Universal Domain Adaptation for Semantic Segmentation

Supplementary Material

In this supplementary material, we provide additional details and analyses to complement the main paper. Appendix A provides a detailed visualization of target pseudo labels generated during training. Appendix B delves into further experiments on our proposed DSPD method, including an analysis of its components, variations in pixel-wise weight (w) forms, and an ablation study about fixed ETF. Appendix C focuses on additional evaluations of the TIM method, analyzing its components and the influence of temperature. Appendix D investigates the robustness of our approach through experiments with different proportions of unknown classes. Appendix D provides a sensitivity analysis for hyperparameters used in both our method and the baseline. Finally, Appendix F introduces the name of source-private classes in Pascal-Context → Cityscapes.

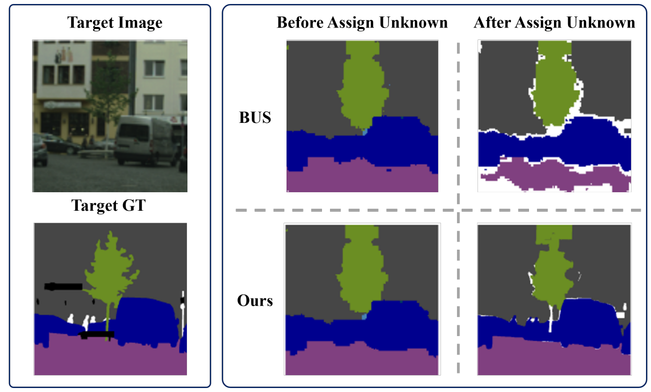

Figure 1. Visualization of target pseudo labels generated by BUS and Ours in the OPDA (Pascal-Context → Cityscapes) scenario at 6000 iterations. The assignment of unknown pixels is performed based on confidence scores.

A. Visualization of Target Pseudo Labels

Figure 1 shows visualization results of pseudo labels generated during training. The comparison highlights that both BUS and Ours generate similar pseudo labels before the assignment of unknown pixels. However, during the confidence score-based assignment of unknown pixels, BUS demonstrates a tendency to incorrectly assign a significant number of pixels as unknown, despite correctly identifying some common class pixels. In contrast, Ours effectively elevates the confidence scores of common classes, successfully reducing the confusion between common and private classes.

B. More Experiments about DSPD

B.1. Performance Comparison of Different Forms of Pixel-Wise Weight Scaling Factor (w)

We compare four pixel-wise weight scaling strategies, each designed to emphasize common classes based on the similarity between pixel embeddings and class prototypes. The formulations are as follows:

$$w^{abs} = 2 - |d_s - d_t|, \tag{1}$$

$$w^{gausian} = 2 \cdot exp(-\frac{(d_S - d_t)^2}{\sigma^2}), \tag{2}$$

$$w^{mean} = \frac{(d_s+1)(d_t+1)}{2} \tag{3}$$

$$w^{ours} = \frac{2(d_s+1)(d_t+1)}{(d_s+1)+(d_t+1)}. \tag{4}$$

In order, we used Absolute Difference (w^{abs}), Gaussian Similarity ($w^{gausian}$), Mean (w^{mean}), and Harmonic Mean (w^{ours}) for Ours. w^{abs} and $w^{gausian}$ approaches assign high weights if the similarities to the source and target prototypes are merely similar. However, this can result in high weights even when the overall similarity to the prototypes is low, making it less effective in distinguishing common classes. w^{mean} approach uses the arithmetic mean of the prototype similarities, incorporating the overall similarity but failing to account for the relative distance between the source and target prototypes. As a result, it is less effective in accurately estimating common classes. We employ a harmonic mean-based formulation that considers both the individual similarities to the prototypes and the relative distance between the source and target prototypes. This approach effectively emphasizes the weights for common classes, improving their estimation and overall performance. As shown in Table 1, w^{ours} achieves the highest performance across all metrics, particularly boosting the H-Score by 3.48 compared to w^{mean}, demonstrating its effectiveness in estimating common classes.

B.2. Ablation Study about L_{proto}

In Table 2, we conducted an ablation study of components of L_{proto} in the OPDA scenario. The components evaluated include L_{CE}, L_{PPC}, and L_{PPD}, each contributing to the effectiveness of handling domain-specific clustering. When L_{CE} is used alone, the model achieves an H-Score of 40.23. Adding L_{PPC} leads to a decrease in H-Score to 37.76, indicating that while L_{PPC} focuses on prototype alignment, it does not significantly enhance performance when used

w	Pascal-Context → Cityscapes		
	Common	Private	H-Score
w^{abs}	53.20	20.10	29.57
$w^{gausian}$	53.76	23.81	33.00
w^{mean}	58.91	27.89	37.85
w^{ours}	**60.94**	**31.27**	**41.33**

Table 1. Experiments result with OPDA scenario (Pascal-Context → Cityscapes) based on different types of w.

L_{proto}			Pascal-Context → Cityscapes		
L_{CE}	L_{PPC}	L_{PPD}	Common	Private	H-Score
✓			58.21	30.74	40.23
	✓		59.49	27.66	37.76
		✓	57.17	26.76	36.96
✓	✓		58.66	31.15	40.51
✓		✓	58.39	30.50	40.32
	✓	✓	57.93	30.47	39.93
✓	✓	✓	**60.94**	**31.27**	**41.33**

Table 2. Ablation study of the components in L_{proto} on OPDA scenario (Pascal-Context → Cityscapes).

Method	Common	Private	H-Score
Ours w/o ETF	54.03	26.79	37.82
Ours	**60.94**	**31.27**	**41.33**

Table 3. Ablation studies of Fixed EFT in DSPD.

without L_{PPD}. Conversely, when L_{PPD} is used in conjunction with L_{CE}, the H-Score drops slightly to 36.96. However, using L_{PPC} and L_{PPD} together with L_{CE} yields the best result, with an H-Score of 41.33, demonstrating the importance of combining these components for optimal domain-specific clustering and performance.Finally, incorporating all three components achieves the best performance a Common mIoU of 60.94, Private IoU of 31.27, and an H-Score of 41.33. This demonstrates effectively capturing domain-specific representations and improving both common and private class performance.

B.3. Ablation Study about Fixed ETF.

To assess the importance of using a fixed Equiangular Tight Frame (ETF) in our prototype structure, we compare performance with and without the ETF constraint. As shown in Table 3), removing ETF results in a noticeable drop in performance (from 41.33 to 37.82 H-Score), confirming that maintaining distance between prototypes is critical for stable learning of domain-specific features.

TIM		Pascal-Context → Cityscapes		
Image Matching	Target RCS	Common	Private	H-Score
		59.46	27.97	38.04
✓		59.45	30.70	40.49
	✓	57.93	30.47	39.93
✓	✓	**60.94**	**31.27**	**41.33**

Table 4. Ablation study of components of TIM on OPDA scenario (Pascal-Context → Cityscapes).

T	Pascal-Context à Cityscapes		
	Common	Private	H-Score
2	56.27	29.18	38.43
1	57.79	30.70	40.10
0.1	58.66	24.89	34.95
0.05	58.32	30.35	39.92
0.01	**60.94**	**31.27**	**41.33**
0.075	55.44	27.05	36.36
0.005	57.50	12.25	20.20

Table 5. Performance comparison based on different temperature values in the OPDA scenario (Pascal-Context → Cityscapes).

C. More Experiments about TIM

C.1. Ablation Study about TIM

In Tabel 4, we conducted an ablation study to evaluate the contributions of Image Matching and Target RCS, the two components of TIM, to the overall performance in the OPDA scenario using Pascal-Context → Cityscapes. When Image Matching is applied individually, the H-Score improves from 38.04 to 40.49. Similarly, when only Target RCS is applied, the H-Score increases to 39.93. In both cases, private class performance improves significantly because the increased confidence score in the common class reduces the amount of unknown pixels that are misassigned in the pseudo label. When both components are applied together, the model achieves the highest H-Score of 41.33, indicating that Image Matching and Target RCS complement each other effectively. The synergy between the two components ensures that both common and private classes are well-represented, leading to significant overall performance improvements.

C.2. Influence of Temperature

Table 5 presented the experimental results of adjusting the temperature parameter T when calculating p_c using the softmax function. A lower temperature value emphasized rare classes by assigning higher weights, whereas a higher temperature value equalized the weights across classes. Based on the results, it was observed that setting T to 0.01 yielded the best performance with an H-Score of 41.33, effectively balancing the weighting of common rare classes.

# of Unknown	Pascal-Context $\rightarrow$ Cityscapes		
	MIC	BUS	Ours
8	9.61	10.50	**18.79**
7	13.51	30.11	**41.33**
6	11.67	18.52	**23.23**

Table 6. Comparison of H-Scores across different numbers of unknown classes in the OPDA scenario (Pascal-Context $\rightarrow$ Cityscapes). MIC used head-expansion for handling unknown classes.

λ_1	H-Score	λ_2	H-Score	τ	H-Score	τ_t	H-Score
0.05	38.43	0.05	39.72	0.15	39.42	0.953	40.80
0.03	39.22	0.03	40.37	0.13	41.04	0.958	41.04
0.02	40.48	0.02	40.22	**0.10**	**41.33**	0.963	41.19
0.01	**41.33**	**0.01**	**41.33**	0.09	40.8	**0.968**	**41.33**
0.005	40.04	0.005	40.58	0.07	39.16	0.973	41.24
0.001	40.25	0.001	41.05	0.05	36.49	0.978	40.01

Table 7. Ablation studies of parameters on the OPDA scenario.

D. Experiments about Various Unknown Proportions

Table 6 presented the experimental results comparing H-Score across MIC, BUS, and Ours under varying numbers of unknown (target-private) classes in the OPDA scenario. The experiments were conducted by designating 6, 7, and 8 unknown classes, with the following configurations:

- 8 unknown classes: “pole”, “sign”, “light”, “terrain”, “person”, “rider”, “train”, and “wall”.
- 7 unknown classes: “pole”, “sign”, “light”, “terrain”, “person”, “rider”, and “train”.
- 6 unknown classes: “pole”, “sign”, “light”, “terrain”, “person”, and “rider”.

The results demonstrate that under diverse proportions of unknown classes, Ours consistently outperforms existing methods, showcasing its robustness and superior performance in the OPDA scenario.

E. Sensitivity of Hyperparameters.

We conduct sensitivity analysis on key hyperparameters used in our method. As shown in Table 7, our method demonstrates robustness to a wide range of hyperparameter values.

F. Source-private Classes in the Pascal-Context.

We define the **43** remaining classes, excluding the common classes, as source-private classes. “aeroplane”, “bag”, “bed”, “bedclothes”, “bench”, “bird”, “boat”, “book”, “bottle”, “cabinet”, “cat”, “ceiling”, “chair”, “cloth”, “computer”, “cow”, “cup”, “curtain”, “dog”, “door”, “floor”, “flower”, “food”, “horse”, “keyboard”, “light”, “mountain”, “mouse”, “plate”, “platform”, “pottedplant”, “rock”, “sheep”, “shelves”, “sign”, “snow”, “sofa”, “table”, “track”, “tvmonitor”, “water”, “window”, and “wood”.

References