# Unsupervised Domain Adaptation for Semantic Segmentation with Pseudo Label Self-Refinement (Supplementary Document)

Xingchen Zhao[2*], Niluthpol Chowdhury Mithun[1*], Abhinav Rajvanshi[1], Han-Pang Chiu[1], Supun Samarasekera[1]

[1]SRI International, Princeton, NJ, USA [2]Northeastern University, Boston, MA, USA

[1]firstname.lastname@sri.com [2]zhao.xingc@northeastern.edu

## 1. Overview

This is the supplementary material to support our manuscript "Unsupervised Domain Adaptation for Semantic Segmentation with Pseudo Label Self-Refinement". It contains additional quantitative and qualitative results related to our experiments that couldn't be included in the main article due to space constraints. In Sec. 2, we provide quantitative results of SYNTHIA→Cityscapes adaptation experiment comparing state-of-the-art methods. In Sec. 3, we present ablation studies on Cityscapes→Dark Zurich and SYNTHIA→Cityscapes adaptation to analyze the impact of different components of our method. We also present experiments to show the effect of varying weights of refinement losses in this section. In Sec. 4, we showcase several qualitative examples of semantic segmentation, comparing our approach and baseline methods.

## 2. SYNTHIA→Cityscapes Results

We compare our method with prior UDA methods on SYNTHIA→Cityscapes adaptation in Table 1. From the last part of the table, it is evident that our method performs significantly better than SOTA methods (mIOU of 62.2 with MIC-DAFormer and 60.9 with DAFormer compared to 63.3 with ours). Same as Cityscapes→Dark Zurich and GTA→Cityscapes results in the main paper, our method consistently achieves higher IoU across most classes. The ResNet-based baseline DACS (w/ our PRN) was trained by combining our PRN module with the prior method DACS. We train this baseline to compare with prior pseudo-label selection or refinement-based UDA methods reported in the first part of Table 1 (e.g., CCM, MetaCor, UAPLR, ProDA). We see our PRN module leads to significant improvement over other pseudo-label selection or refinement-based UDA methods. From the last part of Table 1, we see incorporating HRDA training leads to further improvement in performance, and Ours with HRDA performs better than state-of-the-art DAFormer with HRDA.

## 3. Ablation Studies

We have presented the ablation study of our proposed method on GTA→Cityscapes in Table 3 of the main paper. Here, we present an ablation study on Cityscapes→Dark-Zurich in Table 2 to analyze different components of our method, i.e., Self-Training (ST), Pseudo Label Refinement (PL-R), Noise Mask (NM), Contrastive Learning without or with using the output of PRN (CL w/o R, CL w/ R) and Fourier Adaptation (FA). We again observe that the proposed method leads to a large improvement over the self-training baseline reported in the second row (58.4 in row-2.8 vs. 51.2 in row-2.2). We also observe that our proposed PRN module (with both pseudo-label refinement and noise-mask prediction) leads to significant improvement over the self-training baseline (row-2.4 vs. row-2.2). The impact of noise-mask prediction in PRN shows improvement compared to without it (row-2,4 vs. row-2.2). It is also evident that our pseudo-label refinement is crucial to achieving a performance boost with the contrastive learning module comparing row-2.5 and row-2.6 with row-2.4. We see the use of the PRN module output is crucial for contrastive learning to achieve a performance boost. Comparing row-2.7 with row-2.4, we see performance improvement by applying the FA module. Finally, row-2.8 shows the performance when all the components of our framework are used.

We also perform an ablation study on SYNTHIA→Cityscapes in Table 3. We observe a similar trend to Cityscapes→Dark-Zurich and GTA5→Cityscapes ablation studies that different components of the proposed UDA framework with pseudo-label refinement module consistently help improve performance.

In Fig. 1, we present results on Cityscapes→Dark-Zurich

[*]Equal Contribution

Table 1. Performance evaluation on **SYNTHIA→Cityscapes**. We report mIoU over 16 common categories between these datasets.

| Method | | Road | S.Walk | Build. | Wall | Fence | Pole | T.Light | Sign | Veget. | Sky | Person | Rider | Car | Bus | M.Bike | Bike | mIoU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CBST [10] | ResNet-Based | 68.0 | 29.9 | 76.3 | 10.8 | 1.4 | 33.9 | 22.8 | 29.5 | 77.6 | 78.3 | 60.6 | 28.3 | 81.6 | 23.5 | 18.8 | 39.8 | 42.6 |
| CCM [5] | ResNet-Based | 79.6 | 36.4 | 80.6 | 13.3 | 0.3 | 25.5 | 22.4 | 14.9 | 81.8 | 77.4 | 56.8 | 25.9 | 80.7 | 45.3 | 29.9 | 52.0 | 45.2 |
| MetaCor [1] | ResNet-Based | 92.6 | 52.7 | 81.3 | 8.9 | 2.4 | 28.1 | 13.0 | 7.3 | 83.5 | 85.0 | 60.1 | 19.7 | 84.8 | 37.2 | 21.5 | 43.9 | 45.1 |
| DACS [6] | ResNet-Based | 80.6 | 25.1 | 81.9 | 21.5 | 2.9 | 37.2 | 22.7 | 24.0 | 83.7 | 90.8 | 67.6 | 38.3 | 82.9 | 38.9 | 28.5 | 47.6 | 48.4 |
| UAPLR [8] | ResNet-Based | 79.4 | 34.6 | 83.5 | 19.3 | 2.8 | 35.3 | 32.1 | 26.9 | 78.8 | 79.6 | 66.6 | 30.3 | 86.1 | 36.6 | 19.5 | 56.9 | 48.0 |
| CorDA [7] | ResNet-Based | 93.3 | 61.6 | 85.3 | 19.6 | 5.1 | 37.8 | 36.6 | 42.8 | 84.9 | 90.4 | 69.7 | 41.8 | 85.6 | 38.4 | 32.6 | 53.9 | 55.0 |
| ProDA [9] | ResNet-Based | 87.8 | 45.7 | 84.6 | 37.1 | 0.6 | 44.0 | 54.6 | 37.0 | 88.1 | 84.4 | 74.2 | 24.3 | 88.2 | 51.1 | 40.5 | 45.6 | 55.5 |
| **DACS (w/ our PRN)** | ResNet-Based | **88.1** | **47.1** | **84.8** | **37.5** | **0.9** | **45.0** | **55.4** | **38.6** | **88.2** | **85.2** | **75.2** | **25.5** | **88.4** | **51.9** | **41.3** | **46.4** | **56.2** |
| DAFormer [2] | SegFormer | 84.5 | 40.7 | 88.4 | 41.5 | 6.5 | 50.0 | 55.0 | 54.6 | 86.0 | 89.8 | 73.2 | 48.2 | 87.2 | 53.2 | 53.9 | 61.7 | 60.9 |
| MIC-DAFormer [4] | SegFormer | 83.0 | 40.9 | 88.2 | 37.6 | 9.0 | 52.4 | 56.0 | 56.5 | 87.6 | 93.4 | 74.2 | 51.4 | 87.1 | 59.6 | 57.9 | 61.2 | 62.2 |
| **Ours** | SegFormer | **86.6** | **44.7** | **91.7** | **44.4** | **9.3** | **53.0** | **55.9** | **57.2** | **88.3** | **89.2** | **75.1** | **49.8** | **91.2** | **56.9** | **55.9** | **63.8** | **63.3** |
| DAFormer (w/ HRDA) [3] | SegFormer | 85.2 | 47.7 | 88.8 | 49.5 | 4.8 | 57.2 | 65.7 | 60.9 | 85.3 | 92.9 | 79.4 | 52.8 | 89.0 | 64.7 | 63.9 | 64.9 | 65.8 |
| **Ours (w/ HRDA)** | SegFormer | **87.8** | **49.4** | **88.1** | **49.5** | **5.3** | **59.1** | **65.6** | **62.2** | **85.6** | **94.2** | **79.1** | **53.6** | **87.1** | **65.6** | **65.8** | **66.2** | **66.5** |

Table 2. Ablation study with different components of our proposed method on **Cityscapes→Dark-Zurich**.

| # | ST | PL-R | NM | CL w/o R | CL w/ R | FA | mIoU |
|---|---|---|---|---|---|---|---|
| 2.1 | x | x | x | x | x | x | 37.5 |
| 2.2 | ✓ | x | x | x | x | x | 51.2 |
| 2.3 | ✓ | ✓ | x | x | x | x | 54.9 |
| 2.4 | ✓ | ✓ | ✓ | x | x | x | 55.8 |
| 2.5 | ✓ | ✓ | ✓ | ✓ | x | x | 55.3 |
| 2.6 | ✓ | ✓ | ✓ | x | ✓ | x | 56.5 |
| 2.7 | ✓ | ✓ | ✓ | x | x | ✓ | 58.0 |
| 2.8 | ✓ | ✓ | ✓ | x | ✓ | ✓ | 58.4 |

Table 3. Ablation study with different components of our proposed method on **SYNTHIA→Cityscapes**.

| # | ST | PL-R | NM | CL w/o R | CL w/ R | FA | mIoU |
|---|---|---|---|---|---|---|---|
| 3.1 | x | x | x | x | x | x | 46.5 |
| 3.2 | ✓ | x | x | x | x | x | 60.9 |
| 3.3 | ✓ | ✓ | x | x | x | x | 61.7 |
| 3.4 | ✓ | ✓ | ✓ | x | x | x | 62.1 |
| 3.5 | ✓ | ✓ | ✓ | ✓ | x | x | 62.2 |
| 3.6 | ✓ | ✓ | ✓ | x | ✓ | x | 62.5 |
| 3.7 | ✓ | ✓ | ✓ | x | x | ✓ | 63.1 |
| 3.8 | ✓ | ✓ | ✓ | x | ✓ | ✓ | 63.3 |

by varying weight for target refinement loss (i.e., $\mathcal{L}_{ce}^{RT} + \mathcal{L}_{bce}^{RT}$), while keeping the weight (i.e., $\lambda_2$) of source refinement loss (i.e., $\mathcal{L}_{ce}^{RS} + \mathcal{L}_{bce}^{RS}$) fixed. For this experiment, we use our proposed model without the additional CL and FA modules (i.e., row-2.4 of Table. 2). As reported in row-2.2 of Table 2, the self-training baseline achieves mIoU of 51.2. We observe mIoU improvement compared to the self-training baseline in all the cases. When the target pseudo-label refinement loss is not used (i.e., weight is set to 0), the performance drops to mIoU of 53.3 ($-2.5\%$ compared to the case of loss weight set to 1). It shows that the source refinement loss is effective in improving pseudo-label quality and overall performance (53.3 vs. the self-training baseline result of 51.2). However, the target refinement loss helps to further improve the performance. The best performance is achieved with the target refinement loss weight set to 1.

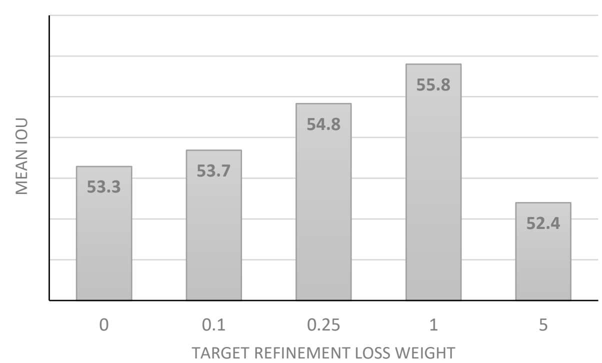


Figure 1. Results on varying weight for target refinement loss (i.e., $\mathcal{L}_{ce}^{RT} + \mathcal{L}_{bce}^{RT}$), while keeping the weight (i.e., $\lambda_2$) of source refinement loss fixed in Cityscapes→Dark-Zurich. For this experiment, we use our proposed model without the CL & FA components.

## 4. Qualitative Results

In this section, we present the qualitative comparison of our approach with the state-of-the-art method DAFormer. The Source-Only baseline results (with no domain adaptation) are also shown for reference. Fig. 2 shows qualitative examples of our method in adapting the model trained on Cityscapes to Dark-Zurich. Similar to the qualitative examples presented in the main paper, we again see that our approach leads to a significant improvement in several classes which can be hard to classify due to changes in domains. We couldn't show the ground truth label in Fig. 2 as we do not have direct access to it for the test set of Dark-Zurich. Fig. 3 shows the qualitative results for adaptation from GTA5 to Cityscapes. These results also include the ground truth semantic labels for reference. We again qualitatively observe that our proposed method consistently performs better than the compared methods.

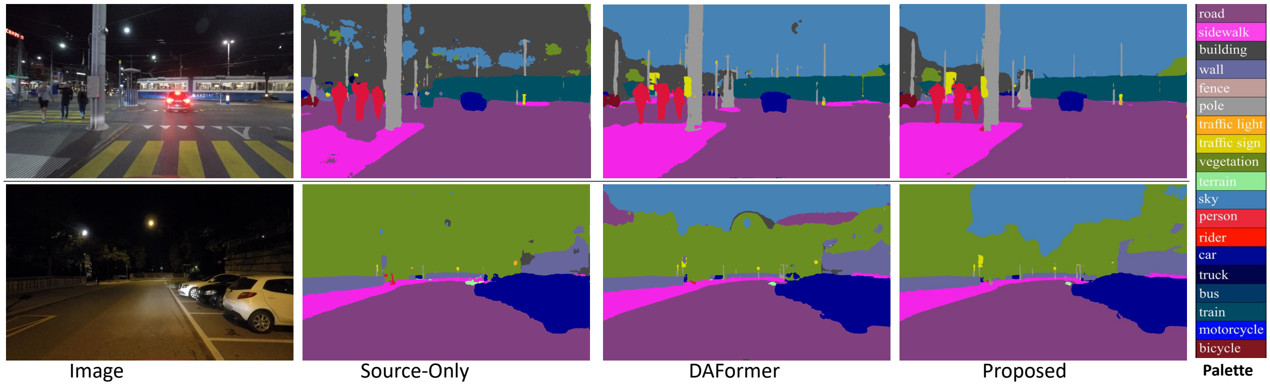


Figure 2. Qualitative Examples of Cityscapes→Dark-Zurich on Dark Zurich test set comparing the Source-Only baseline and DaFormer.

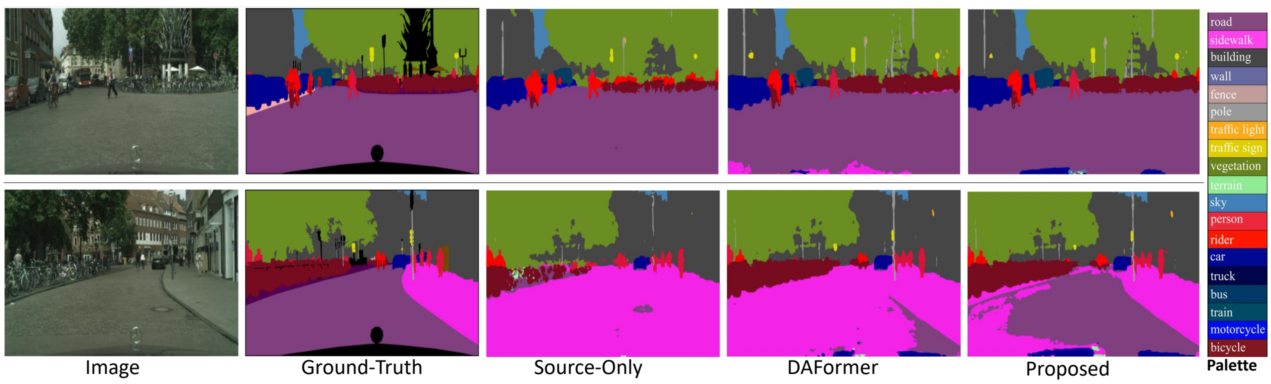


Figure 3. Qualitative Evaluation on GTA5→Cityscapes on Cityscapes val. set comparing GT, Source-Only baseline and DaFormer.

## References

[1] Xiaoqing Guo, Chen Yang, Baopu Li, and Yixuan Yuan. Metacorrection: Domain-aware meta loss correction for unsupervised domain adaptation in semantic segmentation. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3927–3936, 2021. 2

[2] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9924–9935, 2022. 2

[3] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Hrda: Context-aware high-resolution domain-adaptive semantic segmentation. In *Proc. European Conference on Computer Vision (ECCV)*, pages 372–391. Springer, 2022. 2

[4] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. Mic: Masked image consistency for context-enhanced domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11721–11732, 2023. 2

[5] Guangrui Li, Guoliang Kang, Wu Liu, Yunchao Wei, and Yi Yang. Content-consistent matching for domain adaptive semantic segmentation. In *Proc. European Conference on Computer Vision (ECCV)*, 2020. 2

[6] Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson. Dacs: Domain adaptation via cross-domain mixed sampling. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1379–1389, 2021. 2

[7] Qin Wang, Dengxin Dai, Lukas Hoyer, Luc Van Gool, and Olga Fink. Domain adaptive semantic segmentation with self-supervised depth estimation. In *Proc. IEEE/CVF International Conference on Computer Vision*, pages 8515–8525, 2021. 2

[8] Yuxi Wang, Junran Peng, and ZhaoXiang Zhang. Uncertainty-aware pseudo label refinery for domain adaptive semantic segmentation. In *Proc. IEEE/CVF International Conference on Computer Vision*, pages 9092–9101, 2021. 2

[9] 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 *Proc. IEEE/CVF conference on computer vision and pattern recognition*, pages 12414–12424, 2021. 2

[10] Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proc. European Conference on Computer Vision (ECCV)*, pages 289–305, 2018. 2