Master of All: Simultaneous Generalization of Urban-Scene Segmentation to All Adverse Weather Conditions- Supplementary

Nikhil Reddy[1], Abhinav Singhal[2], Abhishek Kumar[2], Mahsa Baktashmotlagh[3], and Chetan Arora[2]

[1] University of Queensland – IIT Delhi Academy of Research (UQIDAR)
[2] Indian Institute of Technology Delhi
[3] The University of Queensland
nikhil.jangamreddy@uqidar.iitd.ac.in

1 Gradient Backpropagation

The Weighted log softmax multi-class normalized cut loss $\mathcal{L}$ for an image $\mathbf{I}$, is defined as:

$$\mathcal{L}(\mathbf{S}) = \sum_{p=0}^{c-1} w_p \frac{(\mathbf{S}_p)^\top \mathbf{A}\,(\mathbf{1} - \mathbf{S}_p)}{\mathbf{d}^\top \mathbf{S}_p} \tag{1}$$

For a standard image, matrices $\mathbf{S}_p$, $\mathbf{A}$ are of large dimensions, which causes a large computational burden for computing the loss. To reduce the computational bottleneck, motivated by [12], we consider equivalent Weighted log softmax multi-class normalized cut loss as defined below:

$$\mathcal{L}(\mathbf{S}) = -\sum_{p=0}^{c-1} w_p \frac{(\mathbf{S}_p)^\top \mathbf{A}\,(\mathbf{S}_p)}{\mathbf{d}^\top \mathbf{S}_p} \tag{2}$$

Gradient of $\mathcal{L}(\mathbf{S})$ w.r.t $\mathbf{S}_p$ can be written as:

$$\frac{\partial \mathcal{L}(\mathbf{S})}{\partial \mathbf{S}_p} = w_p \frac{(\mathbf{S}_p)^\top \mathbf{A}\mathbf{S}_p \mathbf{d}}{(\mathbf{d}^\top \mathbf{S}_p)^2} - w_p \frac{2\mathbf{A}\mathbf{S}_p}{\mathbf{d}^\top \mathbf{S}_p} \tag{3}$$

Eq. (3) is used to backpropagate gradients into network layers by using standard gradient chain rule.

2 Comparison with TENT[15]

In comparison with TENT[15], we demonstrated performance on multiple datasets, including ACDC[11], C-driving[6], etc. Due to the benchmark limitations on the number of submissions, it is not feasible to submit 16 submissions per model to the ACDC-test benchmark website. So we consider the best performing model to be DeepLabv3+ resnet101[1] and submit it to the ACDC benchmark website. Results are shown in Tab. 1. We report that `MALL-domain` improves `mIoU` performance on ACDC-fog by 7%. `MALL-domain` improves average `mIoU` performance on the ACDC benchmark, consisting of fog, rain, night, and snow adverse weather conditions by 17%. ACDC benchmark results are available here.

Method	**ACDC-fog**	**ACDC-Rain**	**ACDC-Snow**	**ACDC-Night**	**Average**
DeepLabv3+ ResNet101[1]	49.0	53.4	40.7	26.2	42.3
with `MALL-domain`	52.4	56.9	51.4	36.8	**49.3**

Table 1: Results of `MALL-domain` on ACDC benchmark website

3 Inference time

For a batch size of 12, with each image resolution of 1024×512, we compare the inference time of TENT[15] with the `MALL` framework consisting of `MALL-sample` and `MALL-domain` methods. We consider the pre-trained daytime models DeepLabv3+ mobilenet and DeepLabv3+ resnet101. Results are reported in table 2. TENT[15] adapts a pre-trained model to a single image similar to the `MALL-sample` method. TENT[15] is comparatively faster than `MALL-sample` as TENT only updates the affine parameters of the batch normalization layer instead of updating the entire network. However, `MALL-sample` outperforms TENT in terms of mIOU% performance.

Method	mobilenet	resnet101
TENT[15]	748	937
`MALL-sample`	1312	1564
`MALL-domain`	1180	1219

Table 2: comparison of inference time per iteration on pre-trained daytime models: mobilenet: DeepLabv3+ mobilenet, resnet101: DeepLabv3+ resnet101, inference time is reported in milliseconds (ms).

Method	`MALL-sample`	`MALL-domain`
IBNNet[7]	1371	1228
SW[8]	1147	990
RobustNet-Resnet50[2]	1013	872
RobustNet-Resnet101[2]	1802	1642

Table 3: comparison of inference time per iteration on pre-trained domain generalization methods using `MALL` framework: RobustNet-resnet50 (ISW), RobustNet-resnet101 (ISW), inference time is reported in milliseconds (ms).

4 Results of MALL-sample.

To demonstrate the efficiency of the `MALL-sample`, we report the average mIOU performance across 13 datasets, described in subsection 4.1: **Table 3:** BiseNetV2: 28.6, ISANet: 41.2, STDC: 39.7, SegFormer: 38.5, GCNet: 41.6, LRASPP: 32.2, Mobilenet V2: 31.8; **Table 4:** IBNNet: 41.1, SW: 36.4, RobustNet-R50: 40.6, RobustNet-R101: 41.3; **Table 5:** Zeroshot-DN (ND, DZ): 42.3, 36.1; MGCDA

(ND, DZ): 49.7, 43.1; **Table 6:** MGCDA: 26.7; **Table 7:** DANNet: 41.8; **Table 8:** `MALL-sample` (ND, DZ): 36.8, 19.8. Similarly, we consider pre-trained state-

Method	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	**mIOU**
RefineNet[5]	68.8	23.2	46.8	20.8	12.6	29.8	30.4	26.9	43.1	14.3	0.3	36.9	49.7	63.6	6.8	0.2	24.0	33.6	9.3	28.5
AdaptSegNet[13]	86.1	44.2	55.1	22.2	4.8	21.1	5.6	16.7	37.2	8.4	1.2	35.9	26.7	68.2	45.1	0.0	50.1	33.9	15.6	30.4
ADVENT[14]	85.8	37.9	55.5	27.7	14.5	23.1	14.0	21.1	32.1	8.7	2.0	39.9	16.6	64.0	13.8	0.0	58.8	28.5	20.7	29.7
BDL[4]	85.3	41.1	61.9	32.7	17.4	20.6	11.4	21.3	29.4	8.9	1.1	37.4	22.1	63.2	28.2	0.0	47.7	**39.4**	15.7	30.8
DMAda[3]	75.5	29.1	48.6	21.3	14.3	34.3	36.8	29.9	49.4	13.8	0.4	43.3	50.2	69.4	18.4	0.0	27.6	34.9	11.9	32.1
GCMA[9]	81.7	46.9	58.8	22.0	20.0	41.2	**40.5**	**41.6**	64.8	31.0	32.1	**53.5**	47.5	**75.5**	39.2	0.0	49.6	30.7	21.0	42.0
MGCDA[10]	80.3	49.3	66.2	7.8	11.0	41.4	38.9	39.0	64.1	18.0	55.8	52.1	**53.5**	74.7	**66.0**	0.0	37.5	29.1	22.7	42.5
DANNet(PSPNet)[16]	**90.4**	**60.1**	**71.0**	33.6	**22.9**	30.6	34.3	33.7	70.5	**31.8**	**80.2**	45.7	41.6	67.4	16.8	0.0	**73.0**	31.6	**22.9**	45.2
`MALL-domain`	90.3	59.8	70.8	**34.2**	22.7	**30.9**	37.2	34.1	**70.6**	31.6	80.1	46.8	43.2	68.1	16.4	**0.3**	72.6	31.9	22.8	**45.5**

Table 4: Per class mIOU scores of the state-of-the-art night image segmentation methods on the Dark Zurich test dataset, `MALL-domain` built on DANNet further improves the results.

of-the-art domain generalization methods IBNNet[7], Switchable Whitening[8], RobustNet-resnet50[2], and RobustNet-resnet101[2], we consider the batch size of 12, with each image resolution of 1024×512. We report the inference time with the `MALL-sample` and `MALL-domain` methods. Results are reported in table 3.

5 Ablation study

To demonstrate the impact of the number of iterations on the mIOU performance, we consider the DeepLabv3+ mobilenet pre-trained model using the `MALL-domain` method. Results are reported in figure 1. mIOU drops after a definite number of iterations to consider this; We use early stopping criteria as defined in `MALL-sample` and `MALL-domain` method. Early stopping criteria are based on Softmax multi-class normalized cut loss between two consecutive iterations.

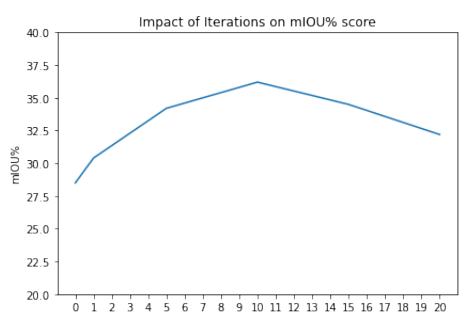
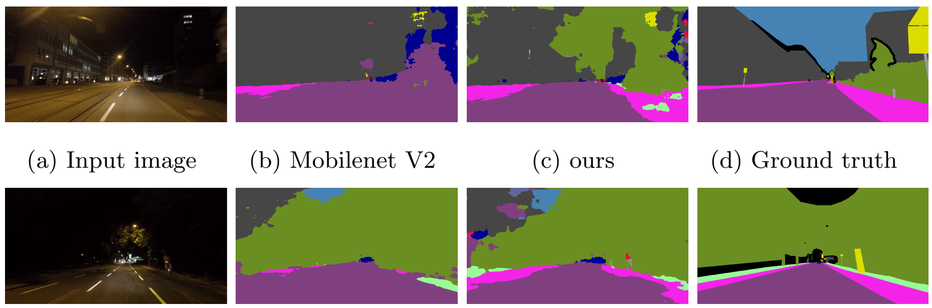


Fig. 1: Impact of number of Iterations on mIOU% score using pre-trained model: Deeplabv3+ mobilenet, dataset: Nighttime Driving test dataset.

6 Qualitative results

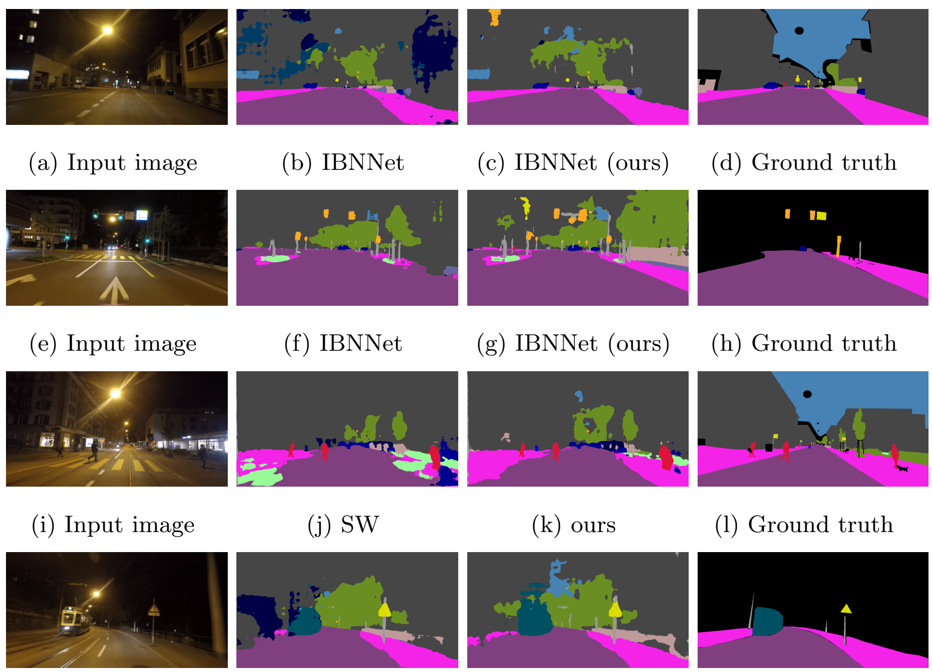
Qualitative visual results for pre-trained models for daylight, DeepLabv3+ mobilenet[1] and DeepLabv3+ resnet101[1] are shown in figure 2. We consider the pre-trained models for state-of-the-art domain generalization methods IBNNet[7], Switchable whitening[8], and RobustNet[2]. Qualitative visual results state-of-the-art domain generalization methods are presented in figure 3 and 4 respectively. Visual results demonstrate the significant improvement in segmentation label predictions by directly using `MALL` framework during inference.



(a) Input image (b) Mobilenet V2 (c) ours (d) Ground truth

(e) Input image (f) DLv3+ resnet101 (g) ours (h) Ground truth

Fig. 2: Qualitative visual comparison of our proposed `MALL` framework on pre-trained daytime models: Mobilenet V2, Deeplabv3+ resnet101 on two images from night image datasets, ours: pre-trained model+ `MALL-domain` method.



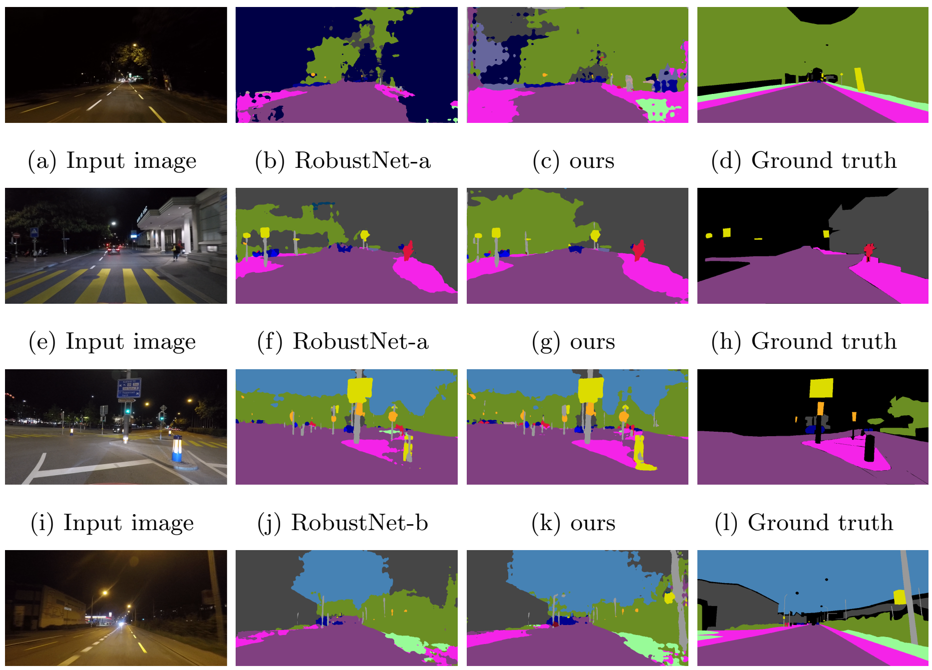
(a) Input image (b) IBNNet (c) IBNNet (ours) (d) Ground truth

(e) Input image (f) IBNNet (g) IBNNet (ours) (h) Ground truth

(i) Input image (j) SW (k) ours (l) Ground truth

(m) Input image (n) SW (o) ours (p) Ground truth

Fig. 3: Qualitative visual comparison of our proposed `MALL` framework on pre-trained Domain generalization models: IBNNet, Switchable Whitening (SW) on two images per model from night image datasets, ours: pre-trained model+ `MALL-domain` method.



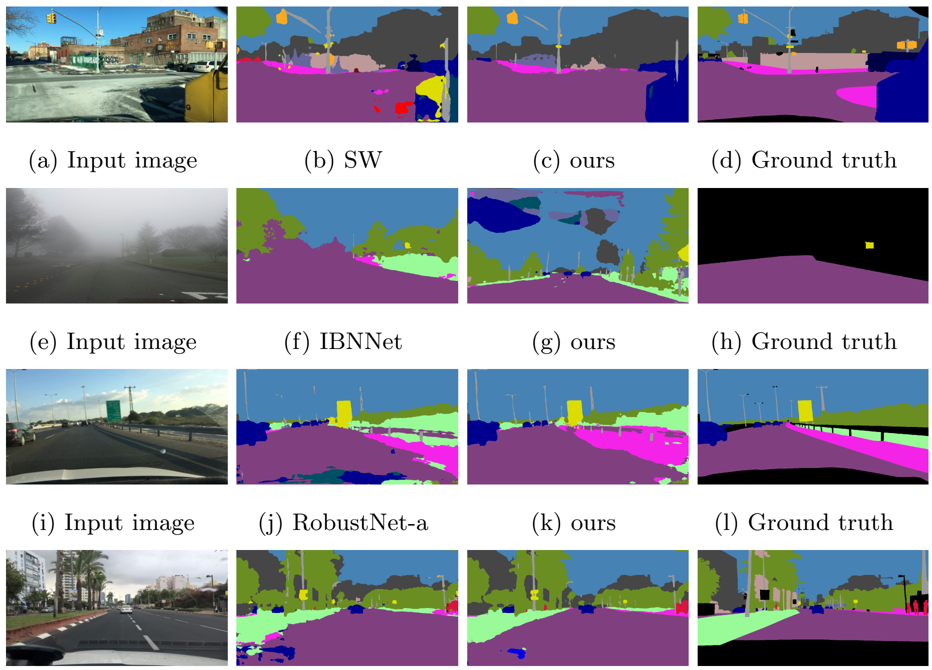
(a) Input image (b) RobustNet-a (c) ours (d) Ground truth

(e) Input image (f) RobustNet-a (g) ours (h) Ground truth

(i) Input image (j) RobustNet-b (k) ours (l) Ground truth

(m) Input image (n) RobustNet-b (o) ours (p) Ground truth

Fig. 4: Qualitative visual comparison of our proposed `MALL` framework on pre-trained Domain generalization models: RobustNet-a: RobustNet-Resnet50 (ISW), RobustNet-b: RobustNet-Resnet101 (ISW) on two images per model from night image datasets, ours: pre-trained model+ `MALL-domain` method.



(a) Input image (b) SW (c) ours (d) Ground truth

(e) Input image (f) IBNNet (g) ours (h) Ground truth

(i) Input image (j) RobustNet-a (k) ours (l) Ground truth

(m) Input image (n) RobustNet-b (o) ours (p) Ground truth

Fig. 5: Qualitative visual comparison of our proposed `MALL` framework on pre-trained Domain generalization models: IBNNet, SW, RobustNet-a: RobustNet-Resnet50 (ISW), RobustNet-b: RobustNet-Resnet101 (ISW) on two images per model from night image datasets, ours: pre-trained model+ `MALL-domain` method.

References

1. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV) (September 2018) 1, 2, 4
2. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11580–11590 (2021) 2, 3, 4
3. Dai, D., Van Gool, L.: Dark model adaptation: Semantic image segmentation from daytime to nighttime. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp. 3819–3824. IEEE (2018) 3
4. Li, Y., Yuan, L., Vasconcelos, N.: Bidirectional learning for domain adaptation of semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6936–6945 (2019) 3
5. Lin, G., Milan, A., Shen, C., Reid, I.: Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1925–1934 (2017) 3
6. Liu, Z., Miao, Z., Pan, X., Zhan, X., Lin, D., Yu, S.X., Gong, B.: Open compound domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 1
7. Pan, X., Luo, P., Shi, J., Tang, X.: Two at once: Enhancing learning and generalization capacities via ibn-net. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 464–479 (2018) 2, 3, 4
8. Pan, X., Zhan, X., Shi, J., Tang, X., Luo, P.: Switchable whitening for deep representation learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1863–1871 (2019) 2, 3, 4
9. Sakaridis, C., Dai, D., Gool, L.V.: Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7374–7383 (2019) 3
10. Sakaridis, C., Dai, D., Van Gool, L.: Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020) 3
11. 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) 1
12. Tang, M., Djelouah, A., Perazzi, F., Boykov, Y., Schroers, C.: Normalized cut loss for weakly-supervised cnn segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1818–1827 (2018) 1
13. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7472–7481 (2018) 3
14. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2517–2526 (2019) 3
15. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. arXiv preprint arXiv:2006.10726 (2020) 1, 2

16. Wu, X., Wu, Z., Guo, H., Ju, L., Wang, S.: Dannet: A one-stage domain adaptation network for unsupervised nighttime semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15769–15778 (2021) 3