DUDA: Distilled Unsupervised Domain Adaptation for Lightweight Semantic Segmentation

Beomseok Kang[†,*], Niluthpol Chowdhury Mithun[‡], Abhinav Rajvanshi[‡], Han-Pang Chiu[‡], Supun Samarasekera[‡]

[†]Seoul National University, Seoul, South Korea
[†]beomseok@snu.ac.kr

[‡]SRI International, Princeton, NJ, USA
[‡]firstname.lastname@sri.com

1. Overview

This is the supplementary material to support our manuscript "DUDA: Distilled Unsupervised Domain Adaptation for Lightweight Semantic Segmentation". It contains additional and detailed results, particularly related to Sec. 4 (Experimental Results) of the main article, that couldn't be included in the main article due to space constraints. In Sec. 3, we note references compared in Fig. 1 in the main paper. In Sec. 4, we provide the training time of DUDA_{MIC} with the various backbone models as an additional implementation detail. In Sec. 5, in addition to Tab. 4 of the main paper, we provide more detailed ablation studies using MiT-B0 backbone across the four different datasets. In Sec. 6, we compare DUDA models with the SOTA UDA methods with the class-wise IoU across four UDA different benchmarks. We also provide experiments comparing DUDA and its baselines (*i.e.*, DAFormer, MIC) with MiT-B2 and MiT-B4 backbones. Additionally, we provide class-wise IoU scores for the compared ResNet-based methods in Tab. 2 of the main paper. Code: https://github.com/beomseokg/DUDA.

2. Implementation Details

Our implementation closely follows SOTA prior works, DAFormer [3] and MIC [5], including backbone and decoder head structures, as well as learning rate, batch size, optimizer, etc. The backbones are pre-trained on ImageNet-1k. Training strategies utilized in DAFormer [3], such as data augmentation, learning rate warm-up, ImageNet feature distance (FD), and rare class sampling (RCS), are integrated into DUDA, but omitted in Fig. 2 (main paper) for simplicity. In addition, the MIC training strategy, such as masked loss, is applied within the DUDA framework when coupled with MIC. We allocate 40k iterations for the pre-adaptation phase and 80k iterations for the fine-tuning. We experiment with 3 different random seeds.

[*]Most of the work was done during BK's SRI International internship.

Method	Backbone	Training Time (iterations)	
		Pre-adaptation (40k)	Fine-tuning (80k)
DUDA_{MIC}	MiT-B0	39 hours	*41 hours
DUDA_{MIC}	MiT-B1	40 hours	*45 hours
DUDA_{MIC}	MiT-B2	43 hours	46 hours
DUDA_{MIC}	MiT-B4	49 hours	58 hours
DUDA_{MIC}	DeepLab-V2	54 hours	*73 hours

Table 1. Training time of DUDA_{MIC} in the various backbones. The training is performed in a single NVIDIA A6000 or A5000 GPU. *Training time is measured in NVIDIA A5000 GPU.

3. Compared Methods

We compare with the following methods in Fig. 1 in the main paper: FDA [25], DACS [16], ProDA [26], CAMix [29], DAFormer [3], HRDA [4] MIC [5], DiGA [15], Fredom [17], SGG [12], CONFETI [7], RTea [27], PRN [28], MoDA [11], and RDAS (*i.e.*, Revisiting Domain Adaptive Semantic Segmentation) [6].

4. Training and Inference Cost

Training Cost. We primarily measure the training time for each of the two training procedures. Our training is performed on a single GPU, either NVIDIA A5000 or A6000, and other training parameters, such as batch size, are the same as introduced in the main text. Tab. 1 summarizes the training time for different backbone models trained by DUDA_{MIC} using the GTA→Cityscapes dataset. Since DUDA is operated on three different networks (LT, LS, and SS) and consists of the two training stages (pre-adaptation and fine-tuning), the training cost is more expensive than that of its baseline, such as HRDA [4] and MIC [5]. For peak memory, DUDA MIC with MiT-B0, B1, B2, B4, and B5 requires 43,520MB, 43,561MB, 43,601MB, 43,742MB, and 43,820MB, respectively—all of which fit comfortably on a single A6000 GPU. In comparison, standard MIC with MiT-B5 consumes 22,444MB. Regarding throughput, in the

Method	Pre-adaptation	Fine-tuning CE	KL	Inconsistency	mIoU (%)	mAcc (%)
Synthetic-to-Real: GTA→Cityscapes (Val.)						
MiT-B0	No Distillation				51.00	62.49
MiT-B0		✓			62.34	71.74
MiT-B0		✓	✓		63.67	72.94
MiT-B0	✓	✓	✓		64.38	73.50
MiT-B0	✓	✓	✓	✓	65.19	75.18
Synthetic-to-Real: Synthia→Cityscapes (Val.)						
MiT-B0	No Distillation				46.09	58.25
MiT-B0		✓			57.06	67.44
MiT-B0		✓	✓		57.01	68.44
MiT-B0	✓	✓	✓		57.72	69.20
MiT-B0	✓	✓	✓	✓	58.31	71.12
Day-to-Nighttime: Cityscapes→DarkZurich (Val.)						
MiT-B0	No Distillation				23.89	40.08
MiT-B0		✓			33.18	49.02
MiT-B0		✓	✓		34.30	50.49
MiT-B0	✓	✓	✓		35.18	51.07
MiT-B0	✓	✓	✓	✓	35.29	51.84
Clear-to-Adverse-Weather: Cityscapes→ACDC (Val.)						
MiT-B0	No Distillation				43.79	58.06
MiT-B0		✓			49.68	62.91
MiT-B0		✓	✓		51.84	65.00
MiT-B0	✓	✓	✓		53.52	66.11
MiT-B0	✓	✓	✓	✓	53.86	67.45

Table 2. Ablation Studies of DUDA on the four different adaptation scenarios with DAFormer as the base across the four datasets. The basic model is configured as DAFormer without DUDA, and subsequently, we incrementally introduce cross-entropy, KL divergence, pre-adaptation, and inconsistency-based balanced losses in the DUDA setup. mIoU for Synthia→Cityscapes is averaged over 16 classes.

Method	Pre-adaptation	Fine-tuning CE	KL	Inconsistency	mIoU (%)	mAcc (%)
Synthetic-to-Real: GTA→Cityscapes (Val.)						
MiT-B0	No Distillation				59.54	69.29
MiT-B0		✓			71.37	79.99
MiT-B0		✓	✓		70.80	79.82
MiT-B0	✓	✓	✓		71.63	80.27
MiT-B0	✓	✓	✓	✓	71.71	81.04
Synthetic-to-Real: Synthia→Cityscapes (Val.)						
MiT-B0	No Distillation				52.96	64.26
MiT-B0		✓			64.79	75.03
MiT-B0		✓	✓		64.56	74.93
MiT-B0	✓	✓	✓		65.02	75.17
MiT-B0	✓	✓	✓	✓	65.25	76.04
Day-to-Nighttime: Cityscapes→DarkZurich (Val.)						
MiT-B0	No Distillation				31.38	47.12
MiT-B0		✓			40.35	60.00
MiT-B0		✓	✓		40.93	60.74
MiT-B0	✓	✓	✓		41.28	60.31
MiT-B0	✓	✓	✓	✓	40.83	60.75
Clear-to-Adverse-Weather: Cityscapes→ACDC (Val.)						
MiT-B0	No Distillation				52.82	64.89
MiT-B0		✓			63.86	74.23
MiT-B0		✓	✓		64.13	74.43
MiT-B0	✓	✓	✓		65.48	75.22
MiT-B0	✓	✓	✓	✓	65.39	75.73

Table 3. Ablation Studies of DUDA on the four different adaptation scenarios with MIC as the base across the four datasets. The basic model is configured as MIC without DUDA, and subsequently, we incrementally introduce cross-entropy, KL divergence, pre-adaptation, and inconsistency-based balanced losses in the DUDA setup. mIoU for Synthia→Cityscapes is averaged over 16 classes.

pre-adaptation stage of DUDA MIC, where three models are jointly updated, the measured throughput on an A6000 GPU is 0.570 imgs/sec (MiT-B0), 0.556 imgs/sec (MiT-B1), 0.517 imgs/sec (MiT-B2), and 0.454 imgs/sec (MiT-B4). These values are lower than the 0.794 imgs/sec achieved by standard MIC with MiT-B5, reflecting the additional computation during joint training. As LT and LS networks are identical regardless of the SS network's backbone, the increase in the training time (both pre-adaptation and fine-tuning) has resulted from the larger SS networks (from top to bottom rows).

Inference Cost. We acknowledge the elevated memory requirement of DUDA during training due to the auxiliary large network, however, it incurs no increase in inference cost. Notably, the primary obstacle from memory issues predominantly emerges at inference time. Considering the other UDA methods [10, 25, 26] take several days [4], the training speed of DUDA is rather similar to them. However, our focus is to obtain SOTA UDA performance in lightweight models (SS networks), not fast training. We can rather expect fast inference as a by-product of lightweight models since DUDA reduces the memory by 1.3∼11.7 times and FLOPs of the backbone by 1.3∼21.2, keeping the mIoU comparable (Tab. 1 of the main paper). Note, the memory and FLOPs of ResNet models are as follows: ResNet-18 (46.2MB and 176.0 GFLOPs), ResNet-50 (99.7MB and 366.1GFLOPs), and ResNet-101 (175.7MB and 638.7 GFLOPs). FLOPs are measured using the same method as in Tab. 1 in the main paper.

In summary, while DUDA's training cost is higher than that of the baseline methods, the inference cost remains the same.

5. Ablation Study

Ablation studies are conducted in MiT-B0 backbone trained by DUDA_{DAF} and DUDA_{MIC} across the four different datasets. Similar to Tab. 4 of the main paper, the performance improvement by involving the four components, pre-adaptation (Pre-adapt) and fine-tuning with the cross-entropy (CE), KL divergence losses (KL), and inconsistency-based loss balancing (Incon.), are investi-

Method	Pre-adaptation	Fine-tuning CE	KL	Inconsistency	mIoU (%)
MiT-B1	No Distillation				60.2
MiT-B1	✓				64.7
MiT-B1		✓			66.8
MiT-B1		✓	✓		67.9
MiT-B1	✓	✓	✓		68.4
MiT-B1	✓	✓	✓	✓	68.5
MiT-B2	No Distillation				63.9
MiT-B2	✓				65.9
MiT-B2		✓			68.4
MiT-B2		✓	✓		68.5
MiT-B2	✓	✓	✓		69.5
MiT-B2	✓	✓	✓	✓	69.8
MiT-B4	No Distillation				66.1
MiT-B4	✓				68.2
MiT-B4		✓			70.0
MiT-B4		✓	✓		70.2
MiT-B4	✓	✓	✓		70.4
MiT-B4	✓	✓	✓	✓	70.5

Table 4. Ablation Studies of DUDA in GTA→Cityscapes with DAFormer. Experimental setups are same with Tab. 2 but MiT-B1, MiT-B2 and MiT-B4 models are used for students.

Method		mIoU (%) ↑ GTA→CS	SYN→CS	CS→DZur	CS→ACDC
DAFormer	MiT-B2	63.9	59.4	44.7	52.2
DUDA$_{DAF}$	MiT-B2	**69.8**	**61.6**	**54.4**	**56.8**
MIC	MiT-B2	72.7	66.4	56.5	60.9
DUDA$_{MIC}$	MiT-B2	**75.5**	**67.3**	**59.6**	**68.6**
DAFormer	MiT-B4	66.1	59.9	46.9	55.5
DUDA$_{DAF}$	MiT-B4	**70.5**	**62.0**	**54.4**	**57.7**
MIC	MiT-B4	74.3	66.0	59.9	64.0
DUDA$_{MIC}$	MiT-B4	**76.7**	**68.3**	**60.3**	**70.2**

Table 5. Comparison with DAFormer and MIC in MiT-B2 and B4 backbones. mIoU for SYN→CS is averaged over 16 classes.

gated. Tabs. 2 and 3 provide the results from DUDA$_{DAF}$ and DUDA$_{MIC}$, respectively. DUDA$_{DAF}$ with the four components consistently achieves the highest mIoU and mAcc in the four datasets. DUDA$_{MIC}$ with the four components in CS→DZur and CS→ACDC shows slightly lower mIoU but highest mAcc.

In summary, DUDA generally shows continuous improvement by including each of the components.

6. Additional Quantitative Results

Comparison with Transformer-based Methods. In Tab. 3 of the main paper, we compare DUDA and its baselines (DAFormer and MIC) in the MiT-B0 and B1 backbones in terms of mIoU scores. Similarly, the comparison with DAFormer and MIC in the MiT-B2 and M4 backbones is presented in Tab. 5. Additionally, we provide the class-wise comparison in Tabs. 6 and 7. It is important to note that, in GTA→CS with MiT-B0 backbone, the methods without DUDA completely fail to segment the Train class, while DUDA$_{DAF}$ and DUDA$_{MIC}$ improve it by the IoU of almost 50% and 70%. Similarly, the accuracy of the Motorbike class shows a large improvement with DUDA (MiT-B0) across SYN→CS, CS→DZur, and CS→ACDC. In a few cases, we see a performance drop, *e.g.*, Road class in SYN→CS and CS→ACDC for DUDA$_{DAF}$. Overall, due to learning from higher-quality labels and inconsistency-based balancing, DUDA performs significantly better in most of the classes across benchmarks, with more substantial improvement generally observed in the minority classes. While the Train class is significantly improved ($\sim$50%) in MiT-B0 and MiT-B1, the improvement is reduced since the baseline accuracy is high. The performance drop of the Road class in SYN→CS and CS→ACDC is again observed in MiT-B2 and MiT-B4. Overall, the observations and trends in the class-wise IoU of MiT-B0 and MiT-B1 are similar in the larger backbones.

In addition to the mIoU presented in Tab. 1 in Sec. 4 of the main paper, we provide the class-wise comparison of DUDA with the SOTA SegFormer-based methods in Tab. 8. It demonstrates the class-wise IoU of the SOTA methods and DUDA with MiT-B0, B1, B2, and B4 backbones across the four different UDA semantic segmentation benchmarks. Our DUDA$_{MIC}$ with MiT-B4 performs slightly better than recent SOTA methods, such as MIC [5] and MICDrop [24], with MiT-B5 in on GTA→CS and SYN→CS. We believe this can be attributed to our inconsistency-weighted loss boosting accuracy on underperforming classes and the efficacy of learning from multiple teachers (LT and LS). DUDA also performs better than recent method CSI [9] in most experiments (*e.g.*, with DAFormer base, mIoU of 69.8 w/ DUDA MiT-B2 vs. 67.9 w/ CSI MiT-B5 in GTA→CS; mIoU of 61.6 w/ DUDA MiT-B2 vs 61.4 w/ CSI MiT-B5 in SYN→CS).

However, we omit comparison with MICDrop and CSI in the Tables to ensure fairness. The performance gain in MICDrop stems from additional geometric information, such as depth predictions, to enhance learning segmentation boundaries. Also, all the compared approaches in the Tables including ours assume consistent taxonomies between source and target domains, typical in traditional UDA semantic segmentation, whereas CSI considers inconsistent taxonomies between domains. Similarly, we do not directly compare with UDA methods leveraging foundation models [9, 13, 23]. Lastly, InforMS [20] proposed Online Informative Class Sampling to dynamically adjust the weights of different semantic classes. However, it is particularly designed for daytime-nighttime adaptation scenarios, considering illumination discrepancies in the scene using spectrogram mean. In contrast, DUDA does not assume specific adaptation scenarios.

Comparison with ResNet-based Methods. The various

Method		Road	S.Walk	Build.	Wall	Fence	Pole	Tr.Light	Sign	Veget.	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	M.bike	Bike	mIoU
Synthetic-to-Real: GTA→Cityscapes (Val.)																					
DAFormer [3]	MiT-B0	92.2	55.9	85.6	25.2	22.3	40.0	39.5	46.2	87.3	43.6	87.7	63.4	31.8	85.4	36.4	40.6	1.4	31.8	52.9	51.0
DUDA$_{DAF}$	MiT-B0	96.3	76.7	88.5	43.0	41.7	48.5	49.5	59.6	89.7	44.4	91.7	68.9	40.3	91.2	72.6	72.6	52.8	52.4	61.2	65.2
MIC [5]	MiT-B0	95.0	68.0	88.4	44.4	29.4	48.6	48.5	62.5	89.8	46.0	92.7	71.0	36.6	87.5	49.2	57.7	0.4	53.6	62.1	59.5
DUDA$_{MIC}$	MiT-B0	97.1	78.3	90.6	56.3	51.0	56.8	58.3	68.1	91.2	49.2	93.7	76.4	50.0	93.3	76.7	82.4	71.2	56.7	65.5	71.7
DAFormer [3]	MiT-B1	94.4	64.8	87.1	34.1	27.2	44.7	47.7	55.0	88.7	47.3	90.3	66.4	32.1	89.2	59.9	55.6	52.0	47.6	58.8	60.2
DUDA$_{DAF}$	MiT-B1	96.7	75.8	89.2	47.7	46.8	50.9	52.9	63.5	90.2	45.3	92.7	70.9	42.8	92.3	76.0	79.1	69.8	56.2	62.2	68.5
MIC [5]	MiT-B1	95.8	72.0	89.9	54.4	40.3	55.8	59.7	70.2	90.9	50.5	93.8	75.3	46.1	91.5	62.8	64.5	35.4	60.6	65.5	67.1
DUDA$_{MIC}$	MiT-B1	97.3	79.6	91.0	54.4	53.9	59.0	62.2	70.8	91.6	50.2	93.9	78.3	53.7	94.2	82.3	84.5	70.8	58.1	67.0	73.3
Synthetic-to-Real: Synthia→Cityscapes (Val.)																					
DAFormer [3]	MiT-B0	57.2	21.6	84.1	9.9	1.0	40.2	34.4	40.6	84.1	-	86.5	65.4	32.3	81.6	-	36.6	-	7.4	54.5	46.1
DUDA$_{DAF}$	MiT-B0	75.9	31.3	88.1	41.9	7.6	48.4	50.4	52.2	84.3	-	91.4	70.3	43.7	86.9	-	54.6	-	45.1	61.0	58.3
MIC [5]	MiT-B0	83.8	39.0	86.1	0.2	0.9	48.1	52.0	49.5	85.9	-	93.5	71.6	35.1	86.3	-	47.9	-	7.3	60.1	53.0
DUDA$_{MIC}$	MiT-B0	85.7	52.5	88.9	43.9	8.6	56.8	62.0	61.2	82.9	-	94.5	78.4	53.1	89.7	-	62.1	-	60.7	63.0	65.3
DAFormer [3]	MiT-B1	85.8	36.6	85.9	32.0	2.3	43.1	47.0	47.5	86.1	-	92.1	69.9	37.2	84.2	-	37.2	-	39.8	59.5	55.4
DUDA$_{DAF}$	MiT-B1	77.5	33.7	88.6	43.2	8.9	51.0	53.6	56.1	84.1	-	90.9	72.8	48.9	86.9	-	59.1	-	49.7	63.4	60.5
MIC [5]	MiT-B1	94.8	69.5	87.2	38.7	1.5	55.0	60.3	57.8	87.5	-	94.3	76.5	46.0	88.9	-	61.4	-	58.2	63.2	65.0
DUDA$_{MIC}$	MiT-B1	85.2	52.4	89.5	46.9	7.8	59.7	65.4	63.7	82.4	-	95.0	80.2	57.9	90.1	-	64.8	-	62.9	64.3	66.8
Day-to-Nighttime: Cityscapes→DarkZurich (Test)																					
DAFormer [3]	MiT-B0	88.4	51.1	61.9	25.8	11.3	45.0	29.8	13.8	44.9	12.0	38.5	31.8	10.0	68.2	19.7	0.0	56.1	9.2	19.2	33.5
DUDA$_{DAF}$	MiT-B0	92.5	64.3	71.7	41.0	16.1	50.4	43.4	45.9	57.9	37.6	64.9	47.3	45.9	76.9	49.6	0.5	78.8	34.2	33.0	50.1
MIC [5]	MiT-B0	91.5	59.2	65.0	44.0	14.8	46.4	10.7	33.3	52.6	34.5	51.5	43.6	17.0	52.2	0.0	0.0	62.8	7.5	26.1	37.5
DUDA$_{MIC}$	MiT-B0	94.7	73.7	79.5	49.5	17.7	57.3	32.0	49.0	57.1	39.7	68.2	58.2	49.1	79.9	78.9	1.8	86.2	31.4	38.7	54.9
DAFormer [3]	MiT-B1	91.0	55.2	50.8	35.2	12.5	38.4	30.4	29.6	29.7	28.5	21.3	32.2	22.5	66.8	59.0	0.0	56.9	9.0	27.7	36.7
DUDA$_{DAF}$	MiT-B1	93.1	65.3	73.1	40.0	18.8	52.3	45.4	46.2	58.7	40.6	65.8	54.3	30.6	79.3	51.3	3.0	86.3	42.9	36.3	51.8
MIC [5]	MiT-B1	91.7	61.8	70.5	44.1	17.8	51.0	19.6	39.0	45.1	34.5	54.0	51.0	14.6	33.8	75.3	0.0	82.1	24.6	29.3	44.2
DUDA$_{MIC}$	MiT-B1	94.5	72.4	80.9	50.7	21.8	60.3	33.3	53.0	58.0	38.4	69.0	62.1	53.0	80.4	72.9	11.5	86.1	38.1	41.3	56.7
Clear-to-Adverse-Weather: Cityscapes→ACDC (Test)																					
DAFormer [3]	MiT-B0	79.0	36.0	66.1	26.9	23.3	41.9	47.2	46.0	80.2	45.4	85.4	38.4	13.2	69.6	37.4	33.3	28.7	18.5	37.0	44.9
DUDA$_{DAF}$	MiT-B0	64.0	55.2	83.0	40.0	33.9	48.2	27.8	54.5	74.1	51.3	60.0	54.1	26.9	80.6	51.6	45.7	80.3	30.6	46.5	53.1
MIC [5]	MiT-B0	66.7	48.9	74.3	42.6	23.4	47.5	58.9	57.0	82.4	53.4	67.5	43.2	18.0	75.0	51.1	42.5	66.2	21.6	42.1	51.7
DUDA$_{MIC}$	MiT-B0	90.2	65.7	87.5	48.8	35.0	53.4	59.1	62.9	75.1	58.8	87.0	62.3	41.5	85.0	60.3	67.8	86.1	42.4	55.4	64.4
DAFormer [3]	MiT-B1	80.6	39.0	73.2	31.9	26.2	44.1	49.3	52.0	69.5	48.1	85.3	44.8	17.4	67.7	36.3	44.6	58.5	25.1	45.6	49.4
DUDA$_{DAF}$	MiT-B1	64.6	57.5	83.6	40.8	34.8	50.2	28.9	56.2	74.7	53.7	60.1	56.9	30.8	81.6	52.6	47.4	82.8	35.0	48.0	54.7
MIC [5]	MiT-B1	56.0	52.4	81.0	45.0	31.1	51.0	60.4	58.0	73.8	54.5	58.5	59.2	39.4	77.5	58.7	55.3	78.7	38.0	53.0	56.9
DUDA$_{MIC}$	MiT-B1	90.7	66.4	88.1	48.5	37.3	55.9	60.2	65.7	75.7	59.4	87.0	66.1	46.4	86.4	63.6	69.0	89.5	48.3	59.7	66.5

Table 6. Comparison of DUDA with DAFormer and MIC in MiT-B0 and B1 backbones in class-wise IoU. mIoU for Synthia→Cityscapes is averaged over 16 classes.

ResNet-based methods reported in Tab. 2 in Sec. 4 (main paper) are compared with the class-wise IoU in Tab. 9, including DUDA$_{MIC}$ with DeepLab-V2 (DLV2) backbone. In particular, the accuracy improvement is noticeable in GTA→CS, and the Train class in DUDA$_{MIC}$ shows $\sim 15\%$ higher IoU than other UDA methods. Similarly, DUDA$_{MIC}$ achieves 10% higher IoU in certain classes, such as the Wall class in SYN→CS and Car class in CS→ACDC.

7. Additional Discussion

Vanilla KD baseline using LT. The vanilla KD baseline (row-3) in the ablation study tables (*e.g.*, Tab. 4 of the main paper and Tab. 4) uses LS for distillation. We observe comparable performance using LT for distillation. For example, with LT, mIoU for MiT-B1 degrades slightly (-0.07%), and mIoU for MiT-B0 improves slightly (+0.04%) in GTA→CS compared to using LS.

KL Loss in pre-adaptation. KL loss enriches learning by capturing class correlations using output logits. Note, we incorporate KL loss in the pre-adaptation. In GTA→CS for DUDA$_{DAF}$, the KL in the pre-adaptation improves the final mIoU of MiT-B0 (65.1→65.2%) and MiT-B1 (68.2→68.5%).

MiT-B5 as both Large (LS and LT) and Small (SS) Networks. As DUDA-B4 outperforms MIC-B5 in GTA→CS and SYN→CS (Tab. 1 in the main script), we evaluate DUDA-B5 to check whether the performance improvement continues. DUDA$_{MIC}$ with MiT-B5 achieves mIoU of 76.7 in GTA→CS and 68.6 in SYN→CS. DUDA$_{DAF}$ with MiT-B5 achieves mIoU of 70.9 in GTA→CS and 61.6 in SYN→CS. DUDA-B5 models perform very similarly to DUDA-B4 (*i.e.,* marginally better overall), indicating the performance gain is now saturated. It makes sense since, in DUDA-B5, all three networks (LS, LT, and SS) are MiT-B5 with the same capacity.

LS/LT as Teacher at Pre-adaptation and Fine-tuning. As we have two large networks (LT and LS) to guide a lightweight network (SS), four different setups can be con-

Method		Road	S.Walk	Build.	Wall	Fence	Pole	Tr.Light	Sign	Veget.	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	M.bike	Bike	mIoU
Synthetic-to-Real: GTA→Cityscapes (Val.)																					
DAFormer [3]	MiT-B2	94.9	63.8	88.6	47.2	34.9	48.3	54.1	57.9	89.6	49.6	91.0	67.8	39.9	90.7	61.2	67.1	59.0	50.4	58.4	63.9
DUDA$_{\text{DAF}}$	MiT-B2	97.0	76.9	89.8	53.4	48.5	52.7	55.5	64.5	90.3	44.5	93.1	72.2	45.3	93.0	79.9	83.3	68.2	54.7	63.0	69.8
MIC [5]	MiT-B2	96.8	76.4	90.9	57.4	51.3	58.8	63.9	70.6	91.4	50.0	94.1	77.0	50.0	93.9	80.0	84.2	70.2	59.1	65.1	72.7
DUDA$_{\text{MIC}}$	MiT-B2	97.5	80.7	91.7	61.7	57.0	60.6	64.3	71.3	91.8	51.5	94.0	79.6	56.1	94.5	83.7	90.0	80.1	61.2	68.1	75.5
DAFormer [3]	MiT-B4	93.9	59.8	88.9	49.6	44.4	48.9	55.7	56.8	89.3	49.3	92.2	71.5	42.1	91.7	59.8	77.0	67.3	57.5	60.0	66.1
DUDA$_{\text{DAF}}$	MiT-B4	97.0	77.2	90.1	54.7	51.3	53.0	57.0	65.1	90.5	45.2	93.0	73.2	45.9	93.2	81.4	82.5	64.6	60.1	64.7	70.5
MIC [5]	MiT-B4	96.4	75.7	91.8	61.0	58.4	59.8	65.3	73.2	92.0	52.7	93.9	79.2	52.5	93.6	76.7	80.9	74.5	65.6	67.6	74.3
DUDA$_{\text{MIC}}$	MiT-B4	97.5	80.7	92.0	63.6	59.5	61.4	65.5	72.0	91.9	51.8	94.1	80.4	57.3	94.7	87.0	91.1	82.9	64.9	68.9	76.7
Synthetic-to-Real: Synthia→Cityscapes (Val.)																					
DAFormer [3]	MiT-B2	89.7	46.9	86.8	36.0	3.8	48.4	52.6	45.1	85.8	-	92.7	72.5	41.6	86.8	-	53.0	-	47.6	60.7	59.4
DUDA$_{\text{DAF}}$	MiT-B2	78.0	32.9	89.0	43.0	8.3	52.3	56.6	56.7	86.1	-	90.7	74.5	49.8	86.7	-	62.2	-	54.4	64.0	61.6
MIC [5]	MiT-B2	91.2	58.5	89.0	44.0	3.1	57.8	65.3	64.8	88.7	-	94.3	79.5	53.4	89.1	-	56.9	-	61.7	64.5	66.4
DUDA$_{\text{MIC}}$	MiT-B2	84.9	51.3	90.0	47.7	8.7	61.7	67.6	64.4	83.1	-	95.0	81.5	60.4	89.2	-	62.2	-	64.6	65.1	67.3
DAFormer [3]	MiT-B4	85.9	41.9	88.4	38.4	6.1	50.1	54.9	56.7	87.4	-	85.2	72.7	45.5	87.1	-	51.8	-	51.6	54.9	59.9
DUDA$_{\text{DAF}}$	MiT-B4	78.0	32.9	88.9	43.0	8.0	52.3	57.7	57.0	86.2	-	90.8	74.7	50.3	86.9	-	66.4	-	54.6	64.0	62.0
MIC [5]	MiT-B4	86.3	49.5	88.4	39.9	9.6	60.2	67.8	63.5	88.8	-	94.2	80.1	56.2	89.5	-	53.3	-	65.1	63.5	66.0
DUDA$_{\text{MIC}}$	MiT-B4	85.5	51.5	90.2	45.5	9.5	62.4	69.1	65.2	84.2	-	95.0	82.0	61.5	89.9	-	69.7	-	67.0	65.3	68.3
Day-to-Nighttime: Cityscapes→DarkZurich (Test)																					
DAFormer [3]	MiT-B2	92.3	57.7	66.5	28.6	18.0	51.3	9.7	40.4	43.7	27.9	46.7	42.7	36.8	74.9	63.6	0.0	77.3	36.5	34.0	44.7
DUDA$_{\text{DAF}}$	MiT-B2	93.6	68.1	75.4	45.4	17.2	53.8	45.6	49.9	58.7	39.8	66.1	50.9	47.5	81.5	53.9	3.2	89.3	55.4	37.4	54.4
MIC [5]	MiT-B2	92.6	70.4	81.3	53.6	21.1	57.3	48.1	53.2	65.1	39.6	79.0	58.4	53.1	53.5	83.5	0.0	86.1	42.3	36.1	56.5
DUDA$_{\text{MIC}}$	MiT-B2	95.0	75.1	82.1	53.6	24.2	61.6	35.0	56.7	58.1	43.1	69.2	64.5	59.9	81.3	81.0	6.1	90.4	53.2	43.0	59.6
DAFormer [3]	MiT-B4	92.7	63.5	65.9	34.7	11.5	48.1	17.0	44.4	44.4	25.1	39.0	54.5	52.7	76.7	47.6	2.7	89.0	42.6	39.9	46.9
DUDA$_{\text{DAF}}$	MiT-B4	93.7	68.2	75.7	42.6	19.3	53.8	43.5	47.0	61.3	37.2	66.8	56.6	54.9	81.0	52.3	3.3	90.1	48.4	38.3	54.4
MIC [5]	MiT-B4	95.3	76.7	83.0	55.4	25.0	63.0	35.4	57.5	59.1	44.2	70.5	66.3	55.1	81.2	80.8	12.7	90.5	42.5	44.5	59.9
DUDA$_{\text{MIC}}$	MiT-B4	95.4	77.2	82.9	55.6	25.6	62.8	35.7	57.8	59.1	43.9	70.5	66.2	58.6	81.2	81.8	13.1	91.4	42.4	44.4	60.3
Clear-to-Adverse-Weather: Cityscapes→ACDC (Test)																					
DAFormer [3]	MiT-B2	55.7	40.3	83.8	42.2	31.8	48.1	39.9	50.5	73.7	48.2	50.6	56.1	31.0	78.6	53.9	55.5	73.0	36.0	43.5	52.2
DUDA$_{\text{DAF}}$	MiT-B2	64.1	57.1	84.1	44.7	36.9	51.8	30.3	58.1	75.0	53.6	59.5	60.3	35.8	82.6	58.5	51.9	84.2	40.6	50.0	56.8
MIC [5]	MiT-B2	53.4	55.4	81.7	53.9	37.8	55.3	59.5	62.3	80.0	55.7	56.7	63.8	39.1	82.7	71.3	67.2	81.9	43.9	55.9	60.9
DUDA$_{\text{MIC}}$	MiT-B2	91.0	67.4	88.7	50.7	39.4	57.8	60.9	67.2	76.2	60.7	87.0	69.5	48.1	88.1	71.7	78.4	90.3	51.3	59.8	68.6
DAFormer [3]	MiT-B4	69.0	34.9	84.4	44.3	32.4	50.9	32.0	57.0	72.2	41.6	72.6	58.5	35.3	81.0	54.1	66.1	81.1	38.3	49.1	55.5
DUDA$_{\text{DAF}}$	MiT-B4	63.2	57.9	85.0	47.6	36.6	52.2	29.6	58.2	75.1	54.4	57.6	61.8	36.9	83.3	59.4	58.7	85.7	41.9	51.0	57.7
MIC [5]	MiT-B4	52.8	62.9	86.1	58.8	41.3	55.9	53.1	59.0	74.9	58.1	47.9	69.6	46.6	86.3	75.7	84.0	89.9	52.7	61.3	64.0
DUDA$_{\text{MIC}}$	MiT-B4	91.4	68.8	89.3	52.3	40.4	59.2	61.3	68.6	76.4	62.1	87.1	71.5	48.6	89.3	76.7	83.8	90.6	55.5	61.5	70.2

Table 7. Comparison of DUDA with DAFormer and MIC in MiT-B2 and B4 backbones in class-wise IoU. mIoU for Synthia→Cityscapes is averaged over 16 classes.

sidered in KD teacher, *e.g.*, LT/LS at the pre-adaptation and LT/LS at the fine-tuning. DUDA employs two large networks (LT and LS) in a multi-teacher setup, where different large networks are used at the pre-adaptation and fine-tuning phases. Our experiments show the multi-teacher setup performs slightly better than the same-teacher setup. Relative mIoU differences are as follows: (multi-teacher) LT→LS (DUDA): +0.0% LS→LT: +0.1% and (same-teacher) LT→LT: -0.5% LS→LS: -0.2%, showing the multi-teacher setup in DUDA is effective.

Upper Bound of Small Model's Performance. In order to better understand the contribution of the UDA side, we investigate the upper bound of small models when distilled by large models in a supervised setup. When distilled from a large MiT-B5 (trained in a *supervised* setup on Cityscapes), the performance of the smaller models is: mIoU of 75.4 for MiT-B0, mIoU of 77.5 for MiT-B1. In contrast, our GTA→Cityscapes DUDA$_{\text{MIC}}$ models, trained *unsupervised*, exhibit only a marginal performance decrease (See Tab. 1 in the main paper, or Tab. 8): MiT-B0 mIoU 71.7, MiT-B1 mIoU 73.3.

Static Inconsistency-based Weighting during Fine-tuning. The general approach to balanced loss involves statically weighting the loss based on the distribution in the entire dataset, and our method follows this practice. Also, during the fine-tuning stage with inconsistency-based balanced loss, the class distribution remains constant since the teacher model is frozen.

Method	Road	S.Walk	Build.	Wall	Fence	Pole	Tr.Light	Sign	Veget.	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	M.bike	Bike	mIoU
Synthetic-to-Real: GTA→Cityscapes (Val.)																				
HRDA [4]	96.4	74.4	91.0	61.6	51.5	57.1	63.9	69.3	91.3	48.4	94.2	79.0	52.9	93.9	84.1	85.7	75.9	63.9	67.5	73.8
DiGA [15]	97.0	78.6	91.3	60.8	56.7	56.5	64.4	69.9	91.5	50.8	93.7	79.2	55.2	93.7	78.3	86.9	77.8	63.7	65.8	74.3
GANDA [8]	96.5	74.8	91.4	61.7	57.3	59.2	65.4	68.8	91.5	49.9	94.7	79.6	54.8	94.1	81.3	86.8	74.6	64.8	68.2	74.5
RTea [27]	97.1	75.2	92.6	63.5	51.8	58.2	66.5	71.2	91.1	49.0	96.8	81.5	54.2	94.2	84.8	86.6	75.7	62.2	66.7	74.7
BLV [21]	96.7	76.6	91.5	61.2	56.9	59.4	62.2	72.8	91.5	51.2	94.3	77.5	54.7	93.5	83.2	84.7	79.7	68.1	67.6	74.9
IR^{2}F-RMM [2]	97.5	80.0	91.0	60.0	53.3	56.2	63.9	72.4	91.7	51.0	94.2	79.0	51.1	94.3	84.7	86.7	75.9	62.6	67.8	74.4
CDAC [19]	97.1	78.7	91.8	59.6	57.1	59.1	66.1	72.2	91.8	53.1	94.5	79.4	51.6	94.6	84.9	87.8	78.7	64.9	67.6	75.3
PiPa [1]	96.8	76.3	91.6	63.0	57.7	60.0	65.4	72.6	91.7	51.8	94.8	79.7	56.4	94.4	85.9	88.4	78.9	63.5	67.2	75.6
MIC [5]	97.4	80.1	91.7	61.2	56.9	59.7	66.0	71.3	91.7	51.4	94.3	79.8	56.1	94.6	85.4	90.3	80.4	64.5	68.5	75.9
MICDrop [24]	97.6	81.5	92.0	62.8	59.4	62.6	62.9	73.6	91.6	52.6	94.1	80.2	57.0	94.8	87.4	90.7	81.6	65.3	67.8	76.6
DUDA$_{\text{MIC}}$ (B4)	97.5	80.7	92.0	63.6	59.5	61.4	65.5	72.0	91.9	51.8	94.1	80.4	57.3	94.7	87.0	91.1	82.9	64.9	68.9	76.7
DUDA$_{\text{MIC}}$ (B2)	97.5	80.7	91.7	61.7	57.0	60.6	64.3	71.3	91.8	51.5	94.0	79.6	56.1	94.5	83.7	90.0	80.1	61.2	68.1	75.5
DUDA$_{\text{MIC}}$ (B1)	97.3	79.6	91.0	54.4	53.9	59.0	62.2	70.8	91.6	50.2	93.9	78.3	53.7	94.2	82.3	84.5	70.8	58.1	67.0	73.3
DUDA$_{\text{MIC}}$ (B0)	97.1	78.3	90.6	56.3	51.0	56.8	58.3	68.1	91.2	49.2	93.7	76.4	50.0	93.3	76.7	82.4	71.2	56.7	65.5	71.7
Synthetic-to-Real: Synthia→Cityscapes (Val.)																				
HRDA [4]	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
DiGA [15]	88.5	49.9	90.1	51.4	6.6	55.3	64.8	62.7	88.2	-	93.5	78.6	51.8	89.5	-	62.2	-	61.0	65.8	66.2
GANDA [8]	89.1	50.6	89.7	51.4	6.7	59.4	66.8	57.7	86.7	-	93.8	80.6	56.9	90.7	-	64.8	-	62.6	65.0	67.0
RTea [27]	87.8	49.0	90.3	50.3	5.5	58.6	66.0	61.4	86.8	-	93.1	79.5	53.1	89.5	-	65.1	-	63.7	64.6	66.5
BLV [21]	87.6	47.9	90.5	50.4	6.9	57.1	64.3	65.3	86.9	-	93.4	78.9	54.9	89.1	-	62.9	-	65.2	66.8	66.8
IR^{2}F-RMM [2]	90.4	54.9	89.4	48.0	7.4	59.0	65.5	63.2	87.8	-	94.1	80.5	55.8	90.0	-	65.9	-	64.5	66.8	67.7
CDAC [19]	93.1	68.5	89.8	51.2	8.9	59.4	65.5	65.3	84.7	-	94.4	81.2	57.0	90.5	-	56.9	-	66.8	66.4	68.7
PiPa [1]	88.6	50.1	90.0	53.8	7.7	58.1	67.2	63.1	88.5	-	94.5	79.7	57.6	90.8	-	70.2	-	65.1	66.9	68.2
MIC [5]	86.6	50.5	89.3	47.9	7.8	59.4	66.7	63.4	87.1	-	94.6	81.0	58.9	90.1	-	61.9	-	67.1	64.3	67.3
MICDrop [24]	82.8	42.6	90.5	51.6	9.6	61.0	65.7	65.0	89.1	-	95.0	81.1	59.7	90.6	-	68.3	-	67.4	66.5	67.9
DUDA$_{\text{MIC}}$ (B4)	85.5	51.5	90.2	45.5	9.5	62.4	69.1	65.2	84.2	-	95.0	82.0	61.5	89.9	-	69.7	-	67.0	65.3	68.3
DUDA$_{\text{MIC}}$ (B2)	84.9	51.3	90.0	47.7	8.7	61.7	67.6	64.4	83.1	-	95.0	81.5	60.4	89.2	-	62.2	-	64.6	65.1	67.3
DUDA$_{\text{MIC}}$ (B1)	85.2	52.4	89.5	46.9	7.8	59.7	65.4	63.7	82.4	-	95.0	80.2	57.9	90.1	-	64.8	-	62.9	64.3	66.8
DUDA$_{\text{MIC}}$ (B0)	85.7	52.5	88.9	43.9	8.6	56.8	62.0	61.2	82.9	-	94.5	78.4	53.1	89.7	-	62.1	-	60.7	63.0	65.3
Day-to-Nighttime: Cityscapes→DarkZurich (Test)																				
HRDA [4]	90.4	56.3	72.0	39.5	19.5	57.8	52.7	43.1	59.3	29.1	70.5	60.0	58.6	84.0	75.5	11.2	90.5	51.6	40.9	55.9
IR^{2}F-RMM [2]	94.7	75.1	73.2	44.4	25.7	60.6	39.0	47.4	70.2	41.6	77.3	62.4	55.5	86.4	55.5	20.0	92.0	55.3	42.8	58.9
MIC [5]	94.8	75.0	84.0	55.1	28.4	62.0	35.5	52.6	59.2	46.8	70.0	65.2	61.7	82.1	64.2	18.5	91.3	52.6	44.0	60.2
DUDA$_{\text{MIC}}$ (B4)	95.4	77.2	82.9	55.6	25.6	62.8	35.7	57.8	59.1	43.9	70.5	66.2	58.6	81.2	81.8	13.1	91.4	42.4	44.4	60.3
DUDA$_{\text{MIC}}$ (B2)	95.0	75.1	82.1	53.6	24.2	61.6	35.0	56.7	58.1	43.1	69.2	64.5	59.9	81.3	81.0	6.1	90.4	53.2	43.0	59.6
DUDA$_{\text{MIC}}$ (B1)	94.5	72.4	80.9	50.7	21.8	60.3	33.3	53.0	58.0	38.4	69.0	62.1	53.0	80.4	72.9	11.5	86.1	38.1	41.3	56.7
DUDA$_{\text{MIC}}$ (B0)	94.7	73.7	79.5	49.5	17.7	57.3	32.0	49.0	57.1	39.7	68.2	58.2	49.1	79.9	78.9	1.8	86.2	31.4	38.7	54.9
Clear-to-Adverse-Weather: Cityscapes→ACDC (Test)																				
HRDA [4]	88.3	57.9	88.1	55.2	36.7	56.3	62.9	65.3	74.2	57.7	85.9	68.8	45.7	88.5	76.4	82.4	87.7	52.7	60.4	68.0
CDAC [19]	87.0	56.7	84.5	53.5	34.3	54.6	43.6	51.4	71.7	58.6	85.4	68.7	45.7	89.0	70.9	81.5	90.1	47.6	59.0	64.9
MIC [5]	90.8	67.1	89.2	54.5	40.5	57.2	62.0	68.4	76.3	61.8	87.0	71.3	49.4	89.7	75.7	86.8	89.1	56.9	63.0	70.4
DUDA$_{\text{MIC}}$ (B4)	91.4	68.8	89.3	52.3	40.4	59.2	61.3	68.6	76.4	62.1	87.1	71.5	48.6	89.3	76.7	83.8	90.6	55.5	61.5	70.2
DUDA$_{\text{MIC}}$ (B2)	91.0	67.4	88.7	50.7	39.4	57.8	60.9	67.2	76.2	60.7	87.0	69.5	48.1	88.1	71.7	78.4	90.3	51.3	59.8	68.6
DUDA$_{\text{MIC}}$ (B1)	90.7	66.4	88.1	48.5	37.3	55.9	60.2	65.7	75.7	59.4	87.0	66.1	46.4	86.4	63.6	69.0	89.5	48.3	59.7	66.5
DUDA$_{\text{MIC}}$ (B0)	90.2	65.7	87.5	48.8	35.0	53.4	59.1	62.9	75.1	58.8	87.0	62.3	41.5	85.0	60.3	67.8	86.1	42.4	55.4	64.4

Table 8. Comparison of DUDA with prior UDA Semantic Segmentation methods in SegFormer-based networks in class-wise IoU. mIoU for Synthia→Cityscapes is averaged over 16 classes.

Method	Road	S.Walk	Build.	Wall	Fence	Pole	Tr.Light	Sign	Veget.	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	M.bike	Bike	mIoU
Synthetic-to-Real: GTA→Cityscapes (Val.)																				
ADVENT [18]	89.4	33.1	81.0	26.6	26.8	27.2	33.5	24.7	83.9	36.7	78.8	58.7	30.5	84.8	38.5	44.5	1.7	31.6	32.4	45.5
CBST [30]	89.6	58.9	78.5	33.0	22.3	41.4	48.2	39.2	83.6	24.3	65.4	49.3	20.2	83.3	39.0	48.6	12.5	20.3	35.3	47.0
CRST [31]	91.7	45.1	80.9	29.0	23.4	43.8	47.1	40.9	84.0	20.0	60.6	64.0	31.9	85.8	39.5	48.7	25.0	38.0	47.0	49.8
DACS [16]	89.9	39.7	87.9	30.7	39.5	38.5	46.4	52.8	88.0	44.0	88.8	67.2	35.8	84.5	45.7	50.2	0.0	27.3	34.0	52.1
ProDA [26]	87.8	56.0	79.7	46.3	44.8	45.6	53.5	53.5	88.6	45.2	82.1	70.7	39.2	88.8	45.5	59.4	1.0	48.9	56.4	57.5
Fredom [17]	90.9	54.1	87.8	44.1	32.6	45.2	51.4	57.1	88.6	42.6	89.5	68.8	40.0	89.7	58.4	62.6	55.3	47.7	58.1	61.3
RTea [27]	95.4	67.1	87.9	46.1	44.0	46.0	53.8	59.5	89.7	49.8	89.8	71.5	40.5	90.8	55.0	57.9	22.1	47.7	62.5	61.9
DiGA [15]	95.6	67.4	89.8	51.6	38.1	52.0	59.0	51.5	86.4	34.5	87.7	75.6	48.8	92.5	66.5	63.8	19.7	49.6	61.6	62.7
CONFETI [7]	96.5	75.6	88.9	45.1	45.9	50.1	61.2	68.2	89.4	45.7	86.3	76.3	49.9	92.2	55.1	62.8	16.7	33.8	63.1	63.3
HRDA [4]	96.2	73.1	89.7	43.2	39.9	47.5	60.0	60.0	89.9	47.1	90.2	75.9	49.0	91.8	61.9	59.3	10.2	47.0	65.3	63.0
MIC [5]	96.5	74.3	90.4	47.1	42.8	50.3	61.7	62.3	90.3	49.2	90.7	77.8	53.2	93.0	66.2	68.0	6.8	38.0	60.6	64.2
DUDA$_{\text{MIC}}$ (R101)	97.7	80.9	91.1	49.8	55.5	57.7	62.6	70.0	91.4	50.9	94.1	78.6	56.7	94.2	81.9	85.5	71.2	59.8	66.9	73.5
DUDA$_{\text{MIC}}$ (R50)	97.4	79.4	91.1	54.6	54.2	56.3	62.1	69.2	91.1	47.3	93.7	76.5	53.1	93.9	79.0	84.5	73.6	59.7	66.3	72.8
DUDA$_{\text{MIC}}$ (R18)	97.0	77.0	89.8	47.8	49.1	54.3	57.8	66.7	90.7	47.1	92.7	73.8	48.7	93.2	73.5	79.1	60.8	54.8	63.1	69.3
Synthetic-to-Real: Synthia→Cityscapes (Val.)																				
ADVENT [18]	85.6	42.2	79.7	8.7	0.4	25.9	5.4	8.1	80.4	-	84.1	57.9	23.8	73.3	-	36.4	-	14.2	33.0	41.2
CBST [30]	53.6	23.7	75.0	12.5	0.3	36.4	23.5	26.3	84.8	-	74.7	67.2	17.5	84.5	-	28.4	-	15.2	55.8	42.5
CRST [31]	67.7	32.2	73.9	10.7	1.6	37.4	22.2	31.2	80.8	-	80.5	60.8	29.1	82.8	-	25.0	-	19.4	45.3	43.8
DACS [16]	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	48.3
ProDA [26]	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
*DAFormer [3]	62.1	24.7	85.2	24.5	3.7	38.6	44.8	50.9	84.9	-	84.1	69.6	40.6	86.1	-	51.7	-	46.5	55.2	55.3
GANDA [8]	87.1	45.8	86.1	28.9	4.8	37.1	40.6	45.0	87.0	-	87.9	69.1	39.8	89.9	-	59.8	-	33.8	57.2	56.3
Fredom [17]	86.0	46.3	87.0	33.3	5.3	48.7	53.4	46.8	87.1	-	89.1	71.2	38.1	87.1	-	54.6	-	51.3	59.9	59.1
RTea [27]	93.2	59.6	86.3	31.3	4.8	43.1	41.8	44.0	88.6	-	90.5	70.4	42.6	89.5	-	56.7	-	40.2	59.9	58.9
DiGA [15]	89.1	53.4	86.1	28.7	3.0	49.6	50.6	34.9	88.2	-	84.9	71.3	40.9	91.6	-	75.1	-	50.3	65.8	60.2
CONFETI [7]	83.8	44.6	86.9	15.4	3.7	44.3	56.9	55.5	84.9	-	86.2	73.8	46.8	90.1	-	57.1	-	46.0	63.2	58.7
HRDA [4]	85.8	47.3	87.3	27.3	1.4	50.5	57.8	61.0	87.4	-	89.1	76.2	48.5	87.3	-	49.3	-	55.0	68.2	61.2
MIC [5]	84.7	45.7	88.3	29.9	2.8	53.3	61.0	59.5	86.9	-	88.8	78.2	53.3	89.4	-	58.8	-	56.0	68.3	62.8
DUDA$_{\text{MIC}}$ (R101)	88.7	54.2	90.0	47.9	8.6	59.0	65.7	62.9	86.8	-	94.3	80.5	59.5	90.4	-	62.6	-	61.1	65.5	67.4
DUDA$_{\text{MIC}}$ (R50)	88.3	50.5	89.7	48.6	9.1	57.6	65.3	62.3	89.3	-	94.0	79.1	58.2	90.2	-	65.5	-	62.6	65.6	67.2
DUDA$_{\text{MIC}}$ (R18)	87.4	48.7	88.5	42.1	9.0	52.3	58.8	58.4	88.9	-	93.6	75.0	52.8	88.8	-	58.1	-	53.9	62.2	63.7
Day-to-Nighttime: Cityscapes→DarkZurich (Test)																				
ADVENT [18]	85.8	37.9	55.5	27.7	14.5	23.1	14.0	21.1	32.1	8.7	2.0	39.9	16.6	64.0	13.8	0.0	58.8	28.5	20.7	29.7
MGCDA [14]	80.3	49.3	66.2	7.8	11.0	41.4	38.9	39.0	64.1	18.0	55.8	52.1	53.5	74.7	66.0	0.0	37.5	29.1	22.7	42.5
DANNet [22]	90.0	54.0	74.8	41.0	21.1	25.0	26.8	30.2	72.0	26.2	84.0	47.0	33.9	68.2	19.0	0.3	66.4	38.3	23.6	44.3
*DAFormer [3]	84.2	56.5	67.4	32.5	14.8	46.1	32.6	44.7	33.8	23.3	1.8	50.2	43.0	74.7	69.4	8.5	54.3	28.6	36.0	44.2
HRDA [4]	88.7	65.5	68.3	41.9	18.1	50.6	6.0	39.6	33.3	34.4	0.3	57.6	51.7	75.0	70.9	8.5	63.6	41.0	38.8	44.9
MIC [5]	82.8	69.6	75.5	44.0	21.0	51.1	43.4	48.3	39.3	37.1	0.0	59.4	53.6	73.6	74.2	9.2	78.7	40.0	37.2	49.4
DUDA$_{\text{MIC}}$ (R101)	94.1	72.2	78.0	45.6	23.8	58.0	32.7	52.8	56.4	33.1	68.9	63.1	55.8	76.4	85.4	12.1	61.9	38.7	40.3	55.2
DUDA$_{\text{MIC}}$ (R50)	93.5	69.1	78.7	47.1	19.6	57.3	32.4	49.2	55.0	36.1	68.0	62.4	52.9	76.9	75.0	1.8	74.8	41.2	39.7	54.2
DUDA$_{\text{MIC}}$ (R18)	92.8	66.0	76.0	42.7	20.1	53.4	30.5	46.5	51.7	33.8	65.4	53.8	38.7	71.3	53.8	3.1	68.0	32.7	35.7	49.3
Clear-to-Adverse-Weather: Cityscapes→ACDC (Test)																				
ADVENT [18]	72.9	14.3	40.5	16.6	21.2	9.3	17.4	21.2	63.8	23.8	18.3	32.6	19.5	69.5	36.2	34.5	46.2	26.9	36.1	32.7
MGCDA [14]	73.4	28.7	69.9	19.3	26.3	36.8	53.0	53.3	75.4	32.0	84.6	51.0	26.1	77.6	43.2	45.9	53.9	32.7	41.5	48.7
DANNet [22]	84.3	54.2	77.6	38.0	30.0	18.9	41.6	35.2	71.3	39.4	86.6	48.7	29.2	76.2	41.6	43.0	58.6	32.6	43.9	50.0
*DAFormer [3]	76.3	48.1	78.0	34.7	26.9	38.3	50.7	52.8	70.0	45.1	78.4	54.5	28.3	78.1	47.9	43.6	70.3	22.6	49.2	52.3
HRDA [4]	84.9	63.2	83.1	33.1	32.3	46.0	42.7	55.4	69.2	52.8	83.1	63.2	37.8	78.1	48.5	58.5	62.4	42.8	57.2	57.6
MIC [5]	88.7	63.9	84.1	38.4	35.7	45.7	51.5	60.3	72.7	52.3	85.8	62.5	39.8	84.7	37.7	68.7	71.9	46.0	56.5	60.4
DUDA$_{\text{MIC}}$ (R101)	89.9	64.7	87.1	40.3	37.4	55.3	61.9	66.9	75.2	58.8	87.0	66.3	40.7	87.5	63.6	75.3	85.9	50.8	57.7	65.9
DUDA$_{\text{MIC}}$ (R50)	89.3	63.2	86.4	40.3	37.1	54.3	59.9	65.8	75.2	57.5	87.0	65.9	39.5	85.9	56.4	66.4	79.4	47.6	56.1	63.9
DUDA$_{\text{MIC}}$ (R18)	88.2	60.0	84.6	38.4	34.3	49.9	53.8	54.2	74.1	54.9	86.8	58.6	33.8	79.1	47.1	36.8	75.7	35.1	50.5	57.7

Table 9. Comparison of DUDA with prior UDA Semantic Segmentation methods in ResNet-based networks in class-wise IoU. mIoU for Synthia→Cityscapes is averaged over 16 classes. SSG [12] and DAFormer [3] in GTA→Cityscapes are omitted since the class-wise IoU is not provided in the literature. *DAFormer results are implemented by ourselves.

References

[1] Mu Chen, Zhedong Zheng, Yi Yang, and Tat-Seng Chua. Pipa: Pixel-and patch-wise self-supervised learning for domain adaptative semantic segmentation. In *Proc. of 31st ACM International Conference on Multimedia*, pages 1905–1914, 2023. 6

[2] Rui Gong, Qin Wang, Martin Danelljan, Dengxin Dai, and Luc Van Gool. Continuous pseudo-label rectified domain adaptive semantic segmentation with implicit neural representations. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7225–7235, 2023. 6

[3] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9924–9935, 2022. 1, 4, 5, 7

[4] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Hrda: Context-aware high-resolution domain-adaptive semantic segmentation. In *European Conference on Computer Vision*, pages 372–391. Springer, 2022. 1, 2, 6, 7

[5] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. Mic: Masked image consistency for context-enhanced domain adaptation. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11721–11732, 2023. 1, 3, 4, 5, 6, 7

[6] Seongwon Jeong, Jiyeong Kim, Sungheui Kim, and Dongbo Min. Revisiting domain-adaptive semantic segmentation via knowledge distillation. *IEEE Transactions on Image Processing*, 2024. 1

[7] Tianyu Li, Subhankar Roy, Huayi Zhou, Hongtao Lu, and Stéphane Lathuilière. Contrast, stylize and adapt: Unsupervised contrastive learning framework for domain adaptive semantic segmentation. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop*, pages 4868–4878, 2023. 1, 7

[8] Yinghong Liao, Wending Zhou, Xu Yan, Zhen Li, Yizhou Yu, and Shuguang Cui. Geometry-aware network for domain adaptive semantic segmentation. In *Proc. of AAAI Conference on Artificial Intelligence*, pages 8755–8763, 2023. 6, 7

[9] Jeongkee Lim and Yusung Kim. Cross-domain semantic segmentation on inconsistent taxonomy using vlms. *arXiv preprint arXiv:2408.02261*, 2024. 3

[10] Yahao Liu, Jinhong Deng, Xinchen Gao, Wen Li, and Lixin Duan. Bapa-net: Boundary adaptation and prototype alignment for cross-domain semantic segmentation. In *Proc. of IEEE/CVF international conference on computer vision*, pages 8801–8811, 2021. 2

[11] Fei Pan, Xu Yin, Seokju Lee, Axi Niu, Sungeui Yoon, and In So Kweon. Moda: Leveraging motion priors from videos for advancing unsupervised domain adaptation in semantic segmentation. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2649–2658, 2024. 1

[12] Duo Peng, Ping Hu, Qiuhong Ke, and Jun Liu. Diffusion-based image translation with label guidance for domain adaptive semantic segmentation. In *Proc. of IEEE/CVF International Conference on Computer Vision*, pages 808–820, 2023. 1, 7

[13] Wenqi Ren, Ruihao Xia, Meng Zheng, Ziyan Wu, Yang Tang, and Nicu Sebe. Cross-class domain adaptive semantic segmentation with visual language models. In *Proc. of 32nd ACM International Conference on Multimedia*, pages 5005–5014, 2024. 3

[14] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):3139–3153, 2020. 7

[15] Fengyi Shen, Akhil Gurram, Ziyuan Liu, He Wang, and Alois Knoll. Diga: Distil to generalize and then adapt for domain adaptive semantic segmentation. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15866–15877, 2023. 1, 6, 7

[16] Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson. Dacs: Domain adaptation via cross-domain mixed sampling. In *Proc. of IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1379–1389, 2021. 1, 7

[17] Thanh-Dat Truong, Ngan Le, Bhiksha Raj, Jackson Cothren, and Khoa Luu. Fredom: Fairness domain adaptation approach to semantic scene understanding. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19988–19997, 2023. 1, 7

[18] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proc. of IEEE/CVF conference on computer vision and pattern recognition*, pages 2517–2526, 2019. 7

[19] Kaihong Wang, Donghyun Kim, Rogerio Feris, and Margrit Betke. Cdac: Cross-domain attention consistency in transformer for domain adaptive semantic segmentation. In *Proc. of IEEE/CVF International Conference on Computer Vision*, pages 11519–11529, 2023. 6

[20] Shiqin Wang, Xin Xu, Xianzheng Ma, Kui Jiang, and Zheng Wang. Informative classes matter: Towards unsupervised domain adaptive nighttime semantic segmentation. In *Proc. of 31st ACM International Conference on Multimedia*, pages 163–172, 2023. 3

[21] Yuchao Wang, Jingjing Fei, Haochen Wang, Wei Li, Tianpeng Bao, Liwei Wu, Rui Zhao, and Yujun Shen. Balancing logit variation for long-tailed semantic segmentation. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19561–19573, 2023. 6

[22] Xinyi Wu, Zhenyao Wu, Hao Guo, Lili Ju, and Song Wang. Dannet: A one-stage domain adaptation network for unsupervised nighttime semantic segmentation. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15769–15778, 2021. 7

[23] Weihao Yan, Yeqiang Qian, Hanyang Zhuang, Chunxiang Wang, and Ming Yang. Sam4udass: When sam meets unsupervised domain adaptive semantic segmentation in intelligent vehicles. *IEEE Transactions on Intelligent Vehicles*, 9 (2):3396–3408, 2023. 3

[24] Linyan Yang, Lukas Hoyer, Mark Weber, Tobias Fischer, Dengxin Dai, Laura Leal-Taixé, Marc Pollefeys, Daniel Cremers, and Luc Van Gool. Micdrop: masking image and depth features via complementary dropout for domain-adaptive semantic segmentation. In *European Conference on Computer Vision*, pages 329–346. Springer, 2025. 3, 6

[25] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proc. of IEEE/CVF conference on computer vision and pattern recognition*, pages 4085–4095, 2020. 1, 2

[26] 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. of IEEE/CVF conference on computer vision and pattern recognition*, pages 12414–12424, 2021. 1, 2, 7

[27] Dong Zhao, Shuang Wang, Qi Zang, Dou Quan, Xiutiao Ye, Rui Yang, and Licheng Jiao. Learning pseudo-relations for cross-domain semantic segmentation. In *Proc. of IEEE/CVF International Conference on Computer Vision*, pages 19191–19203, 2023. 1, 6, 7

[28] Xingchen Zhao, Niluthpol Chowdhury Mithun, Abhinav Rajvanshi, Han-Pang Chiu, and Supun Samarasekera. Unsupervised domain adaptation for semantic segmentation with pseudo label self-refinement. In *Proc. of IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2399–2409, 2024. 1

[29] Qianyu Zhou, Zhengyang Feng, Qiqi Gu, Jiangmiao Pang, Guangliang Cheng, Xuequan Lu, Jianping Shi, and Lizhuang Ma. Context-aware mixup for domain adaptive semantic segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(2):804–817, 2022. 1

[30] Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proc. of European conference on computer vision (ECCV)*, pages 289–305, 2018. 7

[31] Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. Confidence regularized self-training. In *Proc. of IEEE/CVF international conference on computer vision*, pages 5982–5991, 2019. 7