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

Supplementary Material

This document provides more details of our approach and additional experimental results.

1. Implementation Details

Cityscapes → Foggy Cityscapes/BDD100K. The division between known and unknown classes, as outlined in Table 1, follows four distinct protocols that consider both semantic overlap and instance frequency. These protocols are designed based on the following considerations: 1) In real-world contexts, the semantics of unknown classes may either overlap with or differ from known classes. Recent research in open-set object detection (OSOD) [2, 3, 7] tends to focus on cases of minimal overlap, failing to consider the full spectrum of potential real-world conditions. To address this gap, we evaluate both heterogeneous (non-overlapping) and homogeneous (overlapping) semantics within the SG-OSOD framework, offering a more comprehensive assessment. 2) Instance scale diversity: In real-world environments, the distribution of unknown-class objects varies significantly, with some classes being more prevalent while others are relatively rare. This important aspect is often overlooked in the existing OSOD research. To address this, we rank the classes based on the number of objects and select the three most frequent and three least frequent classes as the known classes, thereby better representing the diversity seen in real-world scenarios.

For evaluation, we follow prior OSOD works [3, 5] by introducing sub-tasks involving 3, 4, and 5 unknown classes. To ensure the accuracy of the evaluation, we exclude images containing objects from unknown classes not included in a particular sub-task.

Pascal VOC → Clipart1k. We adopt the unknown class splitting methodology from [4], where the first 10 classes, listed in alphabetical order, are designated as known classes. The remaining classes are then grouped into unknown-class sets Ω_n with sizes $|\Omega_n| \in \{6, 8, 10\}$, yielding three distinct sub-tasks. To ensure a fair evaluation for each sub-task, we exclude images containing unknown-class objects that do not belong to Ω_n. This step prevents unknown-class objects, correctly detected, from being misclassified as false positives.

2. Experimental Results

2.1. Comparison Methods

For Cityscapes → BDD100K, we present experimental results across different settings (hom-sem, freq-dec and freq-inc) in Table 2. We observe that CAT achieves mAP$_k$ scores of 6.05% (9.92%), 6.08% (10.87%), and 8.65% (11.04%) under the hom-sem setting, all of which are lower than those of D-DETR. In contrast, the proposed ASGS demonstrates substantial improvements, with mAP$_k$ scores of 10.41%, 12.69%, and 12.23%, significantly outperforming the baseline methods. Although in some settings, our AR$_u$ is slightly lower than PROB (e.g., in the hom-sem setting with 3 unknown classes, PROB's AR$_u$ exceeds ASGS 's by 0.31%), our other metrics have shown significant improvements, particularly mAP$_k$, which increased by 4.52%. These results indicate that our model effectively addresses the SG-OSOD setting, achieving a well-balanced detection performance for both known and unknown classes.

Table 1. Detailed class splitting settings for Cityscapes → Foggy Cityscapes.

Settings	Known Classes	Unknown Classes
het-sem	car, truck, bus	person, motor, train person, motor, train, bicycle person, motor, train, bicycle, rider
hom-sem	person, bicycle, bus	car, truck, train car, truck, train, motor car, truck, train, motor, rider
freq-dec	person, car, rider	bicycle, train, truck bicycle, train, truck, motor bicycle, train, truck, motor, bus
freq-inc	motor, truck, bus	person, train, car person, train, car, bicycle person, train, car, bicycle, rider

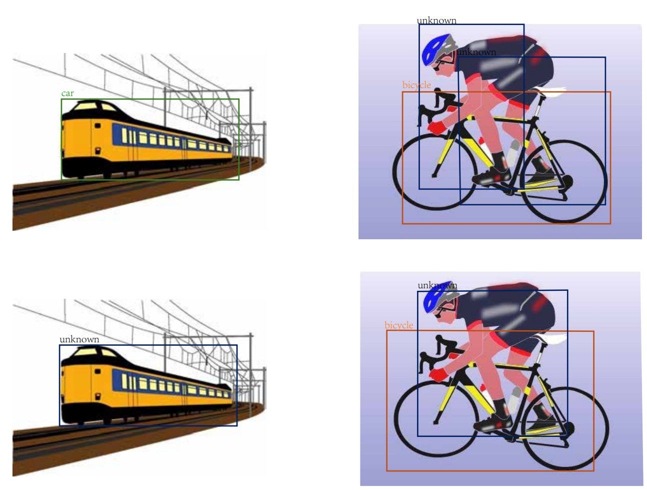

Figure 1. Qualitative comparisons on Pascal VOC → Clipart between OW-DETR (top) and ASGS (bottom).

Table 2. Results on Cityscapes → BDD100k dataset under different task settings.

Task	Method	num.unknown-class: 3				num.unknown-class: 4				num.unknown-class: 5			
		$\text{mAP}_k\uparrow$	$\text{AR}_u\uparrow$	WI↓	AOSE↓	$\text{mAP}_k\uparrow$	$\text{AR}_u\uparrow$	WI↓	AOSE↓	$\text{mAP}_k\uparrow$	$\text{AR}_u\uparrow$	WI↓	AOSE↓
hom-sem	D-DETR(ICLR'21) [1]	9.92	0.00	4.723	26612	10.87	0.00	3.869	22031	11.04	0.00	3.992	30074
	OpenDet(CVPR'22) [3]	8.81	4.11	4.214	18829	12.48	4.51	4.490	17921	11.35	4.29	4.241	24690
	OW-DETR(CVPR'22) [2]	8.62	4.45	4.441	17602	11.20	4.47	4.467	18134	11.49	4.63	4.723	20742
	PROB(CVPR'23) [8]	5.89	**7.12**	3.768	15661	5.53	6.44	3.800	15861	9.82	6.47	4.007	17349
	CAT(CVPR'23) [6]	6.05	5.77	3.997	15692	6.08	5.81	4.021	16170	8.65	5.88	4.229	18450
	ASGS (ours)	**10.41**	6.81	**3.326**	**10281**	**12.69**	**6.75**	**3.512**	**15576**	**12.23**	**6.91**	**3.613**	**16878**
freq-dec	D-DETR(ICLR'21) [1]	14.43	0.00	0.855	7090	14.44	0.00	0.904	7846	14.47	0.00	1.240	12888
	OpenDet(CVPR'22) [3]	14.81	10.66	0.946	4276	15.03	9.91	1.120	4978	15.02	12.11	1.286	7918
	OW-DETR(CVPR'22) [2]	15.10	10.29	0.979	4236	14.97	9.86	1.040	4706	14.98	9.98	1.390	7640
	PROB(CVPR'23) [8]	13.07	**11.89**	0.721	3403	11.95	**14.32**	0.987	4488	14.01	**14.62**	1.082	6987
	CAT(CVPR'23) [6]	13.97	10.96	0.742	3302	13.07	12.13	1.021	4012	14.66	13.31	1.129	7013
	ASGS (ours)	**15.28**	11.23	**0.705**	**3231**	**16.60**	13.17	0.706	**3709**	**16.21**	13.88	**1.010**	**6677**
freq-inc	D-DETR(ICLR'21) [1]	10.85	0.00	3.264	29484	10.87	0.00	3.28	30266	11.11	0.00	3.416	33910
	OpenDet(CVPR'22) [3]	10.47	3.42	5.591	21392	10.42	3.60	5.981	21768	10.58	3.57	6.023	25501
	OW-DETR(CVPR'22) [2]	10.00	3.65	5.412	19732	9.99	3.65	5.469	20632	10.03	3.62	5.521	22430
	PROB(CVPR'23) [8]	6.98	8.12	**3.380**	16129	7.01	**7.12**	4.012	17099	6.58	7.42	4.128	18794
	CAT(CVPR'23) [6]	8.16	5.07	3.804	15810	8.26	5.11	**3.836**	16306	8.76	5.17	4.025	**18612**
	ASGS (ours)	**11.07**	**8.21**	3.421	**15538**	**11.37**	6.47	3.938	**16038**	**12.27**	**7.61**	**4.013**	19864

2.2. Qualitative Comparison

The visualization results in Figure 1 demonstrate a clear performance advantage of ASGS over OW-DETR [2] on the Pascal VOC → Clipart datasets. Our proposed model exhibits two key strengths: effective detection of unknown samples and superior bounding box regression accuracy for both known and unknown classes.

Further qualitative comparisons between (a) OW-DETR and (b) ASGS are presented in Figures 2,3 and4. These figures highlight ASGS's consistent performance across scenarios with varying numbers of unknown classes. In each case, our model maintains robust unknown sample detection while delivering precise bounding box regression for all object categories.

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. 2

[2] 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. 1, 2

[3] 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. 1, 2

[4] 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. 1

[5] 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. 1

[6] 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. 2

[7] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need? 2021. 1

[8] 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. 2

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

Figure 2. Qualitative comparison on Cityscapes → Foggy Cityscapes under het-sem setting between (a) OW-DETR and (b) `ASGS`, the number of unknown classes is 3. The red, blue, green, and black boxes represent the car, bus, truck, and unknown classes.

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

Figure 3. Qualitative comparison on Cityscapes → Foggy Cityscapes under het-sem setting between (a) OW-DETR and (b) ASGS, the number of unknown classes is 4. The red, blue, green, and black boxes represent the car, bus, truck, and unknown classes.

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

Figure 4. Qualitative comparison on Cityscapes $\rightarrow$ Foggy Cityscapes under het-sem setting between (a) OW-DETR and (b) `ASGS`, the number of unknown classes is 5. The red, blue, green, and black boxes represent the car, bus, truck, and unknown classes.