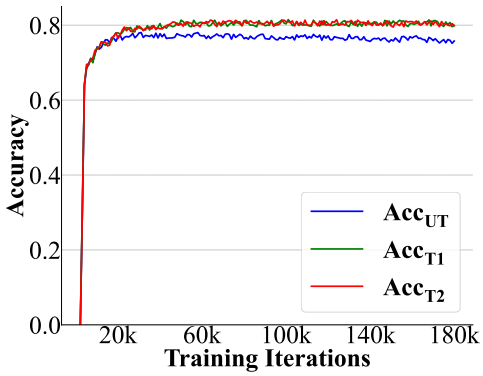
Supplementary Material: Cycle Self-Training for Semi-Supervised Object Detection with Distribution Consistency Reweighting

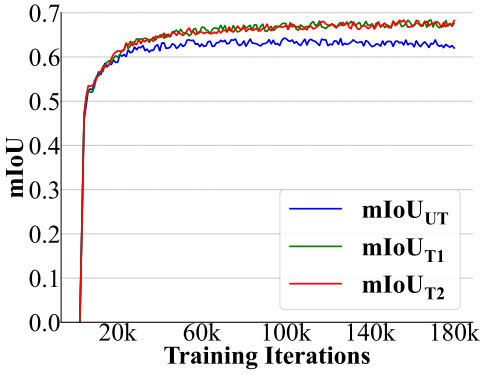
Paper ID: 1213

A ACCURACY AND MIOU COMPARISON

According to the general quality evaluation of the pseudo-labels, we adopt two metrics to assess the performance of different frameworks: the *Accuracy* and the *mIoU*. A comparison is made between the baseline Unbiased Teacher and our CST as shown in Figure 1 (a)(b). From the figures we can observe that our CST can generate more precise pseudo-labels compared to the baseline. Besides, the gaps of the *Accuracy* and the *mIoU* between the baseline and our CST are becoming larger during training, and the *Accuracy* and the *mIoU* of Unbiased Teacher begin to drop after the early stage of training, while those of our CST still keep increasing and tend to be stable. Consequently, the coupling effect and the confirmation biases are alleviated by the effective and robust CST model.



(a)



(b)

Figure 1: (a) The Accuracy of pseudo-labels. (b) The mIoU of pseudo-labels. "UB" means the baseline Unbiased Teacher. "T1" ("T2") denotes the evaluation performed on the pseudo-labels generated by the teacher T1 (T2).

B WEIGHTS COMPARISON BETWEEN TWO TEACHERS

Our CST framework looses the coupling effect compared to the traditional Teacher-Student framework. We have demonstrated the fact by given the euclidean distance of weights between one student model and another teacher model in the main paper. To further verify that there are adequate differences between the knowledge learned by the two teachers, we also give the weights distance between the two teachers T1 and T2. The results are shown in Figure 2. The euclidean distance of the weights W_{T1-T2} between the teacher T1 and the teacher T2 keeps far away as the weights distance W_{T1-S2} and W_{T2-S1}. This reveals that the two teachers have more different meaningful knowledge, which confirms the effectiveness in overcoming the coupling effect of our CST framework.

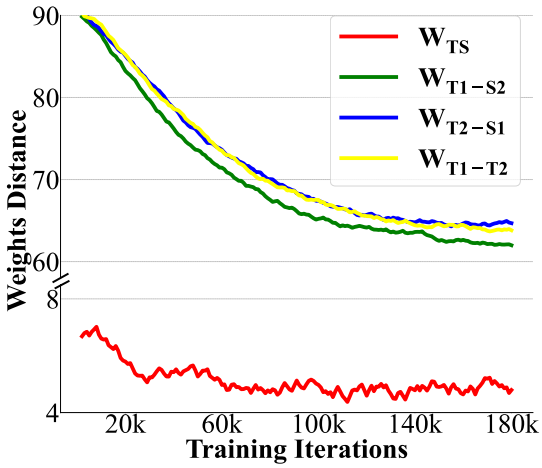


Figure 2: The euclidean distance of weights. An additional curve in yellow is added to represent the weight distance between the teachers T1 and T2.

C ANALYSIS ON THE RESULTS OF PASCAL VOC DATASET

Our CST framework has achieved a comparable evaluation result on PASCAL VOC [1] dataset referred in the Experiments section in the main paper. Nevertheless, only one distinctive method Soft Teacher [4] preforms better than our proposed CST on PASCAL VOC because of less images and categories. Soft Teacher performs better when there are rich labeled data or more simple instances with less categories. Coincidentally, there are only 20 categories and almost 5k images with 2.3 instances per image on average in PASCAL VOC 2007 and PASCAL VOC 2012 dataset, while MS-COCO [2] dataset has 80 categories in total with 7.7 instances per

image on average. Hence it is easy to cause the over-fitting problem in the training process of CST, which is also reflected in our practical experiments. We give the AP_{50} evaluation metric on the PASCAL VOC dataset as an example shown in Figure 3, it is obvious that the performance achieves the best performance by our CST framework in the early training stage after the Burn-In Stage, and then degrades gracefully without gaining more benefits. Although we can decay the learning rate to slow down and produce more gains during training, a learning rate in Unbiased Teacher [3] is still adopted for a fair and plausible comparison. With the same hyper-parameters, the results on PASCAL VOC dataset solidly present the superiority of the proposed CST compared to the baseline.

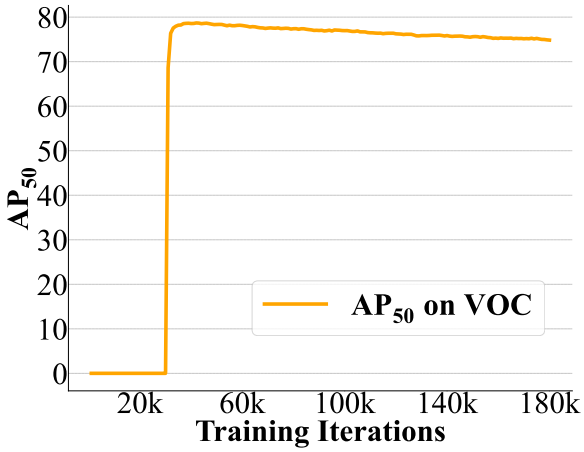


Figure 3: The AP_{50} evaluation result on PASCAL VOC dataset by our CST* framework.

D ADDITIONAL QUALITATIVE RESULTS

For lack of space, we give a few qualitative results of different methods in the main paper to make a comparison. To show the generality of our conclusions, more general qualitative results are visualized as shown in Figure 4. These additional qualitative results further emphasize the superior detection performance of our CST and CST* framework, including the larger percentage of the correct classification and the precise localization with more detected objects, compared to the prior supervised and the Unbiased Teacher methods.

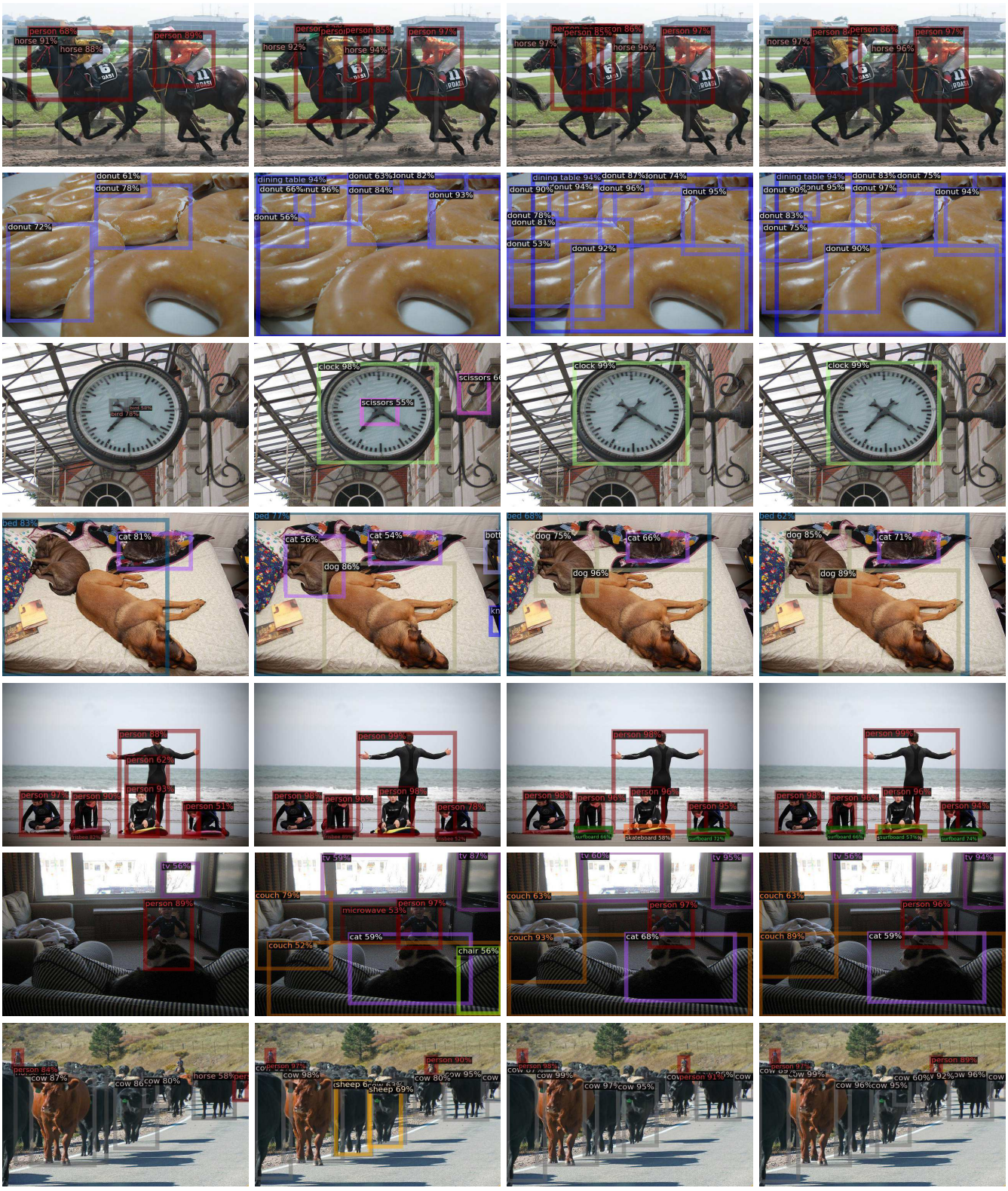
REFERENCES

[1] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88, 2 (2010), 303–338.

[2] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[3] Yen-Cheng Liu, Chih-Yao Ma, Zijian He, Chia-Wen Kuo, Kan Chen, Peizhao Zhang, Bichen Wu, Zsolt Kira, and Peter Vajda. 2021. Unbiased Teacher for Semi-Supervised Object Detection. In *International Conference on Learning Representations*.

[4] Mengde Xu, Zheng Zhang, Han Hu, Jianfeng Wang, Lijuan Wang, Fangyun Wei, Xiang Bai, and Zicheng Liu. 2021. End-to-end semi-supervised object detection with soft teacher. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3060–3069.



(a) Supervised (b) Unbiased Teacher (c) Our CST (d) Our CST*

Figure 4: Additional qualitative results of different methods. (a) Supervised. (b) Unbiased Teacher. (c) Our CST. (d) Our CST*.