Supplementary Material for "Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision"

Xiaokang Chen[1*] Yuhui Yuan[2] Gang Zeng[1] Jingdong Wang[2]
[1]Key Laboratory of Machine Perception (MOE), Peking University [2]Microsoft Research Asia

1. More Implementation Details

Training details. The crop size for PASCAL VOC 2012 and Cityscapes are 512×512 and 800×800, respectively. For the multi-scale data augmentation, we randomly select scale from $\{0.5, 0.75, 1, 1.25, 1.5, 1.75\}$. For Cityscapes dataset, we use OHEM loss as the supervision loss ($\mathcal{L}_s$), and cross entropy loss as the cross pseudo supervision loss ($\mathcal{L}_{cps}$).

Training strategy. We use the similar training strategy as GCT [1] for semi-supervised segmentation. In the supervised baseline for all the partition protocols, we use the batch size 8. We ensure that the iteration number is the same as semi-supervised methods[1]. For semi-supervised methods, at each iteration, we sample additional 8 unlabeled samples. Our method and all the other semi-supervised methods in Table 1 and Table 2 of the main paper follow the same training strategy.

2. Network Perturbation

Our cross pseudo supervision approach (CPS) includes two perturbed segmentation networks, $f(\theta_1)$ and $f(\theta_2)$, which are of the same architecture and initialized differently. In the main paper, we pointed out that the pseudo segmentation results from the two networks are perturbed.

We empirically show the perturbation using the overlap ratio between them during training. The overlap ratio on the labeled set, the unlabeled set and the whole set are given in Figure 1. We can see that (1) the overlap ratio is small at the early training stage and (2) increases during the later training stage. The small overlap ratio at the early stage helps avoid the case the segmentation network converges towards a wrong direction. The large overlap ratio at the later stage implies that the pseudo segmentation results of the two segmentation networks are more accurate.

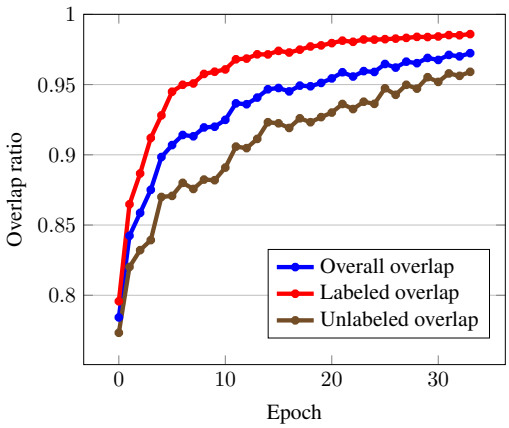

Figure 1: **Prediction overlap of the two networks on PASCAL VOC 2012 under the $1/8$ partition**. We use DeepLabv3+ with ResNet-50 as the segmentation network. We only calculate the overlap ratio in the object region, and the pixels belong to the 'background' class are ignored.

References

[1] Zhanghan Ke, Di Qiu, Kaican Li, Qiong Yan, and Rynson WH Lau. Guided collaborative training for pixel-wise semi-supervised learning. In *ECCV*, 2020. 1

*This work was done when Xiaokang Chen was an intern at Microsoft Research, Beijing, P.R. China

[1]In GCT [1], the supervised baseline uses the batch size 16, and the number of iterations is much smaller than half of the number of iterations in the semi-supervised methods. Therefore, their supervised baseline results are worse than ours (we sure the same number of iterations and each iteration has the same number, 8, of labeled samples).