REAL-OW: REhearsAL-free Open World Object Detection with Low-Rank Adaptation and Dual-Stage Objectness Modeling

Huazhong Zhang[1], Yang Zhang[1]*, Xiaowen Fu[1], Linlin Shen[1], and Jinbao Wang[1]

Shenzhen University, Shenzhen, China
yangzhang@szu.edu.cn

Abstract. Open-World Object Detection (OWOD) requires detectors to identify previously unseen objects as unknown and incrementally incorporate them into the set of known categories, while preserving previously acquired knowledge. Existing frameworks rely heavily on exemplar replay to mitigate catastrophic forgetting, but in some real applications, storing raw data conflicts with data access restrictions and leads to data exposure risks, while incurring significant memory overhead. In this paper, we propose REAL-OW, a novel rehearsal-free framework that decouples incremental knowledge through a collaborative adapter architecture based on Low-Rank Adaptation (LoRA). Specifically, we deploy General Adapters (GAs) in the backbone to enable the significance-aware refinement of cross-task universal representations, while Specific Adapters (SAs) in the decoder provide orthogonal storage for task-specific expertise. To resolve representation drift in objectness modeling under rehearsal-free constraints, we introduce Dual-Stage Objectness Modeling (DSOM), which alternates between feature aggregation and boundary consolidation to stabilize objectness distributions while maintaining the separation between known and unknown categories. Furthermore, DSOM is supported by a Calibrated Gaussian Negative Log-Likelihood (CG-NLL) distance tailored for the dispersed feature distributions inherent in rehearsal-free settings. Extensive evaluations demonstrate that REAL-OW achieves state-of-the-art performance, surpassing existing exemplar replay methods in both detection precision and unknown discovery. Our approach establishes a new baseline for rehearsal-free OWOD.

1 Introduction

Traditional object detection assumes a closed-set environment [2,36], and it can only recognize categories that were defined during the initial training phase. However, real-world settings are dynamic and unpredictable [37]. To bridge this gap, Open-World Object Detection (OWOD) [19] was introduced and it allows systems to operate in unconstrained settings [13]. An OWOD detector must

* Corresponding author.

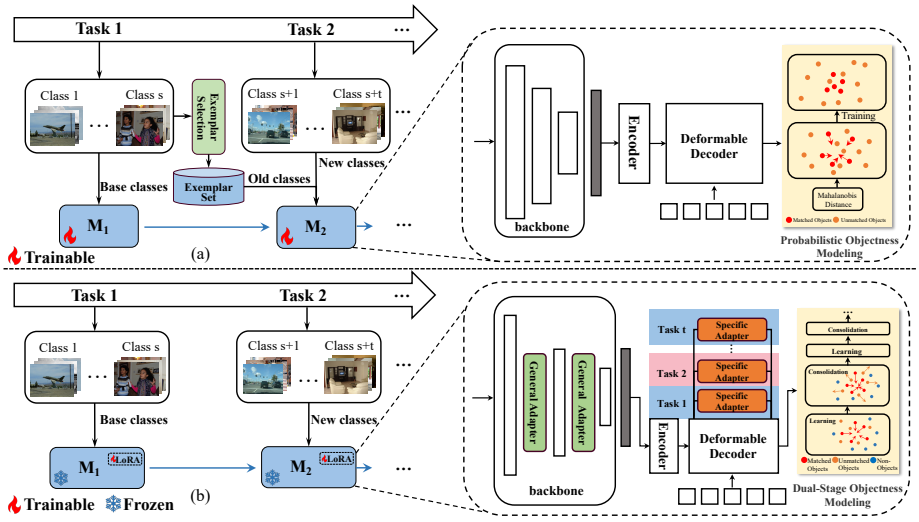

Fig. 1: Training process and architecture overview of our rehearsal-free framework. We remove the exemplar replay mechanism from the typical training process in (a), achieving a more secure and streamlined rehearsal-free architecture design as shown in (b). To resist catastrophic forgetting without data replay, we employ a collaborative adapter architecture to separately capture general and task-specific features. Our Dual-Stage Objectness Modeling alternates between learning and consolidation phases, preventing representation drift that typically occurs in traditional probabilistic objectness modeling under rehearsal-free constraints.

detect unknown objects and take them as novel classes, incrementally learning these new categories without forgetting previous ones [22,28,45,49]. This ability is vital for computer vision scenarios like autonomous driving and open-ended robotics [20,23,29].

Despite its importance, current OWOD frameworks rely heavily on exemplar replay [19,21,58], which mitigates catastrophic forgetting by maintaining a buffer of representative samples from previously learned classes and replaying them during new task training to preserve old knowledge [3,15,35,38].

However, exemplar replay incurs several practical challenges in real-world applications. First, privacy-sensitive fields like medical imaging prevent models from retaining raw data after a task ends to avoid data exposure. This makes it prohibitive to store samples for exemplar replay [32]. Second, as the task sequence expands, the buffer must accumulate more images to cover every previous class, thus incurring growing storage costs [9]. Furthermore, from the perspective of the OWOD objective, which is the continuous exploration and discovery of unknown categories, the learning process should prioritize modeling novel classes rather than repeatedly revisiting previously learned ones [19,23].

These concerns and limitations motivate us to explore a rehearsal-free approach that avoids storing raw exemplar buffers. Low-Rank Adaptation (LoRA) [16] emerges as a compelling solution for its underlying parameter-efficient fine-

tuning mechanism [12, 24, 27, 34, 56]. By constraining each added task updates into low-rank subspaces, LoRA enables the model to acquire new category knowledge while minimizing interference with previously learned representations [34, 44, 47, 50, 53], and most importantly, it achieves this without relying on data replay.

In rehearsal-free OWOD, learning and preserving a continuous stream of knowledge without accessing old data is a fundamental challenge. A single shared representation cannot distinguish task-specific features, while completely independent modules hinder knowledge accumulation and transfer, leading to catastrophic forgetting [12, 56]. To address this, we decouple general and specific knowledge by adopting a collaborative adapter framework, which separates general representations from task-specific expertise. As illustrated in Fig. 1, our training process does not include any data replay. This avoids repeated data access, significantly enhancing data privacy protection and reducing storage requirements. To balance general knowledge and task-specific expertise, our architecture employs two collaborative adapters within the frozen pre-trained model. Specifically, General Adapters (GAs) are placed in the backbone to maintain cross-task universal representations, while Specific Adapters (SAs) are added to the transformer decoder incrementally for each new task to capture knowledge for newly introduced categories.

Beyond incremental learning, OWOD further requires the discovery of unknown objects, which are not annotated but need to be discovered during training [19, 23, 37]. Recent methods employ probabilistic objectness [7, 39, 43, 58] and energy-based modeling [26, 55] to discover unknown objects in unmatched queries and separate them from known classes and background, respectively. However, these strategies depend on exemplar replay to stay distribution stable [23]. In rehearsal-free settings, the model continuously updates its parameters with emerging new data without access to historical samples, leading to potential feature drift, which causes two major problems: (1) unknown decision bias, where emerging unknown class features are pulled toward current known class regions, causing them to inherit known class similarity bias, while unknown objects with significantly different textures from current known categories are often overly suppressed and misclassified as background; and (2) catastrophic forgetting.

To solve these issues, we develop a Dual-Stage Objectness Modeling (DSOM) strategy. As shown in Fig. 1, this approach uses learning phases to aggregate the known feature embeddings and consolidation phases to sharpen the boundary between the knowns and unknowns. Furthermore, we observe that under this rehearsal-free condition, the dual-stage update disperses the known object distribution, rendering the standard Mahalanobis distance [31, 55, 58] ineffective for uncertainty estimation due to its sensitivity to distribution compactness. To address this, we introduce a Calibrated Gaussian Negative Log-Likelihood (CG-NLL) distance, which accounts for the increased variance, better optimizes the expanded objectness density and stabilizes the energy landscape.

Experimental results demonstrate that our framework outperforms all compared exemplar replay methods in both precision and unknown discovery. Re-

markably, our model achieves these results without replaying a single image. To the best of our knowledge, we are the first to propose a strictly rehearsal-free methodology for OWOD, thereby establishing a new baseline for the field. Our main contributions can be summarized as follows:

- We design General and Specific Adapters to continuously refine shared representations while isolating task-specific knowledge, enabling rehearsal-free open-world object detection without storing historical samples.
- We propose a Dual-Stage Objectness Modeling to avoid catastrophic forgetting caused by representation drift, together with a CG-NLL distance to calibrate objectness scores and stabilize the energy landscape.
- Extensive evaluations demonstrate that our approach outperforms all compared exemplar replay based methods, establishing a new and efficient baseline for rehearsal-free open-world object detection.

2 Related Works

2.1 Open World Object Detection

Open-World Object Detection (OWOD) shifts the focus from static closed-set recognition to dynamic environment adaptation [19, 37]. In this paradigm, a detector must not only identify known categories but also detect previously unseen objects as unknown, subsequently incorporating them into the known set as new labels emerge [21, 23, 29]. Joseph et al. [19] formally introduced the OWOD framework by synthesizing concepts from Open Set Recognition [37] and Incremental Object Detection [21]. To alleviate catastrophic forgetting, they employed an exemplar replay (ER) mechanism, where a set of representative samples from earlier tasks is stored and replayed during incremental updates. Many follow-up approaches, such as OW-DETR [13], PROB [58], CAT [28], and ORTH [43], continue to rely on exemplar buffers to preserve performance on prior classes during knowledge expansion.

However, storing raw data raises privacy and storage concerns like we mention above. Since current methods have not fully eliminated this dependency, a truly rehearsal-free approach is necessary.

2.2 Parameter-Efficient Fine-Tuning

Parameter-Efficient Fine-Tuning (PEFT) adapts large-scale models by updating only a small subset of parameters [34, 42]. Among these techniques, Low-Rank Adaptation (LoRA) [16] is a prominent approach [1, 4, 11, 12, 24, 41, 53]. Formally, LoRA freezes the pre-trained weight matrix $W_0 \in \mathbb{R}^{d\times k}$ and introduces a low-rank update $\Delta W = BA$, where $B \in \mathbb{R}^{d\times r}$ and $A \in \mathbb{R}^{r\times k}$ are trainable matrices with rank $r \ll \min(d, k)$. This strategy significantly reduces the parameter space while preserving the generalization capability of the backbone. Recent research [8, 33, 34, 54] focuses on enhancing LoRA through structural constraints and decoupled modularity. In object detection, some works [5, 18, 33]

use LoRA for domain-specific tasks, its potential for rehearsal-free Open-World Object Detection remains largely unexplored. The need to balance universal feature refinement with task-specific isolation in a strictly non-exemplar setting presents a unique challenge for current PEFT designs in detection domains.

2.3 Class-agnostic Objectness Modeling

In OWOD, class-agnostic objectness modeling decouples object existence from category identity by estimating whether a region corresponds to any valid object instance, thereby enabling the discovery of previously unseen categories [17, 30, 36, 40, 55]. Early methods such as PROB [58] and OrthogonalDet [39] estimate objectness through probabilistic modeling and orthogonal decomposition. More recent works, such as PASS [51], leverage attribute subsets to refine proposal scoring, while OWOBJ [55] assumes a dynamic Gaussian prior over objectness features and scores proposals using Mahalanobis distance. However, the optimization of these methods is typically restricted to current task data, relying heavily on exemplar replay to prevent forgetting. Without data rehearsal, continuous model updates cause representation drift, which destabilizes the objectness modeling and leads to a collapse in novelty discovery.

3 Methods

REAL-OW is a decoupled framework that synergizes General Adapters and Specific Adapters with Dual-Stage Objectness Modeling (DSOM) for rehearsal-free OWOD. Built on D-DETR-based detectors [2, 13], this architecture balances general and task-specific knowledge refinement for each incremental task with stable feature accumulation and robust unknown discovery. Together, these components enable incremental learning without relying on historical samples.

General Adapters. We adopt General Adapters to support the continuous refinement of shared representations within the Pyramid Vision Transformer (PVT) backbone. As illustrated in Fig. 2, we inject LoRA modules into the query q and value v projection matrices of the self-attention layers. During training, we keep the original backbone weights W_0 frozen and update only the lightweight GAs parameters. The forward pass for each GA can be formulated as:

$$f_G(x) = W_0 x + A_G B_G x, \tag{1}$$

where x is input at any task t. $B_G \in \mathbb{R}^{r \times k}$ is a fixed, random orthogonal down-projection matrix, and $A_G \in \mathbb{R}^{d \times r}$ is a trainable up-projection matrix initialized to zero. This asymmetric design is motivated by recent work [57] showing that tuning B is more impactful than tuning A. To prevent standard regularization from indiscriminately suppressing all parameters through a uniform penalty, which leads to shrinkage bias [14, 46], we propose a novel Adaptive Gated Sparse Loss ($\mathcal{L}_{AGS}$) to optimize the GAs. This significance-aware mechanism allows the

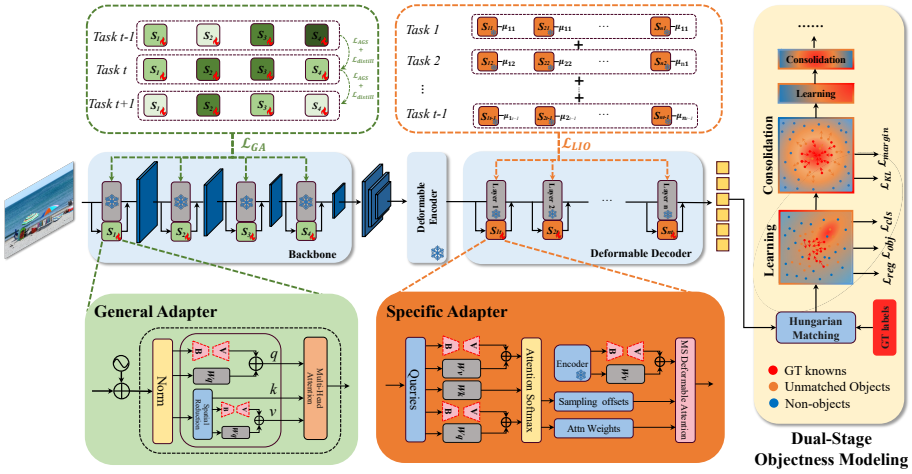

Fig. 2: Pipeline of our REAL-OW. Our architecture consists of three core components: (1) **General Adapters** in the backbone, which employ the $\mathcal{L}_{AGS}$ to enable a significance-aware refinement of cross-task universal representations; (2) **Specific Adapters** in the decoder layers, which are added incrementally to store task-unique features in non-overlapping subspaces; and (3) **Dual-Stage Objectness Modeling (DSOM).** The DSOM module alternates between a learning phase for feature aggregation and a consolidation phase for boundary refinement. This synergistic design enables the model to accumulate knowledge and discover novel objects without relying on historical samples.

model to selectively accumulate impactful knowledge by dynamically modulating the regularization pressure. The loss function is formulated as:

$$\mathcal{L}_{AGS} = \sum_{\ell=1}^{L} \left(\frac{1}{1 + \gamma \cdot (\|\mathbf{A}_\ell\|_F + \|\mathbf{B}_\ell\|_F)} \right) \cdot (\|\mathbf{A}_\ell\|_F + \|\mathbf{B}_\ell\|_F), \tag{2}$$

where L is the number of backbone layers. The fractional term serves as an adaptive gate, which acts as a dynamic switch by sensing the Frobenius norm ($\|\cdot\|_F$) of each adapter layer. Specifically, the gate reduces regularization pressure for essential layers to preserve them as stable anchors for universal knowledge, and it applies higher sparse pressure to redundant layers, driving them to optimize, adapting and capturing shared features in later tasks of the sequence. To ensure high-fidelity consistency with the teacher model as well, we employ $\mathcal{L}_{distill}$, an output-aligned distillation loss, formulated as:

$$\mathcal{L}_{distill} = \sum_{i \in \mathcal{C}_t} s_{t-1,i} \log(s_{t,i}), \tag{3}$$

where s_t represents the normalized feature distribution from backbone over task t classes $\mathcal{C}_t$. Combined with Eq. (2) and Eq. (3), the whole GA loss formulated

as:

$$\mathcal{L}_{GA} = \mathcal{L}_{distill} + \mathcal{L}_{AGS}. \tag{4}$$

Specific Adapters. While the GAs focuses on refining cross-task universal representations in the backbone, the feature interaction part requires specialized expertise to handle task-specific category features. To this end, we employ the Specific Adapters in the decoder. As shown in Fig. 2, the SAs employ the same LoRA placement in the q and v projections and the same asymmetric initialization strategy. But for each new task, a dedicated set of SA modules is added in parallel. This incremental expansion enables the decoder to store task-specific knowledge in isolated branches.

However, this multi-branch architecture is susceptible to feature entanglement. Because LoRA updates only a small fraction of the parameters relative to the pre-trained model, discriminating between unique task features remains difficult. Current incremental methods [34,56] often address this by using scalar control gates but these mechanisms are content-agnostic, as they ignore actual parameter magnitudes, making them insufficient to fully prevent interference between tasks. To ensure that newly added adapters do not interfere with the specialized knowledge stored during preceding task stages, we optimize the SAs using the Layer-wise Interaction Orthogonal Loss ($\mathcal{L}_{LIO}$). Unlike standard geometric constraints, our approach is capacity-aware by coupling the gating signal with the actual parameter intensity. The loss is formulated as:

$$\mathcal{L}_{LIO} = \sum_{i=1}^{t-1} \sum_{k=1}^{N} \left(\mu_t^k \cdot \mathrm{sg}\left[\mu_i^k \cdot \|\mathbf{A}_i^k\|_F \right] \right)^2, \tag{5}$$

where N represents the number of layers in decoder. μ_t^k denotes the learnable weight for task t at the k-th block, and $\|\mathbf{A}_i^k\|_F$ represents the Frobenius norm of the adapter parameters from previous task i. The operator $\mathrm{sg}[\cdot]$ indicates a stop-gradient operation to protect historical weights. The product $\mu_i^k \cdot \|\mathbf{A}_i^k\|_F$ defines the effective strength of a layer, which quantifies both its activation level and stored feature intensity. $\mathcal{L}_{LIO}$ dynamically guides the current task to utilize layers that are either structurally inactivated or parametrically sparse in history. This design ensures that each adapter group selects a non-overlapping set of significant layers, effectively isolating task-specific knowledge and preventing the corruption of prior feature representations.

Dual-Stage Objectness Modeling. Objectness probability and energy-based modeling are essential for identifying candidate proposals and distinguishing between known objects, unknown objects, and background. In a rehearsal-free setting where replay is unavailable, feature drift occurs, which destabilizes the modeling process. To address this, we propose Dual-Stage Objectness Modeling (DSOM), which partitions the optimization process into alternating Learning and Consolidation phases. To further improve objectness quantification, we introduce CG-NLL distance.

CG-NLL Distance. As shown in Fig. 3, we observe that known category distribution becomes more dispersed over time as semantic diversity grows without exemplar replay. Standard Mahalanobis distance fails to account for this expansion, leading to inaccurate objectness estimation. To bridge the gap between geometric distance and probabilistic objectness, we formulate the Calibrated Gaussian Negative Log-Likelihood (CG-NLL) distance to incorporate distribution density into a scaled objectness modeling. Our CG-NLL distance is defined as:

$$d_C(\mathbf{x}) = \frac{(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{1+\beta\ln(1+|\Sigma|)}, \quad (6)$$

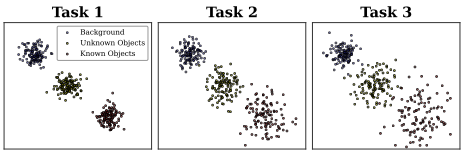

Fig. 3: t-SNE visualization of embeddings from Task 1 to Task 3. Using the standard Mahalanobis distance without data replay, the known category distribution becomes progressively dispersed as tasks increase.

where μ and Σ represent the mean and covariance of the known category distribution. We utilize the term $\tau(\Sigma) = 1+\beta\ln(1+|\Sigma|)$ as an adaptive scaling factor to normalize the distance based on the dispersion. β is a calibration hyperparameter. By adaptively rescaling the Mahalanobis Distance, CG-NLL distance maintains a strict manifold for compact distributions to suppress background noise, while providing a calibration for dispersed distributions to increase tolerance against catastrophic forgetting. Based on it, for unmatched queries $\boldsymbol{q}_i \in \mathcal{Q}_{\text{unk}}$, the objectness score

$$S^i_{obj} = \exp(-d_C(\mathbf{q}_i)), \quad (7)$$

serves as the confidence score for the unknown class pseudo-labels.

Learning Phase. This phase focuses on aggregating the feature representations of currently known categories and recognizes unknown objects. $\mathcal{L}_{Learning}$ is formulated as:

$$\mathcal{L}_{Learning} = \mathcal{L}_{obj} + \mathcal{L}_{cls} + \mathcal{L}_{reg}, \quad (8)$$

where $\mathcal{L}_{obj} = \sum_{q_i \in \mathcal{Q}_{\text{known}}} d_C(q_i)$ minimizes the CG-NLL distance for matched known queries to improve objectness modeling, thereby enforcing a compact clustering of known features around the estimated distribution center. $\mathcal{L}_{cls} = \text{FL}(\hat{l}, l)$ is the sigmoid focal loss for classification. The predictions $\hat{l} \in \mathbb{R}^{Z+1}$ are supervised by targets l, which consist of discrete known class labels $l_{1:Z} \in \{0,1\}^Z$ and a continuous objectness score $l_{Z+1} \in [0,1]$. For unmatched queries, the $Z+1$-th dimension is assigned soft objectness pseudo-labels $l_i[Z+1] = S^i_{\text{obj}}$, given by Eq. (7), allowing the model to regard high-scoring regions as potential unknown objects. $\mathcal{L}_{reg}$ is a combination of the ℓ_1 loss and the GIoU loss.

Consolidation Phase. To maintain stability and sharpen the boundary between known and unknown energy distribution for preventing misclassification, the consolidation phase employs two optimization objectives:

$$\mathcal{L}_{Consolidation} = \mathcal{L}_{margin} + \mathcal{L}_{KL}, \quad (9)$$

Table 1: Quantitative comparison for OWOD on M-OWODB (top) and S-OWODB (bottom). Notably, our REAL-OW is **the only rehearsal-free approach** among all compared methods, yet it outperforms all other high-parameter CNN-based and D-DETR-based models that utilize exemplar replay. Best results are in **bold**.

Task IDs (→)	**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
UC-OWOD [52]	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
ORE [19]	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
2B-OCD [48]	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
OCPL [30]	8.3	56.6	7.7	50.6	27.5	39.1	11.9	38.7	14.7	30.7	30.7	14.4	26.7
OW-DETR [13]	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
CAT [28]	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
PROB [58]	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
OWOBJ [55]	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
Hyp-OW [7]	23.5	59.4	20.6	-	-	44.0	26.3	-	-	36.8	-	-	33.6
RandBox [45]	10.6	61.8	6.3	-	-	45.3	7.8	-	-	39.4	-	-	35.4
ORTH [39]	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
Ours:REAL-OW	**25.5**	**62.6**	**26.9**	**59.4**	**38.7**	**49.0**	**29.6**	**47.7**	**31.2**	**42.2**	**42.8**	**26.7**	**38.7**
ORE [19]	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 [13]	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 [58]	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 [28]	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
OWOBJ [55]	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
Hyp-OW [7]	23.9	72.7	23.3	-	-	50.6	25.4	-	-	46.2	-	-	44.8
ORTH [39]	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
Ours:REAL-OW	**24.9**	**76.9**	**29.8**	**71.5**	**42.3**	**56.9**	**33.3**	**53.4**	**44.6**	**50.5**	**49.7**	**39.3**	**47.1**

Table 2: Comparison of trainable parameters. Our method requires significantly fewer trainable parameters compared to existing OWOD paradigms.

Methods	Trainable Parameters
Transformer-based	39.7M
CNN-based	106.6M
Ours	2.0M

where $\mathcal{L}_{margin}$ stands for energy margin loss, separating known and unknown energy distributions. $\mathcal{L}_{margin}$ is formulated as:

$$\mathcal{L}_{margin} = \mathbb{E}_{q^+ \in Q_{\text{known}}} \mathbb{E}_{q^- \in Q_{\text{unk}}} \left[\max(0, E(q^+) - E(q^-) + \delta) \right], \tag{10}$$

where $E(q) = -\log \sum_{i=1}^{C} \exp(f^i(q))$ denotes the energy level based on the classifier logits $f^i(q)$. This hinge loss pushes the energy of unknown candidates above a margin δ relative to knowns, establishing a discriminative surface. $\mathcal{L}_{KL}$ stands for KL divergence loss [55], acting as a variance regularizer to prevent latent space collapse and maintaining embedding diversity, thereby mitigating over-fitting bias. Through the alternation of these two phases, DSOM achieves a stable and rehearsal-free optimization of objectness knowledge across incremental tasks.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate REAL-OW on M-OWODB [19] and S-OWODB [13], derived from MS-COCO [25] and PASCAL VOC [10]. S-OWODB groups classes by

Table 3: Comparison for incremental object detection (iOD) on PASCAL VOC. We evaluate our model using three standard incremental scenarios. In these settings, new groups of classes (10, 5, or the final 1 class) are added to a detector that has already been trained on the initial set of classes (10, 15, or 19 classes, respectively). Best overall result for each setting are **bold**.

10 + 10 setting	old cls.	new cls.	final mAP	15 + 5 setting	old cls.	new cls.	final mAP	19 + 1 setting	old cls.	new cls.	final mAP
ILOD [38]	63.2	63.1	63.2	ILOD [38]	68.3	58.4	65.8	ILOD [38]	68.5	62.7	68.2
ORE [19]	60.4	68.8	64.5	ORE [19]	71.8	58.7	68.5	ORE [19]	69.4	60.1	68.8
OW-DETR [13]	63.5	67.9	65.7	OW-DETR [13]	72.2	59.8	69.4	OW-DETR [13]	70.7	62.0	70.2
PROB [58]	66.0	67.2	66.5	PROB [58]	73.2	60.8	70.1	PROB [58]	73.9	48.5	72.6
CAT [28]	67.9	67.4	67.7	CAT [28]	75.6	59.3	72.2	CAT [28]	74.5	61.1	73.8
OWOBJ [55]	70.5	72.0	69.9	OWOBJ [55]	76.5	63.7	73.3	OWOBJ [55]	76.6	53.8	75.8
ORTH [39]	74.5	70.2	72.3	ORTH [39]	78.3	71.8	74.7	ORTH [39]	75.6	74.9	75.6
Ours:REAL-OW	73.8	74.3	**74.0**	**Ours:REAL-OW**	81.2	68.1	**75.9**	**Ours:REAL-OW**	78.2	60.7	**77.3**

their super-categories like animals or vehicles, which is more challenging. Both benchmarks comprise four incremental tasks (T_1 to T_4), with 20 new classes introduced per task. During training, only labels for the current task are provided, while testing requires detecting all previously encountered categories. Most importantly, our approach is strictly rehearsal-free and stores no historical samples.

Evaluation Metrics. We use mean Average Precision (mAP) to measure detection accuracy on known classes, and Unknown Recall (U-Recall) to evaluate the ability to discover unseen objects. Additionally, A-OSE and WI are employed to quantify misclassifications between known and unknown categories.

Implementation Details. We implement our framework in PyTorch and conduct all experiments on 3 NVIDIA RTX A6000 GPUs. Our model uses PVT as the backbone and Deformable DETR as the detection framework. Following standard practice, we pre-train the model on ImageNet [6] using a self-supervised approach to prevent premature exposure to future unknown classes [13]. For a fair comparison, this pre-training setup is applied to all evaluated methods. The decoder consists of 6 layers. We set the ranks for the GAs and SAs to 16 and 10, respectively. Within the DSOM module, each iteration employs 2 learning cycles and 10 consolidation cycles. The hyperparameters δ, γ, β are set to 0.2, 0.01, 0.1 respectively. To ensure the reliability, all reported results represent the average performance calculated over five different random seeds.

4.2 Experimental Results

For a fair comparison, we focus on models that generate unknown pseudo-labels internally, excluding approaches that leverage external large models, such as vision language models (VLMs), to provide pseudo-label assignments.

Quantitative. As summarized in Tab. 1, our REAL-OW framework achieves state-of-the-art performance on both M-OWODB and S-OWODB benchmarks. Notably, REAL-OW is the **only strictly rehearsal-free** method among all compared techniques. Despite this constraint, our model consistently achieves the top performance across all results, surpassing heavily parameterized CNN-based (e.g.RandBox, ORTH) and Transformer-based (e.g.CAT, PROB) methods. As shown in Tab. 2, these baseline methods require over 50 times and nearly

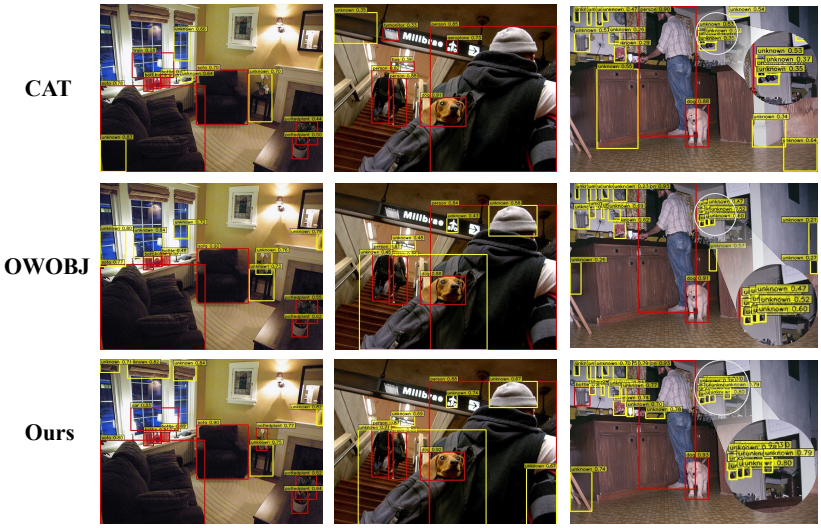

Fig. 4: Qualitative results comparison between CAT, OWOBJ and our REAL-OW. Object detections with Red-known Yellow-unknown are displayed. Our method discovers more novel objects, such as the curtain in column 1 and the ladder in column 3. It also demonstrates more precise localization for multiple small objects, such as the cup cluster in column 3. In contrast, CAT misidentifies the tvmonitor in column 2, and OWOBJ mistakes a known pottedplant for an unknown object in column 1. These results show that our model provides better discovery and discrimination capabilities.

20 times more trainable parameters than our approach, respectively. Furthermore, REAL-OW reaches the highest precision in Task 4, proving its superior resistance to catastrophic forgetting without any data replay. The performance gap between REAL-OW and OWOBJ also suggests that our dual-stage modeling is more effective than single-stage with known-unknown separation. REAL-OW also achieves the best results in A-OSE and WI among all compared methods. Refer to Sec. 2 in supplementary material for more details.

We evaluate REAL-OW on three standard iOD settings to assess knowledge retention. As shown in Tab. 3, our framework achieves the highest final mAP across all settings, outperforming SOTA methods. These results validate our architecture: GAs maintain stable feature foundations, while SAs and DSOM ensure task isolation and energy stability. Collectively, these components enable robust rehearsal-free learning with superior knowledge retention and exceptional resistance to catastrophic forgetting. Per-class details are in supplementary material Sec. 1.

Qualitative. As illustrated in Fig. 4, REAL-OW demonstrates higher U-Recall and more precise localization. It successfully discovers additional novel objects missed by other methods, such as the curtain in column 1 and ladder in column 3, while accurately separating individual cups in column 3. These improvements show the CG-NLL distance effectively quantifies unknown object-

Table 4: Results of different rank settings on Task 3. We evaluate U-Recall and mAP across various rank combinations for GAs and SAs. The corresponding number of trainable parameters is also reported.

GAs ↓ SAs → (rank)	$r=1$		$r=4$		$r=8$		$r=16$		$r=32$		$r=64$	
	U-Recall↑	mAP↑	U-Recall↑	mAP↑	U-Recall↑	mAP↑	U-Recall↑	mAP↑	U-Recall↑	mAP↑	U-Recall↑	mAP↑
$r=1$	23.32	35.31	24.88	38.69	26.92	40.37	27.61	40.44	27.90	39.23	27.88	38.84
	0.29×10^5		0.66×10^5		1.15×10^5		2.13×10^5		4.10×10^5		8.03×10^5	
$r=4$	24.18	36.28	26.14	39.60	27.50	41.86	28.47	41.75	28.83	41.02	28.90	40.54
	0.79×10^5		1.16×10^5		1.65×10^5		2.63×10^5		4.60×10^5		8.53×10^5	
$r=8$	24.98	36.90	27.03	39.72	28.82	42.40	28.90	42.33	28.85	42.18	28.87	41.63
	1.45×10^5		1.82×10^5		2.31×10^5		3.30×10^5		5.26×10^5		9.20×10^5	
$r=16$	25.32	37.78	28.11	40.80	**29.64**	**42.23**	29.88	42.11	29.65	41.20	29.70	40.75
	2.79×10^5		3.15×10^5		3.65×10^5		4.63×10^5		6.60×10^5		10.53×10^5	
$r=32$	24.64	38.57	28.46	41.75	28.70	42.33	29.42	42.08	29.48	41.86	29.60	39.50
	5.45×10^5		5.82×10^5		6.31×10^5		7.29×10^5		9.26×10^5		13.19×10^5	
$r=64$	22.76	38.61	26.21	41.81	27.54	42.21	28.72	41.97	28.32	40.22	28.38	38.43
	10.77×10^5		11.14×10^5		11.63×10^5		12.62×10^5		14.58×10^5		18.51×10^5	

Table 5: Performance comparison between standard distillation $\mathcal{L}_{distill}$ and our $\mathcal{L}_{AGS}$. Best results are in **bold**.

Task IDs (→)	**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
w/o $\mathcal{L}_{AGS}$	**26.7**	**63.2**	25.4	58.6	**38.9**	48.7	28.0	46.5	28.5	40.5	40.2	25.8	36.6
w/ $\mathcal{L}_{AGS}$	25.5	62.6	**26.9**	**59.4**	38.7	**49.0**	**29.6**	**47.7**	**31.2**	**42.2**	**42.8**	**26.7**	**38.7**

ness. Additionally, our model avoids the misclassification errors seen in CAT (tvmonitor, column 2) and OWOBJ (pottedplant, column 1). High confidence scores for these detections confirm that DSOM maintains a stable energy landscape for reliable boundary separation. For more visualization results on the strong resistance of REAL-OW to catastrophic forgetting and its superior capability for discovering unknown classes, please refer to supplementary material Sec. 3.

4.3 Ablation Study

LoRA. We evaluate various LoRA rank configurations for the GAs and SAs. Following previous PEFT methods [14, 33, 47], we apply LoRA to q and v projections. Further validation for this placement provided in supplementary material Sec. 4. As shown in Tab. 4, we observe that performance does not scale linearly with the increase in trainable parameters. Instead, the configuration ($R_{GA}=16, R_{SA}=8$) achieves the optimal trade-off between performance and parameter capacity. Excessively high ranks lead to a decline in U-Recall and offer only marginal gains for mAP. This degradation is more pronounced as R_{GA} increases, suggesting that an over-parameterized general feature space induces an unknown decision bias.

Adaptive Gated Sparse Loss. Tab. 5 compares $\mathcal{L}_{AGS}$ with standard distillation $\mathcal{L}_{distill}$ that relies only on output alignment. As expected, during the initial task, $\mathcal{L}_{AGS}$ lags slightly behind the baseline due to its sparsity constraint.

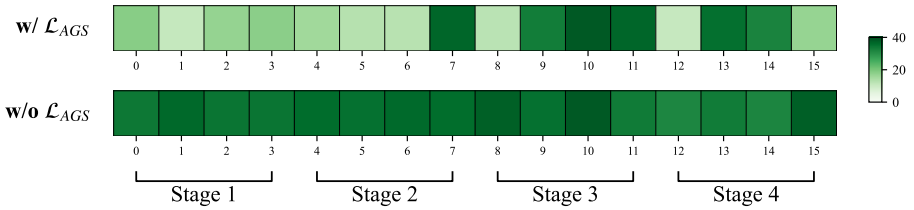

Fig. 5: Visualization of adapter weight magnitudes (ℓ_2 norm) across backbone stages. Deeper colors indicate higher layer significance. Our method (w/ $\mathcal{L}_{AGS}$) achieves a sparse and targeted update pattern, prioritizing significant layers while suppressing redundant ones.

Table 6: Ablation study of the Dual-Stage Objectness Modeling. We report the results on Task 1 and Task 2. D-DETR upper limit denotes training with ground-truth unknown labels. D-DETR-ER uses standard exemplar replay, while D-DETR-RF employs our proposed GA and SA adapters for a rehearsal-free setting. Both baselines lack pseudo-labeling for unknown objects. We evaluate the impact of the CG-NLL distance, and the two training phases.

Methods	Task 1		Task 2			
	U-Recall	mAP	U-Recall	Prev.	Curr.	Both
D-DETR-upper limit	32.9	63.8	40.5	62.9	41.2	52.1
D-DETR-ER	–	60.3	–	54.5	34.4	44.7
D-DETR-RF	–	61.9	–	57.9	38.5	48.2
REAL-OW w/o CG-NLL	23.3	58.3	25.1	57.5	37.9	47.7
REAL-OW w/o Learning	25.4	60.5	26.0	47.8	37.5	42.7
REAL-OW w/o Consolidation	20.1	63.1	21.0	58.9	39.5	49.2
REAL-OW - complete	**25.5**	**62.6**	**26.9**	**59.4**	**38.7**	**49.0**

However, by suppressing redundant layers, this constraint ensures that critical layers carry general features applicable to the encountered known categories, while other layers remain plastic. The performances on later tasks show the superior refinement of $\mathcal{L}_{AGS}$. Fig. 5 further confirms this targeted update pattern that critical layers preserve foundational representations with $\mathcal{L}_{AGS}$, while standard distillation leads to a uniform and undifferentiated weight distribution.

Dual-Stage Objectness Modeling. Tab. 6 validates the core components of our DSOM. **(1) CG-NLL Distance.** Replacing CG-NLL distance with the standard Mahalanobis distance reduces U-Recall, demonstrating that standard metrics fail to calibrate energy scores as features disperse. This result is further explained in Fig. 6. By accounting for distribution dispersion, CG-NLL facilitates more accurate objectness modeling. This optimization enables the model to establish a significantly more distinct energy boundary between known and unknown classes. **(2) The Learning and Consolidation Phases.** Removing the learning phase causes previous mAP to drop from 59.4% to 47.8%, proving that feature aggregation is essential to prevent new knowledge from overwriting

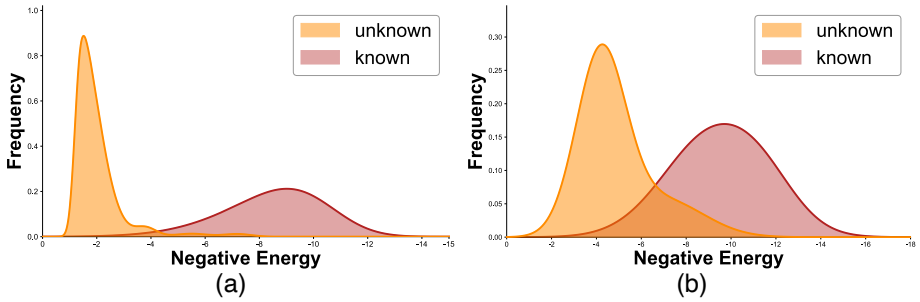

Fig. 6: Comparison between (a) CG-NLL distance and (b) standard Mahalanobis distance on negative energy distributions after Task 1. CG-NLL distance effectively creates a clearer discriminative margin between known and unknown classes compared to the Mahalanobis distance.

old representations. Conversely, skipping the consolidation phase slightly inflates current precision but collapses U-Recall from 26.9% to 21.0%. This indicates that without boundary separation, unknown objects are incorrectly biased toward known distributions and causes misclassification of the known and unknown.

(3) Cycle Ratios. Fig. 7 illustrates the performance after Task 2 under different cycle ratios of two phases. While mAP increases with more learning cycles, U-Recall peaks at a 2:10 ratio before declining. This confirms that while the learning phase maintains precision, allocating too many cycles to it creates a bias that hinders unknown discovery. We select 2:10 as the optimal balance to ensure both high stability and robust novelty detection.

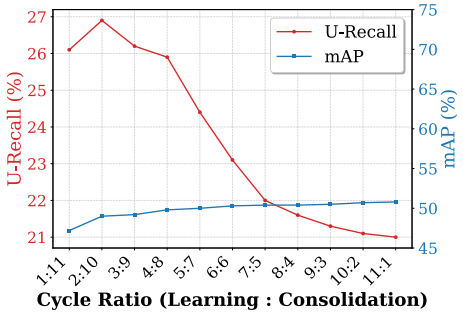

Fig. 7: Performance of different cycle ratios between learning and consolidation phases with total cycles fixed at 12.

5 Conclusion

In this paper, we propose REAL-OW, a novel rehearsal-free framework for Open-World Object Detection. REAL-OW decouples incremental knowledge through a collaborative adapter architecture, utilizing General Adapters for significance-aware feature refinement and Specific Adapters for orthogonal task isolation. To counteract representation drift, we introduce Dual-Stage Objectness Modeling (DSOM) supported by CG-NLL distance, which ensures robust objectness modeling and unknown discovery without data replay. Extensive experiments

validate that REAL-OW outperforms all compared SOTA methods while maintaining a strictly rehearsal-free constraint. The continuous stability of DSOM provides a reliable foundation for the future integration of external large models like vision-language models (VLMs), enabling the generation of high-quality pseudo-labels for future open-world research.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 62176163), the Shenzhen Higher Education Stable Support Program General Project (Grant No. 20231120175215001), the Scientific Foundation for Youth Scholars of Shenzhen University, the National Natural Science Foundation of China (Grant No. 62576218), the Guangdong Provincial Key Laboratory (Grant No. 2023B1212060076), and the Intelligent Computing Center of Shenzhen University.

References

1. Agiza, A., Neseem, M., Reda, S.: Mtlora: Low-rank adaptation approach for efficient multi-task learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16196–16205 (2024)
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European conference on computer vision. pp. 213–229. Springer (2020)
3. Castro, F.M., Marín-Jiménez, M.J., Guil, N., Schmid, C., Alahari, K.: End-to-end incremental learning. In: Proceedings of the European conference on computer vision (ECCV). pp. 233–248 (2018)
4. Chavan, A., Liu, Z., Gupta, D., Xing, E., Shen, Z.: One-forall: Generalized lora for parameter-efficient fine-tuning. arXiv preprint arXiv:2306.07967 **3** (2023)
5. Chen, Z., Duan, Y., Wang, W., He, J., Lu, T., Dai, J., Qiao, Y.: Vision transformer adapter for dense predictions. arXiv preprint arXiv:2205.08534 (2022)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
7. Doan, T., Li, X., Behpour, S., He, W., Gou, L., Ren, L.: Hyp-ow: Exploiting hierarchical structure learning with hyperbolic distance enhances open world object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 1555–1563 (2024)
8. Edalati, A., Tahaei, M., Kobyzev, I., Nia, V.P., Clark, J.J., Rezagholizadeh, M.: Krona: Parameter-efficient tuning with kronecker adapter. In: Enhancing LLM Performance: Efficacy, Fine-Tuning, and Inference Techniques, pp. 49–65. Springer (2025)
9. Ermis, B., Zappella, G., Wistuba, M., Rawal, A., Archambeau, C.: Memory efficient continual learning with transformers. Advances in Neural Information Processing Systems **35**, 10629–10642 (2022)
10. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision **88**(2), 303–338 (2010)

11. Gandikota, R., Materzyńska, J., Zhou, T., Torralba, A., Bau, D.: Concept sliders: Lora adaptors for precise control in diffusion models. In: European Conference on Computer Vision. pp. 172–188. Springer (2024)
12. Guo, H., Zhu, F., Liu, W., Zhang, X.Y., Liu, C.L.: Pilora: Prototype guided incremental lora for federated class-incremental learning. In: European Conference on Computer Vision. pp. 141–159. Springer (2024)
13. Gupta, A., Narayan, S., Joseph, K., Khan, S., Khan, F.S., 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)
14. He, J., Duan, Z., Zhu, F.: Cl-lora: Continual low-rank adaptation for rehearsal-free class-incremental learning. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 30534–30544 (2025)
15. Hou, S., Pan, X., Loy, C.C., Wang, Z., Lin, D.: Learning a unified classifier incrementally via rebalancing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 831–839 (2019)
16. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. ICLR **1**(2), 3 (2022)
17. Jaiswal, A., Wu, Y., Natarajan, P., Natarajan, P.: Class-agnostic object detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 919–928 (2021)
18. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European conference on computer vision. pp. 709–727. Springer (2022)
19. Joseph, K., 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)
20. Joseph, K., Paul, S., Aggarwal, G., Biswas, S., Rai, P., Han, K., Balasubramanian, V.N.: Novel class discovery without forgetting. In: European Conference on Computer Vision. pp. 570–586. Springer (2022)
21. Joseph, K., Rajasegaran, J., Khan, S., Khan, F.S., Balasubramanian, V.N.: Incremental object detection via meta-learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(12), 9209–9216 (2021)
22. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences **114**(13), 3521–3526 (2017)
23. Li, Y., Wang, Y., Wang, W., Lin, D., Li, B., Yap, K.H.: Open world object detection: A survey. IEEE Transactions on Circuits and Systems for Video Technology (2024)
24. Lin, L., Fan, H., Zhang, Z., Wang, Y., Xu, Y., Ling, H.: Tracking meets lora: Faster training, larger model, stronger performance. In: European Conference on Computer Vision. pp. 300–318. Springer (2024)
25. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
26. Liu, W., Wang, X., Owens, J., Li, Y.: Energy-based out-of-distribution detection. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 21464–21475. Curran Associates, Inc. (2020)

27. Lu, H., Zhao, C., Xue, J., Yao, L., Moore, K., Gong, D.: Adaptive rank, reduced forgetting: Knowledge retention in continual learning vision-language models with dynamic rank-selective lora. arXiv preprint arXiv:2412.01004 (2024)
28. Ma, S., Wang, Y., Wei, Y., Fan, J., Li, T.H., Liu, H., 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)
29. Ma, Z., Yang, Y., Wang, G., Xu, X., Shen, H.T., Zhang, M.: Rethinking open-world object detection in autonomous driving scenarios. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 1279–1288 (2022)
30. Maaz, M., Rasheed, H., Khan, S., Khan, F.S., Anwer, R.M., Yang, M.H.: Class-agnostic object detection with multi-modal transformer. In: European conference on computer vision. pp. 512–531. Springer (2022)
31. Mahalanobis, P.C.: On the generalized distance in statistics. Sankhyā: The Indian Journal of Statistics, Series A (2008-) **80**, pp. S1–S7 (2018)
32. Prabhu, A., Torr, P.H., Dokania, P.K.: Gdumb: A simple approach that questions our progress in continual learning. In: European conference on computer vision. pp. 524–540. Springer (2020)
33. Pu, X., Xu, F.: Low-rank adaption on transformer-based oriented object detector for satellite onboard processing of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing (2025)
34. Qiao, F., Mahdavi, M.: Learn more, but bother less: parameter efficient continual learning. Advances in Neural Information Processing Systems **37**, 97476–97498 (2024)
35. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: icarl: Incremental classifier and representation learning. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 2001–2010 (2017)
36. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence **39**(6), 1137–1149 (2016)
37. Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boult, T.E.: Toward open set recognition. IEEE transactions on pattern analysis and machine intelligence **35**(7), 1757–1772 (2012)
38. Shmelkov, K., Schmid, C., Alahari, K.: Incremental learning of object detectors without catastrophic forgetting. In: Proceedings of the IEEE international conference on computer vision. pp. 3400–3409 (2017)
39. Sun, Z., Li, J., Mu, Y.: Exploring orthogonality in open world object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17302–17312 (2024)
40. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019)
41. Valipour, M., Rezagholizadeh, M., Kobyzev, I., Ghodsi, A.: Dylora: Parameter-efficient tuning of pre-trained models using dynamic search-free low-rank adaptation. In: Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. pp. 3274–3287 (2023)
42. Wang, H., Sun, T., Fan, C., Gu, J.: Customizable combination of parameter-efficient modules for multi-task learning. arXiv preprint arXiv:2312.03248 (2023)
43. Wang, X., Chen, T., Ge, Q., Xia, H., Bao, R., Zheng, R., Zhang, Q., Gui, T., Huang, X.J.: Orthogonal subspace learning for language model continual learning.

In: Findings of the Association for Computational Linguistics: EMNLP 2023. pp. 10658–10671 (2023)
44. Wang, X., Zhao, H., Wang, S., Wang, H., Liu, Z.: Malora: Mixture of asymmetric low-rank adaptation for enhanced multi-task learning. In: Findings of the Association for Computational Linguistics: NAACL 2025. pp. 5609–5626 (2025)
45. Wang, Y., Yue, Z., Hua, X.S., Zhang, H.: Random boxes are open-world object detectors. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6233–6243 (2023)
46. Wang, Z., Zhang, Z., Lee, C.Y., Zhang, H., Sun, R., Ren, X., Su, G., Perot, V., Dy, J., Pfister, T.: Learning to prompt for continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 139–149 (2022)
47. Wang, Z., Che, C., Wang, Q., Li, Y., Shi, Z., Wang, M.: Smolora: Exploring and defying dual catastrophic forgetting in continual visual instruction tuning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 177–186 (2025)
48. Wu, Y., Zhao, X., Ma, Y., Wang, D., Liu, X.: Two-branch objectness-centric open world detection. In: Proceedings of the 3rd international workshop on human-centric multimedia analysis. pp. 35–40 (2022)
49. Wu, Z., Lu, Y., Chen, X., Wu, Z., Kang, L., Yu, J.: Uc-owod: Unknown-classified open world object detection. In: European conference on computer vision. pp. 193–210. Springer (2022)
50. Xue, B., Cheng, H., Yang, Q., Wang, Y., He, X.: Adapting segment anything model to aerial land cover classification with low-rank adaptation. IEEE Geoscience and Remote Sensing Letters **21**, 1–5 (2024)
51. Yang, M., Goenawan, G.J., Qin, H., Han, K., Peng, X., Yang, Y., Zhu, H.: Detecting open world objects via partial attribute assignment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20318–20328 (June 2025)
52. Yu, J., Ma, L., Li, Z., Peng, Y., Xie, S.: Open-world object detection via discriminative class prototype learning. arXiv preprint arXiv:2302.11757 (2023)
53. Zhang, J., You, J., Panda, A., Goldstein, T.: Lori: Reducing cross-task interference in multi-task low-rank adaptation. arXiv preprint arXiv:2504.07448 (2025)
54. Zhang, Q., Chen, M., Bukharin, A., Karampatziakis, N., He, P., Cheng, Y., Chen, W., Zhao, T.: Adalora: Adaptive budget allocation for parameter-efficient fine-tuning. arXiv preprint arXiv:2303.10512 (2023)
55. Zhang, S., Ni, Y., Du, J., Xue, Y., Torr, P., Koniusz, P., van den Hengel, A.: Open-world objectness modeling unifies novel object detection. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 30332–30342 (2025)
56. Zhang, X., Bai, L., Yang, X., Liang, J.: C-lora: Continual low-rank adaptation for pre-trained models. arXiv preprint arXiv:2502.17920 (2025)
57. Zhu, J., Greenewald, K., Nadjahi, K., de Ocáriz Borde, H.S., Gabrielsson, R.B., Choshen, L., Ghassemi, M., Yurochkin, M., Solomon, J.: Asymmetry in low-rank adapters of foundation models (2024)
58. Zohar, O., Wang, K.C., 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)