Towards Domain-Aware Knowledge Distillation for Continual Model Generalization - Supplementary material

Nikhil Reddy[1] Mahsa Baktashmotlagh[2]
Chetan Arora[3]
[1]UQ-IIT Delhi Research Academy (UQIDAR) [2]UQ, Australia [3]IIT Delhi, India
https://dose-iitd.github.io/

1. Comparison with Existing TTA methods

Method	Inference time
IBNNet [10]	73
DoSe (ours)	872
w/o Self-Distillation	324
w/o Exemplar-based Buffer	548

Table 1. Inference time comparison of DoSe framework on IBN-Net [10] with different loss variants, Inference time values are in seconds.

We compare the existing SOTA Test-time Adaptation methods on the RefineNet [5] pre-trained model trained on the source (daytime) dataset. Results are reported in Tab. 4. We compare our proposed DoSe with existing TTA Adaptation methods TENT [16], EATA [8], MEMO [19], SAR [9], NOTE [3], RoTTA [18], MECTA [4], RATP [6], RMT [2], MALL [12], ONDA [11], and CoTTA [17]. We consider four adverse weather conditions such as fog, rain, night and snow. We consider 10 iterations on ACDC-val [13] dataset. Our DoSe outperforms existing TTA methods and improves the performance of RefineNet [5] by 5.5% better mIoU in the task of continual test-time adaptation.

2. Ablation study

2.1. Computational cost on C-Driving dataset

We analyze the computational cost of the proposed DoSe framework, we consider the RefineNet [5] model and adapt it to 100-overcast images from C-Driving [7] and perform continual test-time adaptation for 10 iterations. Results are reported in Tab. 3. Proposed DoSe framework adapts to continually changing target environment faster than existing SOTA method CoTTA [17].

Method	Inference time
DeepLabv3+ ResNet101 [1]	110
CoTTA [17]	1080
DoSe (ours)	910

Table 2. Inference time comparison of DoSe framework on RefineNet [5] on C-Driving dataset, Inference time values are in seconds.

Method	Inference time
DeepLabv3+ mobilenet V2 [14]	40
CoTTA [17]	350
DoSe (ours)	220

Table 3. Inference time comparison of DoSe framework on DeepLabv3+ Mobilenet V2 [14] on SHIFT dataset, Inference time values are in seconds.

2.2. Computational cost on SHIFT dataset

We analyze the computational cost of the proposed DoSe framework, we consider the DeepLabv3+ mobilenet V2 [14] model and adapt it to 100-Fog images from SHIFT dataset [15] and perform continual test-time adaptation for 10 iterations. Results are reported in Tab. 3. Proposed DoSe framework adapts to continually changing target environment faster than existing SOTA method CoTTA [17].

2.3. Computational cost - Loss function

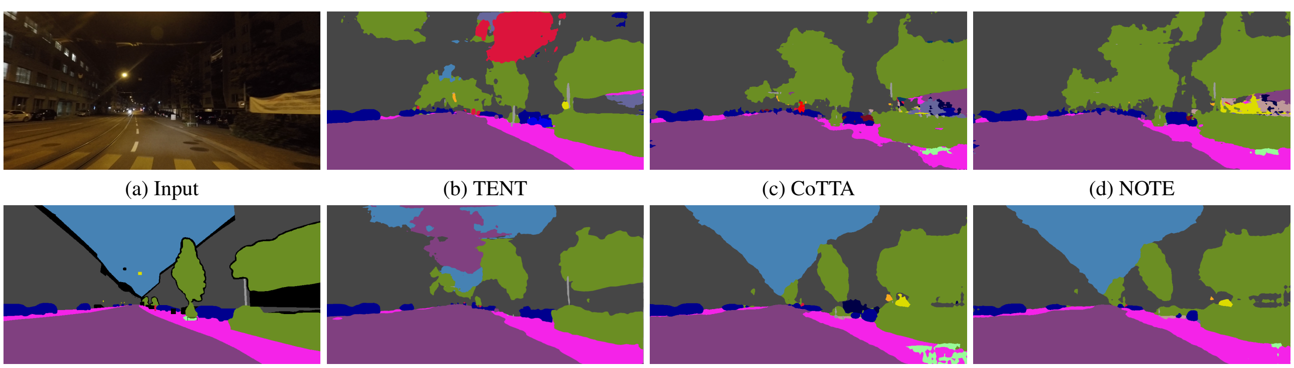
We analyze the computational cost of the proposed DoSe framework, with different individual components of the total loss function. We consider SOTA DG method IBN-Net [10] as our base pre-trained model. We took 100 images from ACDC-snow [13] and apply the total loss function with and without loss components. Results are reported in Tab. 1.

Time	t →																
Round	1				4				7				10				All
Model+Method	Fog	Night	rain	snow	Fog	Night	rain	snow	Fog	Night	rain	snow	Fog	Night	rain	snow	Mean
RefineNet [5]	55.9	42.6	44.2	21.5	55.9	42.6	44.2	21.5	55.9	42.6	44.2	21.5	55.9	42.6	44.2	21.5	41.1
• BN Stats Adapt	54.2	41.1	43.9	20.4	53.7	40.7	43.7	20.2	53.6	40.1	43.6	20.1	53.4	39.8	43.5	19.9	39.5
• TENT-continual [16]	57.5	43.1	45.8	22.8	53.8	42.7	42.9	21.3	51.4	41.2	42.6	21.1	50.7	40.4	41.4	20.7	40.1
• EATA-continual [8]	57.8	44.9	46.4	21.7	53.1	44.8	43.7	22.5	52.7	42.3	43.8	22.4	51.1	41.8	42.4	22.8	40.8
• MEMO [19]	59.7	45.2	45.8	22.7	54.8	45.7	44.8	23.7	53.9	44.8	45.6	23.9	52.7	42.7	43.4	34.8	42.5
• SAR [9]	54.6	43.8	43.7	26.8	55.9	45.6	46.3	24.9	55.8	46.5	46.1	24.5	53.9	44.7	45.6	35.5	42.3
• NOTE [3]	52.1	39.7	40.1	27.8	56.7	45.3	46.8	22.1	55.1	44.5	44.4	24.5	53.7	42.8	44.3	34.1	42.1
• RoTTA [18]	50.8	38.9	40.3	25.9	55.1	45.2	45.4	21.9	54.8	44.2	44.5	24.1	50.9	43.1	44.5	34.7	41.5
• MECTA [4]	50.6	37.7	41.1	26.4	54.5	44.8	46.2	20.3	53.7	44.8	44.7	23.7	50.6	42.8	43.6	32.8	41.1
• RATP [6]	51.1	38.3	40.9	25.9	53.9	43.7	45.9	21.1	52.7	43.6	44.4	22.9	51.9	43.2	44.2	33.4	41.3
• RMT [2]	52.4	38.9	40.6	25.1	52.7	44.8	45.7	22.5	52.6	44.5	45.3	23.4	52.9	42.2	43.1	34.5	41.4
• MALL [12]	53.5	39.1	45.1	26.5	51.1	42.8	43.8	24.2	54.3	44.8	47.8	23.9	53.7	44.2	43.1	36.2	42.2
• ONDA [11]	53.1	39.9	45.4	26.1	51.4	42.1	43.1	24.8	55.1	44.2	47.1	23.2	54.2	45.1	44.2	34.8	42.5
• CoTTA [17]	55.2	39.4	45.8	28.9	54.3	43.2	44.2	24.7	56.6	45.5	47.5	24.3	55.3	46.7	46.2	37.1	43.6
• DoSe (ours)	**57.4**	**41.6**	**46.4**	**30.2**	**55.9**	**45.8**	**48.8**	**26.1**	**58.8**	**45.9**	**48.8**	**25.6**	**57.8**	**49.1**	**47.1**	**37.3**	**46.5**

Table 4. Comparison of DoSe with existing TTA methods on pre-trained daytime model RefineNet [5]. The experiment setup is described in the paper.

Time	t →																
Round	1				4				7				10				All
Model+Method	snow	rain	fog	night	snow	rain	fog	night	snow	rain	fog	night	snow	rain	fog	night	Mean
DeepLabv3+ ResNet101 [1]	50.6	52.3	67.6	21.9	50.6	52.3	67.6	21.9	50.6	52.3	67.6	21.9	50.6	52.3	67.6	21.9	48.1
• DoSe (ours)	**51.4**	**53.8**	**68.9**	**22.3**	**53.8**	**54.3**	**69.6**	**23.8**	**55.7**	**55.6**	**71.8**	**25.2**	**57.9**	**56.8**	**73.4**	**27.9**	**51.4**

Table 5. Ablation study of Proposed framework DoSe using DeepLabv3+ ResNet101 [1] on ACDC [13] dataset with change in order of sequence. The experiment setup is described in the paper.



(a) Input (b) TENT (c) CoTTA (d) NOTE

(e) Ground truth (f) MECTA (g) RATP (h) DoSe (Ours)

Figure 1. Comparison of our approach, DoSe, with existing SOTA Test-Time Adaptation methods.

3. Qualitative Results

In Fig. 1, we demonstrate the qualitative visual results of the DoSe and compare the visual results with existing SOTA Test-time Adaptation methods. Results demonstrate that DoSe can identify minority classes, such as traffic lights etc., effectively.

References

[1] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 1, 2

[2] Mario Döbler, Robert A Marsden, and Bin Yang. Robust mean teacher for continual and gradual test-time adaptation.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7704–7714, 2023. 1, 2

[3] Taesik Gong, Jongheon Jeong, Taewon Kim, Yewon Kim, Jinwoo Shin, and Sung-Ju Lee. NOTE: Robust continual test-time adaptation against temporal correlation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 1, 2

[4] Junyuan Hong, Lingjuan Lyu, Jiayu Zhou, and Michael Spranger. Mecta: Memory-economic continual test-time model adaptation. In *The Eleventh International Conference on Learning Representations*, 2023. 1, 2

[5] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1925–1934, 2017. 1, 2

[6] Chenxi Liu, Lixu Wang, Lingjuan Lyu, Chen Sun, Xiao Wang, and Qi Zhu. Deja vu: Continual model generalization for unseen domains. *arXiv preprint arXiv:2301.10418*, 2023. 1, 2

[7] Ziwei Liu, Zhongqi Miao, Xingang Pan, Xiaohang Zhan, Dahua Lin, Stella X. Yu, and Boqing Gong. Open compound domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[8] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *The International Conference on Machine Learning*, 2022. 1, 2

[9] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. In *Internetional Conference on Learning Representations*, 2023. 1, 2

[10] 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

[11] Theodoros Panagiotakopoulos, Pier Luigi Dovesi, Linus Härenstam-Nielsen, and Matteo Poggi. Online domain adaptation for semantic segmentation in ever-changing conditions. In *European Conference on Computer Vision (ECCV)*, 2022. 1, 2

[12] Nikhil Reddy, Abhinav Singhal, Abhishek Kumar, Mahsa Baktashmotlagh, and Chetan Arora. Master of all: Simultaneous generalization of urban-scene segmentation to all adverse weather conditions. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*, pages 51–69. Springer, 2022. 1, 2

[13] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. *arXiv preprint arXiv:2104.13395*, 2021. 1, 2

[14] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018. 1

[15] Tao Sun, Mattia Segu, Janis Postels, Yuxuan Wang, Luc Van Gool, Bernt Schiele, Federico Tombari, and Fisher Yu. Shift: a synthetic driving dataset for continuous multi-task domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21371–21382, 2022. 1

[16] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020. 1, 2

[17] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7201–7211, 2022. 1, 2

[18] Longhui Yuan, Binhui Xie, and Shuang Li. Robust test-time adaptation in dynamic scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15922–15932, 2023. 1, 2

[19] M. Zhang, S. Levine, and C. Finn. MEMO: Test time robustness via adaptation and augmentation. 2021. 1, 2