

2025 International Conference on Multimedia Analysis and Pattern Recognition (MAPR 2025)

**Khanh Hoa, Vietnam
14-15 August 2025**



**IEEE Catalog Number: CFP25MAP-POD
ISBN: 979-8-3315-5467-5**

**Copyright © 2025 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

| | |
|-------------------------|-------------------|
| IEEE Catalog Number: | CFP25MAP-POD |
| ISBN (Print-On-Demand): | 979-8-3315-5467-5 |
| ISBN (Online): | 979-8-3315-5466-8 |
| ISSN: | 2770-6842 |

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

MAPR 2025 Conference Program

Venue: Novotel Hotel, Nha Trang, Vietnam

Dates: August 14-15, 2025

Total Papers: 67 (Oral & Poster Presentations)

Day 1: August 14, 2025

| | |
|--|--|
| 07:45 - 08:15 | Registration |
| 08:15 - 08:30 | Opening and Welcome Addresses Welcome Speech , Prof. Huu-Thanh Nguyen (SEEE-HUST), General Co-Chair Conference Summary , Prof. Duc-Dung Nguyen, PC Co-Chair |
| 08:30 - 09:15 | Keynote Talk 1: Vision Based AI and DX Service Professor Kang-Hyun Jo (University of Ulsan, Korea) <i>Chair: Prof. Hai Vu, HUST</i> |
| 09:15 - 10:15 5 mins/paper presentation | Oral Session 1: Computer Vision for Detection, Recognition and Segmentation (12 papers) <i>Chair: Prof. Thanh-Hai Tran, HUST</i> <ol style="list-style-type: none">98 - A Semi-Supervised Sparsity-Aware Loss Function for Crack Segmentation...111 - Detection of Varroa Mites from Bee Images Using YOLO Architecture...759 - Generative One-shot Camouflage Instance Segmentation...1390 - 3DSMCOS: A 3D Model-Based Synthetic Data Pipeline for Military Camouflaged Object Segmentation with Distractor-Augmented Realism...1993 - Robust Traffic Vehicle Detection under Real-World Conditions with Misclassified Vehicles Minimization and Weighted Box Fusion...25 |

| | |
|---|---|
| | <ol style="list-style-type: none"> 6. 104 - Enhanced Class Incremental Semantic Segmentation with New Classifier Pre-Tuning...31 7. 106 - Eliminating False Positives in Single-Human Parsing via Prompt Filtering and Pose-Aware Segmentation...37 8. 127 - Few-Shot Instance Segmentation: An Exploration in the Frequency Domain for Camouflage Instances...43 9. 72 - Plant Species Identification from Pollen Grain Images Using Fusion of Advanced Deep Neural Networks...49 10. 130 - Enhancing Ship Detection in Remote Sensing: A Data Augmentation Approach Using State-of-the-Art Text-to-Image Diffusion...55 11. 1 - Systematic Evaluation of Deep Learning-based Vehicle Detection Methods...61 12. 22 - ResEViT-Road: An Efficient Model for Road Quality Classification...67 |
| <p>10:15 - 11:30</p> | <p>Coffee Break & Poster Session 1 (24 papers)</p> <p>All papers from Oral Session 1 and Oral Session 2 will be presented as posters.</p> |
| <p>11:30 - 12:30</p> <p>5 mins/paper presentation</p> | <p>Oral Session 2: Vision-Language Models and Multimedia Retrieval (12 papers)</p> <p><i>Chair: Prof. Thi-Lan Le, HUST</i></p> <ol style="list-style-type: none"> 1. 3 - A Multi-Loss Hybrid CNN-ViT Model for Efficient Image Retrieval...73 2. 15 - Att2Search: an Attribute-based Pedestrian Search...79 3. 27 - SATURN: Autoregressive Image Generation Guided by Scene Graphs...85 4. 28 - SAMURAI: Shape-Aware Multimodal Retrieval for 3D Object Identification...91 5. 44 - Movie Character Retrieval with Few Examples...97 6. 46 - UDP-Edit: Union Dual-Prompt Attention for Local Image Editing in Fast-sampling Diffusion Models...103 7. 24 - Integrating Deep Learning And Explainable AI for Interpretable Educational Intervention Analysis...109 8. 86 - Improving Table Structure Recognition Based on Content-Based Post-Processing...115 |

| | |
|---|---|
| | <p>9. 124 - PerSynth: Personalized Character Video Synthesis from Text Prompts...121</p> <p>10. 144 - Scalable Fashion Product Retrieval with Multi-Task Fine-Tuned Vision-Language Models and Real-Time Distributed Architecture...127</p> <p>11. 151 - In Defense of Character-Level Answer Generation Methods for Text-based Visual Question Answering...133</p> <p>12. 107 - Enhancing User Intent Detection in Chatbots through LLM Distillation...139</p> |
| <p>12:30 - 14:00</p> | <p>Lunch Break</p> |
| <p>14:00 - 14:45</p> | <p>Keynote Talk 2: Model Inversion in Deep Neural Networks</p> <p>Professor Ngai_Man(Man) Cheung, Singapore University of Technology and Design, Singapore <i>Chair: Thanh-Duc Ngo, UIT</i></p> |
| <p>14:45 - 15:35 5 mins/paper presentation</p> | <p>Oral Session 3: Tracking and Activity Recognition (10 papers) <i>Chair: Prof. Thi-Lan Le, HUST</i></p> <ol style="list-style-type: none"> 1. 4 - Enhanced MobileNets via Augmentation of Filtering-based Features...145 2. 96 - Enhancing Continuous Student Activity Recognition through Virtual Trajectory and Appearance Matching...151 3. 115 - DSC-SNN: A Depthwise Separable Convolutional Spiking Neural Network for Efficient, Privacy-Preserving Action Recognition from Event-Based Data...157 4. 61 - HAPX-CLIP: Human Activity Recognition with Visual Sequences and Language Prompts...163 5. 35 - Enhancing Multi-Camera People Tracking with Conflict-aware Cosine Tracking and Human Matching Re-assignment...169 6. 37 - UnsMOT: Unified Framework for Unsupervised Multi-Object Tracking with Geometric Topology Guidance...175 7. 103 - Development of an Indoor Position Monitoring System for Rescue Personnel during Emergencies...181 8. 51 - Improving Micro-Expression Recognition with Phase-Aware Temporal Augmentation...187 |

| | |
|---|---|
| | <p>9. 147 - Efficient Sign Language Recognition with Skeleton Data: A Study of Keypoint Selection, Pose Estimators, and GCN Models...193</p> <p>10. 109 - Enhancing Video-Only Sign Language Recognition with Wasserstein Knowledge Distillation on the VSL Dataset...199</p> |
| 15:35 - 16:50 | <p>Coffee Break & Poster Session 2 (20 papers)</p> <p>All papers from Oral Session 3 and Oral Session 4 will be presented as posters.</p> |
| <p>16:50 - 17:40</p> <p>5 mins/paper presentation</p> | <p>Oral Session 4: AI in Healthcare and Medical Imaging (10 papers)</p> <p><i>Chair: Prof. Hai Vu, HUST</i></p> <ol style="list-style-type: none"> 1. 23 - Hybrid approach for Chest X-ray diagnosis with automated MLOps pipeline...205 2. 34 - HySUP: Seam-Aware Hand-Body Fusion for Accurate Whole-Body 3-D Reconstruction from a Single Image...211 3. 36 - OTADiff: Ovarian Tumor-Aware Diffusion Model for Ultrasound Image Augmentation and Detection...216 4. 67 - Ovarian Malignance Risk Prediction from Medical Notes: A Case Study in Vietnam...222 5. 68 - Vision-Language Transformer Framework For Automated Medical Imaging Reporting...228 6. 80 - OSNet-DCN: Integrating Deformable Feature Learning for Lesion Tracking in Endoscopic Videos...234 7. 81 - Towards Robust Cancer Diagnosis: A Hybrid Deep Learning Pipeline for Imbalanced Multi-Label Microscopic Image Analysis...240 8. 102 - Multimodal Deep Learning for ECG Heartbeat Classification with SHAP-Based Interpretability...246 9. 108 - CR4Re: Reformulating Binary COPD Classification Using Contrastive Representations...252 10. 99 - Superpixel-based Graph Neural Network with Hierarchical Topology for Retinal Image Grading...258 |
| 18:30 - 21:00 | <p>Gala Diner</p> <p><i>Chair: Prof. Duc-Dung Nguyen, VAST</i></p> |

Day 2: August 15, 2025

| | |
|--|--|
| 08:00 - 08:30 | Registration |
| 08:30 - 09:15 | Keynote Talk 3: Sensor Fusion Applications Professor Hiệp Q. Luong, (Ghent University, Belgium) <i>Chair: Prof. Thi-Lan Le, HUST</i> |
| 09:15 - 10:10 5 mins/paper presentation | Oral Session 5: Speech, Language, and Multimodal Interaction (11 papers) <i>Chair: Prof. Thanh-Hai Tran, HUST</i> <ol style="list-style-type: none">5 - Vietnamese Automatic Speech Recognition Utilizing Audio and Visual Data...26431 - Adapting WavLM for Vietnamese Speaker Diarization in Real-world Conversations...27063 - Truth or Trick? A Vietnamese Dataset of Adversarial Logic in Geometry...27678 - Vietnamese Emotion Recognition from Voice and Text: A Confidence-Based Approach...28282 - SwahiliVQA: A Dataset for Visual Question Answering in Swahili Language...28833 - Unified Acoustic Representation Learning for Vietnamese Speech Classification Tasks...29492 - Context-Aware Question Answering for Vietnamese University Admissions via Multi-LLM Architecture...300105 - Cross-lingual XLSR-Wav2Vec2-based Spoofing Speech Detection for Vietnamese Speech...306145 - VietX-NLI: A Cross-lingual Natural Language Inference Dataset with Vietnamese as the Source Language...31212 - Can LLMs Play Ô Ăn Quan Game? A Study of Multi-Step Planning and Decision Making...31832 - Analyzing the Correlation and Impact of Speech Evaluation Metrics on Real-World Speaker Verification and Speech Recognition...324 |

| | |
|---|--|
| <p>10:10 - 11:25</p> | <p>Coffee Break & Poster Session 3 (23 papers)</p> <p>All papers from Oral Session 5 and Oral Session 6 will be presented as posters.</p> |
| <p>11:25 - 12:25</p> <p>5 mins/paper presentation</p> | <p>Oral Session 6: Cybersecurity, and Blockchain (12 papers)</p> <p><i>Chair: Prof. Minh-Triet Tran, HCMUS</i></p> <ol style="list-style-type: none"> 1. 38 - PowerGNN: A Source Code Structural And Textual Awareness Approach For Identifying Malicious PowerShell Scripts...330 2. 52 - RAG-SmartVuln: Enhancing Smart Contract Vulnerability Detection via Retrieval-Augmented LLMs...336 3. 53 - Smart Contract Vulnerability Detection using Prompt Engineering with Reasoning Models...342 4. 66 - Multimodal Windows Malware Detection via Hybrid Analysis and Enriched Graphs: Effectiveness and Explainability...348 5. 71 - ChainFLIP: A Unified Framework Integrating Blockchain, Federated Learning, and IPFS for Secure Supply Chain Management...354 6. 112 - Multimodal Fusion for Vulnerability Detection: Integrating Sequence and Graph-Based Analysis with LLM Augmentation...360 7. 122 - LLM-Based Compliance Analysis of Declared Policies and Actual Cookie Usage on Websites...366 8. 140 - Unsupervised Malicious Domain Detection Using Auto-Encoder Models with Reconstruction Error...372 9. 49 - Dropout Attacks on Knowledge Distillation...378 10. 54 - Undermining Trust: How Bit-Flip Attacks Compromise Anomaly-based Network Intrusion Detection Systems...384 11. 77 - Forensic Challenges in Face Manipulated Videos...390 12. 139 - Adversarial Robustness Evaluation of a Vietnamese Handwriting OCR System...396 |
| <p>12:25 - 14:00</p> | <p>Best Paper Awards</p> <p>Closing Ceremony</p> |
| <p>12:45 - 14:00</p> | <p>Farewell Lunch</p> |

Additional Paper

Trusted Proof: Verifying Photographic Authenticity via Multi-Camera Stereo Vision for Insurance and Forensic Applications...402