NEIGHBORHOOD-AWARE REPRESENTATION LEARNING FOR OPEN WORLD OBJECT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Open World Object Detection (OWOD) extends conventional object detection beyond a fixed label space, enabling unknown object discovery and incremental learning of new classes. Representative objectness-based approaches commonly rely on global probabilistic foreground modeling or reduce the coupling between objectness and known class prediction. However, they typically treat known class discrimination and foreground modeling separately, leaving the connection between the two objectives underexplored. We therefore introduce Prototype Relations for Open World object detection (PROW), a neighborhood-aware representation learning framework that exploits neighborhood relations among known class prototypes as a shared structure for local class discrimination and multi-component foreground modeling. PROW maintains a dynamic graph over class prototypes, with neighborhood relations updated from matched positive queries. The graph identifies locally confusable prototypes as informative negatives for local class discrimination, while the same graph organizes class prototypes into multiple neighborhood-conditioned foreground components. For each query, PROW selects the most compatible foreground component to estimate class-independent objectness. In subsequent tasks, confidence-filtered pseudo-labels from the previous detector and replay samples provide complementary supervision to preserve previously learned knowledge. Compared with the best prior results, PROW achieves absolute gains of 2.4–8.1 points in unknown object discovery, with a 39% reduction in confusion between known and unknown objects.

1 INTRODUCTION

Conventional object detection is formulated under a closed-world assumption, where the set of object classes encountered at test time is fixed and known during training (Ren et al., 2017; Carion et al., 2020; Dai et al., 2023; Joseph et al., 2021). Open World Object Detection (OWOD) (Joseph et al., 2021; Gupta et al., 2022; Zohar et al., 2023; Sun et al., 2024; Li et al., 2025; Ma et al., 2024) relaxes this assumption by requiring a detector to detect and classify objects from known classes, identify unknown objects, and incrementally learn newly introduced classes without catastrophic forgetting. Following the standard OWOD setting, detected unknown objects can be annotated by an oracle and subsequently introduced as known classes, enabling the detector to progressively expand its label space over successive tasks.

A prominent direction in OWOD learns class-independent objectness from known objects to support unknown object discovery (Gupta et al., 2022; Zohar et al., 2023; Sun et al., 2024). Representative approaches model objectness through probabilistic foreground representations (Zohar et al., 2023) or reduce the dependence and interference between objectness and class prediction (Wu et al., 2022a; Sun et al., 2024; Inoue et al., 2025). Figure 1 summarizes these representative designs and the motivation behind PROW.

Despite this progress, two limitations motivate our representation learning design. First, a single global foreground model may be too coarse to capture the heterogeneous structure of foreground representations. Second, objectness modeling and class discrimination are typically addressed separately, so relational structure among known classes is not explicitly reused across the two objectives. Beyond these representation limitations, incremental learning of newly introduced classes can

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

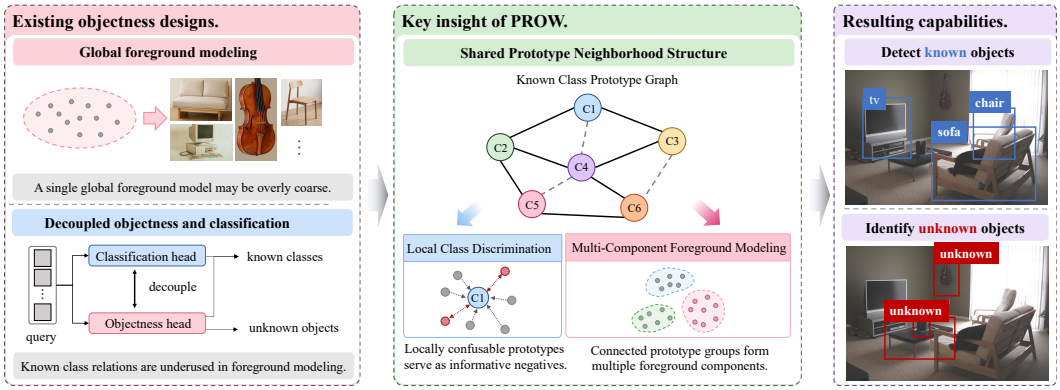


Figure 1: Motivation and key insight of PROW. Representative objectness-based OWOD designs commonly rely on global foreground modeling or decouple objectness from classification. PROW instead uses prototype neighborhoods as a shared relational structure for local class discrimination and multi-component foreground modeling, supporting both known object detection and unknown object discovery.

disrupt previously learned representations and decision boundaries, leading to catastrophic forgetting (McCloskey & Cohen, 1989; Goodfellow et al., 2014).

Our key insight is that neighborhood relations among known class prototypes provide a shared structure for local class discrimination and multi-component foreground modeling. Based on this insight, we propose Prototype Relations for Open World object detection (PROW), a neighborhood-aware representation learning framework. PROW maintains learnable prototypes for known classes and constructs a dynamic prototype graph that captures their evolving local relationships during training.

PROW uses the prototype graph in two complementary ways. For local class discrimination, the graph focuses metric supervision on locally confusable prototypes used as informative negatives, rather than treating all class pairs uniformly. For foreground modeling, the same neighborhood structure organizes class prototypes into multiple foreground components, providing local alternatives to a single global foreground model. For each query, PROW selects the component with the minimum variance-normalized deviation and uses this match for class-independent objectness estimation. During incremental learning, confidence-filtered predictions from the preceding detector supplement supervision for previously learned classes, while replay samples from preceding tasks further mitigate forgetting.

We summarize our contributions as follows:

- We introduce PROW, a neighborhood-aware representation learning framework that uses prototype neighborhoods as a shared relational structure to connect local class discrimination with multi-component foreground modeling in OWOD.
- PROW maintains a dynamic prototype graph that focuses local discrimination on locally confusable prototypes as informative negatives and organizes class prototypes into multiple foreground components. For each query, PROW selects the component with the minimum variance-normalized deviation for class-independent objectness estimation.
- Extensive experiments demonstrate that PROW achieves state-of-the-art OWOD performance. PROW achieves improvements ranging from 2.4 to 8.1 on unknown recall and reduces confusion between known and unknown objects by 39%.

2 Related Work

2.1 Open World Object Detection

Open World Object Detection (OWOD) requires a detector to recognize known classes, identify objects outside the current label space, and incrementally learn newly introduced classes without for-

getting previously acquired knowledge (Joseph et al., 2021). ORE (Joseph et al., 2021) established the standard OWOD protocol by combining class-agnostic proposals with contrastive clustering and energy-based unknown identification. Subsequent DETR-based methods improve unknown object discovery through strategies such as unmatched-query mining, cascaded localization and identification, and annealed label transfer (Gupta et al., 2022; Ma et al., 2023a;b).

Another line of work focuses on class-independent objectness. PROB (Zohar et al., 2023) estimates objectness by probabilistically modeling foreground embeddings, while Two-Branch OWOD (Wu et al., 2022a) and OrthogonalDet (Sun et al., 2024) reduce the dependence of objectness estimation on class prediction. Decoupled PROB (Inoue et al., 2025) further revisits query initialization and decoder supervision to reduce learning interference between objectness and class prediction. More recently, OWOBJ (Zhang et al., 2025) explicitly models objectness together with category information to improve class-agnostic objectness estimation. Knowledge preservation across incremental tasks is commonly addressed through distillation, pseudo-labeling, or replay (Liu et al., 2023; Feng et al., 2022; Kim et al., 2024).

Different from these directions, PROW focuses on how relational structure among known class prototypes can be reused as a shared structure for local class discrimination and foreground modeling.

2.2 Structured Class Representations and Foreground Modeling

Class prototypes provide compact representations of known classes and have been explored in OWOD to improve inter-class discrimination. OCPL (Yu et al., 2022) learns discriminative class prototypes to enhance the separation between known and unknown representations. Hyp-OW (Doan et al., 2024) further exploits hierarchical structure in hyperbolic space with superclass information to organize known class representations. These approaches demonstrate the benefit of structured class representations, but primarily exploit such structure for representation organization or discrimination.

Foreground-oriented methods address a complementary problem. PROB (Zohar et al., 2023) models foreground embeddings with a single class-agnostic multivariate Gaussian for objectness estimation, whereas OrthogonalDet (Sun et al., 2024) reduces the correlation between objectness and class information. OWOBJ (Zhang et al., 2025) further improves probabilistic objectness modeling by jointly considering objectness and category information. Although these methods improve foreground and objectness modeling, they do not explicitly reuse local neighborhood relations among known class prototypes as a shared structure for both class discrimination and foreground representation.

PROW connects these two directions by learning a dynamic graph of known class prototypes and reusing its neighborhood relations for informative negative learning and multi-component foreground modeling. The resulting shared relational structure connects local class discrimination with class-independent objectness estimation rather than treating the two objectives independently.

3 Method

3.1 Problem Setting and Overview

Open World Object Detection (OWOD) proceeds over a sequence of incremental tasks. In task t, the detector is trained on a dataset $\mathcal{D}^t$ with the current set of known classes $\mathcal{K}^t = \{1, \ldots, C_t\}$, while objects outside the current label space may appear without annotations. The detector is required to detect and classify known objects, identify unknown objects, and retain previously learned knowledge as the label space expands. After task t, an oracle may annotate previously unknown classes and introduce them as known classes in task $t + 1$.

PROW is a neighborhood-aware representation learning framework built on Deformable DETR (Zhu et al., 2021). As illustrated in Figure 2, PROW maintains learnable prototypes for known classes and constructs a dynamic prototype graph from matched positive queries. The resulting graph provides a shared relational structure for two complementary objectives: local class discrimination and multi-component foreground modeling. For local class discrimination, the graph uses locally confusable prototypes as informative negatives. For foreground modeling, the same graph organizes class prototypes into multiple foreground components for class-independent objectness estimation.

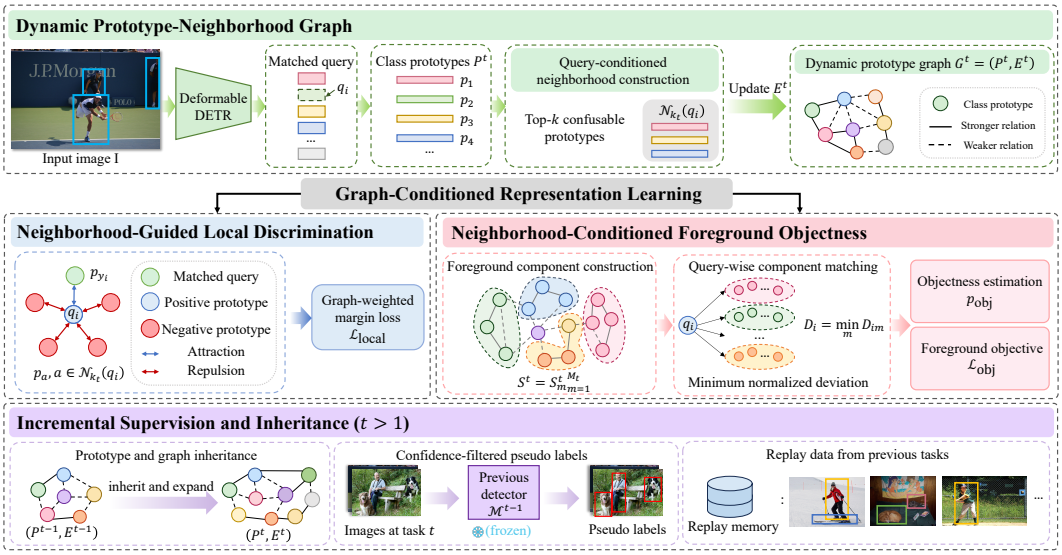


Figure 2: Overview of PROW. PROW first constructs a dynamic prototype-neighborhood graph from matched positive queries. The learned neighborhood structure supports neighborhood-guided local discrimination by selecting locally confusable prototypes and neighborhood-conditioned foreground objectness by organizing class prototypes into multiple foreground components. In subsequent tasks, previously learned prototypes and graph relations are inherited, while pseudo-label supervision and replay help preserve previously learned knowledge.

Following PROB (Zohar et al., 2023), the resulting objectness score is combined with the detector's class scores at inference. During incremental learning, previously learned prototypes and graph relations are inherited across tasks. Confidence-filtered predictions from the previous detector supplement supervision for previously learned classes, while replay samples further mitigate forgetting.

3.2 Dynamic Prototype-Neighborhood Graph

Let $C_t = |\mathcal{K}^t|$ denote the number of known classes in task t. PROW maintains a prototype matrix

$$P^t = [p_1, \ldots, p_{C_t}]^\top \in \mathbb{R}^{C_t \times d}, \tag{1}$$

where each prototype provides a class reference for decoder query embeddings. In task 1, the prototypes are randomly initialized. In subsequent tasks, prototypes of previously learned classes are inherited from the preceding task and continue to be optimized, while prototypes of newly introduced classes are also randomly initialized. All prototypes are learnable parameters and are jointly optimized with the detector by back-propagation.

As illustrated in Figure 2, PROW dynamically models neighborhood relations among known class prototypes using matched positive queries. We maintain a symmetric weighted adjacency matrix $E^t \in [0,1]^{C_t \times C_t}$, where each entry records the neighborhood relation between a pair of class prototypes. In task 1, the graph is initialized as $E^{1,(0)} = I_{C_1}$. In subsequent tasks, the existing graph is inherited from task $t-1$ and expanded for newly introduced classes.

For each matched positive query q_i with class label y_i, the label y_i determines the class anchor for graph updating. We select the $k_t = \min(k, C_t - 1)$ nearest non-matching prototypes to the query:

$$\mathcal{N}_{k_t}(q_i) = \text{TopK}_{c \in \mathcal{K}^t \setminus \{y_i\}} \left(-\|q_i - p_c\|_2^2; k_t\right), \qquad k_t = \min(k, C_t - 1). \tag{2}$$

The selected prototypes constitute a local neighborhood of class references that are most similar to the current query and therefore more likely to be confused with its matched class.

The relations associated with the class anchor are updated as

$$E^t_{y_i,c} \leftarrow \begin{cases} 1, & c \in \mathcal{N}_{k_t}(q_i), \\ \epsilon E^t_{y_i,c}, & c \in \mathcal{K}^t \setminus \left(\{y_i\} \cup \mathcal{N}_{k_t}(q_i)\right), \end{cases} \qquad 0 < \epsilon < 1. \tag{3}$$

The corresponding entries E^t_{c,y_i} are updated symmetrically, while the diagonal entries remain one and relations unrelated to the current class anchor remain unchanged. Thus, neighborhood relations supported by the current query are refreshed, whereas previously established relations gradually decay when they are not supported by new queries from the same anchor class.

The prototypes evolve through gradient optimization together with the detector, while the adjacency matrix is explicitly updated from the observed relations between queries and prototypes. The resulting graph $G^t = \langle P^t, E^t \rangle$ provides the shared relational structure used by local class discrimination and multi-component foreground modeling.

3.3 Graph-Conditioned Representation Learning

As illustrated in Figure 2, the dynamic prototype graph provides a shared relational structure for local class discrimination and multi-component foreground modeling.

Neighborhood-Guided Local Discrimination. For a matched positive query q_i with class label y_i, the corresponding prototype p_{y_i} serves as the positive reference. The Top-k neighborhood $\mathcal{N}_{k_t}(q_i)$ defined in Equation equation 2 provides locally confusable negative prototypes. We weight each negative according to its learned graph relation $E^t_{y_i a}$ and define the local discrimination objective as

$$\mathcal{L}_{\text{local}} = \frac{1}{N_+} \sum_{i=1}^{N_+} \sum_{a \in \mathcal{N}_{k_t}(q_i)} E^t_{y_i a} \left[\|q_i - p_{y_i}\|_2^2 - \|q_i - p_a\|_2^2 + \alpha \right]_+ , \tag{4}$$

where N_+ denotes the number of matched positive queries, $\alpha > 0$ is the margin, and $[z]_+ = \max(z, 0)$. The Top-k neighborhood focuses metric supervision on locally confusable classes, while the graph weights encode their learned neighborhood relations. Negatives that already satisfy the margin incur no penalty.

Neighborhood-Conditioned Foreground Objectness. To capture heterogeneous foreground representations, the same prototype graph is reused to construct multiple foreground components. We use the mean off-diagonal adjacency weight as the grouping threshold:

$$\bar{e}^t = \frac{1}{C_t(C_t - 1)} \sum_{\substack{a,b \in \mathcal{K}^t \\ a \neq b}} E^t_{ab}. \tag{5}$$

Relations satisfying $E^t_{ab} > \bar{e}^t$ are retained, and the connected components of the resulting graph are denoted by $\mathcal{S}^t = \{\mathcal{S}^t_m\}_{m=1}^{M_t}$.

For each component $\mathcal{S}^t_m$, we compute a graph-weighted prototype mean and element-wise variance

$$\begin{aligned} \omega_{mc} &= \frac{\sum_{u \in \mathcal{S}^t_m} E^t_{cu}}{\sum_{v \in \mathcal{S}^t_m} \sum_{u \in \mathcal{S}^t_m} E^t_{vu}}, \\ \mu_m &= \sum_{c \in \mathcal{S}^t_m} \omega_{mc} p_c, \\ v_m &= \sum_{c \in \mathcal{S}^t_m} \omega_{mc} (p_c - \mu_m) \odot (p_c - \mu_m) + \delta \mathbf{1}, \end{aligned} \tag{6}$$

where $\delta > 0$ ensures numerical stability.

Given a query q_i, we compute its variance-normalized deviation from each component:

$$D_{im} = \frac{1}{d} \sum_{\ell=1}^{d} \frac{(q_{i\ell} - \mu_{m\ell})^2}{v_{m\ell}}, \qquad D_i = \min_{1 \le m \le M_t} D_{im}. \tag{7}$$

For each query, PROW selects the component with the minimum variance-normalized deviation for class-independent objectness estimation.

The corresponding foreground objective is

$$\mathcal{L}_{\text{obj}} = \frac{1}{|\mathcal{I}^{+}_{\text{ann}}|} \sum_{i \in \mathcal{I}^{+}_{\text{ann}}} D_i, \tag{8}$$

where $\mathcal{I}^{+}_{\text{ann}}$ denotes the set of matched foreground queries. Minimizing this objective encourages each matched foreground query to remain close to at least one compatible foreground component. These components provide multiple local foreground references rather than models of individual unknown class distributions.

Following PROB (Zohar et al., 2023), D_i is converted into a class-independent objectness score and combined with the detector's classification score at inference; the exact scoring rule is provided in Appendix A.1.

3.4 Incremental Supervision and Overall Objective

As illustrated in Figure 2, in task $t > 1$, previously learned prototypes and graph relations are inherited from task $t - 1$. Annotations are provided for classes introduced in the current task, while instances of previously learned classes may remain unlabeled. To preserve previously acquired knowledge, we retain the detector from task $t - 1$ as a frozen teacher. Predictions with high confidence for previously learned classes are retained as pseudo labels and combined with the available annotations of the current task. We further include a limited set of replay samples from previous tasks to mitigate forgetting.

The complete framework is optimized with

$$\mathcal{L} = \mathcal{L}_{\text{det}} + \lambda_{\text{local}}\mathcal{L}_{\text{local}} + \lambda_{\text{obj}}\mathcal{L}_{\text{obj}}, \tag{9}$$

where $\mathcal{L}_{\text{det}}$ denotes the standard Deformable DETR detection objective. The coefficients λ_{local} and λ_{obj} balance the local discrimination and foreground objectness objectives, respectively.

4 Experiments

4.1 Experimental Setup

Benchmarks. We evaluate PROW on two standard OWOD benchmarks, M-OWODB (Joseph et al., 2021) and S-OWODB (Gupta et al., 2022), following the evaluation protocols of previous work (Zohar et al., 2023; Sun et al., 2024). M-OWODB is constructed from PASCAL VOC (Everingham et al., 2010) and MS-COCO (Lin et al., 2014) and consists of four incremental tasks with semantic overlap across tasks. S-OWODB is constructed from MS-COCO and separates semantic superclasses across tasks, resulting in less semantic overlap between known and future unknown classes.

Evaluation metrics. For known classes, we report Known-class mAP (K-mAP). In incremental tasks, K-mAP is further decomposed into Previously Known-class mAP (PK-mAP) and Current Known-class mAP (CK-mAP). Unknown object discovery is evaluated using Unknown Recall (U-Recall). We additionally report Absolute Open-Set Error (A-OSE) (Zohar et al., 2023), which counts unknown objects incorrectly classified as known class. Higher U-Recall and mAP and lower A-OSE indicate better performance.

Implementation details. PROW is built on Deformable DETR (Zhu et al., 2021) with a ResNet-50 backbone pretrained using DINO (Caron et al., 2021), following the detector configuration of PROB (Zohar et al., 2023). We use 100 object queries with a feature dimension of $d = 256$ and set the prototype neighborhood size to $k = 15$. The graph decay factor is set to $\epsilon = 0.9$, and the margin for local class discrimination is set to $\alpha = 0.2$. We set $\lambda_{\text{local}} = 1$ and $\lambda_{\text{obj}} = 10^{-3}$. The confidence threshold for pseudo-label selection is set to 0.7. Training uses a batch size of 3 with AdamW on two NVIDIA RTX 3090 GPUs. During inference, the 100 highest scoring predictions per image are retained for evaluation. Further training and replay details are provided in the appendix.

Table 1: Comparison with state-of-the-art OWOD methods on M-OWODB. Task 4 contains all benchmark classes and therefore has no U-Recall.

Method	Task 1		Task 2				Task 3				Task 4		
	U-Recall	K-mAP	U-Recall	PK-mAP	CK-mAP	K-mAP	U-Recall	PK-mAP	CK-mAP	K-mAP	PK-mAP	CK-mAP	K-mAP
ORE* (Joseph et al., 2021)	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
OST (Yang et al., 2022)	–	56.2	–	53.4	26.5	39.9	–	38.0	12.8	29.6	30.1	13.3	25.9
UC-OWOD (Wu et al., 2022b)	2.4	50.7	3.4	33.1	30.5	31.8	8.7	28.8	16.3	24.6	25.6	15.9	23.2
OCPL (Yu et al., 2022)	8.26	56.6	7.65	50.6	27.5	39.1	11.9	38.7	14.7	30.7	30.7	14.4	26.7
2B-OCD (Wu et al., 2022a)	12.1	56.4	9.4	51.6	25.3	38.5	11.6	37.2	13.2	29.2	30.0	13.3	25.8
OW-DETR (Gupta et al., 2022)	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
ALLOW (Ma et al., 2023b)	13.6	59.3	10.0	53.2	34.0	45.6	14.3	42.6	26.7	38.0	33.5	21.8	30.6
Randbox (Wang et al., 2023)	10.6	61.8	6.3	–	–	45.3	7.8	–	–	39.4	–	–	35.4
PROB (Zohar et al., 2023)	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 (Ma et al., 2023a)	23.7	60.0	19.1	55.5	32.7	44.1	24.4	42.8	18.7	34.8	34.4	16.6	29.9
OrthogonalDet (Sun et al., 2024)	24.6	61.3	26.3	55.5	**38.5**	47.0	29.1	46.7	30.6	41.3	**42.4**	24.3	37.9
Decoupled PROB (Inoue et al., 2025)	20.3	59.8	18.4	56.4	36.7	46.6	20.3	44.6	26.1	38.5	37.8	22.1	33.8
PROB+OWOBJ (Zhang et al., 2025)	23.6	61.4	23.8	**58.4**	34.4	45.7	25.1	44.8	27.8	38.8	36.4	20.7	32.0
PROW	**31.9**	**61.9**	**34.4**	58.2	38.3	**48.3**	**31.5**	**47.0**	**32.4**	**42.1**	41.7	**29.0**	**38.5**

Table 2: Comparison with state-of-the-art OWOD methods on S-OWODB. Task 4 contains all benchmark classes and therefore has no U-Recall.

Method	Task 1		Task 2				Task 3				Task 4		
	U-Recall	K-mAP	U-Recall	PK-mAP	CK-mAP	K-mAP	U-Recall	PK-mAP	CK-mAP	K-mAP	PK-mAP	CK-mAP	K-mAP
ORE* (Joseph et al., 2021)	1.5	61.4	3.9	56.5	26.1	40.6	3.6	38.7	23.7	33.7	33.6	26.3	31.8
OW-DETR (Gupta et al., 2022)	5.7	71.5	6.2	62.8	27.5	43.8	6.9	45.2	24.9	38.5	38.2	28.1	33.1
PROB (Zohar et al., 2023)	17.6	73.4	22.3	66.3	36.0	50.4	24.8	47.8	30.4	42.0	42.6	31.7	39.9
CAT (Ma et al., 2023a)	24.0	74.2	23.0	67.6	35.5	50.7	24.6	51.2	32.6	45.0	45.4	35.1	42.8
OrthogonalDet (Sun et al., 2024)	**24.6**	71.6	27.9	64.0	39.9	51.3	31.9	**52.1**	42.2	48.8	48.7	38.8	46.2
Decoupled PROB (Inoue et al., 2025)	17.4	73.7	24.4	66.3	45.7	55.5	23.5	50.9	43.3	48.4	46.1	39.8	44.6
PROB+OWOBJ (Zhang et al., 2025)	22.3	**76.2**	**28.7**	**69.8**	41.0	54.8	30.9	50.6	35.7	46.8	46.7	36.9	43.2
PROW	**24.6**	74.3	28.0	64.6	**48.6**	**56.2**	**33.1**	51.8	**44.8**	**49.4**	**50.3**	**46.5**	**49.4**

4.2 Comparison with State-of-the-Art Methods

Unknown Object Discovery. Table 1 shows that PROW consistently improves unknown object discovery on M-OWODB, reaching U-Recall scores of 31.9, 34.4, and 31.5 on Tasks 1–3, respectively. Compared with OrthogonalDet (Sun et al., 2024), these results correspond to gains of 7.3, 8.1, and 2.4 points. PROW also outperforms the more recent Decoupled PROB (Inoue et al., 2025) and PROB+OWOBJ (Zhang et al., 2025) across all three tasks.

On S-OWODB, PROW obtains U-Recall scores of 24.6, 28.0, and 33.1. It matches the highest result on Task 1, remains competitive with PROB+OWOBJ on Task 2, and achieves the highest U-Recall on Task 3. The smaller margin on S-OWODB is consistent with its superclass-separated protocol, where known and future unknown classes exhibit less semantic overlap.

Known Class Detection. PROW maintains strong performance on known classes while improving unknown object discovery. On M-OWODB, it achieves K-mAP scores of 61.9, 48.3, 42.1, and 38.5 across Tasks 1–4, respectively, outperforming the compared methods on all four tasks.

On S-OWODB, PROB+OWOBJ achieves the highest K-mAP on Task 1, whereas PROW performs best on Tasks 2–4 with scores of 56.2, 49.4, and 49.4. In particular, PROW reaches a CK-mAP of 48.6 on Task 2, compared with 39.9 for OrthogonalDet and 41.0 for PROB+OWOBJ. These results indicate that the gains in unknown object discovery do not come at the expense of known class detection.

Table 3: Absolute Open-Set Error (A-OSE $\downarrow$) on M-OWODB.

Method	Task 1	Task 2	Task 3
ORE* (Joseph et al., 2021)	10,459	10,445	7,990
OW-DETR (Gupta et al., 2022)	10,240	8,441	6,803
Randbox (Wang et al., 2023)	4,498	1,880	1,452
PROB (Zohar et al., 2023)	5,195	6,452	2,641
OrthogonalDet (Sun et al., 2024)	4,148	1,791	1,345
PROB+OWOBJ (Zhang et al., 2025)	3,102	4,032	**821**
PROW	**1,891**	**1,465**	1,067

Known–Unknown Confusion. Table 3 compares A-OSE on M-OWODB. PROW achieves the lowest A-OSE on Tasks 1 and 2, with 1,891 and 1,465 errors, respectively. Compared with PROB+OWOBJ (Zhang et al., 2025), PROW reduces A-OSE by 39.0% on Task 1. On Task 3, PROW further reduces A-OSE from 1,345 to 1,067 relative to OrthogonalDet, although PROB+OWOBJ (Zhang et al., 2025) obtains a lower value of 821. Together with the U-Recall results, these comparisons show that PROW consistently reduces the confusion of unknown objects with known classes while maintaining strong unknown object discovery across incremental tasks.

4.3 Ablation Studies

Table 4 presents the ablation of the main components of PROW on M-OWODB. Task 1 isolates the effects of neighborhood-guided local discrimination and neighborhood-conditioned foreground objectness, while Task 2 further examines incremental supervision.

Table 4: Ablation of the main learning components of PROW on M-OWODB. Task 1 evaluates the two graph-conditioned representation objectives, while Task 2 additionally evaluates incremental supervision.

Components			Task 1			Task 2				
Local	Foreground	Incremental	U-Recall	K-mAP	A-OSE	U-Recall	PK-mAP	CK-mAP	K-mAP	A-OSE
–	–	–	18.3	59.3	11,560	14.9	52.1	31.6	41.1	9,769
–	✓	✓	25.0	61.8	11,222	20.6	55.7	30.8	43.2	8,237
✓	–	✓	27.8	59.7	2,065	23.9	57.3	34.2	45.7	1,686
✓	✓	–	**31.9**	**61.9**	**1,891**	29.3	53.9	36.9	45.4	1,494
✓	✓	✓	**31.9**	**61.9**	**1,891**	**34.4**	**58.2**	**38.3**	**48.3**	**1,465**

On Task 1, without local discrimination, U-Recall drops from 31.9 to 25.0 and A-OSE increases from 1,891 to 11,222, while K-mAP remains nearly unchanged. Without foreground objectness, U-Recall decreases from 31.9 to 27.8 and K-mAP from 61.9 to 59.7. These results indicate that the two graph-conditioned objectives provide complementary contributions to unknown object discovery and known class detection.

On Task 2, incremental supervision improves U-Recall from 29.3 to 34.4, PK-mAP from 53.9 to 58.2, and K-mAP from 45.4 to 48.3. The improvement in PK-mAP is consistent with the role of pseudo labels and replay samples in preserving previously learned knowledge. A detailed analysis of pseudo-label and replay supervision is provided in the appendix.

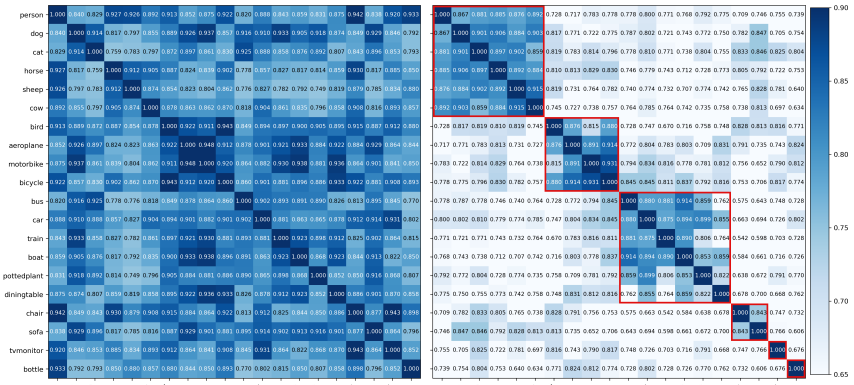
4.4 Further Analysis

Prototype Neighborhood. The neighborhood size k determines the number of non-matching prototypes selected for each matched positive query during graph updating. Table 5 reports the results on M-OWODB Task 1.

Table 5: Effect of the prototype neighborhood size k on M-OWODB Task 1.

k	U-Recall (↑)	K-mAP (↑)	A-OSE (↓)
5	27.7	59.5	2,082
10	27.9	59.1	1,922
15	**31.9**	**61.9**	**1,891**
20	27.5	60.5	1,971

Performance improves as k increases from 5 to 15, with U-Recall increasing from 27.7 to 31.9 and A-OSE decreasing from 2,082 to 1,891. A further increase to $k = 20$ reduces U-Recall to 27.5. This indicates that overly small neighborhoods provide limited local relations, while overly large neighborhoods weaken the locality of the prototype graph. We therefore set $k = 15$ in all experiments.



(a) PROB (b) PROW

Figure 3: Pairwise prototype similarities on M-OWODB Task 1. PROB (Zohar et al., 2023) is shown on the left and PROW on the right. PROW exhibits more localized prototype neighborhoods, with red boxes highlighting representative local structures.

Prototype Neighborhood Structure. Figure 3 compares pairwise prototype similarities between PROB (Zohar et al., 2023) and PROW on M-OWODB Task 1. Compared with PROB, PROW exhibits stronger similarity within localized groups of prototypes and lower similarity across groups. This pattern is consistent with the neighborhood structure captured by the dynamic prototype graph.

Multi-Component Foreground Modeling. Table 6 compares different foreground representations on M-OWODB Task 1.

Table 6: Ablation of foreground representations on M-OWODB Task 1.

Foreground representation	U-Recall (↑)	K-mAP (↑)	A-OSE (↓)
Query-based global	31.0	41.1	2,651
Prototype-based global	26.2	60.0	1,899
Unweighted components	29.5	61.2	**1,879**
Weighted components	**31.9**	**61.9**	1,891

The query-based global representation achieves high U-Recall but substantially lower K-mAP. Using a single prototype-based global representation improves K-mAP to 60.0 but reduces U-Recall to 26.2. In contrast, multiple foreground components improve both metrics, and graph-based weighting further increases U-Recall and K-mAP to 31.9 and 61.9, respectively. These results support the use of multiple foreground references derived from the learned prototype neighborhood for class-independent objectness estimation.

5 Conclusions

We introduce PROW, a neighborhood-aware representation learning framework for open world object detection. PROW learns a dynamic graph over known class prototypes and reuses its neighborhood structure for two complementary purposes: selecting locally confusable prototypes for local class discrimination and constructing multiple foreground components for class-independent objectness estimation. During incremental learning, inherited prototypes and graph relations are complemented by confidence-filtered pseudo labels and replay to preserve previously learned knowledge. Experiments on M-OWODB and S-OWODB demonstrate the effectiveness of PROW in unknown object discovery, known class detection, and reducing known–unknown confusion.

AI USE STATEMENT

Generative AI tools were used solely for language polishing and readability improvements during manuscript preparation. All AI-assisted revisions were carefully reviewed by the authors. The authors take full responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

We provide detailed formulations of all components of PROW together with the training and inference procedures used in our experiments. The experimental setup specifies the model configuration, optimization settings, loss weights, and key hyperparameters. Additional details on incremental supervision, hyperparameter sensitivity, and performance variation across random seeds are provided in the appendix. These details are intended to facilitate reproduction of the reported results.

REFERENCES

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *Proceedings of the European Conference on Computer Vision*, pp. 213–229, 2020.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9650–9660, 2021.

Linhui Dai, Hong Liu, Hao Tang, Zhiwei Wu, and Pinhao Song. Ao2-detr: Arbitrary-oriented object detection transformer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(5): 2342–2356, 2023. doi: 10.1109/TCSVT.2022.3222906.

Thang Doan, Xin Li, Sima Behpour, Wenbin He, Liang Gou, and Liu Ren. Hyp-OW: Exploiting hierarchical structure learning with hyperbolic distance enhances open world object detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(2):1555–1563, 2024.

Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88: 303–338, 2010.

Tao Feng, Mang Wang, and Hangjie Yuan. Overcoming catastrophic forgetting in incremental object detection via elastic response distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9427–9436, 2022.

Ian J. Goodfellow, Mehdi Mirza, Xia Da, Aaron C. Courville, and Yoshua Bengio. An empirical investigation of catastrophic forgeting in gradient-based neural networks. In *International Conference on Learning Representations*, 2014.

Akshita Gupta, Sanath Narayan, K J Joseph, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. OW-DETR: Open-world detection transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9235–9244, 2022.

Riku Inoue, Masamitsu Tsuchiya, and Yuji Yasui. Decoupled prob: Decoupled query initialization tasks and objectness-class learning for open world object detection. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, pp. 8196–8205, 2025.

K J Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5830–5840, 2021.

Junsu Kim, Hoseong Cho, Jihyeon Kim, Yihalem Yimolal Tiruneh, and Seungryul Baek. SDDGR: Stable diffusion-based deep generative replay for class incremental object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28772–28781, 2024.

Yiming Li, Yi Wang, Wenqian Wang, Dan Lin, Bingbing Li, and Kim-Hui Yap. Open world object detection: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(2): 988–1008, 2025. doi: 10.1109/TCSVT.2024.3480691.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Proceedings of the European Conference on Computer Vision*, pp. 740–755, 2014.

Yaoyao Liu, Bernt Schiele, Andrea Vedaldi, and Christian Rupprecht. Continual detection transformer for incremental object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23799–23808, 2023.

Shuailei Ma, Yuefeng Wang, Ying Wei, Jiaqi Fan, Thomas H. Li, Hongli Liu, and Fanbing Lv. 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, 2023a.

Yuqing Ma, Hainan Li, Zhange Zhang, Jinyang Guo, Shanghang Zhang, Ruihao Gong, and Xianglong Liu. Annealing-based label-transfer learning for open world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11454–11463, 2023b.

Zeyu Ma, Ziqiang Zheng, Jiwei Wei, Yang Yang, and Heng Tao Shen. Instance-dictionary learning for open-world object detection in autonomous driving scenarios. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5):3395–3408, 2024. doi: 10.1109/TCSVT.2023.3322465.

Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pp. 109–165. 1989.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.

Zhicheng Sun, Jinghan Li, and Yadong Mu. Exploring orthogonality in open world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17302–17312, 2024.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11), 2008.

Yanghao Wang, Zhongqi Yue, Xian-Sheng Hua, and Hanwang Zhang. Random boxes are open-world object detectors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6233–6243, 2023.

Yan Wu, Xiaowei Zhao, Yuqing Ma, Duorui Wang, and Xianglong Liu. Two-branch objectness-centric open world detection. In *Proceedings of the 3rd International Workshop on Human-Centric Multimedia Analysis*, pp. 35–40, 2022a.

Zhiheng Wu, Yue Lu, Xingyu Chen, Zhengxing Wu, Liwen Kang, and Junzhi Yu. UC-OWOD: Unknown-classified open world object detection. In *Proceedings of the European Conference on Computer Vision*, pp. 193–210, 2022b.

Shuo Yang, Peize Sun, Yi Jiang, Xiaobo Xia, Ruiheng Zhang, Zehuan Yuan, Changhu Wang, Ping Luo, and Min Xu. Objects in semantic topology. In *International Conference on Learning Representations*, 2022.

Jinan Yu, Liyan Ma, Zhenglin Li, Yan Peng, and Shaorong Xie. Open-world object detection via discriminative class prototype learning. In *2022 IEEE International Conference on Image Processing*, pp. 626–630, 2022.

Shan Zhang, Yao Ni, Jinhao Du, Yuan Xue, Philip Torr, Piotr Koniusz, and Anton van den Hengel. Open-world objectness modeling unifies novel object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30332–30342, 2025.

Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable DETR: Deformable transformers for end-to-end object detection. In *International Conference on Learning Representations*, 2021.

Orr Zohar, Kuan-Chieh Wang, and Serena Yeung. PROB: Probabilistic objectness for open world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11444–11453, June 2023.

A Additional Experimental Details

This appendix provides additional implementation details, supplementary ablation studies, stability analyses, qualitative results, and further discussion that complement the main paper.

A.1 Training and Inference Details

PROW follows the Deformable DETR (Zhu et al., 2021) configuration used by PROB (Zohar et al., 2023), with a ResNet-50 backbone pretrained using DINO (Caron et al., 2021). The detector uses 100 object queries with a feature dimension of $d = 256$. Unless otherwise specified, we set the prototype neighborhood size to $k = 15$, the graph decay factor to $\epsilon = 0.9$, and the margin for local class discrimination to $\alpha = 0.2$. The loss weights are set to $\lambda_{\text{local}} = 1$ and $\lambda_{\text{obj}} = 10^{-3}$, and the confidence threshold for pseudo-label selection is set to 0.7. All models are optimized using AdamW with a batch size of 3 on two NVIDIA RTX 3090 GPUs.

The complete training objective is

$$\mathcal{L} = \mathcal{L}_{\text{det}} + \lambda_{\text{local}}\mathcal{L}_{\text{local}} + \lambda_{\text{obj}}\mathcal{L}_{\text{obj}}, \tag{10}$$

where $\mathcal{L}_{\text{det}}$ denotes the standard Deformable DETR detection loss. As described in Section 3.2, the foreground objective is applied to matched foreground queries using their minimum variance-normalized deviations from the foreground components.

At inference, following PROB (Zohar et al., 2023), the minimum deviation D of each query is converted into a class-independent objectness score:

$$p_{\text{obj}} = \exp(-D). \tag{11}$$

A smaller deviation therefore corresponds to a higher foreground objectness score. The final prediction score is obtained by combining the classification and objectness scores:

$$p_{\text{final}} = p_{\text{cls}}\, p_{\text{obj}}. \tag{12}$$

The 100 highest-scoring predictions per image are retained for evaluation. PROW introduces approximately 2% additional parameters relative to the base detector.

A.2 Incremental Supervision

In task $t > 1$, instances of previously learned classes may remain unannotated in the current training data. As described in the main paper, PROW retains the detector from task $t - 1$ as a frozen teacher. High-confidence predictions for previously learned classes are retained as pseudo labels and combined with the available annotations of the current task. A limited set of replay samples from previous tasks is further included to preserve previously acquired knowledge.

Table 7 analyzes the contribution of these two sources of incremental supervision on M-OWODB Task 2.

Table 7: Ablation of incremental supervision on M-OWODB Task 2.

Setting	U-Recall (↑)	PK-mAP (↑)	CK-mAP (↑)	K-mAP (↑)
w/o pseudo labels	28.1	57.1	37.0	47.1
w/o replay	27.6	**58.3**	38.1	48.2
PROW	**34.4**	58.2	**38.3**	**48.3**

Without pseudo-label supervision, U-Recall decreases from 34.4 to 28.1, while PK-mAP and K-mAP decrease from 58.2 to 57.1 and from 48.3 to 47.1, respectively. Removing replay leaves known-class performance largely unchanged but reduces U-Recall from 34.4 to 27.6. The full setting provides the strongest overall performance, indicating that pseudo-label and replay supervision provide complementary support during incremental learning.

B Additional Ablation Studies

B.1 Hyperparameter Sensitivity

Local Discrimination Weight. We examine the effect of the loss weight λ_{local} on M-OWODB Task 2.

Table 8: Sensitivity to λ_{local} on M-OWODB Task 2.

λ_{local}	0.1	0.5	1.0	1.5	2.0
U-Recall (↑)	27.7	28.0	34.4	**34.6**	34.1
K-mAP (↑)	46.9	47.2	**48.3**	47.9	47.3

As shown in Table 8, increasing λ_{local} from 0.1 to 1.0 improves both U-Recall and K-mAP. Although $\lambda_{\text{local}} = 1.5$ yields a marginally higher U-Recall, its K-mAP decreases from 48.3 to 47.9. We therefore set $\lambda_{\text{local}} = 1.0$ in the main experiments, which provides the best overall balance between unknown object discovery and known class detection.

Graph Decay Factor. We investigate the effect of the graph decay factor ϵ, which controls the persistence of neighborhood relations that are not supported by the current matched queries.

As shown in Table 9, performance improves as ϵ increases from 0.05 to 0.9, with higher U-Recall and lower A-OSE. A small ϵ causes unsupported neighborhood relations to decay rapidly, limiting the persistence of useful local structure. In contrast, increasing ϵ to 0.95 leads to a clear degradation across all three metrics, suggesting that overly persistent relations can reduce the graph's ability to adapt to evolving neighborhood structure. We therefore set $\epsilon = 0.9$ in all experiments.

Pseudo-Label Confidence Threshold. We evaluate the confidence threshold θ_{pl} used to retain predictions of previously learned classes on M-OWODB Task 2.

As shown in Table 10, a threshold of $\theta_{\text{pl}} = 0.7$ provides the strongest overall performance, achieving the highest U-Recall, PK-mAP, and K-mAP among the evaluated settings. A lower threshold retains more low-confidence predictions, while a higher threshold removes more predictions of previously learned classes. We therefore use $\theta_{\text{pl}} = 0.7$ throughout the experiments.

C Stability Analysis

To examine the sensitivity of PROW to training randomness, we repeat the experiment on M-OWODB Task 1 using three independent random seeds. Table 11 reports the mean and standard deviation of U-Recall, K-mAP, and A-OSE across the three runs.

Known-class detection remains particularly consistent across different runs, as reflected by the small variation in K-mAP. U-Recall and A-OSE exhibit somewhat larger variations, indicating greater sensitivity in unknown object discovery and known–unknown separation. Nevertheless, the overall performance remains consistent across the three random seeds, supporting the reproducibility of PROW on M-OWODB Task 1.

D Qualitative Analysis

D.1 Representation Visualization

We further visualize the learned query representations on M-OWODB Task 1 using t-SNE (Van der Maaten & Hinton, 2008). Figure 4 compares PROB (Zohar et al., 2023) and PROW. The two methods exhibit noticeably different organizations of the representation space. In particular, PROW shows a more localized grouping pattern among several known classes, which is consistent with the prototype neighborhood structure observed in the main paper. We use this visualization only as qualitative evidence and rely on the quantitative results in the main paper for performance comparison.

Table 9: Sensitivity to the graph decay factor ϵ on M-OWODB Task 1.

ϵ	U-Recall (↑)	K-mAP (↑)	A-OSE (↓)
0.05	21.8	59.8	5,294
0.1	29.2	61.5	3,871
0.5	31.7	60.7	2,133
0.9	**31.9**	**61.9**	**1,891**
0.95	27.1	58.1	5,341

Table 10: Sensitivity to the pseudo-label confidence threshold θ_{pl} on M-OWODB Task 2.

θ_{pl}	U-Recall (↑)	PK-mAP (↑)	CK-mAP (↑)	K-mAP (↑)
0.6	30.7	53.1	36.8	45.0
0.7	**34.4**	**58.2**	38.3	**48.3**
0.8	24.9	55.4	**38.6**	47.0

D.2 Open World Detection

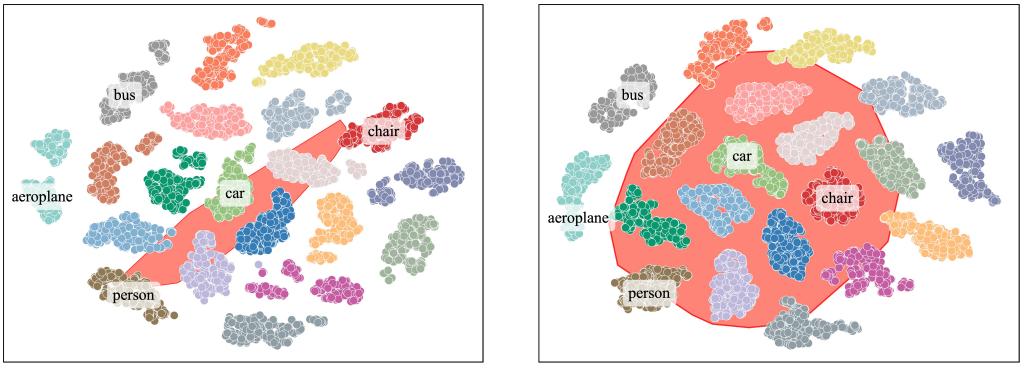
Figure 5 compares PROW with OrthogonalDet on M-OWODB Task 1. The examples illustrate cases in which PROW detects unknown foreground objects that are missed or assigned known class labels by the baseline. These observations are consistent with the quantitative improvements in U-Recall and A-OSE reported in the main paper.

D.3 Unknown Discovery during Incremental Learning

Figure 6 provides qualitative examples over successive incremental tasks. As the known label space expands, PROW continues to identify foreground instances outside the current known class space. The examples complement the quantitative results on Tasks 1–3 reported in the main paper.

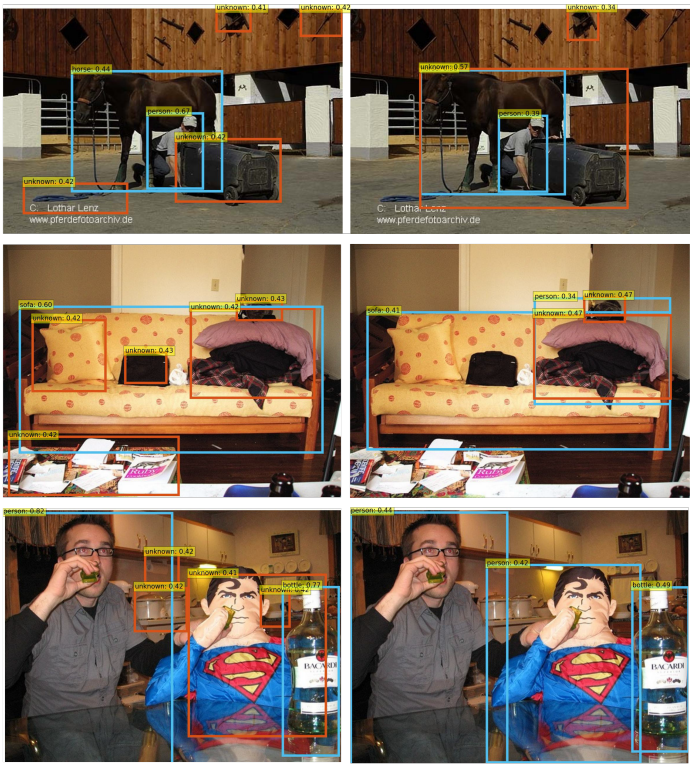
Table 11: Performance variation of PROW across three independent random seeds on M-OWODB Task 1. Results are reported as mean ± standard deviation.

Setting	U-Recall (↑)	K-mAP (↑)	A-OSE (↓)
PROW	31.9 ± 1.7	61.9 ± 0.2	$1,891 \pm 83$



(a) PROB (Zohar et al., 2023). (b) PROW.

Figure 4: t-SNE visualization of query representations on M-OWODB Task 1. Different colors correspond to different known classes, while the highlighted red regions denote unknown predictions. PROW exhibits a more localized organization of known-class representations than PROB (Zohar et al., 2023).



PROW OrthogonalDet

Figure 5: Qualitative comparison on M-OWODB Task 1. Known detections are shown in blue and unknown detections in orange. PROW detects additional unknown foreground objects and reduces several cases of known–unknown confusion compared with OrthogonalDet (Sun et al., 2024).

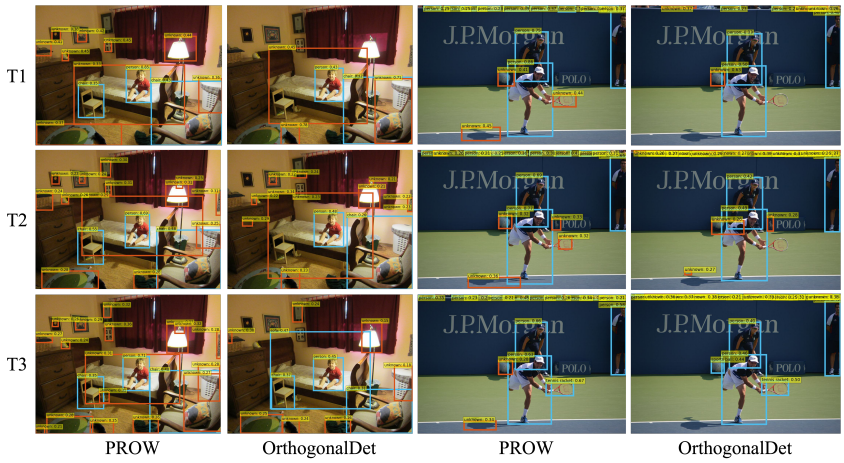


Figure 6: Qualitative examples of unknown object discovery during incremental learning on M-OWODB. T1–T3 denote Tasks 1–3.