Supplementary Materials: PS-TTL: Prototype-based Soft-labels and Test-Time Learning for Few-shot Object Detection

Anonymous Authors

This supplementary material provides more experiment results on PS-TTL in Sec. A and more visualization results in Sec. B.

A MORE EXPERIMENT RESULTS

A.1 One Batch vs. One Epoch

We implement test-time learning by fine-tuning on the test data for one epoch, followed by testing on the same test data. Hence, we require a specific storage capacity to accommodate the test data. In a more realistic scenario, we need to conduct test-time learning for sequentially streamed test data. When a batch of testing samples arrives, we first use the model to make predictions. And then, the model weight are updated on this batch of testing samples. We compared two testing strategies, one batch and one epoch, on the novel split 1 of the PASCAL VOC benchmark, as shown in Table A. By comparing row 1 and row 2, we found that under the One Batch testing strategy, test-time learning can still bring stable improvements to the baseline. Although the performance of the One Batch strategy is generally inferior to that of the One Epoch strategy, we believe this is because in the early stages of testing, the model does not learn enough knowledge from the test data.

A.2 The performance trend of Test-Time Learning

As shown in Fig. A, we plotted the performance trend of the model under different shot settings on the novel split 1 of the PASCAL VOC dataset as training iterations progressed. As expected, with the increase of training iterations, the performance of the model improves progressively. This demonstrates the effectiveness of test-time learning. Through test-time learning, we endow the model with the ability to continuously learn. By learning on the test data, the model can better utilize the novel instances in the test data to capture the data distribution of novel classes.

A.3 Results on MS COCO Under Low-shot Settings

Table B shows the detection results on MS COCO under low-shot settings. The MS COCO dataset contains 80 categories, with each image typically containing multiple instances. This leads to a degradation in the performance of FSOD detectors on the test data, especially in low-shot settings, which hinders the ability to conduct test-time learning. However, our method consistently improves across various low-shot settings, especially in extremely low-sample scenarios, demonstrating notable enhancements. For example, in the 1-shot scenario, our method improved the mAP on novel classes by 9% compared to DeFRCN.

B MORE VISUALIZATION RESULTS

We visualize more detection results of 1-shot on PASCAL VOC in Fig. B. Our method can alleviate the misclassification issue between base and novel classes, as shown in the top group of Fig. B. DeFRCN misclassifies base class dogs and horses as novel class cows, and misclassifies novel class motorbikes as base class bottles, and novel class buses as base class trains. Although our method is not optimized for the regression branch, as more novel class instances are observed, our method can improve the regression performance of novel classes, as shown in the first two columns of the bottom group in Fig. B. Additionally, in the construction of the base data for FSOD, there are many unlabeled novel instances in the base data. This may result in some novel instances being misclassified as background. Our method continuously learns on the test data, which helps alleviate this issue. As shown in the column 3 and column 4 of the bottom group in Fig. B, our method can prevent the omission of novel classes, such as cows and birds.

Table A: Ablation study of the test strategy.

Test Strategy	nAP50		
	1-shot	2-shot	3-shot
DeFRCN	55.4	62.1	65.0
One Batch	56.4	64.0	66.4
One Epoch	58.4	65.7	67.9

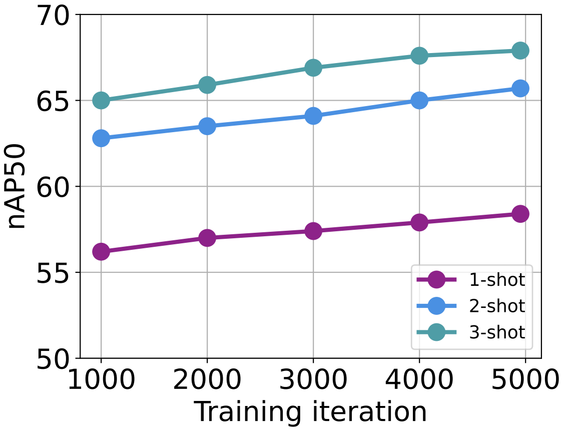

Figure A: The performance trend (nAP50) of Test-Time Learning on PASCAL VOC.

REFERENCES

[1] Yuhang Cao, Jiaqi Wang, Ying Jin, Tong Wu, Kai Chen, Ziwei Liu, and Dahua Lin. 2021. Few-shot object detection via association and discrimination. *Advances in neural information processing systems* 34 (2021), 16570–16581.

[2] Guangxing Han, Yicheng He, Shiyuan Huang, Jiawei Ma, and Shih-Fu Chang. 2021. Query adaptive few-shot object detection with heterogeneous graph convolutional networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3263–3272.

Table B: Few-shot object detection performance on MS COCO.

Method	1-shot		2-shot		3-shot		5-shot	
	nAP	nAP75	nAP	nAP75	nAP	nAP75	nAP	nAP75
TFA [4]	3.4	3.8	4.6	4.8	6.6	6.5	8.3	8.0
MPSR [5]	2.3	2.3	3.5	3.4	5.2	5.1	6.7	6.4
FADI [1]	5.7	6.0	7.0	7.0	8.6	8.3	10.1	9.7
QA-FewDet [2]	4.9	4.4	7.6	6.2	8.4	7.3	9.7	8.6
DeFRCN* [3]	5.5	5.7	9.8	9.9	12.3	**12.6**	14.2	13.7
Ours	**6.0**	**6.5**	**10.1**	**10.3**	**12.5**	12.5	**14.4**	**13.8**

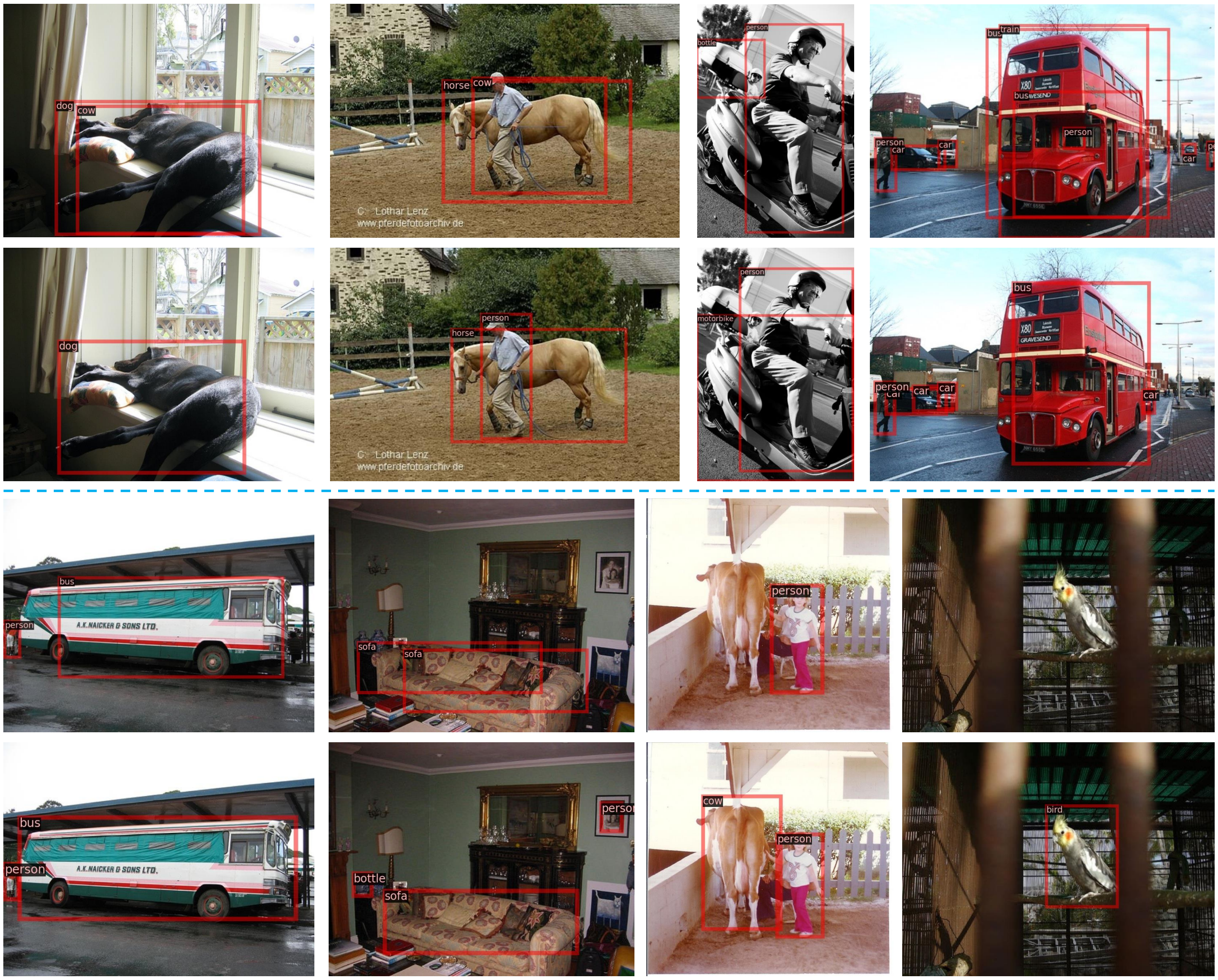

Figure B: Qualitative visualization comparison on PASCAL VOC. The top and bottom lines in each group respectively show the detection results from DeFRCN and our PS-TTL.

[3] Limeng Qiao, Yuxuan Zhao, Zhiyuan Li, Xi Qiu, Jianan Wu, and Chi Zhang. 2021. Defrcn: Decoupled faster r-cnn for few-shot object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 8681–8690.

[4] Xin Wang, Thomas E Huang, Trevor Darrell, Joseph E Gonzalez, and Fisher Yu. 2020. Frustratingly simple few-shot object detection. In *Proceedings of the 37th International Conference on Machine Learning*. 9919–9928.

[5] Jiaxi Wu, Songtao Liu, Di Huang, and Yunhong Wang. 2020. Multi-scale positive sample refinement for few-shot object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16*. Springer, 456–472.