SINGLE-DOMAIN GENERALIZABLE OPEN WORLD OBJECT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

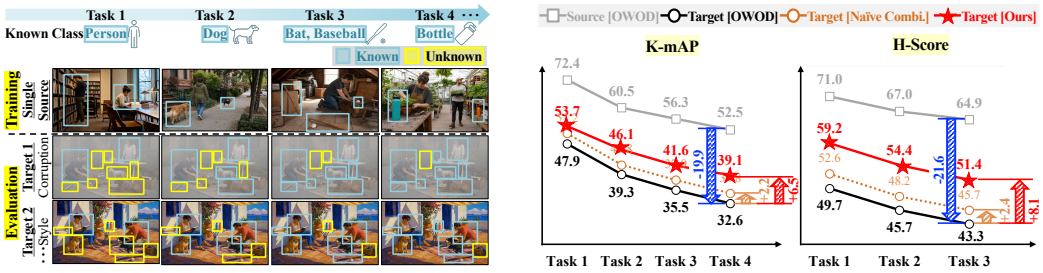
Open world object detection (OWOD) extends closed-set detection by identifying unknown objects and incrementally expanding the known class set. However, its assumption that training and evaluation data share the same domain limits applicability in real-world deployment, where visual domains continually change. Our experiments showed that even a state-of-the-art OWOD detector suffered performance degradation under such shifts, and a straightforward combination with a domain-generalized object detection (DGOD) method provided only marginal recovery. To bridge this gap, we introduce **Single-Domain Generalizable Open World Object Detection (SG-OWOD)**, where a detector trained on a single source domain must preserve its open world capability across multiple unseen target domains. To investigate these challenges, we analyzed a vision-language-based OWOD baseline. We found that domain shifts were associated with feature drift across the backbone and weaker known-class responses. Its attribute-selection statistics were also dominated by background, obscuring object-related evidence for unknown discovery. These observations motivate **RADAR** (**R**obust **A**ttribute-guided **D**etection via **A**ligned **R**epresentations), a framework built on a pretrained vision-language object detector (VLOD) with two complementary components. *Consistency-driven Low-rank Adaptation (CLA)* adapts early backbone layers by enforcing prediction consistency under style perturbations, improving known-class detection under domain shifts. *Background-Decoupled Attribute Selection (BDAS)* builds an object-centric attribute bank by separating object-like regions from background during selection, improving unknown recall when combined with CLA. Across three target domain benchmarks, RADAR improved over the strongest published baselines by at least $+6.5$ K-mAP at Task 4 and $+8.1$ H-Score at Task 3, while outperforming a naive OWOD and DGOD combination at every incremental task.

1 INTRODUCTION

Conventional object detectors operate in a *closed-set* setting, recognizing only the categories observed during training. Open world object detection (OWOD) addresses this limitation by identifying unseen objects as *unknown* and incrementally expanding the known class set without catastrophic forgetting (Joseph et al., 2021). Based on this formulation, recent OWOD studies have improved the detection of known and unknown objects in class-incremental learning (Zohar et al., 2023; Sun et al., 2024; Yavuz & Güney, 2024; Zhang et al., 2025a; Xi et al., 2025b; Li et al., 2025; Heo et al., 2026).

However, OWOD generally assumes that training and evaluation data come from the *same visual domain*, and standard benchmarks do not explicitly evaluate domain shift. In deployment, a detector trained on a *source domain* may encounter unseen *target domains* that differ in image quality, visual style, or scene composition. Whether OWOD detectors remain reliable under such shifts has not been systematically studied. Related domain-generalization settings address different parts of this gap. Domain-generalized object detection (DGOD) trains on multiple source domains to generalize to unseen target domains (Lin et al., 2021; Zhang et al., 2022), whereas Single-domain generalized object detection (Single-DGOD) (Wu & Deng, 2022; Vidit et al., 2023; Danish et al., 2024; Zhang et al., 2025b; 2026) assumes only one labeled source domain. Both settings, however, remain closed-set. Single-domain generalizable open-set object detection (SG-OSOD) (Yuan et al., 2025; 2026)

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.



(a) Description of the proposed SG-OWOD setting.

(b) Representative results of SG-OWOD on COCO-C.

Figure 1: **SG-OWOD overview and motivation.** (a) A detector learns classes incrementally from a single source domain and is evaluated on unseen target domains. (b) OWOD degrades sharply under domain shift and naïve DGOD integration recovers little, whereas RADAR improves both K-mAP and H-Score.

Table 1: **Comparison between SG-OWOD and existing scenarios.** ✓: supported, ✗: not supported.

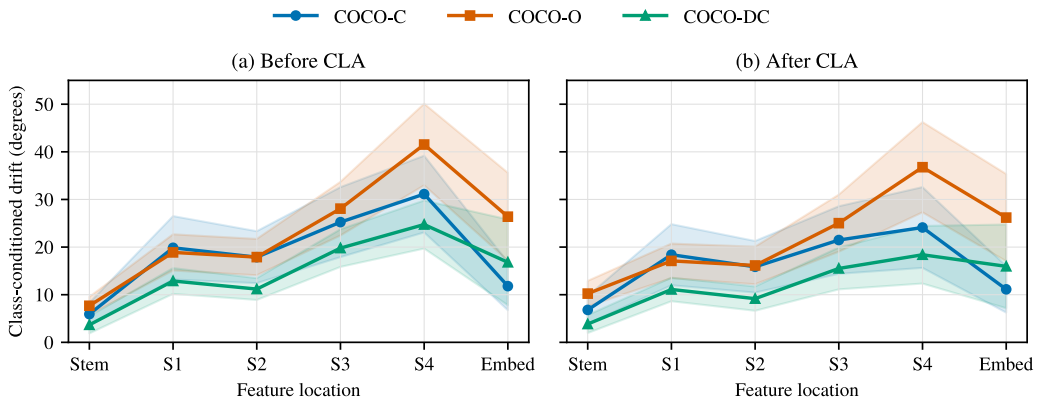
Scenario	Source Domains	Target Domains	Unknown Detection	Incremental Learning
OWOD	Single	Seen	✓	✓
DGOD	Multiple	Unseen	✗	✗
Single-DGOD	Single	Unseen	✗	✗
SG-OSOD	Single	Unseen	✓	✗
SG-OWOD (Ours)	**Single**	**Unseen**	✓	✓

additionally expands open-set to detect unknown objects under a single-source domain shift, but its known-class vocabulary is fixed and cannot expand across incremental tasks.

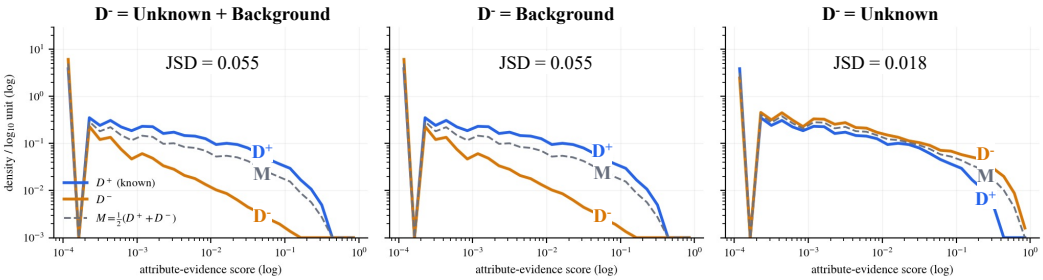
To assess whether existing OWOD and DGOD methods remain effective under unseen domain shifts, we evaluated them on unseen target domains (Figure 1b). We report known-class mAP (K-mAP), unknown recall (U-Recall), and the per-domain harmonic mean of these two metrics, averaged across target domains (H-Score). A state-of-the-art OWOD detector (Li et al., 2025) exhibited a -21.6 H-Score on COCO-C at Task 3 relative to its source-domain performance. Combining it with a single-domain DGOD method (Danish et al., 2024) yielded only a $+2.4$ H-Score, suggesting that this straightforward integration provides limited robustness against domain shifts.

To bridge this gap, we introduce **Single-Domain Generalizable Open World Object Detection (SG-OWOD)**. As illustrated in Figure 1a, a detector trained on a single source domain must detect both known and unknown objects across multiple unseen target domains while incrementally learning new classes. Table 1 summarizes the key characteristics of SG-OWOD. To investigate the challenges posed by this setting, we examined feature responses and attribute-selection statistics in a vision-language object detector (VLOD)-based OWOD (Xi et al., 2025b). Our analysis yielded two observations (Appendix C). First, domain shifts were associated with feature discrepancies from the network stem through later backbone layers and weaker known-class responses (Figure 2(a)). Second, the negative pool used for attribute selection was dominated by background, obscuring object-related evidence for unknown discovery (Figure 2(b)).

We therefore propose **RADAR** (**R**obust **A**ttribute-guided **D**etection via **A**ligned **R**epresentations), built on this base model with two complementary components. **Consistency-driven Low-rank Adaptation (CLA)** adapts early backbone layers with low-rank residuals (Hu et al., 2022) while enforcing prediction consistency across style-perturbed views, improving known-class detection under domain shifts. **Background-Decoupled Attribute Selection (BDAS)** separates object-like regions from background during attribute selection, constructing an object-centric bank for unknown scoring and improving unknown recall when combined with CLA.



(a) Feature drift with and without CLA.



(b) Attribute-response distributions across negative pools.

Figure 2: **Base model analysis motivating RADAR.** (a) Backbone drift under domain shift accompanies weaker known-class responses, motivating CLA. (b) Background statistics can obscure shared object attributes for unknown discovery, motivating BDAS.

Across all three benchmarks that cover synthetic corruption (COCO-C) (Michaelis et al., 2019), natural style shift (COCO-O) (Mao et al., 2023), and background-composition shift (COCO-DC) (Malik et al., 2024), RADAR improved over the strongest published baselines by at least $+6.5$ K-mAP at Task 4 and $+8.1$ H-Score at Task 3, while outperforming the straightforward OWOD and DGOD combination at every incremental task.

In summary, our main contributions are as follows:

- We introduce **SG-OWOD**, a novel scenario that requires unknown discovery and class-incremental learning under single-source domain shift.
- Motivated by our analysis, we propose **RADAR**, with **CLA** for known-class robustness under domain shift and **BDAS** for object-centric attribute selection and improved unknown recall when combined with CLA.
- We demonstrate that RADAR achieved the best Task 4 K-mAP and Task 3 H-Score on all three benchmarks, with minimum gains of **+6.5** K-mAP and **+8.1** H-Score.

2 Related Work

Open World Object Detection (OWOD). Since the original formulation (Joseph et al., 2021), OWOD methods have improved unknown detection through pseudo-labeling (Gupta et al., 2022), probabilistic objectness (Zohar et al., 2023), cascaded localization (Ma et al., 2023), geometric separation (Sun et al., 2024), energy-based objectness (Zhang et al., 2025a; Heo et al., 2026), and concept decomposition (Lv et al., 2026). Recent methods build OWOD on VLODs, applying LoRA to the CLIP text encoder for adaptive decision learning (Liu et al., 2024) or using their semantic

spaces for unknown detection (Xi et al., 2025a;b; Li et al., 2025). This line of work, however, inherits the assumption that training and evaluation share a single domain.

Domain-Generalized Object Detection (DGOD). DGOD generalizes detectors to unseen target domains from multiple source domains (Lin et al., 2021; Zhang et al., 2022), and Single-DGOD from a single one (Wu & Deng, 2022). Later methods improve generalization through semantic augmentation (Vidit et al., 2023), object-aware or prompt-driven learning (Lee et al., 2024; Li et al., 2024), source diversification (Danish et al., 2024), style evolution (Zhang et al., 2025b), and cross-pipeline consistency (Zhang et al., 2026). SG-OSOD adds unknown detection via adaptive subgraph searching (Yuan et al., 2025) and style–content decomposition with objectness attribution (Yuan et al., 2026), but its known-class set remains fixed. SG-OWOD couples unknown discovery and incremental class learning with generalization to unseen domains from a fixed source.

Style Perturbation and Consistency Learning. AdaIN (Huang & Belongie, 2017) transfers style through channel-wise feature statistics, while MixStyle (Zhou et al., 2021), Normalization Perturbation (Fan et al., 2023), and StableFDG (Park et al., 2023) diversify styles through mixing, perturbation, and extrapolation, respectively. DRPC (Yue et al., 2019) combines domain randomization with pyramid consistency; FDA (Yang & Soatto, 2020) transfers low-frequency image amplitudes using unlabeled target images, whereas FSDR (Huang et al., 2021) randomizes domain-variant frequency components for generalization. For Single-DGOD, DivAlign (Danish et al., 2024) aligns classification and localization predictions across diversified views. CLA builds on style perturbation and consistency learning to address the joint demands of known-class detection and attribute-based unknown scoring in SG-OWOD. Guided by observed backbone drift, it combines low-rank adaptation of early visual layers with consistency regularization of class logits, box predictions, and responses to a task-specific frozen attribute bank.

3 Problem Formulation: SG-OWOD

Single-Domain Generalizable Open World Object Detection (SG-OWOD) requires open world detection to hold under domain shift, from a single source domain $\mathcal{S}$ over T class-incremental tasks. At Task $t \in \{1, \ldots, T\}$, training uses only the labeled source dataset $\mathcal{D}_{\mathcal{S}}^{t}$, which annotates the newly introduced class subset $\mathcal{C}^t$ and leaves objects of $\mathcal{K}^{t-1}$ unlabeled, where $\mathcal{C}^p \cap \mathcal{C}^q = \varnothing$ for all distinct p, q. The cumulative known set $\mathcal{K}^t$, full class set $\mathcal{C}_{\text{all}}$, and unknown set $\mathcal{U}^t$ are

$$\mathcal{K}^t = \bigcup_{r=1}^{t} \mathcal{C}^r, \quad \mathcal{C}_{\text{all}} = \bigcup_{r=1}^{T} \mathcal{C}^r, \quad \mathcal{U}^t = \mathcal{C}_{\text{all}} \setminus \mathcal{K}^t. \tag{1}$$

SG-OWOD evaluates the detector on M unseen target domains $\{\mathcal{T}_m\}_{m=1}^{M}$ whose image distributions differ from the source, $P_{\mathcal{S}} \neq P_{\mathcal{T}_m}$ for all m, while the label semantics remain unchanged. No images or annotations from target domains are used for training, model selection, or calibration. At Task t, the detector predicts the classes in $\mathcal{K}^t$ and one *unknown* category for $\mathcal{U}^t$. Let K_m^t and R_m^t denote K-mAP and U-Recall on target domain $\mathcal{T}_m$, respectively. We report their domain-averaged K-mAP, U-Recall, and H-Score:

$$\text{K-mAP}^t = \frac{1}{M}\sum_{m=1}^{M} K_m^t, \quad \text{U-Recall}^t = \frac{1}{M}\sum_{m=1}^{M} R_m^t, \quad \text{H-Score}^t = \frac{1}{M}\sum_{m=1}^{M} \frac{2K_m^t R_m^t}{K_m^t + R_m^t}. \tag{2}$$

At Task T, $\mathcal{K}^T = \mathcal{C}_{\text{all}}$ and $\mathcal{U}^T = \varnothing$, so U-Recall and H-Score are not reported.

4 Methodology: RADAR

As shown in Figure 3, we propose **RADAR** (**R**obust **A**ttribute-guided **D**etection via **A**ligned **R**epresentations), a source-only framework that adapts a VLOD-based model (OW-OVD, Sec. 4.1) to SG-OWOD without target-domain images. Its two complementary components target feature drift (CLA, Sec. 4.2) and background-dominated attribute selection (BDAS, Sec. 4.3). At each task, BDAS first constructs the attribute bank, which is then held fixed during CLA training and inference.

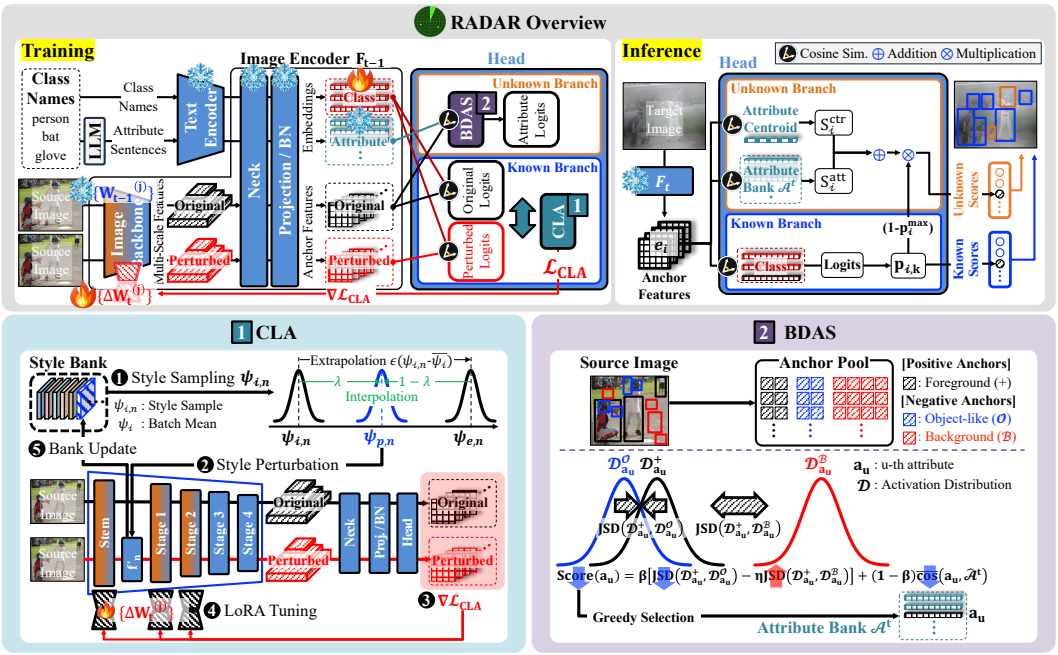


Figure 3: **Overview of RADAR.** BDAS selects an object-centric attribute bank, and CLA adapts early backbone layers under style perturbations.

4.1 Base Model: OW-OVD

We adopt OW-OVD (Xi et al., 2025b) with a YOLO-World-L backbone. Let F_t denote its image encoder after Task t which is composed of backbone, neck, and the projection with batch normalization (BN). Here F_0 is the pretrained VLOD encoder. The Task-t encoder yields a D-dimensional feature $\mathbf{e}_i \in \mathbb{R}^D$ at anchor i. We index the backbone convolutions by $j \in \mathcal{J}$; their kernels $\{W_t^{(j)}\}_{j\in\mathcal{J}}$ are the only parameters of F_t that change across tasks. The neck, projection, and BN statistics remain fixed. Let $\mathbf{c}_k \in \mathbb{R}^D$ denote the text embedding of known class $k \in \mathcal{K}^t$, and let $\mathcal{A}^t = \{\mathbf{a}_u \in \mathbb{R}^D\}_{u=1}^{|\mathcal{A}^t|}$ be the attribute embedding bank. At Task 1, the selected bank $\mathcal{A}^0$ is empty, while CLIP-encoded attribute candidates are available for selection. The known-class logit $z_{i,k}$ and attribute logit $r_{i,u}$ are

$$z_{i,k} = s\,\bar{\mathbf{c}}_k^\top \mathbf{e}_i + b, \qquad r_{i,u} = s\,\bar{\mathbf{a}}_u^\top \mathbf{e}_i + b, \tag{3}$$

where an overbar denotes L2 normalization and the frozen s, b are OW-OVD's shared logit scale and bias. The sigmoid $\sigma(\cdot)$ gives the known-class probability $p_{i,k} = \sigma(z_{i,k})$ and the attribute activation $\rho_{i,u} = \sigma(r_{i,u})$. With Γ_i indexing the $\gamma = 10$ largest activations, OW-OVD forms the unknown score by a softmax-weighted sum:

$$S_i^{\text{att}} = \sum_{u\in\Gamma_i} \frac{\exp(\rho_{i,u})}{\sum_{u'\in\Gamma_i}\exp(\rho_{i,u'})}\rho_{i,u}. \tag{4}$$

RADAR reuses Eq. 4 unchanged as the attribute-score aggregation; only the bank it reads and the features it is applied to change, and the final unknown score is Eq. 11.

4.2 Consistency-driven Low-rank Adaptation (CLA)

Our analysis showed domain discrepancy at the network stem and through the deeper backbone (Figure 2(a); Appendix C.1). Accordingly, CLA inserts LoRA residuals (Hu et al., 2022) into the stem and stage-1–2 convolutions, indexed by $\mathcal{J}_{\text{L}} \subset \mathcal{J}$. This restricts visual adaptation to early layers while leaving deeper backbone weights fixed (Appendix E.2). CLA jointly regularizes class logits, box predictions, and responses to the frozen BDAS bank under style perturbations. For $j \in \mathcal{J}_{\text{L}}$ the adapted forward pass is

$$\mathbf{y}^{(j)} = \big(W_{t-1}^{(j)} + \Delta W_t^{(j)}\big)\mathbf{x}^{(j)}, \;\; \Delta W_t^{(j)} = \tfrac{\alpha_{\text{L}}}{r_{\text{L}}} B_t^{(j)} A_t^{(j)}, \tag{5}$$

with flattened input patch $\mathbf{x}^{(j)}$, output $\mathbf{y}^{(j)}$, and trainable factors $A_t^{(j)} \in \mathbb{R}^{r_{\mathrm{L}} \times (C_{\mathrm{in}} k_h k_w)}$ and $B_t^{(j)} \in \mathbb{R}^{C_{\mathrm{out}} \times r_{\mathrm{L}}}$, where $k_h \times k_w$ is the convolution kernel size and kernels are written in flattened form. We use rank $r_{\mathrm{L}} = 8$ and scale $\alpha_{\mathrm{L}} = 8$, with $B_t^{(j)}$ initialized to zero at each task.

CLA pairs each source image with a style-perturbed view using channel-wise feature statistics (Huang & Belongie, 2017; Zhou et al., 2021). For the stem feature $\mathbf{f}_n$ of image n with channel-wise mean and standard deviation $(\boldsymbol{\mu}_n, \boldsymbol{\sigma}_n)$, let $\boldsymbol{\psi}_{\mathrm{i},n} = (\boldsymbol{\mu}_{\mathrm{i},n}, \boldsymbol{\sigma}_{\mathrm{i},n})$ be an input style drawn from a first-in-first-out bank of such statistics collected from source images. Each input style is extrapolated away from the sampled batch mean $\bar{\boldsymbol{\psi}}_{\mathrm{i}}$ (Park et al., 2023) as $\boldsymbol{\psi}_{\mathrm{e},n} = \boldsymbol{\psi}_{\mathrm{i},n} + \epsilon(\boldsymbol{\psi}_{\mathrm{i},n} - \bar{\boldsymbol{\psi}}_{\mathrm{i}})$. CLA then interpolates the two styles and applies the result to the channel-normalized feature $\hat{\mathbf{f}}_n = (\mathbf{f}_n - \boldsymbol{\mu}_n)/\boldsymbol{\sigma}_n$:

$$\boldsymbol{\psi}_{\mathrm{p},n} = \lambda \boldsymbol{\psi}_{\mathrm{i},n} + (1-\lambda)\boldsymbol{\psi}_{\mathrm{e},n}, \quad \mathbf{f}'_n = \boldsymbol{\sigma}_{\mathrm{p},n} \odot \hat{\mathbf{f}}_n + \boldsymbol{\mu}_{\mathrm{p},n}, \tag{6}$$

where $\epsilon = 5$ is the extrapolation strength, $\odot$ is channel-wise multiplication, and $\lambda \sim \mathrm{Beta}(\nu, \nu)$ with $\nu = 0.3$ (Johnson et al., 1994) pushes each view toward either the sampled or the extrapolated style. With probability 0.3, the paired view is left unperturbed.

For cross-view consistency, let Ω collect the $N_\Omega = 100$ anchors per image from each feature-pyramid level, ranked highest within that level by the original view's maximum known-class probability, and let $\mathbf{v}_i^h$ be the known-class logits, box prediction, and attribute logits for $h \in \{\mathrm{cls}, \mathrm{box}, \mathrm{att}\}$. The consistency loss is

$$\mathcal{L}_h^{\mathrm{align}} = \frac{1}{|\Omega|} \sum_{i \in \Omega} \mathrm{SL}_1\Big(\mathbf{v}_i^{h,\mathrm{pert}}, \mathrm{sg}[\mathbf{v}_i^{h,\mathrm{orig}}]\Big), \tag{7}$$

where SL_1 is the Smooth-L1 loss averaged over response dimensions. The same number of anchors is selected at each level, so all levels contribute equally. The superscripts orig and pert index two views of the same current model, and $\mathrm{sg}[\cdot]$ stops gradients through the original-view consistency targets. With the total consistency loss $\mathcal{L}_{\mathrm{align}} = \mathcal{L}_{\mathrm{cls}}^{\mathrm{align}} + \mathcal{L}_{\mathrm{box}}^{\mathrm{align}} + \mathcal{L}_{\mathrm{att}}^{\mathrm{align}}$ and the OW-OVD detection losses $\mathcal{L}_{\mathrm{det}}^{\mathrm{orig}}$ and $\mathcal{L}_{\mathrm{det}}^{\mathrm{pert}}$, CLA minimizes

$$\mathcal{L}_{\mathrm{CLA}} = \lambda_{\mathrm{det}} \mathcal{L}_{\mathrm{det}}^{\mathrm{orig}} + \lambda_{\mathrm{aug}} \mathcal{L}_{\mathrm{det}}^{\mathrm{pert}} + \lambda_{\mathrm{align}} \mathcal{L}_{\mathrm{align}}, \tag{8}$$

with $\lambda_{\mathrm{det}} = 0.05$, $\lambda_{\mathrm{aug}} = 0.0125$, and $\lambda_{\mathrm{align}} = 0.1$. Only the LoRA factors and current-task class embeddings are trainable; earlier class embeddings and $\mathcal{A}^t$ remain frozen. After Task t, each LoRA update is merged as $W_t^{(j)} = W_{t-1}^{(j)} + \tau_t \Delta W_t^{(j)}$, where $\tau_1 = 1$ and $\tau_{t>1} = 0.1$ (Appendix E.3); convolutions outside $\mathcal{J}_{\mathrm{L}}$ keep $W_t^{(j)} = W_{t-1}^{(j)}$, and the merged set $\{W_t^{(j)}\}_{j \in \mathcal{J}}$ maps F_{t-1} to F_t. The reduced update after Task 1 limits the magnitude of each new task's visual update.

4.3 Background-Decoupled Attribute Selection (BDAS)

OW-OVD selects attributes by minimizing the Jensen–Shannon divergence (JSD) between their activations on positive anchors, which are assigned to a current-task object, and on all remaining negative anchors. Our analysis showed background-dominated negative statistics (Figure 2(b); Appendix C.2). Selection can therefore favor attributes that respond alike to known objects and background, and may discard those shared by known and unknown objects.

BDAS instead splits negative anchors into an object-like set $\mathcal{O}$ and a background set $\mathcal{B}$. At Task 1, the detector initialized with F_0 mines candidate unknown boxes from source images using the full set of CLIP-encoded attribute candidates. At later tasks, the frozen Task-$(t-1)$ detector uses F_{t-1} and $\mathcal{A}^{t-1}$ for mining. After filtering large boxes and overlaps with confident known detections, negative anchors inside the retained boxes form $\mathcal{O}$ and the rest form $\mathcal{B}$. Labeled current-task foreground anchors are excluded from both $\mathcal{O}$ and $\mathcal{B}$. The object-like regions provide a proxy for unlabeled objects.

Following OW-OVD, for each candidate attribute $\mathbf{a} = \mathbf{a}_u$, let $\mathcal{D}_{\mathbf{a}}^{+}$, $\mathcal{D}_{\mathbf{a}}^{\mathcal{O}}$, and $\mathcal{D}_{\mathbf{a}}^{\mathcal{B}}$ denote the normalized distributions of its activations $\rho_{i,u}$ over the positive, object-like, and background pools, respectively. These activation distributions are collected with the frozen pretrained encoder F_0 at every task. The working bank $\hat{\mathcal{A}}^t$ starts from $\mathcal{A}^{t-1}$, or is empty at Task 1. BDAS scores each

candidate by

$$\text{score}(\mathbf{a}) = \beta\Big[\text{JSD}\big(\mathcal{D}_{\mathbf{a}}^{+}, \mathcal{D}_{\mathbf{a}}^{\mathcal{O}}\big) - \eta\,\text{JSD}\big(\mathcal{D}_{\mathbf{a}}^{+}, \mathcal{D}_{\mathbf{a}}^{\mathcal{B}}\big)\Big] + (1-\beta)\,\overline{\cos}\big(\mathbf{a}, \hat{\mathcal{A}}^{t}\big), \tag{9}$$

where $\overline{\cos}$ is the mean cosine similarity to the working bank and is zero for an empty bank. Minimizing the first JSD term favors similar responses on labeled objects and object-like regions; the subtracted background JSD favors distinct foreground and background responses, while the cosine term discourages redundancy. We use $(\beta, \eta) = (0.5, 1)$ (Appendix E.1).

Candidates with the lowest scores are added greedily, with the redundancy term recomputed after each addition, using OW-OVD's budget of 25 attributes per newly introduced class. The resulting bank $\mathcal{A}^t$ is then frozen together with the earlier class embeddings.

4.4 Inference

A test image passes once through F_t. Known-class probabilities $p_{i,k}$ come directly from the CLA-adapted feature $\mathbf{e}_i$. For unknowns, the same frozen BDAS bank $\mathcal{A}^t$ used during CLA training provides S_i^{att} through Eq. 4.

To capture the direction shared across the selected attributes, we form a native-view score using the normalized mean of the BDAS bank:

$$\mathbf{g}_t = \frac{1}{|\mathcal{A}^t|}\sum_{\mathbf{a}\in\mathcal{A}^t}\mathbf{a}, \qquad S_i^{\text{ctr}} = \sigma(s\,\bar{\mathbf{g}}_t^{\top}\mathbf{e}_i + b). \tag{10}$$

Let $p_i^{\max} = \max_{k\in\mathcal{K}^t} p_{i,k}$. The final unknown score is

$$S_i^{\text{unk}} = \xi_{\text{u}}(1 - p_i^{\max})\left(S_i^{\text{att}} + S_i^{\text{ctr}}\right), \qquad \xi_{\text{u}} = 15. \tag{11}$$

The factor $1 - p_i^{\max}$ suppresses anchors already explained by a known class, while the scale ξ_{u}, fixed across tasks and domains, sets the unknown branch's score scale relative to the known branch and therefore affects filtering and ranking.

5 Experiments

5.1 Experimental Setting

Datasets and Metrics. We evaluate three SG-OWOD benchmarks under the S-OWODB split (Gupta et al., 2022; Zohar et al., 2023): 80 COCO classes over four tasks (19/21/20/20). The unseen target domains are 15 corruptions at severity 3 (COCO-C (Michaelis et al., 2019)), six appearance shifts (COCO-O (Mao et al., 2023)), and four background-composition shifts (COCO-DC (Malik et al., 2024)). At IoU 0.5 we report K-mAP, U-Recall, and H-Score (columns K, U-R, H) of Sec. 3, with K split at Tasks 2–4 into AP_{50} for previously learned (Prev) and current-task (Cur) classes. Source metrics use the full clean COCO test set for COCO-C and COCO-O, and COCO-DC's paired 1,127-image clean subset. Target metrics are unweighted domain means.

Implementation. We implemented RADAR in PyTorch and MMDetection on OW-OVD with a YOLO-World-L backbone, trained on a single NVIDIA H200 GPU. CLA used AdamW for one source epoch with batch size 16, at 10^{-4} for the LoRA factors and 2×10^{-5} for current-task embeddings, with the training bank frozen (Appendices A and D).

5.2 Experimental Results

Table 2 reports target results. On COCO-C, RADAR led K-mAP, U-Recall, and H-Score at Tasks 1–3 and K-mAP at Task 4; its Task 3 H-Score was 51.4 ($+8.1$ H-Score) and Task 4 K-mAP was 39.1 ($+6.5$ K-mAP). On COCO-O, it again led all reported metrics, with Task 3 H-Score of 51.0 ($+11.1$ H-Score) and Task 4 K-mAP of 34.7 ($+8.5$ K-mAP). On COCO-DC, Task 1 U-Recall reached 91.5 although K-mAP trailed OWOBJ and OW-OVD; RADAR led U-Recall and H-Score at Tasks 1–3, while stronger previous-class retention helped keep K-mAP best at Tasks 2–4. Its Task 3 H-Score was 75.8 ($+10.0$ H-Score), and Task 4 K-mAP was 61.5 ($+11.0$ K-mAP). Overall, Task 4 K-mAP and Task 3 H-Score gains over the strongest published results (OVOW) were at least $+6.5$ K-mAP and $+8.1$ H-Score, respectively. Source results are in Appendix B.

Table 2: **SG-OWOD results on the target domains.** Each block is the unweighted mean over its domains and H follows Eq. 2. Best and second-best are bold and underlined, excluding the Reference row, our own OWOD+DGOD construction. † marks methods we re-implemented, as no code was public at the time of our evaluation. Ties share marks.

			Task 1			Task 2					Task 3					Task 4		
Domain	Scenario	Model	Cur	U-R	H	Prev	Cur	K	U-R	H	Prev	Cur	K	U-R	H	Prev	Cur	K
Target (COCO-C)	SG-OSOD	ASGS†	43.1	22.2	28.5	27.3	25.3	26.2	14.4	18.0	11.0	24.7	15.6	9.5	11.3	17.1	21.1	18.1
	OWOD	PROB	43.6	22.6	28.3	35.4	15.7	25.0	29.6	26.1	23.0	14.3	20.1	27.4	22.2	18.6	12.3	17.1
		OWOBJ	48.4	18.0	25.4	39.2	20.3	29.3	19.9	22.9	26.9	18.3	24.0	14.6	17.3	23.2	17.2	21.7
		DEUS†	40.7	60.8	47.5	26.4	18.5	22.3	54.9	30.7	21.4	17.5	20.1	55.4	28.8	20.1	15.5	18.9
	OWOD+VLOD	OW-OVD	51.9	23.8	32.3	41.4	38.3	39.8	29.6	33.8	39.3	32.7	37.1	25.6	30.1	29.6	31.5	30.1
		OVOW	47.9	52.4	49.7	47.8	31.5	39.3	55.3	45.7	38.6	29.4	35.5	56.6	43.3	35.3	24.4	32.6
	Reference	OVOW+DivAlign	51.2	54.5	52.6	51.1	33.4	41.8	57.2	48.2	40.9	31.9	37.9	58.0	45.7	37.8	25.8	34.8
	SG-OWOD	**RADAR (Ours)**	**53.7**	**67.1**	**59.2**	**53.1**	**39.8**	**46.1**	**67.1**	**54.4**	**45.4**	**34.1**	**41.6**	**68.0**	**51.4**	**40.8**	**34.0**	**39.1**
Target (COCO-O)	SG-OSOD	ASGS†	28.7	35.6	30.1	16.8	18.0	17.4	35.8	21.9	6.8	11.3	8.3	27.7	12.3	8.6	11.9	9.5
	OWOD	PROB	27.4	32.4	28.2	21.1	9.6	15.1	47.3	21.5	12.6	4.3	9.8	42.1	15.2	9.3	6.8	8.7
		OWOBJ	32.1	28.2	28.4	24.1	12.5	18.0	36.4	22.1	15.9	6.4	12.7	23.4	13.3	11.5	9.0	10.8
		DEUS†	29.8	86.2	43.1	18.8	18.3	18.5	83.1	29.3	16.5	10.2	14.9	83.6	24.7	15.0	14.2	14.7
	OWOD+VLOD	OW-OVD	41.2	42.1	38.9	31.2	35.0	33.3	54.1	38.7	32.9	25.6	30.7	42.8	30.6	24.3	26.2	24.9
		OVOW	35.4	81.1	46.3	33.7	27.4	30.6	83.9	42.1	30.4	20.7	27.7	85.6	39.9	27.4	22.1	26.2
	Reference	OVOW+DivAlign	39.1	81.9	50.0	37.2	29.9	33.8	84.4	45.9	32.6	21.4	29.9	86.3	42.8	29.5	21.3	27.8
	SG-OWOD	**RADAR (Ours)**	**44.0**	**92.2**	**58.1**	**42.5**	**38.4**	**41.0**	**90.8**	**54.9**	**40.2**	**27.1**	**36.5**	**92.6**	**51.0**	**35.7**	**30.7**	**34.7**
Target (COCO-DC)	SG-OSOD	ASGS†	69.3	38.2	49.2	48.4	42.8	45.5	33.0	38.2	18.1	43.8	26.6	35.9	30.5	30.7	43.1	33.8
	OWOD	PROB	72.0	31.2	43.4	61.0	27.8	43.6	50.9	46.8	41.2	26.0	36.1	52.5	42.4	36.5	27.2	34.2
		OWOBJ	**75.1**	30.2	43.1	65.4	39.1	51.6	42.0	46.1	48.1	35.7	44.0	39.3	41.3	43.5	39.2	42.5
		DEUS†	70.1	86.9	77.5	56.7	47.1	51.6	80.0	62.6	50.8	43.7	48.6	92.7	63.5	47.5	46.0	47.1
	OWOD+VLOD	OW-OVD	73.6	54.4	62.6	62.0	**55.7**	58.7	61.6	60.1	57.8	**60.5**	58.7	67.8	62.9	46.8	57.4	49.4
		OVOW	69.7	80.6	74.7	69.6	42.2	55.2	81.9	65.9	54.3	50.2	53.0	87.1	65.8	52.5	44.4	50.5
	Reference	OVOW+DivAlign	46.6	54.4	50.1	46.9	27.3	36.6	53.8	43.5	36.1	31.3	34.5	59.7	43.6	34.4	25.0	32.1
	SG-OWOD	**RADAR (Ours)**	73.1	**91.5**	**81.2**	**73.4**	55.5	**64.0**	**90.2**	**74.9**	**64.4**	58.4	**62.4**	**96.6**	**75.8**	**61.5**	**61.7**	**61.5**

5.3 Ablation and Analysis

Table 3: **Source and COCO-C ablations at Tasks 1–4.** Bold/underline mark best/second-best values, excluding OW-OVD. Additional target results are in Table 19.

	Source										**Target** (COCO-C)									
	Task 1			Task 2			Task 3			Task 4	Task 1			Task 2			Task 3			Task 4
Setting	Cur	U-R	H	K	U-R	H	K	U-R	H	K	Cur	U-R	H	K	U-R	H	K	U-R	H	K
OW-OVD	76.3	29.1	42.1	61.0	39.7	48.1	58.1	33.8	42.7	48.2	51.9	23.8	32.3	39.8	29.6	33.8	37.1	25.6	30.1	30.1
Base†	**76.3**	72.7	74.4	60.4	79.2	68.5	57.2	77.8	66.0	47.0	51.8	60.0	55.0	39.1	66.4	48.9	36.5	64.9	46.4	29.6
+CLA	75.7	69.2	72.3	**66.4**	76.7	71.2	**61.7**	77.0	68.5	**59.6**	**53.9**	57.3	55.2	**46.5**	63.4	53.5	**42.0**	62.7	50.1	**39.6**
+BDAS	76.0	77.8	**76.9**	60.2	**79.5**	68.5	56.8	81.1	66.8	46.7	51.5	66.5	57.4	38.9	**67.2**	49.0	36.1	67.6	46.7	29.1
+BDAS+CLA (Ours)	75.2	**78.0**	76.6	65.9	78.5	**71.6**	61.1	**81.2**	**69.7**	59.0	53.7	**67.1**	**59.2**	46.1	67.1	**54.4**	41.6	**68.0**	**51.4**	39.1

Ablation. Table 3 reports the source and COCO-C ablations; the COCO-O and COCO-DC results are in Table 19 of Appendix E. The +BDAS+CLA row is RADAR in Table 2. These comparisons examine CLA and BDAS, which address the diagnosed feature drift and background-dominated selection statistics. Base† swaps OW-OVD's unknown score for ours (Eq. 11); rows prefixed by + are independent combinations on top of Base. On COCO-C, adding CLA to BDAS yielded $+7.2$ K-mAP at Task 2, $+5.5$ K-mAP at Task 3, and $+10.0$ K-mAP at Task 4. Adding BDAS to CLA yielded $+9.8$ U-Recall at Task 1, $+3.7$ U-Recall at Task 2, and $+5.3$ U-Recall at Task 3. These comparisons supported complementary roles for CLA in later-task known-class detection and BDAS in unknown recall. Hyperparameter analyses using source data are provided in Appendix E.

Qualitative Analysis. Figure 4 compares RADAR with OW-OVD and OVOW on COCO-C and COCO-O. In the highlighted COCO-C example, RADAR detected the backpack missed by both baselines. In the COCO-O example, it detected the highlighted unknown object missed by OVOW.

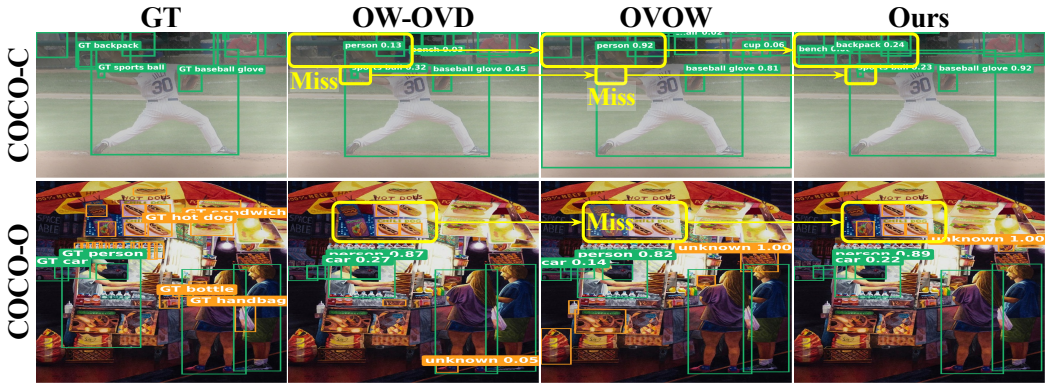


Figure 4: **Qualitative comparison.** Green and orange boxes are known-class and unknown detections; yellow boxes highlight the largest differences across methods.

Table 4: **Cityscapes → Foggy Cityscapes under *het-sem*.** Bold/underline mark best/second-best values. † denotes VLM results recomputed with the top 100 detections per image.

		$u=3$				$u=4$				$u=5$			
Pretraining	Method	mAP_k↑	AR_u↑	WI↓	AOSE↓	mAP_k↑	AR_u↑	WI↓	AOSE↓	mAP_k↑	AR_u↑	WI↓	AOSE↓
Vision-only (closed-vocab.)	D-DETR [ICLR'21]	37.7	0.0	0.376	454	36.7	0.0	0.508	899	34.0	0.0	0.598	1248
	OpenDet [CVPR'22]	37.1	2.3	0.371	268	35.5	1.5	0.499	524	33.3	2.0	0.621	1048
	OW-DETR [CVPR'22]	34.3	2.2	0.482	251	35.4	2.0	0.702	513	30.5	2.0	0.813	922
	PROB [CVPR'23]	28.7	2.8	0.393	209	28.5	3.4	0.564	340	33.3	2.7	0.674	777
	CAT [CVPR'23]	33.1	2.1	0.427	202	33.0	2.3	0.581	416	33.8	2.1	0.681	725
	SOMA [ICCV'23]	34.6	1.5	0.327	193	33.5	1.8	0.433	371	31.1	1.6	0.655	719
	SS-OWFormer [AAAI'24]	40.5	2.9	0.424	182	38.7	2.2	0.629	375	35.7	2.8	0.697	682
	ASGS [ICCV'25]	41.1	2.9	**0.314**	185	38.9	2.7	0.431	309	37.5	2.4	**0.518**	672
	DOAT [AAAI'26]	**42.7**	4.7	0.353	159	**42.1**	4.5	**0.428**	339	**39.2**	4.3	0.581	662
VLM (open-vocab.)	OW-OVD [CVPR'25]†	41.0	29.2	2.773	1578	39.6	27.2	3.874	3294	36.3	25.0	5.082	6673
	OVOW [BMVC'25]†	37.9	26.5	1.677	1515	36.2	22.8	2.645	3552	33.5	20.3	3.641	7506
	RADAR (Ours)†	42.5	**46.7**	1.049	**158**	41.0	**43.9**	1.411	**308**	35.6	**41.0**	1.884	**597**

These examples illustrated known-class detection and unknown discovery under domain shifts.

5.4 Generalization under Weather Shift

Beyond our main benchmarks, we evaluated Cityscapes → Foggy Cityscapes (Cordts et al., 2016; Sakaridis et al., 2018), which applies synthetic fog to real driving images to emulate a plausible weather shift. Table 4 shows that RADAR achieved the highest unknown recall (AR_u) among all baselines. It also recorded the lowest Absolute Open-Set Error (AOSE), which counts unknown objects misclassified as known, and the second-best known-class mAP (mAP_k) at $u = 3$ and 4. Despite higher Wilderness Impact (WI) than vision-only baselines, these results demonstrate competitive transfer beyond our main benchmarks to a weather shift.

6 Conclusion

In this paper, we introduce SG-OWOD, which requires a detector to learn incrementally from one source domain and detect known and unknown objects across unseen domains. A strong OWOD baseline degraded under domain shift, while adding a DGOD method yielded only marginal recovery. Our analysis found backbone drift alongside weaker known-class responses and background-dominated negative statistics for attribute selection. To address these challenges, we propose RADAR, combining CLA for style-consistent early-backbone adaptation with BDAS for object-centric attribute selection. Experiments on COCO-C, COCO-O, and COCO-DC show improved known-class detection and unknown recall without target-domain training data, providing a step toward reliable open world detection across unseen domains.

AI USE STATEMENT

In this work, generative AI tools were used to provide feedback on research methodology and experimental design, assist with the interpretation of experimental results, and support translation. They were additionally used for literature search and summarization, code editing, manuscript language polishing, and LaTeX formatting. The authors independently verified all AI-assisted analyses, experimental claims, citations, and code, and manually reviewed the complete manuscript. The authors take full responsibility for the final content of this work.

ETHICS STATEMENT

This work uses existing publicly available object-detection benchmarks and does not collect new data or involve interaction with human participants. Object-detection systems may be used in surveillance or other sensitive monitoring applications, where domain bias and detection errors could cause harm. Our experiments are intended for benchmark research rather than deployment. Any real-world deployment should include domain-specific validation, privacy safeguards, human oversight, and compliance with the licenses of the original datasets and applicable regulations.

REPRODUCIBILITY STATEMENT

The SG-OWOD protocol is defined in Section 3, and the full method is described in Section 4. Notation and implementation details are given in Appendix A; the benchmark construction, incremental task configurations, and per-domain results are given in Appendix B; the complete training and inference pipeline is given in Appendix D; configured hyperparameters, source-only CLA sensitivity controls, and component ablations are reported in Appendix E; and trainable parameter counts and inference computation are described in Appendix F. Source code, the frozen attribute bank, and trained checkpoints will be released upon publication.

REFERENCES

Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Muhammad Sohail Danish, Muhammad Haris Khan, Muhammad Akhtar Munir, M. Saquib Sarfraz, and Mohsen Ali. Improving single domain-generalized object detection: A focus on diversification and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Qi Fan, Mattia Segu, Yu-Wing Tai, Fisher Yu, Chi-Keung Tang, Bernt Schiele, and Dengxin Dai. Towards Robust Object Detection Invariant to Real-World Domain Shifts. In *International Conference on Learning Representations*, 2023.

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 (CVPR)*, 2022.

Jun-Woo Heo, Keonhee Park, and Gyeong-Moon Park. Detecting unknown objects via energy-based separation for open world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27558–27567, 2026.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. FSDR: Frequency Space Domain Randomization for Domain Generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

Xun Huang and Serge Belongie. Arbitrary Style Transfer in Real-Time With Adaptive Instance Normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

Norman L. Johnson, Samuel Kotz, and N. Balakrishnan. Beta distributions. *Continuous Univariate Distributions*, 2:210–275, 1994.

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 (CVPR)*, 2021.

Wooju Lee, Dasol Hong, Hyungtae Lim, and Hyun Myung. Object-aware domain generalization for object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

Deng Li, Aming Wu, Yaowei Wang, and Yahong Han. Prompt-driven dynamic object-centric learning for single domain generalization in object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Zizhao Li, Zhengkang Xiang, Joseph West, and Kourosh Khoshelham. From open vocabulary to open world: Teaching vision language models to detect novel objects. In *36th British Machine Vision Conference 2025, BMVC 2025, Sheffield, UK, November 24–27, 2025*. BMVA, 2025.

Chuang Lin, Zehuan Yuan, Sicheng Zhao, Peize Sun, Changhu Wang, and Jianfei Cai. Domain-invariant disentangled network for generalizable object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Lihao Liu, Juexiao Feng, Hui Chen, Ao Wang, Lin Song, Jungong Han, and Guiguang Ding. YOLO-UniOW: Efficient universal open-world object detection. *arXiv preprint arXiv:2412.20645*, 2024.

Xueqiang Lv, Shizhou Zhang, Yinghui Xing, Di Xu, Peng Wang, and Yanning Zhang. Knowing the unknown: Interpretable open-world object detection via concept decomposition model. In *Forty-third International Conference on Machine Learning*, 2026.

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 (CVPR)*, 2023.

Hashmat Shadab Malik, Muhammad Huzaifa, Muzammal Naseer, Salman Khan, and Fahad Shahbaz Khan. Objectcompose: Evaluating resilience of vision-based models on object-to-background compositional changes. In *Proceedings of the Asian Conference on Computer Vision*, pp. 372–389, 2024.

Xiaofeng Mao, Yuefeng Chen, Yao Zhu, Da Chen, Hang Su, Rong Zhang, and Hui Xue. Coco-o: A benchmark for object detectors under natural distribution shifts. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6339–6350, 2023.

Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S Ecker, Matthias Bethge, and Wieland Brendel. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*, 2019.

Jungwuk Park, Dong-Jun Han, Jinho Kim, Shiqiang Wang, Christopher G. Brinton, and Jaekyun Moon. StableFDG: Style and Attention Based Learning for Federated Domain Generalization. In *Advances in Neural Information Processing Systems*, 2023.

Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018.

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 (CVPR)*, pp. 17302–17312, June 2024.

Vidit Vidit, Martin Engilberge, and Mathieu Salzmann. CLIP the gap: A single domain generalization approach for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Aming Wu and Cheng Deng. Single-domain generalized object detection in urban scene via cyclic-disentangled self-distillation. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pp. 847–856, 2022.

Xing Xi, Xing Fu, Weiqiang Wang, and Ronghua Luo. OW-VAP: Visual attribute parsing for open world object detection. In *Forty-second International Conference on Machine Learning*, 2025a.

Xing Xi, Yangyang Huang, Ronghua Luo, and Yu Qiu. OW-OVD: Unified open world and open vocabulary object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25454–25464, June 2025b.

Yanchao Yang and Stefano Soatto. FDA: Fourier Domain Adaptation for Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

Mısra Yavuz and Fatma Güney. O1o: Grouping of known classes to identify unknown objects as odd-one-out. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pp. 614–629, December 2024.

Yuxuan Yuan, Luyao Tang, Yixin Chen, Chaoqi Chen, Yue Huang, and Xinghao Ding. ASGS: Single-domain generalizable open-set object detection via adaptive subgraph searching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 20911–20921, October 2025.

Yuxuan Yuan, Lichen Wei, Luyao Tang, Chaoqi Chen, Zheyuan Cai, Yue Huang, and Xinghao Ding. Decompose and attribute: Boosting generalizable open-set object detection via objectness score. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12286–12294, 2026.

Xiangyu Yue, Yang Zhang, Sicheng Zhao, Alberto Sangiovanni-Vincentelli, Kurt Keutzer, and Boqing Gong. Domain Randomization and Pyramid Consistency: Simulation-to-Real Generalization Without Accessing Target Domain Data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.

Peng Zhang, Xiang Yuan, and Gong Cheng. Unified interaction consistency learning for single-source domain-generalized object detection in urban scene. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12674–12682, 2026.

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, June 2025a.

Xingxuan Zhang, Zekai Xu, Renzhe Xu, Jiashuo Liu, Peng Cui, Weitao Wan, Chong Sun, and Chen Li. Towards domain generalization in object detection, 2022.

Zihao Zhang, Aming Wu, and Yahong Han. Style evolving along chain-of-thought for unknown-domain object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14225–14234, 2025b.

Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain Generalization with MixStyle. 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 (CVPR)*, 2023.

SINGLE-DOMAIN GENERALIZABLE OPEN WORLD OBJECT DETECTION

APPENDIX

Supplementary Outline

A. Notation and Implementation . . . 14
B. Benchmarks and Results . . . 15
B.1 Weather-Shift Evaluation Protocol . . . 15
B.2 [Ours] Configuration . . . 15
B.3 [Ours] Source Domain Results . . . 15
B.4 [Ours] Per-Domain Results . . . 16
C. Motivation for RADAR . . . 17
C.1 [CLA] Early Feature Drift . . . 17
C.2 [BDAS] Background-Dominated Selection . . . 18
D. Overall RADAR Pipeline . . . 18
E. Hyperparameters and Ablations . . . 19
E.1 BDAS Coefficients (β, η) . . . 20
E.2 CLA Scope $\mathcal{J}_{\mathrm{L}}$ . . . 20
E.3 CLA Merge τ_t . . . 21
E.4 Inference Unknown-Score Scale ξ_{u} . . . 22
E.5 Unknown-Score Decomposition . . . 22
E.6 Component Ablations . . . 23
F. Computational Costs . . . 25

A NOTATION AND IMPLEMENTATION

Table 5 summarizes the core notation used throughout this supplementary material.

Table 5: Notation.

Symbol	Meaning
$t, \mathcal{D}_S^t, \mathcal{C}^t, \mathcal{K}^t, \mathcal{U}^t$	task, labeled source data, and new/known/unknown class sets
$\mathbf{e}_i, \mathbf{c}_k, \mathbf{a}_u, p_{i,k}$	anchor feature, known/attribute embeddings, and known probability
$\rho_{i,u}, S_i^{\mathrm{att}}$	attribute activation and aggregated attribute score
$\mathcal{O}, \mathcal{B}, \beta, \eta$	BDAS anchor pools, diversity balance, and subtraction weight
$\mathcal{J}_{\mathrm{L}}, \Delta W_t^{(j)}$ $r_{\mathrm{L}}, \alpha_{\mathrm{L}}, \tau_t$	CLA scope, low-rank update, rank, scale, and merge coefficient
$\epsilon, \nu, \Omega, N_\Omega$	style parameters and selected consistency anchors
S_i^{ctr}	native response to the normalized attribute-bank centroid
$p_i^{\max}, S_i^{\mathrm{unk}}, \xi_{\mathrm{u}}$	maximum known probability, final unknown score, and its scale
K, U-R, U-AP, U-P, H	known mAP, unknown recall/AP/precision, and harmonic mean

Implementation. We implement RADAR on the OW-OVD (Xi et al., 2025b) with the YOLO-World-L backbone. The method-wise settings are summarized below.

BDAS. BDAS represents the source activation distributions with 10,000-bin histograms. At Task t, it retains 25 attributes for each newly introduced class in $\mathcal{C}^t$.

CLA. CLA applies rank-8, scale-8 LoRA residuals to 25 convolutions in the stem and Stages 1–2. During adaptation, batch-normalization layers use the frozen running statistics of the base detector. This prevents style-perturbed views from changing the backbone's normalization state, so the backbone update is restricted to the designated LoRA residuals. We optimize 191,192 LoRA parameters and $512|\mathcal{C}^t|$ current-task embedding parameters. At Task 1, these comprise 191,192 and 9,728 parameters, respectively, for 200,920 parameters in total (0.4% of the 47.3M-parameter model).

The source-style FIFO bank stores 256 channel-statistic samples after a 64-sample warm-up, and a mini-batch remains unperturbed with probability 0.3. The consistency objective uses the top $N_\Omega = 100$ anchors at each pyramid level. We train for one epoch with AdamW using learning rates of 10^{-4} for LoRA and 2×10^{-5} for the class embeddings of current task. The loss weights are $(\lambda_{\mathrm{det}}, \lambda_{\mathrm{aug}}, \lambda_{\mathrm{align}}) = (0.05, 0.0125, 0.1)$, with a 3% linear warm-up, cosine decay, and a gradient-norm limit of 10.

Inference. All learned weights and attribute banks are frozen during evaluation, and every image is resized to 640×640. Detection candidates with scores above 0.001 are processed by class-aware NMS at IoU 0.7, after which the 300 highest-scoring detections per image are retained. The OWOD evaluator then includes only retained detections with scores of at least 0.01 in the reported metrics.

Environment. All experiments were conducted on a single NVIDIA H200 GPU using Python 3.10.20, PyTorch 2.1.0, CUDA 12.1, cuDNN 8.9.2, MMDetection 3.3.0, MMCV 2.1.0, MMEngine 0.10.7, and MMYOLO 0.6.0. CLA training uses *seed 0* for Python, NumPy, PyTorch, and data loading; attribute selection uses *seed 42*. cuDNN benchmarking is disabled, although deterministic CUDA algorithms are not enforced. The controlled CLA scope experiments use seed 0, with an additional seed for the two-scope check described in Appendix E.2.

Table 6: Configuration of the COCO-C, COCO-O, and COCO-DC.

S-OWODB incremental configuration					
Task configuration		Task 1	Task 2	Task 3	Task 4
New classes		19	21	20	20
Cumulative known classes		19	40	60	80
Source training images		89,490	55,870	39,402	38,903

Benchmark	Role	# Domains	Domain list	Eval. images/domain
COCO	Source	1	clean	4,952
COCO-C	Target	15	brightness, contrast, defocus blur, elastic transform, fog, frost, gaussian noise, glass blur, impulse noise, JPEG compression, motion blur, pixelate, shot noise, snow, zoom blur	4,952
COCO-O	Target	6	cartoon, handmake, painting, sketch, tattoo, weather	918–1,996
COCO-DC (orig.)	ID ref.	1	clean COCO subset	1,127
COCO-DC	Target	4	adversarial, blip, color, texture	1,127

Benchmark	Task	Classes	#
S-OWODB	T1	aeroplane, bicycle, bird, boat, bus, car, cat, cow, dog, horse, motorbike, sheep, train, elephant, bear, zebra, giraffe, truck, person	19
	T2	traffic light, fire hydrant, stop sign, parking meter, bench, chair, dining table, potted plant, backpack, umbrella, handbag, tie, suitcase, microwave, oven, toaster, sink, refrigerator, bed, toilet, sofa	21
	T3	frisbee, skis, snowboard, sports ball, kite, baseball bat, baseball glove, skateboard, surfboard, tennis racket, banana, apple, sandwich, orange, broccoli, carrot, hot dog, pizza, donut, cake	20
	T4	laptop, mouse, remote, keyboard, cell phone, book, clock, vase, scissors, teddy bear, hair drier, toothbrush, wine glass, cup, fork, knife, spoon, bowl, tv monitor, bottle	20

B Benchmarks and Results

B.1 Weather-Shift Evaluation Protocol

Section 5.4 and Table 4 report the Cityscapes $\rightarrow$ Foggy Cityscapes evaluation under the *het-sem* protocol (Yuan et al., 2025; 2026). This protocol uses 2,846 labeled source images and a pool of 2,975 corrected target images, with three known classes and the first $u \in \{3, 4, 5\}$ of five unknown classes. The released filtering procedure yields 1,314, 1,951, and 2,975 evaluation images for $u = 3$, 4, and 5, respectively. The three VLM methods retain at most 100 post-NMS detections per image for all reported metrics.

B.2 [Ours] Configuration

Table 6 summarizes the configuration of our main benchmark with shared S-OWODB task split. These COCO-family benchmarks use the same labeled clean COCO source.

B.3 [Ours] Source Domain Results

All baseline models in **Table 7** were trained from the same clean COCO source as shown in **Table 6**. COCO-C and COCO-O use the full 4,952-image clean test set as their source domain evaluation. In contrast, the four COCO-DC target domains were constructed from the same 1,127-image subset of clean COCO. We therefore use the corresponding clean COCO images called *original domain* as the paired source domain evaluation result rather than treating them as a separate training source. The 1,127 *original domain* images are byte-identical to their counterparts in the clean COCO test

Table 7: Source results on full clean COCO and the paired COCO-DC original subset. Bold/underline mark best/second-best values, excluding OVOW+DivAlign; ties share marks. † denotes an in-house reproduction.

		Task 1			Task 2					Task 3					Task 4		
Benchmark	Model	Cur	U-R	H	Prev	Cur	K	U-R	H	Prev	Cur	K	U-R	H	Prev	Cur	K
Source: clean COCO																	
COCO-C COCO-O	ASGS† [ICCV'25]	67.4	19.1	29.8	48.1	45.0	46.5	13.0	20.3	20.0	42.7	27.6	9.3	14.0	30.8	37.8	32.6
	PROB [CVPR'23]	70.2	14.7	24.3	64.1	32.0	47.2	25.6	33.2	46.0	28.2	40.1	21.9	28.3	39.6	26.8	36.4
	OrthogonalDet [CVPR'24]	70.7	6.4	11.8	62.8	46.9	54.5	32.0	40.3	54.1	45.6	51.3	40.2	45.1	49.5	42.5	47.8
	O1O [ACCV'24]	69.1	17.2	27.6	58.5	44.3	51.0	16.1	24.5	40.8	40.2	40.6	21.5	28.1	41.2	40.1	40.9
	OWOBJ [CVPR'25]	72.9	13.5	22.8	65.5	38.0	51.0	19.1	27.8	48.8	36.1	44.6	13.0	20.2	43.6	35.5	41.6
	DEUS† [CVPR'26]	72.8	68.2	70.4	62.4	46.0	53.8	60.0	56.7	52.7	45.6	50.4	64.1	56.4	49.3	42.4	47.5
	OW-OVD [CVPR'25]	**76.3**	29.1	42.1	64.3	**58.1**	61.0	39.7	48.1	60.6	**53.2**	58.1	33.8	42.7	46.8	52.4	48.2
	OVOW [BMVC'25]	72.4	69.6	71.0	72.9	49.3	60.5	75.1	67.0	59.4	50.2	56.3	76.7	64.9	56.1	41.7	52.5
	OVOW+DivAlign [CVPR'24]	71.5	69.4	70.5	72.2	48.8	59.9	74.9	66.6	58.7	48.9	55.5	76.1	64.2	55.3	41.1	51.8
	RADAR (Ours)	75.2	**78.0**	**76.6**	**74.9**	57.7	**65.9**	**78.5**	**71.6**	**65.7**	52.0	**61.1**	**81.2**	**69.7**	**60.7**	**53.8**	**59.0**
ID reference: COCO-DC original (clean subset)																	
COCO-DC	ASGS† [ICCV'25]	85.6	47.0	60.7	61.9	58.3	60.0	44.9	51.4	24.0	58.7	35.5	55.0	43.2	41.8	53.7	44.7
	PROB [CVPR'23]	88.0	30.4	45.2	80.8	42.4	60.6	50.3	55.0	59.9	42.1	54.0	53.7	53.8	55.6	38.6	51.3
	OrthogonalDet [CVPR'24]	85.0	14.4	24.7	77.4	59.6	68.1	71.0	69.5	66.5	59.3	64.2	74.0	68.7	63.2	63.1	63.2
	O1O [ACCV'24]	87.7	37.8	52.8	79.1	59.8	68.9	38.1	49.1	54.1	54.8	54.4	66.6	59.9	57.1	56.7	57.0
	OWOBJ [CVPR'25]	**90.4**	32.6	47.9	83.9	50.7	66.5	43.4	52.5	64.0	55.2	61.1	42.1	49.9	61.3	53.3	59.3
	DEUS† [CVPR'26]	87.0	90.1	88.5	75.6	58.3	66.5	86.4	75.2	68.2	61.6	66.1	93.9	77.6	65.8	60.0	64.3
	OW-OVD [CVPR'25]	89.7	64.3	74.9	79.8	**69.0**	74.1	68.1	71.0	73.5	**77.2**	74.7	71.4	73.0	61.3	72.8	64.1
	OVOW [BMVC'25]	86.3	87.2	86.8	87.6	53.5	69.7	89.4	78.3	69.1	67.4	68.6	94.2	79.4	67.7	53.1	64.1
	OVOW+DivAlign [CVPR'24]	85.6	87.3	86.4	85.6	55.0	69.6	89.6	78.3	68.8	67.5	68.3	95.2	79.6	68.5	53.0	64.7
	RADAR (Ours)	89.5	**93.5**	**91.5**	**89.4**	67.9	**78.1**	**92.8**	**84.8**	**78.6**	76.4	**77.9**	**98.7**	**87.1**	**77.9**	**75.3**	**77.3**

Table 8: Per-domain RADAR results on COCO-C at severity 3. Clean COCO is the paired source reference, and the target average is unweighted over all 15 corruption domains.

		Task 1			Task 2					Task 3					Task 4		
Benchmark	Domain	Cur	U-R	H	Prev	Cur	K	U-R	H	Prev	Cur	K	U-R	H	Prev	Cur	K
COCO-C (severity 3)	Source (clean)	75.2	78.0	76.6	74.9	57.7	65.9	78.5	71.6	65.7	52.0	61.1	81.2	69.7	60.7	53.8	59.0
	Target avg.	**53.7**	**67.1**	**59.2**	**53.1**	**39.8**	**46.1**	**67.1**	**54.4**	**45.4**	**34.1**	**41.6**	**68.0**	**51.4**	**40.8**	**34.0**	**39.1**
	brightness	73.4	76.9	75.1	72.5	55.7	63.7	77.4	69.9	63.7	50.1	59.2	80.2	68.1	58.8	50.8	56.8
	contrast	69.5	74.1	71.7	69.0	53.9	61.1	75.1	67.4	60.8	46.2	56.0	77.9	65.1	55.0	48.0	53.2
	defocus blur	54.1	61.7	57.7	54.2	40.7	47.1	63.3	54.0	46.7	36.2	43.2	63.9	51.5	42.0	35.9	40.5
	elastic transform	61.7	72.4	66.6	61.4	32.5	46.2	71.2	56.1	45.9	38.6	43.5	73.7	54.7	43.2	32.7	40.6
	fog	71.6	75.5	73.5	71.2	55.1	62.8	76.7	69.0	62.5	49.6	58.2	79.5	67.2	57.3	50.2	55.5
	frost	61.4	70.9	65.8	60.6	41.4	50.5	71.7	59.3	49.8	38.8	46.1	73.3	56.6	45.5	37.3	43.5
	gaussian noise	50.4	65.5	57.0	50.2	37.2	43.4	66.6	52.5	42.7	31.4	39.0	66.0	49.0	39.1	31.6	37.2
	glass blur	35.3	57.5	43.8	35.6	23.7	29.4	56.1	38.5	27.5	23.5	26.2	55.1	35.5	25.4	15.6	22.9
	impulse noise	52.8	66.4	58.8	52.2	36.9	44.2	67.5	53.4	43.1	31.3	39.2	68.0	49.7	39.4	31.3	37.3
	jpeg compression	56.6	69.0	62.2	55.1	43.8	49.2	69.0	57.4	47.9	35.6	43.8	69.5	53.7	41.5	36.1	40.2
	motion blur	43.4	60.1	50.4	44.3	35.9	39.9	58.8	47.5	39.5	30.0	36.3	58.3	44.7	35.2	30.9	34.2
	pixelate	33.0	68.3	44.5	29.8	36.9	33.5	68.1	44.9	31.6	18.6	27.3	67.7	38.9	24.6	25.1	24.7
	shot noise	53.6	67.2	59.6	53.3	38.9	45.7	66.9	54.3	45.1	31.3	40.5	67.6	50.6	40.7	32.8	38.8
	snow	59.5	71.6	65.0	58.1	38.7	47.9	71.2	57.3	47.4	36.2	43.7	72.8	54.6	42.5	33.7	40.3
	zoom blur	28.9	49.2	36.4	28.6	25.5	27.0	47.3	34.3	26.6	14.0	22.4	46.9	30.3	21.7	18.0	20.8

set. This subset has fewer instances per image than the full clean test set's (2.0 versus 7.4 instances/image) and a larger proportion of large boxes (76.6% versus 33.9%). These differences help explain its higher absolute detection scores.

B.4 [OURS] PER-DOMAIN RESULTS

Tables 8 and 9 expand the averaged results in the main paper into individual target domains. The tables expose domain variation without domain-specific calibration.

Table 9: Per-domain RADAR results on COCO-O and COCO-DC. Clean COCO is the distribution-level source reference for COCO-O. COCO-DC uses its paired 1,127-image original subset, which is excluded from the four-target mean.

		Task 1			Task 2					Task 3					Task 4		
Benchmark	Domain	Cur	U-R	H	Prev	Cur	K	U-R	H	Prev	Cur	K	U-R	H	Prev	Cur	K
	Source (clean)	75.2	78.0	76.6	74.9	57.7	65.9	78.5	71.6	65.7	52.0	61.1	81.2	69.7	60.7	53.8	59.0
	Target avg.	**44.0**	**92.2**	**58.1**	**42.5**	**38.4**	**41.0**	**90.8**	**54.9**	**40.2**	**27.1**	**36.5**	**92.6**	**51.0**	**35.7**	**30.7**	**34.7**
COCO-O	cartoon	34.2	92.3	49.9	31.8	45.1	38.8	91.8	54.5	37.6	21.0	32.1	91.7	47.5	30.8	29.7	30.5
	handmake	32.3	97.0	48.5	29.3	38.3	34.0	99.0	50.6	33.6	27.0	31.4	99.0	47.7	30.7	32.4	31.1
	painting	62.1	91.0	73.8	60.6	54.1	57.2	92.0	70.5	55.9	49.3	53.7	94.8	68.5	52.3	37.2	48.5
	sketch	36.6	85.4	51.2	36.2	35.1	35.7	85.8	50.4	34.8	14.3	28.2	86.2	42.5	27.4	22.2	26.1
	tattoo	31.9	95.3	47.8	30.4	14.4	22.4	96.4	36.3	21.9	15.8	19.8	96.3	32.8	19.6	16.9	18.9
	weather	66.8	92.2	77.5	66.8	43.5	57.8	80.0	67.1	57.2	35.3	54.1	87.5	66.9	53.4	45.9	53.0
	ID ref. (original)	89.5	93.5	91.5	89.4	67.9	78.1	92.8	84.8	78.6	76.4	77.9	98.7	87.1	77.9	75.3	77.3
	Target avg.	**73.1**	**91.5**	**81.2**	**73.4**	**55.5**	**64.0**	**90.2**	**74.9**	**64.4**	**58.4**	**62.4**	**96.6**	**75.8**	**61.5**	**61.7**	**61.5**
COCO-DC	adversarial	65.3	90.9	76.0	66.2	55.6	60.7	90.2	72.5	60.9	52.9	58.3	96.5	72.7	56.4	55.9	56.3
	blip	78.7	92.9	85.2	78.9	57.9	67.9	91.5	77.9	68.0	66.2	67.4	96.8	79.5	67.1	67.5	67.2
	color	75.8	91.1	82.8	75.6	56.0	65.4	89.8	75.7	66.0	57.7	63.4	96.8	76.6	62.3	62.9	62.4
	texture	72.6	91.2	80.9	72.9	52.5	62.2	89.4	73.4	62.6	56.8	60.7	96.5	74.5	60.1	60.6	60.2

Table 10: Task 3 feature drift and correct-class logit drop when toggling the accumulated CLA residual. Target values are domain means; bold marks lower drift or logit drop.

		Feature drift (cosine angle, °) ↓						Logit drop ↓
Benchmark	Setting	Stem	Stage 1	Stage 2	Stage 3	Stage 4	Embed	
COCO-C	w/o CLA	**5.9**	19.9	17.9	25.2	31.1	11.8	1.17
COCO-C	w/ CLA	6.8	**18.4**	**15.9**	**21.5**	**24.1**	**11.1**	**0.92**
COCO-O	w/o CLA	**7.7**	18.9	17.9	28.1	41.5	26.4	3.50
COCO-O	w/ CLA	10.2	**17.1**	**16.1**	**25.0**	**36.8**	**26.2**	**3.29**
COCO-DC	w/o CLA	**3.7**	12.9	11.2	19.8	24.7	16.8	0.45
COCO-DC	w/ CLA	3.8	**11.1**	**9.2**	**15.5**	**18.4**	**16.0**	**0.30**

C MOTIVATION FOR RADAR

This section expands the reference-model analyses in Figure 2. We first examine how domain shifts affect backbone features and known-class responses, and then investigate how background statistics influence attribute selection. These observations motivate CLA and BDAS, respectively; Appendix E.6 reports component ablations, with +BDAS+CLA matching the benchmark model in Table 2.

C.1 [CLA] EARLY FEATURE DRIFT

Table 10 summarizes the Task 3 source-to-target drift shown in Figure 2(a). Drift is the cosine angle between normalized source and target class prototypes in known-class ground-truth regions, averaged over the 15 COCO-C, 6 COCO-O, and 4 COCO-DC domains. The comparison disables or enables the accumulated CLA residual while keeping the remaining detector components fixed and using native features for scoring. It measures the residual's effect within one detector, rather than comparing independently trained models. Lines show benchmark means, shaded regions indicate ± 1 domain standard deviation, and COCO-C uses severity 3.

Without CLA, drift was present at the stem and generally increased through the backbone, peaking at Stage 4 in all three benchmarks. The embedding reduced the angle after Stage 4, but its output retained a domain discrepancy. With CLA, Stage-4 angles decreased by $7.0°$, $4.7°$, and $6.3°$, while Embed angles decreased by $0.7°$, $0.2°$, and $0.8°$ on COCO-C, COCO-O, and COCO-DC, respec-

tively. Correct-class logit drop, measured as the source-minus-target GT-class logit and averaged over domains, decreased by 0.25, 0.21, and 0.15. The stem angle did not decrease, so these observations support improved downstream feature and prediction stability rather than uniform alignment at every layer.

C.2 [BDAS] Background-Dominated Selection

Our base model (OW-OVD) collects anchors outside the labeled foreground pool D^+ in a single negative pool D^-. On a Task 1 source probe, future-class GT-best anchors formed only 0.042% of the estimated negative population. Consistently, **Figure 2(b)** gives nearly identical JSD values for the mixed-negative and background-only pools (0.055), compared with 0.018 for future objects. In Figure 2(b), D^+ denotes known-object anchors and D^- denotes the negative pool in each panel; JSD compares their attribute-response distributions. The similarity between the mixed-negative and background-only distributions indicated that background statistics dominated this selection signal. BDAS therefore separates object-like negative anchors from background. Using a frozen source detector, it retains at most three post-NMS proposals per image after discarding boxes that occupy more than 20% of the image or overlap confident previous-known detections. Eligible negative anchors inside these boxes form $\mathcal{O}$; the remainder form $\mathcal{B}$. Contrasting their attribute-response distributions provides a way to favor attributes shared by objects while penalizing background responses.

Table 11: BDAS Task 1 diagnostic. Source-pool occupancy compares D^- and $\mathcal{O}$; detection values are the +CLA and +BDAS+CLA rows of Table 21. Bold marks the higher value; ties share marks.

Benchmark	Setting	Future occupancy (%)	Relative	K-mAP	U-R	H
Source	w/o BDAS	7.31	1.00×	**75.7**	69.2	72.3
	w/ BDAS	**29.11**	3.98×	75.2	**78.0**	**76.6**
COCO-C	w/o BDAS	–	–	**53.9**	57.3	55.2
	w/ BDAS	–	–	53.7	**67.1**	**59.2**
COCO-O	w/o BDAS	–	–	**44.2**	82.5	56.0
	w/ BDAS	–	–	44.0	**92.2**	**58.1**
COCO-DC	w/o BDAS	–	–	**73.1**	86.4	79.1
	w/ BDAS	–	–	**73.1**	**91.5**	**81.2**

The source-pool diagnostic in Table 11 shows that future-class anchor occupancy is $3.98\times$ higher in $\mathcal{O}$ than in D^-. Occupancy counts eligible anchor centers inside future-class GT boxes, whereas the 0.042% above counts only GT-best anchors. Future-class annotations were used only for these diagnostics. The matched detection rows show higher unknown recall on all three target benchmarks with BDAS.

Table 21 reports the full Task 1 component comparison, including known-class performance.

D Overall RADAR Pipeline

Algorithms 1 and 2 select the attribute bank and adapt the backbone using source data only. Algorithm 3 uses the resulting frozen detector and the same attribute bank.

Taskwise construction. At Task t, BDAS first extends $\mathcal{A}^{t-1}$ with attributes for the current classes. At Task 1, $\mathcal{A}^0$ is empty and mining uses the pretrained detector with CLIP-encoded candidates. Subsequent tasks use the previous detector and selected bank for mining, while activation histograms are always collected with frozen F_0. The redundancy term is zero when the working bank is empty. CLA then learns the LoRA residuals and current-class embeddings for one source epoch using the

Algorithm 1 BDAS Attribute Selection at Task t

Require: F_0, F_{t-1}, $\mathcal{D}_{\mathcal{S}}^t$, $\mathcal{A}^{t-1}$ ($\mathcal{A}^0 = \varnothing$), and $\mathcal{V}^t$ (combined candidates for $\mathcal{C}^t$)
Fixed settings: $(\beta, \eta) = (0.5, 1)$
Ensure: frozen attribute bank $\mathcal{A}^t$

if $t = 1$ **then**
  mine with pretrained F_0 and all CLIP-encoded attribute candidates
else
  mine with frozen F_{t-1} and $\mathcal{A}^{t-1}$
end if
$\mathcal{X}_t \leftarrow$ post-NMS unknown-label boxes with area ratio ≤ 0.2
if $t > 1$ **then**
  remove boxes overlapping known predictions (score ≥ 0.3, IoU ≥ 0.3)
end if
retain the top three boxes per image
extract attribute responses using frozen F_0
$\mathcal{P} \leftarrow$ anchors with current-task assignment score ≥ 0.55
$(\mathcal{O}, \mathcal{B}) \leftarrow \mathrm{SplitNeg}(\mathcal{X}_t)$ by center inclusion, excluding current-task GT interiors
$\widehat{\mathcal{A}}^t \leftarrow \mathcal{A}^{t-1}$
for $\mathbf{a} \in \mathcal{V}^t$ **do**
  form $(\mathcal{D}_{\mathbf{a}}^+, \mathcal{D}_{\mathbf{a}}^{\mathcal{O}}, \mathcal{D}_{\mathbf{a}}^{\mathcal{B}})$ from $(\mathcal{P}, \mathcal{O}, \mathcal{B})$
  $s_{\mathrm{dist}}(\mathbf{a}) \leftarrow \mathrm{JSD}(\mathcal{D}_{\mathbf{a}}^+, \mathcal{D}_{\mathbf{a}}^{\mathcal{O}}) - \eta\,\mathrm{JSD}(\mathcal{D}_{\mathbf{a}}^+, \mathcal{D}_{\mathbf{a}}^{\mathcal{B}})$
end for
repeat
  $s(\mathbf{a}) \leftarrow \beta s_{\mathrm{dist}}(\mathbf{a}) + (1-\beta)\overline{\cos}(\mathbf{a}, \widehat{\mathcal{A}}^t)$ for $\mathbf{a} \in \mathcal{V}^t \setminus \widehat{\mathcal{A}}^t$
  $\mathbf{a}^\star \leftarrow \arg\min_{\mathbf{a}} s(\mathbf{a})$; $\widehat{\mathcal{A}}^t \leftarrow \widehat{\mathcal{A}}^t \cup \{\mathbf{a}^\star\}$
until $25|\mathcal{C}^t|$ current-task attributes have been added
$\mathcal{A}^t \leftarrow \widehat{\mathcal{A}}^t$
freeze and return $\mathcal{A}^t$

Algorithm 2 CLA Backbone Adaptation at Task t

Require: F_{t-1}, $\mathcal{D}_{\mathcal{S}}^t$, frozen $\mathcal{A}^t$, $\mathcal{J}_{\mathrm{L}}$, and $\{\mathbf{c}_k\}_{k \in \mathcal{K}^t}$
Fixed settings: $(r_{\mathrm{L}}, \alpha_{\mathrm{L}}) = (8, 8)$, $(\epsilon, \nu) = (5, 0.3)$, $N_\Omega = 100$, $\tau_1 = 1$, and $\tau_{t>1} = 0.1$
Ensure: frozen F_t and updated class embeddings

insert LoRA factors at $j \in \mathcal{J}_{\mathrm{L}}$
for each mini-batch $\mathcal{M} \subset \mathcal{D}_{\mathcal{S}}^t$ during one source epoch **do**
  $(\mathcal{M}^{\mathrm{orig}}, \mathcal{M}^{\mathrm{pert}}) \leftarrow \mathrm{StylePair}(\mathcal{M}; \epsilon = 5, \nu = 0.3)$
  $\Omega \leftarrow \mathrm{TopAnchors}(\mathcal{M}^{\mathrm{orig}}; N_\Omega = 100)$ at each pyramid level
  compute $\mathcal{L}_{\mathrm{align}}$ on known-class, box, and frozen-bank $\mathcal{A}^t$ predictions over Ω, stopping gradients through original-view consistency targets
  $\mathcal{L}_{\mathrm{CLA}} \leftarrow 0.05\mathcal{L}_{\mathrm{det}}^{\mathrm{orig}} + 0.0125\mathcal{L}_{\mathrm{det}}^{\mathrm{pert}} + 0.1\mathcal{L}_{\mathrm{align}}$
  update only the LoRA factors and current-task embeddings
end for
$F_t \leftarrow \mathrm{Merge}(F_{t-1}, \{\tau_t \Delta W_t^{(j)}\}_{j \in \mathcal{J}_{\mathrm{L}}})$
freeze and return F_t and the updated class embeddings

fixed bank $\mathcal{A}^t$ and merges the residuals into F_t. Earlier class embeddings remain frozen. Target images do not update the bank, backbone, or class embeddings.

E Hyperparameters and Ablations

The component comparisons use the same +BDAS+CLA configuration as Table 2, toggling the bank and accumulated CLA adaptation with their constituent checkpoints fixed. The BDAS coefficient and score-scale tables include source-only development controls measured with the maximum of native and transformed-feature attribute responses. Their shaded configured entries use the current

Algorithm 3 RADAR Inference at Task t

Require: test image I, F_t, $\mathcal{A}^t$, known classes $\mathcal{K}^t$, and frozen attribute scale and bias (s, b)
Fixed setting: $\xi_{\mathrm{u}} = 15$
Ensure: post-processed known and unknown detections

Step 1: Native predictions
$\{\mathbf{e}_i, p_{i,k}\}_{i,k} \leftarrow F_t(I), k \in \mathcal{K}^t$
$\mathbf{g}_t \leftarrow |\mathcal{A}^t|^{-1} \sum_{\mathbf{a} \in \mathcal{A}^t} \mathbf{a}$; $\bar{\mathbf{g}}_t \leftarrow \mathbf{g}_t / \|\mathbf{g}_t\|_2$
Step 2: Unknown scoring
for each anchor i at pyramid level $\ell(i)$ **do**
  compute $S_i^{\mathrm{att}}(\mathbf{e}_i)$ from $\mathcal{A}^t$ using Eq. 4
  $S_i^{\mathrm{ctr}} \leftarrow \sigma(s \bar{\mathbf{g}}_t^{\top} \mathbf{e}_i + b)$; $p_i^{\max} \leftarrow \max_{k \in \mathcal{K}^t} p_{i,k}$
  $S_i^{\mathrm{unk}} \leftarrow \xi_{\mathrm{u}}(1 - p_i^{\max})(S_i^{\mathrm{att}}(\mathbf{e}_i) + S_i^{\mathrm{ctr}})$
end for
Step 3: Detection post-processing (without score clipping)
retain candidates above 0.001; apply class-aware NMS at IoU 0.7; keep the top 300 detections
return the retained known and unknown detections

native score of Eq. 11 and are not directly comparable to the other cells. The source-only CLA controls use H as the primary comparison criterion, subject to a K-mAP decrease of at most 0.5 point; merge coefficients are compared by previous-class AP (Prev AP). No target-domain result determines a configured value.

E.1 BDAS Coefficients (β, η)

We use $(\beta, \eta) = (0.5, 1)$ in Eq. 9. The value $\beta = 0.5$ assigns equal coefficients to the distributional and redundancy terms, while $\eta = 1$ assigns the object-like and background divergence terms equal-magnitude coefficients with opposite signs. The bank retains OW-OVD's budget of 25 attributes per newly introduced class. These values remain fixed across tasks and target domains. Table 12 shows a source-only 3×3 development study. Each pair reselected the Task 1 attribute bank and centroid with the detector fixed. The shaded configured pair reports native-score H; other cells use the earlier development score.

Table 12: Task 1 source H across BDAS coefficients. The shaded configured cell uses the native score; other cells use the earlier development score.

	η		
β	0	1	2
0.2	73.0	76.6	74.9
0.5	72.6	76.6	74.3
0.8	71.6	76.1	74.1

E.2 CLA Scope $\mathcal{J}_{\mathrm{L}}$

Table 13 reports the source-only CLA scope comparison with the bank and native inference score held fixed. These source controls evaluate CLA placement in a fixed development setup; all three scopes satisfy the K-mAP guardrail. The stem with Stages 1–2 gave the highest K-mAP, U-AP, and H, while adaptation through Stage 1 gave marginally higher U-R. We therefore select the stem and Stages 1–2.

Table 13: Task 1 source sensitivity of CLA scope. Bold marks the unrounded maximum within each column; the shaded row is selected.

CLA location	K-mAP	U-R	U-AP	H
Stem	74.9	77.9	19.9	76.4
Stem + Stage 1	75.0	**78.0**	20.0	76.5
Stem + Stages 1–2	**75.2**	78.0	**20.1**	**76.6**

A controlled two-seed check of the top two scopes retained the same selection, although the H margin remained small.

CLA scalar sensitivity. These Task 1 development controls retain the original OW-OVD bank during CLA training and use the BDAS bank at evaluation. Their within-setup comparisons examine local sensitivity, separately from the bank-consistent training pipeline used for the benchmark results. **Tables 14 and 15** vary ϵ, ν, N_Ω, and λ_{align} one at a time around their configured values. K-mAP remains 75.0–75.4, H remains 76.4–76.8, U-R remains 77.4–78.6, and U-AP remains 19.3–21.5. Additional source checks use $r_{\text{L}} = \alpha_{\text{L}} \in \{4, 8, 16\}$, $\lambda_{\text{det}} \in \{0.025, 0.05, 0.1\}$, and $\lambda_{\text{aug}} \in \{0.00625, 0.0125, 0.025\}$; their H ranges are 76.5–76.6, 76.5–76.7, and 76.3–76.5, respectively. These results place the configured values in locally stable source regions.

Table 14: Task 1 source sensitivity of CLA style parameters. Bold marks the unrounded maximum within each parameter block; shaded rows are configured.

Parameter	Value	K-mAP	U-R	U-AP	H
ϵ	2.5	**75.1**	**78.4**	20.1	**76.7**
ϵ	5	75.0	78.1	20.0	76.5
ϵ	10	75.1	78.0	**20.3**	76.5
ν	0.1	**75.2**	77.8	19.3	76.5
ν	0.3	75.0	78.1	20.0	76.5
ν	1.0	75.1	**78.3**	**20.3**	**76.6**

Table 15: Task 1 source sensitivity of CLA consistency parameters. Bold marks the unrounded maximum within each parameter block; shaded rows are configured.

Parameter	Value	K-mAP	U-R	U-AP	H
N_Ω	50	75.1	**78.4**	**20.1**	**76.7**
N_Ω	100	75.0	78.1	20.0	76.5
N_Ω	200	**75.4**	77.4	19.9	76.4
λ_{align}	0.05	75.0	78.3	19.8	76.6
λ_{align}	0.1	75.0	78.1	20.0	76.5
λ_{align}	0.2	**75.1**	**78.6**	**21.5**	**76.8**

E.3 CLA MERGE τ_t

Holding the Task 2 embeddings and learned residual fixed, **Table 16** varies only the coefficient used to merge that residual into the accumulated backbone update. The value $\tau_2 = 0.1$ gave the highest source Prev AP and K-mAP. Increasing the coefficient to 1.0 substantially reduced previous-class retention. Accordingly, we use $\tau_1 = 1$ for the initial update and $\tau_{t>1} = 0.1$ for subsequent tasks.

Table 16: Task 2 source sensitivity of τ_2; bold marks the selected value and unrounded metric maxima.

τ_2	Prev AP	Cur AP	K-mAP
0.1	**75.5**	58.5	**66.6**
0.25	74.1	**58.5**	65.9
0.5	69.8	57.8	63.5
1.0	43.3	53.8	48.8

E.4 Unknown-Score Scale ξ_{u}

The scale $\xi_{\mathrm{u}} = 15$ sets the unknown branch's score relative to the known branch and remains fixed across tasks and domains. The score is not a calibrated probability and is not clipped to $[0, 1]$. Although a positive scale preserves ordering within the unknown branch before filtering, it changes threshold crossings and competition for the top 300 detections. The native-score comparison in Table 18 therefore reports both known mAP and unknown AP alongside recall; increased recall alone does not establish improved ranking or calibration. Table 17 shows source-only scale controls. Its shaded $\xi_{\mathrm{u}} = 15$ row reports the current native-score result, with K-mAP 75.2, U-R 78.0, and H 76.6. The other scale rows use the earlier development score and are not a controlled sweep of the final native score.

Table 17: Task 1 source scale controls. The shaded configured row uses the native score; other rows use the earlier development score.

ξ_{u}	K-mAP	U-R	U-AP	U-P	H
1	75.7	49.1	18.5	7.7	59.6
3	75.6	64.4	19.6	4.1	69.6
5	75.6	70.1	19.9	3.1	72.8
10	75.3	76.0	19.9	2.1	75.6
15	75.2	78.0	20.1	1.9	76.6
30	75.1	80.3	20.0	1.4	77.6

E.5 Unknown-Score Decomposition

Let $G_i = 1 - p_i^{\max}$, where $p_i^{\max}$ is the largest known-class confidence. Table 18 traces the unknown score on the frozen Task 1 detector reported in Table 2. It starts with the native attribute response, then adds the centroid response, knownness gate, and configured scale. The Gate control isolates the centroid contribution under G_i. Only the inference formula changes in this comparison.

Table 18: Task 1 native-score decomposition on the fixed RADAR detector. Target entries are equal-domain means; the shaded rows use the configured formula.

Benchmark	Step	Unknown score	K-mAP	U-R	U-AP	H
Source	Attribute	S_i^{att}	75.7	48.1	15.8	58.8
	+ Centroid	$S_i^{\text{att}} + S_i^{\text{ctr}}$	75.7	48.3	15.9	58.9
	Gate control	$G_i S_i^{\text{att}}$	75.7	48.1	18.6	58.8
	+ Gate	$G_i(S_i^{\text{att}} + S_i^{\text{ctr}})$	75.7	48.2	18.6	58.9
	+ Scale	$15G_i(S_i^{\text{att}} + S_i^{\text{ctr}})$	75.2	78.0	20.1	76.6
COCO-C	Attribute	S_i^{att}	53.9	38.3	14.2	44.6
	+ Centroid	$S_i^{\text{att}} + S_i^{\text{ctr}}$	53.9	38.4	14.2	44.7
	Gate control	$G_i S_i^{\text{att}}$	53.9	38.2	15.6	44.5
	+ Gate	$G_i(S_i^{\text{att}} + S_i^{\text{ctr}})$	53.9	38.4	15.6	44.6
	+ Scale	$15G_i(S_i^{\text{att}} + S_i^{\text{ctr}})$	53.7	67.1	16.7	59.2
COCO-O	Attribute	S_i^{att}	44.2	64.8	18.8	51.1
	+ Centroid	$S_i^{\text{att}} + S_i^{\text{ctr}}$	44.2	65.0	18.9	51.2
	Gate control	$G_i S_i^{\text{att}}$	44.2	64.7	20.9	51.1
	+ Gate	$G_i(S_i^{\text{att}} + S_i^{\text{ctr}})$	44.2	65.0	21.0	51.2
	+ Scale	$15G_i(S_i^{\text{att}} + S_i^{\text{ctr}})$	44.0	92.2	22.4	58.1
COCO-DC	Attribute	S_i^{att}	73.1	74.5	23.3	73.7
	+ Centroid	$S_i^{\text{att}} + S_i^{\text{ctr}}$	73.1	74.7	23.6	73.8
	Gate control	$G_i S_i^{\text{att}}$	73.1	74.5	26.5	73.7
	+ Gate	$G_i(S_i^{\text{att}} + S_i^{\text{ctr}})$	73.1	74.7	26.6	73.8
	+ Scale	$15G_i(S_i^{\text{att}} + S_i^{\text{ctr}})$	73.1	91.5	27.0	81.2

Adding S_i^{ctr} increased U-R and H in 24 of the 25 target domains, with equal-domain mean changes of 0.17 and 0.09 point, respectively, and negligible K-mAP change. Under the gate, adding the centroid increased U-R by 0.14–0.22 point across the three target benchmarks. The gate primarily improved U-AP, while the scale produced a much larger change in recall. These comparisons distinguish the small centroid contribution from the operating-point effect of score scaling.

Base† applies $15G_i(S_i^{\text{att}} + S_i^{\text{ctr}})$ to the frozen OW-OVD backbone, class embeddings, and original attribute bank. It provides a common-score reference for the module comparisons below, while the OW-OVD row retains its native unknown score.

E.6 Component Ablations

Component comparisons. Tables 3 and 19 report source and target component controls. The latter extends Table 3 to COCO-O and COCO-DC. These frozen-component ablations switch the selected bank and accumulated CLA adaptation under the common score. The combinations share their constituent checkpoints; +BDAS+CLA is the Full RADAR model reported in Table 2.

Table 19: **Component ablations on COCO-O and COCO-DC, continuing Table 3.** Bold/underline denote best/second-best unrounded values, excluding OW-OVD.

	Target (COCO-O)										**Target** (COCO-DC)									
	Task 1			Task 2			Task 3			Task 4	Task 1			Task 2			Task 3			Task 4
Setting	Cur	U-R	H	K	U-R	H	K	U-R	H	K	Cur	U-R	H	K	U-R	H	K	U-R	H	K
OW-OVD	41.2	42.1	38.9	33.3	54.1	38.7	30.7	42.8	30.6	24.9	73.6	54.4	62.6	58.7	61.6	60.1	58.7	67.8	62.9	49.4
Base†	41.1	85.1	53.7	33.0	89.8	46.9	30.3	85.3	42.5	24.4	73.6	88.5	80.3	58.4	**93.0**	71.7	58.5	**97.8**	73.1	49.2
+CLA	**44.2**	82.5	56.0	**41.2**	87.8	54.3	**36.8**	89.1	50.5	**34.9**	73.1	86.4	79.1	64.0	90.0	74.8	**62.5**	96.2	75.7	**61.6**
+BDAS	41.0	91.9	55.2	32.8	**91.3**	47.0	30.1	90.1	43.6	24.1	**73.7**	**92.2**	**81.9**	58.3	91.3	71.1	58.2	97.5	72.8	48.8
+BDAS+CLA	44.0	**92.2**	**58.1**	41.0	90.8	**54.9**	36.5	**92.6**	**51.0**	34.7	73.1	91.5	81.2	**64.0**	90.2	**74.9**	62.4	96.6	**75.8**	61.5

Average module effects. Table 20 averages two matched on/off differences for each module. For BDAS, these are Base† to +BDAS and +CLA to +BDAS+CLA; for CLA, they are Base† to +CLA

and +BDAS to +BDAS+CLA. Thus, each effect averages over the two states of the other module rather than treating the rows as a cumulative sequence.

Table 20: Average effects over the two matched pairs per module. C/O/DC denote target-domain means; Task 4 reports only ΔK.

		Task 1			Task 2			Task 3			Task 4
Target	Factor	ΔK	ΔU-R	ΔH	ΔK	ΔU-R	ΔH	ΔK	ΔU-R	ΔH	ΔK
C	BDAS	−0.3	+8.2	+3.2	−0.3	+2.2	+0.5	−0.4	+4.0	+0.8	−0.5
C	CLA	+2.1	−1.0	+1.0	+7.3	−1.5	+5.0	+5.5	−0.9	+4.2	+10.0
O	BDAS	−0.2	+8.3	+1.8	−0.2	+2.3	+0.4	−0.2	+4.1	+0.8	−0.3
O	CLA	+3.0	−1.2	+2.6	+8.2	−1.3	+7.7	+6.4	+3.2	+7.7	+10.6
DC	BDAS	+0.0	+4.5	+1.8	−0.0	−0.8	−0.2	−0.1	+0.0	−0.1	−0.2
DC	CLA	−0.6	−1.4	−0.9	+5.7	−2.0	+3.4	+4.1	−1.2	+2.8	+12.5

At Task 1, BDAS increased target U-R by 4.5–8.3 points on average over the two matched pairs. CLA provided larger later-task K-mAP gains: 5.7–8.2 at Task 2, 4.1–6.4 at Task 3, and 10.0–12.5 at Task 4. The effects were not uniform across benchmarks or metrics. In particular, BDAS did not improve every later-task recall value, and its recall gains did not consistently translate into higher U-AP.

Detailed component ablations. Tables 21–24 report the four combinations of BDAS and CLA. Rows prefixed by + are independent combinations, and target values are domain means. U-AP is included to distinguish unknown ranking quality from recall. The +BDAS+CLA rows match the Full RADAR results in Tables 2 and 7.

Table 21: Task 1 component ablations; Cur is the known-class metric. OW-OVD is a reference and is excluded from the marking; bold and underline denote the best and second-best unrounded values, respectively.

	Source				COCO-C				COCO-O				COCO-DC			
Setting	Cur	U-R	U-AP	H	Cur	U-R	U-AP	H	Cur	U-R	U-AP	H	Cur	U-R	U-AP	H
OW-OVD	76.3	29.1	13.3	42.1	51.9	23.8	10.0	32.3	41.2	42.1	15.5	38.9	73.6	54.4	19.6	62.6
Base†	**76.3**	72.7	16.3	74.4	51.8	60.0	13.3	55.0	41.1	85.1	19.6	53.7	73.6	88.5	20.9	80.3
+CLA	75.7	69.2	17.1	72.3	**53.9**	57.3	15.0	55.2	**44.2**	82.5	19.7	56.0	73.1	86.4	23.9	79.1
+BDAS	76.0	77.8	18.1	**76.9**	51.5	66.5	15.3	57.4	41.0	91.9	20.4	55.2	**73.7**	**92.2**	23.7	**81.9**
+BDAS+CLA	75.2	**78.0**	**20.1**	76.6	53.7	**67.1**	**16.7**	**59.2**	44.0	**92.2**	**22.4**	**58.1**	73.1	91.5	**27.0**	81.2

Table 22: Task 2 component ablations; marking follows **Table 21**.

	Source				COCO-C				COCO-O				COCO-DC			
Setting	K	U-R	U-AP	H	K	U-R	U-AP	H	K	U-R	U-AP	H	K	U-R	U-AP	H
OW-OVD	61.0	39.7	12.7	48.1	39.8	29.6	11.7	33.8	33.3	54.1	12.0	38.7	58.7	61.6	18.3	60.1
Base†	60.4	79.2	14.1	68.5	39.1	66.4	12.7	48.9	33.0	89.8	12.9	46.9	58.4	**93.0**	18.7	71.7
+CLA	**66.4**	76.7	**15.9**	71.2	**46.5**	63.4	**13.2**	53.5	**41.2**	87.8	**13.5**	54.3	64.0	90.0	**20.5**	74.8
+BDAS	60.2	**79.5**	8.2	68.5	38.9	**67.2**	6.5	49.0	32.8	**91.3**	8.2	47.0	58.3	91.3	11.9	71.1
+BDAS+CLA	65.9	78.5	7.2	**71.6**	46.1	67.1	6.7	**54.4**	41.0	90.8	6.5	**54.9**	**64.0**	90.2	12.7	**74.9**

Table 23: Task 3 component ablations; marking follows **Table 21**.

	Source				COCO-C				COCO-O				COCO-DC			
Setting	K	U-R	U-AP	H	K	U-R	U-AP	H	K	U-R	U-AP	H	K	U-R	U-AP	H
OW-OVD	58.1	33.8	10.3	42.7	37.1	25.6	8.4	30.1	30.7	42.8	6.0	30.6	58.7	67.8	14.2	62.9
Base†	57.2	77.8	11.0	66.0	36.5	64.9	**9.1**	46.4	30.3	85.3	6.6	42.5	58.5	**97.8**	14.5	73.1
+CLA	**61.7**	77.0	**12.6**	68.5	**42.0**	62.7	8.4	50.1	**36.8**	89.1	**6.8**	50.5	**62.5**	96.2	**15.7**	75.7
+BDAS	56.8	81.1	3.9	66.8	36.1	67.6	3.7	46.7	30.1	90.1	4.0	43.6	58.2	97.5	6.4	72.8
+BDAS+CLA	61.1	**81.2**	5.3	**69.7**	41.6	**68.0**	5.2	**51.4**	36.5	**92.6**	3.9	**51.0**	62.4	96.6	8.0	**75.8**

Table 24: Task 4 component ablations. No unknown class remains, so U-R, U-AP, and H are undefined; marking follows **Table 21**.

	Source			COCO-C			COCO-O			COCO-DC		
Setting	Prev	Cur	K	Prev	Cur	K	Prev	Cur	K	Prev	Cur	K
OW-OVD	46.8	52.4	48.2	29.6	31.5	30.1	24.3	26.2	24.9	46.8	57.4	49.4
Base†	45.3	52.1	47.0	29.0	31.2	29.6	23.7	25.9	24.4	46.5	57.4	49.2
+CLA	**61.4**	**54.4**	**59.6**	**41.2**	**34.6**	**39.6**	**35.9**	**30.9**	**34.9**	**61.5**	61.7	**61.6**
+BDAS	45.0	51.7	46.7	28.5	30.9	29.1	23.4	25.7	24.1	46.1	57.3	48.8
+BDAS+CLA	60.7	53.8	59.0	40.8	34.0	39.1	35.7	30.7	34.7	61.5	**61.7**	61.5

F Computational Costs

Construction. BDAS extracts source activation histograms and selects the bank without gradient optimization. CLA then trains the designated LoRA factors and current-task class embeddings for one source epoch. With rank 8 over 25 convolutions in the stem and Stages 1–2, the LoRA factors contain 191,192 trainable parameters. The current-task embeddings add $512|\mathcal{C}^t|$ parameters, giving 200,920 parameters at Task 1, approximately 0.4% of the 47.3M-parameter detector. Earlier class embeddings and the selected attribute bank remain frozen.

Table 25 reports Task 1 construction costs reconstructed from execution logs, with each method following its own recipe on one H200 GPU with YOLO-World-L. Training peaks are the maximum memory values recorded by MMEngine. In both tables, † denotes the previously profiled RADAR implementation, which includes closed-form re-anchoring (CoRe).

Table 25: Task 1 construction costs on one H200 GPU. Train. is the maximum simultaneously trainable parameter count; state excludes shared detector weights.

Method	Time (h)	Train. (M)	Train peak (GiB)	State (MiB)
OW-OVD	2.77	0.010	7.00	0.99
OVOW	4.65	22.545	61.93	86.19
RADAR†	6.05	0.201	7.00	14.47

Inference. Each image requires one pass through the adapted detector. The attribute and centroid responses use the same native feature at each anchor, with the centroid computed once from the frozen task bank. In the unfused implementation, each frozen CLA update is an explicit residual branch. It can instead be folded into its convolution as $W_t^{(j)} = W_0^{(j)} + \sum_{s=1}^{t} \tau_s \Delta W_s^{(j)}$, which removes the separate branch while preserving the represented function. The folded detector still computes attribute responses, unknown scores, and detection post-processing.

Table 26 reports batch-1 latency on one H200 GPU at 640×640 resolution over 500 source images after 50 warm-up iterations. CUDA was synchronized around each test step; data loading and metric computation were excluded. The fused implementation was measured separately.

Table 26: Matched H200 inference cost at 640^2 and batch 1. Peak is CUDA maximum allocated memory after warm-up; state excludes shared detector weights.

Method	Params (M)	ms/image	Peak (GiB)	State (MiB)
OW-OVD	47.08	82.30	0.21	0.99
OVOW	69.38	107.95	0.29	86.19
RADAR† (unfused)	47.28	100.95	0.23	14.47
RADAR† (fused)	47.08	95.88	0.21	3.99