** | | **11:05-11:20** |
| Keynote Speech 3 | Prof. Kiyoshi Hoshino, Meiji University,<br>(Professor Emeritus) University of Tsukuba, Japan<br>Speech Title: A new methodology for estimating eye rotational movement and gaze using a camera placed almost at the side of human eyes | 11:20-11:50 |
| Invited Speech 1 | Prof. Jun Sakakibara, Meiji University, Japan<br>Speech Title: The 100-Eyes Particle Image Velocimetry Using a Mirror Array | 11:50-12:10 |
| Invited Speech 2 | Prof. Wen-Syan Li, Seoul National University, Korea<br>Speech Title: Building Complex LLM-based Q&A Systems: Challenges, Blueprint, and Case Studies | 12:10-12:30 |
| **Lunch at the Canteen** | | **12:30-13:30** |

<table>
<tr><td colspan="4">Afternoon Invited Speeches<br>Venue: Central School Building(中央校舍)  Room: 310</td></tr>
<tr><td>Invited Speech 3</td><td colspan="2">Prof. Hiromasa Oku, Gunma University, Japan<br>Speech Title: Dynamic Image Control based on high-speed image processing coupled with novel optical devices</td><td>13:30-13:50</td></tr>
<tr><td>Invited Speech 4</td><td colspan="2">Prof. Yoosoo Oh, Daegu University, Korea<br>Speech Title: Educational and Industrial Innovation through No-Code AI Tools: A New Paradigm for Strengthening Digital Competencies</td><td>13:50-14:10</td></tr>
<tr><td>Invited Speech 5</td><td colspan="2">Dr. Ling Xiao, The University of Tokyo, Japan<br>Speech Title: Revolutionizing AI Applications Innovative Approaches with Multimodal Large Language Models</td><td>14:10-14:30</td></tr>
<tr><td>Invited Speech 6</td><td colspan="2">Dr. Taku Itami, Meiji University, Japan<br>Speech Title: Devices supporting daily life focusing on smart mechatronics</td><td>14:30-14:50</td></tr>
<tr><td colspan="3">Coffee Break& Poster Session<br>Topic: Image detection and computational models<br>Presentations: MP0021, MP0058-A, MP4001</td><td>14:50-15:05</td></tr>
<tr><td colspan="3">Afternoon Onsite Parallel Sessions</td><td>Duration</td></tr>
<tr><td>Room 301</td><td>Room 302</td><td>Room 303</td><td></td></tr>
<tr><td>Session 1<br>Topic: Image analysis and methods<br>Session Chair: Prof. Shih-Lin Wu, Chang Gung University, Taiwan<br>Presentations: MP0016, MP0011, MP0013, MP0038, MP0014, MP0024</td><td>Session 2<br>Topic: Intelligent recognition technology and applications<br>Session Chair: Prof. Yoosoo Oh, Daegu University, Korea<br>Presentations: MP5024, MP5009, MP5016, MP5027-A, MP0002, MP0037, MP0055</td><td>Session 3<br>Topic: Image detection models and algorithms<br>Session Chair: Prof. Guo-shiang Lin, National Chin-Yi University of Technology, Taiwan<br>Presentations: MP0004, MP0026, MP0044, MP0041, MP0064-A, MP0008, MP0066-A</td><td>15:05-16:50</td></tr>
</table>

| | | | |
|---|---|---|---|
| **Session 4**<br>**Topic:** Computer vision and image processing<br>**Session Chair:** Prof. Suk-Ho Lee, Dongseo University, Korea<br>**Presentations:** MP0005, MP0015, MP0045, MP0048, MP0065-A, MP0069, MP0043-A | **Session 5**<br>**Topic:** Computer-aided systems and interactive design<br>**Session Chair:** Prof. Kikuo Asai, The Open University of Japan<br>**Presentations:** MP5001, MP5003-A, MP5004-A, MP5010, MP5011, MP0019, MP0060 | **Session 6**<br>**Topic:** Image encryption and security verification<br>**Session Chair:** Prof. Ran-Zan Wang, Yuan Ze University, Taiwan<br>**Presentations:** MP0023, MP0036, MP0034, MP0050, MP0070-A, MP0020, MP0071-A | 16:50-18:35 |
| **Dinner Time** | | | **18:35-20:00** |

Note:

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

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

# March 6, 2025 (GMT+9)

| Morning Sessions<br>Online Speeches and online sessions<br>Zoom ID: 849 2479 0461<br>Zoom link: https://us02web.zoom.us/j/84924790461 | | Duration |
|---|---|---|
| Keynote Speech 4 | Prof. Tae-Kyun Kim, Korea Advanced Institute of Science and Technology, Korea<br>Speech Title: Image and 3D Shape Generation | 10:00-10:30 |
| Invited Speech 7 | Prof. Maxim Bakaev, Novosibirsk State Technical University, Russia<br>Speech Title: Evoking Personas and Specialists from Large Language Models: going beyond the optimistic common-sense assistants | 10:30-10:50 |
| Invited Speech 8 | Dr. Xiangyu Yue, The Chinese University of Hong Kong, China<br>Speech Title: Towards unified Multimodal Learning | 10:50-11:10 |
| **Break Time** | | **11:10-11:25** |
| Session 7 | **Topic:** Intelligent image processing and application technology based on machine learning<br>**Session Chair:** Prof. Maxim Bakaev, Novosibirsk State Technical University, Russia<br>**Presentations:** MP6001, MP0047, MP0007, MP6002, MP0057, MP0010, MP6003, MP5007 | 11:25-13:25 |

# Detailed Program

## March 5, 2025 (Wednesday)

| | |
|---|---|
| **9:55-10:05** | **Opening Remark** |
| **Venue** | **Central School Building(中央校舍), Room 310** |

### Prof. Kiyoshi Hoshino

### Program Chair

**Meiji University,**
**(Professor Emeritus) University of Tsukuba, Japan**

Prof. Kiyoshi Hoshino is now a Full Professor at Meiji University. He is awarded Professor Emeritus of University of Tsukuba in 2023. He served as a member of the "cultivation of human resources in the information science field" WG, Special Coordination Funds for the Promotion of Science and Technology, MEXT, a member of "Committee for Comfort 3D Fundamental Technology Promotion", JEITA, the General Conference Chair of the 43rd Annual Meeting of Japanese Society of Biofeedback Research, and a councilor and director of the Ibaraki Sports Association. He received IJCAI-09 AI Video Award, iFAN 2010 Best Paper Award, Laval Virtual Awards in 2009, 2013 and 2014, ISER 2015 Best Paper Award, and several domestic and international awards.

# Keynote Speech 1

| | | | |
|---|---|---|---|
| **Host** | Prof. Kiyoshi Hoshino<br>Meiji University, (Professor Emeritus) University of Tsukuba, Japan | **Time** | 10:05-10:35, March 5, 2025 |
| | | **Venue** | Central School Building(中央校舍), Room 310 |

## Prof. Xudong Jiang (IEEE Fellow)

**Nanyang Technological University, Singapore**

Dr. Xudong Jiang received the B.Eng. and M.Eng. from the University of Electronic Science and Technology of China (UESTC), and the Ph.D. degree from Helmut Schmidt University, Hamburg, Germany. From 1986 to 1993, he was a Lecturer with UESTC, where he received two Science and Technology Awards from the Ministry for Electronic Industry of China. From 1998 to 2004, he was with the Institute for Infocomm Research, A-Star, Singapore, as a Lead Scientist and the Head of the Biometrics Laboratory, where he developed a system that achieved the most efficiency and the second most accuracy at the International Fingerprint Verification Competition in 2000. He joined Nanyang Technological University (NTU), Singapore, as a Faculty Member, in 2004, and served as the Director of the Centre for Information Security from 2005 to 2011. Currently, he is a professor in NTU. Dr Jiang holds 7 patents and has authored over 150 papers with over 40 papers in the IEEE journals, including 11 papers in IEEE T-IP and 6 papers in IEEE T-PAMI. Three of his papers have been listed as the top 1% highly cited papers in the academic field of Engineering by Essential Science Indicators. He served as IFS TC Member of the IEEE Signal Processing Society from 2015 to 2017, Associate Editor for IEEE SPL from 2014 to 2018, Associate Editor for IEEE T-IP from 2016 to 2020 and the founding editorial board member for IET Biometrics form 2012 to 2019. Dr Jiang is currently an IEEE Fellow and serves as Senior Area Editor for IEEE T-IP and Editor-in-Chief for IET Biometrics. His current research interests include image processing, pattern recognition, computer vision, machine learning, and biometrics.

## Speech Contents

### Speech Title: Critical Foundations and Techniques of Machine Learning That Drive AI

**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 talk 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.

# Keynote Speech 2

| | | | |
|---|---|---|---|
| **Host** | Prof. Kiyoshi Hoshino<br>Meiji University, (Professor Emeritus) University of Tsukuba, Japan | **Time** | 10:35-11:05, March 5, 2025 |
| | | **Venue** | Central School Building(中央校舍), Room 310 |

## Prof. Kenneth K. M. Lam

**The Hong Kong Polytechnic University, China**

Kin-Man Lam received his Associateship in Electronic Engineering with distinction from The Hong Kong Polytechnic University (formerly Hong Kong Polytechnic) in 1986, his M.Sc. degree in Communication Engineering from the Department of Electrical Engineering, Imperial College, U.K., in 1987, and his Ph.D. degree from the Department of Electrical Engineering, University of Sydney, Australia, in 1996. From 1990 to 1993, he was a lecturer at the Department of Electronic Engineering of The Hong Kong Polytechnic University. He rejoined the Department of Electronic and Information Engineering at The Hong Kong Polytechnic University as an Assistant Professor in October 1996. He became an Associate Professor in 1999 and has been a Professor since 2010. Prof. Lam has been actively involved in professional activities. He has served as a member of the organizing committee or program committee for many international conferences. He was the Chairman of the IEEE Hong Kong Chapter of Signal Processing between 2006 and 2008 and served as the Director-Student Services and the Director-Membership Services of the IEEE SPS between 2012 and 2014 and between 2015 and 2017, respectively. He also held the positions of VP-Member Relations and Development and VP-Publications of the Asia-Pacific Signal and Information Processing Association (APSIPA) between 2014 and 2017 and between 2017 and 2021, respectively. Prof. Lam was an Associate Editor of IEEE Transactions on Image Processing from 2009 to 2014 and of Digital Signal Processing from 2014 to 2018. He also served as an Editor of HKIE Transactions between 2013 and 2018 and as an Area Editor of the IEEE Signal Processing Magazine between 2015 and 2017. Currently, he is the IEEE SPS VP-Membership.

Additionally, he serves as a Senior Editorial Board member of APSIPA Transactions on Signal and Information Processing and an Associate Editor of EURASIP International Journal on Image and Video Processing. His current research interests include image and video processing, computer vision, and human face analysis and recognition.

## Speech Contents

### Speech Title: Recent Research on Facial Expression Recognition

**Abstract:** Facial expression recognition (FER) has been a focal point of research for decades, with significant advancements in learning discriminative representations from facial observations. In this talk, we will explore the latest developments in FER, beginning with an overview of loss functions specifically designed to enhance FER. We will then introduce a novel geometry-aware FER framework that leverages both geometric and appearance-based knowledge to improve FER performance. To further enhance the learned representations, we will delve into attention mechanisms employed in FER. These mechanisms play a crucial role in enhancing the robustness and effectiveness of FER algorithms. Extensive experiments have demonstrated that our frameworks achieve promising performance for geometry-based FER, exhibiting remarkable generalization and robustness in real-world applications. However, traditional FER research has primarily focused on the six basic facial expressions: happiness, sadness, fear, anger, disgust, and surprise. In reality, facial expressions can be much more complex, often conveying multiple emotions simultaneously – referred to as compound expressions. Compound facial expression recognition (CFER) in the wild is a much more challenging task, with limited research currently addressing this area. To tackle CFER, we have developed a novel loss function, termed bi-center loss, which is built upon the center loss function. Unlike the center loss, which considers all categories equally, the bi-center loss enables deep neural networks to learn compound emotion features by leveraging basic emotion centers. Furthermore, we will discuss our current research on CFER based on the datasets for basic expression recognition. Preliminary results showcasing the potential of our research in CFER will be presented.

# Keynote Speech 3

| | | | |
|---|---|---|---|
| **Host** | Prof. Xudong Jiang<br>Nanyang Technological University, Singapore | **Time** | 11:20-11:50, March 5, 2025 |
| | | **Venue** | Central School Building(中央校舍), Room 310 |

## Prof. Kiyoshi Hoshino

**Meiji University,**
**(Professor Emeritus) University of Tsukuba, Japan**

He received two doctor's degrees; one in Medical Science in 1993, and the other in Engineering in 1996, from the University of Tokyo respectively. From 1993 to 1995, he was an Assistant Professor at Tokyo Medical and Dental University School of Medicine. From 1995 to 2002, he was an Associate Professor at University of the Ryukyus. From 2002 to 2023, he served at the Biological Cybernetics Lab of University of Tsukuba as an Associate Professor and a Full Professor. He is now a Full Professor at Meiji University. He is awarded Professor Emeritus of University of Tsukuba in 2023. From 1998 to 2001, he was jointly appointed as a senior researcher of the PRESTO "Information and Human Activity" project of the Japan Science and Technology Agency (JST). From 2002 to 2005, he was a project leader of a SORST project of JST. He served as a member of the “cultivation of human resources in the information science field” WG, Special Coordination Funds for the Promotion of Science and Technology, MEXT, a member of “Committee for Comfort 3D Fundamental Technology Promotion”, JEITA, the General Conference Chair of the 43rd Annual Meeting of Japanese Society of Biofeedback Research, and a councilor and director of the Ibaraki Sports Association. He received IJCAI-09 AI Video Award, iFAN 2010 Best Paper Award, Laval Virtual Awards in 2009, 2013 and 2014, ISER 2015 Best Paper Award, and several domestic and international awards.

## Speech Contents

### Speech Title: A new methodology for estimating eye rotational movement and gaze using a camera placed almost at the side of human eyes

**Abstract:** In this presentation, the author would like to introduce to you some useful ways to utilize human biomedical information. The first is the measurement of eye rotation. Our eyes not only move along the x- and y-axes, but also rotate around the z-axis or optical axis. Since these eye rotational movements tend to occur during dizziness, carsickness, motion sickness, or discomfort feeling, they can be used as a physiological indicator of our internal state. The author will show how this can be estimated with high accuracy using a small camera placed almost at the side of the eyeballs. The second is gaze estimation. This is a physiological indicator of what objects and to what extent a person is interested in. However, when a camera is used to capture images of the eyes and estimate the direction of gaze, the camera image is two-dimensional and eye movements are three-dimensional. Therefore, to achieve highly accurate gaze estimation with a camera installed at the side, it is necessary to introduce some new methodology. The author will introduce that as well.

# Keynote Speech 4 (Online)

| | |
|---|---|
| **Time** | **10:00-10:30, March 6, 2025 (GMT+9)** |
| **Zoom ID** | **849 2479 0461** |
| **Zoom Link:** | **https://us02web.zoom.us/j/84924790461** |

## Prof. Tae-Kyun Kim

**Korea Advanced Institute of Science and Technology, Korea**

Tae-Kyun (T-K) Kim is a full Professor and the director of Computer Vision and Learning Lab at School of Computing, KAIST since 2020, and has been an adjunct reader of Imperial College London (ICL), UK for 2020-2024. He led Computer Vision and Learning Lab at ICL during 2010-2020. He obtained his PhD from Univ. of Cambridge in 2008 and Junior Research Fellowship (governing body) of Sidney Sussex College, Univ. of Cambridge during 2007-2010. His BSc and MSc are from KAIST. His research interests primarily lie in machine (deep) learning for 3D computer vision and generative AI, including: articulated 3D hand/body reconstruction, face analysis and recognition, 6D object pose estimation, activity recognition, object detection/tracking, active robot vision, which lead to novel active and interactive visual sensing. He has co-authored over 100 academic papers in top-tier conferences and journals in the field, and has co-organised series of HANDS workshops and 6D Object Pose workshops (in conjunction with CVPR/ICCV/ECCV) since 2015. He was the general chair of BMVC17 in London, the program co-chair of BMVC23, and is Associate Editor of Pattern Recognition Journal, Image and Vision Computing Journal. He regularly serves as an Area Chair for top-tier vision/ML conferences. He received KUKA best service robotics paper award at ICRA 2014, and 2016 best paper award by the ASCE Journal of Computing in Civil Engineering, and the best paper finalist at CVPR 2020, and his co-authored algorithm for face image representation is an international standard of MPEG-7 ISO/IEC.

## Speech Contents

### Speech Title: Image and 3D Shape Generation

**Abstract:** Followed by the motivations and challenges of 3D video generation, we present our recent works published in CVPR and ECCV 2024.
This includes InterHandGen (Two-hand interaction generation via cascaded reverse diffusion), arbitrary-scale upscaling by latent diffusion model with implicit neural decoder, prompt augmentation for self-supervised text-guided image editing, BITT (Bi-directional texture reconstruction of interacting two hands from a single image). We emphasize the use of diffusion model, diffusion sampling, implicit functions, and self-supervised learning.

# Invited Speech 1

| | | | |
|---|---|---|---|
| **Host** | Prof. Xudong Jiang<br>Nanyang Technological University, Singapore | **Time** | 11:50-12:10, March 5, 2025 |
| | | **Venue** | Central School Building(中央校舍), Room 310 |

## Prof. Jun Sakakibara

**Meiji University, Japan**

Dr. Jun Sakakibara received his Ph.D. from the Department of Mechanical Engineering at Keio University in 1996. He was a visiting scholar at the University of Illinois at Urbana-Champaign from 1996 to 1997 before joining the Department of Engineering Mechanics at the University of Tsukuba in 1997. He later returned to the University of Illinois for a second visit from 2000 to 2001. In 2013, he joined the Department of Mechanical Engineering at Meiji University. Dr. Sakakibara is an associate editor of the Journal of Visualization and a member of the editorial board of Experiments in Fluids. He previously chaired the 46th Annual Meeting of the Visualization Society of Japan and is scheduled to chair the 16th International Symposium on Particle Image Velocimetry in 2025. He was serving as the President of the Visualization Society of Japan in 2024. His current research interests include experimental studies of complex turbulent flows, with particular emphasis on flows associated with wall-bounded systems, separated flows, and flows around wings or obstacles. He is also dedicated to the advancement of optical flow measurement techniques, including particle image velocimetry and laser-induced fluorescence.

## Speech Contents

### Speech Title: The 100-Eyes Particle Image Velocimetry Using a Mirror Array

**Abstract:** Particle Image Velocimetry (PIV) is widely used for velocity measurements in fluid flow fields due to its ability to capture spatially resolved flow structures. However, it suffers from relatively high random errors, which limit its accuracy in resolving the high-wavenumber range of turbulence spectra. To address this limitation, we applied the "Multiple Eye PIV" method, which captures images of tracer particles from multiple directions and reduces errors by averaging the positions of detected particles. This study demonstrates the method's effectiveness in measuring the energy spectrum of turbulent pipe flow and resolving higher wavenumbers.

We employed a mirror array comprising 110 flat mirrors (10 mm × 10 mm) arranged in an axisymmetric ellipsoid to capture particle images. Three-dimensional velocity fields were reconstructed using 3D particle tracking velocimetry (PTV) and tomographic PIV/PTV techniques.

Our results indicate that the energy spectrum deviated from that obtained by direct numerical simulation (DNS) at a specific wavenumber, forming a plateau. Increasing the number of mirrors from 10 to 100 shifted this plateau downward by more than an order of magnitude, enabling resolution of finer-scale turbulence. This highlights the effectiveness of the Multiple Eye PIV method in capturing smaller eddies and improving velocity measurements in turbulent flows.

# Invited Speech 2

| | | | |
|---|---|---|---|
| **Host** | Prof. Xudong Jiang<br>Nanyang Technological University, Singapore | **Time** | 12:10-12:30, March 5, 2025 |
| | | **Venue** | Central School Building(中央校舍), Room 310 |

## Prof. Wen-Syan Li

**Seoul National University, Korea**

Dr. Wen-Syan Li joined the Graduate School of Data Science, SNU as a Full Professor in March 2020 and became a Foreign Fellow of the Brain Pool Program under the National Research Foundation of Korea in June 2020. Before joining SNU, he was Senior Vice President of SAP SE and Head of SAP Customer Innovation & Strategic Projects – Asia Pacific, Japan, & Greater China. His team worked on the new applications in the area of digital supply chain and strategic engagements with key accounts such as Huawei, NTT, Intel, and Lenovo in the area of IoT and SAP Hana. His team was also responsible for building Predictive Analytics capabilities in SAP's in-memory database HANA. He received a Ph.D. degree in Computer Science from Northwestern University (USA). He also has an MBA degree in Finance. Before joining SAP, he was with IBM Almaden Research Center located, NEC Research, and NEC Venture Capital in the USA. He has co-edited 3 books published by Springer, co-authored more than 100 journal articles and conference papers in various areas, and co-invented 82 granted US patents. His research interests include AI, LLM applications, data & knowledge management, and applying AI to solving real-world problems.

## Speech Contents

### Speech Title: Building Complex LLM-based Q&A Systems: Challenges, Blueprint, and Case Studies

**Abstract:** In the Internet era, applications were built based on web servers, application servers, and database architecture. As we enter the era of AI, the front end is replaced by the NL (Natural Language) layer, the middle tier is replaced by LLM (Large Language Model) with agents/agentic workflow and many emerging RAG/reasoning/planning/inferencing techniques, and the backend is replaced by documents embedded in vector databases, web search, APIs, and knowledge bases. In this talk, I present the challenges of building complex LLM-based Q&A systems and blueprint / architectural design considerations to address these challenges, including various techniques used in DeepSeek. I use two applications-in-development as case studies: (1) an urban planning application on top of LLM (i.e., regulation) and geospatial systems (i.e., maps and tabular data for points of interest) and (2) a financial document Q&A system. Finally, I present interesting research topics and application areas.

# Invited Speech 3

| | | | |
|---|---|---|---|
| **Host** | Prof. Kenneth K. M. Lam<br>The Hong Kong Polytechnic University, China | **Time** | 13:30-13:50, March 5, 2025 |
| | | **Venue** | Central School Building(中央校舍), Room 310 |

## Prof. Hiromasa Oku

**Gunma University, Japan**

Hiromasa Oku received the Dr. (Eng.) degree in mathematical engineering and information physics from The University of Tokyo, Japan, in 2003. He was a Researcher with PRESTO, Japan Science and Technology Agency, from 2003 to 2005. He was a Research Associate (2005–2007), an Assistant Professor (2007–2011), and a Lecturer/an Assistant Professor (2011–2014) with The University of Tokyo. He has been an Associate Professor with the School of Science and Technology, Gunma University, since 2014, where he is currently a Professor with the Faculty of Informatics. He has received numerous awards in robotics, virtual reality, and human-computer interaction, including the Best Paper Award at VRST 2017, the Advanced Robotics Best Paper Award (2016), and multiple honors from JSME, SICE, and METI. His research interests include high-speed image processing, high-speed optical devices, and dynamic image control.

## Speech Contents

### Speech Title: Dynamic Image Control based on high-speed image processing coupled with novel optical devices

**Abstract:** While image recognition technology has advanced to a practical level with the recent breakthroughs in machine learning, there are still many unexplored areas in image utilization systems. Many optical systems and optical devices used to acquire images do not assume image recognition methods, leaving room for improvement, especially in response time. The author's group have been pointing out that high-speed imaging optics are also required, especially in systems based on high-speed image recognition. In particular, we have proposed a technology called "Dynamic Image Control" that enables dynamic image capture and projection by combining high-speed image processing and high-speed optics. This talk will introduce the basic concepts of Dynamic Image Control and its applications in combination with high-speed gaze controller using rotational mirrors, high-speed variable focus liquid lenses, and edible optical devices. For example, the high-speed gaze controller is an optical system that can control the optical axis of a camera or projector at high speed. It was applied to high-speed tracking of the field of view to dynamic objects, and also new aerial display based on continuous projection onto a flying screen.

# Invited Speech 4

| | | | |
|---|---|---|---|
| **Host** | Prof. Kenneth K. M. Lam<br>The Hong Kong Polytechnic University, China | **Time** | 13:50-14:10, March 5, 2025 |
| | | **Venue** | Central School Building(中央校舍), Room 310 |

## Prof. Yoosoo Oh

**Daegu University, Korea**

Yoosoo Oh received his Bachelor's degree in the Department of Electronics and Engineering from Kyungpook National University in 2002. He obtained his Master's degree in the Department of Information and Communications from Gwangju Institute of Science and Technology (GIST) in 2003. In 2010, he received his Ph.D. degree in the School of Information and Mechatronics from GIST. In the meantime, he was an executed team leader at Culture Technology Institute, GIST, 2010-2012. In 2011, he worked as a visiting scholar at Lancaster University in the UK. In September 2012, he joined Daegu University, where he is currently a professor at the School of Computer & Information Engineering. He worked as a center director of Smart Drone Center, AZIT Makerspace Center, and Gyeongbuk Technopark Daegu University Center. Also, he has served on the Board of Directors of KSIIS since 2019 and HCI Korea since 2016. From 2017~2019, he worked as a center director of the Mixed Reality Convergence Research Center at Daegu University. From 2015-2017, He worked as a director in the Enterprise Supporting Office of LINC Project Group, Daegu University. His research interests include Activity Fusion & Reasoning, Machine Learning, Context-aware Middleware, Human-Computer Interaction, etc. He has received several awards: the Worldwide Lifetime Achievement in 2017 Albert Nelson Marquis Lifetime Achievement Award by MARQUIS WHO's Who, 2017 awards from the Korea Society of Industrial Information Systems, and the 2016 Excellent Research Award from Daegu University.

## Speech Contents

### Speech Title: Educational and Industrial Innovation through No-Code AI Tools: A New Paradigm for Strengthening Digital Competencies

**Abstract:** As nations worldwide recognize Artificial Intelligence (AI) as a key capability, the demand for AI education and EdTech is rising rapidly. However, the shortage of specialized personnel, high development costs, and limited technological access among teachers and SMEs continue to pose significant challenges to AI adoption.

This talk introduces a No-Code–based block-type AI learning tool designed to address these issues. Integrating hardware blocks with a No-Code AI framework allows users to practice machine learning and deep learning algorithms effortlessly without requiring programming knowledge. Through simple "Pick & Drop" of hardware blocks or "Drag & Drop" operations on a software interface, data preprocessing, training, and inference processes can be visually configured.

The global EdTech market is experiencing double-digit growth annually, and No-Code/Low-Code platforms are expected to expand well beyond education into various industrial fields. Unlike existing tools that often rely on basic coding or limited AI features, No-Code AI block tools offer a wide range of built-in ML algorithms, RFID/sensor-based automatic recognition, and an interface accessible to non-experts, all backed by unique patented technology for a distinct competitive edge.

In South Korea, the government's revised curriculum for 2022 emphasizes the importance of enhancing digital literacy and expanding information education. It supports teacher training, develops specialized talents, and creates instructional content to improve AI competence. With such institutional support, No-Code AI learning tools can significantly increase student engagement and motivation, reduce teachers' lesson preparation time, and ultimately facilitate more straightforward AI utilization in industrial sectors, including SMEs.

In this invited talk, we will discuss the multifaceted potential of No-Code AI tools to bring transformative changes across education and industry. We will also propose strategies for a comprehensive digital transition that is inclusive and supported by international collaboration and systematic policy initiatives.

# Invited Speech 5

| | | | |
|---|---|---|---|
| **Host** | Prof. Kenneth K. M. Lam<br>The Hong Kong Polytechnic University, China | **Time** | 14:10-14:30, March 5, 2025 |
| | | **Venue** | Central School Building(中央校舍), Room 310 |

## Dr. Ling Xiao

**The University of Tokyo, Japan**

Ling Xiao (IEEE and ACM member) is an Assistant Professor at the University of Tokyo (UTokyo) (2023.10 to now) and an adjunct Project Assistant Professor at BeyondAI (2023.11 to now). Previously, she was a postdoctoral researcher at UTokyo (2021.06–2023.09) and an adjunct Project Researcher at BeyondAI (2022.07–2023.10). She obtained her Ph.D. from Huazhong University of Science and Technology in Dec. 2020 and was a visiting student at the University of Queensland, Australia (2018.10–2019.11) under the support of the CSC scholarship. Her research interests focus on AI, recommendation systems, continual learning, large multimodal models, anomaly detections, etc. She has published over 20 peer-reviewed journal and conference papers, and more than 10 Japanese domestic conference papers. She is also served as a reviewer for several major conferences (CVPR 2024, AAAI 2024, ICSSSP 2024, MMM 2024, etc.) and journals (IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), IEICE, etc).

## Speech Contents

### Speech Title: Revolutionizing AI Applications Innovative Approaches with Multimodal Large Language Models

**Abstract:** With the widespread adoption of large language models (LLMs) and their exceptional performance in tasks such as language reasoning, semantic understanding, and knowledge generation, research on large models has become a key driver of breakthroughs in artificial intelligence. This talk will present our recent research on multimodal large language models (MLLMs), including GPT-based zero-shot multimodal models, retrieval-augmented generation powered MLLMs, and prompt-optimization-based MLLMs. These models offer strong, general-purpose foundational support for a variety of tasks, including tourism recommendation, video summarization, advertisement design, and outdoor navigation.

# Invited Speech 6

| | | | |
|---|---|---|---|
| **Host** | Prof. Kenneth K. M. Lam<br>The Hong Kong Polytechnic University, China | **Time** | 14:10-14:30, March 5, 2025 |
| | | **Venue** | Central School Building(中央校舍), Room 310 |

## Dr. Taku Itami

**Meiji University, Japan**

Dr. Taku Itami is currently a Senior Assistant Professor at the Department of Electronics and Bioinformatics, School of Science and Technology, Meiji University, specializing in Robotics and Rehabilitation Engineering. Dr. Itami previously conducted research in the field of Mechatronics at the Department of Systems Engineering, Graduate School of Mie University. He is currently pursuing a doctoral degree in Orthopedics at the Graduate School of Medicine, Gifu University. With this background, he is engaged in the development of devices and systems essential for daily life, with a particular focus on ensuring that the devices developed truly meet the needs of the users.

## Speech Contents

### Speech Title: Devices supporting daily life focusing on smart mechatronics

**Abstract:** In recent years, the aging population has become a serious issue worldwide, and Japan, in particular, entered a "super-aged society" after 2010. On the other hand, the number of people requiring care, as of fiscal year 2022, is estimated to be about 6.9 million, and this number is rapidly increasing due to aging. In contrast, the required number of caregivers is estimated to be around 2.16 million in 2020 and approximately 2.45 million by 2025, indicating the need for about 60,000 new caregivers each year. However, there is a severe shortage of caregiving personnel, and the gap is expected to widen year by year. In response to this, assistive devices and power assist robots are being developed worldwide to help both the elderly and caregivers live better lives. However, the people who will actually use these robots or devices mostly want to solve their current physical problems. It's not enough to just have any robot; things like appearance, comfort, price, and most importantly, the benefits it will bring to their future life need to be considered. Only then will users want to use the robot. We believe that creating devices with a focus on the user's future needs and considering different perspectives is the kind of design that truly addresses real needs. To do this, we work closely with hospitals, nurses, and other experts from the start, promoting collaboration between engineering and healthcare.

# Invited Speech 7 (Online)

| | |
|---|---|
| **Time** | **10:30-10:50, March 6, 2025 (GMT+9)** |
| **Zoom ID** | **849 2479 0461** |
| **Zoom Link:** | **https://us02web.zoom.us/j/84924790461** |

## Prof. Maxim Bakaev

**Novosibirsk State Technical University, Russia**

Maxim Bakaev got his PhD degree in Software Engineering in 2012. He currently works as Associate Professor of the Automated Control Systems department of Novosibirsk State Technical University (NSTU), Russia. He is also the Acting Head of the Data Collection and Processing Systems department. Previously, he received his Master Degree in Digital Design from Kyungsung University, South Korea. His research interests include Human-Computer Interaction, Universal Design, Web User Interfaces, User Behavior Models, Knowledge Engineering, Machine Learning, etc. (https://www.researchgate.net/profile/Maxim-Bakaev) His recent research results are related to perception of visual complexity in graphical user interfaces (UIs) and its relation to Gestalt principles and compression algorithms. So, he has proposed the Index of Difficulty for tasks that involve visual-spatial working memory. He oversees the development of the Web UI Measurement Platform (http://va.wuikb.info/) that integrates online services for collecting ML data for UI assessment. He has served as a committee member for several international conferences, particularly as PC Co-Chair for ICMSC 2018 and ICWE 2019, as Demo & Posters Chair for ICWE 2020, and as Workshops Co-Chair for ICWE 2021. He also served as a reviewer for several international conferences and journals, including CHI, UIST, International Journal of Human-Computer Studies, Applied Ontology, Symmetry, etc. He is the Guest Editor for "Complexity in Human-Computer Interfaces: Information-Theoretic Approaches and Beyond", a Special Issue in Mathematics journal. He is also a Section Editor for the Journal of Web Engineering. He has acted as PI or participant in several research grants, domestic and international. In 2016, he received Novosibirsk City Hall award in science and innovations as a "Best

young researcher in higher education institutions". Under his supervision, more than 20 Master and Bachelor students graduated.

## Speech Contents

### Speech Title: Evoking Personas and Specialists from Large Language Models: going beyond the optimistic common-sense assistants

**Abstract:** Output of Large Language Models (LLM) based services tends to be common-sense and homogeneous, unless prompted otherwise. This trait has profound implications both for human users of the content created with the assistance of generative AI models in for the AI/ML field itself. As for the former, I expect the users to develop “generative blindness” (similar to the “banner blindness” at the onset of WWW) to filter out generic and low information density content. On the other hand, the default low diversity of synthetic texts hinders data augmentation for LLM, who are about to run out of the “natural” training data.

In my talk, I overview recent publications and techniques related to contextual prompting aimed at increasing the diversity and information density of LLM-produced texts. For instance, it has been demonstrated that prompting the models into being more seasoned specialists (e.g., senior software engineers) did increase the quality of the output with respect to solving the tasks in the professional domain. On the other hand, there are large-scale projects to create repositories of synthetic “authors” (e.g., with a whole a billion of them by Tencent AI), but relatively few evaluations of the effectiveness of different techniques in increasing the diversity of their “writings”. Correspondingly, I report the results of our own research in evoking “authors” with different associative thesauri from LLMs through contextual prompts based on Cooper’s personas descriptions, well-established in HCI. The conclusion is that so far the models fail to achieve the same level of diversity in the associations as the real humans.

# Invited Speech 8 (Online)

| | |
|---|---|
| **Time** | **10:50-11:10, March 6, 2025 (GMT+9)** |
| **Zoom ID** | **849 2479 0461** |
| **Zoom Link:** | **https://us02web.zoom.us/j/84924790461** |

## Asst. Prof. Xiangyu Yue

**The Chinese University of Hong Kong**

Xiangyu Yue is currently a Tenure-track Assistant Professor in the Department of Information Engineering at The Chinese University of Hong Kong. He received his PhD degree from EECS department at UC Berkeley. Before that, he received his BE and MS from Nanjing University and Stanford University, respectively. His research interest spans Multi-modal Learning, Transfer Learning, and Computer Vision, with 8000+ citations and H-Index of 27. He has served as reviewer for many conferences and journals, e.g. ICML, NeurIPS, CVPR, ICCV, TPAMI, IJCV, etc. He is serving as Area Chair for many top conferences, e.g. CVPR 2024 and NeurIPS 2024. His research has been recognized by the Lotfi A. Zadeh Award, for his outstanding contributions to soft computing and its applications.

## Speech Contents

### Speech Title: Towards unified Multimodal Learning

**Abstract:** In a complex and diverse world, information is presented in various data modalities (images, sounds, videos, text, etc.). By integrating information from multiple modalities, richer and more accurate semantic representations can be obtained, enabling more comprehensive and deeper learning and understanding. This presentation will introduce a series of works in multimodal learning, including designing a unified framework for understanding over ten modalities, using unpaired data for modality fusion, and endowing large language models with the ability to understand nearly ten data modalities.

# Session 1 (15:05-16:50)

**Topic: Image analysis and methods**
Session Chair: Prof. Shih-Lin Wu, Chang Gung University, Taiwan
Time: 15:05-16:50, March 5, 2025
Venue: Central School Building(中央校舍), Room 301

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

**MP0016**

Segmentation of Tumor Regions in BUSI Breast Ultrasound Images Based on DRA-UNet Model with CBAM

**Yi-Hsuan Shih**[1], Chih-Ying Wu[2, 3], Ruei-Chi Lin[2, 4] and Cheng-Ta Huang[1]
1: Yuan Ze University, Taiwan
2: Far Eastern Memorial Hospital, New Taipei, Taiwan
3: New Taipei City Association of Radiological Technologists, Taiwan
4: Taiwan Association of Medical Radiation Technologists, Taiwan

Abstract- Breast cancer remains a significant global health problem, highlighting the need for early screening and accurate diagnosis. Considering the complexity of ultrasound images, lack of clear boundaries, and variability in tumor shape and texture, overly simple models cannot meet the high precision requirements for lesion segmentation. In order to achieve the stability of automatic segmentation of breast lesions in ultrasound images, this paper proposes an attention-based U-Net framework (DRA_CBAM_UNet model), which integrates existing deep learning models. This includes using densely connected networks for feature reuse, adding residual connections to UNet, a residual feature extraction networks to reduce the difference between encoder and decoder feature maps, and combining dilated convolutions with convolutional block attention module (CBAM) attention mechanism to enhance the model's ability to capture local features. This paper uses the publicly available BUSI dataset to train and test the model, and experimental results show that our proposed method achieves higher segmentation accuracy compared to other reported methods.

**MP0011**

Restoration with Base Point Prediction for Deep Point Cloud Geometry Compression

**Hideaki Kimata**
Kogakuin University, Japan

Abstract- The shapes of real-world objects are being acquired as point clouds and are beginning to be used for a variety of services. In order to handle the huge amount of point cloud data, highly efficient compression methods have been studied. In recent years, compression coding of point cloud geometry using deep learning has been studied. It has been reported that it can be used to achieve higher compression efficiency than classical methods that do not use learning. However, in deep learning-based methods, a large reduction in the amount of code results in degradations that look like cracks and holes. In this paper, a novel restoration method from such visually noticeable degradation is proposed. Although many completion and up-sampling methods have been studied in the past, they are not suitable as a solution to this problem, because the decoder side does not know the location of the degradation and it does not know where and how to fill in. Therefore, in the proposed method, the base points for up-sampling are predicted according to the distortion map. The distortion map is generated at the encoder side and the parameters of the process for predicting and restoring points at the decoder side are estimated based on the distortion map and encoded into the stream. At the decoder side, the restoration process for degradation is carried out according to these parameters. Evaluation experiments show that points lost due to degradation can be restored and subjective quality can be improved.

**MP0013**

Fuzzy Bi-Histogram Equalization for Enhancement of Images

**Hafijur Rahman** and Tetsuya Shimamura
Saitama University, Japan

Abstract- The contrast enhancement (CE) of images has been promi-nently utilized in a variety of disciplines, including medi-cal and satellite imaging systems, due to the greater clar-ity of image features. Usually, existing CE methods shift the mean brightness of an input image and introduce arti-facts and brightening and darkening effects in the respective enhanced image. Thus, we propose the fuzzy bi-histogram equalization (FBHE) method to reduce these disadvantages and get a good CE result. First, the FBHE method com-putes a fuzzy image histogram (FIH). Next, the method computes a partitioning level (PL) from the FIH. It then partitions the FIH using the PL to preserve mean brightness and produces two fuzzy sub-histograms. Last, the method equalizes every sub-histogram to

produce an enhanced image. We implement the FBHE method with some conventional and cutting-edge CE methods in the CE of low-contrast images from a standard image dataset to compare their per-formances. The performances regarding visual assessment, ab-solute mean brightness error, peak signal-to-noise ratio, con-trast improvement, mean structural similarity, and entropy are compared. We find that the proposed FBHE method re-duces the aforementioned disadvantages and shows improved performance over the considered methods in this test. The MATLAB○R code of the FBHE method is publicly available at: https://doi.org/10.13140/RG.2.2.32157.83683/1.

**MP0038**

Spatial Distribution Analysis of Chemical Constituents in Artificially Induced Agarwood Using Combined UNet and GC-MS Approaches

Wei-Ning Huan[1], Wen-Ling Li[1], **Bo-Shuo Zhang**[2], Lam Tuyen Le[2], Huang Chi-Lung[1] and Wen-Pinn Fang[2]

1: Yuanpei University of Medical Technology, Taiwan

2: Yuan Ze University, Taiwan

Abstract- This study proposes an integrated approach to analyse agarwood formation by combining advanced imaging techniques, including U-net model image segmentation and 3D reconstruction, with gas chromatography-mass spectrometry (GC-MS). A U-Net model was employed for image segmentation to identify specific regions of agarwood within cross-sectional slices, followed by 3D reconstruction to visualize the spatial distribution of agarwood in a more intuitive manner. These 3D models allow researchers to assess the growth trends and structural patterns of incense wood in a more direct way. Additionally, gas chromatography-mass spectrometry (GC-MS) was used to analyze the chemical profiles of different regions within the agarwood, revealing significant variations in molecular composition. Through combined imaging and chemical analysis, this study provides a framework for understanding the formation process of artificially induced incense and offers suggestions for optimizing cultivation techniques and enhancing product quality for commercial and medicinal applications.

**MP0014**

Maturity classification of blueberry fruit from camera image for cultivation support system

**Ikuma Esaki**, Satoshi Noma, Takuya Ban, Rebeka Sultana and Ikuko Shimizu

Tokyo University of Agriculture and Technology, Japan

Abstract- In this paper, a method for maturity classification of blueberry fruit from camera images is proposed to realize the cultivation support system. According to the methods commonly used in the agricultural field, the maturity of blueberry fruit is classified into five levels proposed by Shutak. To classify the maturity level of blueberry fruit, we first detect the blueberry fruit region from the whole image, and then classify each region into five levels. Our method achieves high accuracy by extracting the local region of each fruit that includes the background and identifying its maturity level by the Transformer, which can learn the relationships not only between inside the fruit region but also outside the fruit. To train and validate the proposed method, the dataset of blueberry fruit image was created. We took images of blueberry fruits with the color chart corresponding to the maturity level of the fruit and labelled for the object detection labels and five maturity levels compared to the color chart. In experiments, we trained several deep learning models to compare with the proposed method. Three experimental patterns with changing combinations of training and test data in the dataset resulted in an average of 93.2% in the maturity classification task by the proposed method.

**MP0024**

Edge Enhanced Scene Understanding and Depth Estimation

**Naeem Ul Islam** and Syed H. Shah

Yuan Ze University, Taiwan

Abstract- An essential requirement for intelligent robots is the ability to perceive and comprehend the 3D structure of their surroundings without relying on expensive sensors. This perception capability is achievable with AI-based monocular depth estimation, in which a model uses an RGB image to predict the corresponding depth map. Researchers have developed various models that can produce accurate depth maps from a given RGB image. For instance, a prominent model is ZoeDepth, which combines relative and metric depth estimation to enhance depth prediction accuracy on various datasets. However, it still has challenges in accurately estimating fine details and has limitations in generalization across different environments with varying textures and lighting conditions. To overcome these limitations and improve the depth estimation accuracy, we have combined two essential image processing techniques: Canny Edge detection and high pass filtering. This improves edge detection and increases depth accuracy in scenes with complex object structures. Numerous experimental analyses have been carried out for the performance evaluation of the proposed approach for accurate depth estimation. The model was applied to indoor and outdoor datasets, significantly improving edge recognition and depth estimation, especially in zero-shot generalization tasks.

# Session 2 (15:05-16:50)

**Topic: Intelligent recognition technology and applications**
Session Chair: Prof. Yoosoo Oh, Daegu University, Korea
Time: 15:05-16:50, March 5, 2025
Venue: Central School Building(中央校舍), Room 302

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

**MP5024**

Image-level Synthesis and Perturbation for Self-supervised Anomaly Detection

**Ryo Kosugi** and Shin Ando
Tokyo University of Science, Japan

Abstract-Synthetic anomaly samples can play a crucial role in unsupervised image anomaly detection. Current image anomaly synthesis methods exploit image-level operations such as overlays or embedding-level operations such as perturbation by Gaussian noise. However, they have limitations in terms of the diversity and realism of the synthesized samples, respectively. In this paper, we integrate these techniques by guiding the impact of embeddinglevel perturbation towards specific regions of the synthesized images. We control the impact of the embedding-level perturbation on the normal regions of the synthetic anomalies using gradient descent, in order to maintain its consistency with the normality. Training feature extractor and localization networks with more diversity over the synthesized regions contributes to enhanced robustness. The empirical results show the advantage of our approach over the state-of-the-art method.

**MP5009**

Recognizing Challenging Behavior for Multiple Children with Intellectual and Developmental Disabilities Using the 3SLC Method

**Jonguk Jung** and Yoosoo Oh
Daegu University, Republic of Korea

Abstract-Challenging behaviors in children with IDD (Intellectual and Developmental Disabilities) can threaten their safety and the safety of others as well as the surrounding environment. Recording challenging behaviors in IDD children is currently done manually, which may lead to errors. We generate videos for each person in a multi-user setting and

extract joint coordinate values to create composite joint information. In this paper, we propose the 3SLC method that combines Self-Attention, Long-Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN). The proposed 3SLC method analyzes the relationships between the joint composite information, extracts temporal features from the feature map, and analyzes spatial characteristics to predict the behaviors of children with IDD.

**MP5016**

FFT-XMem: Enhancing the Boundary Identification of Video Object Segmentation by Fast Fourier Transform Based XMem Network

Tzu-Chiao Lo and **Shin-Jye Lee**

National Yang Ming Chiao Tung University, Taiwan

Abstract-As for video object segmentation, the recognition of various object and the corresponding position is essential to high-performance object segmentation of streaming video data. Meanwhile, semi-supervised video object segmentation aims to improve data preparation and annotation cost, and developing deep neural networks to process streaming images in spectral domain, which applies filter kernels in the spectral domain to recognize the boundaries of an image, is also an effective way to enhance the performance of video object segmentation. Thus, this paper purpose a simple yet effective concept for enhancing the boundary Identification among the segmentation objects, and the structure of the proposed method consists of Fast Fourier Transform and XMem for effectively recognizing the object boundary based on the features extracted from the spectral domain. The proposed module tries to provide all-around features of the object boundary, so this work applies the existing state-of-the-art network as a baseline for integrating the advantage of Fourier module. Through designing a high-resolution object boundary extraction module, the purposed method can decrease the loading of computation hardware. This work trained the model on one RTX 4090 with two datasets, and the experimental results shows a significant boost in performance compared to other state-of-the-art models. Further, this method also demonstrates the scalability in different models and evidently proves the capability of a compact model in enhancing video object segmentation by low-level pixel matching and object boundary recognition.

**MP5027-A**

Comparing Data Augmentation Methods for Hand Gesture Recognition Using Arm's Surface Electromyograpy Signals

**Kikuo Asai**

The Open University of Japan, Japan

Abstract- Data augmentation in machine learning has been widely used for reducing overfitting in training a classification model. The data augmentation is considered one of the solutions to the problem of requiring sufficient training data for high accuracy classification. We investigate on the performance of data augmentation on arm's surface electromyography (sEMG) signals for hand gesture recognition. In this investigation, severl classification models such as SVM, Random Forest, and CNNs are applied to hand gesture recognition based on sEMG signals. The time series data of sEMG signals are fed into the clssification models after data augmentation.

There are mainly four types of data augmentation methods to time series data: random transformation, pattern mixing, generative models, and decomposition. We select the methods with simple algorithms because of the computational cost. The performance of hand gesture recognition is evaluated by comparing the data augmentation of sEMG signals in the classification models. We discuss the classification performance of the hand gestures, clarifying the suitable methods of data augmentation of the sEMG signals for hand gesture recognition.

**MP0002**

ARNet: Enhanced Tracking and Optimization for Video-Based Affect Recognition

**Alice Othmani**[1], Mustaqeem Khan[2], Abdulmotaleb El Saddik[2, 3] and Hugo Cogo Moreira[4]

1: Université Paris-Est Créteil(UPEC), Créteil, France

2: Mohamed bin Zayed University of AI, Abu Dhabi, UAE

3: University of Ottawa, Canada

4: Østfold University College, Halden, Norway

Abstract- Affect recognition plays a vital role in understanding individuals' emotional well-being and social interactions, particularly for chil-dren with autism who often face challenges in expressing their emotions. In this paper, we propose a novel approach to enhance affect recognition in children with autism by combining the power of deep learning with a new optimization method called "Stochastic Average Gradient Augmented with Tracking" (SAGAT). Through extensive experimentation, we demonstrate that the proposed ap-proach significantly improves the accuracy of the Convolutional Neural Network (CNN) model compared to conventional optimiza-tion methods. Notably, our approach demonstrated strong accuracy in predicting arousal and valence in children with autism, with the CNN model achieving a mean squared error of 0.225 for arousal and 0.174 for valence on the SSBD-affect dataset. Pre-training on the AffectNet dataset further improved performance, reducing MSE to 0.187 for arousal and 0.156 for valence, highlighting the benefits of transfer learning. These findings hold significant implications for facilitating better understanding, support, and intervention strate-gies for children with

autism, ultimately fostering their emotional well-being and social integration. This research opens up promis-ing avenues for integrating advanced optimization techniques with deep learning to empower individuals with autism and promote inclusive technologies in affective computing.

**MP0037**

GlobalContextEnhancer: A Lightweight Framework for Fingerprint Denoising and Restoration on Resource-Constrained IoT Devices

Yu-Cheng Tsui, Ting-Yu Tseng and **Shang-Kuan Chen**
Yuan Ze University, Taiwan

Abstract- Fingerprint recognition provides unique and complex texture features for identifying individuals, making it widely applied in safeguarding personal privacy and security. However, the process of fingerprint acquisition is highly susceptible to various factors, such as skin conditions, uneven capture by devices, fingerprint overlap, and cluttered backgrounds. These issues often introduce significant noise and information loss in fingerprint images, which adversely affect the recognition efficiency and accuracy of subsequent systems. Therefore, developing effective denoising and restoration techniques for fingerprint images is critical to enhancing system performance.

To address these challenges, this study proposes a novel fingerprint image restoration framework based on an autoencoder model, integrating two innovative modules: the Multi-Scale Fusion Module (MSF) and the Global Feature Extraction Module (GFE). The MSF module utilizes a feature pyramid structure to extract and fuse multi-level features, effectively enhancing the model's ability to represent local details while ensuring feature integrity and consistency across different scales. The GFE module focuses on frequency domain analysis, leveraging the compact representation of low-frequency components to extract global structural features of fingerprint images, thereby avoiding detail loss while maintaining computational efficiency.

Experimental results demonstrate that the proposed framework achieves superior restoration performance on mainstream fingerprint datasets while maintaining low computational complexity. This provides a practical solution for fingerprint image processing in resource-constrained environments.

**MP0055**

Radar False Target Deception Jamming Suppression based on Linear Canonical Domain Line Segment Detection

**Jia-Mian Li** and Bing-Zhao Li
Beijing Institute of Technology, China

Abstract- In the field of radar signal processing, anti-jamming is an important research branch. Interrupted sampling repeater jamming (ISRJ) is an important type of active deception jamming, which poses a serious threat to radar detection and identification. In order to combat ISRJ, this paper proposes a jamming suppression method based on linear canonical domain by combining time-frequency (TF) analysis with line segment detection. Firstly, the distribution characteristics of the signal in the short-time linear canonical transform (STLCT) domain are derived. The flexibility of the three free parameters of linear canonical transform significantly improves the TF resolution, and a fast STLCT algorithm is proposed to construct the TF image of the radar echo. Then, the TF image is subjected to canny edge detection and dilation operations, and the position of the discontinuous line segment (representing jamming signals) is detected by Hough transform and filtered. Finally, the inverse STLCT is used to obtain the suppressed output result. Simulation experiments verify the effectiveness of the new method.

# Session 3 (15:05-16:50)

**Topic: Image detection models and algorithms**
Session Chair: Prof. Guo-shiang Lin, National Chin-Yi University of Technology, Taiwan
Time: 15:05-16:50, March 5, 2025
Venue: Central School Building(中央校舍), Room 303

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

**MP0004**

Improving Video Surveillance through ViViTRED: A Transformer-Based Detection Model for Rare and Abnormal Events

**Yahaya Idris Abubakar**, Mamadou Dia, Patrick Siarry, and Alice Othmani
Université Paris-Est Créteil (UPEC), Créteil, France

Abstract- Detecting rare and abnormal events in video surveillance presents a significant challenge due to limited data, environmental variability, and complex interactions within video sequences. Existing meth-ods often struggle to detect such events, especially in real-time applications. This paper introduces ViViTRED, a transformer-based model designed to address these challenges by leveraging advanced spatio-temporal feature extraction techniques. Building upon the Video Vision Transformer (ViViT), ViViTRED integrates tubelet embedding, positional encoding, and a stochastic local winner-takes-all (LWTA) layer to enhance the model's focus on critical features. Extensive experiments on benchmark datasets, including the UBI-Fights and Hockey Fights datasets, demonstrate ViViT-RED's superior performance, achieving accuracy rates of 97.48%and 94.00%, respectively, outperforming state-of-the-art methods. Our results highlight ViViTRED's potential for real-world video surveillance applications, with future work focused on incorporat-ing multi-modal data and optimizing for real-time deployment.

**MP0026**

Using Deep Learning to Detect Rehabilitation Exercise for Patients Diagnosed with Sarcopenia

**Po-Wei Huang**[1], Yung-Ching Weng[1], Chin-Hsuan Chia[2] and Tsorng-Lin Chia[1]
1: Ming Chuan University, Taoyuan, Taiwan
2: Department of Rehabilitation Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

Abstract- Sarcopenia is a common condition among older adults, characterized by a progressive decline in skeletal muscle mass, strength, and functional capacity. This deterioration severely impacts patients' ability to perform activities of daily living and reduces their overall quality of life. Traditional rehabilitation methods often require patients to frequently visit medical facilities and undergo training under professional supervision, posing logistical and practical challenges for many individuals. Consequently, developing an effective home-based rehabilitation monitoring system has become an urgent priority, particularly for homebound sarcopenia patients. This study aims to develop an innovative rehabilitation exercise monitoring system utilizing MMPose, a non-contact skeletal detection technology, to detect skeletal structures and assess the execution of prescribed rehabilitation exercises. The proposed system emphasizes convenience and exercise accuracy while providing a comprehensive analysis of movement patterns. Additionally, it employs time-series alignment to integrate real-time feedback and objective recommendations to optimize rehabilitation plans. By leveraging advanced technological methodologies, this study seeks to enhance rehabilitation outcomes, introduce novel assessment tools to the field of rehabilitation medicine, and alleviate the economic burden associated with recovery.

**MP0044**

Incremental Learning for Object Detection of Unmanned Ariel Vehicles

**Qazi Mazhar ul Haq**[1], Nandhagopal Chandrasekaran[1] and Muhammad Sohail[2]

1: Yuan Ze University, Taiwan

2: National University of Sciences and Technology, Islamabad, Pakistan

Abstract- Object detection is applied in many applications, such as unmanned aerial vehicles and autonomous vehicles, for multiple purposes due to its huge advancements. Object detection works as normal convolution neural network models with classification and localization for unmanned aerial vehicles. However, these traditional models of artificial intelligence suffer from catastrophic forgetting when trained on a new dataset, thereby compromising their ability to retain previously learned information. In this paper, we propose a novel integration of knowledge distillation with the famous YOLO object detection framework to overcome catastrophic forgetting. By using knowledge distillation, the model effectively transfers knowledge from previous tasks to new ones, preserving performance in earlier classes while retaining the previous information. The proposed framework is evaluated on Pascal VOC datasets in two classes to present the performance of incremental learning. Multiple experiments on these datasets suggest that our method has significantly improved the accuracy of previous classes in comparison to state-of-the-art methods.

**MP0041**

Applying deep learning to gait abnormality detection and assessment in home care

**Cai-Wei Lin**[1], Chin-Hsuan Chia[2] and Tsorng-Lin Chia[1]
1: Ming Chuan University, Taoyuan, Taiwan
2: Department of Rehabilitation Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

Abstract- With the development of an aging society, early identification of gait abnormalities is of great significance in preventing falls and improving quality of life. Abnormal gait is an early sign of muscle atrophy, neurodegeneration, and decreased balance ability. Existing gait analysis technology mostly relies on expensive equipment and requires professional operation, making it difficult to extend to home settings. This study proposes a gait abnormality detection method based on deep learning, which combines cameras and depth models to construct a video data set and design evaluation indicators. The experimental results prove that it has the potential for home application and real-time monitoring.
The Importance of Continuous Delivery Continuous Delivery is more than just a technical methodology; it's a cultural shift within organizations. The practice emphasizes the importance of automation, real-time feedback, and iterative development, enabling teams to deliver value to customers swiftly. With CD, organizations can release products more frequently, increasing their ability to respond to market demands and foster innovation.

**MP0064-A**

Automated radiology report generation from liver tumor computed tomography imaging

**Wei-Chi Hsu**[1], Chuan-Han Chen[2] and Che-Lun Hung[1]
1: National Yang Ming Chiao Tung University, Taipei, Taiwan
2: Taichung Veterans General Hospital, Taichung, Taiwan

Abstract- In the past decades, the number of new liver cancer cases has been increasing rapidly. Medical imaging technologies, such MRI and CT, have become very important tools to identify malignant tumors. The radiology reports, made by experienced radiologists, are useful for hepatologist to make diagnosis. The workload involved in creating radiology reports is heavy, causing patients to wait longer for their diagnosis results.
In this study, a method, which combines natural language processing and image processing techniques, is proposed to generate radiology reports from CT images of live tumors.
YOLO and BERT are used to extract relevant features from liver CT images and textual

reports, respectively. These extracted features are then input into the Text-to-Text Transfer Transformer model for fine-tuning, allowing the system to better understand and generate accurate radiology reports.

The experimental results demonstrate that the reports generated by the proposed method is able to reduce report generation time, closely align with the physician's writing style, and maintain greater consistency in formatting and terminology compared to the reports generated by radiologists.

**MP0008**

Improving Wildfire Detection Accuracy Using MobileNetV3-YOLOv8n

**Shiyan Du**, Jiacheng Li and Masato Noto
Kanagawa University, Yokohama, Japan

Abstract- Wildfires cause significant damage to ecosystems and human society, making timely fire detection crucial for minimizing losses. With the advancement of deep learning and Internet of Things (IoT) technologies, intelligent wildfire detection systems show promising application prospects. However, deploying high-precision detection models on resource-constrained IoT devices and drones still faces numerous challenges. To address this issue, this paper integrates the MobileNetV3 lightweight neural network with the YOLOv8 object detection model and proposes two improved models: MobileNetV3 Small-YOLOv8n and MobileNetV3 Large-YOLOv8n. Systematic evaluation on our custom wildfire image dataset demonstrates that the MobileNetV3 Large-YOLOv8n model achieves a flame detection Box(P) of 61.1%, a 4.2 percentage point improvement over the original YOLOv8n. To thoroughly investigate model performance, we designed a series of ablation experiments, comparing different YOLOv8n variants and exploring methods to enhance wildfire feature detection capabilities through the introduction of various feature pyramid structures. This research provides an efficient solution for wildfire detection systems in resource-constrained environments, achieving not only significant improvements in detection accuracy but also maintaining sufficient real-time performance, making important contributions to the advancement of intelligent disaster prevention, mitigation, and emergency response systems.

**MP0066-A**

Deep Learning for Ankle Fracture Identification on X-ray Images: Focus on Bimalleolar and Trimalleolar Fractures

**Hsuan-Yun Chin**[1], Shih-Chieh Tang[2], Shun-Ping Wang[2] and Che-Lun Hung[1]
1: National Yang Ming Chiao Tung University, Taiwan
2: Taichung Veterans General Hospital, Taichung, Taiwan

Abstract- Ankle fractures usually occur in the medial, lateral, and posterior regions of the ankle. Bimalleolar fractures affect two of these regions, with the medial and lateral malleoli being the most common combinations, often caused by sprains or external impacts. Trimalleolar fractures are an extension of bimalleolar fractures, with the addition of the posterior malleolus, usually caused by more significant external impact, resulting in higher instability. Currently, most machine learning methods primarily focus on lateral malleolus; however, a thorough diagnosis of ankle fractures should also encompass tibial injuries, especially in more complex cases. The posterior aspects of the tibia are often obscured and challenging to identify, and early detection is crucial for timely treatment and improved rehabilitation outcomes. A fracture classification model based on transfer learning is proposed to facilitate the diagnosis of complex ankle fractures using X-ray images. The proposed model achieves an accuracy of 0.78 ($\sigma$=0.03) and an F1 score of 0.71 ($\sigma$=0.05) when analyzing anteroposterior and mortise views, along with an accuracy of 0.75 ($\sigma$=0.05) and an F1 score of 0.63 ($\sigma$=0.06) when evaluating lateral views, validated through five-fold cross-validation. Overall, these findings demonstrate the potential of leveraging transfer learning techniques to improve the diagnosis of ankle fractures.

# Session 4 (16:50-18:35)

**Topic: Computer vision and image processing**
Session Chair: Prof. Suk-Ho Lee, Dongseo University, Korea
Time: 16:50-18:35, March 5, 2025
Venue: Central School Building(中央校舍), Room 301

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

**MP0005**

Introducing a Novel Protocol for Collecting Annotated Images to Automate Concrete Structure Damage Severity Assessment

**Camille Ruest**, Raef Cherif and Yacine Yaddaden
University of Quebec at Rimouski (UQAR), Rimouski, Quebec, Canada

Abstract- Concrete structures like dams, bridges, and buildings require regular inspections to ensure public safety. Currently, these inspections are carried out manually, which demands significant human and financial resources and poses risks to inspectors due to accessibility challenges. A more promising approach involves automating inspections using drones and algorithms to detect and estimate cracks. This paper introduces a protocol for capturing high-quality images of concrete structures using a tripod and a high-resolution camera. The collected images were used to develop a robust algorithm to estimate crack widths and classify them based on severity according to the Manuel d'inspection des structures du Ministère des Transports et de la Mobilité Durable du Gouvernement du Québec. The algorithm's accuracy was validated by comparing its measurements with actual crack values obtained directly from the inspected structures. Additionally, the new database, now accessible to the public, contains numerous measurements that could prove invaluable for future research endeavors, inspiring breakthroughs in structural health monitoring.

**MP0015**

Two-shot Spatially Varying Bidirectional Reflectance Distribution Function Estimation for Lustrous Material by Deep Learning

**Ayane Hokari**, Nozomu Terada, Rebeka Sultana and Ikuko Shimizu
Tokyo University of Agriculture and Technology, Japan

Abstract- To realistically represent the appearance of real-world materials in computer graphics, it is essential to estimate the spatially-varying bi-directional reflectance distribution function (SVBRDF) of each material. Using deep learning, several convenient methods for SVBRDF estimation have been proposed. However, challenges remain in handling materials with highlights and patterns due to the simplicity of the model and limited input data. We focus on lustrous fabrics, a commonly seen material that exhibits highlights and patterns. To capture visual cues such as highlights and shading, we use a flash-lit image, and to obtain the base color of the material, we use a non-flash-lit image. We propose a deep neural network that takes these two images as inputs and estimates 13 types of SVBRDF parameters based on the Disney principled BRDF model. In the network's loss function, we use a rendering loss to compare rendered images of the predicted and ground-truth maps under multiple lighting conditions from various viewing directions. Additionally, we introduce a sheen mix rendering loss as an auxiliary loss function. For the training dataset, we use approximately 250,000 images, including diverse materials created by artists. Our method demonstrates quantitatively and qualitatively superior results across a wide range of materials, including lustrous fabrics that are challenging to capture with existing SVBRDF estimation methods.

**MP0045**

Sampling Based on Joint Time-Vertex Linear Canonical Transform

**Yu Zhang**[1], Bing-Zhao Li[1] and Hong-Cai Xin[2]

1: Beijing Institute of Technology, China

2: Beijing Electronic Science and Technology Institute, China

Abstract- Joint time-vertex graph signal processing has seen significant ad-vancements recently. This paper investigates the sampling and recovery in the joint time-vertex linear canonical transform (JLCT) domain. We prove that bandlimited signals in the JLCT domain can be perfectly recovered. Experimental design strategies are employed to generate optimal sampling operators on time-vertex graphs. Furthermore, numerical experiments demonstrate superior perfor-mance compared to methods based on the joint time-vertex Fourier transform and fractional Fourier transform.

**MP0048**

Voice Chat Robot with Integrated Floating Projection

**Chia-Hsun Chiang**[1], Sen-Kai Hsu[1], Yi-Jheng Huang[1] and Yu-Hsuan Lin[2]

1: Yuan Ze University,Taiwan

2: Taiwan Instrument Research Institute National Applied Research Laboratories

Abstract- This paper presents a novel lip-sync animated voice chatbot system to enhance user engagement. Leveraging a Wii Balance Board for input triggering, the system utilizes Azure Bot Service for speech recognition and Language Studio for keyword extraction. Unmatched user queries are processed by ChatGPT, with responses converted to speech. A key contribution of this work is modifying the Wav2lip model to achieve continuous standby and real-time lip-sync animation synthesis upon audio detection, significantly accelerating performance by eliminating redundant face detection. The system features four distinct animation states (Idle, Listening, Thinking, and Answering), providing a seamless and responsive interactive experience.

**MP0065-A**

Deep learning approach for diagnosing heart failure using cardiac color doppler ultrasound

**Chia-Hsiu Wang**[1], Chung-Lieh Hung[2] and Che-Lun Hung[1]
1: National Yang Ming Chiao Tung University, Taiwan
2: MacKay Memorial Hospital, Taiwan

Abstract- Heart failure is a life-threatening condition with a five-year mortality rate of approximately 50%, according to medical statistics. Accurate and early diagnosis is crucial for improving treatment outcomes and reducing patient mortality. Doppler ultrasound, as a non-invasive imaging modality, provides essential information on the speed and direction of myocardial motion. However, to identify heart failure by using it usually relies heavily on clinicians' experience and expertise. To address these challenges, this study proposes an innovative approach that segments myocardial regions from Doppler ultrasound images, converts them into signals, and integrates deep learning techniques for auxiliary diagnosis of heart failure. The proposed method automatically detects heart failure abnormalities by extracting key features from myocardial motion signals. Experimental results demonstrate that the method achieves an accuracy of approximately 83% and an F1 score of 87%, showcasing robust performance across various metrics. This solution enhances diagnostic sensitivity and offers scalability and reliability, paving the way for the effective integration of artificial intelligence into cardiac ultrasound diagnostics.

**MP0069**

Neural Radiance Field Reconstruction in Complex Scenes Without Camera Parameter Dependency

**Minh Hoang Vuong Nguyen**, Viet-Hung Nguyen and Trung-Kien Luong
FPT University, Viet Nam

Abstract- Neural Radiance Fields (NeRF), introduced at ECCV 2020, enable realistic 3D reconstruction from calibrated images and have since inspired global research into Neural Volumetric Rendering (NVR). This presentation explores NVR's foundations, its ties to image-based rendering and inverse graphics, and evaluates three methods: NeRF, 3D Gaussian Splatting, and Gaussian Shader. NeRF offers accessible 3D training, while 3D Gaussian Splatting excels in rendering speed. Gaussian Shader effectively handles reflective surfaces but struggles with complex scenes, where NeRF and 3D Gaussian Splatting prove more robust.

**MP0043-A**

Deep Learning based Neonatal Spine Ultrasound Image Stitching and Tethered Cord Syndrome Detection

Yueh-Peng Chen, Pei-Chen Chung, **Shih-Lin Wu**
Chang Gung University, Taiwan

Abstract- Ultrasound screening for Tethered Cord Syndrome (TCS) in neonates requires segmental evaluation of the entire spinal column. However, this process limits the continuity of ultrasound images and increases examination time. Currently, clinical practice relies on manual stitching of segmented images and visual identification of vertebral positions, which may result in inconsistent positioning of standard images and compromise diagnostic accuracy. To address these issues, we developed an automated image stitching and spinal localization tool for neonatal spinal ultrasound screening. Using image segmentation techniques, we identify key anatomical structures, including vertebral bodies and the spinal cord in the lumbar and sacral regions. Through localization technology, we successfully achieved seamless stitching of lumbar and sacral spine images, expanding the diagnostic field for spinal structure evaluation. For spinal ultrasound image segmentation, we employed a modified U-Net architecture with an EfficientNetB7 encoder, integrating Dice Loss and Focal Loss as the loss functions to enhance segmentation performance effectively. To facilitate image stitching and fusion of the lumbar and sacral spine, we designed a similarity-based vertebral level prediction strategy, achieving a 94% accuracy rate in testing. Preliminary clinical validation demonstrated that our system consistently delivers accurate image stitching and reliable detection of TCS, providing significant technical support for the early diagnosis of neonatal spinal disorders.

# Session 5 (16:50-18:35)

**Topic: Computer-aided systems and interactive design**
Session Chair: Prof. Kikuo Asai, The Open University of Japan
Time: 16:50-18:35, March 5, 2025
Venue: Central School Building(中央校舍), Room 302

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

**MP5001**

Research on Cognitive Assistance System of Augmented Reality Technology in Complex Task Execution

**Zerui Guan**
University Of Washington, United States

Abstract-With the rapid development of intelligent manufacturing technology, human-computer interaction interface is undergoing revolutionary changes. Augmented reality (AR) technology, with its unique immersive experience and real-time interaction characteristics, has become one of the key technologies to improve the efficiency of complex tasks. This paper aims to explore how to use AR technology to build an efficient cognitive assistance system to support the execution of complex tasks in intelligent manufacturing environments. Firstly, this paper analyzes the challenges of human-machine interaction in the field of intelligent manufacturing, especially in tasks requiring high risk and high precision, the cognitive load of operators often becomes a bottleneck restricting efficiency. This paper proposes a framework of cognitive assistance system combining augmented reality technology and reinforcement learning. By capturing data from the operating environment in real time and combining reinforcement learning algorithms, the framework dynamically adjusts the display content of the AR interface to minimize the cognitive burden of the operator while maximizing the accuracy and safety of task execution. In terms of algorithm design, a reinforcement learning model based on Deep Q network (DQN) is adopted, which can learn and predict the optimal operation strategy in the complex task environment. Through simulation experiments, it is found that the proposed model can effectively identify and respond to the operator's behavior pattern, so as to provide personalized AR auxiliary information. In addition, in order to ensure the safety of the system, a multi-level security strategy is designed, including emergency stop mechanism, misoperation prevention and fault self-diagnosis function. The simulation results show that the proposed cognitive assistance system can

significantly improve the performance of operators in complex tasks, reduce the error rate and shorten the task completion time. Especially in the simulated high-pressure environment, the system shows good stability and adaptability. Further user studies have shown that operators have a high acceptance of the system and consider it to have significant advantages in terms of improving work efficiency and reducing work stress.

**MP5003-A**

Reverse Chatbot Approach for Synthesizing and Creating Research Plans

**V. Sithira Vadivel**

University of Newcastle, Australia

Abstract-The spread of false information is particularly concerning in today's digital age, where data can quickly circulate and reach a wide online audience. Researchers have a responsibility to carefully consider the potential impact of exposing participants to misinformation and take steps to minimise harm. The use of large language models, like ChatGPT, can inadvertently amplify misinformation due to potential biases or misinterpretations. When planning research involving synthesis and information ownership, thoughtful and methodical considerations must be made. This study aims to promote good academic practices and foster trustworthy judgments among student researchers, that enhances analytical thinking, prompt ideas and suggests keywords for research questions and hypotheses. A customized ChatGPT was developed to assist novice researchers in (1) planning their research; (2) synthesizing ideas; (3) taking ownership of their work by completing ten steps which include developing research questions; defining independent and dependent variables; determining actions for the identified variables; formulate hypotheses, type of experiment; data collection techniques; validation of the research methodology and outcome. The prototype was developed using the Chat Completions API from OpenAI (models: GPT-4o) and was evaluated by post-graduate students who provided positive feedback: supported learning, sparked creativity, and enhanced users' capabilities while reducing misinformation.

**MP5004-A**

Exploring Children’s Attitudes Towards Artificial Intelligence

**Dylan Yamada-Rice**[1], **John Potter**[2], Angus Main[3] and Eleanor Dare[4]

1: University of Plymouth, United Kingdom

2: University College London, United Kingdom

3: Royal College of Art, United Kingdom

4: University of Cambridge, United Kingdom

Abstract-This presentation presents the findings of having worked with approximately 250 children between the ages of 8 and 11-year-olds in the UK to understand their knowledge of and attitudes towards AI. It builds on the research team's previous work that taught children about digital sensors, the data they can collect about them and used speculative design to create tools to subvert/block them (Main & Yamada-Rice, 2022). The findings that will be presented here are part of a project that more broadly sought children's attitudes towards notions of digital good/bad and knowledge of how these may differ from adults. The project which was funded by the Economic and Social Research Council used hybrid arts practices to ask "What should a good digital society look like and how do we get there?", by focusing specifically on children as currently overlooked users of digital technology and emerging digital citizens. Using theories from Kress (2012; 1996), Barad, (2007), Nail (2020) and practices relating to ethical implications of digital, and emerging technologies, this presentation answers this question in relation to the project findings about AI and as such contributes to the MLHMI 2025 conference's desire to include ethnographic studies on human computer interaction.

**MP5010**

Interface Design for Small Vessels in Autonomous Navigation Systems

**Ryota Imai**, Atsushi Ishibashi, Takahiro Takemoto, Ayoung Yang and Tadasuke Furuya
Tokyo University of Marine Science and Technology, Japan

Abstract-As the development of MASS (Maritime Autonomous Surface Ships) progresses, AI(artificial intelligence)-based systems have become a cornerstone of autonomous navigation technologies. However, the complexity and unpredictability of the marine environment limit the reliability of AI-based decision-making. Consequently, even with the advancement of autonomous ship technologies, interfaces that enable continuous monitoring of AI systems and allow human intervention when necessary remain essential. This study focuses on the design and implementation of a tablet-based interface for small vessels, enabling real-time visualization of AI decision-making processes. The interface supports users in monitoring and intervening as required, thereby enhancing the safety and efficiency of autonomous navigation. By addressing these challenges, the proposed interface aims to ensure a reliable operational environment for autonomous ships, even under fully autonomous conditions.

**MP5011**

From Framework to Functionality: Exploring JavaScript and AI Integration in Automated Coloring

**Cheng-Yuan Ho**, Yu-Wen Gong, Kai-Syun Chen, Su Lee, Yu-Ting Gong, Wei-Han Chen and Wei-An Chen
National Taiwan University, Taiwan

Abstract-This article presents an automated coloring system that synergizes JavaScript and Artificial Intelligence (AI) technologies to enhance user experience, drive educational innovation, and improve creative efficiency. The implementation details include using the Canvas element and the Huemint color model to achieve AI coloring, covering steps such as loading layer data, sorting block sizes, monitoring mouse/touch inputs, tool selection, and coloring logic. Additionally, the study outlines appropriate color models for various scenarios and evaluates the advantages and limitations of different techniques. The results demonstrate how a resource-efficient approach can integrate JavaScript and AI technologies to deliver a flexible and efficient automated coloring solution.

**MP0019**

Class-Conditional Human Motion Generation using StyleGAN and Video Classifier

**Makoto Murakami**[1] and Kazuki Yamamoto[2]
1: Toyo University, Japan
2: SUS Co., Ltd.

Abstract- In movies and games utilizing 3D computer graphics, humanoid characters are often expected to appear and behave like real humans. This study aimed to develop a system for generating and controlling these Computer Graphics characters by specifying motion classes. We first employ the StyleGAN architecture to learn from motion data captured using a motion-capture system, enabling the generation of diverse and natural human movements. Next, the motion data randomly generated by this model are converted into video sequences through differentiable rendering. These sequences are then input into a motion-video classifier to predict the probability of each class. Finally, the latent variables of the motion-generation model are iteratively updated, guiding it toward generating motions corresponding to the desired class. Experimental results confirmed the effectiveness of the proposed method in generating appropriate motions across various motion classes.

**MP0060**

The Application of E-paper Bedside Cards Integrating Wired and Wireless Technologies: Dynamic Information Updates in Smart Wards

Yung-Hsiu Yang, Chen-Chi Liao and **Chun-Yuan Lin**
Asia University, Taiwan

Abstract- In medical environments, bedside cards in patient rooms are used to display patient information, physiological data, and precautions. However, traditional paper-based bedside cards face issues such as delayed updates and susceptibility to damage. To address these problems, this paper proposes using electronic paper (E-paper) as the bedside card. E-paper, based on electronic ink (E-ink) technology, offers low power consumption, excellent readability, and reflective display properties, making it an ideal solution. E-paper retains its displayed content without continuous power, making it suitable for hospital environments. This paper introduces an E-paper bedside card system that combines both wired and wireless technologies to address the need for dynamic information updates in hospital rooms. Wired data updates are performed through portable screens connected to a Raspberry Pi 4, ensuring the stability of information transmission. Additionally, wireless update technology enables healthcare personnel to remotely operate the system using laptops, smartphones, or computers at the nurse station, offering flexibility and avoiding the limitations of fixed locations, thus enhancing work efficiency. The system can display patient basic information, physiological data, precautions, and medical images (such as X-rays, ultrasound images, etc.). This combination of wired and wireless update methods not only improves the efficiency of updating information in the hospital but also reduces errors and delays associated with manual updates.

# Session 6 (16:50-18:35)

**Topic: Image encryption and security verification**
Session Chair: Prof. Ran-Zan Wang, Yuan Ze University, Taiwan
Time: 16:50-18:35, March 5, 2025
Venue: Central School Building(中央校舍), Room 303

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

**MP0023**

The Module-Guided Halftone QR Code

Kuo-Chien Chou, Chien-Hua Chou and **Ran-Zan Wang**
Yuan Ze University, Taiwan

Abstract- A halftone quick response (QR) code not only encodes the canonical text string but also provides a visual message to viewers. It inherits the black-and-white characteristic of a standard QR code that allows it to be printed economically. In the literature, halftone QR codes have usually been designed using numerous types of readability test, with a one-third centroid region of a module set to the original module color to ensure that the message in the QR code can be decoded correctly. This paper derives the best size of the centroid region of a module for determining the module values of a QR code, and accordingly proposes a fast and efficient method for generating halftone QR codes. Experiments were conducted to verify the validity of the proposed method, and the results show that the generated halftone QR codes possess appropriate decoding robustness and exhibit higher visual quality compared to previous methods.

**MP0036**

A Blockchain-Based Zero-Watermarking Model for Secure Digital Rights Management without Data Loss

**Jen-Chun Chang** and Ming-Yan Liao
National Taipei University, Taiwan

Abstract- Existing image copyright protection methods primarily rely on digital watermarking. Past research has focused on enhancing the imperceptibility and robustness of digital watermarks but has often overlooked the generation of critical information during watermarking algorithms, such as the keys used in encryption. These keys typically

rely on trusted third parties for secure storage, resulting in watermarks needing third-party verification. Moreover, the embedding process of traditional digital watermarking algorithms modifies the original image data, leading to data loss. To address these issues, this paper employs zero-watermarking technology to avoid direct modification of the original image, enhances the capacity of zero-watermarking, and ensures its robustness against both geometric and non-geometric attacks, thereby overcoming the limitations of traditional digital watermarking. We also propose a digital rights management model integrating blockchain with zero-watermarking. This ensures secure storage and verification of zero-watermarking without relying on third-party. The system effectively addresses the limitations of both zero-watermarking and traditional digital watermarking methods. Additionally, this paper explores the application of Non-Fungible Token (NFT) in transferring image usage rights.

**MP0034**

Block-based Difference Preserving Encryption and Median Preserving Pixel Value Ordering for Reversible Data Hiding in Encrypted Image

**Tsai-Wei Lin**[1], Chi-Yao Weng[2], Hao-Yu Weng[3], Chien-Lung Hsu[4], Shiuh-Jeng Wang[5] and Cheng-Ta Huang[1]

1: Yuan Ze University, Taiwan

2: National Chiayi University, Taiwan

3: National Central University, Taiwan

4: Chang Gung University, Taiwan

5: Central Police University, Taiwan

Abstract- Reversible data hiding in encrypted images (RDHEI) has become increasingly crucial in the era of cloud computing and digital privacy, where the need to securely embed data in encrypted media without compromising original content is paramount. However, traditional reversible data hiding methods often struggle with the challenges posed by encrypted domains, particularly in maintaining high embedding capacity while maintaining security and reversibility. To overcome these limitations, this paper introduces a novel RDHEI approach that integrates a three-stage encryption process with a median-preserving Pixel Value Ordering (PVO) strategy. The encryption phase employs block permutation, pixel scrambling, and Difference Preserving Encryption (DPE) to enhance image security while maintaining essential pixel correlation. The subsequent data embedding utilizes a median-preserving PVO method, which adapts to different block sizes to optimize embedding capacity. The experimental results show improvements in embedding capacity while maintaining high levels of security across various images. The effectiveness of the proposed method is further validated.

**MP0050**

The Scheme of Reversible Data Hiding with Difference-preserved for Encrypted Images

Hao-Yu Weng[1], Yu-Jin Chen[2], Min-Yi Tsai[3], Cheng-Hsing Yang[4] and **Shiuh-Jeng Wang[2]**

1: National Central University, Taiwan

2: Central Police University, Taiwan

3: National Central University, Taiwan

4: National Pingtung University, Taiwan

Abstract- Cloud technology has developed rapidly in recent years to protect data privacy and security in cloud applications. In this paper, we propose a new reversible data hiding in encrypted images (RDHEI) technique based on Difference Preserving Encryption (DPE), Modified Neighbor Mean Interpolation (MNMI), and Prediction Error Expansion (PEE). In the proposed method, the original image is downscaled, and MNMI is applied to generate an interpolated image. Then, both the original image and the interpolated image are encrypted using DPE. Next, prediction errors are calculated between the encrypted original image and the encrypted interpolated image. Afterward, secret data is embedded into the prediction errors using PEE. Finally, block permutation and pixel scrambling are applied to enhance security. An authorized receiver can use the reverse process to extract the secret message and decrypt the image without loss. Experimental results indicate that the proposed method achieves superior embedding payload and more effective encryption performance compared to existing schemes.

**MP0070-A**

Design and Implement of a Secure Health Promotion System based on Computer Vision, Clinical Guideline, and Serious Game Theory

Tzu-Liang Hsu[1], Chieh-Ni Chen[1], Wei-Cheng Lien[1] and **Chien-Lung Hsu[1,2,3]**

1: Chang Gung University, Taiwan

2: Ming-Chi University of Technology, Taiwan

3: Chang-Gung Memorial Hospital, Taiwan

Abstract- This paper introduces the design and implementation of a secure health promotion system based on computer vision, clinical guidelines, and serious game theory, which ensures effective and enjoyable health promotion through the following key features: (i) Clinical Guideline Compliance: The system is based on clinical guidelines, ensuring that all user actions are accurate and beneficial for health, promoting safe exercise and preventing harm. (ii) MediaPipe Computer Vision: By utilizing MediaPipe, the system allows users to engage in health-promoting activities without the need for wearable devices, using real-time computer vision to track movements. (iii) Serious Game

Theory: Incorporating game theory into the system, it offers 20 games designed to address four health topics: healthy physical fitness, osteoporosis prevention, stroke prevention, and reaction training. These games make health promotion entertaining and effective, increasing user engagement. (iv) Cryptographic Methods for Security: The system ensures the security of personal health data, game parameters, and digital rights through attribute-based encryption, safeguarding privacy and preventing unauthorized access. (v) Secure Key Management: A secure key management mechanism is implemented to maintain the integrity of the software, ensuring that the system and users' health data remain safe from potential security threats.

**MP0020**

Visual Two-Layer QR Code

**Chi-Han Lin**, Yu-Hung Su, Chien-Hua Chou and Ran-Zan Wang
Yuan Ze University, Taiwan

Abstract- QR codes have become a popular contactless interface for delivering information to users. A standard QR code encodes a string of text that can be read by scanning it with an optical device, typically the camera of a smartphone using a QR code app. The widespread adoption of QR codes has sparked significant research interest, particularly in enhancing their visual appeal and augmenting their data capacity. This paper presents a novel QR code scheme that integrates three layers of information into a single QR code. Each QR code displays an image, while two distinct public messages can be retrieved using standard QR code applications—one when scanned from a close range and another from a distance. The trade-off between the robustness of the decoding process for the two messages and the visual quality of the displayed image is examined, and experiments are conducted to demonstrate the feasibility of the proposed method.

**MP0071-A**

AI Model Development for Accurate Right and Left Hip Differentiation in Hip Ultrasound Exams

**Yung-An Ku**[1], Yueh-Peng Chen[1, 2], Hsuan-Kai Kao[1, 2]
1: Chang Gung University, Taiwan
2: Chang Gung Memorial Hospital at Linkou, Taiwan

Abstract- Developmental dysplasia of the hip (DDH) is a common congenital disorder that can lead to hip dislocation and requires surgery if left untreated. The preferred method of screening for DDH is ultrasound. When physicians perform the ultrasound and write a report, they sometimes perform ultrasound on the wrong side or write the results on the

wrong side in the report, resulting in an incorrect report or even complications. Therefore, our goal is to develop an AI model that can automatically determine the left and right sides of a baby's hip joint and help doctors reduce errors.

This study presents a deep learning model using ResNet-18 architecture, specifically developed to accurately distinguish between left and right hip joints. The class-specific recall of this AI model is 100% for the left hip and 99.5% for the right hip. The overall accuracy is 99.7%. These results highlight the accuracy and reliability of the model in identifying subtle anatomical differences, providing a significant step forward in reducing wrong-side errors.

The proposed AI model has significant clinical value by improving examination site verification, aiming to reduce preventable errors and improve patient safety standards. Future work will conduct clinical trials to assess the real-world applicability of the model.

# Poster Session

**Topic: Image detection and computational models**
Time: 14:50-15:05

Venue: Central School Building(中央校舍), Room 310

**MP0021**

Online Training of 3D Gaussians for Streamable Dynamic Scenes via Kalman-guided State Estimation

**Fuchen Yan**, Luyang Tang, Shihe Shen and Ronggang Wang
Guangdong Provincial Key Laboratory of Ultra High Definition Immersive Media Technology, Shenzhen Graduate School, Peking University, China

Abstract- Recent neural rendering techniques have achieved impressive re-sults in creating photo-realistic Free-Viewpoint Videos (FVVs) from multi-view inputs. However, these methods either require offline training or cannot achieve real-time performance. To tackle these limitations, we propose a real-time FVV rendering approach that in-troduces explicit latent variables to model frame-to-frame dynamics using structured Neural 3D Gaussians as the underlying representa-tion. We further present a Historical Attribute Refinement (HAR) mechanism, where the latent states are estimated based on two sources of knowledge: historical predictions through a lookback window and direct measurements from current observations. These two sources are adaptively fused through our Kalman filter-inspired state estimation approach for robust dynamic scene reconstruction. Our method was extensively evaluated on the N3DV and MeetRoom datasets. Compared to existing methods, our approach achieves state-of-the-art PSNR performance in online reconstruction while maintaining fast rendering speeds, low inference complexity and minimal transmission bandwidth requirements, even surpassing some offline reconstruction methods.

**MP0058-A**

Variable-Sized Block based Implicit Neural Representation for Super-Resolution

**Suk-Ho Lee**
Dongseo University, Korea

Abstract- Recently, Implicit Neural Representations (INRs) have been gaining attention as an approach where neural networks learn a continuous function that takes coordinates

as input and outputs the color values at those locations. Using INRs allows for reconstructing images of any size without constraints on spatial resolution. As a result, it has emerged as a promising method for super resolution, enabling a single neural network to represent images at all resolutions. However, existing research on INR-based super-resolution still lags behind other deep learning methods in terms of performance. This is because a single neural network, which takes uniform coordinate values as input, faces challenges in representing information of varying complexity across different regions of an image. Therefore, we propose a method to improve super-resolution performance by decomposing an image into variable-sized blocks so that each block has uniform complexity, regardless of the regional variation in complexity. The INR neural network then learns the image information of each block with uniform complexity. By alleviating differences in regional complexity, the neural network is able to learn regional information more stably and accurately, enabling optimal performance even in areas with diverse levels of complexity.

**MP4001**

A Body Part Detection Method Based on Vision-Language Model

Yu-Hong Zheng, Keng-Chun Chang, Fen-Yu Liao and **Guo-Shiang Lin**

National Chin-Yi University of Technology, Taiwan

Abstract- This paper proposes a body part detection method called MTransVG based on vision-language model. The proposed network developed based on TransVG comprises several components: an image encoder, a text encoder, an image-text feature fusion module, and two predictors. The image encoder and text encoder extract critical features from the image and text, respectively. The image-text feature fusion module integrates multi-model features to achieve useful feature representation. The two predictors are used to detect objects and indicate whether objects exist or not.

The proposed MTransVG model is trained by using a combination of web-collected and self-captured image datasets. Here we perform the subjective and objective evaluation for performance analysis. Experimental results demonstrate that the proposed MTransVG model can accurately locate body parts from both front and rear perspectives, showcasing its potential in smart medical applications, such as medical-assisted diagnostics. Compared with TransVG, the proposed MTransVG demonstrates significant performance improvements in the experiments.

# Session 7 (Online)

## March 6, 2025, Thursday (11:25-13:25, GMT+9)

**Topic: Intelligent image processing and application technology based on machine learning**

Session Chair: Prof. Maxim Bakaev, Novosibirsk State Technical University, Russia

Time: 11:25-13:25, March 6, 2025

Zoom ID: 849 2479 0461

Zoom Link: https://us02web.zoom.us/j/84924790461

***Presenters are recommended to enter the meeting room 10 mins in advance.**

***Presenters are recommended to stay for the whole session in case of any absence.**

***After the session, there will be a group photo for all presenters.**

**MP6001**

Prediction of offensive actions from Tankendo video images using relative frequencies of image features

Liyong Tao[1], Naoki Igo[2] and Kiyoshi Hoshino[1, 3]

1: University of Tsukuba, Tsukuba, Japan

2: Tokyo Information Design Professional University, Tokyo, Japan

3: Meiji University, Kawasaki, Japan

Abstract-The final goal of this study is to clarify, through system implementation, where and how an artificial system should look at human actions to be able to predict near-future results with high accuracy. The authors therefore collected video data from martial arts Tankendo practitioners and constructed an action prediction system for them. Its requirement specifications were to achieve 70% accuracy prediction within the first 0.25 seconds after movement onset, where a database was built using video-extracted features, specifically the bamboo sword tip height and area, and incorporating a calibration method to account for individual differences in middle guard stance and execution habits. Data from an expert practitioner was first used to compute a two-dimensional relative frequency distribution, and a Bayesian estimation and calibration approach was introduced, with 25 class interval and a cumulative probability threshold of 2.5. Calibration was performed by adjusting the database using initial sample and middle guard data from unknown subjects. As the results, the system was evaluated using data from the expert subject and two beginners. An accuracy of 75% was achieved with an output time of 0.3 seconds. Although calibration did not fully match expert performance in unknown subjects, but it consistently improved accuracy without delaying output.

**MP0047**

Exploring and Addressing Different Combinations and Variations of Language and Visual Models on Counting Problem in VQA

**Haiyang Zhao**

University Of Georgia, Athens, USA

Abstract- With the rapid development of artificial intelligence, multimodal learning has attracted great attention by integrating different types of data to enhance the development of artificial intelligence. Visual Question Answering (VQA) is one of the most prominent multi-modal tasks that requires the simultaneous fusion of image and text features to accurately answer questions related to the image. Despite recent progress in VQA, challenges still exist. In this study, our goal is to improve the VQA system by enhancing text process-ing and fusion components, and to improve the counting problem through a combination of language and visual models. Our main contribution includes introducing an Long-Short Term Memory (LSTM) model with enhanced attention mechanism, which signif-icantly improves performance. In addition, we have developed a low-rank version of the Multimodal Bilinear Fusion (MLB) module to reduce computational complexity. Furthermore, our research findings indicate that the standard Bidirectional Encoder Represen-tation from Transformers (BERT) model is not well suited for VQA tasks.

**MP0007**

AS400-DET: Detection using Deep Learning Model for IBM i (AS/400)

Thanh Tran[1, 2], Son T. Luu[1, 2], Quan Bui[1] and **Shoshin Nomura**[1]

1: Amifiable Inc., Meguro City, Japan

2: Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan

Abstract- This paper proposes a method for automatic GUI component detec-tion for the IBM i system (formerly and still more commonly known as AS/400). We introduce a human-annotated dataset consisting of 1,050 system screen images, in which 381 images are screenshots of IBM i system screens in Japanese. Each image contains multi-ple components, including text labels, text boxes, options, tables, instructions, keyboards, and command lines. We then develop a detection system based on state-of-the-art deep learning models and evaluate different approaches using our dataset. The experimental results demonstrate the effectiveness of our dataset in constructing a system for component detection from GUI screens. By automatically detecting GUI components from the screen, AS400-DET has the potential to perform automated testing on systems that operate via GUI screens.

**MP6002**

Prediction of attacking motions from video images of martial art Tankendo using support vector machine

Xinyue Zhang[1], Maki Nakamura[2], Naoki Igo[3] and Kiyoshi Hoshino[1, 4]
1: University of Tsukuba, Tsukuba, Japan
2: Iryo Sosei University, Chiba, Japan
3: Tokyo Information Design Professional University, Tokyo, Japan
4: Meiji University, Kawasaki, Japan

Abstract-This study aims to develop a Tankendo offensive actions prediction system that can achieve the same performance level as known offensive actions when an unknown attacker performs a strike. The data were collected from a highly skilled Tankendo practitioner, including 100 strikes to the head, 100 to the throat, and 100 to the torso. The dataset was constructed by extracting data from the one-second period before each strike was completed, with 240 strikes used for system development and 60 for parameter optimization. During video processing, the system tracked the attacker's collar, the tip of the bamboo sword, and the guard to compute features such as angle, distance, velocity, and acceleration. Principal Component Analysis was then applied to reduce the feature dimensions to nine. Additionally, to handle sections that were difficult to classify, an "other" category was defined, and its labeling position was set as a tunable parameter. The number of consecutive frames used before the final output was also set as an optimizable parameter and was evaluated through a full parameter combination analysis using 60 strike sequences. To ensure predictions were made before the mid-phase of the motion, the labeling position was ultimately set to 0.5 seconds before strike completion, with three consecutive frames used as the final decision criterion. When analyzing two unknown attackers, the system first measured their strike durations and selected the optimal parameters based on the known data. The evaluation included two attackers with longer strike durations than the known attacker. The experimental results indicate that for Subject A, the system's average output time was -0.4 ± 0.1575 seconds before strike completion, with correct outputs at -0.3715 ± 0.1328 seconds and an accuracy of 50 percent. For Subject B, the system's average output time was -0.4829 ± 0.1123 seconds before strike completion, with correct outputs at -0.4983 ± 0.1127 seconds and an accuracy of 66.67 percent. Furthermore, an analysis of the optimal parameters for Subjects A and B revealed that the optimal parameters for unknown attackers tend to have a later label assignment time and fewer consecutive frames compared to known attackers. This suggests a potential relationship between optimal parameter variations and strike duration; however, this relationship has not yet been clearly established.

**MP0057**

Depth, Thermal and RGB-Segmented Silhouette Imaging in Human Pose Estimation for Activity Monitoring

Yihao Zhang[1,2], **Haoran Ni[1,2]**, Beichen Sun[1,2], Zheng Yang Chin[2] and Kai Keng Ang[2,3]
1: Hwa Chong Institution, Singapore
2: Institute for Infocomm Research (I²R), Agency for Science, Technology and Research, Singapore
3: Nanyang Technological University, Singapore

Abstract- Human Pose Estimation (HPE) is a computer vision task that involves deep learning algorithms to estimate keypoints of different human joints in images. The keypoints form a skeleton model of the human body with applications in activity monitoring for healthcare, domotics and surveillance systems. However, HPE that uses RGB images deteriorates under inadequate lighting and raises privacy concerns because identifiable facial features are recorded. Thus, training a HPE deep learning model that predicts keypoints on lighting-invariant and privacy-preserving human silhouette images is proposed. Silhouette Imaging Dataset (SID) was collated from numerous public datasets, consisting of depth images, thermal images and simulated human silhouette images generated from RGB images using image segmentation, which were used to fine-tune the YOLO11s-pose HPE model. To classify predicted keypoints from the HPE model into different actions for activity monitoring, this study proposed a Residual Temporal Convolution Network with Stacked Bidirectional Gated Recurrent Unit (ResTCN-SBiGRU) Human Activity Recognition (HAR) deep learning model. The HAR model was trained on a public dataset to recognise 11 actions. The study shows the following results: the fine-tuned YOLO11s-pose yielded an mAP50-95(P) of 0.851; ResTCN-SBiGRU achieved an F1-score of 0.933. Finally, the study developed a real-time activity monitoring application using HPE, HAR and an additional proposed motion detection algorithm. The application gets live video input from a KinectV2 sensor and a laptop webcam, and displays pose, action and motion visualisations. This allows for testing of the different models and demonstrates the feasibility of activity monitoring using silhouette-based HPE.

**MP0010**

Temporal-aware Bidirectional Attention Network for Video-Text Retrieval

**Liujian Li**, Huaxiang Zhang, li Liu, Xinfeng Dong and Shengtang Guo
Shandong Normal University, China

Abstract- Cross-modal retrieval is an important research domain in multimodal data processing. With the proliferation of multimodal data, users' demands for retrieval

technologies have gradually increased, promoting the development of cross-modal retrieval technology. However, existing cross-modal retrieval models need several areas for improvement. For instance, 1) they inadequately utilize the interactive relationship between text features and video features, resulting in redundant information within the learned video and text features, which hinders the alignment of video and text features. 2) They inadequately establish temporal relationships among video frames. To address these weaknesses, we propose a video-text cross-modal retrieval model based on a Temporal-aware Bidirectional Attention Network (TBANet), which interacts with a similarity aggregation network to enable the retrieval model to learn high-quality text and video features. Firstly, we design a Bidirectional Cross-Attention (BCA) module to strengthen the interaction between videos and texts, which reduces the impact of redundant information on feature learning, and enhances the semantic alignment of text and video features. Secondly, we design a temporal-aware embedding (TAE) module to comprehensively establish temporal relationships among video frames. Finally, we validate the effectiveness and feasibility of our model on the MSRVTT, MSVD, and LSMDC datasets.

**MP6003**

An Trial of Predicting Effective Strikes in Martial Arts using Image Features based on Shifting Patterns of RGB Intensity

**Yinta Bu**[1] and Kiyoshi Hoshino[1, 2]

1: University of Tsukuba, Tsukuba, Japan

2: Meiji University, Kawasaki, Japan

Abstract-Our goal is to understand, through system implementation, where and how it should look at human actions to be able to predict near-future results with high accuracy. The authors therefore collected video data from martial arts practitioners and constructed an action prediction system for them. In this study, the authors proposed a new system for high-accuracy and early prediction of martial arts tankendo techniques. This system achieves motion prediction by dividing input images into 64×64 pixel blocks and analyzing the temporal change patterns of RGB values across three consecutive frames of image blocks. We enhanced prediction accuracy by encoding the position information of image blocks using trigonometric functions and fusing it with RGB features to preserve spatial information. Additionally, to accommodate subjects with different physical builds, we introduced an adaptive calibration method that maps the distribution patterns of active blocks to a standard model. This system is designed to utilize finely divided RGB pattern features and perform predictions through deep learning. However, experimental results revealed that the system failed to achieve this initial goal. The earliest prediction times

were 0.462 seconds before completion for head strike, 0.363 seconds for throat thrust, and 0.330 seconds for torso thrust. What the series of results showed was that the proposed system only discriminated between the three types of techniques in the appearance, where the final outcome was generally known, and could not be said to have predicted the final outcome of the technique in the appearance, where the final outcome was still not well known.

**MP5007**

Memory-Augmented Transformer for Relation Extraction of entities

Cristian E. M. Villalobos[1], Ricardo Tanscheit[1], Leonardo Forero[2] and **Marco Aurelio Pacheco[1]**

1: Pontifical Catholic University of Rio de Janeiro, Brasil

2: Rio de Janeiro State University, Brasil

Abstract-This work proposes a model for NLP tasks that extends a Universal Transformer (UT), a parallel-in time self-attentive recurrent sequence model, to take advantage of the dynamic external memory controller of a Differentiable Neural Controller (DNC). The UT has high intrinsic parallelism and uses dynamic halting to control the number of computational steps needed to process each input of the sequence. The proposed model improves the UT by adding a char-n-gram-based word representation in the input encoder and proposes a linear bottleneck residual block with batch normalization as a transition function, thus improving the feature learning inside the Transformer. Furthermore, this work modifies the original memory access controller of the DNC to process parallel-in time recurrent sequences and shows significant speedup over the original. Performance evaluation considers three benchmarks datasets for a relation extraction task. The paper also shows that the proposed model outperforms previous neural network approaches.

# Note

# Note

# Call for Papers

## December 20-22, 2025, Tokyo, Japan

**Tokyo, Japan, December 20-22, 2025** **Submission deadline: 2025-7-10**

**2025 7th Asia Digital Image Processing Conference (ADIP 2025)**

Conference Website: **https://adip.org/**

ADIP 2025 papers will be published into ***international conference proceedings, which will be submitted for indexing by Ei Compendex and Scopus.***

Submission methods: e-mail: **adip@acm-sg.org**

Zmeeting Submission System: **https://www.zmeeting.org/submission/adip2025**

Keynote Speaker: **Prof. Kenji Suzuki from Institute of Integrated Research at Institute of Science Tokyo, Japan.**

Conference will be held in **Tokyo, Japan**

## June 13-15, 2025, Tsukuba, Japan

**Tsukuba, Japan, June 13-15, 2025** **Submission deadline: 2025-3-20**

**2025 7th Asia Symposium on Image Processing (ASIP 2025)**

Conference Website: **https://www.asip.net/**

ASIP 2025 papers will be published into ***will be published in ASIP 2025 CPS Conference Proceedings, which will be submitted to IEEE Xplore, and be submitted for indexing to Ei Compendex, Scopus.***

Submission methods: e-mail: **asip@acm-sg.org**

Zmeeting Submission System: **https://www.zmeeting.org/submission/asip2025**

Conference will be held in **Tsukuba International Congress Center, Japan**

Keynote Speaker: **Prof. Changsheng Xu (IEEE & IAPR Fellow) from Chinese Academy of Sciences, China, Prof. Hiroyuki Kudo from University of Tsukuba, Japan, Prof. Laurent David Cohen (IEEE Fellow) from Universite Paris Dauphine, France and Prof. Keisuke Kameyama from University of Tsukuba, Japan**

# Call for Papers

## June 25-27, 2025, Tsukuba, Japan

**Tsukuba, Japan, June 25-27, 2025** **Submission deadline: 2025-3-25**

**2025 7th Blockchain and Internet of Things Conference (BIOTC 2025)**

Conference Website: **https://biotc.net/**

BIOTC 2025 papers will be published into ***international conference proceedings, which will be submitted for indexing by Ei Compendex and Scopus.***

Submission methods: e-mail: **biotc.contact@gmail.com**

Zmeeting Submission System: **https://www.zmeeting.org/submission/biotc2025**

Keynote Speaker: **Prof. Qun Jin from Waseda University, Japan and Prof. Chin-Chen Chang (IEEE Fellow/ IEE Fellow) from Feng Chia University, Taiwan**

Conference will be held in **University of Tsukuba, Japan**

## September 24-26, 2025, Okinawa, Japan

**Okinawa, Japan, September 24-26, 2025** **Submission deadline: 2025-4-20**

**2025 7th International Conference on Big Data Engineering (BDE 2025)**

Conference Website: **https://www.bde.net/**

BDE 2025 papers will be published into ***will be published in International Conference Proceedings, which will be submitted for indexing to Ei Compendex, Scopus.***

Submission methods: e-mail: **bde.conference@gmail.com**

Zmeeting Submission System: **https://www.zmeeting.org/submission/BDE2025**

Conference will be held in **University of the Ryukyus, Okinawa, Japan**

Keynote Speaker: **Prof. Zheng Yan (IEEE Fellow, IET Fellow, AAIA Fellow, and AIIA Fellow) from Xidian University, China, Prof. Xiaoli Li (IEEE Fellow and AAIA Fellow) from Institute for Infocomm Research, A*STAR, Singapore, Prof. Irwin King (IEEE Fellow, INNS Fellow and AAIA Fellow) from The Chinese University of Hong Kong, China and Prof. Tomoaki Otsuki (IEICE fellow, AAIA fellow) from Keio University, Japan**