2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

OWOD-FSL: Open-World Object Detection via Few-Shot Learning and Dynamic Prototypes

Zhiwei Li[1], Zhiyu Zhang[1], Yang Zhou[1], Jianping Li[2], Tianyu Shen[1], Li Wang[3], Fengli Lu[4], Huaping Liu[5] and Kunfeng Wang[1]

***Abstract*— Open-World Object Detection (OWOD) presents a critical challenge for modern computer vision systems: detecting known classes, identifying unknown objects, and incrementally learning to recognize them over time. However, current approaches have two fundamental limitations: (1) the fixed-dimensional classification head inherently restricts incremental learning capabilities, and (2) heavy reliance on extensive annotated data hinders adaptability in few-shot settings. To address these limitations, we propose OWOD-FSL that integrates dynamic prototype classification head with few-shot learning. At the core of our approach are two major contributions: a dynamic prototype classification head that supplants traditional fixed classifiers with an expandable prototype classifier for scalable class expansion, and a biologically-inspired bi-phase learning strategy that integrates offline prototype generation with incremental learning refinement. Comprehensive experiments on M-OWODB benchmark shows that OWOD-FSL achieves state-of-the-art performance in both unknown class recall (U-Recall) and known class mAP, significantly outperforming existing methods.**

I. Introduction

As a fundamental computer vision task, object detection focuses on accurate localization and classification of objects within images. The rapid evolution of deep learning has enabled widespread applications of this technology in autonomous driving, robotic vision, and video surveillance systems [1]. However, traditional detection methods operate under a closed-world assumption, limiting recognition capabilities to a predefined set of classes [2]. This paradigm presents a fundamental limitation for real-world deployment, where intelligent systems must inevitably handle novel objects absent from training data.

*This work was supported in part by the Beijing Natural Science Foundation under Grant 4252048, in part by the National Natural Science Foundation of China under Grants 62025304 and 62302047, in part by the Fundamental Research Funds for the Central Universities under Grant buctrc202413, and in part by the National Key Laboratory of Hybrid Human-Machine Augmented Intelligence, Xi'an Jiaotong University, under Grant HMHAI-202416.

[1]Zhiwei Li, Zhiyu Zhang, Yang Zhou, Tianyu Shen and Kunfeng Wang are with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China wangkf@mail.buct.edu.cn

[2]Jianping Li is with the School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China lijianping@mail.dlut.edu.cn

[3]Li Wang is with the School of Machanical Engineering, Beijing Institute of Technology, Beijing 100081, China wangli_bit@bit.edu.cn

[4]Fengli Lu is with the School of Computer Science and Engineering, Guangxi Normal University, Guangxi 541004, China lydianlfl@126.com

[5]Huaping Liu is with the College of Computer Science and Technology, Tsinghua University, Beijing 100084, China hpliu@tsinghua.edu.cn

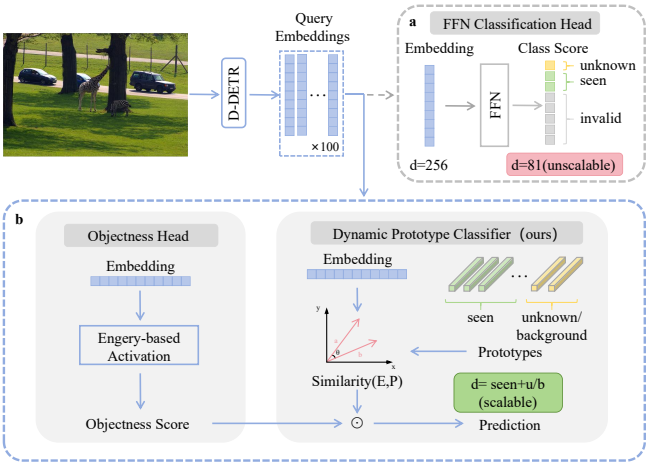

Fig. 1. **Comparative analysis with current OWOD methods.** Current approaches employ fixed-dimensional (d=81) FFN classification heads that inherently lack scalability. Our improved framework introduces a dynamic prototype classifier that enables open-world recognition through embedding-prototype similarity computation (similarity(E,P)). This architecture overcomes the fundamental limitations of fixed-output classifiers by supporting dynamic dimensional expansion, allowing flexible adaptation to novel classes during incremental learning.

To overcome this fundamental limitation, Joseph et al. [3] introduced the open-world object detection (OWOD) paradigm, establishing three critical model requirements: known class detection, unknown object recognition, and continuous expansion of knowledge through incremental learning. This groundbreaking framework has established a new research direction for computer vision systems to adapt to real-world dynamics environments. Current OWOD methodologies predominantly follow two approaches: region proposal-based detection frameworks (e.g., RE-OWOD [4]) and end-to-end Transformer-based methods (e.g., PROB [5]).The former extends the two-stage detection paradigm of Faster R-CNN [6]. The latter employs Deformable DETR [7] for global context encoding and multi-scale feature extraction. Subsequent advancements have yielded improved variants of both Faster R-CNN [8]–[10] and Deformable DETR [11]–[13] architectures, progressively enhancing open-world detection capabilities.

Despite remarkable advances in open-world object detection (OWOD), current methodologies remain constrained by two fundamental limitations that hinder their practical implementation. Firstly, the inherent architectural constraint of fixed-dimensional classification head imposes an artificial upper bound on class capacity [3]. This design choice

results in a superficial incremental learning capability - while models demonstrate competent performance within predetermined class bounds, they fundamentally fail to achieve true incremental learning when confronted with novel classes beyond their predefined scope, as illustrated in Fig. 1(a). This architectural limitation critically undermines the systems' long-term adaptability in real-world dynamic environments.

Secondly, current paradigms exhibit an unsustainable dependence on extensive annotated data, representing a stark departure from the neurobiological efficiency of human visual system. Contemporary neuroscience research has elucidated the remarkable "short-term to long-term memory" consolidation mechanism mediated by hippocampal-cortical interactions [14], [15], which enables rapid class acquisition from minimal exemplars while supporting progressive knowledge refinement. In contrast, current OWOD approaches universally require extensive annotated data for novel classes assimilation, rendering them particularly unsuitable for mission-critical applications where annotated samples are inherently scarce.

To address these limitations, we propose a novel metric learning-based framework incorporating few-shot class-incremental learning, which operates through three synergistic components: (1) A dynamically expandable prototype classification head that supersedes conventional fixed-dimensional classifiers, enabling truly scalable class representation; (2) An offline prototype generation mechanism that exploits structural properties of pre-trained feature spaces to derive novel class prototypes from few-shot samples; (3) An replay-based incremental optimization protocol that coordinates prototype space evolution, ensuring harmonious knowledge integration across learning phases. **We summarize our contributions as follows:**

- We propose a dynamic prototype classification head framework for OWOD, replacing static classifier with extensible prototype classifier to enable sustainable class-incremental learning.
- We develop a biologically-inspired dual-phase prototype learning strategy, combining offline prototype generation with incremental learning refinement, which emulates human memory consolidation processes while eliminating retraining requirements.
- Comprehensive experiments on the M-OWODB benchmark demonstrates state-of-the-art performance, with 51 % mAP and 21.7 % unknown class recall (U-Recall), validated through both quantitative metrics and qualitative analysis.

II. Related Works

A. Open-World Object Detection

The study of open-world object detection (OWOD) stems from a paradigm shift beyond the traditional closed-world assumption. Joseph et al. [3] first systematically formalized the OWOD framework, introducing unknown class detection and incremental learning mechanisms. Their ORE detector, built upon Faster R-CNN, employed contrastive clustering and energy-based scoring for unknown object identification, along with exemplar replay strategy to mitigate catastrophic forgetting.

Subsequent research has expanded this direction. Gupta et al. [11] introduced Transformer into OWOD through Deformable DETR. Their work (OW-DETR) employs pseudo-labeling to handle novel classes. Orr et al. [5] developed PROB that decouples objectness estimation from class-specific prediction by modeling queries as class-agnostic Gaussians distributions. Using Mahalanobis distance for objectness scoring, it selectively retains exemplars to effectively mitigate catastrophic forgetting. Zhao et al. [4] proposed RE-OWOD, an enhanced Faster R-CNN-based framework incorporating a auxiliary Proposal ADvisor (PAD) and a Class-specific Expelling Classifier (CEC), significantly boosting detection robustness in open environments.

Current methods universally assume an upper bound on the maximum number of detectable classes [3]. To overcome this fundamental limitation, our approach introduces a dynamic prototype classification head, enabling unbounded class expansion in incremental learning.

B. Few-Shot Class-Incremental Learning

Few-shot class-incremental learning (FSCIL) aims to learn new classes from limited samples while retaining prior knowledge [16]. Although no unified taxonomy yet exists for FSCIL, several innovative approaches have emerged.

Tao et al. [17] first defined the FSCIL problem and proposed TOPIC, employing Neural Gas (NG) networks to model feature space topology during incremental learning. Zou et al. [18] categorized FSCIL methods into metric learning-based and fine-tuning-based approaches. Metric learning method employs prototype-based similarity metrics (e.g., cosine/Euclidean distance) for classification. Zheng et al. [19] extended this paradigm by introducing class structural regularizer, ensuring feature discriminability through prior knowledge constraints. Hersche et al. [20] approached FSCIL from a hyperdimensional computing perspective, mapping inputs to a quasi-orthogonal prototype space. Their C-FSIL framework combines a frozen feature extractor, a trainable classifier and dynamic memory expansion. Yang et al. [21] identified that direct fine-tuning leads to misalignment between known class features and decision boundaries, exacerbating catastrophic forgetting. Their solution integrates neural collapse theory with few-shot learning, significantly improving classifier generalization.

Current FSCIL methods remain constrained by feature distortion and inflexible prototype representations during incremental learning. Our framework introduces topology-aware initialization and elastic prototype expansion, thereby achieving sustainable incremental learning in open-world settings.

III. Method

We propose OWOD-FSL, a novel framework that integrates metric learning-based dynamic prototype classification head with few-shot class-incremental learning into

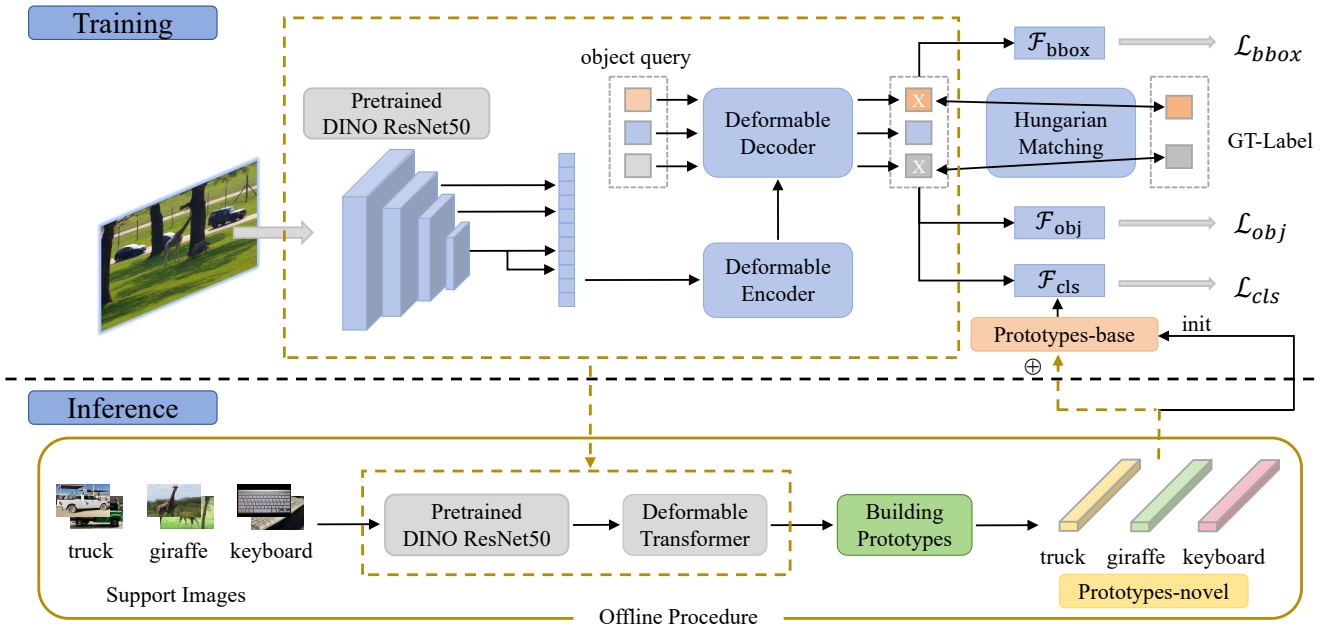

Fig. 2. **Overview of the proposed OWOD-FSL for open-world object detection.** The proposed method builds upon Deformable DETR framework with dynamic prototype classification head and a dual-phase learning strategy. The model first establishes base detection capabilities via multi-task optimization, then progresses through prototype construction (generating novel prototypes from few-shot samples with frozen feature extractors) and incremental learning (integrating novel classes while preserving unknown detection). This architecture achieves 51% mAP and 21.7% U-Recall on M-OWOD benchmark through deformable attention and dynamic prototypes. Diagram notations: "⊕" denotes prototype concatenation during inference, while "init" marks prototype repository initialization for incremental learning.

OWOD. Section III-A details our architectural design, which builds upon Deformable DETR to establish robust feature representations. Notably, we develop a dual-phase prototype learning strategy specifically designed for incremental learning in open-world settings. Section III-B introduces our dynamic prototype classification head, which implements metric learning-based open-world detection through an expandable prototype repository. Section III-C further presents the incremental learning mechanism, where an active sample selection techniques combined with exemplar replay strategy effectively mitigates catastrophic forgetting.

A. Overall Architecture

Fig. 2 shows the overall architecture of OWOD-FSL. The proposed OWOD-FSL incorporates a dynamic prototype classification head into the Deformable DETR framework and employs a dual-phase prototype learning strategy to achieve a human-like "short-term to long-term memory" incremental learning mechanism. The training pipeline consists of three coordinated stages: (1) base model training, (2) offline prototype generation, and (3) incremental learning optimization, which work synergistically to address dynamic learning requirements in open-world settings.

OWOD-FSL builds upon a pre-trained DINO ResNet50 backbone, adapting the Deformable DETR framework for open-world settings. The standard Deformable DETR generates a set of embeddings per image, each processed by detection heads to produce final predictions. Following PROB's design paradigm, we decouple objectness estimation from category classification through dedicated detection heads for "objectness score" and "class probability", complemented by a regression head for bounding box coordinates.

The model undergoes initial training on base classes using a multi-task joint optimization strategy. This combines b-box regression loss ($\mathcal{L}_{bbox}$), objectness loss ($\mathcal{L}_{obj}$), and classification loss ($\mathcal{L}_{cls}$) to simultaneously develop: (1) precise recognition capability for known classes and (2) effective detection capacity for unknown classes. Crucially, the deformable attention mechanism significantly enhances multi-scale object detection performance.

Upon detecting unknown classes, the system enters an offline prototype generation phase. This phase leverages the structural properties of the pre-trained feature space by freezing the feature extractor parameters, computing mean feature vectors from few-shot support images, which is subsequently stored in a novel class prototype repository (Prototype-novel). During inference, the model enhances its detection capability by jointly leveraging the novel-class and base-class prototype repositories.

As novel class data accumulates, the system activates the incremental learning phase by initializing with offline-generated prototypes, unfreezing network parameters for end-to-end fine-tuning. To mitigate catastrophic forgetting, the optimization process incorporates exemplar replay mechanism that maintains a small set of exemplars from known classes. Through this incremental learning process, the model effectively assimilates novel classes into its known class detection capability while preserving its capacity to identify unknown classes.

The dual-phase prototype learning strategy, comprising

offline prototype generation and incremental learning optimization, operates continuously throughout the detector's lifecycle. This approach enables the detector to progressively update its knowledge base with new information while maintaining previously acquired detection capabilities, achieving sustainable open-world object detection performance.

B. Dynamic Prototype-based Classification

Conventional object detection models typically employ fixed-dimensional linear classification heads, where the weight matrix dimensionality is constrained by predefined class quantities. This design exhibits two fundamental limitations when applied to open-world object detection: (1) the requirement for architectural modification of the weight matrix when introducing novel classes, lacking scalability for incremental learning; and (2) vulnerability to parameter overload in few-shot scenarios. To address these challenges, we propose a metric learning-based dynamic prototype classification head that replaces fixed linear classification head with extensible prototype vectors.

For each object instance, the prototype feature is computed as the mean feature vector extracted from prediction boxes matching ground truth annotations. Subsequent class-level prototypes are obtained by averaging features across all instances of each class. Notably, base-class prototypes are learned from the entire training set rather than constructed from support images, which are exclusively reserved for novel class representation.

1) Prototype Classification Head: Given embeddigs $\mathbf{z} \in \mathbb{R}^d$ generated by extractor f_θ, we construct a dynamic prototype repository $\mathcal{P} = \{\mathbf{p}_c\}_{c=1}^{N}$, where each prototype vector $\mathbf{p}_c \in \mathbb{R}^d$ corresponds to a specific class. Classification probabilities are computed via cosine similarity between embeddings and prototypes:

$$s(y_c|z) = \sigma(\alpha \cdot \cos(z, p_c)) = \frac{1}{1 + \exp(-\alpha \cdot \frac{z^T p_c}{|z||p_c|})} \quad (1)$$

where $\sigma(\cdot)$ denotes the sigmoid activation function and $\alpha \in \mathbb{R}^+$ represents a learnable temperature parameter that scales the logit distribution. Both embedding and prototype vectors undergo ℓ_2-normalization, ensuring similarity measurement in unit hypersphere space. This design offers significant advantages: when introducing new class y_{c+1}, simply augmenting $\mathcal{P}$ with corresponding prototype vectors suffices, eliminating architectural modifications.

2) Dual-Phase Prototype Learning Strategy: To achieve both few-shot learning and incremental learning, we devise a dual-phase prototype learning strategy that effectively balances feature space stability with model adaptability.

Offline prototype generation. This phase constructs initial prototypes for novel classes through parameter-free methods. Given K support samples $\{\mathbf{x}_i^{\text{novel}}\}_{i=1}^{K}$ for novel class y_{novel}, the initial prototype is computed as:

$$\mathbf{p}_{\text{novel}} = \frac{1}{K} \sum_{i=1}^{K} f_\theta(\mathbf{x}_i^{\text{novel}}) \quad (2)$$

with the extractor f_θ remaining frozen to preserve feature space integrity. Pretrained features maintain intra-class compactness and inter-class distinguishability even when $K = 1$.

When novel class samples reach threshold H, the incremental learning phase initiates end-to-end model optimization. This phase first populates the prototype repository with offline-generated prototypes, employing momentum updates ($\alpha = 0.9$) for smoothing and prototype orthogonality constraints to maintain feature space separation. The optimization follows a progressive unfreezing strategy: initially updating only classification head parameters (lr $= 5 \times 10^{-4}$), then deeper backbone layers (lr $= 1 \times 10^{-4}$), and finally the entire network, preventing disruptive feature space alterations.

C. Active Sample Selection for Incremental Learning

In OWOD scenarios, models must incrementally assimilate novel classes while preserving existing knowledge. Conventional exemplar replay methods typically employ random sampling for historical data retention, suffering from two critical limitations: (1) inability to discriminate sample importance and (2) difficulty in balancing knowledge transfer between base and novel classes. Our solution introduces an objectness score-based active sample selection mechanism.

To enhance OWOD performance, we actively select instances with low/high objectness scores as exemplars. Low-objectness instances represent challenging cases (e.g., occluded or truncated objects) that improve boundary recognition, while high-objectness samples embody class representative features. After base class training on dataset $\mathcal{D}$, we compute objectness probabilities for matched query embeddings, selecting 25 highest/lowest-scoring instances per class. The model then undergoes incremental learning optimization using these retained samples alongside novel class data.

IV. Experiments

We conduct comprehensive evaluation on the M-OWOD benchmark, establishing performance comparisons with reported OWOD methods while providing detailed architectural ablations and qualitative analyses. Furthermore, we perform multiple randomized sampling trials across each subtask to evaluate few-shot detection capabilities.

A. Experimental Setup

Dataset. Following the protocol of Joseph et al. [3], we conduct comprehensive evaluation using the "superclass-mixed OWOD benchmark" (M-OWODB), which innovatively integrates MS-COCO and PASCAL VOC datasets to construct four progressive tasks $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_4\}$. Each task's training set is constructed from MS-COCO and PASCAL VOC training images, while the MS-COCO validation set and PASCAL VOC test set serve for evaluation. Crucially, when learning task $\mathcal{T}_n$, all previously encountered classes are treated as known classes, effectively simulating real-world incremental learning requirements. For few-shot evaluation, we ensure statistical reliability through 10 randomized trials

TABLE I

THE COMPARATIVE EXPERIMENTAL RESULTS ON THE OWOD BENCHMARK. THE COMPARISON IS SHOWN IN TERMS OF UNKNOWN CLASS RECALL (U-RECALL) AND KNOWN CLASS MAP@0.5 (FOR PREVIOUS, CURRENT, AND ALL KNOWN OBJECTS). SYMBOLS ↑ /↓ DENOTE HIGHER/LOWER VALUES INDICATING BETTER PERFORMANCE RESPECTIVELY.

Method	Task 1		Task 2				Task 3				Task 4		
	U-Recall	mAP(↑)	U-Recall	mAP(↑)			U-Recall	mAP(↑)			mAP(↑)		
	(↑)	Current known	(↑)	Previously known	Current known	Both	(↑)	Previously known	Current known	Both	Previously known	Current known	Both
ORE [3]	4.9	56.0	2.9	52.7	26.0	39.4	3.9	38.2	12.7	29.7	29.6	12.4	25.3
UC-OWOD [8]	2.4	50.7	3.4	33.1	30.5	31.8	8.7	28.8	16.3	24.6	25.6	12.9	23.2
2B-OCD [9]	12.1	56.4	9.4	51.6	25.3	38.5	11.7	37.2	13.2	29.2	30.0	13.3	25.8
OW-DETR [11]	7.5	59.2	6.2	53.6	33.5	42.9	5.7	38.3	15.8	30.8	31.4	17.1	27.8
Fast-OWDETR [22]	9.2	56.6	8.8	51.3	28.6	39.4	7.8	39.2	15.7	32.2	28.2	11.4	25.0
Open World DETR [12]	21.0	**59.9**	15.7	51.8	36.4	44.1	17.4	38.9	24.7	34.2	32.0	19.7	29.0
OCPL [10]	8.3	56.6	7.7	50.7	27.5	39.1	11.9	38.6	14.7	30.7	30.8	14.4	26.7
PROB [5]	19.4	59.5	17.4	**55.7**	32.2	44.0	19.6	43.0	22.2	36.0	35.7	18.9	31.5
CAT [13]	21.8	59.9	18.6	54.0	33.6	43.8	23.9	42.1	19.8	34.7	35.1	17.1	30.6
RE-OWOD [4]	9.1	59.7	9.9	54.1	37.3	45.6	11.4	43.1	24.6	37.6	38.0	**28.7**	35.7
OWOD-FSL	**22.4**	59.2	**19.6**	54.6	**38.1**	**46.4**	**24.3**	**45.6**	**30.6**	**40.6**	**39.3**	26.5	**36.1**

TABLE II

UNKNOWN OBJECT CONFUSION ON M-OWODB. THE COMPARISON MEASURED BY UNKNOWN CLASS RECALL (U-RECALL), WILDNESS INDEX (WI), AND ABSOLUTE OPEN-SET ERROR (A-OSE).

Method	Task 1			Task 2			Task 3		
	U-Recall (↑)	WI (↓)	A-OSE (↓)	U-Recall (↑)	WI (↓)	A-OSE (↓)	U-Recall (↑)	WI (↓)	A-OSE (↓)
ORE [3]	4.9	0.0621	10459	2.9	0.0282	10445	3.9	0.0211	7990
UC-OWOD [8]	2.4	**0.0136**	9294	3.4	0.0116	5602	8.7	0.0073	3801
2B-OCD [9]	12.1	0.0481	–	9.4	0.0160	–	11.7	0.0137	–
OW-DETR [11]	7.5	0.0571	10240	6.2	0.0278	8441	5.7	0.0156	6803
Open World DETR [12]	21.0	0.0549	5909	15.7	0.0210	4378	17.4	0.0133	2641
OCPL [10]	8.3	0.0423	5670	7.7	0.0220	2690	11.9	0.0162	5166
PROB [5]	19.4	0.0569	5195	17.4	0.0344	6452	19.6	0.0151	2641
CAT [13]	21.8	0.0581	7070	18.6	0.0263	5902	23.9	0.0177	5189
RE-OWOD [4]	9.1	0.0449	–	9.9	0.0331	–	11.4	0.0241	–
OWOD-FSL	**22.4**	0.0396	**1210**	**19.6**	**0.0112**	**383**	**24.3**	**0.0067**	**295**

with different seeds, reporting averaged results across three sampling strategies (5-shot, 10-shot, and 30-shot).

Metrics. Our evaluation framework examines three key aspects: (1) Known class detection performance measured by mean Average Precision (mAP@0.5), with further breakdowns into base classes and novel classes to assess incremental learning; (2) Unknown class detection evaluated through standard unknown class recall (U-Recall@IoU$\geq$0.5), supplemented by Absolute Open-Set Error (A-OSE) and Wilderness Impact (WI) indices to quantify unknown class misclassification and known class interference respectively; (3) Few-shot learning capability assessed via mAP and Recall metrics under limited supervision.

Implementation Details. The architecture builds upon Deformable DETR with a DINO-pretrained ResNet-50 FPN backbone extracting multi-scale features. These features are flattened and processed through a Deformable Transformer to generate 256-dimensional query embeddings. Our detection head employs a tripartite design: (i) a bbox regression head, (ii) a dynamic prototype classification head, and (iii) an objectness prediction head. To reduce false positives, we construct background prototypes using typical background classes (e.g., sky, road) from COCOStuff. Few-shot support sets are created through randomized sampling from each subtask's training data.

B. Open-World Object Detection Performance

1) Main Experimental Results: Table I presents a comprehensive comparison between OWOD-FSL and other reported OWOD methods on the M-OWODB benchmark. Experimental results demonstrate that OWOD-FSL achieves consistent improvements across multiple key metrics, notably exhibiting superior performance in unknown class recall (U-Recall) and significant gains in mAP for novel classes. These findings confirm that our dynamic prototype classification head, coupled with incremental learning strategy, not only preserves detection accuracy on known classes but also delivers a marked advancement in recognizing unknown objects.

In unknown object detection, OWOD-FSL achieves the highest U-Recall scores across all four sequential tasks, with performance of 22.4% and 24.3% in Task 1 and Task 3 respectively - surpassing the second-best methods by 0.6 and 0.4 percentage points. Compared to the Deformable DETR-based PROB method, our approach demonstrates consistent improvements of 3.0, 2.2, and 4.7 percentage points in U-Recall across the tasks, with Task 4 excluded from this metric

Fig. 3. **Unknown detection comparison: OWOD-FSL (top) vs PROB (bottom).** Yellow/blue boxes mark known/unknown objects. OWOD-FSL detects more unknowns (zebras, hydrants, stop signs) with higher confidence (e.g., skateboards), proving better open-world robustness.

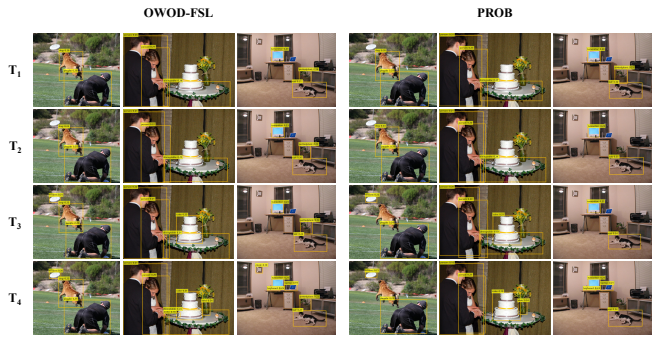

Fig. 4. **OWOD-FSL vs PROB Incremental Detection Performance.** OWOD-FSL maintains stable accuracy on known classes while successfully detecting novel objects (frisbee, cake/knife/vase, clock/keyboard/laptop across tasks). PROB shows catastrophic forgetting, failing to detect previously learned objects (dog, diningtable, pottedplant).

as all 80 classes become known.

For known class detection, OWOD-FSL maintains competitive performance with 59.2% mAP in Task 1 and exhibits superior incremental learning capabilities in subsequent incremental tasks. Notably, our method achieves 36.1% mAP in Task 4, outperforming PROB by 4.6 percentage points.

Table II further examines model performance in unknown class confusion. OWOD-FSL achieves substantially lower Absolute Open-Set Error (A-OSE) values of 1210, 383, and 295 across three evaluation tasks. Similarly, our method obtains excellent Wilderness Impact (WI) indices of 0.0396, 0.0112, and 0.0067, confirming its effectiveness in minimizing unknown class interference.

2) Qualitative Results: We conduct visual analysis on the MS-COCO test set to evaluate OWOD-FSL's performance in unknown class detection and incremental learning, using PROB as the baseline. Fig. 3 demonstrates OWOD-FSL's superior unknown class detection, correctly identifying zebra, fire hydrant, and skateboard with higher confidence than PROB. The incremental learning analysis in Fig. 4 reveals OWOD-FSL's advantages in both incremental learning capacity for novel classes and examining catastrophic forgetting in known classes. While PROB fails to detect newly objects like knives and clocks, and shows catastrophic forgetting of previously seen classes (e.g., dogs in Task 3).

TABLE III

ABLATION STUDY OF DYNAMIC PROTOTYPE CLASSIFICATION HEAD COMPONENTS.

Dynamic Prototype Classifier Head	Prototype Init	Background Prototype	Base mAP (%)	Novel mAP (%)
✓			42.7	29.2
✓	✓		42.5	32.6
✓	✓	✓	**54.6**	**38.1**

TABLE IV

INCREMENTAL DETECTION PERFORMANCE COMPARISON UNDER FEW-SHOT SETTINGS.

Task IDs	Task 2		Task 3		Task 4	
	mAP	Recall	mAP	Recall	mAP	Recall
5-shot	5.4	8.1	4.9	7.5	4.0	6.6
10-shot	6.6	9.9	5.6	8.6	5.1	7.7
30-shot	6.7	10.2	5.5	8.5	5.2	7.9

C. Ablation Studies

Systematic ablation studies in Table III investigate the contributions of each component. The base dynamic prototype mechanism achieves 42.7% and 29.2% mAP for base and novel classes respectively in $\mathcal{T}_1$, validating its effectiveness. The support set-based prototype initialization boosts novel class performance to 32.6% mAP, while background prototype modeling further elevates base and novel class mAP to 54.6% and 38.1%. These results reveal the synergistic effects among components: dynamic prototypes provide scalability, proper initialization ensures optimization stability, and background modeling refines feature space structure.

D. Few-shot Learning Performance

To thoroughly assess the model's few-shot learning capability, we conduct systematic evaluations under three settings (5-shot, 10-shot, and 30-shot) as shown in Table IV. Each experimental configuration undergoes 10 randomized trials with averaged results reported. The results reveal that under the most challenging 5-shot condition, the model maintains mAP scores of 5.4%, 4.9%, and 4.0% in Tasks 2-4 respectively. Performance shows steady improvement with increased samples - the 10-shot setting yields an average 1.2 percentage point gain over 5-shot, while 30-shot demonstrates performance saturation with diminishing returns compared to 10-shot, indicating the model approaches its learning potential upper bound with limited samples.

V. CONCLUSION

This paper presents OWOD-FSL, an novel open-world object detection method. Our model effectively addresses fundamental limitations in incremental learning and data efficiency. Future work will focus on enhancing few-shot detection performance and exploring integration with embodied intelligent systems to validate the framework's potential in embodied evolution.

REFERENCES

[1] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[2] Y. Li, Y. Wang, W. Wang, D. Lin, B. Li, and K.-H. Yap, "Open world object detection: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[3] K. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, "Towards open world object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 5830–5840.

[4] X. Zhao, Y. Ma, D. Wang, Y. Shen, Y. Qiao, and X. Liu, "Revisiting open world object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 5, pp. 3496–3509, 2023.

[5] O. Zohar, K.-C. Wang, and S. Yeung, "Prob: Probabilistic objectness for open world object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11 444–11 453.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[7] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection," *arXiv preprint arXiv:2010.04159*, 2020.

[8] Z. Wu, Y. Lu, X. Chen, Z. Wu, L. Kang, and J. Yu, "Uc-owod: Unknown-classified open world object detection," in *European conference on computer vision*, Springer, 2022, pp. 193–210.

[9] Y. Wu, X. Zhao, Y. Ma, D. Wang, and X. Liu, "Two-branch objectness-centric open world detection," in *Proceedings of the 3rd international workshop on human-centric multimedia analysis*, 2022, pp. 35–40.

[10] J. Yu, L. Ma, Z. Li, Y. Peng, and S. Xie, "Open-world object detection via discriminative class prototype learning," *arXiv preprint arXiv:2302.11757*, 2023.

[11] A. Gupta, S. Narayan, K. Joseph, S. Khan, F. S. Khan, and M. Shah, "Ow-detr: Open-world detection transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 9235–9244.

[12] N. Dong, Y. Zhang, M. Ding, and G. H. Lee, "Open world detr: Transformer based open world object detection," *arXiv preprint arXiv:2212.02969*, 2022.

[13] S. Ma, Y. Wang, Y. Wei, *et al.*, "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*, 2023, pp. 19 681–19 690.

[14] P. W. Frankland and B. Bontempi, "The organization of recent and remote memories," *Nature reviews neuroscience*, vol. 6, no. 2, pp. 119–130, 2005.

[15] A. C. Toader, J. M. Regalado, Y. R. Li, *et al.*, "Anteromedial thalamus gates the selection and stabilization of long-term memories," *Cell*, vol. 186, no. 7, pp. 1369–1381, 2023.

[16] S. Tian, L. Li, W. Li, H. Ran, X. Ning, and P. Tiwari, "A survey on few-shot class-incremental learning," *Neural Networks*, vol. 169, pp. 307–324, 2024.

[17] X. Tao, X. Hong, X. Chang, S. Dong, X. Wei, and Y. Gong, "Few-shot class-incremental learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 183–12 192.

[18] Y. Zou, S. Zhang, Y. Li, and R. Li, "Margin-based few-shot class-incremental learning with class-level overfitting mitigation," *Advances in neural information processing systems*, vol. 35, pp. 27 267–27 279, 2022.

[19] G. Zheng and A. Zhang, "Few-shot class-incremental learning with meta-learned class structures," in *2021 International Conference on Data Mining Workshops (ICDMW)*, IEEE, 2021, pp. 421–430.

[20] M. Hersche, G. Karunaratne, G. Cherubini, L. Benini, A. Sebastian, and A. Rahimi, "Constrained few-shot class-incremental learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 9057–9067.

[21] Y. Yang, H. Yuan, X. Li, Z. Lin, P. Torr, and D. Tao, "Neural collapse inspired feature-classifier alignment for few-shot class incremental learning," *arXiv preprint arXiv:2302.03004*, 2023.

[22] X. Chen, "Fast owdetr: Transformer for open world object detection," Ph.D. dissertation, Nanyang Technological University, 2022.