TARGET-ONLY RELATIONAL LEARNING FOR OBJECT DETECTION UNDER DOMAIN AND CATEGORY SHIFTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Domain Adaptive Object Detection (DAOD) transfers a detector from a labeled source domain to an unlabeled target domain. Deployment becomes substantially harder when target appearance changes, unknown categories occur, and source images cannot be retained. We study these conditions jointly and propose GraphGen, a source-free framework that builds sparse relation graphs over target-region features. Graph propagation regularizes target representations, uncertainty-aware pseudo-labeling separates known and unknown proposals, and graph consistency stabilizes teacher–student adaptation. On Cityscapes→Foggy Cityscapes and Pascal VOC→Clipart, GraphGen improves known-class detection and novel-object recall over the methods included in our common protocol, with gains of up to 2.22 mAP and 1.19 AR points. The results support target-side relational reasoning, while the remaining dependence on pseudo-label quality motivates broader multi-seed and cross-domain evaluation.

1 INTRODUCTION

Domain Adaptive Object Detection (DAOD) aims to transfer detection capabilities from a labeled source domain to an unlabeled target domain with different visual characteristics Chen et al. (2018); Saito et al. (2019). This task is crucial for real-world deployment of detection systems, where variations in lighting, weather conditions, and imaging sensors inevitably cause visual distribution shifts between training and operational environments. Traditional DAOD approaches typically focus on aligning feature distributions between domains through adversarial training and domain-invariant representation learning Xu et al. (2020); Chen et al. (2020).

However, existing DAOD methods face significant limitations for practical use. Most approaches assume object categories are identical across domains Wang et al. (2019); Chen et al. (2020), making them unable to handle novel categories in deployment. Conventional methods also require access to both source and target data during adaptation Li et al. (2020), which is often infeasible in industrial settings. Current techniques struggle to distinguish between domain-shifted known objects and genuinely novel categories Dhamija et al. (2020); Li et al. (2023).

To address these challenges, we distinguish three axes of the problem. Domain distribution shift changes the target image and feature distributions Wang et al. (2019); Chen et al. (2020); open-set class shift changes the target label space by introducing unknown categories Joseph et al. (2021); Li et al. (2023); and the source-free constraint changes the available supervision by removing source images during adaptation Kundu et al. (2020); Liang et al. (2020).

To address these challenges, we propose GraphGen, a unified framework that operates only on unlabeled target images and a source-trained detector during adaptation. Its three objectives share one target-instance graph but regularize different signals: representation preservation, known/unknown separation, and teacher–student consistency. This coupling is the main distinction from pipelines that apply graph alignment, novelty scoring, and self-training as independent stages. Experiments on Cityscapes→Foggy Cityscapes and Pascal VOC→Clipart compare the framework under weather and style shifts. Our key contributions include:

- A target-only relation graph that supports representation preservation, novelty separation, and teacher–student consistency under a strict source-free protocol.

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

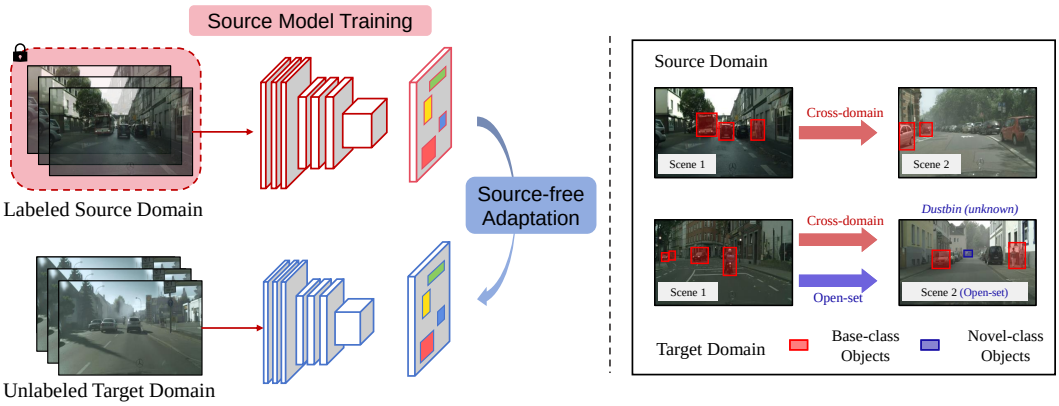

Figure 1: Problem setting for source-free domain-adaptive open-set object detection. Source images are used only to train the initial detector and are unavailable during adaptation. The right panel illustrates domain and category shifts for evaluation; it does not depict source–target matching during adaptation.

- An uncertainty-aware pseudo-labeling rule and graph contrastive objective with an explicit ignored region for ambiguous target proposals.
- Evaluation across two domain shifts and multiple unknown-class splits, with ablations that isolate the contribution of each graph objective.

2 Related Work

Unsupervised Domain Adaptation for Object Detection. Adapting object detectors across domains transfers capabilities from annotated source data to target scenarios with distribution shifts Chen et al. (2018); Xu et al. (2020). Representative strategies include adversarial alignment Chen et al. (2018); Zhu et al. (2019), pseudo-label self-training Kim et al. (2019); Rezaeianaran et al. (2021), and feature-distribution regularization Saito et al. (2019). These methods generally assume that source and target share the same categories Wang et al. (2019); Chen et al. (2020). Adaptive open-set object detection instead evaluates both target-domain shift and unknown target categories Li et al. (2023); GraphGen adds the source-free constraint to this setting.

Source-Free Domain Adaptation for Object Detection. Source-free domain adaptation (SFDA) assumes that only a source-trained model and unlabeled target data remain available Kundu et al. (2020); Liang et al. (2020). For detection, LODS removes domain style information Li et al. (2022a), PETS periodically refreshes patch-level teacher knowledge Liu et al. (2023), and recent methods strengthen self-training and low-confidence supervision Hao et al. (2024); Yoon et al. (2024). IRG-SFDA is the closest graph-based precedent: it models relations among target instances in a teacher–student detector VS et al. (2023). GraphGen differs by coupling the target relation graph to an explicit open-set novelty objective and evaluating known detection together with unknown-object recall. We therefore claim a joint source-free open-set formulation rather than graph reasoning itself as new.

Object Detection in Open-world Environments. Open-set and open-world detection require a detector to reject or discover objects outside its labeled vocabulary Bendale & Boult (2016); Dhamija et al. (2020); Joseph et al. (2021). Methods based on low-density feature regions or energy scores improve unknown detection Han et al. (2022); Liu et al. (2020). Source-free open-set DAOD further requires target adaptation without source images; SF-OSDA is the most directly matched baseline in our evaluation Luo et al. (2023). Unlike open-vocabulary detectors, GraphGen retains a fixed known-class vocabulary and predicts a class-agnostic unknown category.

Graph-based and Uncertainty-aware Open-set DA/DAOD. Graph-based DAOD methods such as dual bipartite graph learning and SIGMA explicitly match source and target graphs and consequently require source-domain examples during adaptation Chen et al. (2021); Li et al. (2022b). IRG-SFDA removes that requirement but addresses closed-set adaptation VS et al. (2023). GraphGen instead constructs only a target-side proposal graph and uses it for three coupled operations:

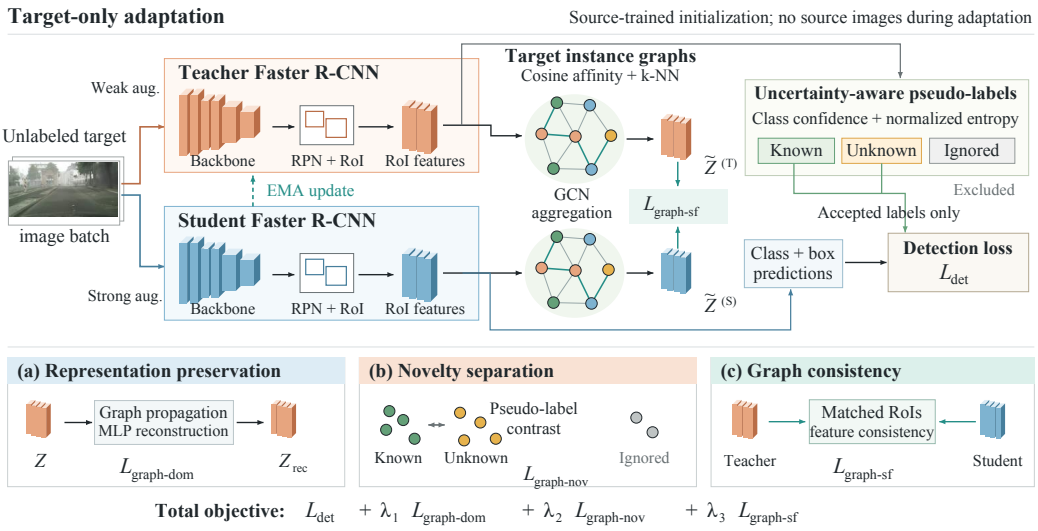

Figure 2: GraphGen adaptation pipeline. Weak and strong target views pass through teacher and student Faster R-CNN detectors. RoI features branch to sparse target-instance graphs and prediction heads. Teacher confidence and normalized entropy determine known, unknown, or ignored pseudo-labels; accepted labels supervise student detection. Insets detail (a) feature reconstruction, (b) pseudo-label contrastive separation, and (c) graph-feature consistency on matched RoIs. The dashed arrow denotes the student-to-teacher EMA update. All adaptation inputs come from the target domain.

preserving source-initialized representations, separating confident known and unknown proposals, and enforcing teacher–student consistency. No source–target graph or paired source image is used during adaptation.

3 Proposed Method

Problem Formulation and Motivation. We separate three axes of the deployment setting. Domain shift changes the marginal input distribution, $\mathcal{P}_s(I) \neq \mathcal{P}_t(I)$. Open-set shift changes the label support: the target contains the known source classes $\Omega_b = \{1, \ldots, K\}$ and unknown categories Ω_u, which are evaluated as one class $K+1$. The source-free constraint removes the labeled source set $\mathcal{D}_s = \{(I_s^i, Y_s^i)\}_{i=1}^{n_s}$ after source training, leaving only the source-trained model Θ_s and unlabeled target images $\mathcal{D}_t = \{I_t^j\}_{j=1}^{n_t}$ during adaptation.

Our goal is to adapt Θ_s to $\mathcal{D}_t$ for accurate recognition of known classes Ω_b and effective detection of unknown objects Ω_u. We propose a unified framework to address all three shifts. The following sections describe each component.

Target Graph Representation Preservation. During adaptation, we regularize target-region features relative to the representation encoded by the source-trained detector, without constructing a source–target graph. The working hypothesis is that relations among target proposals can provide context that is less sensitive to local appearance corruption than an isolated proposal feature; this is an empirical inductive bias rather than an invariance guarantee.

We use Faster R-CNN with a ResNet-50 backbone and extract RoI features $\{\mathbf{z}_i \in \mathbb{R}^d\}_{i=1}^N$ from target proposals whose teacher confidence exceeds the proposal-filtering threshold. A graph is constructed for each target mini-batch from these filtered proposals. Each node $v_i \in \mathcal{V}$ represents one RoI feature and each node retains its predicted box geometry for teacher–student matching.

We define edge weights from cosine similarity in the learned RoI feature space. This graph encodes feature affinity; it does not explicitly encode physical distance or prove spatial invariance.

$$\omega_{ij} = \exp\Big(\gamma \cdot \frac{\langle \mathbf{z}_i, \mathbf{z}_j \rangle}{\|\mathbf{z}_i\|_2 \, \|\mathbf{z}_j\|_2}\Big), \tag{1}$$

where $\gamma > 0$ controls the sharpness of the affinity distribution. Larger values assign relatively more weight to highly similar proposals; novelty decisions are made separately by the rule below. For each node we retain its k highest-affinity neighbors and set all other entries to zero. The resulting sparse affinity matrix $\Omega = [\omega_{ij}]$ is degree-normalized as

$$\widetilde{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}}\,\Omega\,\mathbf{D}^{-\frac{1}{2}}, \tag{2}$$

where $\mathbf{D}$ is a diagonal matrix with $D_{ii} = \sum_j \omega_{ij}$. The normalization limits the direct effect of node degree on aggregation.

With the normalized adjacency matrix, we propagate information across the graph to incorporate contextual information from related objects, enriching feature representations with structural knowledge that helps bridge the domain gap. We implement this through a graph convolution operation:

$$\tilde{\mathbf{Z}} = \Psi\Big(\widetilde{\mathbf{A}}\,\mathbf{Z}\,\mathbf{W}\Big), \tag{3}$$

where $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N]^\top$ is the matrix of all object embeddings, $\mathbf{W}$ is a learnable weight matrix that transforms the features into a suitable space for propagation, and $\Psi(\cdot)$ is a non-linear activation function (in our implementation, GELU) that introduces non-linearity into the propagation process. This graph convolution essentially computes weighted averages of neighboring features, with the weights determined by the normalized adjacency matrix $\widetilde{\mathbf{A}}$.

To discourage graph propagation from excessively changing the source-initialized representation, we introduce a cyclic consistency constraint:

$$\mathcal{L}_{\text{graph}-\text{dom}} = \frac{1}{N}\sum_{i=1}^{N}\Big\|\mathbf{z}_i - \Xi\big(\tilde{\mathbf{z}}_i\big)\Big\|_2^2, \tag{4}$$

where $\Xi(\cdot)$ is a projection operator (two-layer MLP) that maps graph-processed features back to the original space. This cyclic consistency prevents feature drift and preserves semantic content from the source domain.

Graph construction, propagation, and reconstruction regularize target features against drift from the source initialization. We use the name graph alignment for consistency with the module name, while emphasizing that no source features enter this adaptation-time graph.

Graph-guided Novelty Discovery. Graph propagation alone does not distinguish known classes from unknown target categories. We therefore combine detector confidence with normalized predictive entropy and update known-class prototypes from reliable teacher predictions.

For each known class $k \in \Omega_b$, we maintain a prototype vector $\mathbf{p}^k$. To determine whether embedding $\mathbf{z}_i$ is known or unknown, we define a novelty score $\wp(\mathbf{z}_i)$ based on class probability and prediction entropy:

$$\wp(\mathbf{z}_i) = 1 - \max_{c\in\Omega_b}\Big\{p(c \mid \mathbf{z}_i)\Big[1 - \bar{\mathcal{H}}\big(\mathbf{p}(\mathbf{z}_i)\big)\Big]\Big\}, \tag{5}$$

where $p(c \mid \mathbf{z}_i)$ is the detector's probability for class c. The term $\max_{c\in\Omega_b} p(c \mid \mathbf{z}_i)$ reflects model confidence, while normalized entropy $\bar{\mathcal{H}}(\mathbf{p}(\mathbf{z}_i))$ measures uncertainty:

$$\bar{\mathcal{H}}\big(\mathbf{p}(\mathbf{z}_i)\big) = -\frac{1}{\log K}\sum_{c\in\Omega_b} p(c \mid \mathbf{z}_i)\log p(c \mid \mathbf{z}_i), \tag{6}$$

where K is the number of known classes, so $\bar{\mathcal{H}} \in [0, 1]$. The multiplicative term $[1 - \bar{\mathcal{H}}(\mathbf{p}(\mathbf{z}_i))]$ gives low novelty scores to predictions that are both confident and low-entropy, while uncertain predictions receive higher scores.

The prototypes evolve using only target proposals assigned a confident known-class pseudo-label by the teacher. For class k,

$$\mathbf{p}^k \leftarrow \tau\,\mathbf{p}^k + (1-\tau)\,\mathbb{E}\Big[\mathbf{z}_i \mid y_i = k\Big], \tag{7}$$

where the conditional mean is computed over teacher-selected proposals and $\tau \in (0, 1)$ is the prototype momentum. Restricting updates to confident known proposals reduces contamination by uncertain target instances.

To further boost discrimination, we add contrastive learning within the graph. Unknown objects may share features with known classes but have distinct relational patterns. We define a graph-based contrastive loss to pull same-class embeddings closer and push different-class ones apart:

$$\mathcal{L}_{\text{graph}-\text{nov}} = -\frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \log \frac{\exp\big(\text{sim}(\tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j)/\tau_c\big)}{\sum_{a\in\{j\}\cup\mathcal{N}(i)} \exp\big(\text{sim}(\tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_a)/\tau_c\big)}, \tag{8}$$

where sim is cosine similarity between ℓ_2-normalized graph features, τ_c is the temperature, $\mathcal{P}$ contains teacher-consistent positive pairs, and $\mathcal{N}(i)$ contains negatives for anchor i. Including the positive in the normalization term gives a bounded supervised-contrastive objective. Ambiguous proposals do not participate in this loss.

This contrastive mechanism operates on graph-transformed features $\tilde{\mathbf{z}}_i$. It complements prototype-based novelty scoring by encouraging proposals with consistent pseudo-labels to remain close while separating assigned known and unknown proposals.

Graph-aware Self-regularization. We next adapt the detector without source images by enforcing consistency between two augmented target views.

Our solution creates a self-supervised framework that maintains consistency between different views of the same target data. Inspired by Mean-Teacher paradigm, we establish a teacher-student knowledge distillation framework uniquely adapted to our graph-based representation learning approach. For each target image $I_t \in \mathcal{D}_t$, we generate two distinct augmented versions using different stochastic transformation operators:

$$\widetilde{I}_t = \mathcal{T}_s(I_t) \quad \text{and} \quad \widehat{I}_t = \mathcal{T}_w(I_t), \tag{9}$$

where $\mathcal{T}_s(\cdot)$ represents a strong augmentation operator applying aggressive transformations (color jittering, cropping, blur), and $\mathcal{T}_w(\cdot)$ represents a weak augmentation operator applying milder transformations (horizontal flipping, slight color adjustments). This asymmetric strategy challenges the student network while maintaining stable teacher predictions.

Our framework employs two identical networks with different update mechanisms: a student network Θ_S processing strongly augmented inputs $\widetilde{I}_t$, and a teacher network Θ_T processing weakly augmented inputs $\widehat{I}_t$. The teacher's parameters are updated using an exponential moving average (EMA) of the student's parameters:

$$\Theta_T \leftarrow \alpha\,\Theta_T + (1-\alpha)\,\Theta_S, \tag{10}$$

where $\alpha \in (0,1)$ is the EMA decay. The teacher therefore changes more gradually than the student and supplies the pseudo-labels used for target adaptation.

For each teacher proposal q_t, let $m_t = \max_{c\in\Omega_b} p(c \mid q_t)$. We use two confidence thresholds, $\tau_{\text{low}} < \tau_{\text{high}}$, so uncertain proposals remain explicitly ignored:

$$\tilde{y}_t = \begin{cases} \arg\max\limits_{c\in\Omega_b} p(c \mid q_t), & \text{if } m_t \geq \tau_{\text{high}}, \\ K+1, & \text{if } m_t < \tau_{\text{low}} \text{ and } \wp(q_t) \geq \tau_{\text{unknown}}, \\ \text{ignored}, & \text{otherwise} \end{cases} \tag{11}$$

The interval $[\tau_{\text{low}}, \tau_{\text{high}})$ forms an ambiguity band. Proposals in this band, and low-confidence proposals that do not satisfy the novelty criterion, are excluded from prototype updates and supervised target losses. The probabilities $p(c \mid q_t)$ come from the teacher classification head.

The graph-transformed features $\tilde{\mathbf{z}}_i$ capture contextual relationships between objects, providing a holistic classification basis. We formulate a consistency loss between teacher and student networks:

$$\mathcal{L}_{\text{graph}-\text{sf}} = \frac{1}{N} \sum_{i=1}^{N} \left\| \tilde{\mathbf{z}}_i^{(T)} - \tilde{\mathbf{z}}_i^{(S)} \right\|_2^2, \tag{12}$$

where $\tilde{\mathbf{z}}_i^{(T)}$ and $\tilde{\mathbf{z}}_i^{(S)}$ are graph-transformed RoI features for teacher–student proposals matched after accounting for the geometric augmentations.



Figure 3: Qualitative comparison on source-free AOOD from Cityscapes to Foggy Cityscapes among (a) DETR Carion et al. (2020), (b) OW-DETR Gupta et al. (2022), and (c) GraphGen. Box colors follow the class legend embedded in the figure.

Table 1: Quantitative comparison on source-free open-set object detection from Cityscapes to Foggy Cityscapes. **Top-1** and Top-2 values are marked within each split and metric. Method citations are given in Section 4.

Method	Set	num. novel-class: 3				num. novel-class: 4				num. novel-class: 5			
		$mAP_b \uparrow$	$AR_n \uparrow$	WI↓	AOSE↓	$mAP_b \uparrow$	$AR_n \uparrow$	WI↓	AOSE↓	$mAP_b \uparrow$	$AR_n \uparrow$	WI↓	AOSE↓
OW-DETR	inst-sem	37.12	3.81	0.493	25	37.92	3.61	0.858	65	34.85	3.09	0.977	144
PETS	inst-sem	38.11	4.17	0.525	24	38.90	3.73	0.884	66	37.04	3.31	1.011	151
LODS	inst-sem	38.27	3.81	0.506	22	39.75	3.61	0.950	63	38.99	2.99	1.102	143
SF-OSDA	inst-sem	38.31	3.45	0.405	25	38.60	3.61	0.787	73	38.55	3.03	0.940	171
GraphGen (Ours)	inst-sem	**40.53**	**4.38**	**0.378**	**20**	**40.27**	**3.92**	**0.743**	**60**	**39.84**	**3.42**	**0.854**	**140**
OW-DETR	hom-sem	28.07	2.20	0.679	69	26.63	2.81	1.053	123	25.52	6.92	3.276	818
PETS	hom-sem	28.30	2.90	1.138	53	27.39	3.76	1.962	116	25.91	7.64	5.234	811
LODS	hom-sem	28.85	4.19	1.498	**48**	28.35	5.08	2.297	**75**	27.21	8.49	4.571	**75**
SF-OSDA	hom-sem	29.00	3.64	1.360	52	28.14	4.42	1.980	130	26.70	8.02	3.954	786
GraphGen (Ours)	hom-sem	**30.24**	**4.48**	**1.269**	55	**29.83**	**5.28**	**1.853**	120	**28.57**	**8.78**	**3.758**	790
OW-DETR	freq-dec	45.06	7.67	0.855	168	42.97	10.00	1.364	331	42.95	11.03	1.609	460
PETS	freq-dec	45.95	8.27	0.889	173	44.69	10.34	1.362	355	43.63	11.30	1.749	505
LODS	freq-dec	46.13	8.63	1.090	216	45.03	11.01	1.651	409	44.80	11.91	1.943	577
SF-OSDA	freq-dec	46.32	9.35	1.222	245	45.56	12.02	1.763	453	45.13	11.32	2.099	637
GraphGen (Ours)	freq-dec	**47.58**	**10.54**	**0.825**	**165**	**46.23**	**12.57**	**1.253**	**320**	**45.86**	**12.83**	**1.557**	**450**
OW-DETR	freq-inc	25.26	3.47	1.936	135	22.17	3.17	1.889	212	20.82	2.92	2.549	558
PETS	freq-inc	25.40	3.84	1.812	113	22.25	3.28	2.165	191	21.46	2.98	2.700	528
LODS	freq-inc	25.92	3.58	3.094	109	23.69	3.46	2.841	181	21.66	3.07	3.197	515
SF-OSDA	freq-inc	25.99	3.37	1.764	124	22.80	3.22	1.824	208	21.50	2.85	2.571	552
GraphGen (Ours)	freq-inc	**26.83**	**3.97**	**1.017**	**67**	**24.94**	**3.68**	**0.993**	**113**	**22.58**	**3.29**	**1.487**	**314**

Together, pseudo-label filtering and graph consistency provide the target-only supervision used during source-free adaptation.

Overall Optimization Objective. After addressing each challenge, we integrate all components into a unified optimization framework. The total objective combines detection loss and our regularization terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det}} + \lambda_1 \mathcal{L}_{\text{graph-dom}} + \lambda_2 \mathcal{L}_{\text{graph-nov}} + \lambda_3 \mathcal{L}_{\text{graph-sf}}. \tag{13}$$

Here, $\mathcal{L}_{\text{det}}$ contains the Faster R-CNN classification and box-regression losses on accepted pseudo-labels; the remaining terms implement graph reconstruction, novelty separation, and graph consistency. The coefficients λ_1, λ_2, and λ_3 balance these terms.

The objectives share the same target proposal features and graph. In practice, we first enable graph reconstruction and then introduce novelty separation and teacher–student consistency after warm-up, reducing the influence of early uncertain pseudo-labels.

4 Experiments

Datasets and Evaluation Metrics We evaluate GraphGen on Cityscapes→Foggy Cityscapes Cordts et al. (2016); Sakaridis et al. (2018) and Pascal VOC→Clipart Everingham et al. (2010); Inoue et al.

Table 2: Quantitative results on source-free open-set object detection from Pascal VOC to Clipart, with **Top-1** and Top-2 marked. Method citations are given in Section 4.

Method	Ω_n	mAP$_b$ ↑	AR$_n$ ↑	WI↓	AOSE↓
OW-DETR		15.56	30.50	8.028	3431
PETS		16.05	25.38	9.085	4872
LODS	6	16.65	21.44	9.413	5108
SF-OSDA		16.67	31.93	7.927	4699
GraphGen (ours)		**17.20**	**32.15**	**7.654**	**3401**
OW-DETR		14.85	31.85	8.467	3871
PETS		15.98	25.71	9.674	5496
LODS	8	15.69	22.25	9.966	5790
SF-OSDA		16.16	32.38	8.289	5250
GraphGen (ours)		**16.88**	**33.12**	**8.001**	**3551**
OW-DETR		14.27	34.06	9.866	4995
PETS		14.37	26.89	10.752	6874
LODS	10	14.12	22.42	10.538	7271
SF-OSDA		15.22	34.77	9.314	6640
GraphGen (ours)		**15.88**	**35.77**	**9.114**	**4781**

(2018), representing weather and style shifts. We follow the adaptive open-set object-detection protocol of Li et al. Li et al. (2023). Its four split families, denoted `inst-sem`, `hom-sem`, `freq-dec`, and `freq-inc`, vary the semantic relation and frequency trend between known and unknown categories; all methods use the same published split files.

For each dataset pair, source categories form the known set and the held-out target categories are mapped to a class-agnostic unknown label. We use the published category files, split names, IoU conventions, averaging rules, NMS settings, and evaluation implementation of Li et al. Li et al. (2023); no category is redefined by us. We report mean average precision over known classes (mAP$_b$), average recall over unknown objects (AR$_n$), Wilderness Impact (WI), and Absolute Open-Set Error (AOSE) under that implementation. Lower WI and AOSE indicate fewer unknown-to-known errors. We evaluate 3, 4, and 5 unknown classes on Cityscapes→Foggy Cityscapes and 6, 8, and 10 on Pascal VOC→Clipart.

Implementation Details We use Faster R-CNN with a ResNet-50 backbone. Source training and target adaptation use SGD with learning rate 10^{-3}, momentum 0.9, weight decay 10^{-4}, and batch size 4 on NVIDIA A100 GPUs. Adaptation lasts 12 epochs; graph reconstruction is enabled first and novelty and teacher–student losses are introduced after the warm-up stage. We use $k = 5$ neighbors, prototype momentum $\tau = 0.9$, teacher decay $\alpha = 0.99$, confidence thresholds $\tau_{\text{low}} = 0.6$ and $\tau_{\text{high}} = 0.8$, and novelty threshold $\tau_{\text{unknown}} = 0.6$. Hyperparameters are selected using labeled source training/validation data or values from the cited baselines and are fixed before target evaluation; no target labels are used for model selection.

Source-free Adaptation of Baselines The comparison includes open-world detector OW-DETR Gupta et al. (2022) and source-free detection methods PETS Liu et al. (2023), LODS Li et al. (2022a), and SF-OSDA Luo et al. (2023). OW-DETR is adapted to the target-only phase using teacher-generated pseudo-labels. The other three methods retain their source-free objectives and native detector components. All methods use the same data splits, target images, evaluation metrics, and adaptation budget. Table 4 summarizes the comparison protocol.

4.1 Main Empirical Results

We compare GraphGen with OW-DETR Gupta et al. (2022), PETS Liu et al. (2023), LODS Li et al. (2022a), and SF-OSDA Luo et al. (2023) under the protocol above. The tables support comparison with these methods only; we do not claim exhaustive coverage of all recent source-free detectors.

Table 1 shows that GraphGen obtains the highest mAP$_b$ and AR$_n$ among the compared methods across the reported Cityscapes→Foggy Cityscapes splits. In the `freq-inc` setting with five un-

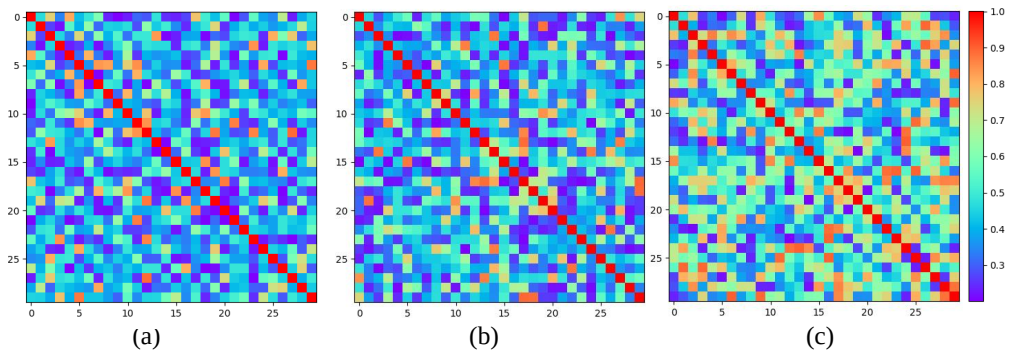

Figure 4: Prediction-to-ground-truth matching matrices on Cityscapes→Foggy Cityscapes for (a) DETR, (b) OW-DETR, and (c) GraphGen. Warmer diagonal entries indicate stronger matched-box affinity.



Figure 5: Post-hoc correspondence visualization between paired Cityscapes examples (a) and their Foggy Cityscapes counterparts (b). Lines are computed only for analysis after adaptation; source images are never used by the adaptation algorithm or for model selection.

known classes, it reaches 22.58% mAP_b and 3.29% AR_n. The result supports the benefit of target relation modeling under this evaluation protocol, but does not by itself establish robustness beyond the two tested domain pairs.

Table 2 provides a second domain pair. With eight unknown target categories, GraphGen achieves 16.88% mAP_b and 33.12% AR_n, while reporting WI of 8.001 and AOSE of 3551. The ablation in Table 3 further shows that the complete configuration has higher mAP_b and AR_n than each reported removal variant; uncertainty across repeated runs remains unmeasured.

Figure 3 presents qualitative comparisons among DETR Carion et al. (2020), OW-DETR Gupta et al. (2022), and GraphGen on Cityscapes→Foggy Cityscapes. The embedded legend maps yellow, red, purple, and blue boxes to car, person, rider, and unknown detections, respectively. The examples illustrate how GraphGen changes known/unknown assignments under fog, but they are qualitative evidence rather than a substitute for the aggregate metrics.

4.2 Further Empirical Study

Ablation Study Table 3 isolates the three graph objectives under a fixed ablation protocol with a shorter schedule and default thresholds. This protocol is intentionally different from the final, tuned evaluation in Table 1, so its absolute values should not be compared row-by-row with Table 1. Within the ablation, removing graph alignment, novelty discovery, or self-regularization lowers mAP_b and AR_n relative to the full model. Cyclic consistency contributes beyond graph propagation alone. WI and AOSE are unchanged between the full model and the no-self-regularization variant at the reported precision, so the evidence supports an accuracy and recall benefit for self-regularization but not a measurable open-set error benefit in this setting.

Box Matching Analysis Figure 4 visualizes prediction-to-ground-truth affinities for DETR, OW-DETR, and GraphGen. GraphGen produces a more concentrated diagonal pattern in these examples, consistent with fewer mismatched boxes; this diagnostic is not used during training.

Table 3: Ablation study on source-free AOOD from Cityscapes to Foggy Cityscapes. All rows use the same ablation protocol: a shorter training schedule and default thresholds, unlike the final tuned protocol in Table 1. The table shows the impact of each component, with **Top-1** and Top-2 performance marked.

Model Variants	**mAP**$_b$ ↑	**AR**$_n$ ↑	**WI** ↓	**AOSE** ↓
Source Only	37.10	1.83	0.684	203
No Graph-dom	37.86	2.11	0.739	204
Graph-dom w/o CC	38.22	3.33	0.991	169
No Graph-nov	38.55	3.03	0.940	171
No Graph-sf	39.96	3.33	**0.684**	**169**
GraphGen (Full)	**41.06**	**4.42**	**0.684**	**169**

Post-hoc Correspondence Analysis Figure 5 links detections between paired clear and foggy versions of the same scenes after training. This diagnostic visualizes whether the learned target representation preserves class-consistent correspondences under corruption. It is not an inter-domain graph used by GraphGen, and neither the source image nor the displayed pair influences adaptation.

Complexity and Limitations With k-NN sparsification, graph message passing scales as $O(Nk)$ in the number of retained proposals rather than requiring dense $O(N^2)$ propagation. End-to-end latency, memory, multi-seed variation, and threshold sensitivity are not reported in the present study and remain necessary for a complete deployment analysis. The evaluation is also limited to two domain pairs and a fixed Faster R-CNN backbone. These limits motivate future tests on larger shifts, additional detector families, and stronger recent source-free baselines such as IRG-SFDA, SF-UT, and LPLD.

Failure Modes and Practical Implications GraphGen is most likely to fail when early pseudo-label errors form locally coherent but semantically incorrect neighborhoods. Message passing can then amplify an erroneous assignment instead of correcting it, particularly for small objects, rare known classes, or unknown objects that resemble a known category. The ignored uncertainty region limits direct supervision from ambiguous proposals, but it does not guarantee that all retained neighbors are reliable. In practice, graph adaptation should therefore be monitored jointly through known-class confidence, unknown-object recall, and the proportion of ignored proposals rather than through a single detection metric. Conservative thresholding is preferable when false unknown alarms are costly, whereas recall-oriented settings may admit more uncertain proposals. These operating choices are deployment-specific and should be calibrated on target-like validation data without reintroducing source images.

5 Conclusion

This paper presented GraphGen, a target-only graph framework for source-free open-set domain-adaptive object detection. Across the two evaluated domain pairs, the complete model improves known-class detection and unknown-object recall over the included baselines. The ablations show lower mAP$_b$ and AR$_n$ when each graph objective is removed, while the unchanged WI and AOSE for one removal cautions against attributing every open-set gain to self-regularization. The approach is applicable when source images cannot be retained, but the evidence is limited to two benchmarks, a Faster R-CNN backbone, and single reported runs. Future work should add modern source-free baselines, multi-seed uncertainty, threshold and graph-density sensitivity, and measured runtime and memory before making broader robustness or deployment claims. A stronger evaluation should also stress-test the reliability of known/unknown pseudo-labels when target shifts are larger, category overlap is weaker, or many proposals remain ambiguous; extending the graph construction to adaptively determine its density is a particularly important step toward deployment. In particular, the current study does not isolate the effect of neighbor quality or the interaction between pseudo-label noise and message passing. A controlled comparison with simpler confidence-based regularization, together with per-class and failure-case analyses, would clarify where structural transfer helps and where noisy relations may reinforce incorrect assignments.

REFERENCES

Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1563–1572, 2016.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *ECCV*, 2020.

Chaoqi Chen, Zebiao Zheng, Xinghao Ding, Yue Huang, and Qi Dou. Harmonizing transferability and discriminability for adapting object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8869–8878, 2020.

Chaoqi Chen, Jiongcheng Li, Zebiao Zheng, Yue Huang, Xinghao Ding, and Yizhou Yu. Dual bipartite graph learning: A general approach for domain adaptive object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2703–2712, October 2021.

Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3339–3348, 2018.

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*, pp. 3213–3223, 2016.

Akshay Dhamija, Manuel Gunther, Jonathan Ventura, and Terrance Boult. The overlooked elephant of object detection: Open set. In *WACV*, 2020.

Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 88(2):303–338, 2010.

Akshita Gupta, Gabriela Csurka, Nicolas Thome, and Matthieu Cord. Ow-detr: Open-world object detection with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9378–9387, 2022.

Jiaming Han, Yuqiang Ren, Jian Ding, Xingjia Pan, Ke Yan, and Gui-Song Xia. Expanding low-density latent regions for open-set object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9581–9590, 2022. doi: 10.1109/CVPR52688.2022.00937.

Yan Hao, Florent Forest, and Olga Fink. Simplifying source-free domain adaptation for object detection: Effective self-training strategies and performance insights. In *European Conference on Computer Vision*, pp. 196–213. Springer, 2024. doi: 10.1007/978-3-031-72949-2_12.

Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, and Kiyoharu Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5001–5009, 2018.

KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In *CVPR*, 2021.

Seunghyeon Kim, Jaehoon Choi, Taekyung Kim, and Changick Kim. Self-training and adversarial background regularization for unsupervised domain adaptive one-stage object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6092–6101, 2019.

Jogendra Nath Kundu, Naveen Venkat, R Venkatesh Babu, et al. Universal source-free domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4544–4553, 2020.

Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9641–9650, 2020.

Shuaifeng Li, Mao Ye, Xiatian Zhu, Lihua Zhou, and Lin Xiong. Source-free object detection by learning to overlook domain style. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8014–8023, 2022a.

Wuyang Li, Xinyu Liu, and Yixuan Yuan. SIGMA: Semantic-complete graph matching for domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5281–5290, 2022b. doi: 10.1109/CVPR52688.2022.00522.

Wuyang Li, Xiaoqing Guo, and Yixuan Yuan. Novel scenes & classes: Towards adaptive open-set object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15734–15744, 2023. doi: 10.1109/ICCV51070.2023.01446.

Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, pp. 6028–6039. PMLR, 2020.

Weitang Liu, Xiaoyun Wang, John D Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 33:21464–21475, 2020.

Zekun Liu, Xinyu Peng, Zikun Liu, Li Shen, Jian Li, Hong Lu, Chun Chen, and John See. Periodically-refreshed patch-based knowledge distillation for source-free domain adaptive object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17833–17843, 2023.

Xiran Luo, Guoliang Kang, and Jian Liang. Source-free open-set domain adaptation for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23971–23980, 2023.

Farzaneh Rezaeianaran, Pascal Mettes, and Cees GM Snoek. Seeking for all-round objects: A pseudo-gt-based self-training approach for domain adaptive object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3191–3201, 2021.

Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6956–6965, 2019.

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

Vibashan VS, Poojan Oza, and Vishal M. Patel. Instance relation graph guided source-free domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3520–3530, 2023. doi: 10.1109/CVPR52729.2023.00343.

Yu-Neng Wang, Tsun-Yi Wu, Yu-Chiang Frank Yeh, and Yu-Chiang Frank Wang. Towards domain adaptive object detection on a foggy day. In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4270–4274. IEEE, 2019.

Chang-Dong Xu, Xing-Ran Zhao, Xin Jin, and Xiu-Shen Wei. Exploring categorical regularization for domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11724–11733, 2020.

Ilhoon Yoon, Hyeongjun Kwon, Jin-Kuk Kim, Junyoung Park, Hyunsung Jang, and Kwanghoon Sohn. Enhancing source-free domain adaptive object detection with low-confidence pseudo label distillation. In *European Conference on Computer Vision*, pp. 337–353. Springer, 2024. doi: 10.1007/978-3-031-72907-2_20.

Xinge Zhu, Jiangmiao Pang, Ceyuan Yang, Jianping Shi, and Dahua Lin. Adapting object detectors via selective cross-domain alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 687–696, 2019.

A SOURCE-FREE BASELINE PROTOCOL

Shared evaluation protocol. The comparison uses the published category files and split names of Li et al. Li et al. (2023). We do not define alternative known/unknown categories or metric conventions; their evaluation implementation specifies the IoU thresholds, averaging rules, and NMS settings used for all methods. Consequently, exact category lists should be read from the released split files rather than inferred from the number of unknown classes.

Baseline policy. For every baseline, the detector is first trained on the labeled source domain. Source images and labels are then discarded. Any training component that originally relies on source supervision is replaced during adaptation by the corresponding target-only operation: high-confidence teacher pseudo-labels replace source annotations, and source–target alignment terms operate only on differently augmented target views. All methods receive the same target images, split files, evaluation code, and adaptation budget. The table below summarizes the component retained by each method under this policy.

Table 4: Role of each baseline in the common target-only evaluation.

Method	Original scope	Evaluation in this paper
OW-DETR	Open-world detection	Source-trained initialization followed by target-only pseudo-label adaptation.
PETS	Source-free detection	Original target-only patch-refresh teacher–student objective under the common data split and budget.
LODS	Source-free detection	Original target-only style-overlooking objective under the common data split and budget.
SF-OSDA	Source-free open-set detection	Original target-only open-set objective under the common data split and metrics.

B DECLARATION OF LLM USAGE

During the writing of the manuscript, we utilized a Large Language Model (ChatGPT) as a writing assistant. The scope of its usage was limited to **improving grammar, polishing sentences, and enhancing the clarity and fluency of this manuscript**. The method, claims, experimental results and conclusions are developed by the authors.