

PROCEEDINGS OF SPIE

International Conference on Computer Vision, AI, and Intelligent Automation (ICCVAA 2026)

Yong Yue
Editor

10–12 April 2026
Zhengzhou, China

Organized by
University of Haute Alsace (France)

Published by
SPIE

Volume 14260

Proceedings of SPIE 0277-786X, V. 14260

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 Computer Vision, AI, and Intelligent Automation (ICCVAA 2026)*, edited by Yong Yue, Proc. of SPIE 14260, Seven-digit Article CID Number (DD/MM/YYYY); (DOI URL).

ISSN: 0277-786X

ISSN: 1996-756X (electronic)

ISBN: 9798902325918

ISBN: 9798902325925 (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*

COMPUTER VISION

- 14260 02 **A minority-friendly learning method for object detection: joint optimization of oversampling-driven data balancing and loss reconstruction** [14260-9]
- 14260 03 **Adaptive prompt-guided local cross-modal alignment for zero-shot vision recognition** [14260-20]
- 14260 04 **TexPAT-YOLO: a partial channel fusion backbone for textile defect detection based on YOLOv11** [14260-8]
- 14260 05 **Coal gangue image recognition technology based on the improved YOLOv7 algorithm** [14260-22]
- 14260 06 **A multiscale feature fusion algorithm for enhancing the clarity of underwater images** [14260-23]
- 14260 07 **CMPT: crack mining with periodic-frequency grouping and twin-domain collaborative attention fusion** [14260-28]
- 14260 08 **Performance evaluation and comparative experiment of machine vision algorithms for digital inspection of Guangxi bronze drums** [14260-29]
- 14260 09 **Lightweight road crack detection by fusing multiscale features and attention mechanisms** [14260-41]
- 14260 0A **A lightweight vision transformer for edge devices and its real-time applications in industrial quality inspection** [14260-34]
- 14260 0B **Research on traffic sign image recognition and classification methods based on CNN** [14260-1]
- 14260 0C **Mamba and diffusion-based structure-guided arbitrary-scale video super-resolution** [14260-32]
- 14260 0D **AST-ToMe: adaptive step-aware thresholding for efficient temporal-spatial feature reuse in diffusion models** [14260-12]
- 14260 0E **Commercial applications and optimization of deep learning image recognition based on artificial intelligence** [14260-5]

MEDICAL IMAGE ANALYSIS

- 14260 OF **A review of reliability challenges and countermeasures of biomedical image analysis in surgical robotics** [14260-4]
- 14260 OG **TWD-Det: uncertainty-aware three-way decision framework for robust medical object detection** [14260-7]
- 14260 OH **Research on vision-sensor-language multimodal model for knee joint risk early warning in squat training** [14260-17]
- 14260 OI **Lightweight 3D network for ischemic stroke lesion segmentation** [14260-27]
- 14260 OJ **Comparative study of MRI deblurring optimization algorithms based on total variation regularization** [14260-24]

HUMAN-COMPUTER INTERACTION AND BEHAVIOR RECOGNITION

- 14260 OK **Research on gesture recognition and feedback system for art interactive installations using deep learning** [14260-10]
- 14260 OL **Exploration of AI vision tracking technology in tennis teaching and competition** [14260-30]
- 14260 OM **Research on intelligent interaction triggering in digital media interactive scenarios based on YOLO object detection algorithm** [14260-38]
- 14260 ON **Research on student behavior recognition based on machine learning and its applications** [14260-39]
- 14260 OO **Computer vision-based human motion recognition: technological evolution, key challenges, and cutting-edge trends** [14260-26]

AI SECURITY AND PHYSICAL MODELING

- 14260 OP **Frequency-augmented multiscale graph neural networks for time-series industrial control network attack detection** [14260-13]
- 14260 OQ **Tackling extreme class imbalance in NIDS: a hybrid architecture integrating SMOTE and CNN-BiLSTM with focal loss** [14260-16]
- 14260 OR **Physics-guided convolutional neural network-transformer for robust sea clutter modeling in high sea states** [14260-21]

AUTONOMOUS DRIVING PERCEPTION

- 14260 OS **Research on quantitative optimization of video desensitization for autonomous vehicles based on artificial intelligence** [14260-3]

- 14260 OT **Robust multitask learning framework for panoptic autonomous driving perception**
[14260-11]
- 14260 OU **Visual perception with object detection for autonomous driving under complex weather conditions** [14260-18]
- 14260 OV **A multimodule collaborative optimization method for YOLOv11-based vehicle detection in complex traffic scenes** [14260-31]
- 14260 OW **DQN-based 3D path planning for UAVs in urban airspace** [14260-15]
- 14260 OX **MRFO-based 3D path planning for UAV rescue missions in complex mountainous environments** [14260-37]
- 14260 OY **Improved YOLOv8n with feedback linearization-based fuzzy predictive functional control for rice transplanting machine row detection and adaptive path tracking** [14260-35]
- 14260 OZ **CSFormer: cross-spatial attention in vision transformer for road extraction from satellite imagery** [14260-25]