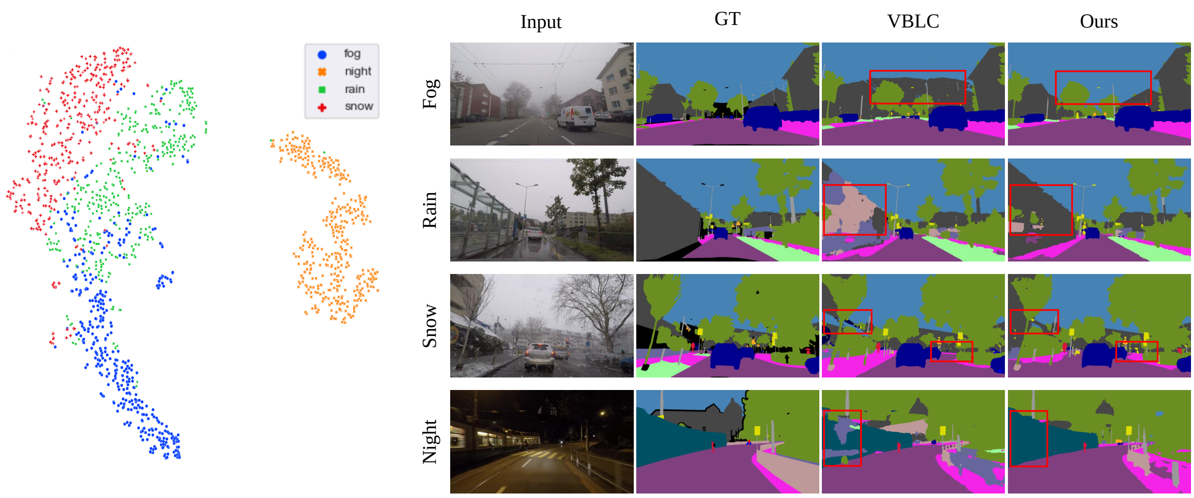
A Subdomain-Specific Knowledge Distillation for Unsupervised Domain Adaptation in Adverse Weather Conditions

Gyuwon Choi[1†], Yejin Lee[1‡], Donggon Jang, Dae-Shik Kim*

Department of Electrical Engineering, KAIST

September 11, 2023



(a) t-SNE map for ACDC dataset (b) Qualitative results on ACDC dataset

Figure 1: (a) The feature maps from the Resblock-1 of ResNet-101 are visualized using t-SNE after applying the Gram matrix. (b) Visualization of segmentation results on ACDC validation set.

Semantic segmentation is a branch of scene understanding that can be used to perceive urban driving scenes in applications such as self-driving cars. However, existing models are trained on clear weather images, and they suffer from performance degradation when weather, season, and brightness changes.

To address this issue, unsupervised domain adaptation methods have been proposed to adapt models trained on clear weather to adverse weather without ground-truth. Refign [1] uses reference images where images are obtained from the same locations as the target images but in clear weather to reduce domain gap. VBLC [2] utilizes Visibility Boost

*Corresponding author

[1]These authors contributed equally to this work.

†gyuwonchoi@kaist.ac.kr

‡yejin@kaist.ac.kr

Module to enhance the visibility of target images to make them similar to the source images, thereby reducing the domain gap. However, all of these methods assume that the target domain follows a single distribution.

In this paper, we show that there is a significant gap between distributions in adverse weather conditions. To this end, we visualize the style statistics of target domain ACDC dataset [3], which includes images of fog, rain, snow, and night using t-SNE [4]. As shown in Figure 1, the style distribution of fog, rain, and snow (dubbed FRS) are clustered, indicating that these three weather conditions have similar style distributions. On the other hand, the night stands apart from the other three conditions. This suggests that there is a gap between the distributions within a single target domain in adverse weather dataset, and the distributions with a large gap need to be distinguished and adapted separately.

We propose to train two subdomain teachers, one for FRS and another for night subdomain individually. The knowledge of the two different subdomains is distilled to the student in an online manner with symmetric cross-entropy. This allows the student to have complementary knowledge from each weather domain.

Method. The architecture consists of two subdomain teachers, a student, and a student exponential moving average (EMA) model. Our method performs self-training on each subdomain with the teacher and then distills the learned subdomain knowledge to the student. The student EMA weights are updated with the exponential moving average of the student weights, which accumulates overall target domain knowledge. With student EMA, we can accumulate specialized knowledge for each subdomain, and utilize it to obtain less noisy pseudo labels for self-training. At inference time, we only use the student model, without considering the gap between the sub-target domains.

Subdomain-specialized Teacher. We observed that the FRS and night domains have different distributions. To address this, we divide the target domain into two subdomains: FRS and night. We then train FRS and night teacher respectively to obtain subdomain-specific knowledge. To extract domain-specific knowledge to each teacher, the source samples are matched to the target style using histogram matching. This helps subdomain teachers to learn domain-specific knowledge by making the source images closer to the target images. We train the teachers with converted source samples and cross-entropy in supervised manner. In the case of target images, we apply ClassMix [5] on target and source to reduce the gap. We input the mixed images to the teachers and EMA to obtain predictions and pseudo labels (PLs) each, and then applied cross-entropy. From this process, we can obtain teachers which have domain-specific knowledge and we call them subdomain-specialized teachers (SSTs).

Online Knowledge Distillation. We employ online knowledge distillation to transfer complementary domain information to the student. Two teachers distill their knowledge to the student via symmetric cross-entropy loss. This allows the student to learn domain-specific knowledge that has been extracted by the SSTs. Additionally, symmetric cross-entropy is known to prevent abrupt gradient changes caused by overconfident predictions and to stabilize the optimization process [6].

Experiments. We conduct experiments on the Cityscapes [7] dataset as the source domain and the ACDC dataset [3] as the target domain. We utilize the DAFormer [8] and DeepLab-v2 [9] architectures and compare our method to MIC [10] and VBLC [2]. The experiment is conducted on the validation set. We perform source warm-up on two teachers and a student at the beginning of training. This ensures that the model is initialized with a proper starting point [11].

Our method achieves 64.82 mIoU on the DAFormer and 49.97 mIoU on the DeepLab-v2. This is a performance gain of 2.98 and 2.9 over VBLC and MIC, respectively, on the DAFormer. We achieve mIoU of 76.84 on the fog, 68.55 on the rain, 64.78 on the snow, and 40.24 on the night while VBLC achieves mIoU of 72.78 on the fog, 65.53 on the rain, 60.45 on the snow, and 41.74 on the night. Our method outperforms VBLC in the fog, rain, and snow domains. This shows that dividing the target domain into subdomains and learning specialized knowledge with each teacher is effective in improving performance. Our method is comparable to VBLC in the night. We hypothesize this is due to the lack of night images.

Conclusion. We propose an unsupervised domain adaptation method for semantic segmentation in adverse weather conditions. We discovered that there are differences in distribution within the target domain ACDC. We then propose a method to adapt each subdomain individually, by creating two subdomain-specialized teachers and distilling complementary domain information to a student. Our method outperforms VBLC and MIC on the ACDC dataset. Although our method outperforms on fog, rain, and snow (dubbed FRS), it achieves a relatively low mIoU on the night domain. This is because the number of night samples is scarce compared to FRS, and there is a large difference in illumination between the sky in the night and the sky in the FRS. We will address this issue in future work.

References

[1] D. Brüggemann, C. Sakaridis, P. Truong, and L. Van Gool, "Refign: Align and refine for adaptation of semantic segmentation to adverse conditions," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 3174–3184.

[2] M. Li, B. Xie, S. Li, C. H. Liu, and X. Cheng, "Vblc: Visibility boosting and logit-constraint learning for domain adaptive semantic segmentation under adverse conditions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 7, 2023, pp. 8605–8613.

[3] C. Sakaridis, D. Dai, and L. Van Gool, "Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 765–10 775.

[4] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[5] V. Olsson, W. Tranheden, J. Pinto, and L. Svensson, "Classmix: Segmentation-based data augmentation for semi-supervised learning," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1369–1378.

[6] M. Döbler, R. A. Marsden, and B. Yang, "Robust mean teacher for continual and gradual test-time adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7704–7714.

[7] M. Cordts, M. Omran, S. Ramos, T. Scharwächter, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset," in *CVPR Workshop on the Future of Datasets in Vision*, vol. 2. sn, 2015.

[8] L. Hoyer, D. Dai, and L. Van Gool, "Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9924–9935.

[9] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[10] L. Hoyer, D. Dai, H. Wang, and L. Van Gool, "Mic: Masked image consistency for context-enhanced domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11 721–11 732.

[11] O. Chakraborty, A. Sahoo, R. Panda, and A. Das, "Anyda: Anytime domain adaptation," in *The Eleventh International Conference on Learning Representations*, 2022.