Domain-Adaptive Semantic Segmentation with Memory-Efficient Cross-Domain Transformers — Supplementary Material

Ruben Mascaro
rmascaro@ethz.ch
Lucas Teixeira
lteixeira@mavt.ethz.ch
Margarita Chli
chlim@ethz.ch

Vision for Robotics Lab
ETH Zurich, Switzerland
University of Cyprus, Cyprus

Qualitative Analysis

The supplementary material provides an extended analysis on example predictions obtained with our trained models on the evaluated UDA benchmarks. As we report results averaged over 3 random seeds in the paper, for each benchmark we take the model with the median accuracy to generate the predictions. Our results are compared with the predictions made by the same models trained using the DAFormer [1] UDA pipeline, which we take as baseline. These predictions are generated using the weights and code provided in the original DAFormer repository[1].

Fig. 1 shows qualitative results for synthetic-to-real UDA in both the GTA→Cityscapes (rows 1-3) and the SYNTHIA→Cityscapes (rows 4-6) benchmarks. Predictions are computed on the validation set of the Cityscapes dataset. The increase in performance of our approach can be observed in the ability of our models to better distinguish between *road* and *sidewalk* (especially on SYNTHIA→Cityscapes), even in images where their appearance is very similar. In addition, our approach leads to finer segmentation of classes such as *terrain* and a reduction of misclassification errors.

Fig. 2, on the other hand, shows qualitative results for clear-to-adverse-weather UDA on Cityscapes→ACDC. Although we report the performance on the hidden ACDC test set in the paper, here we compute the predictions on the validation set for comparison with the ground-truth labels. In this case, the increase in performance of our approach with respect to DAFormer is much more evident, especially for classes such as *sky*, *road*, and *sidewalk*. Besides, we observe that our approach reduces the misclassifcation of parts of *cars*, *busses*, and *trains*, which is a common problem DAFormer suffers from. This is an indication of the ability of our approach to better guide the learning of context relationships in the target domain.

[1] https://github.com/lhoyer/DAFormer

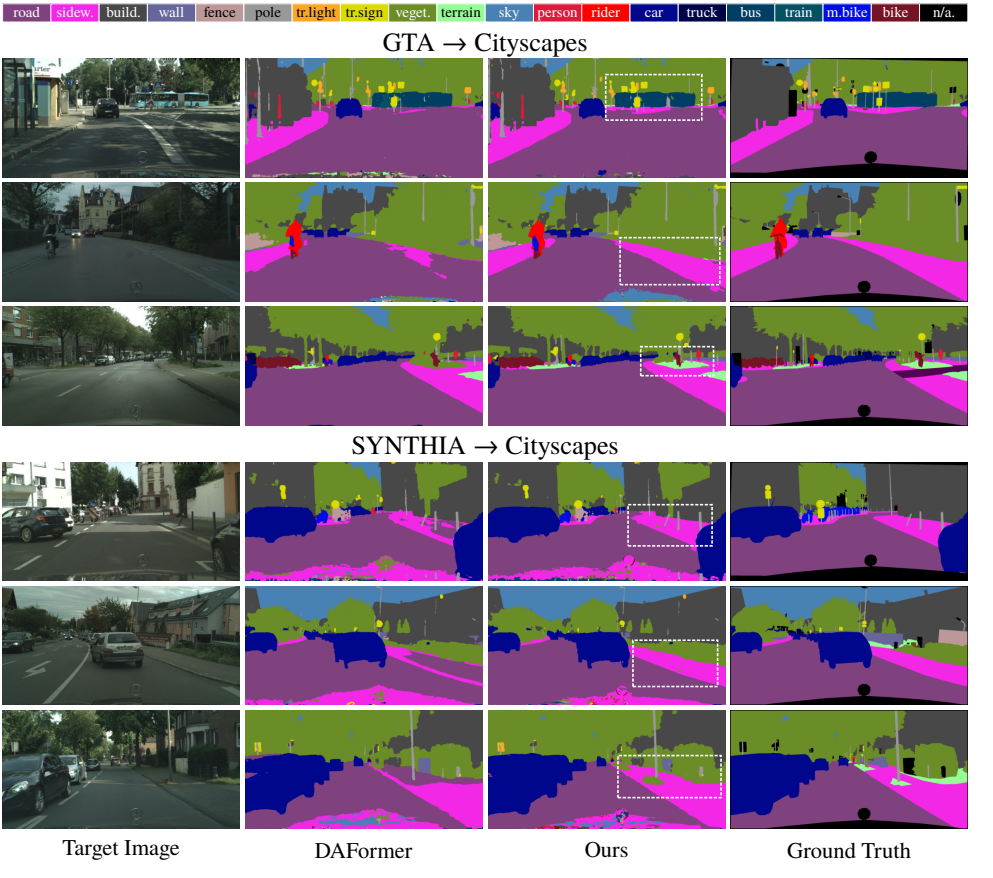

Figure 1: Qualitative semantic segmentation results of our method compared to the baseline DAFormer [11] on GTA→Cityscapes (rows 1-3) and SYNTHIA→Cityscapes (rows 4-6). Our approach shows greater robustness against the confusion of *sidewalk* and *road* (rows 2, 4-6), *vegetation* and *terrain* (row 3), or *bus* and *train* (row 1).

Figure 2: Qualitative semantic segmentation results of our method compared to the baseline DAFormer [11] on Cityscapes→ACDC. Examples of fog (rows 1-2), rain (rows 3-4), snow (rows 5-6) and night (rows 7-8) are provided. Our approach shows greater robustness against problems such as the confusion of *sidewalk* and *road* (rows 5, 6, 7), the confusion of *wall* and *fence* (rows 2, 3), the confusion of *sky* and *road* (row 1), and the misclassification of parts of *cars* (rows 1, 6), *busses* (row 4) and *trains* (row 8). Better segmentation accuracy is also observed in particularly complex regions, such as the *car* occluded behind two *people* in row 3.

References

[1] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.