PROGRAM
CONFERENCE

IWPR 2025

2025 5th International Workshop on IMAGE PROCESSING

IWIP 2025

2025 10th International Workshop on PATTERN RECOGNITION

JUNE 13-15, 2025 SINGAPORE

Conference Program

2025 10th International Workshop on Pattern Recognition (IWPR 2025)

2025 5th International Workshop on Image Processing (IWIP 2025)

Singapore | June 13-15, 2025

Sponsored by

Sensors and Systems Society of Singapore

NEWCASTLE AUSTRALIA INSTITUTE OF HIGHER EDUCATION | A WHOLLY OWNED ENTITY OF THE UNIVERSITY OF NEWCASTLE

Supported by

Colorado Technical University

VALDOSTA STATE UNIVERSITY

UNIVERSIDADE DE COIMBRA

Published by

SPIE. DIGITAL LIBRARY

Table of Content

Conference Information and Tips

1) Materials Provided by the Presenter

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

※ Official language: English.

2) Duration of Each Presentation

※ Keynote Speech: 45min, including Q&A

※ Invited Speech: 30min, including Q&A

※ Oral Presentation: 15min, including Q&A

3) Notice

※ **UTC+8. Please be aware of time difference between this and your region/country.**

4) Online Presentation Tips

	Meeting ID	Link
zoom Zoom Download	Room ID: 895 3784 4640	https://us02web.zoom.us/j/89537844640

Note:

We recommend that you install the Zoom platform on your computer before the conference starts. New users can participate in the Zoom meeting without registration.

Participants who are going to do an online presentation are required to join the rehearsal in Zoom on Friday, June 13, 2025. Duration: 3min apiece. Feel free to leave after you finish the test.

◆Name Setting

Keynote Speaker: KN-Name

Committee: Position-Name

Author: Paper ID-Name

Delegate: Delegate -Name

◆Useful Links

- Conference Banner
- Zoom Background

Welcome Message

On behalf of Conference Committee, we welcome you to attend 2025 10th International Workshop on Pattern Recognition (IWPR 2025) & 2025 5th International Workshop on Image Processing (IWIP 2025) held in Singapore(changed to online conference) during June 13-15, 2025, sponsored by Sensors and Systems Society of Singapore, hosted in Newcastle Australia Institute of Higher Education, Singapore.

IWPR 2025 & IWIP 2025 welcomes author submission of papers from any branch of pattern recognition and image processing, and their applications or other topic areas. The areas covered by the include, but not limited to: Computer Vision, Biomedical Image Analysis, Pattern Recognition and Machine Learning, Image, Speech, Signal and Video Processing, Document Analysis, Biometrics and Pattern Recognition Applications, Image & Video Analysis and Segmentation, Image and Video Compression, Face Recognition, Medical Image processing, Pattern Recognition in New Modalities.

The conference aims to provide an interactive communication platform for practitioners to learn about the most cutting-edge academic and industrial application trends, to share the latest scientific research and technological achievements, innovative ideas and scientific methods in the field of pattern recognition and image processing, to improve the level of academic research and industrial application in the field of intelligence so as to serve the global strategic deployment of new and old kinetic energy conversion, and promotes technology research, development, and application home and abroad.

We feel deeply grateful to all that have contributed to make this event possible: authors, the conference steering committees, the conference speakers, and the peer reviewers. Thanks are also extended to the conference administrative committee and the supporters for their tireless efforts throughout the course of the conference.

We hope that all participants benefit from these two conferences.

With Warmest Regards,

Conference Organizing Committee

Conference Committee *(in no particular order)*

Conference General Chair
Xudong Jiang (IEEE Fellow), Nanyang Technological University, Singapore

Conference Local Organizing Chair
Kok Chiang Liang, Newcastle Australia Institute of Higher Education, Singapore

Conference Program Chairs
Qiu Chen, Kogakuin University, Japan
Pedro Furtado, University of Coimbra, Portugal
Houjin Chen, Beijing Jiaotong University, China
Bok-Min Goi, Universiti Tunku Abdul Rahman, Malaysia

Conference Program Co-Chair
Kin Choong Yow, University of Regina, Canada

Conference Publicity Chairs
Howard TANG, Nanyang Technological University, Singapore
Ho Chee Kit, Cushman & Wakefield, Singapore

Conference Technical Program Committees
Shiwei Ma, Shanghai University, China
Songyuan Li, University of Exeter, UK
Yoshihiro Mitani, Ube National College of Technology, Japan
Stefano Berretti, University of Florence, Italy
Xianfang Sun, Cardiff University, UK
Alex Ong, Republic Polytechnic, Singapore
Nan Zhu, Xi'an Technological University, China
Touhid Bhuyian, Daffodil International University, Bangladesh
Tsung-Nan Chou, Chaoyang University of Technology, Taiwan
Luepol Pipanmaekaporn, King Mongkut's University of Technology North Bangkok, Thailand
Osama Halabi, Qatar University, Qatar
Birjodh Tiwana, Linkedin Inc., USA
B. S. Harish, JSS Science & Technology University, India

Agenda Overview (UTC+8)

Friday, June 13, 2025		
Zoom Test	**14:00-14:50**	**Room ID: 89537844640** **Link: https://us02web.zoom.us/j/89537844640**

Zoom Test Timetable

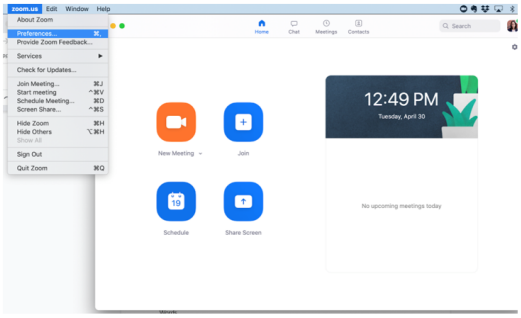
- ✧ Participants who are going to do an online presentation are required to join the rehearsal in Zoom on Friday, June 13, 2025. Duration: 3min apiece. Feel free to leave after you finish the test.
- ✧ We will test control panel including screen sharing, audio, video and "Raise Hand" feature, etc. Please get your presentation slides and computer equipment prepared beforehand.

14:00-14:30	WP501, WP502, WP507, WP517, WP520, WP524, WP604, WP523
14:30-14:50	Alternative time for participants who are unavailable at allocated time. Other online participants, includes but not limited to keynote speaker, invited speaker, session chair, committee member, delegate

Zoom Guidance

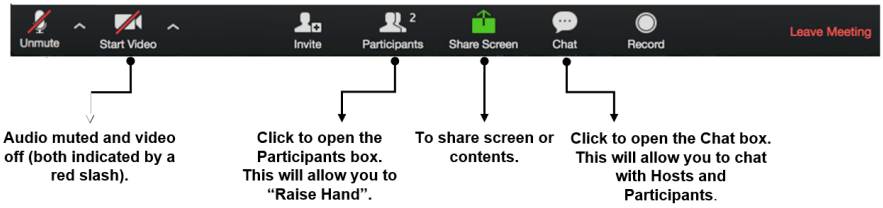
You can join the meeting without sign-in process.
Just put the meeting ID and join us.

URL: https://zoom.us/download



Each meeting has a unique **9, 10, or 11-digit** number called a **meeting ID** that will be required to join a Zoom meeting.

For any questions on the meeting day, you can text privately to "Assistant" for help.



Saturday, June 14, 2025		
Plenary Session	**Room ID: 895 3784 4640**	**Link: https://us02web.zoom.us/j/89537844640**
Host: Kok Chiang Liang, Newcastle Australia Institute of Higher Education, Singapore		
09:00-09:10	Opening Speech: Kok Chiang Liang, Newcastle Australia Institute of Higher Education, Singapore	
09:10-09:55	Keynote Speech I: From Discrimination to Dense Prediction by Deep Learning with Examples of Semantic Segmentation **Xudong Jiang** (IEEE Fellow), Nanyang Technological University, Singapore	
09:55-10:40	Keynote Speech II: Explainable Artificial Intelligence (xAI) for Pattern Recognition **Kin-Choong Yow,** University of Regina, Canada	
10:40-11:00	Online Group Photo & Break	
11:00-11:30	Invited Speech: Biomedical Limb Lengthening Implant **Kok Chiang Liang,** Newcastle Australia Institute of Higher Education, Singapore	
11:30-13:00	Break Time	
Saturday, June 14, 2025 | Parallel Session		
13:00-15:00	**Oral Session: Digital Image Detection and Analysis Method Based on Machine Learning** WP501, WP502, WP507, WP517, WP520, WP524, WP604, WP523	Room ID: 895 3784 4640

Keynote Speaker I

Saturday, June 14, 2025 09:10-09:55	Room ID: 895 3784 4640

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

Speech Title: From Discrimination to Dense Prediction by Deep Learning with Examples of Semantic Segmentation

Abstract: With the increasing powerfulness of deep learning, artificial intelligence moves from simple discriminating task to challenging dense prediction. Visual scene segmentation is an example of dense prediction problem as it need to classify every pixel in the image into different classes. It is crucial to exploit context information and aggregate multi-scale features to achieve dense prediction. Context is essential for dense prediction. Due to the diverse shapes of objects and their complex layout in various scene images, the spatial scales, locations and shapes of contexts for different objects have very large variation and are unknown. There is no way to do any kind of alignment or normalization over different inputs. It is thus ineffective or even impossible to aggregate various context information from a predefined fixed region or simply learnt from training data. This talk will show how to solve these problems by elaborating a number of ideas from two CVPR oral papers and a T-PAMI paper.

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 200 papers with 2 papers in Nature Communications, 20 papers in Pattern Recognition and over 40+ papers in the IEEE journals, including 6 papers in IEEE Transactions on Pattern Analysis and Machine Intelligence and 14 papers in IEEE Transactions on Image Processing. Four 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 Signal Processing Letter from 2014 to 2018, Associate Editor for IEEE Transactions on Image Processing 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 Transactions on Image Processing and Editor-in-Chief for IET Biometrics. His current research interests include image processing, pattern recognition, computer vision, machine learning, and biometrics.

Keynote Speaker II

Saturday, June 14, 2025 09:55-10:40	Room ID: 895 3784 4640

Prof. Kin-Choong Yow
University of Regina, Canada

Speech Title: Explainable Artificial Intelligence (xAI) for Pattern Recognition

Abstract: Deep learning has recently achieved great success in many visual recognition tasks. However, deep neural networks are often perceived as black-boxes, making their decision less understandable to humans and prohibiting their usage in safety-critical applications. In this talk, we will discuss some of the basic ideas in explainable deep learning and how we can incorporate them into state-of-the-art pattern recognition techniques. We will study the use of a prototypical part network and show how we can use it to improve the reliability of respiratory disease detection from chest x-ray images. We will also examine the application of explainable techniques to state-of-the-art pattern recognition algorithms such as You Only Look Once (YOLO) to detect faulty insulators in transmission lines.

Kin-Choong Yow obtained his B.Eng (Elect) with 1st Class Honours from the National University of Singapore in 1993, and his Ph.D. from Cambridge University, UK in 1998. He joined the University of Regina in September 2018, where he is presently a Professor in the Faculty of Engineering and Applied Science. Prior to joining UofR, he was an Associate Professor in the Gwangju Institute of Science and Technology (GIST), Republic of Korea, (2013-2018), Professor at the Shenzhen Institutes of Advanced Technology (SIAT), P.R. China (2012-2013), and Associate Professor at the Nanyang Technological University (NTU), Singapore (1998-2013). In 1999-2005, he served as the Sub-Dean of Computer Engineering in NTU, and in 2006-2008, he served as the Associate Dean of Admissions in NTU.

Kin-Choong Yow's research interest is in Artificial General Intelligence and Smart Environments. Artificial General Intelligence (AGI) is a higher form of Machine Intelligence (or Artificial Intelligence) where the intelligent agent (or machine) is able to successfully perform any intellectual task that a human being can. Kin-Choong Yow has published over 100 top quality international journal and conference papers, and he has served as reviewer for a number of premier journals and conferences, including the IEEE Wireless Communications and the IEEE Transactions on Education. He has been invited to give presentations at various scientific meetings and workshops, such as ACIRS, in 2018 and 2019; ICSPIC, in 2018; and ICATME, in 2021. He is the Editor-in-Chief of the Journal of Advances in Information Technology (JAIT), a Managing Editor of the International Journal of Information Technology (IntJIT), and a Guest Editor of MDPI Applied Sciences. He is also a member of APEGS and ACM, and a senior member with the IEEE.

His pioneering work in Mobile and Interactive Learning won the HP Philanthropy grant in 2003 for applying Mobile Technologies in a Learning Environment. Only 7 awards were given to the 21 Asia Pacific Countries who were invited, and his project was the only one from Singapore to win it. Also, in 2003, he was one of the only 2 Singaporeans to be awarded participation to the ASEAN Technology Program on Multi Robot Cooperation Development held in KAIST, Korea.

He was the winner of the NTU Excellence in Teaching Award 2005, and he won the Most Popular SCE Year 1 lecturer for 4 consecutive years 2004-2007. He has led numerous student teams to National and International victories such as the IEEE Computer Society International Design Competition (CSIDC) (2001), the Microsoft Imagine Cup (2002, 2003 and 2005), and the Wireless Challenge (2003).

Invited Speaker

Saturday, June 14, 2025 11:00-11:30	Room ID: 895 3784 4640

Dr. Kok Chiang Liang
Newcastle Australia Institute of Higher Education, Singapore

Speech Title: Biomedical Limb Lengthening Implant

Abstract: This presentation outlines the development and optimization of a Biomedical Limb Lengthening Implant with wireless integration, designed to address limb length discrepancies (LLD) affecting over 35% of adults. The implant combines an intramedullary nail with Bluetooth-enabled active feedback, enabling precise control via a patient-centric mobile app. Key innovations include a low-power PCB-integrated system (microcontroller, H-bridge, DC motor) for real-time adjustments and saline/muscle-simulated attenuation tests validating signal reliability (optimal ⩽3m range). Finite Element Analysis (FEA) of thread designs (0.25–0.40mm pitch) identified Titanium Ti-6Al-4V as optimal, minimizing stress concentrations (⩽0.40mm pitch) and displacement under torque (1.1 N • m), outperforming stainless steel. Clinical data analysis (1007 cases) highlighted mechanical failures (36% of complications), guiding design refinements to reduce risks like thread or distraction mechanism failure. Funded by Singapore's NRF (S$200k), future work expands to IoT medical devices (e.g., drug infusion systems), with collaborations spanning orthopedic surgeons (Mount Elizabeth Hospital) and Medot Pte Ltd. This research bridges engineering precision with clinical needs, enhancing patient outcomes in limb reconstruction.

In 2010, Dr Kok graduated with First Class Honours in Bachelor of Electrical & Electronic Engineering from Nanyang Technological University (NTU). His exceptional performance earned him the highly coveted Singapore EDB Integrated Circuit Design PhD Scholarship to pursue his PhD at NTU. In 2014, Dr Kok was awarded his PhD Degree in Electrical & Electronic Engineering in which he delved deeply into power management units, AI, sensors and energy harvesting systems. He also served as an NTU undergraduate tutor and teaching assistant for NTU-TUM Master courses. In 2014, Dr Kok joined the Ministry of Défense, Mindef DSO National Lab as a senior member of technical staff. Here, he spearheaded several state-of-the-art projects, earning acclaim with the prestigious Design Innovation Award (Individual) at the Electronics division level. Dr Kok was invited to be the Adjunct Professor (Faculty Member) at Singapore University of Social Sciences (SUSS), where he teaches electronics courses with passion. In 2021, he was bestowed with the prestigious Gold Medal Award for Teaching Excellence (University level) at SUSS.

In 2020, Dr Kok joins the Newcastle Australia Institute of Higher Education as a lecturer and program coordinator for the Bachelor of Electrical and Electronic Engineering (BEEE). His influence extends far beyond the classroom, as evidenced by his exclusive invitation to the Channel News Asia (CNA) Money Mind programme in May 2021, where he shared his expertise on blockchain technology and sustainable energy solutions. In Nov 2021, he receives the Best Paper award at the 3rd ICESA. Dr Kok also serves as chairman for the STEM Industrial Advisory Board Committee and a committee member for the PEI Exam Board Council. His expertise is sought after on the international stage, with invitations as keynote/plenary speaker and local organising chair for GMASC 2023, MSM 2024, CCCN 2024, ASET 2024, ACEE 2024 and PCDS 2024. Furthermore, he is in the technical program committee for ICET 2024, ITET 2024, ICICDT 2024, TENCON 2024 and RASSE 2024. He is also the chairperson and moderator for WES 2023, session chair for AGBRP 2024, TENCON 2024 and ISCAS 2024. He also serves as the publicity chair for MCSoC 2024 and RASSE 2024. Recently, he also serves as the special session chair and co-trainer for workshop titled "Modern Technologies for Sustainability and Asset Management" in McSOC 2024. In July 2024, he is appointed to the Topical Advisory Panel for MDPI Electronics, Circuit and Signal Processing Section. He also serves as guest editor and reviewer for esteemed Q1/Q2 ranking journals such as MDPI Sensors/Electronics/Applied Sciences, IEEE Access, Circuits, Systems, and Signal Processing and IEEE Transactions on Industrial Electronics. Till date, he has been awarded research funding of more than S$240K in both PI and co-PI capacity. With over 50 publications in Q1/Q2 ranking journals, top conferences, and several book chapters, Chiang Liang's scholarly impact continues to reverberate across the global engineering landscape.

Oral Session

Saturday, June 14, 2025 13:00-15:00	Room ID: 895 3784 4640

Digital Image Detection and Analysis Method Based on Machine Learning

Chairperson:

WP501 13:00-15:15	Contextual Information Induced Attention Network for Infrared Small Target Detection **Junying Li,** University of Electronic Science and Technology of China, China Abstract-Infrared small target detection is a challenging task due to the low contrast, small size and complex background. Despite considerable advancements in accuracy, current methods frequently come with high computational costs. Moreover, single modeling that relies on global features or local details usually cannot take into account the effective fusion of global and local information, which affects the detection accuracy and robustness. To solve the above problems, this paper suggests a contextual information induced attention network (CIANet). The network designs a global semantic interaction module (GSI) to perceive the global contextual information and roughly determine the possible small target area. Then, the local context localization module (LCL) carries out refined processing under the guidance of the coarse localization, reducing the computational complexity while ensuring high accuracy. Finally, fusing feature maps of various scales keeps small target's position and detail information better. Experimental results show that the proposed method outperforms current methods in terms of efficiency and accuracy.
WP502 13:15-13:30	Chatphasia: A Personalized End-to-End System for Aphasia Therapy **Farhan Azmi,** Singapore Institute of Technology, Singapore Abstract-This paper introduces an AI-powered system designed to revolutionize therapy for aphasia patients by enhancing accessibility and personalization. The solution integrates a mobile application for patients to perform word-retrieval tasks and a web dashboard for practitioners to monitor progress and tailor therapy. Key components include a fine-tuned Automatic Speech Recognition (ASR) model optimized for Singaporean aphasic speech and a Large Language Model (LLM)-driven cue generation framework based on hierarchical techniques used in speech therapy. The combination of these components forges a comprehensive tool for speech rehabilitation, demonstrating significant improvements in usability, transcription accuracy, and therapy outcomes.
WP507 13:30-13:45	Complementary horizontal visibility patches: from texture classification to remote sensing image classification **Laifan Pei,** China University of Geosciences (Wuhan), China Abstract-Texture information has been widely used in remote sensing image classification to improve the accuracy of feature recognition. This paper presents new applications of complex network theory and tools to texture classification and remote sensing image classification to recognize texture patterns. The proposed complementary horizontal visibility patches (CHVP) algorithm naturally extends the horizontal visibility graph (HVG) algorithm. The initial model of the complex network is first constructed by taking the image pixel points as network nodes. Then, the initial network model dynamically evolved using the CHVP method to extract network features at various stages of evolution and describe image structural features. Finally, it was tested on standard Kylberg, UMD, and SIRI-WHU data sets, which have general accuracies of 100. 00%, 95.70%, and 70. 5%, respectively, compared to the local binary pattern (LBP), adaptive center pixel selection (ACPS), simple siamese (SimSiam), and local ternary pattern (LTP), among others. This new feature discovery method has important potential implications in texture modeling.

WP517 13:45-14:00	Multi-Robot Reinforcement Learning Navigation Method Based on Credit Assignment **Sitong Qian,** Jilin University, China Abstract-In multi-agent tasks, the credit assignment problem has garnered widespread attention, as an effective credit assignment mechanism can not only promote cooperation among agents but also significantly enhance the efficiency and quality of task completion. Reasonable credit assignment ensures that each agent receives its due rewards, thus motivating them to participate and coordinate more actively, which is crucial for successfully completing complex tasks. To address this issue, we propose an improved algorithm based on the MAPPO algorithm. This algorithm employs a centralized training and decentralized execution framework, incorporates a counterfactual baseline, and uses an enhanced action value network as a centralized critic. By fixing the behaviors of other agents and marginalizing the actions of a single agent, we can more accurately assess each agent's contribution to the overall team performance. This approach not only ensures the fairness of credit assignment but also allows for a granular assessment of each agent's specific performance, achieving a more refined reward mechanism. To validate the effectiveness of our improved algorithm, we conducted experiments in a collaborative robotic navigation task. The results indicate that our method outperforms existing approaches in overall performance.
WP520 14:00-14:15	Deep Learning vs Traditional Methods for Echocardiogram Analysis on Heart disease **Chiang Liang Kok,** University of Newcastle Australia, Singapore Abstract-Heart disease remains a leading cause of mortality worldwide, necessitating accurate and efficient diagnostic tools. Echocardiograms are a primary imaging modality for assessing cardiac function, but their manual interpretation is time-consuming and subject to variability. Recent advancements in deep learning (DL) have shown promise in automating echocardiogram analysis, potentially outperforming traditional image processing and machine learning methods. This study compares the effectiveness of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent architectures, against conventional approaches like speckle tracking, Doppler analysis, and feature-based machine learning in detecting and diagnosing heart disease from echocardiographic data. By referencing to other similar result that uses different type of methods we can see the improvement of a higher accuracy in detection rate in heart disease with deep learning.
WP524 14:15-14:30	Multi dimensional analysis based on historical Olympic data - medal prediction and trend exploration **Haiyan Tian and Bo Zhang,** College of Computer Science, Inner Mongolia University, Hohhot, China Abstract-The analysis of historical Olympic data holds significant importance for research and practical application. This study constructs an Olympic athlete dataset using Kaggle and a self-built dataset. Employing data cleaning and multidimensional data mining, it investigates the relationship between height, weight, age, and medal acquisition. The RoBERTa-CNN-BiLSTM-ATT-CRF model is introduced to effectively extract and categorize entities like athlete and event information, enhancing analytical accuracy. This model shows improvements in P-value, R-value, and F1-value compared to previous entity recognition models, particularly in F1-value. The results indicate the model's capacity to capture the interplay of factors such as athletes' height, weight, and age with medal outcomes. These findings reveal the impact of scientific training and physical conditioning on competitive performance. This study emphasizes the role of physical well-being in sports, aligning with the Olympic principles of 'green, health, and sustainable development. By summarizing factors influencing winning, training a predictive model, and offering medal optimization recommendations, it provides data-driven support for scientific training and holistic health. The sophisticated RoBERTa-CNN-BiLSTM-ATT-CRF model advances the state-of-the-art in entity recognition through its innovative fusion of multi-

	level semantic representation and attention-driven contextualization, ultimately contributing to a more informed and sustainable approach to the Olympic Games.
WP604 14:30-14:45	Context-aware Feature Map Calibration for CNN Generalization Improvement **Min-Yan Tsai,** National Cheng Kung University, Taiwan Abstract-Robust object recognition remains a challenge when visual models rely on background context that co-occurs with target objects. Such dependence can severely degrade performance under distribution shifts. In this work, we examine the degree to which CNNs rely on background cues, and propose a training-time strategy to reduce such reliance. Our method combines Grad-CAM with segmentation masks to quantify the spatial focus of individual feature channels. Channels that consistently activate on background regions are selectively suppressed using an adaptive dropout mechanism, encouraging the model to prioritize object-centric representations while preserving regularization benefits. Experiments under standard and background-altered conditions show improved robustness, with accuracy rising from 82.45% to 89.92% for ResNet-18, 81.76% to 84.80% for ShuffleNet, and 80.42% to 86.96% for VGG16BN.
WP523 14:45-15:00	Comparative Evaluation of Specialized Deep Learning and Large Multimodal Models for Image-based Sea State Recognition on the Beaufort Scale **Marina Ivasic-Kos,** University of Rijeka and Centre for Artificial Intelligence, University of Rijeka, Croatia Abstract-Recent advancements in large language models (LLMs) have demonstrated strong performance across a variety of language-based tasks. With the ability of LLMs to process multimodal inputs, including images, their potential applications are expanding. In this study, we explored this potential by focusing on a specialized task typically addressed by deep learning models. We fine-tuned the ResNet101d, XCIT-small, and Swin Base models to recognize sea states on the Beaufort scale (ranging from 1 to 8) using our dataset of images captured from ocean-going ships. Their performance was then compared to that of the OpenAI 4o model on a newly created test set. After manual calibration, the general LLM produced some encouraging results; however, the specialized deep neural network models demonstrated superior and more consistent performance across all metrics, along with faster real-time inference. The optimized transformer model Swin Base achieved the highest F1 score of 46.62%, while the calibrated GPT-4o model reached 27%. Likewise, the optimized Resnet101d achieved the lowest mean absolute error of 0.739, compared to 0.842 for the calibrated GPT-4o model.

Note

Sponsored by

Sensors and Systems Society of Singapore

NEWCASTLE AUSTRALIA INSTITUTE OF HIGHER EDUCATION | A WHOLLY OWNED ENTITY OF THE UNIVERSITY OF NEWCASTLE

Supported by

Colorado Technical University

VALDOSTA STATE UNIVERSITY

UNIVERSIDADE DE COIMBRA