Learning Pseudo-Relations for Cross-domain Semantic Segmentation Supplementary Material

A. Details on Pseudo-Relations.

We show the situation where unreliable relationship pairs come into play, as shown in Figure 1. U1, U2 and U3 are considered unreliable samples due to the lower highest confidence. However, U1 contains strong clues of category 0, i.e. U1 does not belong to category 0, U2 contains strong clues of category 0, i.e. U2 does not belong to class 1. If we establish a positive sample relationship between U1 and U2, that is, they are of the same class, then class clues can be exploited to assist the model in distinguishing ambiguous samples. A similar situation can be found between U2 and U3. It shows that valuable class clues can be propagated through reliable relations, which provides a effective way to utilize unreliable samples.

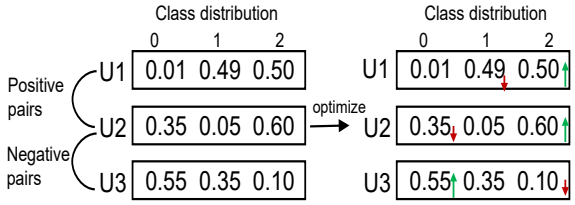

Figure 1: Examples for efficient relation learning between unreliable samples.

B. Qualitative Results.

We visualize our results in comparison with the baseline model (an improved version of SAC [1]) and the existing self-training methods in synthetic-to-real scene, as shown in Fig. 2. Compared with SAC [1], our method shows advantages at object edges and on small objects. For example, at the edges of 'sidewalk' and 'road', our segmentation results are smooth and consistent with the shallow relations. Compared with ProDA [2], our method produces more structured and reasonable segmentation results. For example, in the fifth line, for the 'building' and 'road', the ProDA method learns category information but this information is not complete, resulting in incomplete segmentation. This is because we take full advantage of the relations in the pseudo-labels and use the shallow (RGB) relations to produce strong regularization on the segmentation. This verifies the effectiveness of our method, which can reasonably utilize pseudo-relations to learn reliable category relation representations.

We present T-SNE results on GTA5 $\rightarrow$ Cityscapes task, as shown in Fig. 3. It shows that the feature distribution map of the baseline model has severe false mixing between classes. In contrast, our method pushes the inter-class features and narrows the intra-class features, which constructs a better feature space of the target domain.

C. Hyper-parameters Sensitivity.

In this section, we verify the sensitivity of the trade-off coefficient α and the grid size N in RTea.

The grid size N are verified on both adaptation tasks in Table. 1. We explore the grid size N with a step of 4 pixel (equivalent to a step size of 32 pixels on the original image size). When the grid is small, the modeled pixel associations are less, which is not enough to propagate the effective information of confident pixels and achieves less gains. As the grid size N increases, effective associations are propagated and the adaptation performance of the model gradually becomes better. When the grid size N is set to 8, the best performance is achieved on the two tasks, achieving 3.8% and 2.9% performance improvements, respectively. When the grid is larger, the range of the pixel relation is expanded but massive noise will be introduced into the high-level relations S_{high}, resulting in performance degradation. This is also consistent with our observation of two pseudo-relationship building priors. Overall, when the grid size is set between 8-16, the performance improvement of the model is relatively stable. In the end, we set the grid size N as 8 on both tasks.

In Table. 2, we search for the balance parameter α. The larger the value of α, the greater the weight of the low-level pseudo-relation. When α is small, high-level relations S_{high} dominate and the performance shows limited improvement, which shows noisy relation in S_{high} prevents pseudo-relations from better working. As the value of α increases, low-level relations S_{low} are introduced to correct noise in high-level relations S_{high} and the performance of adaptation is further improved. When the grid size N is set

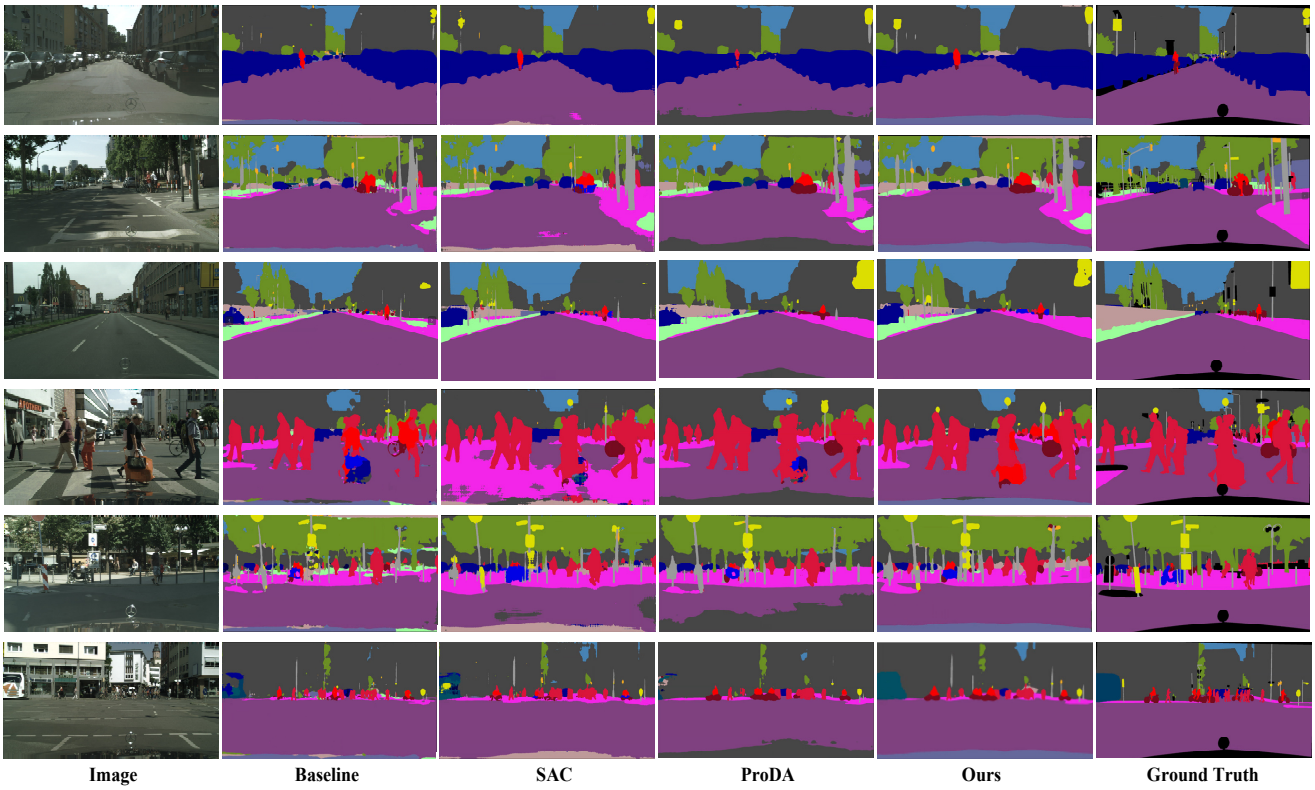

Figure 2: Visualization of prediction results from different methods. Baseline denotes the pseudo-label learning, *e.g.* L_s+L_{st}. SAC [1] and ProDA[2] are the existing state-of-the-art self-training methods. Ours denotes the proposed method RTea. Best viewed in color.

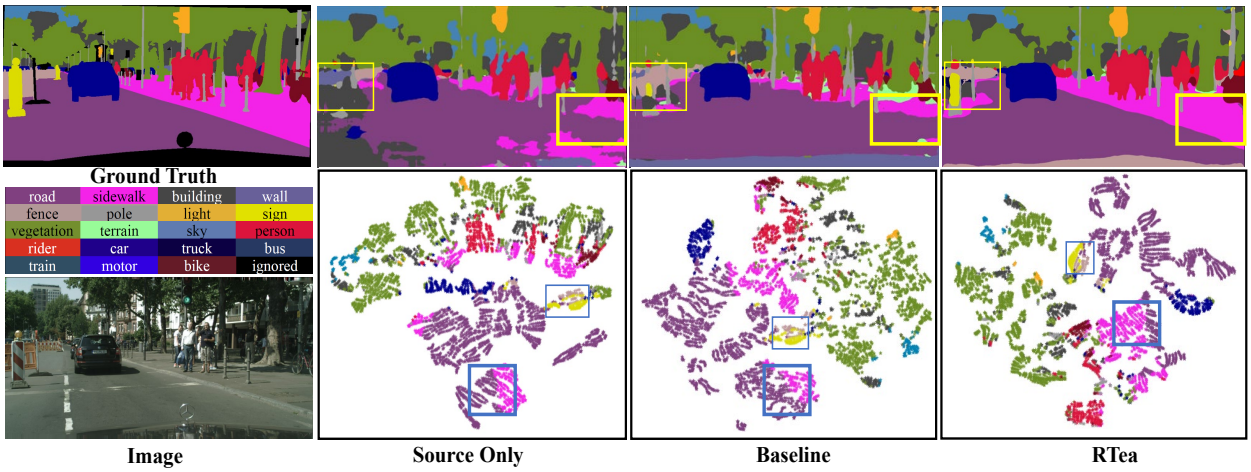

Figure 3: Qualitative visualization of the feature distribution of the target domain from different models on GTA5 $\rightarrow$ Cityscapes task, *i.e.* source only model *v.s.* baseline model (an improved version of SAC [1]) *v.s.* RTea model (Ours). Different colors of the scatter points represent different categories. Best viewed in color.

to 0.5, the best performance is achieved on the two tasks. But when α exceeds 0.9, the model is degraded heavily, which illustrates that semantic information in high-level relations is essential for adaptation. It shows that high- and low-level relations should be mixed to achieve better performance improvement, and the fusion effect is relatively stable within a certain range. Therefore, we finally set the α to 0.5.

N	2	4	8	12	16
G2C	56.1	58.2	**59.6**	59.1	58.9
S2C	54.1	55.3	**56.9**	56.1	55.9

Table 1: Sensitivity experiments with varying N on GTA5 $\rightarrow$ Cityscapes (G2C) and SYNTHIA $\rightarrow$ Cityscapes (S2C) tasks.

α	0	0.1	0.3	0.5	0.7	0.9	1
G2C	56.8	57.4	58.9	**59.6**	59.2	58.1	57.8
S2C	54.6	55.1	55.9	**56.9**	56.3	54.9	54.6

Table 2: Sensitivity experiments with varying α on GTA5 $\rightarrow$ Cityscapes (G2C) and SYNTHIA $\rightarrow$ Cityscapes (S2C) tasks.

References

[1] Nikita Araslanov and Stefan Roth. Self-supervised augmentation consistency for adapting semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15384–15394, June 2021. 1, 2

[2] Pan Zhang, Bo Zhang, Ting Zhang, Dong Chen, Yong Wang, and Fang Wen. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12414–12424, 2021. 1, 2