Supplemental Material for "SiamDoGe: Domain Generalizable Semantic Segmentation using Siamese Network"

Zhenyao Wu[1], Xinyi Wu[1], Xiaoping Zhang[2], Lili Ju[1,†], and Song Wang[1,†]

[1] University of South Carolina
{zhenyao, xinyiw}@email.sc.edu ju@math.sc.edu songwang@cec.sc.edu
[2] Wuhan University
xpzhang.math@whu.edu.cn

In this supplementary material, we first report some additional quantitative (Sec. A.1) results and qualitative results (Sec. A.2). In Sec B, we further discuss the limitation of our method and some directions for future work. Finally, we provide a summary of the datasets we employed in Sec. C.

A Additional Results

A.1 Quantitative results

Detailed results on the ACDC. To study the impact of SiamDoGe on individual adverse conditions, we further evaluate our model on each condition of ACDC dataset including fog, nighttime, rain and snow. The results for the settings GTAV → ACDC (G → A) and Cityscapes → ACDC (C → A) are reported in Table 1 and Table 2, respectively. It can be seen that the proposed SiamDoGe surpasses the state-of-the-arts on all conditions under both settings. Besides, across the four conditions, we find that our method achieves more performance gains on the nighttime condition compared with the second best, *i.e.*, 4.28 mIoU on G → A and 6.01 mIoU on C → A. In addition, both IBN-Net [4] and RobustNet [1] decrease the performance on this condition.

Table 1. Quantitative comparison results of our SiamDoGe with the existing state-of-the-art DG approaches for semantic segmentation under the setting G → A. All the methods use ResNet-50 as backbone. Note that the performance on "all" conditions is not an simple average of the four individual conditions.

Method	Fog	Night	Rain	Snow	All
IBN-Net [4]	30.35	4.80	13.63	9.99	22.55
RobustNet [1]	32.56	6.32	33.02	29.97	25.46
SiamDoGe	**36.45** (+3.89)	**10.60** (+4.28)	**35.84** (+2.82)	**30.71** (+0.74)	**29.25** (+3.79)

Table 2. Quantitative comparison results of our SiamDoGe and the existing state-of-the-art DG approaches for semantic segmentation under the setting C $\rightarrow$ A. All the methods use ResNet-50 as backbone.

Methods	Fog	Night	Rain	Snow	All
DeepLabv3+ [1]	45.7	25.0	50.0	42.0	41.6
IBN-Net [4]	63.56	21.5	51.91	49.46	44.05
RobustNet [1]	64.62	23.18	54.07	50.00	46.91
SiamDoGe	**67.42** (+2.80)	**31.01** (+6.01)	**54.44** (+0.37)	**55.92** (+5.92)	**52.34** (+5.43)

The per-categories comparison. We also provide the per-category results for the main comparison experiments (supplement to Table 1 and Table 2 in the main paper) in Table 3 and Table 4.

A.2 Qualitative results

In this section, we provide more visual comparison of our SiamDoGe with RobustNet [1] under the setting G $\rightarrow$ {C, B, M, S, A} in Figure 1 and the setting C $\rightarrow$ {G, B, M, S, A} in Figure 2, respectively.

B Limitation and discussion

Compared with alternatives that improve the domain generalization ability with a single branch, our method needs more training time. This limitation is caused by the inherent architecture of Siamese networks.

C Dataset details

Finally, we also provide an overview of the datasets used in our experiments in Table 5 for a clear illustration.

[1] Note that the numbers are borrowed from ACDC [7] which serves as a baseline of the setting Cityscapes $\rightarrow$ ACDC.

Table 3. The per-category comparison results of our SiamDoGe with the state-of-the-arts under the setting G → {C, B, M, S, A}.

	Method	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	**mIoU**
C	IBN-Net [4]	51.3	24.1	59.7	14.1	25.9	23.0	30.9	15.7	85.0	40.6	67.8	60.6	4.9	76.7	23.7	16.3	0.8	11.9	10.1	33.9
	RobustNet [1]	60.5	25.5	65.4	21.6	23.7	25.8	33.3	15.5	85.4	38.5	70.3	61.9	9.3	82.7	25.4	21.1	0.0	16.8	12.2	36.6
	SiamDoGe	84.4	35.5	79.2	27.9	25.2	29.7	40.5	23.1	84.8	29.4	69.1	64.3	29.8	86.5	32.7	27.9	2.5	24.3	19.5	43.0
B	IBN-Net [4]	68.9	28.8	56.7	6.0	21.4	31.9	35.0	21.7	66.6	22.1	66.9	50.2	5.8	70.3	12.8	7.2	0.0	23.2	18.1	32.3
	RobustNet [1]	74.9	30.4	65.2	11.5	22.6	34.9	36.9	22.2	69.0	26.2	77.4	50.3	10.0	72.4	16.2	6.1	0.0	31.3	11.4	35.2
	SiamDoGe	81.1	39.3	60.4	10.0	33.3	29.7	34.6	25.0	74.0	27.1	72.9	54.5	15.7	80.9	25.7	6.0	0.0	34.7	8.5	37.5
M	IBN-Net [4]	66.4	32.9	57.1	10.5	24.9	31.5	38.0	38.9	73.5	28.4	82.0	56.4	9.1	73.1	30.7	16.0	12.6	16.6	18.6	37.8
	RobustNet [1]	74.8	36.2	66.1	15.2	26.0	35.4	37.9	36.6	74.3	30.2	88.2	58.6	11.1	78.7	33.3	14.4	11.0	20.4	17.8	40.3
	SiamDoGe	84.7	42.1	40.4	19.2	28.9	36.1	40.6	40.3	76.5	29.3	57.0	60.4	23.8	83.2	34.6	11.8	6.3	33.2	23.9	40.6
S	IBN-Net [4]	48.3	46.3	77.2	7.5	6.9	20.8	11.0	9.9	62.2	0.0	89.1	55.3	3.6	54.3	0.0	12.7	0.0	15.2	10.1	27.9
	RobustNet [1]	52.0	49.4	78.0	6.5	5.1	21.2	11.1	7.8	60.0	0.0	89.6	53.3	4.3	54.0	0.0	9.2	0.0	17.3	8.9	28.3
	SiamDoGe	54.8	46.8	78.0	8.4	2.9	25.4	15.0	8.8	56.2	0.0	90.0	52.0	9.8	53.3	0.0	12.8	0.0	14.2	10.2	28.3
A	IBN-Net [4]	44.8	19.9	47.9	13.6	9.9	15.8	38.5	8.2	46.2	19.0	72.0	30.5	4.1	34.5	5.0	5.4	0.1	11.8	1.1	22.6
	RobustNet [1]	50.6	18.4	49.2	14.6	11.1	20.6	39.4	8.7	53.6	16.7	70.4	33.6	3.1	60.9	9.1	6.8	0.3	15.9	0.5	25.5
	SiamDoGe	69.1	20.0	43.8	25.4	14.2	22.9	40.8	12.3	59.8	22.8	63.7	37.2	11.9	70.6	17.4	7.2	2.7	12.8	1.1	29.3

Table 4. The per-category comparison results of our SiamDoGe with the state-of-the-arts under the setting C $\rightarrow$ {G, B, M, S, A}.

	Method	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	**mIoU**
G	IBN-Net [4]	77.7	27.8	78.7	40.5	19.8	40.5	49.0	28.5	72.4	43.5	84.6	63.8	40.8	75.7	55.7	6.7	0.0	41.8	8.5	45.1
	RobustNet [1]	78.8	32.6	80.1	41.9	20.1	40.1	47.1	24.5	71.3	42.1	88.5	61.5	34.8	73.2	56.7	11.4	0.0	42.6	7.8	45.0
	SiamDoGe	78.3	32.6	80.2	39.1	22.2	35.9	43.8	33.0	71.5	26.4	87.7	62.5	39.9	78.1	53.5	19.9	0.0	41.4	10.5	45.1
B	IBN-Net [4]	90.8	53.2	77.9	16.8	31.2	41.4	42.5	43.0	81.8	34.8	89.3	60.9	38.6	84.0	32.3	24.4	0.0	41.8	37.8	48.6
	RobustNet [1]	91.3	55.0	78.8	20.1	31.7	41.1	41.1	43.3	81.9	35.6	90.2	61.7	40.8	84.1	36.4	38.8	0.0	48.7	43.1	50.7
	SiamDoGe	92.0	56.8	76.6	16.7	37.7	32.9	38.9	44.6	82.3	40.6	89.3	61.9	30.2	86.2	34.4	52.0	2.7	56.8	46.7	51.5
M	IBN-Net [4]	88.8	38.8	78.5	36.5	45.6	47.8	51.0	67.3	87.1	49.8	92.2	69.0	45.2	88.1	19.0	47.7	26.6	47.3	57.4	57.0
	RobustNet [1]	88.9	40.9	80.1	37.2	43.9	47.4	50.6	65.1	86.8	49.3	95.8	70.1	49.1	87.8	45.6	47.0	29.0	47.2	52.4	58.6
	SiamDoGe	88.7	40.6	80.4	29.6	45.4	42.1	46.8	65.3	86.1	44.9	95.7	68.9	39.8	88.8	58.9	57.8	37	47.7	56.5	59.0
S	IBN-Net [4]	50.2	19.0	73.8	1.7	1.0	27.9	19.3	5.0	62.7	0.0	89.4	52.3	21.6	38.5	0.0	4.9	0.0	14.0	15.5	26.1
	RobustNet [1]	49.8	25.3	74.4	2.8	1.1	28.6	20.7	5.8	60.0	0.0	89.2	49.8	19.4	36.3	0.0	12.3	0.0	9.8	12.2	26.2
	SiamDoGe	48.1	20.4	71.7	3.7	0.7	21.9	22.4	6.4	56.6	0	89.9	49.6	19	45.5	0.0	15.4	0.0	15.6	20.0	26.7
A	IBN-Net [4]	82.3	43.8	58.0	28.1	29.1	36.7	63.2	47.1	76.3	26.6	75.1	43.7	17.5	64.0	5.0	48.2	46.2	24.4	21.6	44.1
	RobustNet [1]	83.3	47.7	63.4	30.0	26.3	38.1	63.1	49.1	73.7	25.6	76.4	44.8	20.3	60.1	37.2	49.5	50.8	27.8	23.9	46.9
	SiamDoGe	86.9	55	64.4	33.0	31.9	34.0	61.3	51.6	79.4	31.5	80.8	51.1	30.4	78.8	46.7	49.9	55.0	37.9	34.8	52.3

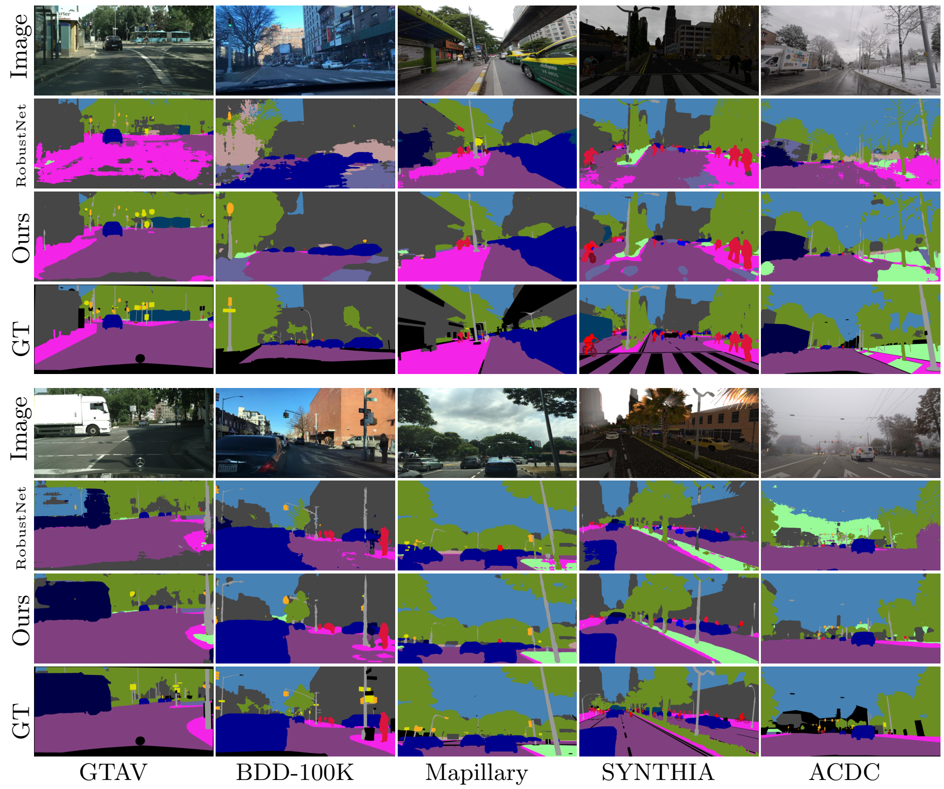

Fig. 1. Qualitative comparison results of our SiamDoGe with RobustNet [1] under the setting G $\rightarrow$ {C, B, M, S, A}.

Table 5. Illusrtation of datasets. GTAV and Cityscapes are treated as either seen or unseen datasets and the rest are only treated as unseen dataset for evaluation. *The minimum resolution is reported.

Dataset	Synthetic	Real	Train	Eval	Resolution
GTAV [5]	✓		18,785	6,382	1,914×1,052
Cityscapes [2]		✓	2,975	500	2,048×1,024
SYNTHIA [6]	✓		-	2,820	960×720
Mapillary [3]		✓	-	2,000	1,920×1,080*
BDD-100K [8]		✓	-	1,000	1,280×720
ACDC [7]		✓	-	406	1,920×1,080

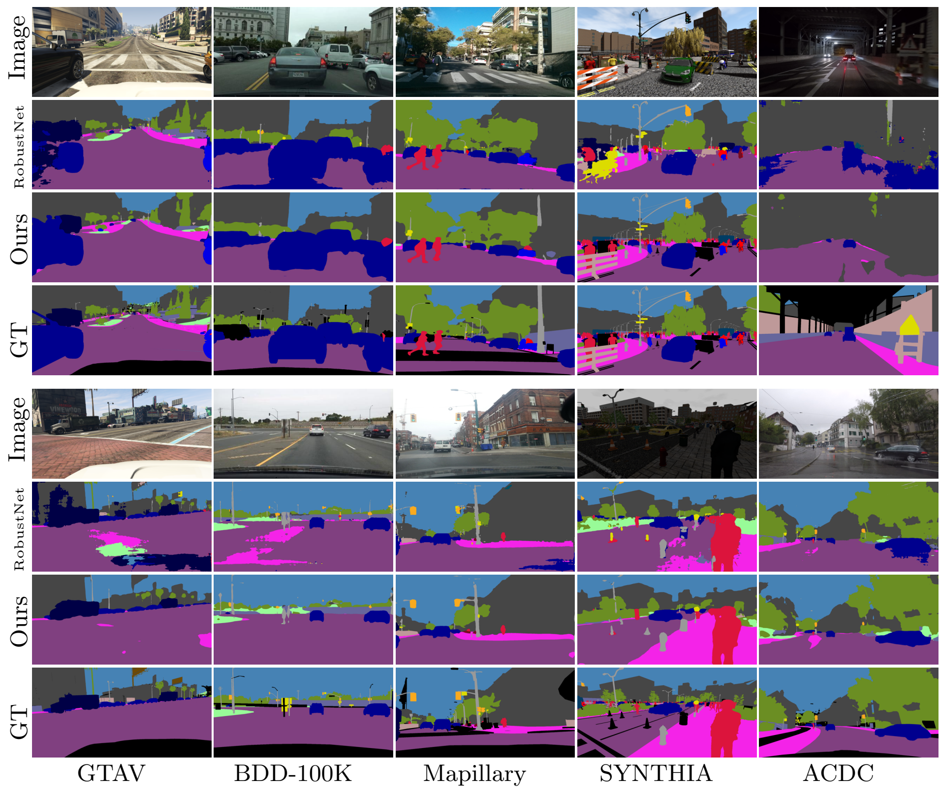

Fig. 2. Qualitative comparison results of our SiamDoGe with RobustNet [1] under the setting C $\rightarrow$ {G, B, M, S, A}.

References

1. Choi, S., Jung, S., Yun, H., Kim, J.T., Kim, S., Choo, J.: Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In: CVPR. pp. 11580–11590 (2021) 1, 2, 3, 4, 5, 6
2. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: CVPR. pp. 3213–3223 (2016) 5
3. Neuhold, G., Ollmann, T., Rota Bulo, S., Kontschieder, P.: The mapillary vistas dataset for semantic understanding of street scenes. In: ICCV (2017) 5
4. Pan, X., Luo, P., Shi, J., Tang, X.: Two at once: Enhancing learning and generalization capacities via ibn-net. In: ECCV. pp. 464–479 (2018) 1, 2, 3, 4
5. Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: Ground truth from computer games. In: ECCV (2016) 5
6. Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A.M.: The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In: CVPR. pp. 3234–3243 (2016) 5
7. Sakaridis, C., Dai, D., Van Gool, L.: Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. arXiv preprint arXiv:2104.13395 (2021) 2, 5
8. Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T.: Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In: CVPR (2020) 5