21st International Conference on Computer Vision Theory and Applications (VISAPP 2026)

Volume 2

Marbella, Spain
9-11 March 2026

ISBN: 979-8-3313-3352-2

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2026) by SCITEPRESS – Science and Technology Publications, Lda.
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

For permission requests, please contact SCITEPRESS – Science and Technology Publications, Lda. at the address below.

SCITEPRESS – Science and Technology Publications, Lda.
Avenida de S. Francisco Xavier, Lote 7 Cv. C,
2900-616 Setúbal, Portugal

Phone: +351 265 520 185
Fax: +351 265520 186

info@scitepress.org

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
Email: curran@proceedings.com
Web: www.proceedings.com

CONTENTS

INVITED SPEAKERS

KEYNOTE SPEAKERS

Rethinking Multimodal AI Models: Beyond Accuracy, Towards Trust 3
Elisa Ricci

Self-Supervised Learning Is Dead, Long Live Self-Supervised Learning (in the Age of MLLMs) 5
Yuki Asano

Seeing, Speaking, and Reasoning in a Visual World 7
Cees G. Snoek

RECOGNITION & DETECTION

FULL PAPERS

Toward Automated Bed Safety: Comparative Study of Vision Algorithms for Bed-Rail State Detection 13
Navid Aslankhani Khameneh, Michela Farenzena and Marco Carletti

Simplify Your Fusion: Reducing Complexity for Multimodal Sensor Fusion 22
Thorsten Herd, Philipp Heidenreich and Christoph Stiller

Stylized Synthetic Augmentation Further Improves Corruption Robustness 31
Georg Siedel, Rojan Regmi, Abhirami Anand, Weijia Shao, Silvia Vock and Andrey Morozov

G2TM: Single-Module Graph-Guided Token Merging for Efficient Semantic Segmentation 43
Victor Bercy, Martyna Poreba, Michal Szczepanski and Samia Bouchafa

Group Emotion Recognition Explained! A Human-Centric Approach Using Multi-Modal VGER Dataset and Framework 55
Harinadh Sivaramakrishna Patamsetti, Kamakshya Prasad Nayak, Kamalakar Vijay Thakare, Debi Prosad Dogra, Heeseung Choi, Haesol Park and Ig-Jae Kim

UMergeNet: Exploring Lightweight Mechanisms for High-Performance Semantic Segmentation 67
Alan Klinger Sousa Alves and Bruno Motta de Carvalho

ViQAgent: Zero-Shot Video Question Answering via Agent with Open-Vocabulary Grounding Validation 79
Tony Montes and Fernando Lozano

Multi-Objective Quantum-Inspired Firefly Algorithm for Medical Image Segmentation 91
Alokeparna Choudhury, Sourav Samanta, Sanjoy Pratihar and Oishila Bandyopadhyay

Human-in-the-Loop Refinement of Zero-Shot Object Detection for Domain-Specific Artwork Datasets 103
Alex Alonso Dalbøl, Christofer Meinecke and Stefan Jänicke

Spatio-Temporal Attention for Consistent Video Semantic Segmentation in Automated Driving 117
Serin Varghese, Kevin Roß, Fabian Hüger and Kira Maag

AVNet: Cross-Spectral Attention-Vision Model for Camouflaged Object Detection in Ecological Conservation 128
Henry O. Velesaca, Andrea Mero, Rafael E. Rivadeneira, Guillermo A. Castillo and Angel D. Sappa

A Simple and Lightweight Model for Person Re-Identification 138
Stuart A. Montes, Edward Cayllahua and Rensso Mora-Colque

Real-Time 6DoF Pallet Pose Estimation with Monocular Metric Depth 149
Azuma Miura, Hideaki Uchiyama, Masahiro Yamaguchi, Natsuki Kai, Takahiro Shiroshima and Hideo Saito

Geometric Skeletal Distance Learning for Self-Supervised Sign Language Recognition 162
Evangelos G. Sartinas, Dimitrios Kosmopoulos, Emmanouil Z. Psarakis, Kostas Blekos, Bikram Kumar De and Vangelis Metsis

SHORT PAPERS

HyperNut: Hyper Spectral Dataset of Nuts for Unsupervised Defect Detection and Segmentation 177
Afshin Dini, Farnaz Delirie and Esa Rahtu

Trade-Offs of Contrast Enhancement and Denoising for Low-Resolution Images: An Empirical Study on DIV2K 185
Morales García Emmanuel, Domínguez Isidro Saúl, Rojano Cáceres José Rafael and Cruz López Cecilia

Robust Cell Segmentation in Urine Cytology Images for Bladder Cancer Diagnosis 193
Mariana L. Teixeira, Hugo S. Oliveira, Raquel L. Monteiro, Daniela Ferreira, Tania Pereira, Raphaël F. Canadas and Hélder P. Oliveira

Domain-Specific Synthetic Data Generation for Person Re-Identification in Public Transport 201
Perry op 'T. Landt, Ulrich Krispel and Torsten Ullrich

Order-Level Arthropod Detection Using Deep Learning: Addressing Scale Variability through Synthetic Data 210
Chenchang Liu, Svetlana Ionova, Patrick Mäder and Marco Seeland

Enhancing Temporal Stability in Small Object Detection: A Post-Processing Approach for YOLOv8 218
Junyu Xiao and Shohei Yokoyama

A Parametric 3D Bee Model for Scalable Synthetic Data Generation in Animal Behavior Studies 226
Christian Matos Rivera, Remi Megret and David Flores

High Semantic Features for the Continual Learning of Complex Emotions: A Lightweight Solution 234
Thibault Geoffroy, Gauthier Gerspacher and Lionel Prevost

INFORM: A Food Monitoring and Tracking System for Sustainable Healthcare Facilities 243
James Rainey, John Wannan, Douglas MacLachlan, Boguslaw Obara and Deepayan Bhowmik

An Automated Deep Learning Pipeline for Panel Segmentation, Defect Detection, and 3D Reconstruction in Impact Test Analysis 251
Uğur Can Karaca and Toygar Akgün

Estimating Person Positions Using a Camera and Wireless Devices in a Space with Temporary Shielding 259
Yuji Makimura, Daichi Hayashi, Teppei Kobiki, Shouki Sakemoto and Masashi Nishiyama

Synthetic Data in the Context of Automotive In-Cabin: A Review 267
Julia Frick, Patrick Laufer, Roman Seidel and Gangolf Hirtz

Rendered vs AI-Generated vs Real Data: Training Industrial Object Detection Models with Multi-Source Data 276
Chafic Abou Akar, Hadi Koubeissy, Joe Khalil, Jimmy Tekli, Marc Kamradt and Abdallah Makhoul

From Label Error Detection to Correction: A Modular Framework and Benchmark for Object Detection Datasets 288
Sarina Penquitt, Jonathan Klees, Rinor Cakaj, Daniel Kondermann, Matthias Rottmann and Lars Schmarje

When Surgery Meets the Unknown: Uncertainty-Aware Open-Set Recognition for Surgery Phase Classification 300
Stefan Geyer, Vicky Kalogeiton and Alina Roitberg

Microstructural Classification of Non-Oriented Electrical Steels: A Machine Learning and Computer Vision-Based Approach
Leonardo Adriano Vasconcelos de Oliveira, José Daniel de Alencar Santos, Francisco Nélio Costa Freitas, Pedro Pedrosa Rebouças Filho, Hamilton Ferreira Gomes de Abreu and Luis Flávio Gaspar Herculano 308

ReAL-YOLO: Reinforcement-Driven Active Learning for Training YOLO Object Detectors 316
Gabriel Souto Ferrante and Priscila Tiemi Maeda Saito

ASRDNet: A New Image-Segmentation Neural Network Model for Detecting Asian Rust on Soybean Leaves 324
Paulo Henrique Bueno Lopes and Leandro A. F. Fernandes

Class-Based Adaptive Training for Facial Expression Recognition Using a Deep Convolutional Network
Brian Luís Coimbra Maia, Lucas Silva Santana, Gabriel Rezende da Silva, Mathews Edwirds Gomes Almeida, Paulo Victor de Magalhães Rozatto, Luiz Maurílio da Silva Maciel, Saulo Moraes Villela, Marcelo Bernardes Vieira and Bruno Campos de Carvalho 332

Computer Vision-Assisted Literacy: Recognizing Children's Handwritten Words
Simone Bello Kaminski Aires, Maria Eduarda Guedes Pinto Gianisella, Simone Nasser Matos, Marcos Vinicius Santos Passos and Thales Janisch Santos 341

Palm Trees Meets Deep Learning: A Real-Time Detection of Boufaroua with YOLOv8n and Embedded Optics 351
Amal Ben Amor, Sonia Mosbah and Jihene Hlel

Long-Tailed Species Recognition in the NACTI Wildlife Dataset 357
Zehua Liu and Tilo Burghardt

Retail Shelf Monitoring Using Deep Hough Transform and Object Detection
Luiz Fernando Merli de Oliveira Sementille, Marcos Cleison Silva Santana, Douglas Rodrigues, Danilo Samuel Jodas, Kelton Augusto Pontara da Costa, Fabio Luiz de Oliveira and João Paulo Papa 366

Improving Underwater Fish Detection and Tracking via Temporal Modeling and Deformable Convolutions 374
Josep Sánchez, Jose-Luis Lisani and Antoni Bennasar Garau

Hepatic Vessel Segmentation: A Narrative Review 382
Nassim Kaddouri, Afifa Dahmane and Hadja Faiza Haned Khellaf

Hazard-Aware Duration Prior and F1-Adaptive Loss for Temporal Action Segmentation 390
Hinako Mitsuoka and Kazuhiro Hotta

Kinship Verification with Custom Sampling and Hard Contrastive Loss 398
Warley Barbosa and Tiago Vieira

Improved MambdaBDA Framework for Robust Building Damage Assessment across Disaster Domains 406
Alp Eren Gençoğlu and Hazım Kemal Ekenel

Visibility-Gated ConvGRU for Robust Multi-View Pedestrian Detection and Tracking in BEV Space 415
Taigo Sakai, Takeshi Nakamura, Hiroshi Shimizu and Kazuhiro Hotta

ASO PatchCore: Memory-Efficient and Fast Anomaly Detection via Automatic Sampling Optimization 423
Satoshi Kamiya, Kota Yamashita and Kazuhiro Hotta

Gabor-Guided Tiny Bird Detection for Drone-Based Agricultural Bird Deterrence 431
Mathijs Lens and Toon Goedemé

Modeling Student Engagement in the Wild: Analysis from Classroom Video 439
Uroš Petković, Jonas Frenkel, Rebecca Lazarides and Olaf Hellwich

Billboard in Focus: Estimating Driver Gaze Duration from a Single Image 447
Carlos Pizarroso, Zuzana Berger Haladová, Zuzana Černeková and Viktor Kocur

Classification of Normal versus Leukemic Cells with Swin Transformer and Balanced Data Augmentation 454
Douglas Costa Braga, Albert de Jesus Souza, Samuel Bastos Borges Pinho and Daniel Oliveira Dantas

SignIT: A Comprehensive Dataset and Multimodal Analysis for Italian Sign Language Recognition 461
Alessia Micieli, Giovanni Maria Farinella and Francesco Ragusa

Adversarial Robustness of Proxy-Based Metric Learning Models 469
Marcin Maciąg and Grzegorz Sarwas

EventAction: Vision Mamba-Based Event-Driven Action Recognition 477
Qingyu Wang, Xingzhen Song, ChungSheng Chang, Feiyu Ge, Eisuke Sato and Masato Tsukada

Generalizable Hyperparameter Optimization for Federated Learning on Non-IID Cancer Images 485
Elisa Gonçalves Ribeiro, Rodrigo Moreira, Larissa Ferreira Rodrigues Moreira and André Ricardo Backes

An Ensemble Approach to Climate Misinformation Detection 493
Lei Lei, Marc Lalonde, Hamed Ghodrati and Azur Handan

Efficient Transformer-Based Spatio-Temporal Action Recognition for Industrial Safety Surveillance 500
Ruslan Zaripov, Anastasiya Shpileva, Maksim Koltakov, Georgii Petrov, Viacheslav Shalamov and Valeria Efimova

A Deterministic Edge-AI System for Early Wildfire Smoke Detection: From Lightweight Neural Models to Operationally Reliable Surveillance 509
Damian Kmiecik and Adrian Dziembowski

IoMB: Evaluating Object Detectors on Occluded and Imbalanced Seabird Populations 518
Charis Hanna, Kasim Terzić, Mark James and Karen Spencer

Comparative Evaluation of Vision Transformer Architectures for Video-Based Weather Intensity Recognition 526
Yuma Kokubu and Tomokazu Ishikawa

LOW-LEVEL VISION & COMPUTATIONAL IMAGING

FULL PAPERS

Defense that Attacks: How Robust Models Become Better Attackers 539
Mohamed Awad, Mahmoud Mohamed and Walid Gomaa

End-to-End Optimization of Polarimetric Measurement and Material Classifier 550
Ryota Maeda, Naoki Arikawa, Yutaka No and Shinsaku Hiura

Training-Free Plant Leaf Disease Severity Estimation Using Fuzzy C-Means Clustering and Reference Palette Validation 560
Om Chatterjee, Prachi Das, Amiya Kumar Bhowmik, Avighyan Chakraborty, Jacob Tauro and Sanjoy Pratihar

Diffusion-Based HDR Reconstruction from Mosaiced Exposure Images 572
Seeha Lee, Dongyoung Choi and Min H. Kim

Solving Large Square Jigsaw Puzzles with Quasi-Linear Candidate Selection Using DC-L1 Norm 584
Him Kafle and Amit Banerjee

Horizontally-Aware Deep Network for Seismic Multiple Attenuation 595
José Ribamar Durand Rodrigues Júnior, Paulo Ivson Netto Santos, Aristófanes Corrêa Silva, Francisco de Assis Silva Neto, Carlos Rodriguez Suarez and Deane Roehl

Adaptive IG-ODAM: Efficient Attribution Maps for Object Detection via Spatial-Guided Sampling 606
Yuma Nakai, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi

Efficient Multi-Temporal Building Change Detection with Reduced-Rank Linear Attention 615
Ricardo A. Acuña-Villogas, Germain García-Zanabria, Cristian Lopez and Rensso Mora-Colque

SHORT PAPERS

Quality of Brain Tumour Detection in Hyperspectral Imaging Based on Ground Truth Representation 627
Martin Rydlo, Max Verbers, Carlos Vega, Raquel Leon, Himar Fabelo, Gustav Liu Burström, Alfonso Lagares, Jesus Morera Molina, Gustavo M. Callico, Francesca Manni and Svitlana Zinger

GazeUnconstrained: A Multimodal Dataset for Visual Attention and Gaze Estimation in Natural Video Viewing 635
Wenjuan Zhou, Haibin Cai, Baihua Li and Qinggang Meng

Computing a Characteristic Orientation for Rotation-Independent Image Analysis 644
Cristian Valero-Abundio, Emilio Sansano-Sansano, Raúl Montoliu and Marina Martínez García

Supervised Deep Feature-Based Industrial Defect Detection in Optical Lenses with Minimal Data 652
Tarek Zenati

Fusing Thermal and Event Data for Visible Spectrum Image Reconstruction 661
Simone Melcarne and Jean-Luc Dugelay

Centrality-Driven Prufer-Code Based Graph Encoding to Capture Structural Relationships 670
Atashi Saha and Sanjoy Pratihar

Similarity Analysis of AI-Generated Images from Spanish and English Prompts via Feature Extraction 678
Paulina Morillo, Christopher Morales, Christian Castro and Diego Vallejo-Huanga

Improving Video Object Detection Performance in Rainy Conditions Using Edge Preserving Image Deraining 687
Sibani Panigrahi, Debi Prosad Dogra and Jothi Ramalingam

Depth Estimation in Scattering Media due to Synchronization Delay in Projector Camera Systems 695
Ryosuke Komatsu, Masaya Oishi, Takafumi Iwaguchi, Hiroshi Kawasaki and Hiroyuki Kubo

Event Camera Deraining Using Transformer Networks with Time-Space Attention 703
Pin-Yuan Yang and Shih-Chieh Chang

Dance Style Classification Using Laban-Inspired and Frequency-Domain Motion Features 715
Ben Hamscher, Arnold Brosch, Nicolas Binninger, Maksymilian Jan Dejna and Kira Maag

CP HDR: A Feature Point Detection and Description Library for LDR and HDR Images 724
Artur Santos Nascimento, Daniel Oliveira Dantas, Valter Guilherme Silva de Souza and Beatriz Trinchão Andrade

Self-Supervised Real-World Image Denoising with Noise-Level-Aware Dynamic Receptive Fields 735
Yoichi Furukawa, Takahiro Maruyama and Kazuhiro Hotta

No Mountain, no Building, no Cue? Synthetic Data Generation of Digital Surface Models and Their Application to Visual Geo-Localization 743
Nicolai Skutsch, Olaf Hellwich and Frank Fuchs-Kittowski

Video Denoising Still Needs Quality Benchmarks 752
Sofiia Dorogova, Amir Shamsutdinov, Aleksei Khalin and Egor Ershov

Edge Bundling with Divergence and Convergence 760
Naoki Hashimoto, Ian Gomasaki and Ryosuke Saga

Few-Shot Data Sampling Strategies for Insect Pest Maturity Classification with YOLO Family Models 768
Kelly Abreu, Aclecio Costa and Dibio Borges

Compensating Light Source Defocus for Per-Pixel Surface Roughness Measurement 776
Ryohei Ohmori, Michitaka Yoshida, Ryo Kawahara and Takahiro Okabe

Unveiling AI-Manipulated Medical Images: Detecting and Localizing Tampered Areas 786
T. Ashlesha, Vignesh, S. Varun, Vats Shubhangi and Narayan Surabhi

SynShapes: A Synthetic Image-Annotation Dataset for Edge Detection 794
Guillermo A. Castillo, Xavier Soria and Angel D. Sappa

Pixel-Driven Image Representation through Rule-Based Cellular Automata Dynamics 803
Lorenzo C. Maia, Priscila T. M. Saito and Pedro H. Bugatti

AUTHOR INDEX 811