

Open-World Object Detection and Segmentation: Current Challenges and Emerging Opportunities

Muhammad Iqbal Ali¹, In-Chul Han¹, Soo Kyun Kim^{2*}

¹Student, Dept of Computer Engineering, Jeju National University

²Professor, Dept of Computer Engineering, Jeju National University

오픈 월드 객체 탐지 및 분할 연구: 현재의 과제와 새로운 가능성

무하마드 이크발 알리¹, 한인철¹, 김수균^{2*}

¹제주대학교 컴퓨터공학과 학생, ²제주대학교 컴퓨터공학과 교수

Abstract Real-world vision systems often encounter categories unseen during training, challenging the closed-set limitation of traditional detectors. This review summarizes advances in Open-World Object Detection (OWOD) and its extension to open-world segmentation, where models must recognize known classes, reject novel instances, and incrementally learn new categories. We categorize OWOD approaches into pseudo-labeling, class-agnostic proposal learning, metric-based novelty scoring, and hybrid methods leveraging vision-language or video cues, connecting them to open-world semantic and instance segmentation. Key challenges include proposal bias, open-set misclassification, continual learning degradation, and inconsistent evaluation. Finally, we highlight future directions such as integration with 3D perception, self/semi-supervised learning, and calibrated uncertainty, outlining a roadmap toward reliable open-world perception.

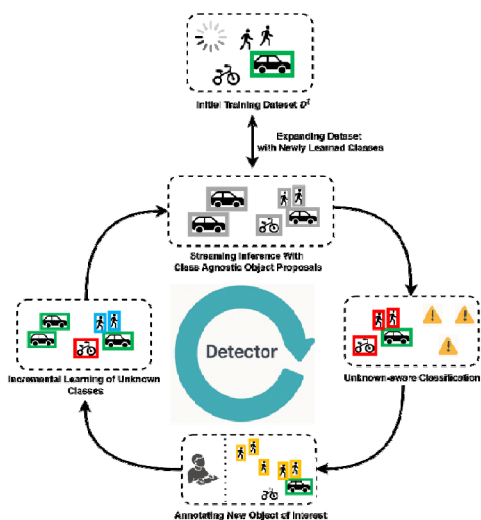
Key Words : Open-world Object Detection (OWOD), Open-World Segmentation, Open-Set Recognition, Unknown Detection, Incremental/Continual Learning

요약 현실 세계의 비전 시스템은 학습 시 보지 못한 객체 범주를 자주 마주하게 되며, 이는 기존 폐쇄형 탐지기의 한계를 드러낸다. 본 논문은 오픈 월드 객체 탐지(Open-World Object Detection, OWOD)와 그 확장 분야인 오픈 월드 분할 연구를 종합적으로 분석하였다. OWOD는 기존 클래스 인식, 미지 객체 식별, 새로운 클래스의 점진적 학습을 요구하며, 주요 접근법은 의사 라벨링, 클래스 비의존 제안 학습, 특징 기반 미지 객체 판별, 비전-언어 및 영상 기반 하이브리드 기법으로 분류된다. 본 연구는 OWOD의 핵심 과제(제안 편향, 오분류, 점진 학습의 성능 저하, 평가 불일치)를 논의하고, 3D 인식 통합, 자기/반지도 학습, 불확실성 정규화 등을 포함한 향후 연구 방향을 제시한다.

주제어 : 오픈 월드 객체 탐지, 오픈 월드 분할, 오픈 셋 인식, 미확인 물체 탐지, 점진적/지속적 학습

1. Introduction

Conventional detection frameworks assume a closed world in which the train/test label set is fixed, yet practical deployments (e.g., autonomous driving and robotics) encounter unseen categories. Open-World Object Detection (Open-World Object Detection (OWOD)) aims to enable detection frameworks to (i) localize previously trained classes, (ii) designate unfamiliar instances as novel object, and (iii) incrementally absorb newly labeled classes via episodic oracle updates. Unlike closed-world models that force-fit novelties, OWOD rejects them and couples novel object handling with continual-learning mechanisms to curb performance degradation from forgetting prior knowledge, aligning with non-stationary data streams. Since Joseph et al[1]. formalized OWOD, progress spans improved novel object proposals, novel-class discovery, and incremental update strategies for instance OW-DETR[2], PROB[3], class-agnostic calibration, and metric-learning-based novel object scoring[4-6]. This review consolidates these advances and extends to open-world object segmentation, outlining challenges and promising directions. An overview of the OWOD problem is presented in Fig. 1.



[Fig. 1] An Overview of the Open World Object Detection Problem

1.1 Definitions and scope

Open-world object detection detects known classes, flags out-of-set objects as unknown, and later learns them as new classes [20] [21]. Open-world segmentation applies the same idea at the pixel level (semantic, instance, or panoptic) by marking unknown regions [22]. This setting is needed in practice because label sets are incomplete and environments change, which can lead to overconfident errors.

1.2 Review method

We screened peer-reviewed papers related to OWOD and open-world segmentation that published in top conferences such as CVPR. We included works with standard open-set or incremental splits and explicit unknown handling. For each, we noted, how unknowns are scored or mined, use of pseudo-labels, and reported metrics, then organized the field into a simple taxonomy.

1.3 Issues in Existing Literature

Open-world detection and segmentation still face five practical gaps. First, models are trained on incomplete label sets, so unseen objects are mislabeled and hard to add later without forgetting. Second, unknown scoring and thresholding are unstable across domains; proposal and small-object recall remain weak, and pseudo-labels add noise. Third, for segmentation, per-pixel discovery is tougher, with boundary errors and mask leakage into known classes. Fourth, benchmarks and metrics are inconsistent, which blurs comparisons. Finally, real deployments face domain shift and tight latency/memory budgets that further reduce unknown recall and mask quality.

2. Related Work

2.1 Open-Set Recognition

Open-World Object Detection (OWOD) is

grounded in open-set recognition (OSR), wherein a classifier must identify samples from unseen categories and decline confident assignment to known labels. Scheirer et al. [7] formalized OSR via the notion of open space risk the danger of labeling novel inputs as members of previously trained classes. Within OSR, the objective is twofold: maintain accuracy on previously trained classes while reliably and robustly rejecting novel objects. Classical approaches (e.g., 1-vs-Set SVM) imposed penalties for decisions outside a class's support region. With deep networks, methods such as OpenMax [8] modeled activation distributions to flag novelty, while CROSR [18] coupled classification with reconstruction to better separate novel objects. These principles out-of-distribution detection, calibrated confidence, and conservative decision boundaries underpin OWOD) Crucially, however, OSR addresses isolated classification, whereas OWOD integrates novel object handling into structured prediction: the detector must localize objects and then decide between known and novel object, and subsequently support incremental incorporation of newly labeled classes.

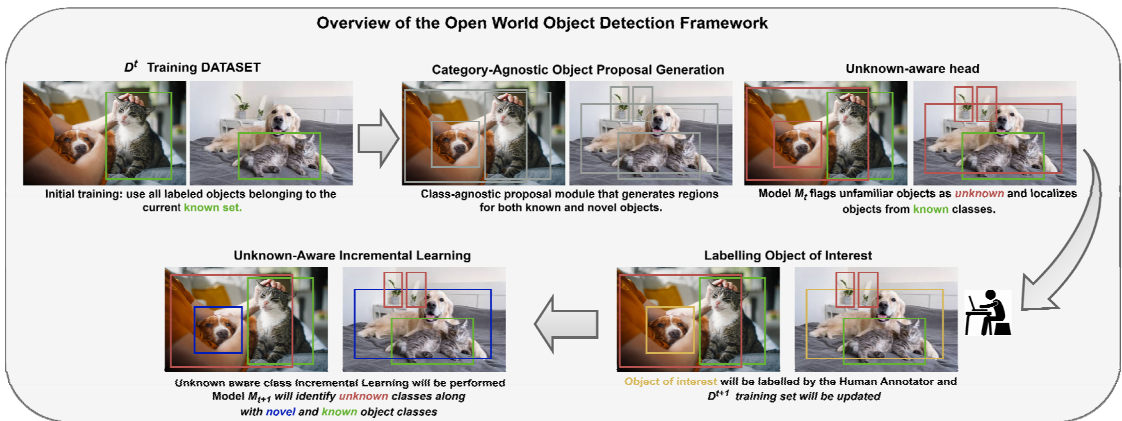
2.2 Incremental and Lifelong Object Detection

Incremental object detection also termed class-incremental detection (CID) updates a detector to recognize novel classes while mitigating forgetting of previously learned categories, yet it does not explicitly handle novel objects at inference. Peng et al. [18] fine-tuned detectors for new classes while preserving prior knowledge via distillation; Faster ILOD further improved stability and accuracy. Subsequent works [9, 10, 11, 12, 13, 14, 15, 16] incorporated exemplar replay to retain past instances, and Joseph et al. [1] introduced a meta-learning strategy for incremental detection. In CID, the test distribution is assumed to contain only previously trained classes (i.e., no novel objects to flag). By contrast, Open-World Object

Detection (OWOD) can be viewed as the fusion of OSR and CID: the system first performs OSR-style rejection to identify novel object objects, and then assimilates them incrementally in CID fashion. This combined formulation is often described more broadly as open-world recognition, coupling novel object rejection with continual expansion of the label space. Table 1 provides an overview of the existing methods and their underlying frameworks.

2.3 OWOD vs OVOD

Vocabulary Detection: It is important to distinguish OWOD from open-vocabulary object detection (OVOD). OVOD seeks to recognize categories beyond the training set by exploiting semantic priors, for instance, word embeddings or large vision-language models, to generalize to unseen labels. Consequently, open-vocabulary methods [19] can recognize novel classes at test time using language cues but typically do not perform continual learning: the detector is not updated to permanently incorporate these classes, and it presumes some semantic guidance (names/descriptions) for them. In contrast, OWOD assumes no prior knowledge of novel object categories and requires learning them from scratch once annotations are provided. Thus, the two paradigms target different facets of openness: Open-World Object Detection (OWOD) emphasizes the detection of novel objects with the continual expansion of the label space, whereas OVOD emphasizes zero-shot recognition via language grounding. They are complementary, and emerging hybrids leverage pre-trained vision-language models to assist novel object discovery and labeling while retaining OWOD's incremental loop. Fig. 2 provides the milestone chart for the Open-World Object Detection (OWOD) problem development.



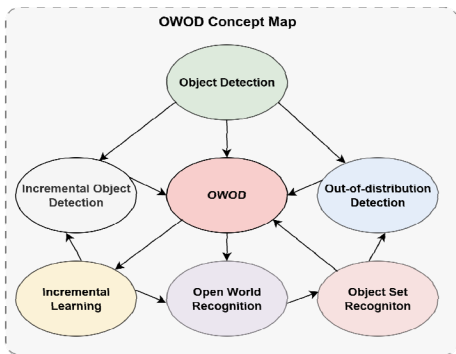
[Fig. 3] Overview of the Overall Process of Open World Object Detection

predicts a centerness-style localization quality and filters overconfident errors via a Gaussian mixture model. These designs reliably deliver higher U-Recall than pseudo-labeling approaches, but at the cost of more false positives; accordingly, many systems add post-hoc checks to confirm that high-objectness candidates are inconsistent with every known class.

search. UC-OWOD [6] partitions novel objects into multiple clusters (e.g., Unknown-1/2) and reports UC metrics. These approaches sharpen novel object discrimination and enable early novel-class grouping but add complexity and show mixed gains over simpler novel object heads.

3.4 Hybrid/other Methods

Methods increasingly leverage external priors and modalities. MAVL and DOWB use vision-language models (e.g., GLIP/CLIP-like backbones) to propose or rescore candidate novel objects; the video-centric STUD exploits temporal consistency and energy filtering to stabilize novel object discovery; other works introduce non-learned or randomized proposals (e.g., selective search, random boxes) to counter bias in region proposals. These strategies typically improve novel object recall, but introduce dependencies on auxiliary models, increase computational and tuning overhead, and may require domain-specific calibration for robust deployment.



[Fig. 4] Concept Map and Related Field of OWOD

3.3 Metric/representation learning

Embed instances in a feature space where previously trained classes form compact, prototype-centered clusters: samples with low prototype affinity or high energy are flagged as novel object. ORE [1] couples contrastive prototype learning with an energy-based decision head; RE-OWOD employs a class-specific “expelling” classifier and augments proposals via selective

3.5 Abbreviations for Table 1

We provide the Abbreviations for Table 1. as follows: mAP = mean Average Precision; U-Recall = Unknown Recall; WI = Wilderness Impact; A-OSE = Absolute Open-Set Error; UDP = Unknown Detection Precision; UDR = Unknown

Detection Recall. T1-T4 denote incremental tasks 1-4.

4. Benchmarks and Evaluation

The prevailing OWOD protocol partitions COCO into four sequential tasks (t_1, t_2, t_3, t_4) using either VOC+COCO or COCO super category splits. Studies report mAP on previously trained classes and U-Recall for novel object detection; many additionally include WI and A-OSE to quantify open-set errors, alongside retention measures across increments. However, heterogeneity in splits and metric sets hinders like-for-like comparison, reinforcing the need for standardized protocols and unified reporting.

5. Open-World Object Segmentation

The open-world setting extends beyond bounding boxes to dense prediction. Under open-world assumptions, both semantic (pixel-wise) and instance (object-wise) segmentation must handle regions belonging to classes unseen during training, marking such regions as novel object and incorporating them once annotations become available.

5.1 Open-World Semantic Segmentation

The objective is to segment known categories while detecting pixels that do not align with any known class and, ideally, separating distinct novel object groups. Practically, this couples (i) anomaly/open-set segmentation (known vs. novel object pixels) with (ii) novel-class discovery (clustering novel object regions). Recent architectures jointly predict closed-set semantics and a novel objectness map, constrain known-class features to improve rejection, and use similarity cues to cluster putative novel objects. Empirical results

indicate strong novel object detection without degrading mIoU on previously trained classes critical in safety-centric applications such as autonomous driving.

5.2 Open-World Instance Segmentation

Hybrid open-world segmentation frameworks combine bottom-up proposals with top-down refinement to mitigate class bias and over-segmentation, enabling generalization to unseen categories. Class-agnostic proposal heads and heuristic region generators further improve recall for novel objects. These approaches naturally extend to open-world panoptic segmentation, which must address both novel “things” and “stuff.”

6. Challenges

Despite significant progress, open-world object detection (and segmentation) remains a challenging problem. We highlight several key challenges that current OWOD algorithms face.

6.1 Unknown bias in region proposals

Proposal generators trained on labeled knowns inherit class bias, missing truly novel objects. Class-agnostic or localization-driven proposals increase U-Recall but risk elevated false positives; achieving robust, class-independent “objectness” without over-triggering remains open.

6.2 Unknown-aware classification & open-set errors

A single “novel object” logit rarely captures the fine boundary between known and novel. Contrastive/prototype methods and statistical filters (e.g., energy, GMM) help, but higher sensitivity to novel objects can raise false alarms on hard knowns; balancing detection and specificity is unresolved.

6.3 Catastrophic forgetting under incremental updates

Adding new classes degrades earlier ones despite replay and distillation. Memory/compute constraints limit rehearsal; OWOD further shares capacity with novel object handling. More efficient, bias-corrected continual learning is needed for long incremental sequences.

7. Future Directions

The field of open-world detection and segmentation is ripe with opportunities. We outline several future directions and research opportunities that could drive progress.

7.1 Task and modality integration

Unify OWOD with open-world segmentation (semantic/panoptic) and 3D perception (LiDAR/point clouds) so systems both localize and delineate novel objects. Leverage self-/semi-supervision, multi-view/video cues, and multi-sensor inputs (depth, radar) to surface class-agnostic objectness and stabilize decisions.

7.2 Representation and generalization

Learn features that separate previously trained classes while preserving capacity for novel-category discovery. Promising avenues include contrastive/meta-learning, episodic training that simulates novel objects, and generative augmentation to expose models to out-of-distribution patterns, with calibrated uncertainty for “I don’t know.”

8. Conclusion

Open-world object detection and segmentation are evolving toward systems that detect, abstain, and learn. Recent advances span pseudo-labeling, class-agnostic proposals, energy- and prototype-

based novel scoring, and early open-world segmentation techniques. Yet challenges remain: novel-class recall, open-set errors, forgetting, and the lack of standardized evaluation.

Progress in OSR, continual learning, VLM pretraining, and multimodal/3D fusion will enable more reliable open-world perception for autonomous driving and embodied robotics.

REFERENCES

- [1] Joseph.K.J, Khan.S, Khan.F.S Balasubramanian. V.N, “Towards open world object detection”, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5830-5840, 2021.
- [2] Gupta.A, Narayan.S, Joseph.K.J, Khan.S, Khan.F.S and Shah.M, “Ow-detr: Open-world detection transformer”, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.9235-9244, 2022.
- [3] Zohar.O, Wang.K.C and Yeung.S, “Prob:Probabilistic objectness for open world object detection”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.11444-11453, 2023.
- [4] Ma.S, Wang.Y, Wei.Y, Fan.J, Li.T.H, Liu.H and Lv.F, “Cat: Localization and identification cascade detection transformer for open-world object detection”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.19681-19690, 2023.
- [5] Zhao.X, Ma.Y, Wang.D, Shen.Y, Qiao.Y and Liu.X, “Revisiting open world object detection”, IEEE Transactions of Circuits and Systems for Video Technology, 34(5), pp.3496-3509, 2023.
- [6] Wu.Z, Lu.Y, Chen.X, Wu.Z, Kang.L and Yu.J, “UC-OWOD: Unknown-classified open world object detection”, In European conference on computer vision., Cham: Springer Nature Switzerland, pp.193-210, 2022.
- [7] Scheirer.W.J, de Rezende Rocha.A, Sapkota.A and Boulton.T.E, “Toward open set recognition”, transactions on pattern analysis and machine intelligence, 35(7), pp.1757-1772, 2012.
- [8] Bendale.A and Boulton.T.E, “Toward open set deep networks”, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp.1563-1572, 2016.
- [9] Sodano.M, Magistri.F, Nunes.L, Behley.J and Stachniss.C, “Open-world semantic segmentation including class similarity”, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.3184-3194, 2024.

[10] Kalluri.T, Wang.W, Wang.H, Chandraker.M, Torresani.L and Tran.D, "Open-world instance segmentation: Top-down learning with bottom-up supervision", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.2693-2703, 2024.

[11] Iqbal.M.A, Yoon.Y.C, Khan.M.U and Kim.S.K, "Improved open world object detection using class-wise feature space learning", IEEE Access, 11, pp.131221-131236, 2023.

[12] Iqbal.M.A, Yoon.Y.C and Kim.S.K, "Redefining Object Detection for Open-World Settings: A Framework for Simultaneous Identification of Known and Unknown Classes", IEEE Access, 2024.

[13] Pershouse.D, Dayoub.F, Miller.D and Sunderhauf.N, "Addressing the challenges of open-world object detection", arXiv preprint arXiv:2303.14930, 2023..

[14] Wang.Y, Yue.Z, Hua.X.S and Zhang.H, "Random boxes are open-world object detectors", In Proceedings of the IEEE/CVF international coference on computer vision, pp.6233-6243, 2023.

[15] Du.X, Wang.X, Gozum.G and Li.Y, "Unknown-aware obje ct detection: Learning what you don't know from videos in the wild", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.13678-13688, 2022.

[16] Liang.W, Xue.F, Liu.Y, Zhong.G and Ming.A, "Unknown sniffer for object detection: Don't turn a blind eye to unknown objects", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.3230-3239, 2023.

[17] Yoshihashi.R, Shao.W, Kawakami.R, You.S, Iida.M and Naemura.T, "Classification-reconstruction learning for open-set recognition", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.4016-4025, 2019.

[18] Peng.C, Zhao.K, and Lovell.B.C, "Faster ilod: Incremental learning for object detectors based on faster rcnn", Pattern recognition letters, 140, pp.109-115, 2020.

[19] Zareian.A, Rosa.K.D, Hu.D.H, and Chang.S.F, "Open-vocabulary object detection using captions", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.14393-14402, 2021.

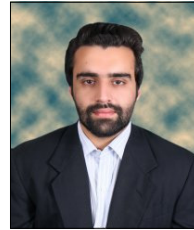
[20] Iqbal, Muhammad Ali, Yeo-Chan Yoon, and Soo Kyun Kim. "Tail-Calibrated Transformer Autoencoding with Prototype-Guided Mining for Open-World Object Detection." Applied Sciences 15.20 (2025): 10918.

[21] Bibi, M., Faseeh, M., Nawaz Khan, A., Rizwan, A. , Waqas Khan, Q., Ahmad, R. and Kim, D.H., 2025. An Auto-Annotation Approach for Object Detection and Depth-Based Distance Estimation in Security and Surveillance Systems. IEEE Access, 13, pp.45240-45251.

[22] Iqbal, Muhammad Ali, Waqas Jadoon, and Soo Kyun Kim. "Synthetic image generation using conditional GAN-provided single-sample face image." Applied Sciences 14, no. 12 (2024): 5049.

무하마드 알리 이크발(Muhammad Ali Iqbal)

[정회원]



- 2016년 6월 : Federal urdu university of Arts, Science & Technology 컴퓨터공학과 (공학학사)
- 2019년 6월 : Comsats university 컴퓨터공학과 (공학석사)

▪ 2022년 9월 ~ 현재 : 제주대학교 컴퓨터공학과 박사과정

<관심분야>

컴퓨터 비전, 기계학습, 딥러닝

한 인 철(Han In Chul)

[준회원]



- 2025년 2월 : 제주대학교 컴퓨터 공학과 (공학학사)
- 2025년 3월 ~ 현재 : 제주대학교 컴퓨터공학과 석사과정

<관심분야>

인공지능, 혼합현실, 컴퓨터 비전, 그래픽 상호작용

김 수 균(Kim Soo Kyun)

[종신회원]



- 2006년 2월 : 고려대학교 컴퓨터 학과 (이학박사)
- 2025년 3월 ~ 현재 : 제주대학교 컴퓨터공학과 교수

<관심분야>

멀티미디어, 패턴 인식, 이미지 전처리, 모바일 그래픽, 기하 모델링, 그래픽 상호작용