ASGS: Single-Domain Generalizable Open-Set Object Detection via Adaptive Subgraph Searching

Yuxuan Yuan[1]*, Luyao Tang[1,2,*], Yixin Chen[1,2], Chaoqi Chen[3,†], Yue Huang[1,2] and Xinghao Ding[1,2]
[1] Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University
[2] School of Informatics, Xiamen University [3] Shenzhen University
{lytang, yuanyuxuan0908}@stu.xmu.edu.cn, cqchen1994@gmail.com, yhuang2010@xmu.edu.cn

Abstract

Albeit existing Single-Domain Generalized Object Detection (Single-DGOD) methods enable models to generalize to unseen domains, most assume that the training and testing data share the same label space. In real-world scenarios, unseen domains often introduce previously unknown objects, a challenge that has been largely overlooked. In this paper, we tackle the practical problem of Single-domain Generalizable Open-Set Object Detection (SG-OSOD), which addresses both unseen domains and unknown classes. We identify two key challenges: (1) detecting unknown classes with only known-class data, and (2) learning robust features to mitigate domain shift. To address these challenges, we propose the framework termed `ASGS`*, which leverages adaptive subgraph structures to enhance the understanding of unknown scenes and classes.* `ASGS` *consists of Subgraph-wise Unknown-class Learning (SUL) and Class-wise Embedding Compaction (CEC). SUL employs non-parametric methods to detect unknown samples and performs Adaptive Subgraph Searching (ASS) for high-order structural feature extraction, enabling domain-robust unknown class learning. Moreover, the CEC module enhances class discrimination robustness through contrastive learning, which results in more compact class clusters in unknown scenarios. Experimental results demonstrate the effectiveness of the proposed* `ASGS`.

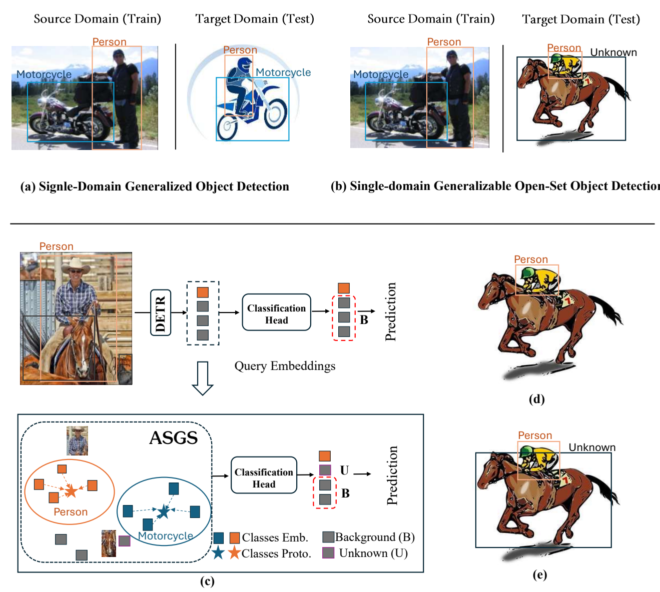


Figure 1. **Top:** Illustration of (a) the Single-DGOD setting, and (b) the proposed SG-OSOD setting. **Bottom:** Comparison of `ASGS` with Single-DGOD detector. (c) `ASGS`'s scheme. (d) Detection examples from a vanilla Single-DGOD detector, where "horse" as an unknown class is misclassified as background, and (e) detection examples from `ASGS`, which successfully detects "horse" as an unknown class due to its integrated unknown-class awareness.

1. Introduction

Object detection is a foundational task in computer vision, with widespread applications in areas such as autonomous driving and robotics. However, in practical applications, deep neural networks are often vulnerable to distribution shifts [6]. For instance, in the field of autonomous driving, the performance of object detection systems can be significantly degraded due to varying weather and lighting conditions. This challenge has spurred research into unsupervised domain adaptation object detection (DAOD) [2–6, 27, 30, 38, 55], which aims to transfer an object detector from a labeled source domain to a novel one with a shared class space. However, DAOD requires simultaneous access to source and target domain data, which is often infeasible in real-world scenarios.

To overcome these limitations, domain generalizable ob-

* Equal Contribution; † Corresponding Author.

ject detection (DGOD) [8, 22, 44, 47] has emerged to enable detectors to generalize to unseen domains without accessing target data. Among DGOD settings, Single-DGOD [43, 46], which requires only one source domain for training, is particularly appealing due to its data efficiency and practical deployability. However, current DGOD methods assume that source and target domains share the same label space, which rarely holds in practice as novel environments often introduce previously unknown objects. In response to this issue, recent works on open-set object detection (OSOD) [10, 13, 15, 35, 53, 54] have concentrated on detecting unknown classes, aiming to enhance the ability of detection models to identify not only known objects but also previously unseen categories in a given dataset. Despite this progress, addressing the dual challenge of unknown domains (domain shift) and unknown objects (class shift) remains an open and pressing problem. To this premise, an open question arises: *Can detection models be trained to recognize unknown objects and apply this knowledge reliably in unseen environments?*

Hence, we focus on the more difficult yet underexplored problem of Single-domain Generalizable Open-Set Object Detection (SG-OSOD), in which both domain shift and open classes occur in the unseen target data. To attain this objective, in this paper, we present a simple yet effective framework (ASGS) for SG-OSOD, as illustrated in Fig. 1. Specifically, ASGS introduces two key components: (i) Subgraph-wise Unknown-class Learning (SUL) module, which adaptively constructs subgraphs of varying orders to achieve cross-domain stable modeling of unknown classes, and (ii) Class-wise Embedding Compaction (CEC) module, which learns compact within-class feature representations to improve robustness in unseen scenarios.

Effective detection of unknown classes hinges on utilizing unlabeled background information. Compared to meaningless backgrounds, unknown classes often share more feature similarities with known classes [41]. Building upon this, in SUL, we first employ a non-parametric method to identify boundary points from known classes and potential unknown samples. However, these unknown points constitute only low-order cues derived from individual samples, which are vulnerable to domain shift. To overcome this, we develop Adaptive Subgraph Searching (ASS), which introduces a paradigm shift from individual sample-based detection to structural relationship modeling. Specifically, ASS dynamically constructs and analyzes subgraphs of varying orders to capture the intrinsic relationships between known and unknown instances. By modeling relationships among multiple entities, the model can capture more stable structural information, resulting in structured representations that are more robust to domain shift compared to individual visual features. To further mitigate the effects of domain shift, we propose CEC as a complementary module. CEC introduces contrastive learning to obtain compact intra-class feature representations while maintaining clear separation between class clusters. This design works in synergy with SUL by ensuring that the discovered structural patterns remain discriminative even under significant domain variations, as the preserved inter-domain distances help maintain stable classification boundaries.

Our key contributions are summarized as follows:

- We study a real-world friendly problem, Single-domain Generalizable Open-Set Object Detection (SG-OSOD), and propose a principled framework (ASGS) to jointly consider domain and class shift.
- Technically, we introduce SUL, which leverages non-parametric methods and novel adaptive subgraph structures to detect unknown classes and capture domain-invariant relationships. Complementarily, we develop CEC, which employs contrastive learning to create compact class-wise feature representations that maintain discriminative power across different domains.
- Experimentally, through evaluations on three distinct SG-OSOD benchmarks, we demonstrate that ASGS achieves state-of-the-art performance across various scenarios.

2. Related Work

2.1. Domain Generalized Object Detection

Domain Generalizable Object Detection (DGOD) [21, 23–25, 29, 31, 32, 37, 45, 46, 52, 57] has recently garnered significant attention due to its enhanced performance over domain-adaptive object detection and its ability to operate independently of target domain data. For example, GUOD [32] introduced a data augmentation technique aimed at increasing the domain diversity of a limited dataset, thereby enhancing its overall generalization performance. DIDN [31] proposed a domain-invariant disentangled network, which disentangles image-level and instance-level features across various source domains to build a universal object detector. RAPT [52] utilized a comprehensive evaluation benchmark along with an innovative technique termed “region-aware proposal reweighting”. Further research has concentrated on Single-DGOD [43, 45, 46], initially introduced through CDSD [46], which utilizes a cyclic-disentangled self-distillation technique to learn domain-invariant representations without requiring domain-specific labels. How to endow source-domain detectors with the capability to detect unknown open-set categories remains an area requiring further investigation.

2.2. Open-Set Domain Generalization

Open-Set Domain Generalization (OSDG) is still in its early stages. DAML [40] proposes a framework based on domain augmentation and meta-learning. Since the model is not explicitly or implicitly informed by the existence of potential

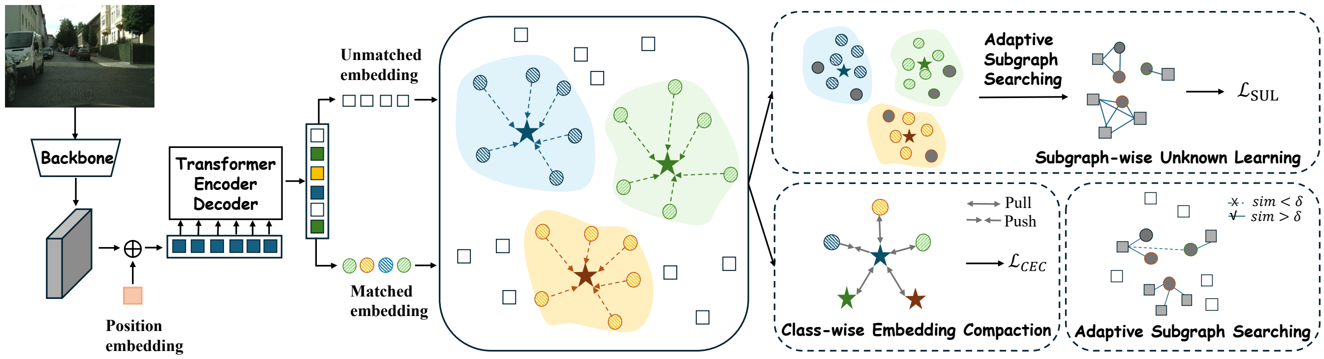


Figure 2. **Overview of the proposed ASGS.** ASGS comprises two key components: Subgraph-wise Unknown-class Learning (SUL) module adaptively constructs subgraphs to detect unknown objects, where Adaptive Subgraph Searching (ASS) dynamically builds connections between samples based on similarity thresholds, while the Class-wise Embedding Compaction (CEC) module enforces compact feature representations using contrastive learning.

unknown classes in both training and inference phases, this work is in a way more focused on close-set domain generalization instead. [20] borrows metric learning to diffuse the feature representations of unknown classes but relies on existing DG baselines to acquire domain-invariant features. In addition, Open-Set single domain generalization (OS-SDG) has emerged [48, 58], which differs from OSDG in that only one source domain is available for training. CrossMatch [55] recommends imitating unknown classes via adversarial learning, and OneRing [48] synthesizes unknown class samples by masking the labels of original ones. Nevertheless, these existing approaches primarily focus on recognition tasks, and there remains a notable absence of research addressing detection tasks, particularly the critical intersection of domain generalization and open-set detection.

2.3. Open-Set Object Detection

Open-Set Object Detection (OSOD) [10, 13, 15, 34, 35, 50, 53, 54, 60] aims to train object detectors capable of identifying both known and unknown objects, drawing significant attention due to its potential for real-world applications. The concept of the OSOD problem was first introduced in [9], where the authors established a benchmark for the OSOD problem and discussed various open-set detectors based on classifier design. OpenDet [15] proposes focusing on location for unknown classes to generate high-quality pseudo labels. OW-DETR [14] extends the open-set learning into detection transformer [59] and selects the activated and unmatched object queries for self-training. Additionally, some works [11, 56] leverage two base-class samples to synthesize representations for unknown classes during model optimization. PROB [60] enhanced the recall of unknown classes by modeling the OSOD problem as a class-agnostic Gaussian distribution. CAT [34] enhances the detection of unknown classes by simulating human-like scene observation in an open world. Despite this, these existing methods heavily rely on low-order embeddings to perceive unknown features, making them particularly susceptible to domain shifts in cross-domain scenarios.

3. Methodology

3.1. Preliminaries

Problem Statement. We have a single labeled source domain $\mathcal{D}_S = \{X_s^i, Y_s^i\}_{i=1}^{n_s}$ of n_s labeled samples and unlabeled target domain $\mathcal{D}_t = \{X_t^i, Y_t^i\}_{i=1}^{n_t}$. $\mathcal{D}_S$ and $\mathcal{D}_t$ are sampled from different data distributions, $P_s \neq P_t$. The labels of the source domain images consist of a set of parameters $Y_s = \{(b_x, b_y, b_w, b_h), y_s\}$, where (b_x, b_y, b_w, b_h) are the coordinates of the bounding box, and y_s represents the target category. Unlike Single-DGOD, the source and target domains are defined within the same class space. SG-OSOD considers a set of classes C_s for the source domain and C_t for the target domain, where $C_s \subset C_t$ and $C_u = C_t \backslash C_s$ represents the *unknown* classes that appear in the target domain. The objective of SG-OSOD is to train a model on $\mathcal{D}_S$ that can detect all objects from $\mathcal{D}_t$, including both the known classes $|C_s|$ and the unknown classes, achieving detection across a total of $|C_s| + 1$ classes [14].

Overall Architecture. The overview of ASGS is demonstrated in Fig. 2, which is based on Deformable DETR (D-DETR) [59]. For a batch-wise source image $\{X_s^i, Y_s^i\}_{i=1}^{B}$, D-DETR first uses a feature extraction backbone to extract image-level features $\{F_s^i\}_{i=1}^{B}$. Then the image-level features along with supplement positional encoding are sent into the deformable transformer encoder and decoder [42] to obtain a set of N query embedding $Q \in \mathbb{R}^{N \times D}$. Subsequently, based on bipartite graph matching [1], we identify the ground-truth matched object queries Q_m and unmatched object queries Q_{um}. We then compute the known-class prototypes $\{\mu_s^k\}_{k=1}^{|C_s|}$, i.e., the class centroids, based on Q_m. In SUL, we first identify the boundary known samples for each class $\mathbb{Q}^k$ by selecting the K intra-class samples farthest from the class center. Then, using KNN distance, we locate the top M unmatched embeddings $\hat{Q}_{um}$ closest to these boundary samples. After that, we apply the Adaptive Subgraph Searching (ASS) algorithm to construct subgraphs $\mathcal{G}$ of varying orders adaptively, leveraging these subgraphs to facilitate unknown class learning $\mathcal{L}_{SUL}$. Meanwhile, as for CEC, we utilize the contrastive loss in a fully supervised way. Specifically, each class prototype μ_s^k serves as a query, with matched embeddings within the class as positive queries $\boldsymbol{c}^+$, and prototypes and embeddings from other classes as negative queries $\boldsymbol{c}^-$. CEC promotes compact feature representations within known classes via $\mathcal{L}_{CEC}$.

3.2. Subgraph-wise Unknown-class Learning (SUL)

Standard D-DETR generates a set of N query embeddings for each image, with each query embedding used by the detection head to produce the final predictions. However, extending D-DETR to detect unknown objects poses a significant challenge: without explicit labels for unknown objects, these samples are treated as background during training, severely hindering accurate detection. While existing OSOD approaches attempt to address this by filtering unmatched queries under certain conditions as pseudo-labels for unknown classes, their reliance on single-sample embedding features makes them vulnerable to domain shifts.

To overcome these limitations, we propose SUL, a novel module that constructs adaptive high-order subgraphs to capture intrinsic relationships between known and unknown embeddings. By modeling the structural connections rather than isolated features, this approach enables more stable representations that remain robust despite significant domain variations. These high-order subgraphs preserve essential relational patterns that single-sample methods fail to capture, making our approach particularly effective when generalizing to unseen domains.

Specifically, given a batch of source images $\{X_s^i, Y_s^i\}_{i=1}^B$ containing $|C_s|$ known classes. A feature extraction backbone and deformable transformer are deployed to obtain N query embedding $Q_s \in \mathbb{R}^{N\times D}$, for simplicity, we omit the dimension of batch-size. Q_s are then fed into the detection head to generate prediction results. Based on bipartite graph matching, certain query embeddings are matched to ground truth boxes $Q_m \subseteq Q_s$, corresponding to features of known classes. The remaining query embeddings $Q_{um} = Q_s \backslash Q_m$ encompass both unknown classes and background objects. To effectively identify unknown samples, we first utilize Q_m to estimate the class-conditional object prototype $\{\mu_s^k\}_{k=1}^{|C_s|}$ (the pentagram in Fig. 2), where $\mu_s^k \in \mathbb{R}^D$. Then the class-conditional prototypes are updated in an Exponential-Moving-Average (EMA) [26] manner for each known-class $k \in \{1, 2, ..., |C_s|\}$ with matched queries Q_m^k:

$$\mu_s^k := \text{Normalize}\left(\alpha\mu_s^k + (1-\alpha)\,\text{Mean}(Q_m^k)\right), \qquad (1)$$

where $\text{Mean}(\cdot)$ indicates the statistical mean of object query embeddings, $\text{Normalize}(\cdot)$ is normalization on the object embeddings, and $\alpha \in (0, 1)$ is the prototype update factor.

Boundary-aware Known/Unknown Sample Discovery. To distinguish unknown class features from background, we exploit that unknown classes typically share similarities with known classes at their boundaries. We first identify the boundary known samples for each class based on the L_2 distance between the class prototype μ_s^k and all matched query embeddings within the class Q_m^k. If a matched query embedding has a large distance from its prototype, it is likely to be on the boundary of the known sample in the feature space. Thus, we select embeddings with the largest L_2 distance and denote this set of boundary samples for each class k as $\mathbb{Q}^k = (q_1^k, q_2^k, ..., q_K^k)$, where $q_i^k \in Q_m^k$.

Then we leverage the non-parametric K-nearest neighbor (KNN) [18] to identify the unknown samples near the boundary. Assume that the L_2-normalized feature of unmatched query embeddings is $\mathbf{z}_i = q_{um,i}/\parallel q_{um,i} \parallel$ and $q_{um,i} \in Q_{um}$. For a normalized boundary known query embedding $\mathbf{q}_i^k$ we compute its KNN distance in $\mathbb{Z}$:

$$d_M(\mathbf{q}_i^k, \mathbb{Z}) = \parallel \mathbf{q}_i^k - \mathbf{z}_{(m)} \parallel_2, \qquad (2)$$

where $\mathbf{z}_{(m)}$ denotes the m-th nearest neighbor in $\mathbb{Z}$. Through the above steps, we identify M nearest unmatched query embeddings $\hat{Q}_{um} \subseteq Q_{um}$ for each boundary known sample. These boundary-adjacent embeddings $\hat{Q}_{um}$ likely contain unknown class information rather than merely representing the background.

Adaptive Subgraph Searching. Although these unmatched query embeddings near the boundary can be directly used to simulate unknown class data, as in mainstream OSOD approaches [14, 15, 35, 54]. In SG-OSOD, however, domain shifts can distort the embeddings of known and unknown classes, which are simulated using source domain data. This distortion can cause these embeddings to overlap with known-class embeddings in the target domain, leading to confusion and resulting in suboptimal detection performance. To this end, `ASGS` employs an Adaptive Subgraph Searching (ASS) algorithm to construct high-order subgraphs adaptively, enabling more robust cross-domain learning for unknown classes.

Specifically, for each $\hat{q}_i \in \hat{Q}_{um}$ and its corresponding boundary known embedding q_i^k, we calculate the cosine similarity. If $\text{sim}(\hat{q}_i, q_i^k)$ exceeds a predefined threshold δ, a connection is established between $\hat{q}_i$ and q_i^k, as illustrated in the bottom right of Fig. 2. When multiple embeddings are connected, we construct subgraphs by connecting each pair of vertices that satisfy this similarity criterion, resulting in the subgraph set $\mathcal{G} = \{g_1, g_2, \ldots, g_s\}$. It is noteworthy that the order of these subgraphs may vary. Since each subgraph contains multiple similar features from both known and unknown samples, these high-order cues facilitate complex multi-sample interactions, enabling the model to learn more robust cross-domain patterns and enhancing its ability to generalize to unknown classes in new environments. Ultimately, we utilize $\mathcal{G}$ to train the unknown classifier,

$$\mathcal{L}_{SUL} = -\frac{1}{|\mathcal{G}|} \sum_{i=1}^{|\mathcal{G}|} \log \left(p \left(f_{cls} \left(\bar{\mathcal{G}} \right) = |C_s| + 1 \mid \bar{\mathcal{G}} \right) \right), \tag{3}$$

where $f_{cls}(\cdot)$ denotes the classifier in detection heads, $\bar{\mathcal{G}}$ represents the mean of all features within a subgraph, serving as an abstract representation of $\mathcal{G}$.

3.3. Class-wise Embedding Compaction (CEC)

Although SUL successfully activates unknown class regions, the embeddings of these unknown classes are distributed around the known categories. Unfortunately, the expansion into previously known regions inevitably harms the detector's performance on known classes. For this reason, we introduce a contrastive learning objective to obtain more compact feature representations for known classes. This compaction reduces the likelihood of misclassifying unknown objects as known classes in novel domains, while simultaneously improving known class detection.

Inspired by the remarkable success of contrastive learning in representation learning, we leverage and adapt the powerful InfoNCE framework [36]. The canonical InfoNCE loss is formulated as:

$$\mathcal{L}^{\text{info}} = -\log \frac{\exp\left(\boldsymbol{v} \cdot \boldsymbol{v}^+ / \tau\right)}{\exp\left(\boldsymbol{v} \cdot \boldsymbol{v}^+ / \tau\right) + \sum_{\boldsymbol{v}^- \in \mathcal{N}_I} \exp\left(\boldsymbol{v} \cdot \boldsymbol{v}^- / \tau\right)}, \tag{4}$$

where $\boldsymbol{v}^+$ and $\boldsymbol{v}^-$ denote the positive and negative samples of $\boldsymbol{v}$, respectively, and τ is a temperature hyper-parameter.

In CEC, we utilize the contrastive loss in a fully supervised way with class labels. Specifically, for a prototype μ_s^k (query) with its category label k, the positive samples $\boldsymbol{c}^+$ are matched query embeddings from the same class Q_m^k (positive keys). In contrast, the negative samples $\boldsymbol{c}^-$ are prototypes and matched query embeddings from other classes (negative keys). Formally, CEC for a prototype from class k is:

$$\mathcal{L}_{\text{CEC}}^k = -\log \frac{\exp\left(\mu_s^k \cdot \boldsymbol{c}^+ / \tau\right)}{\exp\left(\mu_s^k \cdot \boldsymbol{c}^+ / \tau\right) + \sum_{\boldsymbol{c}^- \in \mathcal{N}_k} \exp\left(\mu_s^k \cdot \boldsymbol{c}^- / \tau\right)}, \tag{5}$$

where $\mathcal{N}_k$ encompasses the negative samples for prototype μ_s^k. In effect, L_{CEC} encourages the object embeddings to be aligned with its class prototype and form a compact cluster for each class, while also enabling more robust handling of unknown scenarios.

3.4. Model Optimization

Our `ASGS` is trained end-to-end using the following joint loss formulation:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{SUL} + \lambda_2 \mathcal{L}_{CEC} + \mathcal{L}_{DETR}, \tag{6}$$

where $\mathcal{L}_{SUL}$ is the proposed subgraph-wise unknown-class learning loss, $\mathcal{L}_{CEC}$ is used for class-wise embedding compaction, and $\mathcal{L}_{DETR}$ denotes the D-DETR detection loss. λ_1 and λ_2 are set to 1.0 and 0.1, respectively.

4. Experiments

4.1. Setup

Datasets. Following [28], we conduct experiments on three benchmarks to validate the effectiveness of `ASGS`. For Cityscapes → Foggy Cityscapes/BDD100K, the division of categories into known and unknown groups follows 12 distinct settings, categorized along two dimensions: (1) **Semantic overlap and instance frequency**, which includes 4 sub-tasks: heterogeneous semantics (**het-sem**), homogeneous semantics (**hom-sem**), frequency decrease (**freq-dec**), and frequency increase (**freq-inc**); (2) **Number of unknown classes**, which includes 3 sub-tasks. For Pascal VOC → Clipart1K, we split known and unknown classes solely based on the number of unknown classes. More splitting details are provided in Appendix.

Cityscapes → Foggy Cityscapes. Cityscapes [7] is an urban street scene dataset containing images from 50 cities across Germany, captured at different times of day, and includes 8 object classes. Foggy Cityscapes [39] is a modified version of Cityscapes, synthesized to simulate foggy conditions with a fog intensity level of 0.02.

Pascal VOC → Clipart1K. Pascal VOC [12] is a widely used real-world object detection dataset consisting of 20 object classes, resulting in about 15k images. Clipart1K [17] includes 1K comical images sourced from search engines.

Cityscapes → BDD100K. BDD100K [49] is a large-scale driving video dataset that features images captured in diverse urban environments and driving conditions. We focus on the daytime subset, which includes $36,728$ images for training and $5,258$ images for evaluation.

Evaluation Protocols. We utilize mean Average Precision (mAP_k) to evaluate the known class performance. In line with prior works [9, 19, 28], we adapt Average Recall (AR_u), Wilderness Impact (WI) [9], and Absolute Open-Set Error (AOSE) [9] for unknown class evaluation.

Implementation Details. We employ Deformable DETR and initialize the feature extractor with ResNet-50 [16] pretrained with DINO [51]. Our model is trained with the AdamW optimizer [33], using a batch size of 4 on an

Task	Method	num.unknown-class: 3				num.unknown-class: 4				num.unknown-class: 5			
		mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓	mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓	mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓
het-sem	D-DETR(ICLR'21) [1]	37.69	0.00	0.376	454	36.67	0.00	0.508	899	33.96	0.00	0.598	1248
	OpenDet(CVPR'22) [15]	37.12	2.32	0.371	268	35.54	1.51	0.499	524	33.26	1.98	0.621	1048
	OW-DETR(CVPR'22) [14]	34.32	2.24	0.482	251	35.38	1.98	0.702	513	30.49	2.03	0.813	922
	PROB(CVPR'23) [60]	36.67	1.27	0.195	20	34.86	1.47	0.354	53	33.32	1.22	0.396	84
	CAT(CVPR'23) [34]	33.12	2.12	0.427	202	33.03	2.26	0.581	416	33.76	2.12	0.681	725
	SOMA(ICCV'23) [28]	34.58	1.54	0.327	193	33.50	1.84	0.433	371	31.07	1.64	0.655	719
	ASGS (ours)	**41.09**	**2.87**	**0.314**	**185**	**38.86**	2.71	**0.431**	**309**	**37.51**	2.41	**0.518**	**672**
hom-sem	D-DETR(ICLR'21) [1]	39.59	0.00	3.148	4283	38.88	0.00	2.700	5135	36.46	0.00	3.110	8959
	OpenDet(CVPR'22) [15]	37.17	4.75	2.521	1735	37.80	5.32	2.663	2205	35.24	6.21	3.031	3505
	OW-DETR(CVPR'22) [14]	36.61	3.75	2.533	1736	36.00	3.59	2.680	1880	33.49	4.95	3.262	3473
	PROB(CVPR'23) [60]	35.64	1.01	0.148	14	33.66	1.23	0.269	37	29.18	1.02	0.299	58
	CAT(CVPR'23) [34]	30.90	8.89	3.577	2335	30.50	9.30	3.773	2839	29.01	10.17	4.701	5626
	SOMA(ICCV'23) [28]	40.62	8.76	2.814	2603	39.62	9.66	2.992	3195	37.14	9.89	3.507	5642
	ASGS (ours)	**40.78**	9.29	**2.075**	**1669**	**40.40**	**9.97**	**2.324**	**1148**	**37.51**	10.89	2.949	**3293**
freq-dec	D-DETR(ICLR'21) [1]	52.01	0.00	0.572	1055	50.46	0.00	0.841	1833	49.18	0.00	0.917	2012
	OpenDet(CVPR'22) [15]	52.01	9.16	0.582	570	50.01	10.02	0.801	1179	49.10	10.28	0.898	1248
	OW-DETR(CVPR'22) [14]	51.81	8.06	0.547	575	50.29	9.13	0.792	1074	49.14	9.33	0.870	1219
	PROB(CVPR'23) [60]	40.12	9.33	**0.423**	488	39.38	10.12	0.671	1012	39.76	10.41	0.824	1201
	CAT(CVPR'23) [34]	44.40	6.30	0.448	495	43.52	8.11	0.742	**941**	42.49	8.08	0.844	1168
	SOMA(ICCV'23) [28]	51.03	9.13	0.467	631	49.74	9.85	**0.669**	1078	47.63	10.72	**0.792**	1403
	ASGS (ours)	**52.36**	**9.62**	0.496	**467**	**51.14**	**11.59**	0.699	1046	**49.22**	**11.38**	0.822	**1122**
freq-inc	D-DETR(ICLR'21) [1]	30.68	0.00	2.393	2719	30.89	0.00	2.417	4499	28.06	0.00	2.679	7663
	OpenDet(CVPR'22) [15]	28.17	4.75	2.521	1735	27.80	5.32	2.663	2205	25.24	6.21	3.031	**4105**
	OW-DETR(CVPR'22) [14]	27.28	5.42	3.343	1806	26.92	5.52	3.428	2941	24.25	5.26	3.698	4798
	PROB(CVPR'23) [60]	21.06	**8.64**	5.637	2289	22.38	9.51	5.571	3478	20.76	8.41	5.612	5757
	CAT(CVPR'23) [34]	25.41	7.91	6.137	2571	26.13	8.03	6.491	4092	23.23	7.60	6.543	6052
	SOMA(ICCV'23) [28]	32.97	3.93	2.673	1158	30.08	6.09	3.428	2654	26.49	5.84	3.436	4290
	ASGS (ours)	**34.52**	8.33	**2.350**	**1130**	**31.94**	**10.28**	**1.430**	**2160**	**28.48**	**8.78**	**2.514**	4488

Table 1. Results on Cityscapes → Foggy Cityscapes dataset under 12 different task settings.

NVIDIA A40 GPU. The initial learning rate is set to 0.0002, the weight decay is 5×10^{-4} and the total epoch is set to 65. The D-DETR architecture incorporates three comprehensive transformer encoders and decoders, with the query number N established at 100. The dimension of the embeddings D is defined as 256. We configure the number of known samples near the boundary to K at 5, and set the number of nearest unmatched embeddings to M at 5. The prototype update factor α is set to 0.9, the temperature hyperparameter τ in Eq. 5 is set to 0.1, and the threshold δ is set to 0.6 in the experiments.

4.2. Main Results

Comparison Methods. We compare ASGS with the state-of-the-art methods, including Open-set Detector (**OpenDet**) [15], Open-World Detection Transformer (**OW-DETR**) [14], Probabilistic Objectness transformer-based open-world detector (**PROB**) [60], Localization and identification Cascade Detection Transformer (**CAT**) [34], and Structured Motif Matching (**SOMA**) [28]. Notably, all baseline methods are reproduced and evaluated under the SG-OSOD setting for fair comparison.

Cityscapes → Foggy Cityscapes. The comparison results are reported in Table 1. The proposed ASGS demonstrates superior performance across all evaluation metrics and task settings. For known class detection, ASGS consistently outperforms baseline methods in mAP$_k$ across all 12 settings. Notably, under the het-sem setting with 3 unknown classes, ASGS achieves 41.09% mAP$_k$, surpassing the baseline D-DETR (37.69%) by a significant margin. As the number of unknown classes increases, detection becomes more challenging. While existing methods show improvements in unknown class detection but at the cost of known class performance. For instance, CAT achieves mAP$_k$ scores of 33.12%, 33.03%, and 33.76% under the het-sem setting, all lower than D-DETR. In contrast, ASGS maintains strong performance with mAP$_k$ scores of 41.09%, 38.86%, and 37.51%. In the more challenging hom-sem setting, ASGS achieves remarkable AR$_u$ scores while maintaining competitive mAP$_k$ performance. Although in some scenarios other methods show marginally better performance in individual metrics (e.g., PROB's AR$_u$ of 11.30% vs ASGS's 10.89% in hom-sem with 5 unknown classes), ASGS consistently achieves better overall performance when considering all metrics jointly.

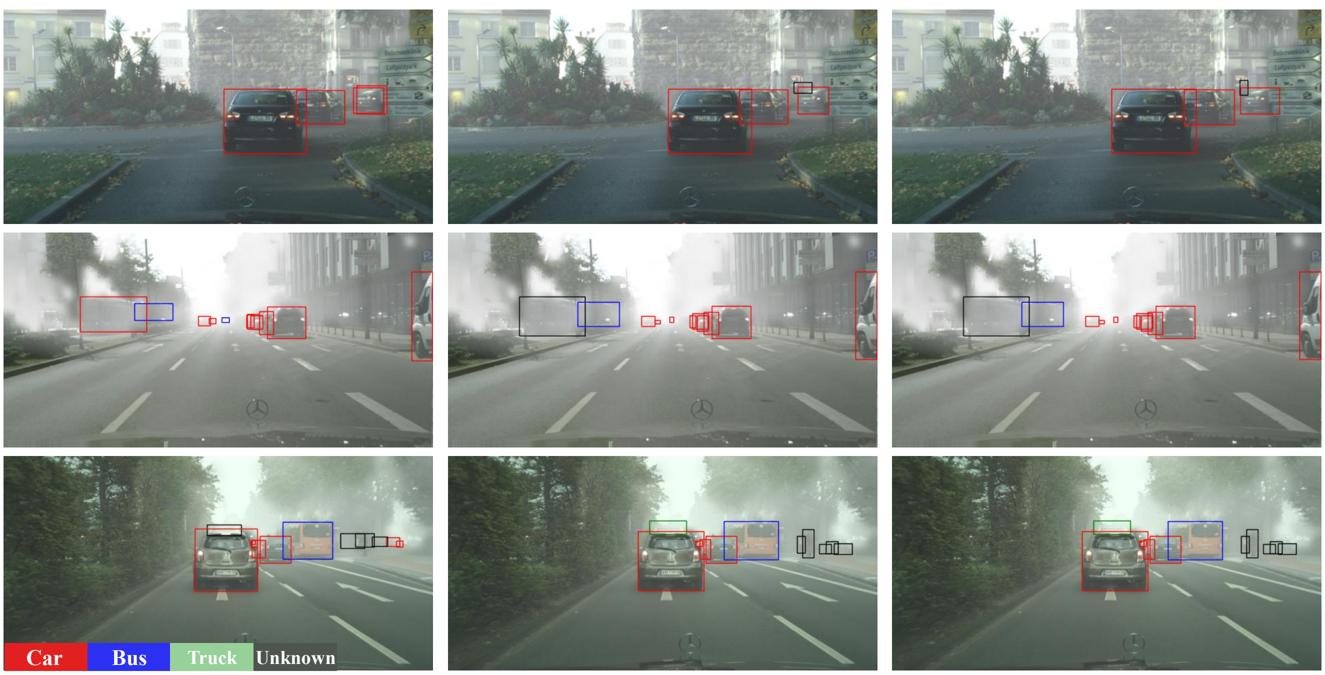
Pascal VOC → Clipart. Table 2 shows that ASGS achieves superior performance with 6, 8, and 10 unknown classes. For known classes, ASGS significantly outperforms D-DETR with mAP$_k$ of 21.66%, 21.95%, 19.06% versus 18.34%, 18.17%, 16.39%. For unknown classes, ASGS

Method	num.unknown-class: 6				num.unknown-class: 8				num.unknown-class: 10			
	mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓	mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓	mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓
D-DETR(ICLR'21) [1]	18.34	0.00	6.057	4567	18.17	0.00	6.459	5379	16.39	0.00	6.893	6402
OpenDet(CVPR'22) [15]	16.57	21.20	6.472	3608	16.02	21.24	7.558	4122	15.88	17.66	7.598	5248
OW-DETR(CVPR'22) [14]	16.67	21.78	6.637	3711	16.40	20.00	7.408	4278	15.71	17.75	7.885	5254
PROB(CVPR'23) [60]	12.12	24.66	6.101	3091	11.51	25.07	6.826	4213	11.76	21.41	6.968	5921
CAT(CVPR'23) [34]	13.06	22.64	6.617	3289	13.38	24.51	6.571	4351	13.87	20.45	7.012	6016
SOMA(ICCV'23) [28]	15.20	**26.64**	7.197	3966	15.04	25.50	7.800	4632	14.48	**24.27**	8.738	6003
ASGS (ours)	**21.66**	25.10	**6.048**	**3061**	**21.95**	**26.57**	**6.435**	**4078**	**19.06**	22.68	**6.863**	**5247**

Table 2. Performance on Pascal VOC → Clipart dataset.

Method	num.unknown-class: 3				num.unknown-class: 4				num.unknown-class: 5			
	mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓	mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓	mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓
D-DETR(ICLR'21) [1]	13.24	0.00	0.133	956	13.27	0.00	0.170	1054	13.29	0.00	0.184	1598
OpenDet(CVPR'22) [15]	13.02	1.10	0.131	712	12.91	1.36	0.197	902	12.92	1.27	0.171	1248
OW-DETR(CVPR'22) [14]	12.77	1.19	0.129	630	12.77	1.21	0.181	700	12.80	1.32	0.162	862
PROB(CVPR'23) [60]	10.71	**1.83**	0.128	609	10.29	1.78	0.166	728	9.82	**1.88**	0.178	801
CAT(CVPR'23) [34]	11.12	1.45	0.135	730	11.38	1.51	0.171	778	11.76	1.41	0.182	812
SOMA(ICCV'23) [28]	8.22	1.22	0.163	764	7.15	0.79	0.169	**512**	7.33	0.88	0.198	736
ASGS (ours)	**13.57**	1.81	**0.126**	**600**	**13.69**	**1.81**	**0.148**	720	**13.62**	**1.88**	**0.161**	**718**

Table 3. Performance on Cityscapes → BDD100K dataset under **het-sem** setting.



(a) OW-DETR **(b) ASGS (ours)** **(c) Ground Truth**

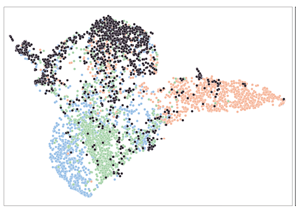
Figure 3. Qualitative comparisons on Cityscapes → Foggy Cityscapes under het-sem setting between (a) OW-DETR and (b) ASGS.

achieves competitive AR$_u$ comparable to SOMA while maintaining the lowest AOSE and WI scores across all settings.

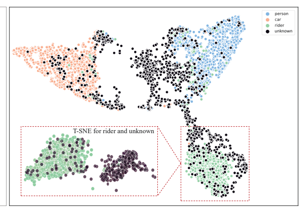
Cityscapes → BDD100K. Table 3 shows the experimental results on Cityscapes → BDD100K under het-sem setting with 3, 4, and 5 unknown classes. ASGS achieves the best performance in most metrics across all settings. Specifically, for known class detection, ASGS obtains the high-

set	mAP$_k$↑	AR$_u$↑	WI↓	AOSE↓
D-DETR	37.69	0.00	0.376	454
w/o. SUL	42.60	0.00	0.292	354
w/o. CEC	36.10	2.63	0.322	201
SUL w/o. ASS	37.31	2.76	0.326	169
ASGS (Full)	41.09	2.87	0.314	185

Table 4. Ablation of ASGS on Cityscapes → Foggy Cityscapes under het-sem and 3 unknown classes

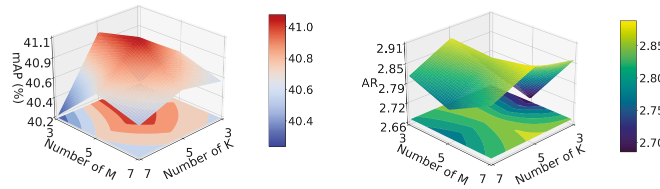


(a) ASGS w/o CEC

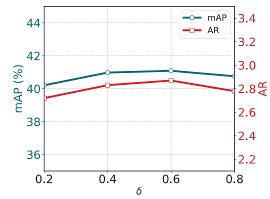


(b) ASGS

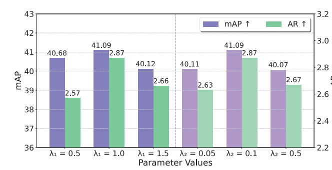
Figure 4. T-SNE feature visualization on the Cityscapes → Foggy Cityscapes under freq-dec setting .



(a) The number of M and K



(b) The threshold δ



(c) The number of λ_1 and λ_2

Figure 5. Ablation of the hyperparameters.

est mAP$_k$ (13.57%, 13.69%, 13.62%), outperforming the second-best method D-DETR. For unknown class detection, ASGS achieves competitive AR$_u$ comparable to PROB.

4.3. Further Empirical Analysis

Qualitative results. Fig. 3 presents visualization results comparing ASGS with OW-DETR [14] on the Cityscapes → Foggy Cityscapes. As can be seen, compared to the existing open-set method OW-DETR, the proposed model effectively and robustly detects unknown samples and achieves superior, accurate bounding box regression performance in both known and unknown classes.

Ablation Study. In Table 4, we evaluate the contribution of the different components of ASGS (i.e., SUL and CEC). When removing SUL, mAP$_k$ improves but a complete inability to detect unknown classes. This confirms SUL's essential role in detecting unknown classes across domains. Without CEC, mAP$_k$ decreases significantly, AR$_u$ also shows a modest decline. This demonstrates that CEC effectively maintains compact known-class representations, which is crucial for robust performance across both known and unknown classes. Without ASS the model's performance on known and unknown classes decreases. This confirms that structural information from adaptive subgraphs provides more domain-robust features than single-sample embeddings alone. Furthermore, as shown in Fig. 4, we visualized the T-SNE plot of the model, which confirms the effectiveness of both modules. Fig. 4a, without CEC shows less distinct separation between classes, with more overlap between different colored points. In contrast, Fig. 4b demonstrates that the full model achieves a clearer separation between known and unknown classes.

Hyper-parameter Analysis. Fig. 5 shows the impact of different hyperparameters. All experiments are conducted on Cityscapes → Foggy Cityscapes under het-sem and 3 unknown classes. Analysis of hyperparameters K and M is shown in Fig. 5a. The model achieves optimal performance mAP$_k$ when selecting K and M at 5. For unknown class, while AR$_u$ reaches its peak with K at 5 and M at 7, the quantitative difference between M values of 5 and 7 is negligible. Therefore, considering mAP$_k$ and AR$_u$ in a comprehensive way, we select the K and M values of 5 as our final settings. Fig. 5b reveals optimal performance with δ at 0.6, balancing meaningful connections while avoiding spurious relationships, with performance remaining stable across a wide range of values, demonstrating our method's robustness. As shown in Fig. 5c, we evaluate the impact of different weight values, which highlight the effectiveness of setting λ_1 at 1.0 and λ_2 at 0.1 for achieving an optimal balance between known and unknown class.

5. Conclusion

We explore the problem of Single-domain Generalizable Open-Set Object Detection (SG-OSOD), a practical yet Unexplored area. To tackle this challenge, we propose ASGS, a novel framework that effectively captures and identifies unknown objects while maintaining robust performance across domain variations. ASGS achieves this through two complementary components, SUL leverages high-order structural relationships between known and unknown samples to create domain-robust representations, while CEC enforces discriminative and compact feature representations that maintain their integrity across domain shifts. This complementary design enables ASGS to effectively identify unknown objects while preserving strong performance on known categories. Experimental results demonstrate that ASGS's effectiveness in handling SG-OSOD.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant 82172033, Grant U19B2031, Grant 61971369, Grant 52105126, Grant 82272071, and Grant 62271430; and in part by the Dreams Foundation of Jianghuai Advance Technology Center; and in part by the Open Fund of the National Key Laboratory of Infrared Detection Technologies.

References

[1] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 3, 6, 7

[2] 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*, pages 8869–8878, 2020. 1

[3] 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*, pages 2703–2712, 2021.

[4] Chaoqi Chen, Zebiao Zheng, Yue Huang, Xinghao Ding, and Yizhou Yu. I3net: Implicit instance-invariant network for adapting one-stage object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12576–12585, 2021.

[5] Chaoqi Chen, Jiongcheng Li, Hong-Yu Zhou, Xiaoguang Han, Yue Huang, Xinghao Ding, and Yizhou Yu. Relation matters: Foreground-aware graph-based relational reasoning for domain adaptive object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3677–3694, 2022.

[6] 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*, pages 3339–3348, 2018. 1

[7] 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*, pages 3213–3223, 2016. 5

[8] 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*, pages 17732–17742, 2024. 2

[9] Akshay Dhamija, Manuel Gunther, Jonathan Ventura, and Terrance Boult. The overlooked elephant of object detection: Open set. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1021–1030, 2020. 3, 5

[10] Xuefeng Du, Gabriel Gozum, Yifei Ming, and Yixuan Li. Siren: Shaping representations for detecting out-of-distribution objects. *Advances in Neural Information Processing Systems*, 35:20434–20449, 2022. 2, 3

[11] Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. Vos: Learning what you don't know by virtual outlier synthesis. *arXiv preprint arXiv:2202.01197*, 2022. 3

[12] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010. 5

[13] Dario Fontanel, Matteo Tarantino, Fabio Cermelli, and Barbara Caputo. Detecting the unknown in object detection. *arXiv preprint arXiv:2208.11641*, 2022. 2, 3

[14] Akshita Gupta, Sanath Narayan, KJ 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*, pages 9235–9244, 2022. 3, 4, 6, 7, 8

[15] 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*, pages 9591–9600, 2022. 2, 3, 4, 6, 7

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[17] 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*, pages 5001–5009, 2018. 5

[18] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547, 2019. 4

[19] KJ 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*, pages 5830–5840, 2021. 5

[20] Kai Katsumata, Ikki Kishida, Ayako Amma, and Hideki Nakayama. Open-set domain generalization via metric learning. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 459–463. IEEE, 2021. 3

[21] Benjamin Kiefer, Martin Messmer, and Andreas Zell. Diminishing domain bias by leveraging domain labels in object detection on uavs. In *2021 20th International Conference on Advanced Robotics (ICAR)*, pages 523–530. IEEE, 2021. 2

[22] 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*, pages 2947–2955, 2024. 2

[23] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI conference on artificial intelligence*, 2018. 2

[24] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M Hospedales. Episodic training for domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1446–1455, 2019.

[25] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5400–5409, 2018. 2

[26] Junnan Li, Caiming Xiong, and Steven CH Hoi. Mopro: Webly supervised learning with momentum prototypes. *arXiv preprint arXiv:2009.07995*, 2020. 4

[27] 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*, pages 5291–5300, 2022. 1

[28] 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*, pages 15780–15790, 2023. 5, 6, 7

[29] Ya Li, Xinmei Tian, Mingming Gong, Yajing Liu, Tongliang Liu, Kun Zhang, and Dacheng Tao. Deep domain generalization via conditional invariant adversarial networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 624–639, 2018. 2

[30] Yu-Jhe Li, Xiaoliang Dai, Chih-Yao Ma, Yen-Cheng Liu, Kan Chen, Bichen Wu, Zijian He, Kris Kitani, and Peter Vajda. Cross-domain adaptive teacher for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7581–7590, 2022. 1

[31] 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*, pages 8771–8780, 2021. 2

[32] Hong Liu, Pinhao Song, and Runwei Ding. Towards domain generalization in underwater object detection. In *2020 IEEE international conference on image processing (ICIP)*, pages 1971–1975. IEEE, 2020. 2

[33] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5

[34] 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*, pages 19681–19690, 2023. 3, 6, 7

[35] Dimity Miller, Lachlan Nicholson, Feras Dayoub, and Niko Sünderhauf. Dropout sampling for robust object detection in open-set conditions. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3243–3249. IEEE, 2018. 2, 3, 4

[36] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 5

[37] Vihari Piratla, Praneeth Netrapalli, and Sunita Sarawagi. Efficient domain generalization via common-specific low-rank decomposition. In *International Conference on Machine Learning*, pages 7728–7738. PMLR, 2020. 2

[38] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6956–6965, 2019. 1

[39] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126:973–992, 2018. 5

[40] Yang Shu, Zhangjie Cao, Chenyu Wang, Jianmin Wang, and Mingsheng Long. Open domain generalization with domain-augmented meta-learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9624–9633, 2021. 2

[41] Leitian Tao, Xuefeng Du, Xiaojin Zhu, and Yixuan Li. Non-parametric outlier synthesis. *arXiv preprint arXiv:2303.02966*, 2023. 2

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 3

[43] 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*, pages 3219–3229, 2023. 2

[44] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and S Yu Philip. Generalizing to unseen domains: A survey on domain generalization. *IEEE transactions on knowledge and data engineering*, 35(8):8052–8072, 2022. 2

[45] Kunyu Wang, Xueyang Fu, Yukun Huang, Chengzhi Cao, Gege Shi, and Zheng-Jun Zha. Generalized uav object detection via frequency domain disentanglement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1064–1073, 2023. 2

[46] 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*, pages 847–856, 2022. 2

[47] Fan Wu, Jinling Gao, Lanqing Hong, Xinbing Wang, Chenghu Zhou, and Nanyang Ye. G-nas: Generalizable neural architecture search for single domain generalization object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5958–5966, 2024. 2

[48] Shiqi Yang, Yaxing Wang, Kai Wang, SHANGLING JUI, and Joost van de Weijer. One ring to bring them all: Model adaptation under domain and category shift. 2022. 3

[49] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous

multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020. 5

[50] Jinan Yu, Liyan Ma, Zhenglin Li, Yan Peng, and Shaorong Xie. Open-world object detection via discriminative class prototype learning. *arXiv preprint arXiv:2302.11757*, 2023. 3

[51] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022. 5

[52] Hanlin Zhang, Yi-Fan Zhang, Weiyang Liu, Adrian Weller, Bernhard Schölkopf, and Eric P Xing. Towards principled disentanglement for domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8024–8034, 2022. 2

[53] Xiaowei Zhao, Yuqing Ma, Duorui Wang, Yifan Shen, Yixuan Qiao, and Xianglong Liu. Revisiting open world object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. 2, 3

[54] Jiyang Zheng, Weihao Li, Jie Hong, Lars Petersson, and Nick Barnes. Towards open-set object detection and discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3961–3970, 2022. 2, 3, 4

[55] Yangtao Zheng, Di Huang, Songtao Liu, and Yunhong Wang. Cross-domain object detection through coarse-to-fine feature adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13766–13775, 2020. 1, 3

[56] Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Learning placeholders for open-set recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2021. 3

[57] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4396–4415, 2022. 2

[58] Ronghang Zhu and Sheng Li. Crossmatch: Cross-classifier consistency regularization for open-set single domain generalization. In *International conference on learning representations*, 2022. 3

[59] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020. 3

[60] 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*, pages 11444–11453, 2023. 3, 6, 7