Co-Student: Collaborating Strong and Weak Students for Sparsely Annotated Object Detection(Supplemental Material)

Lianjun Wu[1], Jiangxiao Han[1], Zengqiang Zheng[2], and Xinggang Wang[1,✉]

[1] School of EIC, Huazhong University of Science and Technology
[2] Wuhan Jingce Electronic Group Co., Ltd.

1 More Main Results

1.1 Comparison with Calibrated-Teacher [5]

We trained our method *Co-Student* on the sparsely annotated MS COCO [2] dataset, `COCO-miss50p`, using FCOS [4] (ResNet-101) and multi-scale training, exactly the same as Calibrated-Teacher. The results are shown in Tab. 1. We achieved a 0.6% higher AP than Calibrated-Teacher. This demonstrates that our method, equipped with two collaborated students and a denoising teacher model, achieves higher performance in the SAOD task. At the same time, Calibrated-Teacher and our method are orthogonal. The former focuses on optimizing the teacher model, while the latter emphasizes both denoising using the teacher model and fully leveraging pseudo-labels from student networks. In principle, the two methods are complementary.

1.2 Results in Pascal-VOC [1]

To ensure more fair comparisons and provide additional reference experimental results for future work, we have additionally presented the performance of our method on the sparsely annotated version of the PASCAL-VOC 2007 dataset, i.e., 50% of annotations are missing, in Tab.2. Experiments on the PASCAL-VOC 2007 dataset were conducted using 5011 trainval images, and performance was computed on 4952 images of the test set. We use the standard PASCAL-VOC style AP50, which is the Average Precision computed at an IoU threshold of 0.5. We use the PASCAL-VOC 2007 train set for training and the PASCAL-VOC 2007 test set for evaluation. We sample using exactly the same hyperparameter settings as those used in MS COCO [2].

2 More Ablation Studies

2.1 Effect of Confidence Threshold θ_t

The most important hyper-parameter in our denoising teacher model is the confidence threshold θ_t used for generating pseudo-labels. This threshold directly

✉ Corresponding author: `xgwang@hust.edu.cn`

Table 1: The main results in `COCO-miss50p` setting. The methods marked with $\star$ are trained on fully annotated `COCO-train2017` set. The results marked with † are inherited from their publications.

Method	AP	$\mathbf{AP}_{50}$	$\mathbf{AP}_{75}$	$\mathbf{AP}_{s}$	$\mathbf{AP}_{m}$	$\mathbf{AP}_{l}$
FCOS* [4]	42.5†	-	-	-	-	-
FCOS [4]	37.9†	-	-	-	-	-
Calibrated-Teacher [5]	39.0†	-	-	-	-	-
Co-Student	**39.6**	**58.0**	**42.8**	**23.8**	**43.7**	**51.3**

Table 2: Results in sparsely annotated VOC. The results marked with † are inherited from their publications.

Method	AP	$\mathbf{AP}_{50}$	$\mathbf{AP}_{75}$
Co-mining [6]	47.3	72.0	50.2
SparseDet [3]	-	74.4†	-
Co-Student	**50.3**	**75.8**	**54.5**

Table 3: The ablation results on the confidence threshold θ_t of denoising teacher model.

θ_t	0.45	0.50	0.55	0.60	0.65
AP	19.8	36.9	36.8	**37.0**	36.2

affects the robustness of the teacher network based on EMA to noise. Thresholds that are too low or too high can both result in a decrease in the denoising capability of the teacher network. Results are shown in Tab. 3. When we adjust this threshold, we have observed that the performance of our method remains relatively stable, ranging between 0.5 and 0.6.

2.2 Effect of Data Augmentation in *Co-Student*

To determine whether strong data augmentation itself plays the primary role in the Co-Student framework, or if it is the way we employ the collaborated students with strong and weak augmentation that predominantly influences the outcome, we sequentially replace the strong&weak augmentation with weak&weak augmentation and strong&strong augmentation. The results are shown in Tab. 4. The results indicate that simply introducing strong data augmentation does not lead to significant improvements in the SAOD task. Compared to using only weak augmentation, incorporating strong augmentation alone resulted in only a 0.2% increase. Instead, our proposed *Co-Student* utilize strong and weak augmentation collaboratively to excavate more pseudo-annotations, which are crucial in SAOD.

3 More Qualitative Results

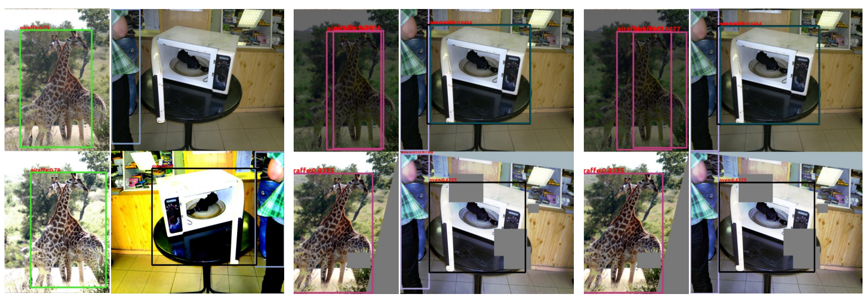
We show additional qualitative cues in Fig. 2 to demonstrate that the *Co-Student* with strong and weak augmentations can excavate more unlabeled objects, and the pseudo-labels generated by our methods are more reliable and accurate.

Table 4: The results of ablation experiments on *Co-Student* using different combinations of data augmentation.

Aug.	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Strong&Weak	**37.0**	**55.2**	**39.9**	**21.4**	**40.4**	**47.9**
Weak&Weak	35.9	53.5	38.4	19.8	39.2	46.7
Strong&Strong	36.1	54.5	39.0	20.9	39.5	46.6

4 Failure Cases

We also show some failure cases in Fig. 1 for both Co-mining and *Co-Student*. They are all affected by the noise in the generated pseudo-labels, and it appears that the denoising teacher model in our approach fails in cases involving occlusion and easily confused objects.



(a) (b) (c)

Fig. 1: Failure cases in both methods. (a) The pseudo-labels generated by Co-mining; (b) The pseudo-labels generated by *Co-Student*; (c) The refined pseudo-labels denoised by the teacher model.

References

1. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision **88**(2), 303–338 (2010)
2. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
3. Suri, S., Rambhatla, S., Chellappa, R., Shrivastava, A.: Sparsedet: Improving sparsely annotated object detection with pseudo-positive mining. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6770–6781 (2023)
4. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9627–9636 (2019)

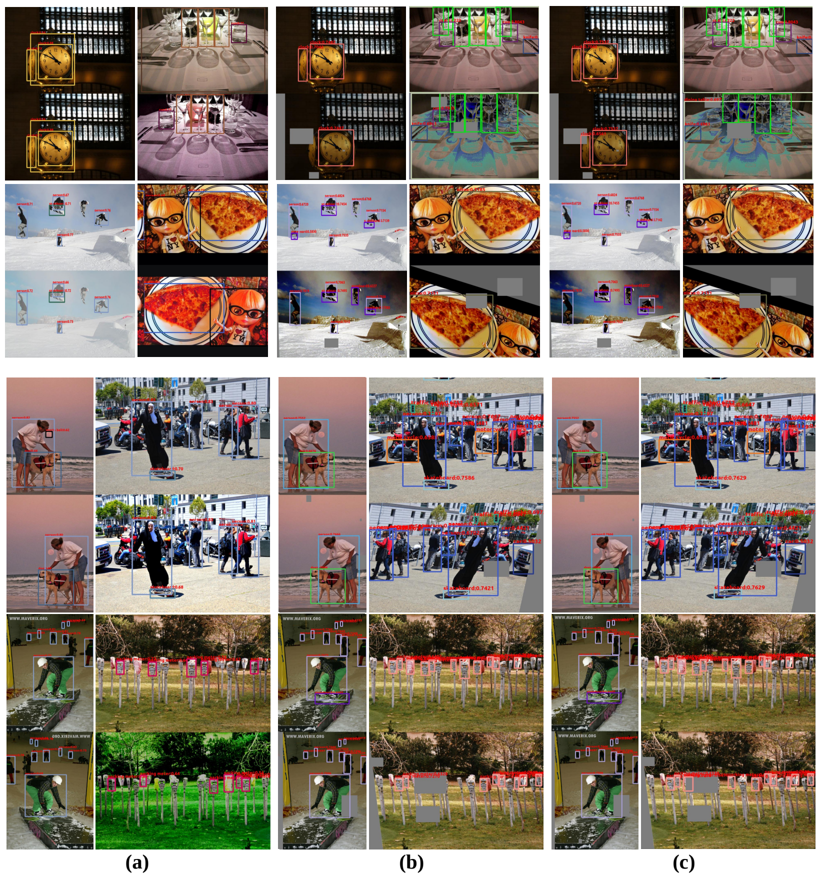


Fig. 2: More qualitative results showing the qualities of the pseudo-labels generated by different methods. (a) The pseudo-labels generated by Co-mining; (b) The pseudo-labels generated by *Co-Student*; (c) The refined pseudo-labels denoised by the teacher model.

5. Wang, H., Liu, L., Zhang, B., Zhang, J., Zhang, W., Gan, Z., Wang, Y., Wang, C., Wang, H.: Calibrated teacher for sparsely annotated object detection (2023)
6. Wang, T., Yang, T., Cao, J., Zhang, X.: Co-mining: Self-supervised learning for sparsely annotated object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 2800–2808 (2021)