Appendix of DUDA: Online-Offline Dual Domain Adaption for Semantic Segmentation

BMVC 2022 Submission # 585

1 Introdution of Online-CBST

This section will introduce the Online-CBST algorithm which is presented by MFA [1] in detail. CBST [2] is a classic class-balanced pseudo-label selecting strategy. It sorts all pixels according to the softmax confidence and selects the top r of the pixels based on the confidence value as pseudo label. The select ratio r and the pseudo labels are updated by stage. However, the CBST strategy is designed for offline pseudo label selection and is not suitable in online pseudo label setting. Therefore, MFA presents the Online-CBST algorithm for batch-level. It applies the CBST algorithm on the batch data and the select ratio r is updated by iteration. The formulation for updating the select ratio r can be written as:

$$r = r_0 + (\frac{n_{cur}}{n_{max}})^p (r_{max} - r_0) \tag{1}$$

where r_0, r_{max}, n_{max} is the predefined initial select ratio, max select ratio and max train iteration respectively. n_{cur} is the current train iteration. p is the hyperparameter to control the growth of r. In our experiments, r_0 and r_{max} is set as 0.2 and 0.75 respectively for online pseudo label part. While for offline pseudo label updating, r_0 and r_{max} is set as 0.4 and 0.95 respectively. Besides, p is set as 1.1. The Algorithm. 1 shows more details for Online-CBST.

Algorithm 1: Online-CBST Algorithm

```
# Online-CBST algorithm
def get_threshold(logits, r, n, num_classes):
    # logits:  model output
    # r:  select ratio, n:  current iteration
    # num_classes:  the number of classes in this dataset
    scores = softmax(logits)
    pred_prob = max(scores, dim=1)
    pred_label = argmax(scores, dim=1)
    threshold = []
    for c in range(num_classes)
        x = pred_prob[pred_label == c]
        cls_num = len(x)
        if cls_num == 0:  continue
        threshold.apend(np.percentile(x, (1 - valid_thresh) * 100))
    r = update(r, n) # update the select ratio according to Eq.1
    return threshold, r
```

© 2022. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

	components			mIoU
WarmUp	FDA Warmup			46.2
	ST-MT	sce loss	F2W operation	mIoU
Baseline	✓			53.9
	✓	✓		54.8
	✓	✓	✓	55.0
	high-confidence acceptance	inconsistent discard	consistent acceptance	mIoU
TOPLU	✓			56.5
	✓	✓		56.8
	✓	✓	✓	56.9

Table 1: Ablation study of each component for more detail.

2 Supplemental Experiments

Ablation studies of our baseline. To further study more components in our DUDA, this section will bring more ablation studies for GTA5-to-Cityscapes benchmark. Firstly, we analyse the components that build our baseline. The "Baseline" item in Table. 1 shows the effectiveness of mean-teacher framework with consistency loss, sce loss and full-to-weak operation. FDA is used as a warmup method firstly and achieves 46.2 mIoU. Then the naive self-training and mean-teacher framework with consistency loss achieve 53.9 mIoU. The noisy robust sce loss further brings 0.9 mIoU improvement and we achieve 54.8 mIoU. Finally. we add full-to-weak operation, and the model arrives at 55.0 mIoU.

Ablation studies of TOPLU. Secondly, we analyse every step in our purposed TOPLU algorithm. The "TOPLU" item in Table. 1 shows the effectiveness of all three steps in our purpose TOPLU algorithm. We can observe that the high confidence acceptance step improves the performance most. We believe this is because the performance improvement during training corrects some mislabelled pixels in the initial pseudo label. Our designed inconsistency discard step and consistency accept step also bring some performance improvement. Especially, the inconsistency discard step brings 0.3 mIoU gain based on the first step, it proves the temporal inconsistency information is helpful to the selection of noise labels.

3 Results Visualization

In this section, more results will be visualized. Figure. 1 shows the predicted result of the valid set for GTA5-to-Cityscapes benchmark. For SYNTHIA-to-Cityscapes benchmark, our DUDA also performs well. Figure. 2 provides a visualization of the updated results of offline pseudo labels. Figure. 3 shows the predicted results of the valid set for SYNTHIA-to-Cityscapes benchmark. We can observe that our DUDA produce visible improvement comparing with the FDA warmup model. In addition, the effect of noise label correction is also very obvious.

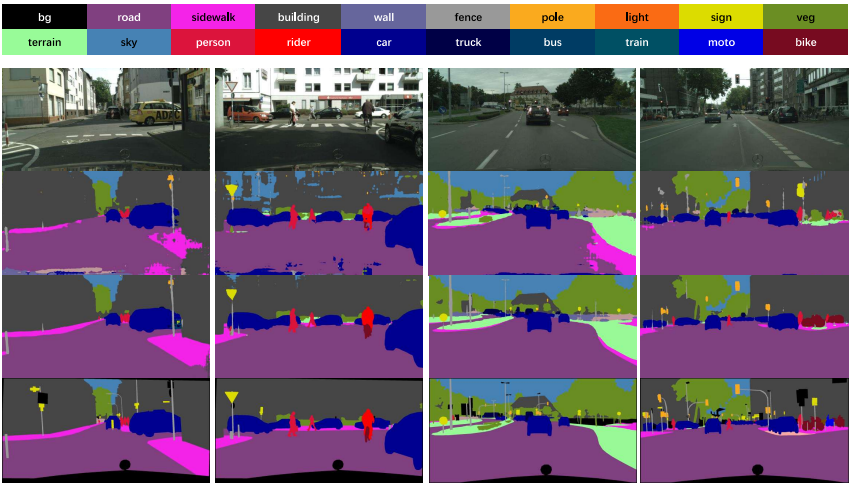

Figure 1: The predicted results of our DUDA for GTA5-to-Cityscapes benchmark. The first row is the valid images, the second and third rows is the FDA warmup results and our DUDA results. The last row is the ground truth.

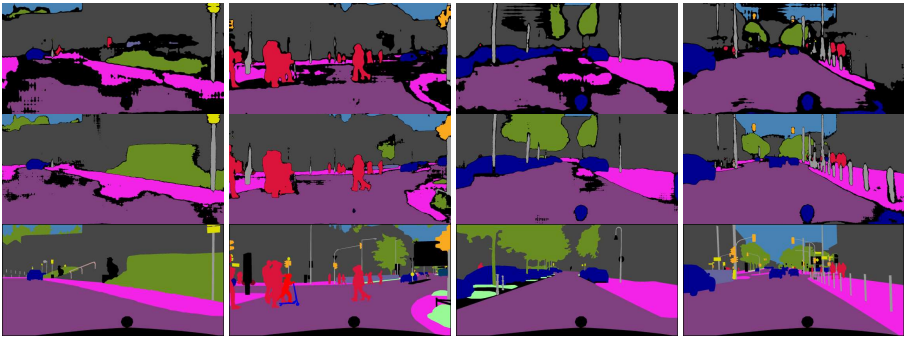

Figure 2: The offline pseudo label update results for SYNTHIA-to-Cityscapes benchmark. The first row is the initial pseudo label generated by CBST. The second and third rows is the updated results and ground truth.

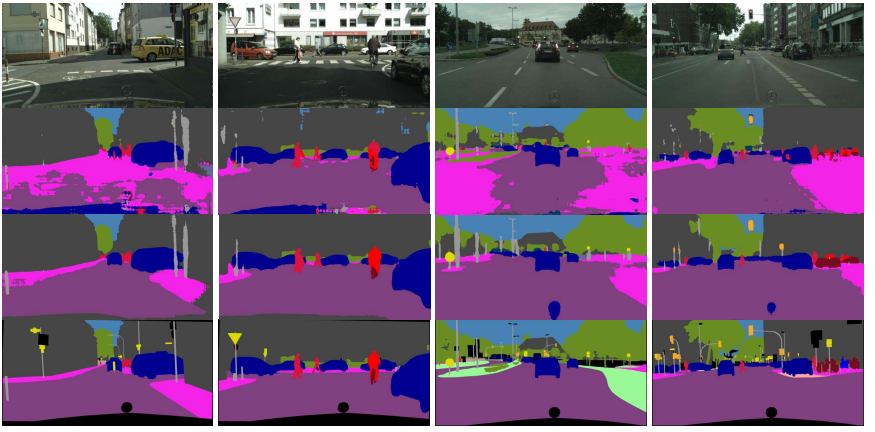

Figure 3: The predicted results of our DUDA for SYNTHIA-to-Cityscapes benchmark. The first row is the valid images, the second and third rows is the FDA warmup results and our DUDA results. The last row is the ground truth.

References

[1] Kai Zhang, Yifan Sun, Rui Wang, Haichang Li, and Xiaohui Hu. Multiple fusion adaptation: A strong framework for unsupervised semantic segmentation adaptation. In *The British Machine Vision Conference (BMVC)*, 2021.

[2] Yang Zou, Zhiding Yu, BVK Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)*, pages 289–305, 2018.