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

Volume 1

Marbella, Spain
9-11 March 2026

ISBN: 979-8-3313-3351-5

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

FOUNDATIONS & REPRESENTATION LEARNING

FULL PAPERS

GMAR++: Efficient Gradient Enhanced Attention Rollout 13
Sohambhai Joita and Tatyana Ivanovska

PoET: Lightweight Pose Encoder Transformer for Online Sign Language Recognition 19
Julie Lascar, Jules Françoise, Michèle Gouiffès, Annelies Braffort and Diandra Fabre

NeRF-Seg3D: Integrating Semantic Segmentation, 3D Reconstruction, and Camera Geometry for Enhanced Rust Analysis on Oil & Gas Platforms 29
Jose L. Huillca, Raphael dos S. Evangelista, Christian Erik Condori Mamani, Horácio Brescia Macêdo Henriques, Caio Marcus Monteiro de Oliveira, João Vitor Mendonça de Moraes, Robinson Luiz Souza Garcia, Michelle Soares Pereira Facina, Maikon Bressani, Eduardo C. Vasconcellos, Lucas Bertelli Martins, Marcos Amaral de do Almeida, Raul Queiroz Feitosa, Patrick Nigri Happ, Esteban Walter Gonzales Clua and Leandro A. F. Fernandes

Anomaly Detection Using a Lite PatchCore for Mobile Robotic Industrial Application 41
Lelio Lardon, H.Guis Vincente and Eric Busvelle

A Conformal Loss Correction Approach for Label-Noise Learning 52
Marcos Mota, Helio Pedrini and André Santanchè

Domain-Aware Diffusion for Synthetic-to-Real Data Augmentation 63
Salahidine Lemaachi, Anaïs Druart and Nicolas Winckler

Realistic Evaluation of Test-Time Adaptation: Unsupervised Model Selection 75
Sebastian Cygert, Damian Sójka, Tomasz Trzciński and Bartłomiej Twardowski

Multimodal Embedding for Scientific Image Caption Generation 87
Jose L. Huillca and Leandro A. F. Fernandes

Self@x -Paced Percentile Distillation for Fast and Accurate LiDAR Person Detection 99
Avesalon Razvan Marian and Miron Radu

ProxiVideoFriends: Revisiting Proxemics through Temporal and Social Reasoning 107
Isabel Jiménez-Velasco, Rafael Muñoz-Salinas, Vicky Kalogeiton and Manuel J. Marín-Jiménez

Contrastive Self-Supervised Learning with Distance-Guided Positive and Negative Pair Selection 117
Bruno A. Shimura, Priscila T. M. Saito and Pedro H. Bugatti

Diversity by Chance: Rethinking the Need for Determinantal Point Processes in Active Learning 126
Brecht Deprez, Michiel Vandecappelle, Steven Lauwereins and Toon Goedemé

Unveiling Degradations with Deep Learning Models in Face Recognition Systems 136
Leandro Dias Carneiro and Flavio de Barros Vidal

Exploiting Test-Time Augmentation in Federated Learning for Brain Tumor MRI Classification 148
Thamara Leandra de Deus Melo, Rodrigo Moreira, Larissa Ferreira Rodrigues Moreira and André R. Backes

From Graphs to Images: Non-Parametric PPI Context Integration for Vision-Based Protein Function Prediction 157
Yeonatan Mauhnoom, Gabriel Bianchin de Oliveira, Helio Pedrini and Zanoni Dias

EMTA: End-to-End Multi-Task Audio-Driven Talking-Head Animation 167
Laxmi Narayen Nagarajan Venkatesan, Divyam Choudhary, Sai Madhavan G, Subhajeet Lahiri, Rahulraj B. R, Rittik Panda, Dinesh Babu Jayagopi, Raj Tumuluri and Magnus Revang

Automatic Classification of Whole-Body PET Scans Using Deep Features 179
Celso Luiz Silva Soares Filho, Darlan Bruno Pontes Quintanilha, Alison Corrêa Mendes, Anselmo Cardoso de Paiva, Alexandre Pessoa, Ramsey D. Badawi, Vivek Swarnakar and Cláudio de Souza Baptista

Comparing Euclidean and Hyperbolic K-Means for Generalized Category Discovery 188
Mohamad Dalal, Thomas B. Moeslund and Joakim Bruslund Haurum

How to Sample High Quality 3D Fractals for Action Recognition Pre-Training? 199
Marko Putak, Thomas B. Moeslund and Joakim Bruslund Haurum

Disentangled Representation Learning with Variational Autoencoders for Facial Image Analysis 211
Lucas Massa, Tiago Vieira, Allan Martins, William Schwartz, Rodrigo Paes and Willy Tiengo

ShareDAB-DETR: Memory-Efficient Object Detection through CentralBank Parameter Sharing 222
Khaled Chikh, Roberto Cavicchioli and Alessandro Capotondi

Early-Retry Regeneration Framework for Improving Object-Sufficiency in Text-to-Image Generation 234
Shogo Ishii, Tomoaki Yamazaki, Kengo Murata, Seiya Ito and Kouzou Ohara

Deep Ordinal Partial-Label Learning for Cost-Effective Small Diamond Quality Grading 244
Arno Waes, Bram Claes and Toon Goedemé

Ego-Exo Temporal Action Segmentation in Industrial Environments 255
Alessandro Passanisi, Giovanni Maria Farinella and Francesco Ragusa

SHORT PAPERS

Breast Cancer Detection Using Multi-View Infrared Images and CNN Models
Gustavo Pioli Resende, Renata dos Santos Melo, Henrique Coelho Fernandes and André Ricardo Backes 267

Combi-CAM: A Novel Multi-Layer Approach for Explainable Image Geolocalization
D. Faget, J. L. Lisani and M. Colom 275

Identification of Water Tanks and Swimming Pools as Potential Dengue Breeding Sites Using Satellite Imagery
Rafael de Camillo Masson and André Ricardo Backes 282

Similarity Analysis in Source Code Using 1D Signal
Kaique Venuto Mancuzo and André Ricardo Backes 290

Noise-Aware Named Entity Recognition for Historical VET Documents
Alexander M. Esser and Jens Dörpinghaus 298

ScrewCount: A Dataset and Benchmark for Exemplar Efficiency and Text-Guided Few-Shot Object Counting
Farnaz Delirie, Afshin Dini, Amirmasoud Molaei and Leila Sadeghi 306

Graph-Based Point Cloud Surface Reconstruction Using B-Splines
Stuti Pathak, Rhys Evans, Gunther Steenackers and Rudi Penne 314

Beyond the Black Box: A Heuristic-Based Framework for Employing Real-World Reasoning towards Explainable and Reliable Results
Murilo Regio, Thomas Trappenberg and Isabel Manssour 322

Frequency-Domain Detection of AI-Generated Images via Radial Spectrum Templates
Yuta Sato, Takayoshi Yamashita, Hironobu Fujiyoshi and Tsubasa Hirakawa 330

QAC: Quality-Aware Clustering for Data-Centric Seismic Segmentation
Rahma Aloui, Pranav Martini, Pandu Devarakota, Apurva Gala and Shishir K. Shah 338

AIGVE-Tool: AI-Generated Video Evaluation Toolkit with Multifaceted Benchmark
Xinhao Xiang, Xiao Liu, Zizhong Li, Zhuosheng Liu and Jiawei Zhang 346

ORDINALBENCH: A Benchmark Dataset for Diagnosing Generalization Limits in Ordinal Number Understanding of Vision-Language Models
Yusuke Tozaki and Hisashi Miyamori 358

Recursive Memory Transformers for Scalable Analysis of Complex Vector Drawings
Andrey Pimenov, Vera Terenteva, Sergey Muravyov and Valeria Efimova 366

Towards Efficient Semantic Segmentation Deep Neural Networks: Case Study of Unstructured Pruning with DeepLabV3
Erika-Melinda Kali and Mihai Negru 374

A Data Centric Framework for Seismic Data Selection Using Latent Space Distribution Analysis
Samiha Mirza, Apurva Gala, Pandu Devarakota and Shishir Shah 382

Identification of Collided Vehicles in Indian Traffic Accidents Using Hierarchical Deep Learning Framework
Shrusti Porwal and Preety Singh 389

Focal-FusionNet: A Dual-Scale Attention Network for Accurate Small Object Detection with False Positive Suppression 397
Shubham Kumar Dubey, J. V. Satyanarayana and C. Krishna Mohan

PRT: A Training-Free Planar Tracking Framework via Point Correspondence and Robust RANSAC Estimation 402
Yong Kim, Hojae Kim, Suhyun Kim, Hyeonji Kim, Huijin Choi and Yonghyun Park

GloFM: A GLORYS Flow-Matching Emulator for Spatio-Temporal Ocean Data Assimilation 411
Pierre Garcia, Théo Archambault, Dominique Béréziat and Anastase Charantonis

FALQON-MST: A Fully Quantum Framework for Graph Optimization in Vision Systems 419
Guilherme E. L. Pexe, Lucas A. M. Rattighieri, Leandro A. Passos, Douglas Rodrigues, Danilo S. Jodas, João P. Papa and Kelton A. P. da Costa

Optimal Camera Placement for Dynamic Scenes via Reinforcement Learning in Virtual Environments 427
Grigorios Tsipouridis, Nikolaos Dimitriou, Efstathia Martinopoulou, Iason Karakostas and Dimitrios Tzovaras

A Computational Approach for Generative Modeling and Supervised Classification: Comparative Evaluation of GAN and Diffusion Models for Histological Data Augmentation 435
Bianca Lanconi de Oliveira Garcia, Guilherme Botazzo Rozendo, Thiago Leal Pozati, Thaína Aparecida Azevedo Tosta, Marcelo Zanchetta do Nascimento and Leandro Alves Neves

DevPrompt: Deviation-Based Prompt Learning for One-Normal Shot Image Anomaly Detection 443
Morteza Poudineh and Marc Lalonde

MCAMOT: Multi-Camera Assisted Multi-Object Tracking for Automated Pig Monitoring 451
Qinghua Guo, Yue Sun, Patrick P. J. H. Langenhuizen, J. Elizabeth Bolhuis, Piter Bijma and Peter H. N. De With

Hybrid Ensemble of Convolutional and Transformer Models for Retinal Image Quality Evaluation 459
Luís Felipe Rocha Pereira, João Marcello Mendes Moreira, João Dallyson Sousa de Almeida, Anselmo Cardoso de Paiva, Aristófanes Corrêa Silva and Elaine de Paula Fiod Costa

Frame Sequence Classification in Maritime UAV Surveillance: A Comparative Study of Temporal Deep Learning Architectures 467
Joniel Bastos Barreto, Filipe F. Caetano, Juliana M. F. da Silva and Carlos H. Q. Forster

The Threshold Paradox: Why Calibrating on the Test Set Introduces Bias in Anomaly Detection 475
Aurélie Cools, Sédrick Stassin and Sidi Ahmed Mahmoudi

Separation-Aware Downsampling for Vehicle Detection 483
Yupei Guo, Yota Yamamoto and Yukinobu Taniguchi

ControlNet-Guided Diffusion for Realistic Plant Leaf Generation 491
Georgios Tsoumplekas, Mia Pham, Yannis Spyridis, Thomas Lagkas, Georgios Th. Papadopoulos, Panagiotis Sarigiannidis and Vasileios Argyriou

Phys-3D: Physics-Constrained Real-Time Crowd Tracking and Counting on Railway Platforms 499
Bin Zeng, Johannes Künzel, Anna Hilsmann and Peter Eisert

Dynamic Kernel Linear Attention Mechanism 509
Zhiwei Guo, Qinxia Hu and Xiao Hu

DINOv3 for DE-ViT: Boosting Few-Shot Object Detection with Scalable and Robust Visual Features 517
Helmut Neuschmied, Werner Bailer and Martin Winter

Inpainting of Sparse Tracks Image Satellite Using Plug and Play and Learned Prior 524
Dmitrii Drozdov, Pierre Garcia, Dominique Béréziat and Anastase Charantonis

Application of GANs for Segmenting Cancerous Masses in Mammograms 532
Afifa Dahmane, Yanis Hammouche and Hadja Faiza Haned Khellaf

Video-Based Locomotion Analysis for Fish Health Monitoring 540
Timon Palm, Clemens Seibold, Anna Hilsmann and Peter Eisert

SITUATE: Synthetic Object Counting Dataset for VLM Training 549
René Peinl, Vincent Tischler, Patrick Schröder and Christian Groth

When Reality Is Not Enough: Synthesizing Single Frame Defects for Improving Defect Segmentation in Film 558
Martin Korb, Stefanie Onsori-Wechtitsch and Peter Schallauer

Lightweight and Efficient Deep Learning Architecture for Retinal Vessel Segmentation 566
Henda Boudegga, Rostom Kachouri and Yaroub Elloumi

Chaos-SSL: An Attention-Based Self-Supervised Learning Framework with Chaotic Transformation for Medical Image Classification 574
Joao Batista Florindo

Non-Contact Respiration Rate Estimation in Cattle from Overhead Videos 582
Purbaditya Bhattacharya, Erik Endlicher, Goutham Ravinaidu, Gerald Bieber, Judith Louise Pieper, Jan Langbein, Birger Puppe and Uwe Freiherr von Lukas

Are Superpixels the Missing Ingredient for Robust ViT Pruning? 590
Ahmed Soulmani, Martyna Poreba and Michal Szczepanski

A PeBERT Model for Generating Keywords for Retinography Images 598
Pedro Victor de A. Fonseca, João Dallyson Sousa de Almeida and Geraldo Braz Junior

Sin Loss for Class Imbalance and Hard Samples 606
Reiji Saito and Kazuhiro Hotta

Porosity Classification in High Pressure Die Casting Using Thermal Images and Sensor Data Fusion via Fuzzy Cognitive Maps 613
Tomasz Michno, Roxana Holom, Sebastian Schmalzer, Pauline Meyer-Heye, Giulia Scampone, Elias Riegler, Matthias Hartmann, Urban Repanšek, Nejc Košir, Peter Šifrer and Katarzyna Poczęta

GlovEgo-HOI: Bridging the Synthetic-to-Real Gap for Industrial Egocentric Human-Object Interaction Detection 623
Alfio Spoto, Rosario Leonardi, Francesco Ragusa and Giovanni Maria Farinella

Beyond Accuracy: A Counterfactual Analysis of Plant Disease Classifiers 631
Hajer Mejri and Sébastien Roy

Domain Specific Multi-Branch Network for Human Pose Estimation Inside a Car 642
Romain Guesdon, Carlos Crispim-Junior and Laure Tougne Rodet

Graph Memory: A Structured and Interpretable Framework for Modality-Agnostic Embedding-Based Inference 652
Artur A. Oliveira, Mateus Espadoto, Roberto M. Cesar Jr. and Roberto Hirata Jr.

Explaining the Inexplicable: An Explainable AI Approach to Boost Knowledge Transfer 660
Pietro Manganelli Conforti, Lorenzo Papa, Irene Amerini and Paolo Russo

Enhanced Class-Expertise Weighted Aggregation 671
Stepan Kasai, Sviatoslav Stumpf, Alexey Zabashta and Valeria Efimova

A Review of Advances in Artificial Intelligence-Driven Technologies for Assessing Food Quality Using Hyperspectral and Multispectral Imaging 679
Michael B. Estrada, Boris X. Vintimilla and Luis E. Chuquimarca

CRAX: Parameter-Efficient Fine-Tuning of SAM2 for Interactive Crack Annotation 689
Filip Hendrichovsky, Raphael Pickl, Lukas Bednar and Daniel Soukup

NEF: Neural Error Fields for Follow-up Training with Fewer Rays 701
Alessandro Luchetti, Kenta Ito, Dieter Schmalstieg, Denis Kalkofen and Shohei Mori

Synthetic Aerial Image Generation with Controllable Diffusion Models for Pinus Segmentation under Data Scarcity 709
Thiago Innani Justus, Rodrigo Minetto, Gilson Giraldi and Mauren Louise Sguario

A Neuro-Symbolic Framework of Behaviors Recognition for Human Robot Interaction 717
Tao Wang and Hao Tang

Efficient Video Segmentation with Differential Networks 726
Artur A. Oliveira, Mateus Espadoto, Roberto M. Cesar Jr. and Roberto Hirata Jr.

From Synthetic Data to Real Restorations: Diffusion Model for Patient-Specific Dental Crown Completion 734
Dávid Pukanec, Tibor Kubík and Michal Španěl

SLAMDUNKS: A Vision-Based Approach for Uncovering Semantic Relations in Dataset Collections 743
Petra Bevandić and Barbara Hammer

AUTHOR INDEX 751