CONFERENCE PROGRAM

January 10-12, 2025
Hong Kong, China

IPMV 2025

2025 7TH INTERNATIONAL CONFERENCE ON IMAGE PROCESSING AND MACHINE VISION

2025 6TH INTERNATIONAL CONFERENCE ON ELECTRONICS AND SIGNAL PROCESSING ICESP

CONFERENCE PROGRAM (UTC+8)

2025 7th International Conference on Image Processing and Machine Vision

(IPMV 2025)

2025 6th International Conference on Electronics and Signal Processing

(ICESP 2025)

Hong Kong, China | January 10-12, 2025

Co-sponsored by

香港大學計算機科學系 THE UNIVERSTIY OF HONGKONG DEPAPTMENT OF COMPUER SCIENCE

澳門大學 UNIVERSIDADE DE MACAU UNIVERSITY OF MACAU

Supported by

武漢理工大學 WUHAN UNIVERSITY OF TECHNOLOGY

대구대학교 DAEGU UNIVERSITY

MAHANAKORN UNIVERSITY OF TECHNOLOGY

Universidad Politécnica de Cartagena

EUT+ EUROPEAN UNIVERSITY OF TECHNOLOGY

CTIM CENTRO DE TECNOLOGÍAS DE LA IMAGEN

TABLE OF CONTENT

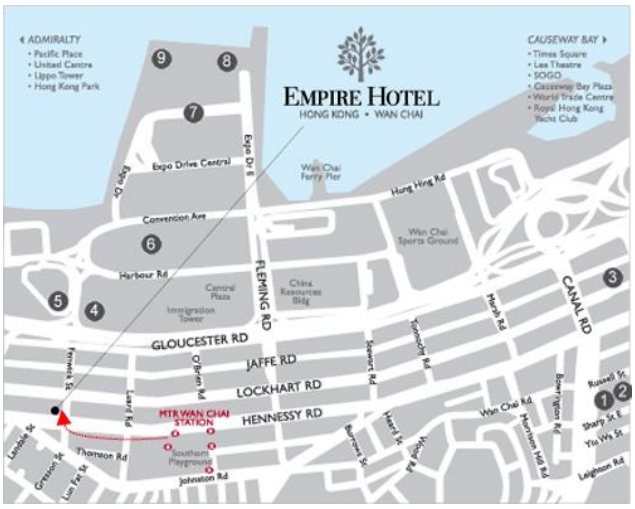
GENERAL INFORMATION

A Conference Venue

香港灣仔皇悦酒店 (Empire Hotel Hong Kong—Wan Chai)

Add: No.33 Hennessy Road, Wan Chai District, Hong Kong, China

Phone Number / Fax Number	+852-36922111 / +852-28613121
Check-In Time	14:00
Check-Out Time	12:00
Parking	PARKING AVAILABLE (15 CARS) PARKING FEE 60HKD/2HOUR NEED BOOKING
Credit Cards	American Express Diner's Club JCB Master Card VISA



B Onsite Registration

Arrive at Registration desk at **Hotel Lobby**→ Inform the staff of your paper ID→ Sign-in→ Claim your conference kit.

C Devices Provided by the Organizer

Offline Oral Session: Laptops (with MS-Office & Adobe Reader) / Projectors & Screen / Laser Sticks

D Materials Provided by the Presenter

Onsite/Online Oral Session: Slides (pptx or pdf version). Format 16:9 is preferred.

E Duration of Each Presentation

Keynote Speech: 45min, including Q&A.

Oral Session: 15min, including Q&A.

Invited Talk: 25min, including Q&A.

F Notice

• Please wear your delegate badge (name tag) for all the conference activities. Lending your participant card to others is not allowed.

• Please take good care of your valuables at any time during the conference. The conference organizer does not assume any responsibility for the loss of personal belongings of the participants during conference day.

• Accommodation is not provided. Participants are suggested make early reservation.

• Please show the badge and meal coupons when dining.

G Online Conference

Zoom Download	ROOM MEETING ID	LINK
	813 0177 8800	https://us02web.zoom.us/j/81301778800

Note:

1. We recommend that you install the Zoom platform on your computer. New Zoom users can skip the registration step and enter the meeting ID directly to participate the online session.
2. Prior to the formal conference, presenter shall join the test room to make sure everything is on the right track.
3. Please rename your Zoom Screen Name in below format before entering meeting room.

Name Setting:

Keynote Speaker: KN-Name

Author: Paper ID-Name

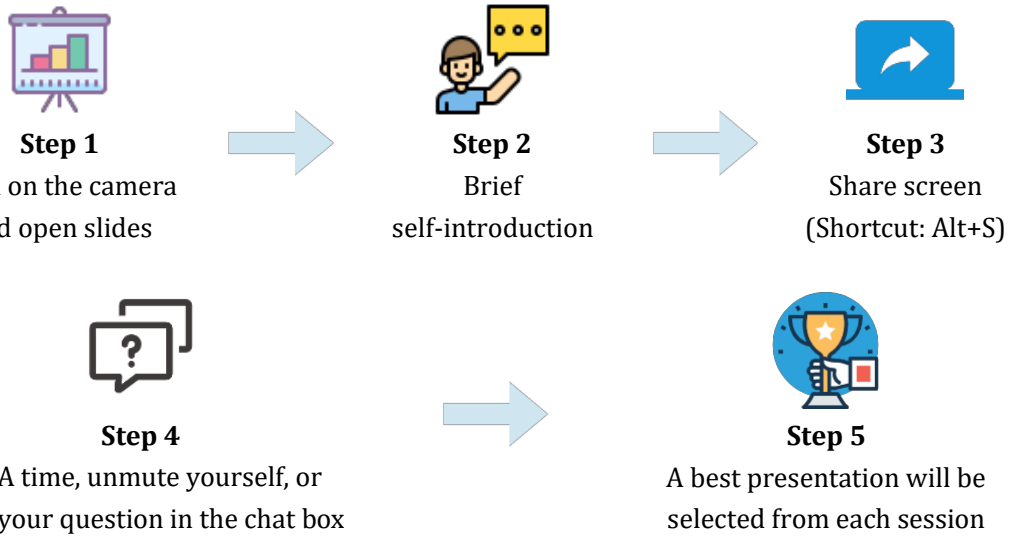
Committee: Position-Name

Delegate: Delegate-Name

Useful Links:

✔ Conference Banner ✔ Zoom Background

Presentation Process by Zoom Meeting



About Online Presentation

• Every presenter has 15 minutes, including Q & A.

• The best presentation certificate and all authors' presentation certificates will be sent after conference by email.

• We'll record the whole conference. If you do mind, please inform us in advance. We'll stop to record during your presentation time.

WELCOME MESSAGE

Dear All,

We are pleased to welcome you to the joint conference of 2025 7th International Conference on Image Processing and Machine Vision (IPMV 2025) and 2025 6th International Conference on Electronics and Signal Processing (ICESP 2025), to be held in Hong Kong, China, January 10-12, 2025. This conference is co-sponsored by Department of Computer Science, University of Hong Kong and University of Macau.

IPMV is an international conference that serves researchers, scholars, professionals, students, and academicians who are looking to both foster working relationships and gain access to the latest research results. It will put special emphasis on the participations of PhD students, Postdoctoral fellows and other young researchers from all over the world. It would be beneficial to bring together a group of experts from diverse fields to discuss recent progress and to share ideas on open questions.

This year's program will consist of 3 keynote speeches, delivered by Prof. James Tin-Yau Kwok (Fellow of IEEE, The Hong Kong University of Science and Technology, Hong Kong), Prof. Junsong Yuan (Fellow of IEEE, State University of New York at Buffalo, USA), and Prof. Chi Man Pun (University of Macau, Macau). 1 invited talk is given by Dr. Muhammad Asif Khan (Qatar Mobility Innovations Center, Qatar University, Qatar), followed by 1 onsite and 1 online oral session.

Our deepest gratitude goes to the Advisory Board, Organizing Committee, International Scientific Committee, institutions, and volunteer who have directly and indirectly supported the success of this seminar. Wish you a very productive conference with exciting and encouraging discussions and exchange of knowledge so that together we can anticipate a future of ground-breaking knowledge, research, and technology.

Finally, we wish you a very successful conference! Hope you will enjoy your stay in Hong Kong.

Conference Organizing Committees,

Hong Kong

IPMV 2025

Ms. Ching Cao

ipmv_conf@yeah.net

ICESP 2025

Ms. Fiona

icesp_conf@163.com

CONFERENCE COMMITTEE

Conference Advisory Committees
Junsong Yuan, State University of New York at Buffalo, USA
Zhongfei (Mark) Zhang, Binghamton University State University of New York, USA

Conference General Chair
Kenneth K.Y. Wong, University of Hong Kong, Hong Kong, China

Conference General Co-Chairs
James Tin-Yau Kwok, The Hong Kong University of Science and Technology, Hong Kong, China
Chi Man Pu, University of Macau, Macau, China

Conference Program Chairs
Ray Chak-Chung Cheung, City University of Hong Kong, Hong Kong, China
Hui Zhang, BNU-HKBU United International College, China
Yew Kee WONG Eric, Hong Kong Chu Hai College, China
Hong Lin, University of Houston-Downtown, USA

Conference Program Co-Chairs
Guanying Chen, Sun Yat-sen University (Shenzhen), China
Seokwon Yeom, Daegu University Gyeongsan, South Korea
Zhen Gao, McMaster University, Canada

Conference Publicity Chair
Kai Han, The University of Hong Kong, Hong Kong, China

Technical Program Committees
Fakhreddine Ababsa, PIMM - ENSAM, France
Chong Wang, Ningbo University, China
Wannida Sae-Tang, King Mongkut's University of Technology North Bangkok, Thailand
Wanwan Li, University of South Florida, USA
Tieyong Zeng, The Chinese University of Hong Kong, Hong Kong, China
Raymond Pang, LSCM R&D center in Hong Kong, Hong Kong, China
Bin Xue, National University of Defense Technology, China
Nguyen Ngoc Phien, Ton Duc Thang University, Vietnam
Pavlo Maruschak, Ternopil Ivan Puluj National Technical University, Ukraine
Luis Gómez, Universidad de Las Palmas de Gran Canaria, Spain
Amir Nakib, Université Paris Est Créteil, France
Serdar Ciftci, Harran University, Turkey
Maleerat Sodanil, King Mongkut’s University of Technology North Bangkok, Thailand
Suphongsa Khetkeeree, Mahanakorn University of Technology, Thailand
Thaweesak Yingthawornsuk, King Mongkut's University of Technology Thonburi, Thailand
Puntani Pongsumpun, King Mongkut’s Institute of Technology Ladkrabang, Bangkok, Thailand
Sheli Sinha Chaudhuri, Jadavpur University, India
Haruna Abdu, Federal Unversity Lokoja, Nigeria
Ashwani Kumar Aggarwal, Sant Longowal Institute of Engineering and Technology, India

AGENDA OVERVIEW

Session Time	Friday, January 10, 2025 (UTC+8)	Venue
14:00-17:00	On-site Registration & Materials Collection	1F Lobby at Empire Hotel Hong Kong—Wan Chai
14:00-15:00	Zoom Rehearsal for All Online Participants	Zoom Meeting ID: 813 0177 8800

Session Time	Saturday, January 11, 2025 (UTC+8) | Plenary Meeting	
Empire Boardroom (Room 3), 1F | Zoom Meeting ID: 813 0177 8800		
08:30-09:00	On-site Registration & Materials Collection	
09:00-09:10	**Opening Remarks**	Conference General Chair: **Prof. Kenneth K.Y. Wong**, University of Hong Kong, Hong Kong, China
09:10-09:55	**Keynote Speech**	**Prof. James Tin-Yau Kwok**, Fellow of IEEE, The Hong Kong University of Science and Technology, Hong Kong, China *Speech Title: Large Language Models: Pre-Training, Fine-Tuning and Application*
09:55-10:40	**Keynote Speech** (Online)	**Prof. Junsong Yuan**, Fellow of IEEE, State University of New York at Buffalo, USA *Speech Title: Intelligent Hand Sensing and Augmented Interaction*
10:40-11:10	Group Photo & Coffee Break	
11:10-11:55	**Keynote Speech**	**Prof. Chi Man Pun**, University of Macau, Macau, China *Speech Title: Image Manipulation Localization with Deep Neural Networks*
12:00-13:30	Set Lunch / COZY Restaurant 1F	

Session Time	Saturday, January 11, 2025 (UTC+8) | Parallel Sessions	
13:30-15:15	**Onsite Session 1A: Vision-based Image Processing and Computational Models** Presentations: P007, P006, P016, P012, P405-A, P418, P014	Empire Boardroom (Room 3), 1F
15:15-15:30	Coffee Break	
15:30-17:15	**Onsite Session 1B: Vision-based Image Processing and Computational Models** Presentations: P015, P034, P401, P413, P411-A, P032, P029-A	Empire Boardroom (Room 3), 1F
17:30-19:00	Dinner Time / Tiffin Restaurant LG2 Floor	

Fully Online

13:30-15:55	**Online Session 1A: Digital Image Detection and Processing Methods** **Invited Talk: Dr. Muhammad Asif Khan** Presentations: P022, P002, P005, P011, P414, P018, P009, P1002	Zoom: 813 0177 8800
16:00-18:00	**Online Session 1B: Digital Image Detection and Processing Methods** Presentations: P1004, P416, P008, P417, P027, P030, P035, P412-A	

KEYNOTE SPEAKER

Prof. James Tin-Yau Kwok

The Hong Kong University of Science and Technology, Hong Kong, China
Fellow of IEEE

Speech Title: Large Language Models: Pre-Training, Fine-Tuning and Application

Abstract: Large language models (LLMs) are now widely used. However, several challenges remain in pre-training and fine-tuning. First, during unsupervised pre-training, semantically irrelevant information can negatively impact downstream tasks, leading to negative transfer. Second, multiple models with various hyperparameter configurations are often created during fine-tuning, but typically only one of these models is utilized in the downstream task. To address the first issue, we introduce a new pre-training method that trains each expert using only semantically relevant data through cluster-conditional gates. This approach allows for the allocation of downstream tasks to customized models pre-trained on data most similar to the downstream data. To tackle the second issue, we consider the learned soup, which combines all fine-tuned models with learned weighting coefficients. While this can significantly enhance performance, it is also computationally expensive. We propose to mitigate this issue by formulating the learned soup as a hyperplane optimization problem and employing block coordinate gradient descent to learn the mixing coefficients. At each iteration, this approach only requires loading a few fine-tuned models and building a computational graph with one combined model. Experimental results show that this can then be run on a single GPU while significantly reducing memory usage. Finally, we present an application of LLMs for knowledge graph completion. We utilize an in-context learning strategy to guide the LLM. Empirical results demonstrate its effectiveness, achieving improved performance with no additional training required.

Biography: James Kwok is a Professor in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. He is an IEEE Fellow. He has served / is serving as an Associate Editor for the IEEE Transactions on Neural Networks and Learning Systems, Neural Networks, Neurocomputing, Artificial Intelligence Journal, International Journal of Data Science and Analytics, and on the Editorial Board of Machine Learning. He is also serving as Senior Area Chairs of major machine learning / AI conferences including NeurIPS, ICML, ICLR, IJCAI. He is on the IJCAI Board of Trustees. He is recognized as the Most Influential Scholar Award Honorable Mention for "outstanding and vibrant contributions to the field of AAAI/IJCAI between 2009 and 2019". Prof Kwok is the IJCAI-2025 Program Chair.

KEYNOTE SPEAKER

Prof. Junsong Yuan

State University of New York at Buffalo, USA
Fellow of IEEE

Speech Title: Intelligent Hand Sensing and Augmented Interaction

Abstract: Humans are the most intelligent beings on the planet not only because of our powerful brain but also due to the unique structure of our hands. Hands have been crucial tools for us to interact and change both the physical world and the virtual world such as metaverse. In this talk, we will discuss real-time hand sensing using optical cameras, and how it can enhance our interactions with physical world and metaverse. Towards 3D hand sensing from single 2D images, we will discuss how to leverage synthetic hand data to address high-dimensional regression problem of articulated hand pose estimation and 3D hand shape reconstruction. To improve the generalization ability of handling hands of various shapes and poses, we will also discuss invariant hand representation through disentanglement. The resulting systems can facilitate intelligent interactions in virtual and real environments using bare hands, as well as via hand object interactions.

Biography: Dr. Junsong Yuan is Professor and Director of Visual Computing Lab at Department of Computer Science and Engineering, State University of New York (SUNY) at Buffalo, USA. Before joining SUNY Buffalo, he was Associate Professor at Nanyang Technological University (NTU), Singapore. He obtained his Ph.D. from Northwestern University, M.Eng. from National University of Singapore, and B.Eng. from Huazhong University of Science Technology. He is a recipient of SONY Faculty Innovation Award (2024), SUNY Chancellor's Award for Excellence in Scholarship and Creative (2022), IEEE Trans. on Multimedia Best Paper (2016), Northwestern Outstanding EECS Ph.D. Thesis (2010), and Nanyang Assistant Professorship (2009). He serves as Editor-in-Chief of Journal of Visual Communication and Image Representation (JVCI), Associate Editor of IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI) and IEEE Trans. on Image Processing (T-IP). He also serves as General/Program Co-chair of ICME and Area Chair for CVPR, ICCV, ECCV, NeurIPS, ACM MM, etc. He is a Fellow of IEEE (2021) and IAPR (2018).

KEYNOTE SPEAKER

Prof. Chi Man Pun

University of Macau, Macau, China
Senior Member of IEEE

Speech Title: Image Manipulation Localization with Deep Neural Networks

Abstract: Creating fake pictures has become more accessible than ever, but tampered images are more harmful because the Internet propagates misleading information so rapidly. Reliable digital forensic tools are, therefore, strongly needed. Traditional methods based on hand-crafted features are only useful when tampered images meet specific requirements, and the low detection accuracy prevents them from being used in realistic scenes. Recently proposed learning-based methods have improved accuracy, but neural networks usually require training on large labeled databases. This is because commonly used deep and narrow neural networks extract high-level visual features and neglect low-level features where there are abundant forensic cues. In this talk, we will discuss some solutions to this problem. Two novel image splicing localization methods are proposed using deep neural networks, which mainly concentrate on learning low-level forensic features and consequently can detect splicing forgery, although the network is trained on a small automatically generated splicing dataset.

Biography: Prof. Pun received his Ph.D. degree in Computer Science and Engineering from the Chinese University of Hong Kong in 2002, and his M.Sc. and B.Sc. degrees from the University of Macau. He had served as the Head of the Department of Computer and Information Science, University of Macau from 2014 to 2019, where he is currently a Professor and in charge of the Image Processing and Pattern Recognition Laboratory. He has investigated many externally funded research Projects as PI, and has authored/co-authored more than 200 refereed papers in many top-tier Journals (including T-PAMI, T-IFS, T-IP, T-DSC, T-KDE, and T-MM) and Conferences (including CVPR, ICCV, ECCV, AAAI, ICDE, IJCAI, MM, and VR). He has also co-invented several China/US Patents, and is the recipient of the Macao Science and Technology Award 2014 and the Best Paper Award in the 6th Chinese Conference on Pattern Recognition and Computer Vision (PRCV2023). Dr. Pun has served as the General Chair for the 10th &11th International Conference Computer Graphics, Imaging and Visualization (CGIV2013, CGIV2014), the 13th IEEE International Conference on e-Business Engineering (ICEBE2016), and the General Co-Chair for the IEEE International Conference on Visual Communications and Image Processing (VCIP2020) and the International Workshop on Advanced Image Technology (IWAIT2022), and the Program/Local Chair for several other international conferences. He has also served as the SPC/PC member for many top CS conferences such as AAAI, CVPR, ICCV, ECCV, MM, etc. He is currently serving as the editorial board member for the journal of Artificial Intelligence (AIJ). Besides, he has been listed in the World's Top 2% Scientists by Stanford University since 2020. His research interests include Image Processing and Pattern Recognition; Multimedia Information Security, Forensic and Privacy; Adversarial Machine Learning and AI Security, etc. He is also a senior member of the IEEE.

INVITED SPEAKER

Dr. Muhammad Asif Khan

Qatar Mobility Innovations Center
Qatar University
Qatar

Speech Title: Real-time Crowd Counting at the Edge

Abstract: Recent research on crowd counting shows the efficacy of deep learning methods such as CNNs due to their strong capability of auto-feature extraction. To achieve higher accuracy in dense scenes, deeper models with a large number of parameters are developed. However, although achieving good accuracy, deeper models create performance bottlenecks in real-time applications due to large memory requirements, higher training complexity, and large inference delay. On the contrary, shallow models are lightweight, incur low inference delay, and require low memory, but are usually disregarded by some due to their limited accuracy in dense crowd scenes. This talk will discuss how to design efficient and more robust lightweight models for crowd-counting applications at the edge providing useful insights for prospective researchers and DL practitioners. This talk provides a retrospective overview of the mainstream research works and new research directions to design, build, train, and deploy crowd-counting models at edge devices for real-world scenarios. The talk will cover novel techniques such as annotation techniques, density estimation, knowledge distillation, curriculum learning, dataset pruning, importance scoring, and sample ranking.

Biography: Muhammad Asif Khan is a Research Scientist at Qatar Mobility Innovations Center (QMIC), Doha, Qatar. He was a postdoctoral research fellow at Qatar University. He received a Ph.D. degree in electrical engineering from Qatar University (2020), an M.Sc. degree in telecommunication engineering from the University of Engineering and Technology Taxila, Pakistan (2013), and a B.Sc. degree in telecommunication engineering from the University of Engineering and Technology Peshawar, Pakistan (2009). He is the recipient of the Postdoctoral Research Award (PDRA) from the Qatar National Research Fund (QNRF) in 2022. He has published over 50 peer-reviewed articles and book chapters. He is a senior member of IEEE, a member of IET, and a Chartered Engineer (CEng) with the Engineering Council (UK).

Dr. Khan also serves as an Associate Editor of IEEE Transactions on Consumer Electronics (TCE), IEEE Transactions on Technology and Society (TTS), and IEEE Future Directions Technology Policy and Ethics Newsletter.

ONSITE SESSION 1

Onsite Session 1: Vision-based Image Processing and Computational Models
Chairperson:

TIME: 13:30-17:15 Saturday, January 11 | VENUE: Empire Boardroom (Room 3), 1F

Part A

Time / Paper ID	Speech Title / Presenter
13:30-13:45 P007	**Multi-vision Encoder with Salient Foreground Separation for Video Anomaly Detection** Suhang Cai, Faculty of Electrical Engineering and Computer Science, Ningbo University North Campus, China Abstract-Video anomaly detection (VAD) plays a crucial role in intelligent surveillance systems across various public spaces and industries. The visual feature encoder is critical to VAD performance. However, previous works typically employ either a video encoder or an image encoder, limiting their ability to simultaneously recognize both motion and spatial semantics. Additionally, the significant redundancy in surveillance video data remains largely unaddressed. In this paper, we propose a novel multi-vision encoder architecture that combines an image encoder and a video encoder to effectively capture both motion and spatial information. Furthermore, we introduce a salient foreground separation pooling (SFSP) module that directs the model's attention to motion-significant regions in surveillance videos, enhancing the detection of potentially anomalous events. Extensive experiments demonstrate that our approach achieves competitive performance on two benchmark datasets, UCF-Crime and TAD, validating the effectiveness of our proposed modules.
13:45-14:00 P006	**Physics-Based Image Generation Using a Variational Autoencoder for Efficient SHM and Damage Detection in Aeronautical Composite Plate** Fakhreddine Ababsa, PIMM - ENSAM, France Abstract-Structural Health Monitoring (SHM) aims to monitor the condition of industrial structures through autonomous and non-destructive in situ measurements. For SHM purposes, thin aeronautical composite structures are equipped with piezoelectric elements (PZT) that function both as sensors and actuators, transmitting and receiving ultrasonic waves. The input data for SHM consists of 1D time series associated with various actuator-sensor PZT path geometries (ASPG) and are fed into machine learning algorithms (MLA). Their dimensionality can be reduced using damage indices (DI) that describe associated energy, time of flight, or standard deviation, among other features. In this paper, we propose to augment and aggregate the DI with ASPG to construct spatially organized images (SOI) that can be used as input by image-processing MLAs. We also propose the implementation of an autoencoder to generate synthetic SOI images, thereby expanding the database as much as possible to improve the performance of the defect classification model. Experimental results are presented to demonstrate the effectiveness of the proposed approaches.

14:00-14:15 P016	**ROS-Based LiDAR Mapping and Camera-Driven People Tracking for Autonomous Robots** Jean Sindayigaya, Warsaw University of Technology, Poland Abstract-The project of a robot intended for autonomous mapping and environmental exploration is described in the paper. Whether the robot navigates automatically or manually, it is always beneficial to create maps of the area. The robot uses sensors to sense its environment; sophisticated sensors like LiDAR (Light Detection and Ranging) are especially important because they offer information for creating intricate maps. Simultaneous Localization and Mapping (SLAM) is one of the methods used in the creation of environment maps. Together with sophisticated motion planning and obstacle avoidance strategies, the produced maps are utilized for real-time robot navigation, which guarantees effective and secure environment exploration. Determining the distribution of persons in rooms was the goal. The project created rooms, detection items (people), the robot itself, and its sensors using the Gazebo simulation environment. The ROS localization package's SLAM technique was used to map the environment using information from the wheel odometry and IMU sensor. The SORT tracking module and the YOLOv5 object recognition technique were used to track things in camera photos. The robot and object had a 99% success rate at the closest distance and a 74% success rate at the furthest distance in the object detection tests. Additionally, tests in which the robot and objects remained immobile had the best tracking accuracy.
14:15-14:30 P012	**Adaptive Laser Stripe Extraction Method Based on Semantic Segmentation for Industrial Measurement** Zhipeng Song, School of Instrumentation and Optoelectronic Engineering, Beihang University, China Abstract-The 3-D measurement based on line structured light sensors plays an important role in industrial measurement, but the harsh environment (lighting, physical properties, noise) during industrial measurement seriously affects the accuracy of laser stripe extraction, resulting in inaccurate measurement. This paper introduces an adaptive ROI laser stripe extraction algorithm that combines deep learning and traditional methods to address the challenge of stripe extraction in industrial measurement. Aiming at the problem of small proportion of laser stripes in images, a network structure based on ResNet encoder and MultiAttention FusionNet based on FCN Net decoder is proposed, which combines multi-scale feature fusion and channel attention mechanism to improve the robustness and extraction accuracy of stripes. Finally, the final extraction result is obtained by filtering the stripe length and direction vector. The experimental verification was carried out using wheelset measurement under lathe re-profiling conditions as an example, and the results showed that the extracted stripes met the measurement requirements, indicating the advantages of the proposed method.
14:30-14:45 P405-A	**Color image steganalysis based on improved EfficientNet and hybrid downsampling** Tong Fu, Southeast University, China Abstract-Currently, with the development of convolutional neural networks (CNN), numerous deep learning-based steganalysis models for grayscale im-ages have emerged. However, as color images are more prevalent on the Internet, using CNNs for color image steganalysis has gradually become a re-search focus. Nevertheless, existing

	steganalysis models often exhibit insufficient feature extraction capabilities after extracting noise residuals in the preprocessing layer. Additionally, their downsampling stages are typically designed based solely on convolutional structures, failing to integrate the visual Transformer structure, which can capture global features. To address these issues, we propose a color image steganalysis network (IHNet) based on improved EfficientNet and hybrid downsampling built upon UCNet. This model comprises a feature extraction module, feature aggregation module, and classification module. In the feature extraction module, we improve the structure of EfficientNet by replacing the original Squeeze-and-Excitation (SE) module with SKConv, enabling multi-scale extraction of noise residuals from the preprocessing layer. In the feature aggregation module, while pre-serving the convolutional structure of UCNet, we introduce a Transformer downsampling block, effectively fusing convolutional and Transformer structures in series to capture and process both local and global information within images. The experimental results demonstrate that the proposed IH-Net achieves the best detection performance compared with other start-of-the-art methods. Besides, we design ablation experiments to verify the rationality of each component.
14:45-15:00 P418	**Research on Quantum Phase Perceptron for Solving XOR and XNOR Problems** Shuang Cong, University of Science and Technology of China, China Abstract-A quantum phase perceptron (QPP) based on phase representation of quantum states is proposed in this paper. The proposed quantum perceptron is constructed by a single-layer structure with two inputs and one out-put. The input and qubits are transformed into phase representations with amplitude 1 through H gate phase coding. after quantum coding, all operations of the network are performed phase shift and phase sum separately to obtain a single output of the network, and all operations satisfy unitary operation. The operator that rotates around the y-axis is used to realize the phase shift, and the quantum controlled non-gate (CNOT) is used to realize the quantum phase superposition of two qubits. The probability amplitude of 1 in the output quantum state of the network is measured, and the real output is obtained. This paper also deduces the learning algorithm of quantum perceptron network weights to solve the "XOR" problem. The experimental results show that for many initial weight phases, the proposed QPP can realize the XOR and XNOR functions which the classical perceptron cannot solved.
15:00-15:15 P014	**Transfer-Function-based Text Recognition for Condition Monitoring** Dailin Zhang, Huazhong University of Science and Technology, China Abstract-In the industrial and medical sectors, the condition information of many devices is often not directly transmittable due to security or commercial considerations, which poses challenges for real-time monitoring and data analysis of text recognition. This paper optimizes the training of the DBNet model for text detection and the Convolutional Recurrent Neural Network (CRNN) for text recognition, and a transfer-learning-based method is proposed to improve the accuracy and processing speed of the text recognition. Experiments show that the proposed method can achieve a higher accuracy and processing speed than traditional method in text recognition.

15:15-15:30 Break Time

Part B	
Time / Paper ID	Speech Title / Presenter
15:30-15:45 P015	**2D Vision-Based with Monocular Depth Estimation for Pose Estimation** Pitijit Charoenwuttikajorn, Sirindhorn International Institute of Technology, Thammasat University, Thailand Abstract-Pose estimation with a vision system is a major factor during bin picking on the ability of the system to grasp the object. It helps the robot visualize the position and orientation of each object, through the pose that has been identified correctly. 3D vision systems have accuracy but are expensive which restricts them in budget-constrained circumstances. To solve these limitations, this paper proposes a 2D vision-based approach that uses RGB systems for pose estimation in the bin-picking approach. The core of the approach consists of deep-learning techniques that use Mask R-CNN for object detection and segmentation. The Monocular depth estimation using a multi-scale network (MiDaS) model for effective depth estimation in the absence of 3D sensors. Through the estimation of the relative depth between the camera and the objects. Distance calculation of the object from the camera using the metric based on the object's width and the focal length. The experiments show that the model is capable of correctly predicting pickup points, demonstrating its suitability for tasks where minor positional deviations are acceptable.
15:45-16:00 P034	**Transfer learning for 3D reconstruction system of upper-limb anatomy** Yanzhen Dong, School of Mechanical Engineering, Shanghai Jiao Tong University, China Abstract-The 3D reconstruction of upper limb anatomy plays a crucial role in various biomedical fields, such as ergonomics, motion rehabilitation, and prosthesis design. In recent years, advancements in 3D acquisition devices and computer-aided design have enabled rapid, high-precision digital reconstruction of upper limb structures. However, real-time data capture of human body parts remains challenging due to the non-static and texture-less nature of the scanning target. Moreover, the 3D models constructed from images must meet high standards of completeness and accuracy for practical applications. In this study, a binocular speckle-projected optical acquisition device was designed for data capture. A transfer learning method based on a matching network was proposed, where fine-tuning enabled the stereo matching network to output high-quality depth maps, better suited for reconstructing upper limbs. Results indicate that our proposed system successfully reconstructs various parts of the upper limb in 3D, offering better point cloud completeness compared to traditional structured light methods, with a single shot reconstruction time of approximately 0.2s. This system provides an effective solution for the acquisition and reconstruction of 3D upper limb data, achieving high-quality models suitable for custom medical device design.
16:00-16:15 P401	**Hardware Architecture Design of the Canny Edge Detection Algorithm for UHD Videos** Jiesen Feng, Guangdong University of Technology, China Abstract-Edge detection is a crucial technique in the field of image processing. The Canny edge detection algorithm is an efficient edge detection algorithm with high accuracy. However, the high computational complexity of the Canny edge detector hinders its real-time hardware implementation. Therefore, this paper presents a hardware-friendly Canny edge detector, which aims to process 4K@30fps UHD videos.

	This Canny edge detector provides a block-level processing method to reduce latency and introduces a fast arctangent approximation algorithm to simplify the complex computation of the conventional Canny algorithm. Moreover, we propose an efficient parallel hardware architecture for each module in the Canny detector to achieve high throughput. Experiments demonstrate that the detection performance of our proposed fast Canny algorithm is similar to that of the traditional Canny algorithm. Additionally, our designed hardware architecture achieves the requirement of processing 4K@30fps UHD videos at 100 MHz clock frequency.
16:15-16:30 P413	**Soft--ROI created through sigmoid functions** Wei-Jun CHEN, Carl Zeiss Group, Germany Abstract-This paper proposes to create a new type of soft-ROI (window) for signal and image processing through sigmoid functions. The S-shaped ROI boundary enables the sidelobes of a hard rectangular ROI to be largely reduced, while the relatively flat ROI area makes it better to maintain the raw discrete signal resolution compared to the common bell-shaped ROIs (windows). At the same time, such a new type of ROI is infinitely differentiable and parameterizable, making it powerful and adaptable to meet many types of application requirements.
16:30-16:45 P411-A	**A Lightweight Image Forgery Prevention Scheme for IoT using GAN-based Steganography** Xiao Li, Southeast University, China Abstract-Computer Vision (CV) applications empower various Internet of Things (IoT) scenarios. However, their advancements in image generation and manipulation tools make it increasingly easy to produce highly deceptive forged images, escalating the risk of image forgery. Cryptography-based methods can secure images but can not support direct CV applications with compromised visual legibility. Existing Generative Adversarial Network (GAN)-based steganography methods can effectively facilitate CV applications and image forgery prevention with high indistinguishability between stego and cover images. However, they are inefficient in resource-constrained IoT scenarios. Therefore, we propose a lightweight image forgery prevention scheme for IoT using GAN-based steganography. Our scheme embeds identity data within images. If forged, it fails to recover, triggering alerts. Our scheme can significantly im-prove efficiency with a lightweight generator designed by incorporating blue-print separable convolutions, sum connections and discrete wavelet transform while ensuring high effectiveness. These designs ensure high effectiveness by exploiting features with reduced redundancy and significantly improve efficiency by requiring few computational resources and storage costs. Real world IoT experimental results demonstrate this. Our scheme provides flexible user customization, ensuring security through the confidentiality of the generator parameters and keys for pre-processing. Users can determine the locations of identity data using these keys and decide their quantity themselves without compromising security. They can also flexibly set the identity data with extensive capacity (e.g., 16,384 bits/2,048 bytes in images with a size of 128×128). Moreover, identity data can be recovered directly from stego images without requiring the original cover images, ensuring an error-free recovery.

16:45-17:00 P032	**Integrating YOLOv8 and Eigen-CAM: A Novel Framework for Explainable Brain Tumor Segmentation** Ahmed Bouridane, Sharjah University, UAE Abstract-Brain tumor segmentation is a critical task in medical imaging, providing essential information for diagnosis, treatment planning, and monitoring of brain cancer. While deep learning models, such as YOLO (You Only Look Once), have demonstrated remarkable performance in object detection and segmentation, their black-box nature limits their applicability in medical contexts where transparency and explainability are paramount. To address this challenge, we propose a novel framework that integrates YOLOv8 with Eigen-CAM, termed YOLO-Eigen-CAM, for accurate and explainable segmentation of brain tumor subregions, specifically focusing on Core, Edema, and Enhanced Tumor. The YOLO architecture is utilized for efficient detection and segmentation of tumor subregions, leveraging its end-to-end, one-stage detection capabilities to achieve high precision and speed. Eigen-CAM, an advanced visualization technique, is integrated to generate class activation heatmaps that elucidate the model's decision-making process, providing interpretability for each predicted segmentation. The framework was evaluated on BraTS 2023 dataset, demonstrating its effectiveness in both segmentation accuracy and explainability. The YOLOv8-Eigen-CAM model achieved competitive performance metrics, with a mean average precision (mAP) of 0.879 across all tumor subregions and consistently high recall rates. The integration of Eigen-CAM enabled visual assessment of model predictions, which revealed that the framework accurately localizes tumor subregions while offering insights into the regions of the image contributing to the decisions. Comparative analysis showed that YOLOv8, an earlier version of the architecture, outperformed YOLOv11 in terms of training stability and precision, suggesting that further optimizations in YOLOv11 are required. Our results underscore the potential of combining advanced detection models with explainability techniques to develop robust and interpretable tools for brain tumor segmentation. The proposed YOLO-Eigen-CAM framework sets a foundation for future research in explainable AI for medical imaging, paving the way for more transparent and reliable AI-assisted diagnostic systems.
17:00-17:15 P029-A	**Processing shallow profiling images based on peak detection method** Zhiwei Zhao, Southeast University, China Abstract-Shallow profiling technology plays a crucial role in the fields of marine science and engineering, with extensive applications in areas such as measuring the acoustic properties of the Earth and detecting and identifying objects buried beneath the seafloor. In practical engineering applications, the structural contours of underwater sediment layers shown in images obtained by shallow profiling instruments often lack the clarity required for detailed analysis. A primary challenge has been improving the detection resolution of shallow profiling images, which remains a focal point of research. To achieve this goal, this article presents a novel method designed to enhance the resolution of shallow profiling images. This method uses peak detection technology to process the detected underwater shallow profiling images obtained by instruments. Then comparing the processed images with the actual sampling results. Through the process, the structure of the seafloor sediment layers can be clearly displayed, while the shape and location characteristics of the underwater sediments can be accurately detected. This method provides a new technological approach for the research of high-resolution shallow profiling technology.

ONLINE SESSION 1

Online Session 1: Digital Image Detection and Processing Methods
Chairperson:

TIME: 13:30-18:00 Saturday, January 11 (UTC+8) | ZOOM ID: 813 0177 8800
Meeting Link: https://us02web.zoom.us/j/81301778800

Part A

Time / Paper ID	Speech Title / Presenter
13:30-13:55 **Invited Talk**	**Real-time Crowd Counting at the Edge** **Dr. Muhammad Asif Khan, Qatar Mobility Innovations Center, Qatar University, Qatar** Abstract-Recent research on crowd counting shows the efficacy of deep learning methods such as CNNs due to their strong capability of auto-feature extraction. To achieve higher accuracy in dense scenes, deeper models with a large number of parameters are developed. However, although achieving good accuracy, deeper models create performance bottlenecks in real-time applications due to large memory requirements, higher training complexity, and large inference delay. On the contrary, shallow models are lightweight, incur low inference delay, and require low memory, but are usually disregarded by some due to their limited accuracy in dense crowd scenes. This talk will discuss how to design efficient and more robust lightweight models for crowd-counting applications at the edge providing useful insights for prospective researchers and DL practitioners. This talk provides a retrospective overview of the mainstream research works and new research directions to design, build, train, and deploy crowd-counting models at edge devices for real-world scenarios. The talk will cover novel techniques such as annotation techniques, density estimation, knowledge distillation, curriculum learning, dataset pruning, importance scoring, and sample ranking.
13:55-14:10 P022	**Enhancing Super-Resolution Models: A Comparative Analysis of Real-ESRGAN, AESRGAN, and ESRGAN** Avinash Senthil, Cambridge Centre for International Research, Ltd., United States Abstract-This paper conducts a comparative analysis of three prominent image super-resolution models: ESRGAN, Real-ESRGAN, and AESRGAN. Utilizing Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as primary quantitative metrics, we assess the performance of each model against ground truth images. Our findings reveal that ESRGAN achieves an average PSNR of 26.45 dB and an average SSIM of 0.74, indicating its foundational role in super-resolution. In contrast, Real-ESRGAN, designed to optimize performance in real-world scenarios, scores an average PSNR of 29.15 dB and an SSIM of 0.78, demonstrating its ability to reduce artifacts and preserve structural details effectively. The standout model, AESRGAN, achieves the highest scores with an average PSNR of 29.70 dB and an SSIM of 0.82, reflecting its advanced feature extraction and attention mechanisms. We present a detailed comparison of PSNR and SSIM across ten instances, confirming that AESRGAN consistently outperforms both Real-ESRGAN and ESRGAN in image quality metrics. This analysis not only highlights the strengths and limitations of each model but also provides insights into their practical applications in enhancing image quality in various contexts.

14:10-14:25 P002	**Enhancing Hyperspectral Image Classification with Integrated CNN-GAN Models and Attention Mechanisms** Zhinan Xie, ShanghaiTech University, China Abstract-In the evolving landscape of remote sensing and computer vision, Hyperspectral Image (HSI) Classification has gained prominence, largely driven by advancements in neural networks. Despite significant progress, a critical aspect often overlooked is the effective utilization of spectral features extracted by Convolutional Neural Networks (CNN). Addressing this gap, our study introduces innovative methodologies that synergize CNN with Generative Adversarial Networks (GAN) to optimize their integration and fully harness their perceptual capabilities. GAN can produce new images to enhance the dataset and CNN can efficiently capture spatial spectral information. This fusion not only enhances the feature extraction from pixels during generation and discrimination phases but also elevates the model's sensitivity and adaptability, which are crucial in image classification. Furthermore, our research extends to the development of a model that incorporates an Attention mechanism within the network framework. This addition allows the model to dynamically adjust its focus based on specific input data, thereby improving the precision of feature analysis. Through comprehensive experiments and analysis, including various models such as Hybrid Net and Hybrid Net + Attention, our findings demonstrate a significant improvement in overall accuracy. These enhancements, particularly the integration of GAN and Attention mechanisms with CNN, present a promising direction for the field of HSI Classification, offering notable contributions to both theoretical and practical applications in image analysis.
14:25-14:40 P005	**Enhanced Change Detection in SAR Images Using DWT and Optimised Loss Functions** Mohamed Ihmeida, Buckinghamshire New University, UK Abstract-Synthetic aperture radar (SAR) image change detection (CD) involves identifying changes between images captured at different times over the same geographical region. SAR provides significant advantages for disasterrelated remote sensing due to its all-weather capabilities and ability to penetrate clouds and darkness. Nevertheless, the presence of speckle noise presents a major challenge for accurate change detection. This paper introduces a robust method utilising a dual-domain model that integrates the DiscreteWavelet Transform (DWT) with biorthogonal wavelets to effectively suppress speckle noise. Furthermore, we propose an improved loss function that combines Mean Squared Error (MSE) and Kullback-Leibler Divergence (KL) to further enhance change detection accuracy. Extensive evaluations on three SAR datasets demonstrate that our approach outperforms state-of-the-art methods, significantly improving change detection accuracy.
14:40-14:55 P011	**Photometric Stereo-based Defect Detection for Lithium Battery Top Covers** Li Huayong, Dongguan University of Technology, China Abstract-Detecting surface defects on lithium battery top covers is critical for ensuring their performance, reliability, and longevity. Photometric stereo (PS) methods, known for their ability to capture detailed surface topography with high precision and speed, have been widely applied in industrial defect detection. However, these methods often face challenges in dynamic scenes with varying lighting conditions and in detecting subtle defects such as shallow scratches. To address these limitations, we propose an enhanced approach that combines the advantages of photometric stereo with a line-

	scan camera and weighted least squares filtering. This novel combination improves the contrast of surface defects, particularly in fine or shallow scratches, by enhancing the surface detail and minimizing noise. Experimental results demonstrate the effectiveness of this method, showing significant improvements in the detection accuracy and sensitivity to fine defects on the battery end surfaces.
14:55-15:10 P414	**A low noise bandgap reference with high PSR for EEG acquisition chip** Lu Liu, Xidian University, China Abstract-The bandgap reference voltage is the vital module of the analog circuit power supply system. For on-chip EEG acquisition systems, a power supply design characterized by low noise, low power consumption, and a high-power supply rejection (PSR) is crucial. Based on the traditional bandgap circuit, this paper realizes low noise and high PSR by adopting the PSR enhancement technology and low noise design to satisfy the requirements of the EEG acquisition chip. In this design, the startup circuit and the trimming circuit are also implemented to address the degenerate working point problem and the manufacturing process error. The circuit design has been fabricated using a 40nm 1P7M CMOS process. The simulation results show that the bandgap reference voltage can produce a low-frequency PSR of -99.83dB, a temperature drift factor of 8.341ppm/°C across the temperature range of 0~80°C and an integrated noise of 10.49 μVrms within the frequency range of 10-100 KHz
15:10-15:25 P018	**Feasibility of identifying two phase flow quality by Shadowgraphy** Ian Kemp, School of Architecture Technology and Engineering, University of Brighton, UK Abstract-Two-phase process conditions are applied across numerous applications due to their effectiveness in energy transfer, production and storage. Two-phase is important for heat transfer in cryogenics, dense expansion in power production, and storage in liquid-cooled batteries. Shadowgraphy capture offers a non-invasive technique that allows a real-time measurement of characteristics, such as steam wetness. If phase conditions could be determined during operation, optimum fluid wetness could be maintained, resulting in a more efficient energy system. This paper employs the optical technique of shadowgraphy, utilising a Photron SA4 camera for image capture, coupled to the developed algorithm in the imaging software tool, ImageJ, to develop a method for the process of estimating steam quality and void fraction captured in a horizontal orientation. Void fraction was achieved using chordal widths of 20 pixels, both horizontal and vertical across a 30 mm tubular sight glass, estimating the chordal liquid content with a time-averaged quantitative comparison, using processed greyscale images. Techniques for further optimisation of accuracy are detailed with respect to image visulisation, image processing, and importantly, results from a gravitational condensate trap are presented for validation. Results for a flow velocity of 50 m/s and at a pressure of 1.3 bar were calculated as near 0.98 gaseous content. This work demonstrates the feasibility of shadowgraphy for wet steam process conditions, and offers a new measurement technique for estimating the two-phase fraction in flows of high vapour content, thus complementing the advancement in parallel techniques such as vortex and orifice flow measurements.

15:25-15:40 P009	**Evaluation of the YOLO architecture in Object Detection from aerial images** Luis Rodrigo Barba Guaman, Universidad Técnica Particular de Loja, Ecuador Abstract-This paper evaluates the effectiveness with which YOLO structures perform object detection in aerial imagery, with an emphasis on applications related to intelligent traffic control. The study compares the performance of Yolov5, YOLOv7, YOLOv8 and YOLOv9 to determine which model is the most accurate. The process involved collecting and annotating a dataset of aerial imagery and then training and evaluating each model. According to the results, the accuracy of Yolov5Aug, which was so named because of its data augmentation and more robust training, was 85% as was that of YOLOv8Aug, while that of YOLOv9Aug and YOLOv5 was 84%. These results demonstrate the effectiveness of YOLO designs in intricate contexts such as aerial imagery. In summary, Yolov5Aug and YOLOv8Aug outperforms Yolo5 and Yolo8 designs in identifying aerial objects, which supports the high accuracy and effectiveness of YOLO architectures. These results hold great promise for future research and real-world applications in traffic control and management systems.
15:40-15:55 P1002	**Medical image segmentation with hybrid neural networks** Ling-Fang Li, Southwest Jiaotong University, China Abstract-Recent advancements in CNN and Transformer-based models have significantly advanced medical image segmentation. However, these approaches have inherent limitations in capturing long-range and local relationships. In this paper, we introduce a Dual-Encoder Fusion Model that incorporates a novel Correlation Fusion Module (CFM) for medical image segmentation tasks. This model leverages the strengths of Convolutional Neural Networks (CNN) for local context modeling and Transformers for comprehending long-range dependencies in pixel interactions. Experimental results demonstrate a substantial improvement over existing models on the Synapse dataset, achieving enhancements of 2.28% and 3.47% on the dice metric for Aorta and Pancreas organs respectively. Additionally, our model attains the highest mean HD95 score of 9.05 on the Synapse dataset while utilizing fewer parameters. When evaluated on the MSD datasets, our model outperforms a fine-tuned nnUNet in three out of five tumor detection tasks and maintains competitive performance in three out of four organ boundary delineation tasks. Notably, on the MSD-Lung dataset, our model surpasses a fine-tuned nnUNet on the dice metric by 6.4%. These results underscore the effectiveness of the CFM module within the dual-encoder architecture.

15:55-16:00 Break Time

Part B	
Time / Paper ID	Speech Title / Presenter
16:00-16:15 P1004	**Image Captioning Using Sparse Semantic Features** LIANG junbin, Jinan University, Zhuhai, China Abstract-In the task of image caption generation, it is necessary to pay attention to both the extraction of visual information and the expression of linguistic information, and the use of multiple different descriptive methods applied to the training process of the image caption generation model at the same time can make the image caption generation model take into account the sparse semantics that exists in the multiple descriptive methods, so as to improve the quality of the generated utterances. In this paper, we propose to add a sparse semantic vector generation module (Context Query Generator, CQG), which is composed of a multi-attention mechanism, between the image encoder and the language decoder to generate sparse semantic vectors, which are aligned with the sparse semantic features of multiple descriptive statements and then decoded to obtain the utterances. In addition, in order to further improve the efficiency of sparse semantic alignment, the paper also designs a corresponding supervised training method and BCEFocalLoss loss function with reference to the target detection model DETR[1]. The method is applied to several test models and has been tested on three datasets, Flickr8k, Flickr30k, and MS COCO, resulting in 5-40% improvement.
16:15-16:30 P416	**A High Precision Temperature Sensor for Biological Medical Devices** Shuoshuo Zhu, Xidian University, China Abstract-This paper presents a temperature sensor for biological medical devices using 40-nm CMOS technology to prevent the high operating temperature of devices from causing harm to human organs. Bipolar PNP transistors are employed to sense the temperature and generate a voltage proportional to the absolute temperature. To improve the resolution, the variation in output voltage per degree Celsius is amplified by the programmable gain amplifier (PGA) with adjustable gain. The structure of PGA is capacitor feedback and the input capacitor is controlled by switching to achieve different gains. The conversion of temperature to digital output is accomplished by a single-slope analog-to-digital converter (SS ADC). The SS ADC mainly consists of a comparator and a ramp generator. Considering the power, the ramp generator uses an integrated structure of switched capacitor, and a dynamic latch combined with a static comparator forms the comparator. The proposed temperature sensor can achieve a high resolution of 0.045°C within the temperature range of 20°C to 60°C, with a conversion rate is 14 KHz.
16:30-16:45 P008	**Evaluation and Comparison of Deep Learning Models for Vehicle Detection in Aerial Imagery of Urban Environments** Luis Rodrigo Barba Guaman, Universidad Técnica Particular de Loja, Ecuador Abstract-Vehicle detection is essential for applications such as surveillance, traffic control, and autonomous driving. This study aims to evaluate and compare the performance of various deep learning models for vehicle detection in urban environments, utilizing a standardized dataset of aerial images captured by UAVs. The methodology includes data pre-processing, augmentation, and the application of models such as YOLO, TensorFlow SSD, and Roboflow, with each model assessed based on key metrics: precision, recall, mean average precision (mAP), F1 Score, and

	training time. Results indicate that the Roboflow 3.0 COCOs model achieved the highest performance, providing an optimal balance between accuracy, recall, and efficiency. Other models, such as TensorFlow SSD and YOLOv5, also demonstrated competitive results, highlighting their potential for real-time applications. However, challenges remain in improving detection accuracy under adverse conditions, such as low visibility and complex traffic scenarios. This research underscores the promise of integrating UAVs and deep learning for enhanced vehicle detection and identifies areas for future research to further improve model robustness and real-time applicability.
16:45-17:00 P417	**Optimal Design of LED Array Based on Simulated Annealing Algorithm** Ying Huang, University of electronic science and technology of China, China Abstract-With the continuous development of the automobile industry, head-up dis-play (HUD) technology has been widely used in modern auto-mobiles. In the Picture Generation Unit (PGU) of a HUD system, since the liquid crystal it-self does not emit light, the backlight module is required to provide lighting. The illumination uniformity of the LED array light source in the backlight module is crucial to the performance and display effect of the backlight module. In order to improve the illumination uniformity of the backlight LED array, this paper optimizes the LED array arrangement using the simulated annealing algorithm to achieve uniform irradiance on the receiving plane. First, the irradiance distribution function of the LED array is derived, and an evaluation function is constructed to measure the uniformity of the illumination. The simulated annealing algorithm is applied to calculate the optimal LED array spacing, and the geometric model is imported in-to optical simulation software for validation. The simulation results demonstrate that the algorithm effectively improves the uniformity of irradiance. Furthermore, the introduction of a diffuser film further improves this uniformity.
17:00-17:15 P027	**Artificially Intelligent Diagnosis of Lumbar Spine Muscle Spasm from Sagittal MRI Scans: Clinical Significance Classification Using Convolutional Neural Network** Olajide Odelanu, Solenis Industries Limited, United Kingdom Abstract-Lower back pain (LBP) is a common and costly medical condition with significant implications. This research focused on identifying muscle spasm in the lumbar spine, which is a prevalent and indicative condition. This means muscle spasm can serve to indicate other underlying abnormal conditions, which can vary from no to serious levels of clinical significance. Traditional diagnostic methods, which rely on manual analysis of MRI images, are time-consuming and resource-intensive, often causing appointment delays. To improve diagnostic accuracy, reduce radiologists' workload, and boost confidence in diagnoses, this study developed a multiclass Convolutional Neural Network (CNN) model to classify muscle spasms in MR images into four levels of clinical significance: None, Mild, Moderate, and Severe. This classification is based on the severity of abnormal condition the spasm is indicating. The model achieved 99% across accuracy, recall, precision, and F1-score.

17:15-17:30 P030	**USPNet: A novel network for intrapartum ultrasound standard plane classification** Yifan Zhang, Jinan University, China Abstract-This paper presents USPNet, a novel model designed for the classification of standard planes in intrapartum ultrasound images, focusing on improving the classification accuracy of small anatomical targets. Intrapartum ultrasound, as a non-invasive, objective, and reproducible method, is increasingly utilized to evaluate labor progress. However, due to the small size of key anatomical features, accurate classification remains challenging. USPNet addresses this challenge by introducing a simplified architecture that incorporates a Shift block, SC-MLP, and ORIConv blocks to enhance feature extraction. Experimental results on two datasets, Zj-AOP and Zj-CD, demonstrate that USPNet outperforms existing advanced models in terms of accuracy, precision, recall, and F1-score. The proposed model effectively improves the classification of small anatomical targets, providing enhanced support for clinical decision-making during labor.
17:30-17:45 P035	**Explaining Arabic Sign Language Recognition: A YOLOv11 and Transformer Hybrid Model with Grad-CAM Visualizations** Nabil Hezil, Computational Linguistics department CRSTDLA, Algeria Abstract-Arabic Sign Language (ArSL) recognition plays a vital role in facilitating communication between deaf communities and the hearing population in Arab countries. This paper presents a novel approach to ArSL recognition from video using a combination of YOLOv11 and Transformer models. YOLOv11 is employed to detect and track hands in each video frame, enabling efficient extraction of spatial features, while a Transformer is used to model the temporal dependencies between frames for accurate gesture recognition. This hybrid model allows for the recognition of both isolated and continuous gestures, making it highly adaptable to real-world scenarios. To further enhance the explainability of the model, we integrated Gradient-weighted Class Activation Mapping (Grad-CAM), which provides visual insights into the areas of the input frames that the model focuses on during gesture recognition. By generating heatmaps, Grad-CAM helps identify whether the model accurately captures the relevant hand movements or if it mistakenly focuses on background elements, improving the interpretability of the model's decision-making process. The proposed model was evaluated on the KArSL-100 dataset, achieving a training accuracy of 99.5% and a test accuracy of 99.5%, with a loss of 0.02. These results demonstrate the effectiveness of combining YOLOv11 and Transformers for ArSL recognition, with strong performance in both training and testing phases, minimal overfitting, and robust explainability through Grad-CAM visualizations.
17:45-18:00 P412-A	**Real-Time Detection and Classification of Heart Arrhythmia using ECG Feature Detection with Moving Statistic Adaptive Thresholding Algorithm in Microcontroller Systems: A Comparison of Long-Time Cigarette and Vape Smokers** Vinze Lawrence B. Reyes, Mapua University, Philippines Abstract-This study presents the development and validation of a microcontroller-based electrocardiogram (ECG) system designed to detect cardiac arrhythmias, specifically Sinus Tachycardia and Atrial Fibrillation. The validation involved a comparative analysis with standard ECG readings to assess the system's accuracy, sensitivity, specificity, and F1 score. The results demonstrated high accuracy across

	the tested conditions, with the system achieving 88.10% accuracy, 77.78% sensitivity, 91.67% specificity, and 82.22% F1 score. The study included 20 participants, divided into two groups: cigarette smokers and vape smokers. The cigarette smokers, with an average age of 24.4 years, an average nicotine intake of 6.9 mg per day, and an average smoking history of 76.8 months, showed a varied occurrence of arrhythmias, with one case of Sinus Tachycardia and five (5) cases of Sinus Rhythm detected. Furthermore, the vape smokers, averaging 23.6 years of age with a higher nicotine intake of 7.99 mg per day and a shorter vaping history of 21.6 months, showed no cases of Sinus Tachycardia or Atrial Fibrillation and five (5) cases of Sinus Rhythm. These findings suggest the impacts of cigarette and vape smoking on cardiovascular health, with potential implications for arrhythmia occurrence. Based on these results, the researchers recommend further developing the microcontroller ECG system to enhance its portability, including adding a display screen and battery power source. Moreover, expanding the system's diagnostic capabilities to include a broader range of arrhythmias is advised to help diverse patient populations better. The study emphasizes the need for ongoing system refinement and validation to ensure continued accuracy and reliability in clinical settings.

NOTE

Co-sponsored by:

香港大學計算機科學系 THE UNIVERSITY OF HONG KONG DEPARTMENT OF COMPUTER SCIENCE

澳門大學 UNIVERSIDADE DE MACAU UNIVERSITY OF MACAU

Supported by:

武漢理工大學 WUHAN UNIVERSITY OF TECHNOLOGY

대구대학교 DAEGU UNIVERSITY

MAHANAKORN UNIVERSITY OF TECHNOLOGY

Universidad Politécnica de Cartagena

EUT+ EUROPEAN UNIVERSITY OF TECHNOLOGY

CTIM CENTRO DE TECNOLOGÍAS DE LA IMAGEN