[Supplementary Material] Deliberated Domain Bridging for Domain Adaptive Semantic Segmentation

Lin Chen[1,2*] **Zhixiang Wei**[1*] **Xin Jin**[3*] **Huaian Chen**[1†] **Miao Zheng**[2] **Kai Chen**[2] **Yi Jin**[1†]
[1] University of Science and Technology of China
[2] Shanghai AI Laboratory [3] Eastern Institute for Advanced Study
{chlin, zhixiangwei}@mail.ustc.edu.cn, jinxin@eias.ac.cn, anchen@mail.ustc.edu.cn
{zhengmiao, chenkai}@pjlab.org.cn, jinyi08@ustc.edu.cn

A Appendix

In the supplementary material, we provide more experimental results and summary them as follows:

- In Sec. A.1, we utilize UMAP [4] to generate more visualization results for a comprehensive analysis of the feature space learned by different models.
- In Sec. A.2, we provide a comprehensive training procedure of our proposed method.
- In Sec. A.3, we provide detailed ablation studies on the key components of our proposed method w.r.t single-source, multi-source, and multi-target benchmarks.
- In Sec. A.4, we compare our proposed method with previous state-of-the-art methods by qualitative results on the GTA5 $\rightarrow$ Cityscapes benchmark.
- In Sec. A.5, we provide a detailed comparison of the training efficiency with previous state-of-the-art methods on the GTA5 $\rightarrow$ Cityscapes benchmark.
- In Sec. A.6, we further discuss the limitations and potential negative impacts of our proposed method.

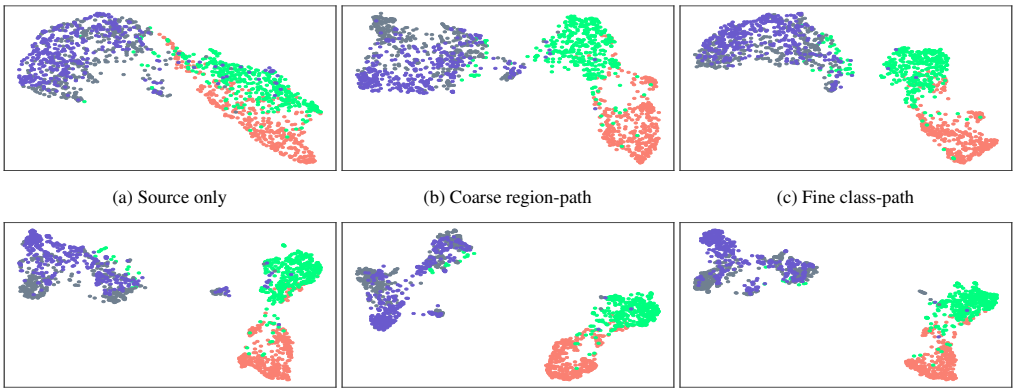
A.1 Visualization of Feature Space

To better understand the intuitions behind the proposed method, we utilize UMAP [4] to visualize the target feature representations before the final classification layer on the GTA5 $\rightarrow$ Cityscapes benchmark. As shown in Fig. 1, the feature representation of the model trained on the source domain is not compact enough within each class, and the inter-class distance of each class pair is also relatively small. While models trained on the coarse region-path (CRP) and fine class-path (FCP) have comparable performance, their feature distributions are quite different. In FCP, due to more attention paid to the object's inherent properties, the feature representation within each class is more compact than in CRP. However, the representations of 'bus' and 'train' in FCP are seriously overlapping. The model trained on CRP tends to exploit the contextual information to discriminate the confused classes, which drives to more separable representations of 'bus' and 'train.'

As depicted in the bottom row of Fig. 1, the student models integrating knowledge from two complementary teacher models not only have more compact feature representations than the teacher model in FCP but also mitigate the overlapping issue of the confused 'bus' and 'train' classes. Furthermore, the alternating optimization strategy can consistently further enhance such properties of the student model.

[*] Equal contribution.
[†] Correspoding author.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).



(a) Source only (b) Coarse region-path (c) Fine class-path

(d) Student of stage1 (e) Student of stage2 (f) Student of stage3

Figure 1: UMAP [4] visualization of the feature space before the classification layer on GTA5 → Cityscapes. For a clear analysis, we select two pairs of categories suffering class bias and confusion, respectively, *i.e.*, salmon for road, green for sidewalk, gray for bus, and purple for train.

A.2 Algorithm

The training procedure of our DDB is summarized in Algorithm 1, which is composed of multiple rounds of alternating dual-path domain bridging (DPDB) and cross-path knowledge distillation (CKD) steps. For detailed equations and loss functions, please refer to the main paper.

Algorithm 1: Training process of DDB

Input: training dataset: (X_s, Y_s, X_t); batch data: (x_s, y_s, x_t); teacher models: M_C^1, M_F^1 and student model: M_S^1; EMA teacher models: M'^r_C, M'^r_F; EMA momentum: α; alternating rounds: R.
Output: final student model: M_S^R.

for $r \leftarrow 1$ **to** R **do**
  # Dual-path Domain Bridging
  EMA models initialization: $\theta_{M'^r_C} \leftarrow \theta_{M_C^r}, \theta_{M'^r_F} \leftarrow \theta_{M_F^r}$;
  for $k \leftarrow 1$ **to** *max_iterations* **do**
    Get source images x_s, target images x_t;
    Get CRP bridging images x_{crp}, labels $\hat{y}_{crp}$ by Eqn. 3
    Get pseudo-label weight maps m_{crp} by Eqn. 6;
    Optimize M_C^r by Eqn. 8;
    Update EMA model: $\theta^{k+1}_{M'^r_C} \leftarrow \alpha\theta^k_{M'^r_C} + (1-\alpha)\,\theta^k_{M_C}$;
    Get FCP bridging images x_{fcp}, labels $\hat{y}_{fcp}$ by Eqn. 3
    Get pseudo-label weight maps m_{fcp} by Eqn. 6;
    Optimize M_F^r by Eqn. 8;
    Update EMA model: $\theta^{k+1}_{M'^r_F} \leftarrow \alpha\theta^k_{M'^r_F} + (1-\alpha)\,\theta^k_{M_F}$;
  # Cross-path Knowledge Distillation
  Calculate the prototypes η_C^r, η_F^r on X_t by Eqn. 10;
  for $k \leftarrow 1$ **to** *max_iterations* **do**
    Get source images x_s;
    Get clean and augmented target images x_t, x_t^{aug};
    Utilize prototypes to get adaptive weight maps w_C, w_F by Eqn. 11;
    Get ensembled pseudo-label $\bar{y}_t$;
    Optimize M_S^r by Eqn. 14;
  if $r \neq R$ **then**
    Teacher models initialization for next round: $\theta_{M_C^{r+1}} \leftarrow \theta_{M_S^r}, \theta_{M_F^{r+1}} \leftarrow \theta_{M_S^r}$;

Table 1: Ablation studies on the momentum of EMA in the DPDB step (fine-class path) on the GTA5 (G) → Cityscapes (C) benchmark.

Momentum	mIoU
0.9	58.3±0.4
0.99	58.5±0.3
0.999	**58.4±0.3**
0.9999	54.2±0.1

Table 2: Ablation studies on the augmentation strategy for the input of the student model in the first CKD step on the GTA5 (G) → Cityscapes (C) benchmark.

Gaussian blur	Color jitter	mIoU
		61.0±0.2
✓		61.2±0.2
✓	✓	**61.5±0.3**

A.3 Detailed Ablation Studies

Study on Baseline. As mentioned in Sec.3.2 of the main paper, we follow [5, 1] to construct a simple self-training pipeline (pseudo labeling) to validate the effectiveness of various DB methods in the DASS task and treat it as our baseline. As shown in Tab. 3, Tab. 4, and Tab. 5, the model trained on baseline tends to give over confidences to some easily discriminated categories, resulting in performance converging to 0 in other categories. Specifically, due to more labeled data in the source domain, the model under the multi-source setting has a extremely over-confidence w.r.t the easily discriminated categories, resulting in a performance decrease of 8.5 in mIoU.

Study on Dual-path Domain Bridging. As shown in Tab. 3, Tab. 4, and Tab. 5, the source-only and baseline models all perform poorly on three benchmarks. Benefited from the proposed dual-path domain bridging scheme, the adapted model on each path can obtain gains at least 24.4, 15.3, and 19.7 in mIoU separately. Although the CRP and FCP models achieve comparable performance, they have pretty different behaviors in each class. For example, on the GTA5 → Cityscapes (single-source) benchmark, the CRP model achieves 39.5 in the 'train' class while the FCP model only achieves 0.0. This is because the CRP model tends to exploit contextual information to mitigate the class confusion issue. Due to the lack of contextual information, the FCP model pays more attentions on the inherent object properties, mitigating the class bias, such as 'sidewalk' and 'terrain' classes. Moreover, as shown in Tab. 1, we further conducted ablation experiments on α, and found that $\alpha = 0.99$ will lead to better and more stable results.

Study on Cross-path Knowledge Distillation. As illustrated in Tab. 3 and Tab. 4, the hard distillation is more effective than the soft manner in the single-source and multi-source domain settings. Especially in the multi-source setting (see Tab. 4), the hard distillation can improve model by 2.1 in mIoU than the soft one. Furthermore, the proposed adaptive ensemble can consistently improve the performance of both soft and hard distillation schemes in all three settings. By leveraging CKD that equipped with the hard distillation and adaptive ensemble schemes, we can consistently obtain a more superior student model than its teachers in all three benchmarks. Especially in the multi-target setting (see Tab. 5), the student model gets gains of 3.6 and 4.4 in mIoU than the CRP and FCP teacher models, separately. Moreover, we further conducted ablation experiments about the augmentation strategies. As shown in Tab. 2, combining the Gaussian blur and color jitter augmentation techniques leads to the best performance.

Influence of Alternating Optimization Strategy. As illustrated in Tab. 3, Tab. 4, and Tab. 5, after integrating knowledge from two complementary teacher models, the student model performs better than each teacher model. Once a superior student is obtained, we further conduct next DPDB step and utilize the weights of the student model to initialize the teacher models in DPDB to obtain two more powerful teachers. The student model performs best across all three domain settings in the second round. However, the student model shows a slight performance degradation after the third round of alternate training in the multi-source and multi-target domain settings. We analyse the degradation is because the non-negligible domain conflict in these two settings.

Table 3: Ablation studies on the key components of our proposed method on the GTA5 → Cityscapes benchmark.

	components		road	sidewalk	building	wall	fence	pole	light	sign	vege.	terrain	sky	person	rider	car	truck	bus	train	motor	bike	mIoU	gain
baseline	source only		60.4	15.1	58.3	8.7	21.3	20.9	33.2	22.4	77.7	8.6	71.3	55.8	13.2	77.0	22.8	22.1	0.4	14.1	6.1	32.1	
	pseudo labeling		92.5	52.6	80.7	6.7	3.1	0.0	0.8	0.6	82.7	36.2	86.1	53.1	0.0	82.7	21.9	41.9	0.0	0.8	0.0	33.8	+1.7
stage1 DPDB	region path	class path																				mIoU	gain
	✓		92.2	51.5	86.6	39.4	36.4	28.1	49.1	41.1	84.4	29.1	86.1	70.8	43.4	90.0	53.6	55.6	39.5	41.0	56.6	56.5	+24.4
		✓	93.3	54.9	88.2	37.6	43.3	41.2	52.0	49.6	88.3	45.7	88.2	70.8	35.3	91.5	63.1	62.4	0.0	45.0	54.4	58.2	+26.1
stage1 CKD	hard distillation	adaptive ensemble																				mIoU	gain
			93.4	55.9	87.6	44.5	43.0	35.4	51.1	45.5	88.4	44.4	88.6	70.9	44.9	91.5	55.1	56.5	15.9	48.1	60.2	59.0	+26.9
	✓		94.1	59.9	88.1	44.2	43.9	36.8	52.9	47.8	88.3	45.3	88.7	70.9	42.9	91.6	66.0	63.2	3.5	49.9	60.4	59.9	+27.8
		✓	94.4	63.3	88.8	43.8	43.5	38.2	54.9	51.4	87.5	38.7	89.2	71.1	43.2	91.1	63.6	65.7	24.1	46.0	62.6	61.1	+29.0
	✓	✓	94.6	64.0	88.9	43.9	44.0	39.6	54.7	52.4	88.0	41.1	89.2	71.6	42.6	91.9	61.4	63.9	22.4	46.7	62.7	61.2	+29.1
stage2 DPDB	region path	class path																				mIoU	gain
	✓		94.4	62.5	88.1	41.7	43.2	35.9	54.0	50.7	87.3	40.8	85.3	70.5	43.4	92.2	67.3	63.9	33.4	49.1	63.3	61.4	+29.3
		✓	94.8	65.0	88.9	36.6	46.0	39.3	52.9	54.0	88.5	46.7	88.2	71.1	41.8	92.4	73.5	68.8	29.5	50.1	62.4	62.6	+30.5
stage2 CKD	hard distillation	adaptive ensemble																				mIoU	gain
	✓	✓	95.3	67.4	89.3	44.4	45.7	38.7	54.7	55.7	88.1	40.7	90.7	70.7	43.1	92.2	60.8	67.6	34.5	48.7	63.7	62.7	+30.6
stage3 DPDB	region path	class path																				mIoU	gain
	✓		94.9	65.8	88.1	36.4	44.5	35.5	53.9	54.3	87.1	36.0	88.5	70.8	42.4	92.0	61.6	67.3	38.8	48.6	63.6	61.6	+29.5
		✓	95.1	66.2	88.9	39.2	46.4	39.7	53.9	54.4	88.3	42.2	90.0	71.0	42.4	91.8	63.1	69.8	35.7	50.2	63.7	62.7	+30.6
stage3 CKD	hard distillation	adaptive ensemble																				mIoU	gain
	✓	✓	95.1	67.2	88.9	39.2	44.8	38.2	54.6	54.7	87.6	37.5	91.2	71.2	44.0	91.8	62.7	70.0	40.1	48.8	63.8	62.7	+30.6

Table 4: Ablation studies on the key components of our proposed method on the GTA5 + Synscapes → Cityscapes benchmark.

	components		road	sidewalk	building	wall	fence	pole	light	sign	vege.	terrain	sky	person	rider	car	truck	bus	train	motor	bike	mIoU	gain
baseline	source only		82.5	42.4	79.0	27.2	31.7	40.8	53.0	45.6	85.3	30.9	80.9	68.7	35.7	78.3	39.0	42.7	9.6	37.3	55.9	50.9	
	pseudo labeling		94.5	65.8	81.1	19.7	0.5	0.1	42.1	31.6	82.7	34.3	87.8	53.0	0.2	82.3	4.7	47.5	2.4	20.9	54.4	42.4	-8.5
stage1 DPDB	region path	class path																				mIoU	gain
	✓		96.0	70.0	88.3	45.5	46.7	44.0	60.1	62.4	88.3	43.6	91.8	74.0	51.6	90.8	57.1	73.2	57.1	56.0	69.9	66.7	+15.8
		✓	96.4	72.8	89.0	48.1	47.0	46.5	57.8	64.8	88.8	49.2	92.4	72.9	49.5	92.3	65.6	70.2	31.3	54.3	69.1	66.2	+15.3
stage1 CKD	hard distillation	adaptive ensemble																				mIoU	gain
			95.2	67.5	89.0	47.5	49.8	43.3	58.6	63.1	88.2	40.0	92.8	72.8	50.3	91.7	61.5	65.6	22.2	55.7	69.9	64.5	+13.6
	✓		95.4	66.9	89.9	54.2	52.5	47.0	61.1	64.6	89.6	50.5	93.2	75.2	52.0	92.6	67.2	69.8	17.6	55.8	70.3	66.6	+15.7
		✓	94.9	69.1	88.5	49.6	47.6	41.9	57.9	62.9	89.2	46.7	93.1	72.9	49.3	91.3	55.8	68.5	32.8	50.2	69.4	64.8	+13.9
	✓	✓	96.7	74.9	90.1	55.0	51.2	47.6	60.6	66.0	89.7	49.4	93.3	74.4	52.3	92.4	62.8	73.9	47.2	55.7	71.0	68.6	+17.7
stage2 DPDB	region path	class path																				mIoU	gain
	✓		96.6	73.2	89.4	49.7	49.9	46.6	61.5	64.8	89.5	45.9	92.4	74.8	52.9	92.6	62.9	77.2	61.9	57.5	71.7	69.0	+18.1
		✓	96.9	75.5	89.6	48.7	50.6	47.2	61.0	66.0	89.4	50.3	92.6	74.2	52.3	92.4	65.7	74.3	49.1	59.0	71.0	68.7	+17.8
stage2 CKD	hard distillation	adaptive ensemble																				mIoU	gain
	✓	✓	96.9	75.6	90.0	54.4	48.6	47.6	61.1	66.3	89.7	48.4	93.4	74.4	52.7	92.3	60.8	74.7	58.9	53.9	71.4	69.0	+18.1
stage3 DPDB	region path	class path																				mIoU	gain
	✓		96.7	73.2	89.4	48.2	49.3	46.6	61.6	65.4	89.3	42.7	92.4	74.3	52.7	92.5	64.5	75.2	62.6	56.2	71.4	68.6	+17.7
		✓	97.0	75.1	89.4	50.6	49.4	46.9	61.1	66.3	89.1	45.4	92.8	73.5	52.1	92.1	64.4	75.8	58.1	58.1	70.5	68.8	+17.9
stage3 CKD	hard distillation	adaptive ensemble																				mIoU	gain
	✓	✓	96.9	75.2	89.8	54.4	48.8	47.0	60.6	65.7	89.4	46.1	93.2	74.0	52.3	92.4	61.5	75.4	59.6	53.1	71.3	68.8	+17.9

Table 5: Ablation studies on the key components of our proposed method on the GTA5 → Cityscapes + Mapillary benchmark. C denotes the Cityscapes dataset, and M denotes the Mapillary dataset.

	components		target	road	sidewalk	building	wall	fence	pole	light	sign	vege.	terrain	sky	person	rider	car	truck	bus	train	motor	bike	mIoU	avg.	gain
baseline	source only		C	53.3	15.2	56.6	8.2	26.2	21.2	30.7	22.2	76.3	9.3	53.3	55.3	15.5	72.9	21.5	4.9	0.9	20.2	7.4	30.1	32.8		
			M	55.7	27.1	55.3	9.9	20.6	22.7	33.3	31.6	68.4	21.1	70.6	53.5	30.9	72.7	32.3	11.6	5.6	36.3	14.9	35.5		
	pseudo labeling		C	87.0	21.2	78.6	15.7	0.0	1.4	13.4	1.9	84.4	35.2	83.7	50.0	0.0	85.1	30.9	36.5	0.0	0.4	0.0	32.9	33.1	+0.3
			M	80.4	29.9	71.4	12.3	3.1	2.2	19.4	3.1	79.0	40.3	95.4	46.4	0.6	78.3	38.2	24.6	0.0	7.3	0.0	33.3		
	region path	class path	target																				mIoU	avg.	gain
stage1 DPDB	✓		C	93.3	58.3	86.3	33.2	36.4	31.3	49.5	45.3	87.7	44.8	88.8	69.1	39.8	88.1	43.7	49.6	18.3	35.0	50.0	55.2	53.3	+20.5
			M	73.9	54.5	79.0	27.6	32.2	30.5	45.8	43.8	71.4	38.0	78.9	61.5	45.6	84.6	52.7	52.2	21.5	48.1	35.9	51.5		
		✓	C	61.7	47.5	87.4	29.9	37.4	35.7	51.4	53.8	88.0	47.1	89.8	67.2	31.1	36.2	59.6	54.0	8.6	33.4	46.0	50.8	52.5	+19.7
			M	86.0	53.3	81.7	28.8	36.1	31.5	46.9	48.4	71.8	45.3	90.7	62.9	50.1	85.3	57.2	54.9	16.9	47.9	32.8	54.1		
	hard distillation	adaptive ensemble	target																				mIoU	avg.	gain
stage1 CKD			C	89.3	52.4	87.8	38.0	40.2	33.4	53.2	51.5	88.5	48.7	90.3	69.1	39.7	83.7	57.3	53.4	2.9	39.5	51.1	56.3	55.9	+23.1
			M	86.8	56.5	81.6	32.6	37.3	31.6	47.1	46.9	75.2	44.3	91.8	63.5	49.6	86.8	58.1	55.9	19.0	52.7	35.3	55.4		
	✓		C	88.8	52.8	87.8	47.8	40.2	34.3	51.3	50.3	88.8	48.2	90.2	68.8	38.4	80.4	57.4	55.5	0.1	38.9	53.8	56.5	55.8	+23.0
			M	87.1	56.3	82.0	33.4	36.8	31.2	46.6	47.2	74.4	43.9	91.9	62.4	49.8	86.5	58.4	52.8	10.8	50.4	42.1	55.0		
		✓	C	87.6	58.3	88.4	42.3	42.7	35.2	53.5	55.7	88.6	46.7	91.2	68.7	41.1	71.1	50.1	58.3	8.8	38.2	54.8	56.9	56.8	+24.0
			M	87.6	57.6	82.0	33.1	37.4	33.8	47.8	49.8	73.0	40.7	91.9	63.2	51.1	85.7	57.1	60.1	29.1	52.3	44.7	56.7		
	✓	✓	C	89.6	58.7	88.3	40.1	44.7	34.8	53.9	54.4	88.8	47.2	90.9	69.4	41.1	77.2	51.9	57.3	6.4	34.9	55.0	57.1	56.9	+24.1
			M	88.3	57.8	81.8	33.8	37.8	33.2	47.8	48.4	73.7	41.9	91.9	63.2	49.9	86.3	59.2	61.3	27.2	51.7	40.4	56.6		
	region path	class path	target																				mIoU	avg.	gain
stage2 DPDB	✓		C	93.9	65.5	87.1	31.5	43.1	30.8	53.2	53.9	88.4	49.1	89.8	69.6	42.0	91.3	60.0	56.9	12.9	42.2	59.3	59.0	57.7	+24.9
			M	85.2	60.4	81.5	32.0	36.7	32.3	47.5	48.8	71.0	38.6	89.9	63.7	51.5	86.4	62.9	63.3	27.8	52.4	40.2	56.4		
		✓	C	93.2	67.2	88.6	41.5	45.4	32.5	53.0	57.3	89.0	49.0	90.8	68.7	40.4	82.5	63.9	57.5	0.8	41.7	58.5	59.0	57.9	+25.1
			M	89.7	58.6	82.3	34.7	36.9	32.7	47.8	49.9	70.3	38.2	91.0	63.6	52.3	86.3	64.2	65.7	20.0	53.6	51.3	56.8		
	hard distillation	adaptive ensemble	target																				mIoU	avg.	gain
stage2 CKD	✓	✓	C	93.5	67.8	88.3	38.4	45.6	32.3	54.2	57.9	89.2	48.6	91.6	69.1	43.2	84.6	63.6	61.8	15.1	44.1	58.6	60.4	58.6	+25.8
			M	89.3	60.8	81.4	35.9	38.4	32.9	48.5	50.5	69.9	37.9	90.1	62.6	49.6	86.0	62.7	62.9	26.1	52.0	42.8	56.9		
	region path	class path	target																				mIoU	avg.	gain
stage3 DPDB	✓		C	93.0	63.0	86.9	28.5	44.9	28.1	53.9	53.2	88.8	50.0	90.7	67.4	41.4	90.7	66.0	61.0	18.2	41.6	43.9	58.5	57.6	+24.8
			M	88.6	62.5	81.8	34.2	36.7	31.8	48.2	48.7	68.7	34.1	90.5	64.0	51.1	86.3	63.3	59.2	29.6	53.6	43.2	56.6		
		✓	C	92.7	64.0	88.1	27.7	46.2	31.8	54.4	58.7	88.9	50.3	91.2	68.5	40.9	86.3	68.9	59.9	0.6	43.9	49.3	58.5	57.6	+24.8
			M	90.4	62.6	82.4	34.7	36.4	32.5	48.2	50.3	67.4	35.7	90.5	63.6	51.9	85.7	64.7	56.1	22.9	54.7	44.3	56.6		
	hard distillation	adaptive ensemble	target																				mIoU	avg.	gain
stage3 CKD	✓	✓	C	92.6	63.4	87.6	39.2	45.1	29.1	54.5	57.7	88.9	49.5	91.6	67.6	43.3	86.8	58.0	64.5	26..3	44.8	40.7	59.5	58.2	+25.4
			M	90.1	61.8	81.9	37.1	36.6	32.3	48.7	50.3	67.8	35.6	90.8	63.4	50.6	86.0	61.9	58.8	26.2	53.9	45.2	56.8		

A.4 Comparison of Qualitative Results

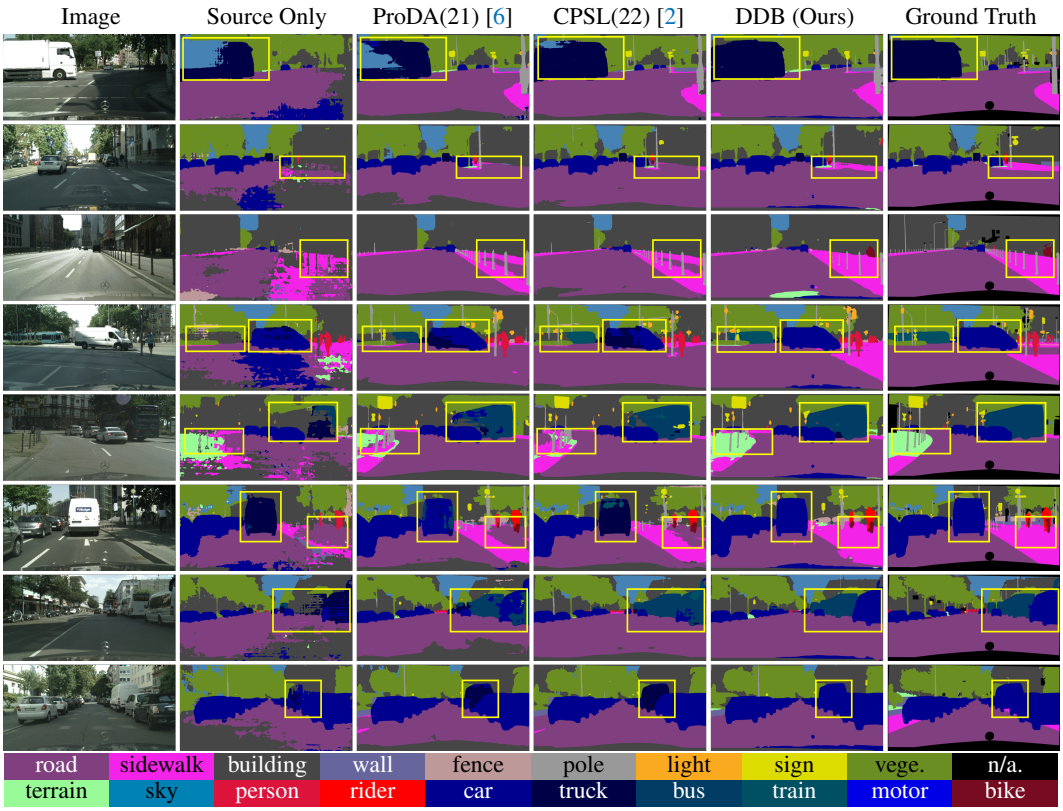


Figure 2: Qualitative comparisons of recent state-of-the-art methods on GTA5 → Cityscapes.

A.5 Comparison of Training Efficiency

As illustrated in Tab. 6, our method achieved better performance after one round of training with smaller input size, fewer GPUs, fewer iterations, and fewer post-processing steps than other SOTA methods. Furthermore, our method could still achieve gains of 1.5 mIoU with one more training round, while still being more efficient than other SOTA methods.

Table 6: Comparison of the training efficiency on the GTA5 → Cityscapes benchmark.

Method	Input size	Total GPUs	Total iterations	Pseudo-label generation	Prototype generation	mIoU
ProDA+distill [6]	(896,512)	4	238K	✓	✓	57.5
UndoDA+distill [3]	(896,512)	4	238K	✓	✓	59.3
CPSL+distill [2]	(896,512)	4	238K	✓	✓	60.8
DDB-round1	(512,512)	1	80K	×	✓	61.2
DDB-round2	(512,512)	1	160K	×	✓	**62.7**

A.6 Limitations

Although our proposed approach achieves impressive performance on the single-source, multi-source, and multi-target domain settings, it still requires multiple training rounds to conduct alternating optimization processes of the proposed dual-path domain bridging and cross-path knowledge distillation steps. We will explore an end-to-end optimization approach in future work. Moreover, the proposed method might be used in undesirable applications like surveillance or military UAVs for the purpose of domain adaptive semantic segmentation. Legal limitations on the applications of semantic segmentation algorithms could be a potential defense.

References

[1] L. Hoyer, D. Dai, and L. Van Gool. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *CVPR*, 2022.

[2] R. Li, S. Li, C. He, Y. Zhang, X. Jia, and L. Zhang. Class-balanced pixel-level self-labeling for domain adaptive semantic segmentation. In *CVPR*, 2022.

[3] Y. Liu, J. Deng, J. Tao, T. Chu, L. Duan, and W. Li. Undoing the damage of label shift for cross-domain semantic segmentation. In *CVPR*, 2022.

[4] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv*, 2018.

[5] W. Tranheden, V. Olsson, J. Pinto, and L. Svensson. Dacs: Domain adaptation via cross-domain mixed sampling. In *WACV*, pages 1379–1389, 2021.

[6] P. Zhang, B. Zhang, T. Zhang, D. Chen, Y. Wang, and F. 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.