

2025 International Conference on Content-Based Multimedia Indexing (CBMI 2025)

**Dublin, Ireland
22-24 October 2025**

IEEE Catalog Number: CFP2514C-POD
ISBN: 979-8-3315-5501-6



Copyright © 2025, IEEE

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:	CFP2514C-POD
ISBN (Print-On-Demand):	979-8-3315-5501-6
ISBN (Online):	979-8-3315-5500-9

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

TABLE OF CONTENTS

MI-Cap: A Multi-Modal Interpretable Model for Video Captioning	1
<i>Antoine Hanna-Asaad, Decky Aspandi-Latif, Titus Zaharia</i>	
Accelerating Vector Search at Scale: BAM-ANN with Batch-Aware Memory-Disk Hybrid Indexing	9
<i>M M Mahabubur Rahman, Jelena Tešić</i>	
Evaluating the Recognisability of AI-Generated Familiar Images in a Closed Environment with a Gamified Approach	16
<i>Marc Gallofré Ocaña, Balazs Mosolygo, Bahareh Fatemi</i>	
Hockey2D: A Keypoint-Based Framework for Ice Hockey Rink Localization and Object Mapping	23
<i>Mehdi Houshmand Sarkhoosh, Cise Midoglu, Saeed S. Sabet, Tomas Kupka, Pål Halvorsen</i>	
Vision Projector: Improving Zero-Shot Composed Image Retrieval at Inference	30
<i>Hoang-Bao Le, Allie Tran, Binh T. Nguyen, Liting Zhou, Cathal Gurrin</i>	
Beyond CNNs: Efficient Fine-Tuning of Multi-Modal LLMs for Object Detection on Low-Data Regimes	35
<i>Nirmal Elamon, Rouzbeh Davoudi</i>	
Personalizing Retrieval Using Joint Embeddings; or “the Return of Fluffy”	41
<i>Bruno Korbar, Andrew Zisserman</i>	
U-Cker: Initial Development of an Interactive Video Retrieval System for Novice Users	49
<i>Kazuya Ueki, Ryo Muto, Takuya Wada, Ryota Akaba, Genesis Faith Fernandez</i>	
Multi-modal Context Reranking for Lifelog Question Answering	54
<i>Quang-Linh Tran, Ly-Duyen Tran, Binh Nguyen, Gareth J.F. Jones, Cathal Gurrin</i>	
GenFlow: Interactive Modular System for Image Generation	62
<i>Duc-Hung Nguyen, Huu-Phuc Huynh, Minh-Triet Tran, Trung-Nghia Le</i>	
Melanoma Segmentation with SAM-Like Models: Assessing the Influence and Limits of Bounding Box Input	69
<i>Nicolas Martin, Philippe Mulhem, Jean-Pierre Chevallet</i>	
Fusion of Global and Local Features with Multi-Inverted Indices for Image Retrieval	76
<i>Li Weng, Xizhe Wang, Qianneng Wang, Bingya Wu</i>	
VoiceVision: AI-Powered Speaker-Aware Cropping and Content Indexing for Multi-Speaker Videos	84
<i>Mehdi Houshmand Sarkhoosh, Cise Midoglu, Saeed S. Sabet, Tomas Kupka, Pål Halvorsen</i>	
Exploring the Effect of Size, Architecture and Fine-Tuning Hyperparameters on Large Visual-Language Model Adaptation for Video Memorability Prediction	89
<i>David Luna-García, Iván Martín-Fernández, Sergio Esteban-Romero, Manuel Gil-Martín Fernando Fernández-Martínez</i>	
Binned-KMeans: Extracting Memorable, Explainable, Weighted and Sorted Colour Palettes from Videos	96
<i>Andreea Pocol, Lesley Istead, Craig Kaplan</i>	

Predicting Moral Values in Lyrics Through Audio	101
<i>Charalampos Saitis, Ben Heyderman, Vjosa Preniqi, K yriaki Kalimeri †, Johan Pauwels</i>	
Historical Postcard Stamp Content Understanding	108
<i>Matthieu Pelingre, Salvatore Tabbone</i>	
Text-Oriented Image Query Representation for Zero-Shot Composed Image Retrieval	115
<i>Pavan Kartheek Rachabathuni, Andrea Ciamarra, Roberto Caldelli, Marco Bertini</i>	
Explanatory Interactive Machine Learning for Bias Mitigation in Visual Gender Classification	122
<i>Nathanya Queby Satriani, Djordje Slijepčević, Markus Schedl, Matthias Zeppelzauer</i>	
Breaking the 2D Dependency: What Limits 3D-Only Open-Vocabulary Scene Understanding	130
<i>Domenico D'Orsi, Fabio Carrara, Fabrizio Falchi, Nicola Tonello</i>	
Label-Efficient Skeleton-Based Recognition with Stable Graph Convnets	135
<i>Hichem Sahbi</i>	
Masked Spikformer: Gaussian based and Random Spike Masking for Energy-Efficient Spiking Transformers	143
<i>Oumaima Marsi, Sebastien Ambellouis, José Mennesson, Cyril Meurie, Anthony Fleury, Charles Tatkeu</i>	
ReViewQwen: An Explainable Vision-Language Model for Discrepancy Detection in Multimodal E-Commerce Reviews	150
<i>Sandeep Kalari, Mohan Krishna Sunkara, Dominik Soós, Vikas Ashok, Ravi Mukkamala</i>	
TREB: Temporal Refinement of Egocentric Body Pose	157
<i>Bruno Henriques, Benjamin Allaert, Nicolas Sutton-Charani, Pierre Slangen, Jean-Philippe Vandeboer</i>	
A New Pipeline for Extracting and Clustering Sub-Images from Unannotated Complex Image Datasets	163
<i>Chafic Abou Akar, Christian Beddawi, Marc Kamradt, Abdallah Makhoul</i>	
Semi-Supervised Approach to Detect Human Discontent from Real-Life Behaviour Data	170
<i>Elena Vildjiounaite, Oulu Finland, Vesa Kyllonen, Johanna Kallio, Pauli Rasanen</i>	
A Mixed-Methods Investigation of XR Security Warnings—Lessons Learned	177
<i>Junyi Zou, Riccardo Bovo, Ali Hamza, Georgios Loukas</i>	
Anonymisation of Visual Lifelogs using Diffusion Models and Large Language Models	185
<i>Minh-Quang Le, Graham Healy, Liting Zhou, Cathal Gurrin</i>	
An Experimental Study on Generating Plausible Textual Explanations for Video Summarization*	192
<i>Thomas Eleftheriadis, Evlampios Apostolidis, Vasileios Mezaris</i>	
First-Person Human Sensing for Upper Limb Neuroprosthesis Control: 6D Pose Estimation of Objects to Grasp	200
<i>Ander Etxezarreta, Jenny Benois-Pineau, Renaud Péteri, Lucas Bardisbanian, Aymar De Rugby</i>	
NN Watermarking for Face Segmentation Task	206
<i>Carl De Sousa Trias, Mihai Mitrea</i>	
A Comparative Study of Conversational and Conventional Search Methods for Image Retrieval	214
<i>Anastasiia Potiagalova, Joemon J. Jose, Benjamin R. Cowan, Gareth J. F. Jones</i>	
SoccerChat: Integrating Multimodal Data for Enhanced Soccer Game Understanding	221
<i>Sushant Gautam, Cise Midoglu, Vajira L. Thambawita, Michael A. Riegler, Pal Halvorsen, Mubarak Shah</i>	

Zero-Shot Vision-Language Model for Event Detection in Smart Surveillance	229
<i>Younes Kebour, Smail Niar, Nacim Ihaddadene, Abdelghani Bekrar, Hammouda Elbez</i>	
DSI-3D: Differentiable Search Index for Point Clouds Retrieval	237
<i>Chahine-Nicolas Zede, Laurent Caraffa, Valerie Gouet-Brunet</i>	
BandNaviHD: Band-Member Backtrack Interface Based on Member History Information	244
<i>Masatoshi Hamanaka</i>	
Understanding Indoor Context in an Office Environment: An Empirical Study on Air Stuffiness Perception	248
<i>Johanna Kallio, Jussi Liikka, Satu-Marja Mäkelä, Atte Kinnula, Elena Vildjiounaite</i>	
ELIP: Enhanced Visual-Language Foundation Models for Image Retrieval	254
<i>Guanqi Zhan, Yuanpei Liu, Kai Han, Weidi Xie, Andrew Zisserman</i>	
MMMS: Multi-Modal Multi-Surface Interactive Segmentation	262
<i>Robin Schön, Julian Lorenz, Katja Ludwig, Daniel Kienle, Rainer Lienhart</i>	
FPN-Based Multi-Scale Feature Fusion for Robust 3D Pedestrian Detection in Crowded Scenes	270
<i>Kiyotaka Matsue, Kenta Umene, Nghia Dao, Hieu Nguyen, Manh Phan</i>	
GeMix: Conditional GAN-Based Mixup for Improved Medical Image Augmentation	277
<i>Hugo Carlesso, Maria Eliza Patulea, Moncef Garouani, Radu Tudor Ionescu, Josiane Mothe</i>	
Lip Reading Across Languages: A Cross-Modal Framework Leveraging Foundation Models	284
<i>Ruxandra TAPU, Bogdan MOCANU</i>	
Robust Multimedia Verification of Cheapfakes and Deepfakes via External Context Leveraging	291
<i>Minh-Nhat Nguyen, Trong-Nghia Tran, Minh-Triet Tran, Duc-Tien Dang-Nguyen, Trong-Le Do</i>	
Probabilistic Fusion Model for Multi-Label Media Content Classification	299
<i>Javier Carreno, Khuong An Nguyen, Zhiyuan Luo, Andrew Fish</i>	
Novice-Friendly Video Retrieval in Mixed Reality with VitriVR- VR	306
<i>Florian Spiess, Heiko Schuldt</i>	
TSalV360: A Method and Dataset for Text-driven Saliency Detection in 360-Degrees Videos*	309
<i>Ioannis Kontostathis, Evlampios Apostolidis, Vasileios Mezaris</i>	
Toward an Energy-Efficient and Explainable Neural Network Architecture for Detection of Breast Cancer in Mammography	317
<i>Alireza Siyavashi, Christian Herglotz</i>	
Dual-Objective Adversarial Disentanglement for Protecting Speech Data used for Diagnosing Parkinson's Disease	324
<i>Mehtab Ur Rahman, Martha Larson, Louis ten Bosch, Cristian Tejedor-García</i>	
Media Search: A Multi-Stage Image Retrieval Framework with Enriched Image Captioning	330
<i>Ayşe Vildan Nurdağ, Mete Mert Birdal, Yusuf Yazıcı, Barış Özcan, Erkut Arıcan</i>	
Rethinking Wine Tasting for Chinese Consumers: A Service Design Approach Enhanced by Multimodal Personalization	336
<i>Xinyang Shan, Yuanyuan Xu, Tian Xia, Yin-Shan Lin</i>	

Toward Content-Based Indexing and Retrieval of Head and Neck CT With Abscess Segmentation	341
<i>Thao Thi Phuong Dao, Tan-Cong Nguyen, Trong-Le Do, Truong Hoang Viet, Nguyen Chi Thanh</i>	
<i>Huynh Nguyen Thuan, Do Vo Cong Nguyen, Minh-Khoi Pham, Mai-Khiem Tran, Viet-Tham Huynh</i>	
<i>Trong-Thuan Nguyen, Trung-Nghia Leo, Vo Thanh Toan, Tam V. Nguyen, Minh-Triet Tran, Thanh Dinh Le</i>	
Enhancing Vision-Language Model Pre-Training with Image-Text Pair Pruning Based on Word Frequency	349
<i>Mingliang Liang, Martha Larson</i>	
Facilitating Interactive Image Labelling Using Fine-Tuned SAM2	356
<i>Hermann Fürntratt, Werner Bailer</i>	
Examining Performance Disparities Between Expert and Novice Users in Interactive Video Retrieval	361
<i>Omar Shahbaz Khan, Ujjwal Sharma, Stevan Rudinac, Björn Þór Jónsson</i>	
Mitigating Shortcut Learning in Online Action Detection and Anticipation via Cross-Modal Semantic Alignment	365
<i>Sensen Wang, Yuehu Liu, Chi Zhang</i>	
Does CLIP Perceive Art the Same Way We Do?	372
<i>Andrea Asperti, Leonardo Dessi, Maria Chiara Tonetti, Nico Wu</i>	
A Survey of Information Disorder on Video-Sharing Platforms	380
<i>Meiyu Li, Wei Ai, Naemul Hassan</i>	
TrueEar: A Lightweight and Accurate Fake Voice Detector for Mobile Devices	390
<i>Cameron Baird, Ke Li, Dan Lin</i>	
MultiHuSE: A Multimodal Dataset for Humour Styles and Emotions	398
<i>Mary Ogbuka Kenneth, Foad Khosmood, Abbas Edalat</i>	
EcoStream: A Resource Utilization and Power Consumption Dataset in Multimedia Streaming for Sustainability Analysis	405
<i>Tariq Al Shoura, Reza Razavi, Mohammad Moshirpour</i>	
MSS: A Multilingual Spoofed Speech Dataset with Code-Switching for Anti-Spoofing Measures	413
<i>Muhammad Hamza, Hafsa Ilyas, Junaid Mir, Ali Javed, Muhammad Haroon Yousaf, Ahmed Zoha</i>	
AgriPotential: A Novel Multi-Spectral and Multi-Temporal Remote Sensing Dataset for Agricultural Potentials	420
<i>Mohammad El Sakka, Caroline De Pourtales, Lotfi Chaari, Josiane Mothe</i>	
Dialogue-AV: A Dialogue-Attended Audiovisual Dataset	427
<i>Luís Vilaca, Paula Viana, Yi Yu</i>	
Music4All A+A: A Multimodal Dataset for Music Information Retrieval Tasks	435
<i>Jonas Geiger, Marta Moscati, Shah Nawaz, Markus Schedl</i>	
SeqBench: Benchmarking Sequential Narrative Generation in Text-to-Video Models	442
<i>Zhengxu Tang, Zizheng Wang, Luning Wang, Zitao Shuai, Chenhao Zhang, Siyu Qian, Yirui Wu</i>	
<i>Bohao Wang, Haosong Rao, Zhenyu Yang, Chenwei Wu</i>	
GTR: General Handwritten Lines Text Recognition Dataset	449
<i>Xu Ji, Haizhao Sun, Yu Ning, Ming Wu, Chuang Zhang</i>	

Towards Graph-Based Federated Learning: ModelNet - A ResNet-based Model Classification Dataset	456
<i>Abhisek Ray, Lukas Esterle</i>	
MERCI: A Multimodal Dataset for Personalised and Emotionally-Aware Dialogues	463
<i>Mohammed Althubyani, Zhijin Meng, Shengyuan Xie, Francisco Cruz, Imran Razzak, Mukesh Prasad</i>	
<i>Eduardo B. Sandoval, Baki Kocaballi</i>	