Style-Hallucinated Dual Consistency Learning for Domain Generalized Semantic Segmentation (Supplementary Material)

Yuyang Zhao[1], Zhun Zhong[2], Na Zhao[1], Nicu Sebe[2], and Gim Hee Lee[1]

[1] National University of Singapore
[2] University of Trento
https://github.com/HeliosZhao/SHADE

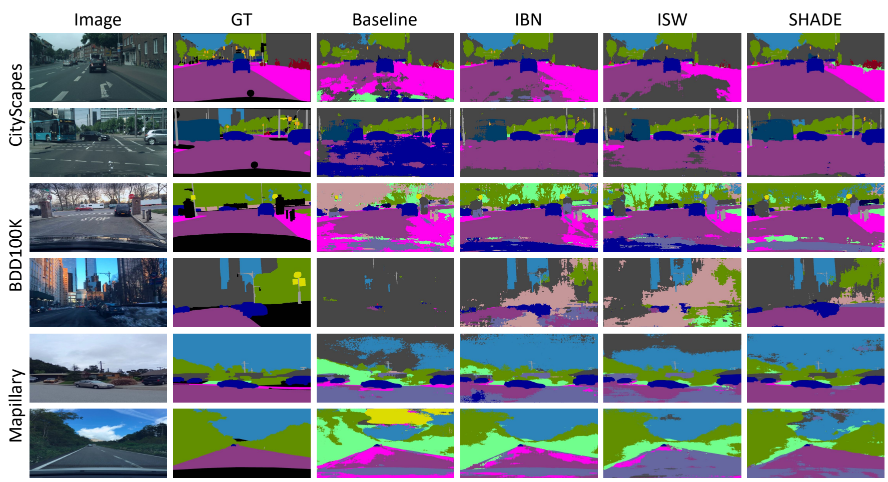

Fig. 1. Qualitative comparison of segmentation results.

1 Visualization

Qualitative results. We compare the segmentation results among baseline, IBN-Net [4], ISW [1] and SHADE on CityScapes [2], BDD100K [5] and Mapillary [3] in Fig. 1. We obtain two observations from Fig. 1. First, SHADE consistently outperforms other methods under different target conditions (*e.g.*, sunny, cloudy and overcast). Second, SHADE can well deal with both "stuff" classes (*e.g.*, road) and "things" classes (*e.g.*, bus and bicycle). The above two observations demonstrate that SHADE is robust to style variation and has strong ability in segmenting unseen real-world images.
Style Visualization. To better understand our SHM, we visualize the style-diversified samples with an auto-encoder. Examples are shown in Fig. 2. SHM

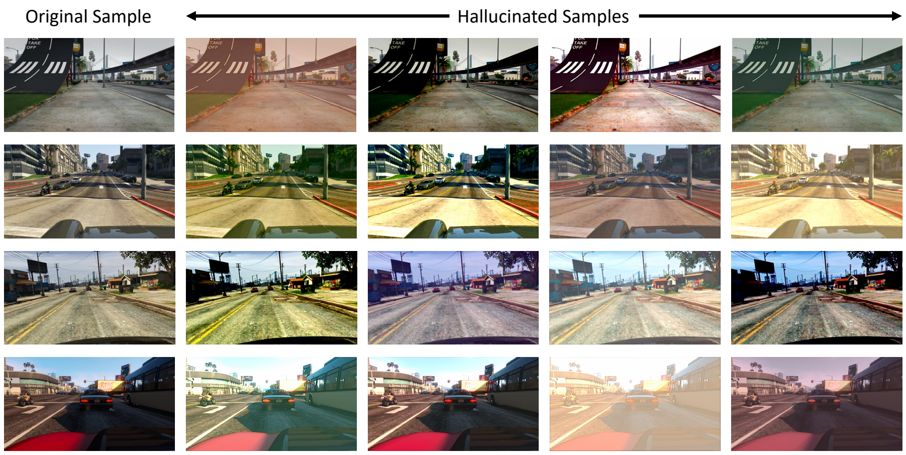

Fig. 2. Visualization of style-diversified samples.

replaces the original style features with the combination of basis styles to obtain new samples of different styles, *e.g.*, weather change (from overcast to sunny) and time change (from dusk to midday).

2 Per-class IoU

We show the per-class IoU in Tab. 1 to verify the effectiveness of retrospection consistency (RC). RC can improve the performance of both "things" classes (+**2.47%**) and "stuff" classes (+**1.54%**), and the improvement on the "things" classes is more significant.

Table 1. Per-class IoU on CityScapes dataset. Models are trained on GTAV with ResNet-50 backbone.

	Road	S.walk	Build.	Wall	Fence	Pole	Tr.Light	Sign	Veget.	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	M.bike	Bike	mIoU
SHADE w.o. RC	81.4	36.3	75.1	**27.2**	25.5	33.0	**34.8**	**19.1**	84.9	35.5	65.3	64.7	26.5	**85.2**	23.7	28.7	5.8	**32.6**	25.2	42.7
SHADE	**82.8**	**37.6**	**77.1**	26.0	**29.3**	**33.3**	33.5	17.9	**85.6**	**36.1**	**69.4**	**67.0**	**30.8**	84.8	**31.9**	**31.4**	**15.5**	31.2	**27.0**	**44.6**

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 (2021) 1
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 (2016) 1
3. Neuhold, G., Ollmann, T., Rota Bulo, S., Kontschieder, P.: The mapillary vistas dataset for semantic understanding of street scenes. In: ICCV (2017) 1
4. Pan, X., Luo, P., Shi, J., Tang, X.: Two at once: Enhancing learning and generalization capacities via ibn-net. In: ECCV (2018) 1
5. 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) 1