Supplementary: Semantic-Aware Domain Generalized Semantic Segmentation

In this supplementary, additional information is provided as following five aspects:

A. comprehensive evaluation on other DG settings;
B. comparison with Domain Adaptation methods;
C. detailed framework structure and implementations;
D. discussion on computational complexity;
E. additional visualization of semantic segmentation.

Appendix A. Evaluation on other DG settings

Existing DG methods [3, 11–13, 20] only focus on three domain generalization settings, *i.e.* (1) G $\rightarrow$ C, B, M, & S; (2) S $\rightarrow$ C, B, M, & G and (3) C $\rightarrow$ G, S, B, & M while losing the sight of the other two DG settings: (4) B $\rightarrow$ G, S, C, & M and (5) M $\rightarrow$ G, S, C, & M. However, recently several studies [2, 6] stress the importance of the last two settings. Therefore, we consider a more comprehensive evaluation which performs generalization from each of them. Results on the last two settings are reported in Tab. 1, suggesting that our model consistently achieves the state-of-the-art results on all settings and backbones.

Table 1. Performance comparison in terms of mIoU (%) between DG methods. The best and second best results are **highlighted** and underlined, respectively. † denotes our re-implemention of the respective method. G, C, B, M and S denote GTA5, Cityscapes, BDDS, Mapillary and SYNTHIA, respectively.

Methods	Backbone	Train on BDDS (B)				Train on Mapillary (M)			
		→G	→S	→C	→M	→G	→S	→C	→B
Baseline		25.30	21.08	38.76	23.48	25.34	22.16	36.13	24.17
IBN† [11]		29.47	26.40	39.72	26.12	29.68	26.31	41.39	29.48
SW† [12]		27.10	25.23	39.54	25.67	28.70	25.57	39.66	28.37
DRPC† [20]	VGG-16	32.83	28.06	40.17	29.00	31.53	28.03	45.15	30.46
GTR† [13]		32.75	27.63	41.06	29.71	32.67	27.32	44.47	31.83
ISW† [3]		32.60	28.58	42.21	30.54	32.65	28.44	45.68	32.06
Ours		**34.11**	**30.13**	**44.62**	**32.06**	**33.86**	**30.67**	**47.51**	**33.07**
Baseline		26.12	21.65	39.03	23.87	25.46	23.41	36.79	26.37
IBN† [11]		28.97	25.42	41.06	26.56	30.68	27.01	42.77	31.01
SW† [12]		27.68	25.37	40.88	25.83	28.47	27.43	40.69	30.54
DRPC† [20]	ResNet-50	33.19	29.77	41.30	31.86	33.04	29.59	46.21	32.92
GTR† [13]		33.25	30.61	42.58	30.73	32.86	30.26	45.84	32.63
ISW† [3]		32.74	30.53	43.50	31.57	33.37	30.15	46.43	32.57
Ours		**34.75**	**31.84**	**44.94**	**33.21**	**34.01**	**31.55**	**48.65**	**34.62**
Baseline		25.84	24.62	42.06	24.70	26.81	23.74	39.68	27.19
IBN† [11]		30.28	29.06	44.92	29.90	32.07	28.83	44.89	30.27
SW† [12]		28.34	26.74	44.28	27.58	30.31	24.06	42.33	28.65
DRPC† [20]	ResNet-101	34.13	31.75	46.73	32.63	36.40	30.27	46.16	32.17
GTR† [13]		35.26	31.98	45.34	33.27	34.65	29.56	47.68	33.98
ISW† [3]		34.87	32.89	46.15	34.17	35.53	30.92	48.54	34.02
Ours		**37.56**	**33.83**	**48.32**	**35.24**	**37.72**	**32.63**	**50.07**	**35.79**

Table 2. Comparison results between ours and Domain Adaptation methods on GTA5→Cityscapes. DA and DG denote Domain Adaption and Domain Generalization respectively.

Backbone	Task	Method	Access Tgt	mIoU
VGG-16	DA	FCN wild [6]	✓	27.1
		CDA [21]	✓	28.9
		CyCADA [5]	✓	34.8
		ROAD [1]	✓	35.9
		I2I [9]	✓	31.8
		AdaptSegNet [14]	✓	35.0
		SSF-DAN [4]	✓	37.7
		DCAN [18]	✓	36.2
		CBST [22]	✓	30.9
		CLAN [8]	✓	36.6
		ADVENT [16]	✓	36.1
		DPR [15]	✓	37.5
		BDL [7]	✓	41.3
		FDA [19]	✓	**42.2**
	DG	Ours	×	38.2
Resnet-101	DA	CyCADA [5]	✓	42.7
		ROAD [1]	✓	39.4
		I2I [9]	✓	35.4
		AdaptSegNet [14]	✓	41.4
		DCAN [18]	✓	41.7
		CLAN [8]	✓	43.2
		ADVENT [16]	✓	43.8
		DPR [15]	✓	46.5
		IntraDA [10]	✓	46.3
		DADA [17]	✓	**47.3**
	DG	Ours	×	45.3

Table 3. Comparison results between ours and Domain Adaptation methods on SYNTHIA→Cityscapes.

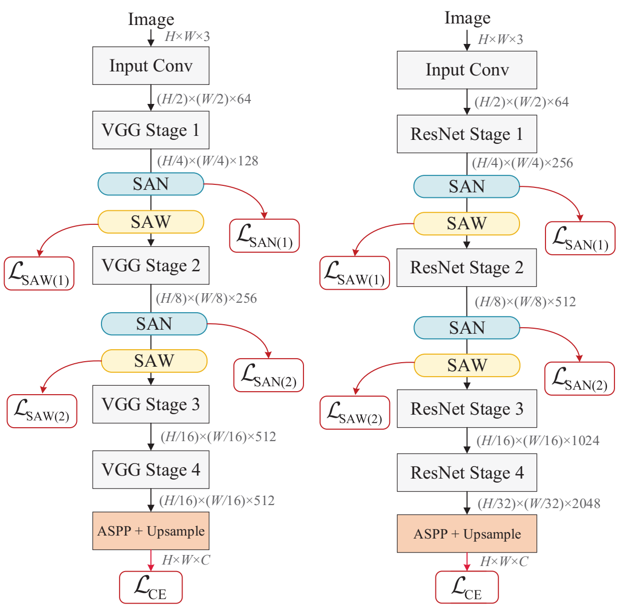
Backbone	Task	Method	Access Tgt	mIoU
VGG-16	DA	FCN wild [6]	✓	20.2
		CDA [21]	✓	29.0
		ROAD [1]	✓	36.2
		DCAN [18]	✓	35.4
		CBST [22]	✓	35.4
		ADVENT [16]	✓	31.4
		DPR [15]	✓	33.7
		BDL [7]	✓	39.0
		FDA [19]	✓	**40.5**
	DG	Ours	×	37.4
Resnet-101	DA	ADVENT [16]	✓	40.8
		DPR [15]	✓	40.0
		IntraDA [10]	✓	41.7
		DADA [17]	✓	**42.6**
	DG	Ours	×	40.9

Appendix B. Comparison with DA methods

Domain Adaptation (DA) methods require access to the target domain to solve domain shift problems. In contrast, our method is designed in Domain Generalization (DG) manner for broad generalization to totally unseen domains without accessing any target domain data. Therefore, the target domain-accessible DA methods have the inherent performance superiority than DG methods which are target domain-agnostic. In order to see whether our approach is up to the performance standard of DA, we compare the results of our method with those reported from several previous state-of-the-art DA methods. From Tab. 2 and Tab. 3, we can see that the generalization performance of our method outperforms the adaptation performance of most other techniques. In addition, no target-domain data is needed in our method, resulting in more extensive applicability.

Appendix C. Further Implementation Details

We follow previous work [3, 11] to adopt normalization and whitening at the first two stages of convolution layers, since shallow layers encode more style information [11]. As shown in Fig. 1, for each backbone network, we impose SAN and SAW after stage 1 and stage 2.



(a) VGG-16 (b) ResNet-50 and ResNet-101

Figure 1. Detailed Architecture of our approach with the backbone of VGG and ResNet.

Appendix D. Computational complexity

As shown in Tab. 4, compared to the baseline, our methods performs domain generalization with negligible addition in both training and inference time. This is because the proposed modules are only implemented in the first two layers of the network, for only four main categories *i.e.* $C = 4$, see Sec. 5.5 of the main paper. The additional memory overhead from our modules is less than 2G.

Table 4. Comparison on computation cost.

Backbone	Methods	Memory (G)	Training Time (s)	Inference Time (ms)
Vgg-16	Baseline	5.28	0.37	48.17
	Ours	7.41	0.39	48.20
Res-50	Baseline	6.43	0.40	48.84
	Ours	8.04	0.41	48.86
Res-101	Baseline	8.24	0.43	50.31
	Ours	10.17	0.45	50.37

Appendix E. More qualitative results

Fig. 2 shows more qualitative results under various unseen domains. We demonstrate the effects of the proposed semantic-aware feature matching by comparing the segmentation results from our proposed approach and the baseline. In the setting of GTA5 $\rightarrow$ Mapillary, the baseline fails to cope with these weather changes, while ours still shows fair results. Under the illumination changes as shown in GTA5 $\rightarrow$ BDDS, our method finds the road and sidewalk clearer than the baseline.

References

[1] Yuhua Chen, Wen Li, and Luc Van Gool. Road: Reality oriented adaptation for semantic segmentation of urban scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7892–7901, 2018. 1

[2] Yi-Hsin Chen, Wei-Yu Chen, Yu-Ting Chen, Bo-Cheng Tsai, Yu-Chiang Frank Wang, and Min Sun. No more discrimination: Cross city adaptation of road scene segmenters. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1992–2001, 2017. 1

[3] Sungha Choi, Sanghun Jung, Huiwon Yun, Joanne T Kim, Seungryong Kim, and Jaegul Choo. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11580–11590, 2021. 1, 2

[4] Liang Du, Jingang Tan, Hongye Yang, Jianfeng Feng, Xiangyang Xue, Qibao Zheng, Xiaoqing Ye, and Xiaolin Zhang. Ssf-dan: Separated semantic feature based domain adaptation network for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 982–991, 2019. 1

[5] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1989–1998, 2018. 1

[6] Judy Hoffman, Dequan Wang, Fisher Yu, and Trevor Darrell. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. *arXiv preprint arXiv:1612.02649*, 2016. 1

[7] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6936–6945, 2019. 1

[8] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2507–2516, 2019. 1

[9] Zak Murez, Soheil Kolouri, David Kriegman, Ravi Ramamoorthi, and Kyungnam Kim. Image to image translation for domain adaptation. In *Proceedings of the IEEE/CVF*

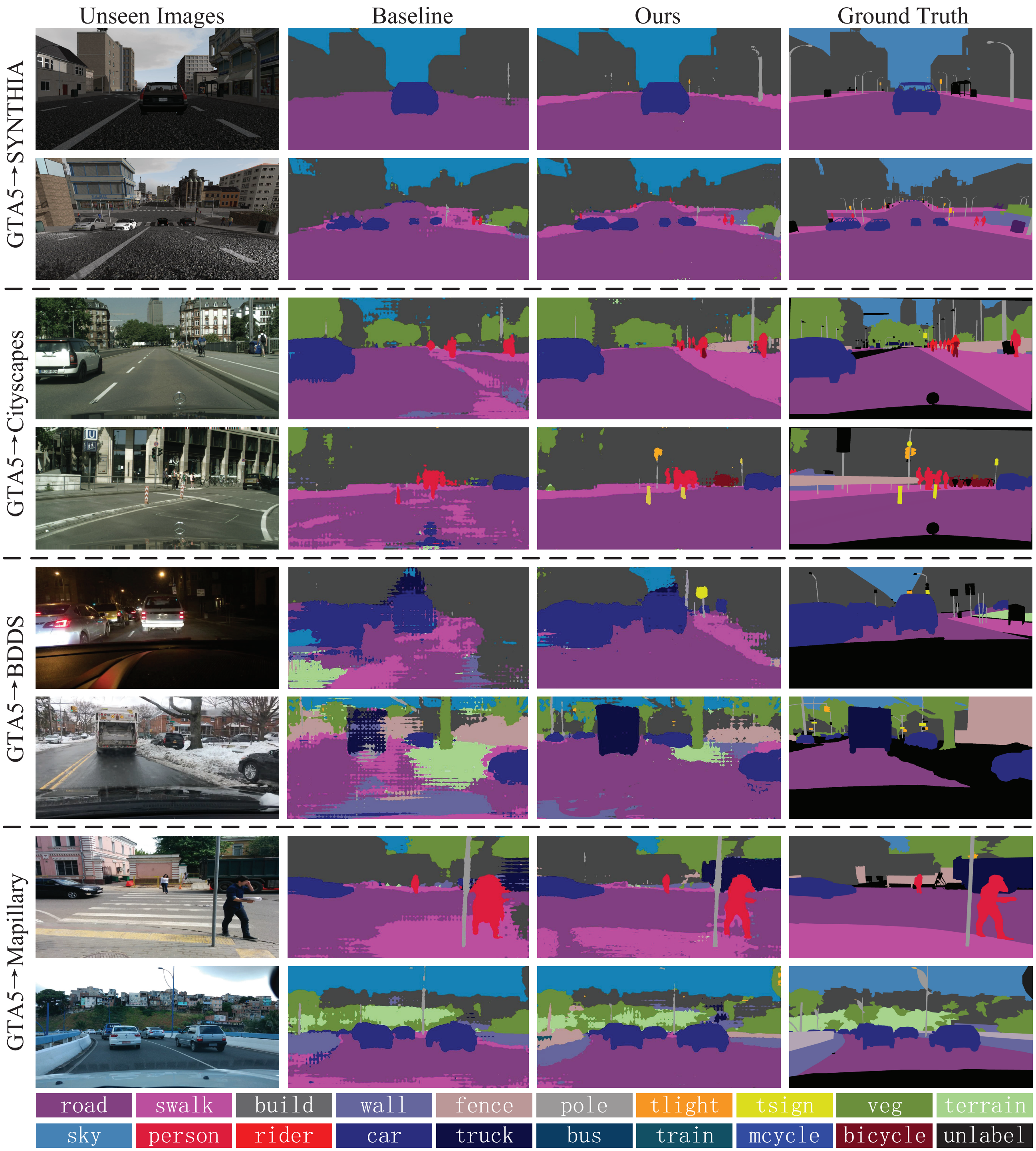


Figure 2. Qualitative results of our approach generalizing from GTA5 to other four domains.

Conference on Computer Vision and Pattern Recognition (CVPR), pages 4500–4509, 2018. 1

[10] Fei Pan, Inkyu Shin, Francois Rameau, Seokju Lee, and In So Kweon. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3764–3773, 2020. 1

[11] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 464–479, 2018. 1, 2

[12] Xingang Pan, Xiaohang Zhan, Jianping Shi, Xiaoou Tang,

and Ping Luo. Switchable whitening for deep representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1863–1871, 2019. 1

[13] Duo Peng, Yinjie Lei, Lingqiao Liu, Pingping Zhang, and Jun Liu. Global and local texture randomization for synthetic-to-real semantic segmentation. *IEEE Transactions on Image Processing (TIP)*, 30:6594–6608, 2021. 1

[14] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7472–7481, 2018. 1

[15] Yi-Hsuan Tsai, Kihyuk Sohn, Samuel Schulter, and Manmohan Chandraker. Domain adaptation for structured output via discriminative patch representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1456–1465, 2019. 1

[16] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2517–2526, 2019. 1

[17] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Dada: Depth-aware domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7364–7373, 2019. 1

[18] Zuxuan Wu, Xintong Han, Yen-Liang Lin, Mustafa Gokhan Uzunbas, Tom Goldstein, Ser Nam Lim, and Larry S Davis. Dcan: Dual channel-wise alignment networks for unsupervised scene adaptation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 518–534, 2018. 1

[19] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4085–4095, 2020. 1

[20] Xiangyu Yue, Yang Zhang, Sicheng Zhao, Alberto Sangiovanni-Vincentelli, Kurt Keutzer, and Boqing Gong. Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2100–2110, 2019. 1

[21] Yang Zhang, Philip David, and Boqing Gong. Curriculum domain adaptation for semantic segmentation of urban scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2020–2030, 2017. 1

[22] Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Domain adaptation for semantic segmentation via class-balanced self-training. *arXiv preprint arXiv:1810.07911*, 2018. 1