



# Guest Editorial: Special Issue on Open-World Visual Recognition

Zhun Zhong<sup>1,2</sup> · Hong Liu<sup>3</sup> · Yin Cui<sup>4</sup> · Shin'ichi Satoh<sup>5</sup> · Nicu Sebe<sup>6</sup> · Ming-Hsuan Yang<sup>7</sup>

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## 1 Introduction

This special issue is devoted to the theme of open-world visual recognition. Visual recognition is a critical task in computer vision, which has gained significant attention in recent years due to its numerous applications in various fields, including image classification, object detection, semantic segmentation, instance retrieval, etc. Despite significant progress, existing visual models continue to face considerable challenges in open-world scenarios. These challenges include recognizing novel classes, adapting to unseen domains, learning under constraints of data privacy, enhancing robustness against adversarial samples, among others. This special issue provides a comprehensive overview of the latest advancements in open-world visual recognition, aiming to address these complex issues.

The call for papers for this special issue attracted a total of 144 submissions, reflecting the community's strong

interest and ongoing research efforts in this area. After a rigorous peer-review process, consistent with the journal's high standards for quality and innovation, 44 papers have been accepted. This results in an acceptance rate of 30.5% (44/144), highlighting the competitive nature of our selection process. The accepted papers showcase cutting-edge research and are mainly organized into seven thematic categories: open-set & open-vocabulary recognition, domain adaptation & generalization, out-of-distribution detection, learning with imperfect training labels, novel class discovery, incremental learning and other open-world applications.

## 2 Overview of Accepted Papers

### 2.1 Open-Set & Open-Vocabulary Recognition

This part of the issue consists of 11 papers. *The first article*, by Zhang et al., introduces the Open-Vocabulary Keypoint Detection (OVKD) task and proposes a novel framework based on semantic-feature matching, which combines vision and language models to link language features with local visual keypoint features. *The second article*, by Shi et al., presents a novel approach for open-vocabulary semantic segmentation by leveraging the capabilities of large language models (LLMs) instead of traditional vision-language (VL) pre-training models like CLIP. *The third article*, by Wang et al., introduces the task of Open-Vocabulary Video Instance Segmentation (OV-VIS) and proposes a transformer-based model to solve this task. *The fourth article*, Chen et al., studies the task of open-vocabulary object detection and attributes recognition and proposes an effective framework that disentangles the task into class-agnostic object proposal and open-vocabulary classification. *The fifth article*, by Thawakar et al., studies the problem of open-world video instance segmentation (VIS) and introduces a novel framework with feature enrichment mechanism and spatio-temporal objectness module. *The sixth article*, by Tang et al., proposes a training-free paradigm for open-world segmentation, which effectively harnesses the power of vision foundational

---

✉ Hong Liu  
hliu@ids.osaka-u.ac.jp  
Zhun Zhong  
zhunzhong007@gmail.com  
Yin Cui  
yinc@nvidia.com  
Shin'ichi Satoh  
satoh@nii.ac.jp  
Nicu Sebe  
niculae.sebe@unitn.it  
Ming-Hsuan Yang  
mhyang@ucmerced.edu

<sup>1</sup> Hefei University of Technology, Hefei 230002, China  
<sup>2</sup> University of Nottingham, Nottingham NG8 1BB, UK  
<sup>3</sup> Osaka University, Osaka 565-0871, Japan  
<sup>4</sup> NVIDIA, Santa Clarita 95051, USA  
<sup>5</sup> National Institute of Informatics, Tokyo 101-8430, Japan  
<sup>6</sup> University of Trento, Trento 38123, Italy  
<sup>7</sup> University of California at Merced, Merced, CA 95344, USA

models. *The seventh article*, by Yang et al., introduces the prototype-based segmentation framework that can combine textual and visual clues, providing comprehensive support for open-world semantic segmentation. *The eighth article*, by Chakravarthy et al., studies the problem of open-world lidar panoptic segmentation. It discovers the drawbacks of existing methods on this task and suggests a balanced approach that can achieve strong performance on both known and unknown classes. *The ninth article*, by Yang et al., introduces the Causal Inference-inspired approach for real-world open-set recognition, addressing challenges posed by covariate and semantic shifts. *The tenth article*, by Zuo et al., integrates vision-language embeddings from foundation models into 3D Gaussian Splatting, which can enhance multi-view semantic consistency and thus facilitate downstream tasks, such as open-vocabulary object detection. *The eleventh article*, by Xie et al., introduces a diffusion-based data augmentation technique for large vocabulary instance segmentation, which operates without training or label supervision.

## 2.2 Domain Adaptation & Generalization

This part of the issue consists of 13 papers. *The first article*, by Wu et al., proposes a domain-aware prompting approach for cross-domain few-shot learning, which learns a hybridly prompted model for enhancing adaptability on unseen domains. *The second article*, by Dai et al., studies the task of cross-domain person re-identification and proposes a novel framework that can generate intermediate domains for improving the knowledge transfer between source and target domains. *The third article*, by Zhao et al., introduces the problem of Multi-Source-Free Domain Adaptive Object Detection (MSFDAOD) and proposes a Divide-and-Aggregate Contrastive Adaptation (DACA) framework that can efficiently leverage the advantages of multiple source-free models and aggregate their contributions to adaptation in a self-supervised manner. *The fourth article*, by Gu et al., introduces an adversarial re-weighting approach for partial domain adaptation, which can reduce the influence of source-private classes and minimize prediction uncertainty in the target domain. *The fifth article*, by Zhang et al., proposes a new framework for source-free domain adaptation, which incorporates pre-trained networks into the adaptation process to improve the quality of target pseudo-labels. *The sixth article*, by Liang et al., provides a survey of test-time adaptation (TTA), which categorizes TTA into several distinct groups, provides a comprehensive taxonomy of advanced algorithms for each group, and analyzes relevant applications of TTA. *The seventh article*, by Wang et al., presents a comprehensive survey on online test-time adaptation (OTTA), re-implements existing OTTA methods with

Vision Transformer, and proposes novel evaluation metrics that consider both accuracy and efficiency. *The eighth article*, by Yang et al., introduces the Hierarchical Visual Transformation (HVT) network to help the model learn domain-invariant representation and narrow the domain gap in various visual matching and recognition tasks. *The ninth article*, by Hu et al., introduces a large-scale benchmark for domain generalizable person re-identification as well as proposes a novel framework based on diverse feature space learning to learn domain-adaptive discriminative representations. *The tenth article*, by Wang et al., studies the problem of domain generalized unmanned aerial vehicle object detection and proposes a novel frequency domain disentanglement method to improve model's generalization ability on this challenging task. *The eleventh article*, by Luo et al., proposes a method based on network pruning for domain generalized semantic segmentation, which can prune the filters or attention heads that are more sensitive to domain shift. *The twelfth article*, by Huang et al., considers a specific domain generalization task, i.e., out-of-distribution generalization and presents an Exploring Variant parameters for Invariant Learning (EVIL) approach to find the parameters that are sensitive to distribution shift. *The thirteen article*, by Li et al., studies the model's robustness to adversarial examples, which can be regarded as a specific category of domain generalization. This article proposes a novel method to automatically learn online, instance-wise, data augmentation policies for improving robust generalization.

## 2.3 Out-of-Distribution Detection

This part of the issue consists of 6 papers. *The first article*, by He et al., considers the task of Out-of-Distribution (OOD) detection with noisy examples in the training set. It introduces the Adversarial Confounder Removing (ACRE) method, which utilizes progressive optimization with adversarial learning to curate collections of easy-ID, hard-ID, and open-set noisy examples as well as to reduce spurious-related representations. *The second article*, by Nie et al., introduces Virtual Outlier Smoothing (VOSo) for OOD detection by generating auxiliary outliers using in-distribution samples. *The third article*, by Cheng et al., takes the advantage of the breakthrough of generative models and demonstrates that training with a large quantity of generated data can eliminate overfitting in reliable prediction tasks, e.g., OOD detection. *The fourth article*, by Fang et al., introduces a novel perspective on OOD detection by exploring the loss landscape and mode ensemble, showing the effectiveness of mode ensemble in enhancing OOD detection. *The fifth article*, by Yang et al., provides a survey of OOD detection methods and presents a unified framework called generalized OOD detection, which encompasses five highly

related tasks, i.e., OOD detection, anomaly detection (AD), novelty detection (ND), open set recognition (OSR), and outlier detection (OD). In addition, it provides a comprehensive discussion of representative methods from other tasks and how they relate and inspire the development of OOD detection methods. *The sixth paper*, by Wang et al., provides a consolidated analysis of OOD detection and open-set recognition (OSR), performing cross-evaluation of state-of-the-art methods, proposing a new large-scale benchmark, and providing empirical analysis on existing methods and correlation between OOD detection and OSR.

## 2.4 Learning with Imperfect Training Labels

This part of the issue consists of 7 papers. *The first article*, by Butt et al., introduces a large-scale dataset for road segmentation in challenging unstructured roadways. It proposes an Efficient Data Sampling (EDS) based self-training framework for semi-supervised learning setting. *The second article*, by Sun et al., introduces Variational Rectification Inference (VRI) to address the problem of learning with noisy labels by formulating adaptive loss rectification as an amortized variational inference problem. *The third article*, by Xie et al., presents the Probabilistic Representation Contrastive Learning (PRCL) framework for semi-supervised semantic segmentation, which enhances the robustness of the unsupervised training process. *The fourth article*, by Zhao et al., addresses the task of open-set semi-supervised learning and proposes a prototype-based clustering and identification algorithm to enhance feature learning. *The fifth article*, by Sun et al., introduces the Open-World Deep-Fake Attribution task and benchmark, where the unlabeled dataset may contain attacks that have never been encountered in the labeled set, and proposes the Multi-Perspective Sensory Learning (MPSL) framework to solve this task. *The sixth article*, by Qiao et al., presents Adaptive Fuzzy Positive Learning (A-FPL) for annotation-scarce semantic segmentation, which can effectively alleviate interference from wrong pseudo labels and progressively refining semantic discrimination. *The seventh article*, by Siméoni et al., provides a survey of different unsupervised object localization methods that discover objects in images without requiring any manual annotation in the era of self-supervised ViTs.

## 2.5 Novel Class Discovery

This part of the issue consists of 2 papers. *The first article*, by Chi et al., studies novel class discovery task under unreliable sampling and proposes a hidden-prototype-based discovery network (HPDN) to handle sampling errors. *The second article*, by Riz et al., introduces the task of Novel Class Discovery (NCD) for 3D point cloud semantic

segmentation. To solve this problem, it proposes a new method utilizing online clustering, uncertainty estimation, and semantic distillation.

## 2.6 Incremental Learning

This part of the issue consists of 2 papers. *The first article*, by Xuan et al., introduces the concept of Incremental Model Enhancement (IME), where training data arrives sequentially and each training split typically corresponds to a set of independent classes, domains, or tasks. It proposes a Memory-based Contrastive Learning framework, which shows superiority on both image classification and semantic segmentation tasks. *The second article*, by Zhou et al., revisits Class-Incremental Learning (CIL) in the context of pre-trained models (PTMs) and shows that the core factors in CIL are adaptivity for model updating and generalizability for knowledge transferring. It also proposes a general framework by aggregating the embeddings of PTM and adapted models for classifier construction.

## 2.7 Other Open-World Applications

This part of the issue consists of 3 papers. *The first article*, by Want et al., proposes a specific open-world visual recognition task, i.e., Pattern-Expandable Image Copy Detection, aiming to identify novel tamper patterns. It proposes Pattern Stripping, which can easily introduce new pattern features with minimal impact on the image feature and previously seen pattern features. *The second article*, by Xu et al., addresses the challenge of visual object tracking in hazy conditions and introduces a feature restoration transformer to improve model's robustness under hazy imaging scenarios. *The third article*, by Shi et al., introduces a model-agnostic Curricular shApe-aware FEature (CAFE) learning strategy for Panoptic Scene Graph Generation (PSG), which is effective on both robust and zero-shot PSG tasks.

## 3 Conclusion

The 44 contributions in this special issue offer a diverse array of perspectives aimed at addressing the challenges in the field of open-world visual recognition. These articles not only underscore the ongoing dialogue within the community but also highlight innovative approaches to tackle real-world issues effectively. Through this special issue, we aim to spark further research and discussion within the community, encouraging continued advancements and practical applications of open-world visual recognition.

Finally, we would like to express our heartfelt gratitude to the dedicated reviewers who devoted their valuable

time and effort to thoroughly review the papers and provide constructive feedback to the authors. We also extend our appreciation to the diligent editorial team at Springer and International Journal of Computer Vision, especially Prof. Yasuyuki Matsushita, Ms. Yasotha Sujeen, and Ms.

Katherine Moretti. Their invaluable assistance was crucial in the successful publication of this special issue.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.