Supplementary Material for Unknown Sniffer for Object Detection: Don't Turn a Blind Eye to Unknown Objects

In this supplementary material, we provide additional details that were not included in the main manuscript due to the page limitations. These details include more ablation studies, qualitative results, failure cases, discussions about Wilderness Impact, as well as the applicability of GBD.

A. More Ablation studies

The results of the ablation studies for the GOC losses are presented in Table 1. It is evident that removing either L_{neg} or L_{pos} leads to a significant drop in performance, indicating that these losses are essential. And L_{con} contributes to further improve the accuracy. As shown in Table 2, a parameter sensitivity experiment was carried out for the sampling parameter T used by the negative energy suppression. Our model achieves the best result when setting T to 100.

L_{neg}	L_{pos}	L_{con}	U-AP	U-F1	U-PRE	U-REC
✓	×	×	0.090	0.001	1.000	0.001
×	✓	×	-	-	-	-
×	×	✓	-	-	-	-
✓	✓	×	0.390	0.456	0.423	0.494
×	✓	✓	0.005	0.018	0.009	0.333
✓	×	✓	-	-	-	-
✓	✓	✓	0.454	0.479	0.433	0.535

Table 1. Ablation study of GOC losses.

T	5	10	50	100	150	200
U-F1	0.460	0.462	0.466	0.479	0.455	0.457

Table 2. Parameter sensitivity experiments of T.

B. More Qualitative Results

To further demonstrate the effectiveness of the contributions, Fig. 1 shows additional results of different methods on several example images of the COCO-Mix (first three rows) and COCO-OOD dataset (last four rows). The 2^{nd} and 4^{th}-7^{th} rows contain only a small number of objects. In these images, the previous methods generate a lot of bounding box predictions, most of which are false positives. In contrast, our method avoids this issue by using graph-based box determination to eliminate redundant boxes. Furthermore, the umbrellas in the 3^{rd} image and the mouse in the 5^{th} image demonstrate that UnSniffer captures unknown objects more effectively than other methods. The ability is due to the clear distinction between objects and non-objects by the generalized object confidence score. In addition, our method also distinguishes the unknown and known objects better than other methods, with the help of negative energy suppression, which is demonstrated by the mattress in the 1^{st} image and the toy in the 6^{th} image.

C. Failure Cases

In this paper, the generalized object detector in UnSniffer utilizes the offset value given by the box regression head to refine the bounding boxes. Since the box regression head lacks the supervision of unknown objects, the unknown predictions slightly differ from the ground truth. For example, in the 2^{nd} image of Fig. 1, the unknown prediction does not entirely cover the fork. Additionally, in the 4^{th} image, the left bounding box predicted by UnSniffer only partially covers the zebra on the left. We hope that our work, along with its identified shortcomings, inspires other researchers to further advance this field.

D. More Discussions about Wilderness Impact

According to [2], TP_c and FP_c denote the true-positive number and false-positive number of objects with known classes, respectively. FP_o denotes the unknown objects which is misclassified as a known class. And Wilderness Impact (WI) can be formulated as follows:

$$\left(\frac{TP_c}{TP_c+FP_c}\Big/\frac{TP_c}{TP_c+FP_c+FP_o}\right)-1=\frac{FP_o}{TP_c+FP_c} \quad (1)$$

Referring to Table 2 in the main text, compared to ORE [6], UnSniffer obtains a higher Absolute Open-Set Error (AOSE) but a lower WI in the quantitative experiment. The reason is that UnSniffer obtains a higher mAP than ORE, resulting in a lower FP_c. Thus, we argue that WI is sensitive

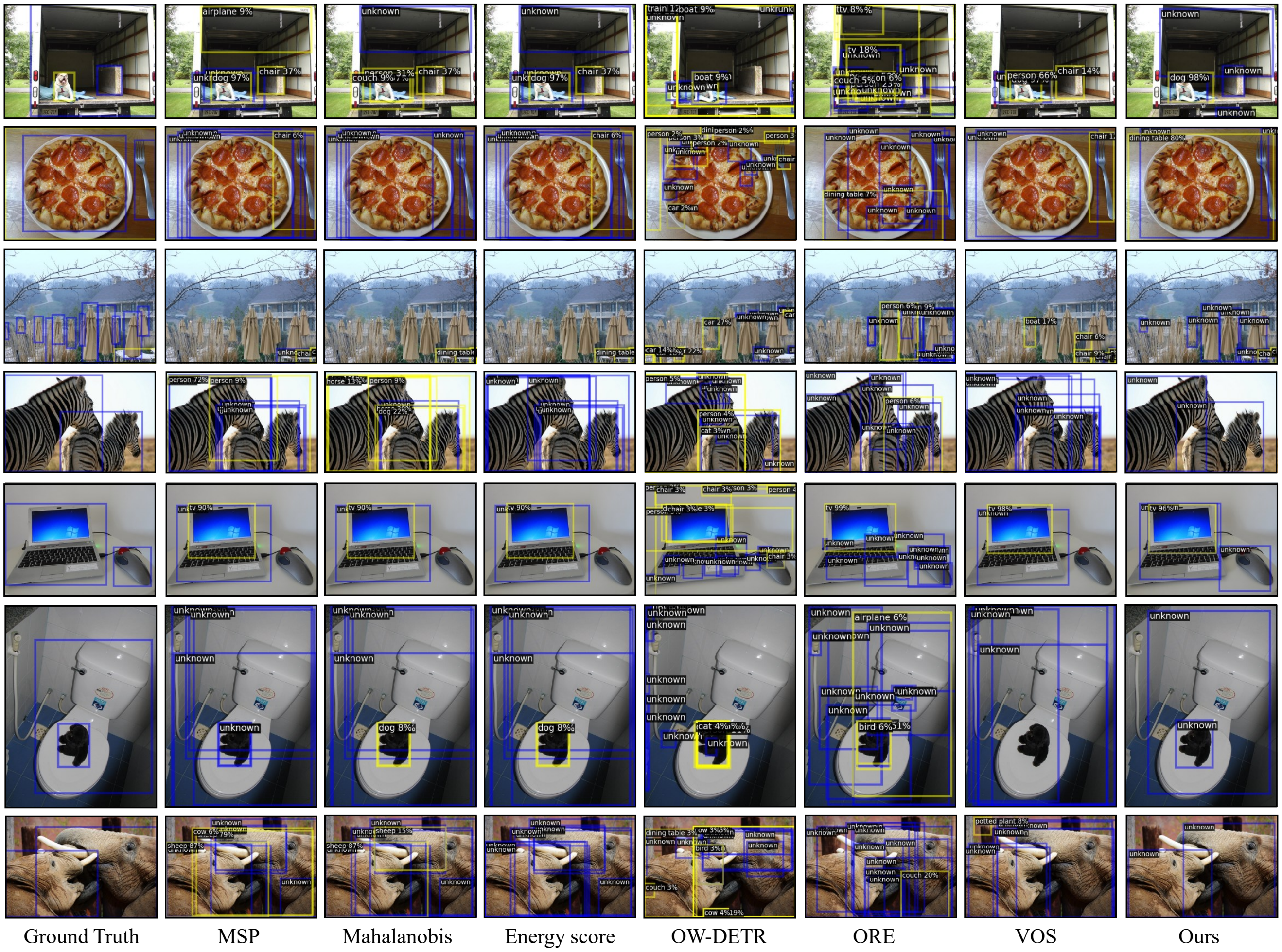

Figure 1. **Example results on COCO-Mix (first three rows) and COCO-OOD datasets (last four rows)**. 1^{st} column: ground truth; 2^{nd}-8^{th} columns: visualization results of MSP [5], Mahalanobis [1], Energy score [7], OW-DETR [4], ORE [6], VOS [3] (with threshold computed on COCO-OOD dataset), and our method. The detections are overlaid on the known (yellow) and unknown (blue) class objects. Since ORE and OW-DETR generate too many results, we only draw the top-10 boxes for each image. And other methods draw all predicted boxes.

Post-processing methods	Number of input	10	20	50	100	150	200
GBD	FPS	9.59	9.28	8.19	6.42	3.77	1.97
	U-F1	0.467	0.491	0.508	0.466	0.409	0.353
NMS	FPS	10.39	10.34	10.35	10.35	10.22	10.13
	U-F1	0.472	0.494	0.468	0.357	0.270	0.208

Table 3. FPS and U-F1 gained by UnSniffer using graph-based box determination (GBD) or non maximum suppression (NMS).

to the number of predictions made for known objects and is advantageous for the models that generate more prediction in an image.

E. Applicability of GBD

Intuitively, the graph partitioning process is usually time-consuming. Indeed, when using the normalized cut algorithm implemented by the Python package 'scikit-image', our method achieves 1.04 frames per second (FPS). To speed up the post-processing unit, we re-implement the normalized cut algorithm by 'PyTorch' and accelerate it using CUDA. The optimized speed of the entire network is 3.62 FPS on the platform of GTX 1080Ti GPU. More specifically, Table 3 compares the performance of the models using GBD and NMS on the COCO-OOD dataset when the number of input proposals for GBD is fixed. It can be observed that the time cost does not significantly increase when the number of input boxes of GBD is less than 50. When the number of input boxes exceeds 100, the inference speed of UnSniffer drops greatly. Intuitively, GBD is more sensitive to speed than accuracy, and achieves a better F1-Score in more cases.

References

[1] Taylor Denouden, Rick Salay, Krzysztof Czarnecki, Vahdat Abdelzad, Buu Phan, and Sachin Vernekar. Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance. *ArXiv*, abs/1812.02765, 2018. 2

[2] Akshay Raj Dhamija, Manuel Günther, Jonathan Ventura, and Terrance E. Boult. The overlooked elephant of object detection: Open set. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020. 1

[3] Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. Vos: Learning what you don't know by virtual outlier synthesis. In *International Conference on Learning Representations (ICLR)*, 2022. 2

[4] Akshita Gupta, Sanath Narayan, KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. OW-DETR: Open-world detection transformer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[5] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *ArXiv*, abs/1610.02136, 2016. 2

[6] KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1, 2

[7] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems (NIPS)*, 2020. 2