Supplemental Materials to
" Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation"

In this supplemental file, we provide the following materials:

- The training procedure of CPSL;
- The definition of mean pixel accuracy (MPA) (referring to Sec4.3-Unequal partition constraint in the main paper);
- Ablation studies in terms of per-category IoU (referring to Sec4.3-Ablation study in the main paper);
- Comparisons on the training process on the GTA5→Cityscapes task;
- More parameter sensitivity analyses (referring to Sec4.3-Parameter sensitivity analysis in the main paper);
- More qualitative results (referring to Sec4.2-Qualitative results in the main paper).

1. Algorithm

The training procedure of our CPSL is summarized in Algorithm. 1. For detailed equations and loss functions, please refer to our main paper.

2. Mean pixel accuracy (MPA)

Denoting by C the number of classes, by p_{ij} the number of pixels which belong to the i-th class but are wrongly classified into the j-th class, and by p_{ii} the number of pixels which belong to the i-th class and are accurately classified into the i-th class, the pixel accuracy (PA) of the i-th class is defined as:

$$PA = \frac{p_{ii}}{\sum_{j=1}^{C} p_{ij}}. \tag{1}$$

Then the mean pixel accuracy (MPA) is defined as:

$$MPA = \frac{1}{C} \sum_{i=1}^{C} \frac{p_{ii}}{\sum_{j=1}^{C} p_{ij}}. \tag{2}$$

As discussed in Sec. 4.3 of our manuscript, under the constraint of equal partition, many pixels belonging to large categories are assigned to small categories, largely improving the pixel accuracy of small classes. However, this constraint has very small influences on large categories because these categories contain a great number of pixels. Therefore, the MPA is improved.

3. Ablation study

Due to the space limit, we only reported the mIoU scores in Tab.3 of the main paper. Here we present in Tab. 1 the per-class IoU scores of ablation studies. Note that "w/o CB" denotes that we do not employ the class-balanced sampling techniques, and constrain that Q should induce an equipartition of data rather than an unequal partition. One can see that this leads to a degradation of 3.9 in terms of mIoU, demonstrating that the equal partition constraint is not reasonable when the class distribution of data is highly imbalanced.

Algorithm 1: Training Procedure of CPSL

Input : Training data $\mathcal{D}_S = \{(X_n^s, Y_n^s)\}_{n=1}^{N_S}$ and $\mathcal{D}_T = \{X_n^t\}_{n=1}^{N_T}$;
Output: The output model f_{SEG} ;

Generate soft pseudo labels P_{ST} with the warmed-up model;
Initialize the weight of f_{SL} and f'_{SL} with the prototypes $[\bar{\mathbf{z}}_1, \cdots, \bar{\mathbf{z}}_C]$ for each category computed by Eq. 7;
for $i = 1$ *to* max_epochs **do**
  for $n = 1$ *to* N_S **do**
    Get source image X_n^s;
    Train the model f_{SEG} using loss $\mathcal{L}_{\text{SEG}}^s$;

    Get target image X_n^t;
    Extract features from X_n^t to obtain $Z \in \mathbb{R}^{H \times W \times D}$ and normalize it with $z_i = \frac{z_i}{||z_i||_2}$;
    Sample a group of pixels $\hat{Z} = [z_1, \cdots, z_M]$ from Z randomly;
    Augment the features $\hat{Z}$ with a memory bank $\mathcal{M}$ and obtain $Z_{aug} = [\hat{Z}; \mathcal{M}]$;
    for $k = 1$ *to sinkhorn_iterations* **do**
      $Q^*_{aug} = \operatorname{diag}(\alpha) \exp(\frac{f_{\text{SL}}(Z_{aug})}{\varepsilon}) \operatorname{diag}(\beta)$;
    end
    Compute the self-labeling loss $\mathcal{L}_{\text{SL}}$ through Eq. 5 using the cluster assignment of current batch Q_{cur};
    Train the self-labeling head f_{SL} using loss $\mathcal{L}_{\text{SL}}$.

    Update the momentum self-labeling head f'_{SL} in an EMA manner;
    Pass X_n^t through f'_{SEG} and f'_{SL} to obtain self-labeling assignment P_{SL};
    Use P_{SL} to rectify P_{ST} and obtain the rectified pseudo labels $\hat{Y}_n^t$ through Eq. 1;
    Update f_{SEG} using loss $\mathcal{L}_{\text{SEG}}^t$;
    Update the momentum segmentation model f'_{SEG} in an EMA manner.
  end
end

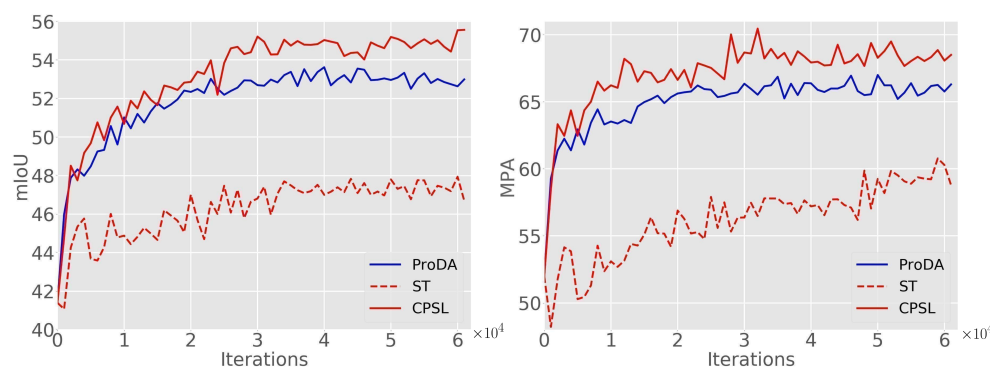

Figure 1. The mIoU (left) and MPA (right) scores evaluated on the validation set during the training.

4. Training process of CPSL and ProDA

To further highlight the improvement of CPSL during training, we plot the curves of mIoU and MPA scores on the GTA5→Cityscapes task in Fig. 1. A large performance improvement of CPSL over ProDA can be observed in terms of both mIoU and MPA.

Method	road	sideway	building	wall	fence	pole	light	sign	vege	terrace	sky	person	rider	car	truck	bus	train	motor	bike	mIoU	Δ
w/o SL	91.9	56.3	82.9	35.9	30.2	37.5	37.4	32.9	85.3	39.2	77.8	51.2	18.6	84.7	37.8	44.6	1.0	20.2	42.7	47.8	-7.9
w/o ST	82.4	39.0	70.5	30.5	16.0	24.1	39.6	37.0	77.8	24.2	78.7	28.5	18.7	75.7	9.2	36.1	4.1	22.9	36.5	39.4	-16.3
w/o CB	91.7	51.3	84.0	33.9	24.3	42.5	43.3	**49.0**	81.5	29.1	75.8	67.0	28.5	87.7	34.3	**63.3**	20.1	36.0	40.5	51.8	-3.9
w/o Init	89.6	**56.1**	80.0	40.3	**36.7**	43.7	45.9	39.6	86.2	39.8	**81.9**	66.7	24.8	**89.0**	**45.4**	50.8	0.0	31.4	9.3	49.9	-5.8
w/o Aug	90.6	45.5	83.8	41.4	33.0	44.3	**52.0**	42.0	**86.4**	40.2	81.6	68.4	28.9	88.0	42.8	58.5	14.9	40.0	47.1	54.2	-1.5
w/o Mom	**92.6**	53.7	**84.1**	41.7	36.6	**44.8**	50.6	41.7	86.2	**40.5**	79.6	68.2	26.6	87.4	37.4	55.9	19.3	43.1	47.5	54.6	-1.1
CPSL	91.7	52.9	83.6	**43.0**	32.3	43.7	51.3	42.8	85.4	37.6	81.1	**69.5**	**30.0**	88.1	44.1	59.9	**24.9**	**47.2**	**48.4**	**55.7**	-

Table 1. Ablation studies on the key components of CPSL in terms of per-category IoU. The top score is highlighted in **bold** font.

5. Parameter analysis

Tab. 2 and Tab. 3 show the segmentation results by using different self-labeling loss weight λ_1 and consistency regularization loss weight λ_2, respectively. One can see that our method is insensitive to these two parameters. Tab. 4 shows the effect of temperature τ. We employ the cluster assignment P_{SL} as a weight map to online modulate the softmax probability of pseudo labels P_{ST}, where the temperature τ controls the modulation intensity. When $\tau \rightarrow 0$, the modulation intensity increases so that the rectified pseudo labels $\hat{Y}^t$ will rely heavily on P_{SL}. When $\tau \rightarrow \infty$, the modulation intensity decreases so that the rectified pseudo labels $\hat{Y}^t$ will rely heavily on P_{ST}.

λ_1	0	0.01	0.1	0.5
mIoU	51.4	54.2	55.7	54.9

Table 2. The influence of parameter λ_1.

λ_2	1	5	10	20	30
mIoU	55.5	55.7	55.2	54.7	54.4

Table 3. The influence of parameter λ_2.

τ	0.05	0.08	0.1	0.15
mIoU	52.8	55.7	55.3	53.6

Table 4. The influence of temperature parameter τ.

6. Qualitative results

PSL *vs.* CPSL. To better illustrate the performance of our method, we implement a variant of CPSL without class-balanced training, *i.e.*, purely Pixel-level Self-Labeling (PSL). The qualitative results of PSL and CPSL are shown in Fig. 2. Overall, CPSL is capable of producing more accurate segments across various scenes. Specifically, our method performs better on long-tailed categories, *e.g.* “bus”, “bicycle”, “person”, “light”. Compared to PSL, the segment boundaries of CPSL tend to be clearer and closer to object boundaries, such as “bicycle” and “person”. Besides, it is noteworthy that PSL wrongly classifies the “road” class into the “sidewalk” class in a large area, which is attributed to the equipartition constraint applied on cluster assignments. This constraint is not useful and would even degrade the performance if the real class distribution is not uniform. However, this issue is solved by aligning class distribution of cluster assignments to that of pseudo labels.

Comparisons with state-of-the-arts. As in Fig. 3 of the main manuscript, we compare our CPSL with other state-of-the-art methods. Here we provide more visualization results in Fig. 3 - Fig. 9. Our method performs better on long-tailed categories, such as “person”, “pole”, “traffic light”, “bus”, and “rider”.

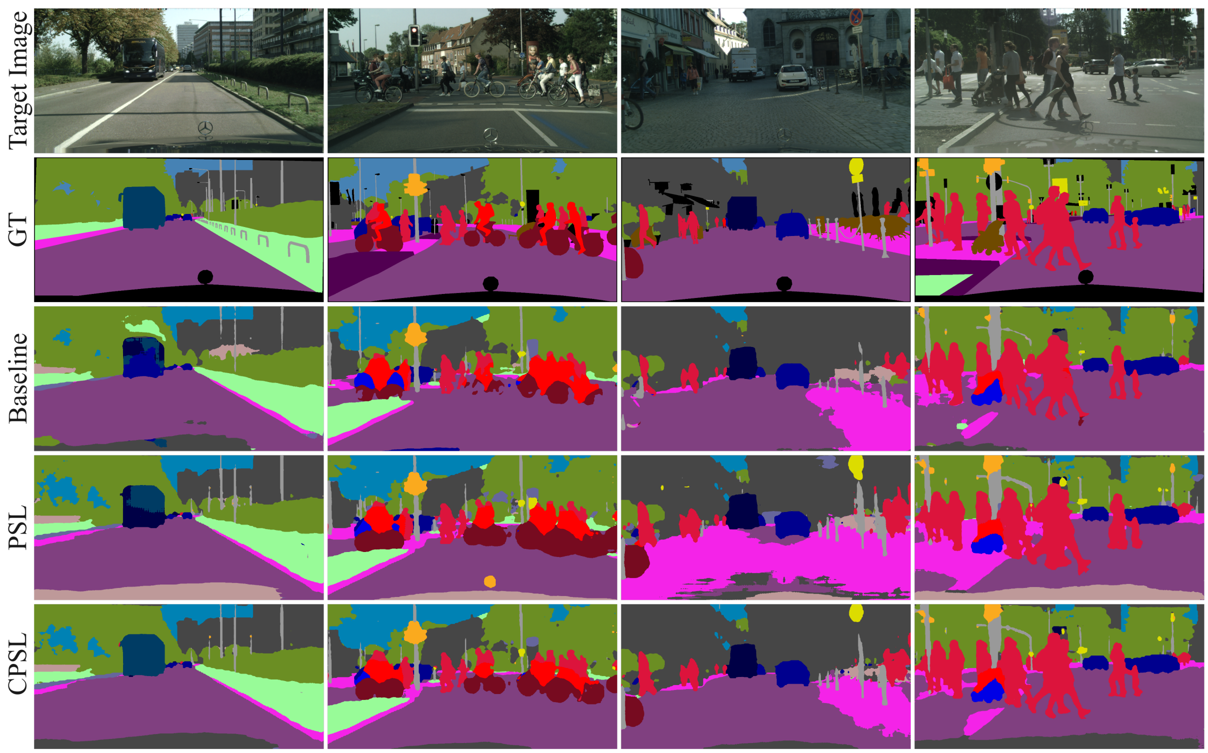

Figure 2. Qualitative results of PSL and CPSL on the GTA5→Cityscapes task.

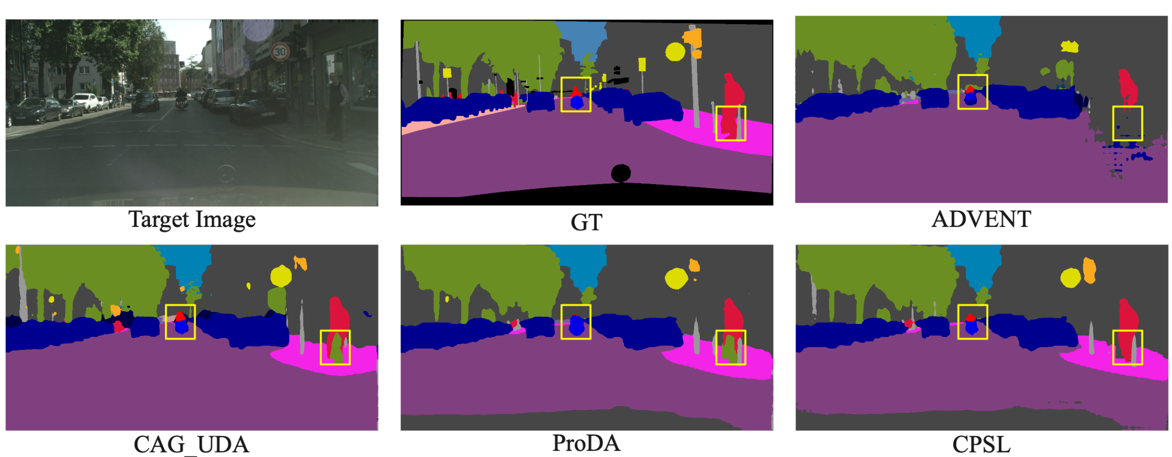

Figure 3. Qualitative comparison of different methods on the GTA5→Cityscapes task.

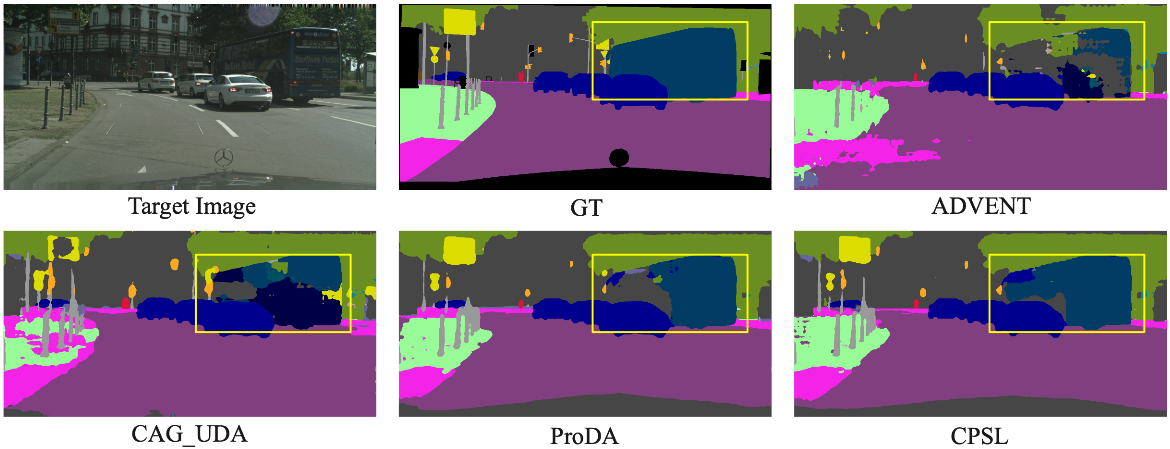

Figure 4. Qualitative comparison of different methods on the GTA5→Cityscapes task.

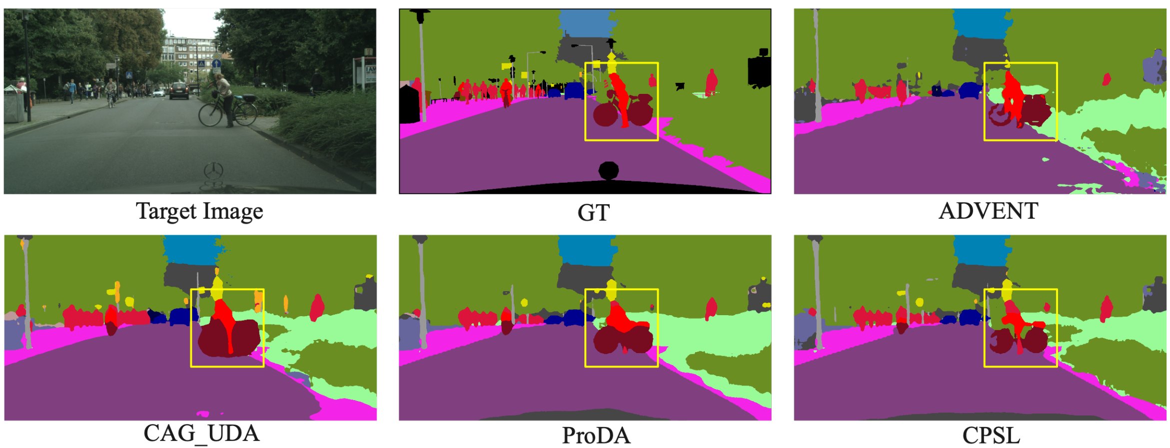

Figure 5. Qualitative comparison of different methods on the GTA5→Cityscapes task.

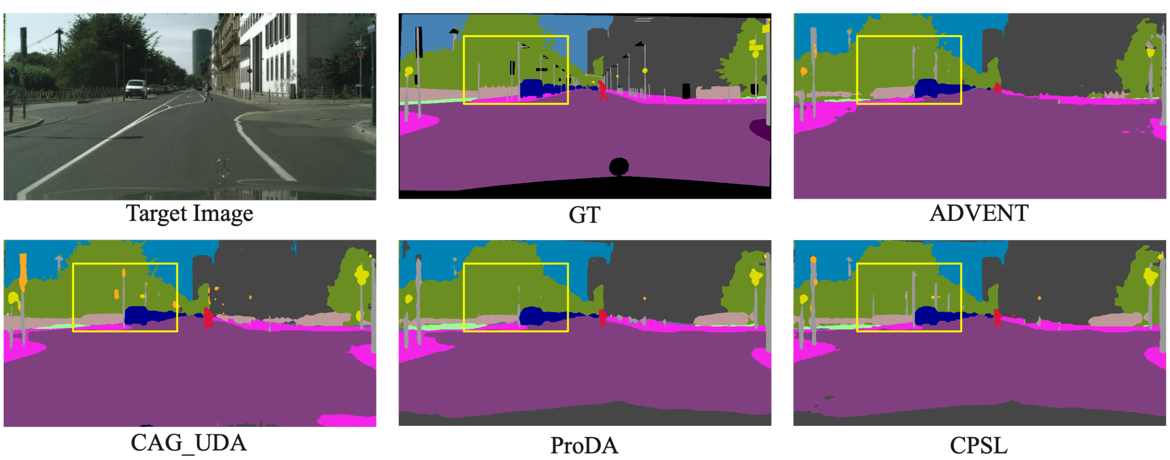

Figure 6. Qualitative comparison of different methods on the GTA5→Cityscapes task.

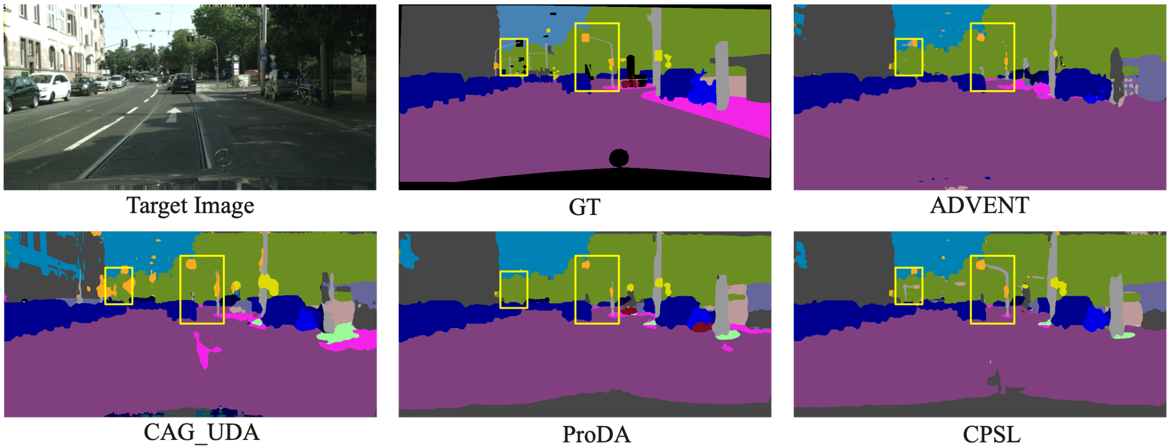

Figure 7. Qualitative comparison of different methods on the GTA5→Cityscapes task.

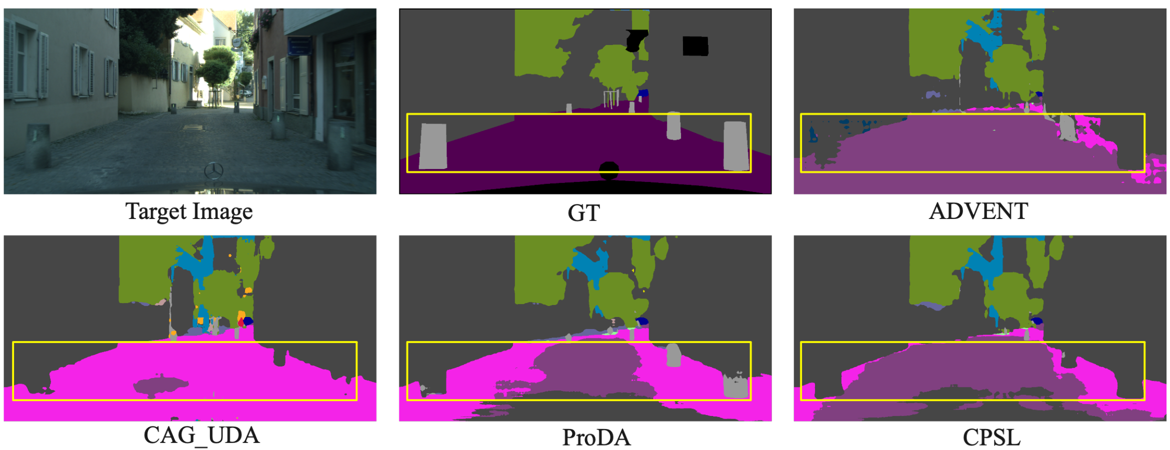

Figure 8. Qualitative comparison of different methods on the GTA5→Cityscapes task.

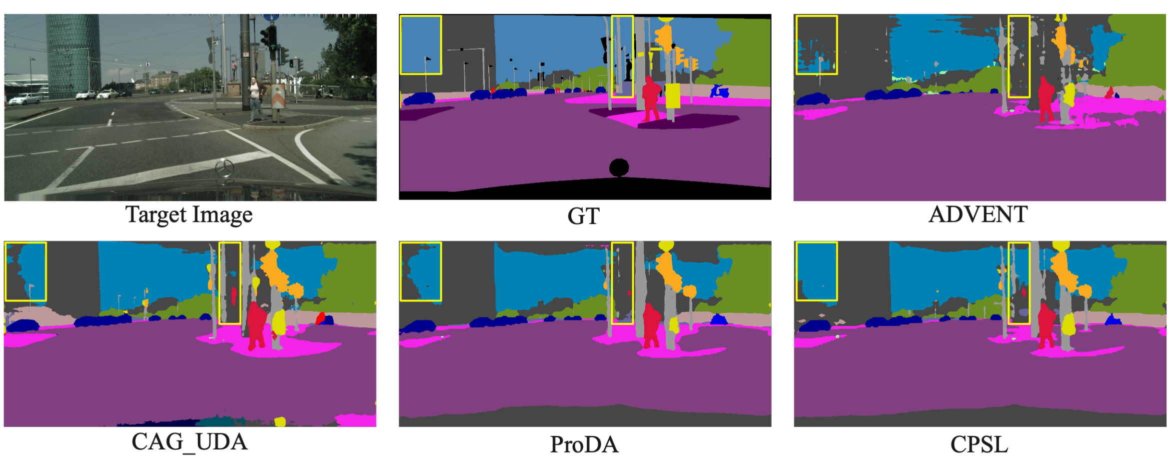

Figure 9. Qualitative comparison of different methods on the GTA5→Cityscapes task.