ATSO: Asynchronous Teacher-Student Optimization for Semi-Supervised Image Segmentation
Appendix

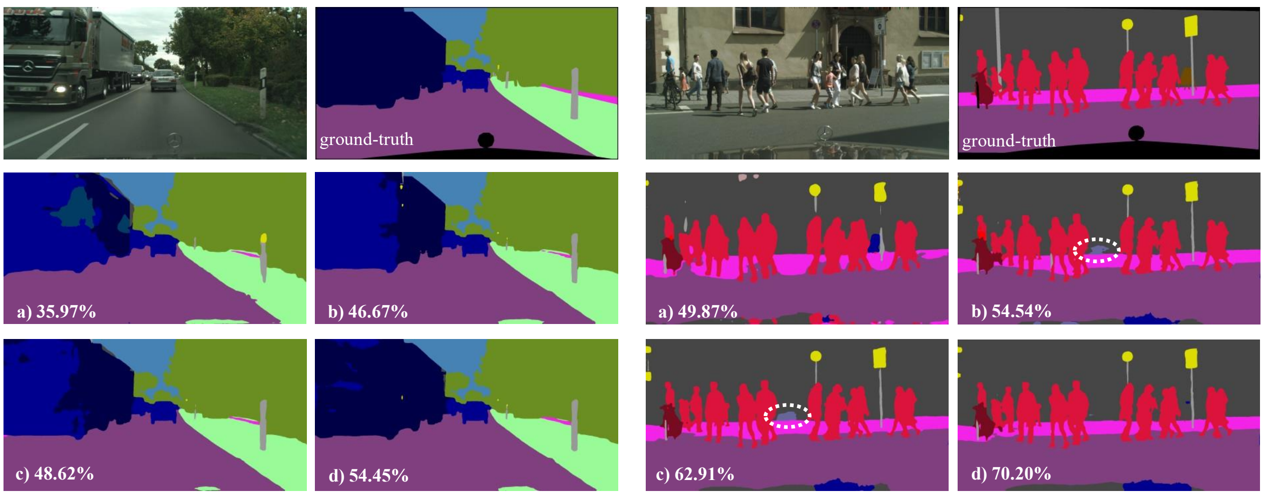

Figure 3: Visualization of the segmentation results of the experiments on the Cityscapes datasets, including the supervised only on the labeled part, self-learning, STSO and ATSO. The four settings are signed by letter a, b, c, d respectively and compared with the mIOU results. We point an obvious error appeared in the self-learning and STSO with the white dashed circles.

1. Visualization on the *Cityscapes* dataset

For a more straightforward comparison among the three learning strategies, *e.g.*, ATSO, STSO and the self-learning, we provide two visualization examples in figure 3. For some hard cases where baseline and STSO would fail to distinguish similar categories due to the limited label size in training set, ATSO could recognize them successfully. For example the truck in the figure 3 is attributed to the car in the model trained only on the labeled set, while the accuracy improved step by step as our methods adding to the training process. In another view, with the addition of the reference set, the quality of the segmentation is obviously improved, but due to the inaccurate of pseudo labels in the reference set, some additional errors may occur that do not appear in the supervised experiment. In the second example of figure 3, both the self-learning baseline and STSO mistakenly segment the wall inside the dashed circle, but ATSO avoids this problem.

2. Transfer from *Cityscapes* to *Mapillary*

First, we report the self-learning baseline for the transfer learning experiments from Cityscapes to Mapillary, which reports an accuracy of 26.97% (averaged over all classes). This is done without using the pre-defined mapping to reduce the number of classes from 19 to 5. Note that this number is lower than the corresponding numbers of STSO (28.11%) and ATSO (28.26%).

To show how our algorithm improves segmentation, we provide some typical examples in the transfer learning task from Cityscapes to Mapillary. We find that the direct transfer results are often below satisfaction, but the semi-supervised learning approaches can improve domain transfer performance dramatically. In particular, ATSO works best among all the solutions. In the first and third example, we find that ATSO$_{5\to19}$ is more stable at producing good results for the large objects. This is mainly because the learning process is stabilized by the pseudo labels gen-

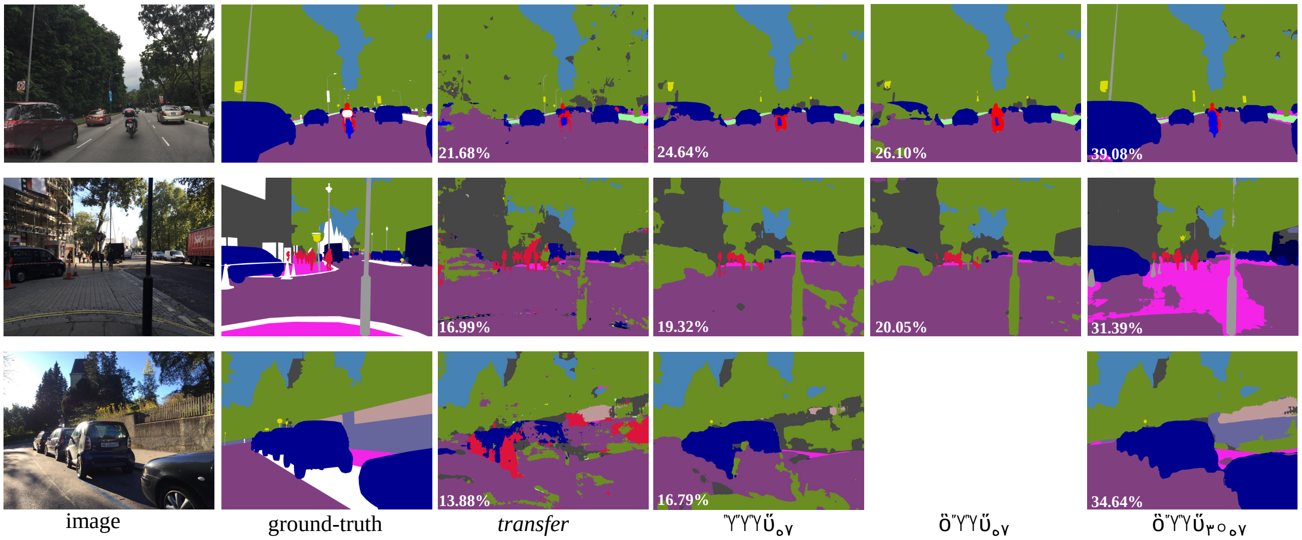

Figure 4: Visualization of the results produced by the direct transfer baseline, STSO, ATSO, and tuning ATSO on the 19-class pseudo label, respectively. Especially, some classes with a small area (*e.g.*, *traffic light*), are improved significantly in $\text{ATSO}_{5\rightarrow19}$. The number at the lower-left corner of each image indicates the IOU averaged over all classes – although pixel-wise accuracy of ATSO seems good, this number can be impacted by some poorly segmented classes.

erated on the reduced 5 classes. On the other hand, when we tune the model by generating the 19-class pseudo label, the performance gain is often more significant on small objects. So, the proposed flowchart (starting from 5 classes and then tuning on 19 classes) is verified effective.

3. Experiments on the PASCAL VOC dataset

We also evaluate ATSO on another natural segmentation dataset PASCAl VOC which only includes $1,462$ training images. Following the setting in previous workings [20], we augment it to $10,582$ training images with SEMANTIC BOUNDARIES. We used the DeepLab V2 network based on ResNet-101 and pre-trained on ImageNet in our experiment. We train the model for 40 epochs in each generation. Other training details and data augmentation operations are same to those in Cityscapes experiments. We follow the standard protocol that 1% of training data is labeled and the remainder is unlabeled. Under this setting, ATSO, STSO, and the self-learning baseline report 45.6%, 41.6%, and 38.2% mIOU, respectively, showing the same trend as other datasets.

4. Add strong data augmentation to ATSO

From our motivation, ATSO obstacles the error from being propagated from teacher to student, while stronger data augmentation alleviates over-fitting. Both of them can be understood as regularization towards reducing the inductive bias, and they can be combined. We add cutout as a stronger augmentation to the Cityscapes experiments using $1/8$ of labeled data, and unsurprisingly, both STSO and ATSO get improved from 60.7%, 61.8% to 61.7%, 62.4%, and ATSO still enjoys an accuracy gain.