

PROCEEDINGS OF SPIE

International Conference on Optics and Computer Vision (ICOCV 2026)

Yang Yue
Editor

16–18 January 2026
Hong Kong, China

Organized by
Xi'an Jiaotong University (China)

Sponsored by
Hong Kong Baptist University (Hong Kong, China)
Central China Normal University (China)
Aurel Vlaicu University of Arad (Romania)

Published by
SPIE

Volume 14233

Proceedings of SPIE 0277-786X, V. 14233

SPIE is an international society advancing an interdisciplinary approach to the science and application of light.

The papers in this volume were part of the technical conference cited on the cover and title page. Papers were selected and subject to review by the editors and conference program committee. Some conference presentations may not be available for publication. Additional papers and presentation recordings may be available online in the SPIE Digital Library at SPIDigitalLibrary.org.

The papers reflect the work and thoughts of the authors and are published herein as submitted. The publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Please use the following format to cite material from these proceedings:

Author(s), "Title of Paper," in *International Conference on Optics and Computer Vision (ICOCV 2026)*, edited by Yang Yue, Proc. of SPIE 14233, Seven-digit Article CID Number (DD/MM/YYYY); (DOI URL).

ISSN: 0277-786X

ISSN: 1996-756X (electronic)

ISBN: 9798902325260

ISBN: 9798902325277 (electronic)

Published by

SPIE

P.O. Box 10, Bellingham, Washington 98227-0010 USA

Telephone +1 360 676 3290 (Pacific Time)

SPIE.org

Copyright © 2026 Society of Photo-Optical Instrumentation Engineers (SPIE).

Copying of material in this book for internal or personal use, or for the internal or personal use of specific clients, beyond the fair use provisions granted by the U.S. Copyright Law is authorized by SPIE subject to payment of fees. To obtain permission to use and share articles in this volume, visit Copyright Clearance Center at copyright.com. Other copying for republication, resale, advertising or promotion, or any form of systematic or multiple reproduction of any material in this book is prohibited except with permission in writing from the publisher.

Printed in the United States of America by Curran Associates, Inc., under license from SPIE.

Publication of record for individual papers is online in the SPIE Digital Library.

**SPIE. DIGITAL
LIBRARY**

SPIDigitalLibrary.org

Paper Numbering: A unique citation identifier (CID) number is assigned to each article in the Proceedings of SPIE at the time of publication. Utilization of CIDs allows articles to be fully citable as soon as they are published online, and connects the same identifier to all online and print versions of the publication. SPIE uses a seven-digit CID article numbering system structured as follows:

- The first five digits correspond to the SPIE volume number.
- The last two digits indicate publication order within the volume using a Base 36 numbering system employing both numerals and letters. These two-number sets start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B ... 0Z, followed by 10-1Z, 20-2Z, etc. The CID Number appears on each page of the manuscript.

Contents

vii *Conference Committee*

INTERNATIONAL CONFERENCE ON OPTICS AND COMPUTER VISION (ICOCV 2026)

- 14233 02 **Transformer pruning and optimization for embodied AI: an embedded implementation on RK3588** [14233-49]
- 14233 03 **Defect identification based on data augmentation and deep learning** [14233-16]
- 14233 04 **SW-CSTCN: an automatic epilepsy detection model based on sliding window convolution, channel attention, and temporal convolutional network** [14233-51]
- 14233 05 **GrabCut-Open3D-PCA: open-vocabulary driven 3D sorting** [14233-54]
- 14233 06 **Sentiment classification method based on data augmentation and neural networks** [14233-18]
- 14233 07 **Optimization of snack image recognition models based on data augmentation and deep learning** [14233-17]
- 14233 08 **Novel MPC-based trajectory tracking approach for unmanned vehicles: balancing tracking accuracy, control smoothness, and real-time performance** [14233-52]
- 14233 09 **DDoS-AIDE: towards explainable and automated DDoS defense using large language models** [14233-21]
- 14233 0A **Research on personnel area intrusion recognition and safety monitoring system in intelligent workshop** [14233-50]
- 14233 0B **A comparative study of ACNet and SPA-ResNet performance under different signal-to-noise ratio conditions** [14233-55]
- 14233 0C **Research on the construction and application of consumer purchasing behavior prediction model based on deep neural networks** [14233-13]
- 14233 0D **An empirical study on self-admitted technical debt (SATD) detection with large language models** [14233-47]
- 14233 0E **AdapFusion: adaptive fusion of multimodal features to improve classification** [14233-43]
- 14233 0F **A random forest model simulating Bayesian thinking: uncertainty quantification and interpretability improvement methods** [14233-41]
- 14233 0G **PRAN: point relation analysis network** [14233-31]

- 14233 OH **Research on the application of pre-trained language model architecture enhancement and fine-tuning based on differential attention mechanism in classical poetry translation** [14233-23]
- 14233 OI **Image classification by fusing convolutional neural networks and lightweight Mamba state space model** [14233-42]
- 14233 OJ **HandiSpeak: lightweight continuous ASL recognition for real-time deployment** [14233-48]
- 14233 OK **Dual-attention residual network for efficient single-image super-resolution** [14233-56]
- 14233 OL **How do different architectural variations affect the performance of VQA models under extremely constrained resources?** [14233-32]
- 14233 OM **Recognition of special vehicle types using data augmentation, attention mechanisms, and neural networks** [14233-62]
- 14233 ON **Research on brain tumor image segmentation based on multiscale context attention** [14233-37]
- 14233 OO **Personalized prediction of affective intensity from eye-tracking via learning-to-rank** [14233-53]
- 14233 OP **Deep reinforcement learning for path optimization in UAV-enabled emergency material distribution** [14233-40]
- 14233 OQ **SonarDet: receptive field-enhanced multiscale sonar image object detection** [14233-19]
- 14233 OR **A multimodal emotion recognition algorithm based on deep convolution and transformer with multihead self-attention** [14233-34]
- 14233 OS **Research on mean-shift-based multiagent formation control optimized by genetic algorithm** [14233-22]
- 14233 OT **ERDet: receptive field enhanced multiscale steel surface defect detection** [14233-46]
- 14233 OU **CHST: CLS-guided hybrid sparse transformer for occluded pedestrian re-identification** [14233-45]
- 14233 OV **Design of short-term power load forecasting method using differential machine learning to improve LSTM algorithm** [14233-15]
- 14233 OW **Rule-guided lightweight deep reinforcement learning policy optimization and convergence analysis** [14233-39]
- 14233 OX **Analysis of airport passenger behavior patterns based on clustering algorithms** [14233-57]

- 14233 0Y **Computer simulation-based optimization of passenger flow in transportation hubs**
[14233-65]
- 14233 0Z **Application of multiobjective optimization algorithms in enhancing green construction strategies** [14233-58]
- 14233 10 **Prompt optimization for large language model-driven object detection in complex scenes**
[14233-24]
- 14233 11 **Research on automatic schedule verification and deviation analysis framework of power infrastructure projects based on 3DGS** [14233-61]
- 14233 12 **The application of federated learning frameworks in distributed data processing for rural households** [14233-28]
- 14233 13 **Implementation of a greedy algorithm in a real-time logistics scheduling teaching tool**
[14233-64]
- 14233 14 **Intelligent diagnosis assistant system of colorectal cancer based on artificial intelligence**
[14233-63]
- 14233 15 **Research and application of zero trust security defense model based on proactive defense concept** [14233-30]
- 14233 16 **Automatic pupil measurement in closed-eye infrared imaging based on ResUNet-CBAM**
[14233-33]
- 14233 17 **Image forgery detection based on YOLOv5 object detection and frequency-domain feature fusion** [14233-27]
- 14233 18 **A computer vision-based multidimensional data modelling study of visual expression of gamified art in interaction design** [14233-26]
- 14233 19 **High-fidelity 3D rendering and intelligent visual interaction for intangible cultural heritage: a metaverse-based approach** [14233-12]
- 14233 1A **Research on intrusion recognition method of small UAV based on communication and perception integration and multimodal AI** [14233-2]
- 14233 1B **Decision transformer for power dispatch: offline pre-training with power flow simulator and online fine-tuning** [14233-1]
- 14233 1C **A BERT-based multitask learning system for automated assessment of advanced business English writing proficiency** [14233-4]
- 14233 1D **Design and construction of a facial recognition attendance platform based on MobileNetV3** [14233-3]
- 14233 1E **Research on deformation prediction of complex steel plate weld seam structure based on graph neural network** [14233-5]

- 14233 1F **Automatic compliance verification of pumped storage power station engineering construction based on CLIP architecture and IoT data fusion [14233-11]**
- 14233 1G **Construction of intelligent guide interaction system structure and instruction resolution mechanism based on multimodal perception fusion [14233-7]**
- 14233 1H **Construction of a machine-learning-based shale cap permeability prediction model and analysis of CO₂ sequestration adaptability [14233-10]**
- 14233 1I **Design and implementation of a data-driven operational performance evaluation system [14233-35]**
- 14233 1J **Innovative research on Lingnan symbolic cotton textile clothing combining laser engraving and computer-aided design [14233-36]**
- 14233 1K **The application and optimization of graph neural networks in enterprise-level data anomaly detection and repair systems [14233-14]**
- 14233 1L **Keyword-cued multimodal fusion for cross-domain target matching: a generalized framework for low-altitude scenarios [14233-29]**