Master of All: Simultaneous Generalization of Urban-Scene Segmentation to <u>All</u> Adverse Weather Conditions

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

Abstract. Computer vision systems for autonomous navigation must generalize well in adverse weather and illumination conditions expected in the real world. However, semantic segmentation of images captured in such conditions remains a challenging task for current state-of-the-art (SOTA) methods trained on broad daylight images, due to the associated distribution shift. On the other hand, domain adaptation techniques developed for the purpose rely on the availability of the source data, (un)labeled target data and/or its auxiliary information (e.g., GPS). Even then, they typically adapt to a single(specific) target domain(s). To remedy this, we propose a novel, fully test time, adaptation technique, named *Master of ALL* (MALL), for simultaneous generalization to multiple target domains. MALL learns to generalize on unseen adverse weather images from multiple target domains directly at the inference time. More specifically, given a pre-trained model and its parameters, MALL enforces edge consistency prior at the inference stage and updates the model based on (a) a single test sample at a time (MALL-sample), or (b) continuously for the whole test domain (MALL-domain). Not only the target data, MALL also does not need access to the source data and thus, can be used with any pre-trained model. Using a simple model pre-trained on daylight images, MALL outperforms specially designed adverse weather semantic segmentation methods, both in domain generalization and test-time adaptation settings. Our experiments on foggy, snow, night, cloudy, overcast, and rainy conditions demonstrate the target domain-agnostic effectiveness of our approach. We further show that MALL can improve the performance of a model on an adverse weather condition, even when the model is already pre-trained for the specific condition.

Keywords: Multi-target Domain Generalization, All Weather Urban-Scene Segmentation, Test-time Adaptation, Normalized Cuts.

1 Introduction

Out-of-domain generalization plays a pivotal role in the success of deep neural networks (DNNs) for safety-critical applications such as autonomous navigation.

Problem Setting	Source Data	Target Label (y^t)	Train Loss	Test Loss	Open Targets
Fine-tuning	–	✓	$\mathcal{L}(x^t, y^t)$	✗	✗
Unsupervised domain adaptation[57]	x^s, y^s	✗	$\mathcal{L}(x^s, y^s) + \mathcal{L}(x^t, x^s)$	✗	✗
Source-free domain adaptation[19]	✗	✗	$\mathcal{L}(x^t)$	✗	✗
Domain Generalization[6]	x^s, y^s	✗	$\mathcal{L}(x^s, y^s)$	✗	✓
Test-time training[45]	x^s, y^s	✗	$\mathcal{L}(x^s, y^s) + \mathcal{L}(x^s)$	$\mathcal{L}(x^t)$	✓
Fully test-time adaptation[52]	–	✗	✗	$\mathcal{L}(x^t)$	✓

Table 1: Comparing different problem settings, the proposed technique fits in as a *fully test-time adaptation* paradigm, which aims to adapt a pre-trained model to an unseen target domain.

Although `DNN` based architectures have achieved tremendous success in many computer vision tasks [25,14,13], they don't perform well when the test data comes from a domain with different sample distribution, referred to as the domain shift [30]. Domain shift can be caused by different factors such as input corruption, adverse weather (rain, snow etc.), illumination changes (e.g. night time), adversarial attacks, sensor malfunction etc. Ensuring robust performance on unseen target domains is critical for the real-world applicability of `DNN`s.

Despite the success of `DNN`s for the task of semantic segmentation in sunny, daytime conditions [4,3], these models suffer from severe performance degradation when applied on night, or adverse weather images due to low illumination, shadows, motion blur, glare/or and overexposure. There is scarcity of labeled ground truth data in night-time and adverse weather conditions for training a `DNN` model. Further, semantic segmentation models are required to exhibit robust performance and reliable operation in a wide range of changing environments, and expecting availability of the data corresponding to all such conditions, and their combinations (e.g. rainy night-time) is unrealistic, and impractical.

To mitigate the aforementioned issues, Domain Generalization (`DG`) approaches, such as RobustNet [6], aim to improve robustness on the unseen target domains. However the performance of `DG` methods is still quite low on adverse weather conditions (see Tab. 4). On the other hand, unsupervised domain adaptation (`UDA`) techniques for semantic segmentation [62,64,59,58,63,50,49,47,41,26,21,18,15,5] typically focus on synthetic-to-real domain adaptation, and are not relevant to our focus. `UDA` techniques specially designed for adapting from normal to adverse weather, require access to target dataset, and are limited to single target domain only, such as nighttime[33,38,53] or fog[35,8]. This limits their real life applicability which requires simultaneous adaptation to all weather conditions.

While effective, it might not be practical in a real-world scenario to collect (even unlabeled) target data corresponding to all adverse visual conditions or their exponential number of combinations. Moreover, there might be a shift in the target data itself due to the dynamic nature of real-world processes. This necessitates designing a model which can adapt on-the-fly when the data is being received during inference. In the light of the above discussion, in this work, we propose `MALL` for semantic segmentation of adverse weather images. `MALL` adapts

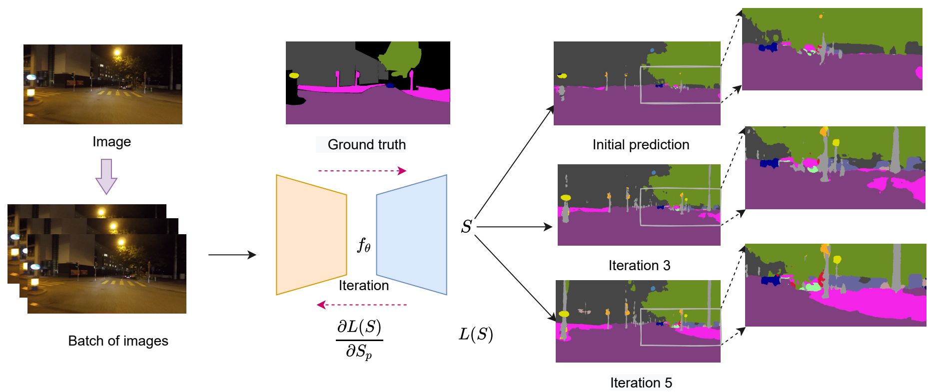


Fig. 1: Our technique improves the prediction of a pre-trained segmentation model by penalizing the loss if the edges in the predicted label image, S, do not align with the visual edges in the input image.

a pre-trained model to unseen weather deteriorated images by enforcing an edge consistency prior. The prior is enforced by aligning the edges in the input `RGB` image with the edges in the output predicted label image using a loss inspired from the normalized cuts [43]. Moreover, motivated by the recent fully-test time adaptation paradigm [52], we enforce the prior only during the inference (and not at the train time). This has a significant advantage in terms of removing our reliance on the availability of unlabeled target data during training, and capability to use any pre-trained model for adaptation during inference. The key contributions of our work include:

- We propose to enforce edge consistency prior during inference for unsupervised test time adaptation of semantic segmentation techniques for adverse weather images. Taking inspiration from [43], we propose a *Weighted Log Multi-class Normalized Cut* loss for the unsupervised test time training.
- We propose a new framework named `MALL` in two different settings: `MALL-sample` adapts a pre-trained model to a single image at a time, whereas `MALL-domain` adapts a model to the target domain at the test time.
- We perform rigorous experiments to demonstrate that `MALL` outperforms all `SOTA` techniques in multiple settings: domain generalization, and unsupervised domain adaptation.
- In a one of its kind experiment, we demonstrate the simultaneous, and domain-agnostic effectiveness of the `MALL` on six different adverse weather conditions: night, overcast, cloudy, snow, fog, and rain.

2 Related Work

Test-Time Adaptation: In a test-time training framework [45], in the first step, one trains a `DNN` with the main task like classification and an auxiliary task like rotation prediction of an image. During inference, the auxiliary task is used to adapt network parameters to handle distribution shifts. TENT [52] has proposed a fully test-time adaptation framework assuming no access to source data and without altering training. It adapts the affine parameters of a batch normalization layer to minimize entropy-based loss for a particular test sample. We also propose a fully test-time adaptation framework specifically for unseen adverse weather conditions such as nighttime, rain, etc. However, unlike TENT, which uses entropy-based loss, we enforce edge prior to achieve our objective.

Domain Generalization (DG): These approaches aim to improve generalization on the unseen target domains. While popular for image classification [60,61,31], very few DG approaches have been proposed for semantic segmentation. IBNNet [27] combines the advantage of instance normalization (to prevent over-fitting on the training data) and batch normalization (learning discriminative intermediate feature representations) to improve generalization in semantic segmentation. Switchable whitening [28] de-correlates intermediate feature maps to remove domain-specific style information. RobustNet [6] extends whitening approach to selectively remove higher values in channel-wise covariance matrix. MALL improves generalization and complement above DG methods by directly adapting to unseen adverse weather images during inference.

UDA Methods for Night-time Image Segmentation: GCMA [33] proposes the use of intermediate twilight domain to adapt to nighttime images gradually from day to twilight, and twilight to the night-time domain. MGCDA [38] extends it using self-training with curriculum learning to adapt gradually to night images. Similarly [44,9,32], rely on image transfer models like cycleGAN as a pre-processing step. However, their performance depend significantly on the quality of images generated in the pre-processing step. To mitigate this, DANNet [53] proposes a one-stage adaptation network without relying on an intermediate domain or image transfer model. DANNet uses reweighing strategy to handle misalignment between day-night image pairs, and adversarial training accompanied by pseudo-label supervision to enhance segmentation predictions. DANNet also uses additional supervision from GPS to enhance semantic label predictions. DANNet generates noisy predictions for moving objects like cars, and small objects like traffic lights. Our proposed method, MALL, does not need any extra information from the target domain dataset.

Adaptation without Target Data: In a real-world scenario, it is often impractical to collect large labeled or unlabeled target data covering all possible illumination scenarios. Recent work Zeroshot-DayNight [20] assumes no access to unlabeled target night images and uses the color invariance idea[12], to propose a new layer in a DNN, referred to as color invariant convolution (CIConv). The method claims to reduce the day-night distribution shift for intermediate feature maps, but requires retraining the entire network from scratch using CIConv, thus restricting its applicability. On the other hand, the proposed MALL technique does not do any retraining of a model, but adapts the model on-the-fly to unseen weather images during inference. MALL improves current SOTA methods in both the scenarios, when having access to unlabeled night images during training, as well as without access to them.

Normalized Cuts: Normalized cuts[43] is a popular graph partitioning algorithm proposed for binary image segmentation. Tang *et al.* [46] has proposed to use the normalized cut loss as a regularizer to improve performance in weakly-supervised segmentation tasks. It uses cross-entropy loss for labeled pixels and continuous normalized cut-based loss for weakly-labeled (scribbles) pixels to improve segmentation predictions. We extend the normalized cut loss as edge consistency prior for unsupervised domain generalization setting to penalize the segmentation predictions if the label edges do not align with the visual edges in the input.

3 Methodology

In this section, we go through the formulation of the proposed `MALL` technique. Specifically, we propose to optimize the model parameters directly during inference using *Weighted Log Multi-class Normalized Cut* (`WL-MNC`) loss. The loss is designed to penalize incorrect semantic label predictions when the consistency of visual edges present at the boundaries of classes in an image is not preserved. Note that, our assumption in the proposed `MALL` framework is that we do not need access to the source data. The `MALL` technique can work in two different settings, namely, `MALL-sample` and `MALL-domain`. The former updates the pre-trained model based on a single nighttime image at a time, whereas the later adapts the model continuously to all samples of the target domain as they are presented in a streaming mode.

3.1 Softmax Multi-Class Normalized Cut Loss

In this section, we extend normalized cut [43] as a loss function for an unsupervised, multi-class setting that can handle class imbalance in the target dataset. Normalized Cut [43] is a popular graph partitioning algorithm for binary image segmentation task. For the binary image segmentation task, we define a graph node corresponding to every pixel in the image. Let $\mathbf{I}$ be an image, and $\mathbf{A}$ be the affinity matrix, with $\mathbf{A} = [\mathbf{A}_{ij}]$ being the similarity between pixel i and pixel j in the image. Let $\mathbf{d}$ be the degree vector defined as $\mathbf{d} = \mathbf{A}\mathbf{1}$, where $\mathbf{1}$ denotes a vector of all ones. For a binary image segmentation task, where pixels are labeled either as a foreground or background pixel, Normalized Cut loss for partitioning pixels of an image into foreground ($\mathcal{F}$) and background pixels ($\mathcal{B}$) is defined as:

$$\text{cut}(\mathcal{F}, \mathcal{B}) = \sum_{x \in \mathcal{F}, y \in \mathcal{B}} \mathbf{A}(x, y) \tag{1}$$

$$\text{Ncut}(\mathcal{F}, \mathcal{B}) = \frac{\text{cut}(\mathcal{F}, \mathcal{B})}{\text{assoc}(\mathcal{F}, \mathcal{V})} + \frac{\text{cut}(\mathcal{F}, \mathcal{B})}{\text{assoc}(\mathcal{B}, \mathcal{V})} \tag{2}$$

where $\mathcal{V} = \mathcal{F} \cup \mathcal{B}$, $\text{assoc}(\mathcal{F}, \mathcal{V}) = \sum_{p \in \mathcal{F}, q \in \mathcal{V}} \mathbf{A}_{pq}$. If we denote binary image segmentation prediction by a matrix $\mathbf{S}$, then Eq. (2) can be rewritten as:

$$\text{Ncut} = \frac{\mathbf{S}^\top \mathbf{A}(\mathbf{1} - \mathbf{S})}{\mathbf{d}^\top \mathbf{S}} \tag{3}$$

Further, Normalized Cut loss can be extended to a multi-class setting as follows. Let c be the total number of classes, and p denote a particular class. Let $\mathbf{S}_p$ be the segmentation label prediction matrix corresponding to class p, such that $\mathbf{S}_p[i][j] = 1$, if the predicted label of pixel (i, j) is p, and $\mathbf{S}_p[i][j] = 0$, otherwise. Then, Multi-class Normalized Cut loss is a non-differentiable loss as shown below:

$$\text{MultiClassNormalizedCut} = \sum_{p=0}^{c-1} \frac{(\mathbf{S}_p)^\top \mathbf{A}\,(\mathbf{1} - \mathbf{S}_p)}{\mathbf{d}^\top \mathbf{S}_p} \tag{4}$$

To make a network end-to-end trainable, one can relax Multi-class Normalized Cut loss to Softmax Multi-class Normalized Cut loss [46], by relaxing the value of $S_p[i][j]$ to predicted probability of pixel (i, j) taking label p.

3.2 Class Imbalance Re-weighting

We observe that most benchmark datasets (e.g., Cityscapes[7]) have highly skewed ratio corresponding to the pixels of minority to majority classes. This leads to reduced accuracy for the pixels belonging to the minority class. To handle the class imbalance, we introduce weighted log class probability weighting into the Multi-class Normalized Cut loss function. Let r_p denote the normalized frequency of class p in the source domain, i.e., number of pixels of class p divided by total number of pixels[4]. We define the weight, w_p, as:

$$w_p = \frac{\log(r_p)}{\sum_{i=0}^{c-1} \log(r_i)}. \tag{5}$$

We use w_p to define a new loss function, Weighted Log Multi-class Normalized Cut (`WL-MNC`) loss, to handle the class imbalance as follows:

$$\mathcal{L}_{\texttt{WL-MNC}} = \sum_{p=0}^{c-1} w_p \frac{(\mathbf{S}_p)^\top \mathbf{A}\,(\mathbf{1} - \mathbf{S}_p)}{\mathbf{d}^\top \mathbf{S}_p}. \tag{6}$$

3.3 Sample Importance Weighting

Given an image at the test time, we first create a mini-batch by multiple distortions (called samples) of the input image. Then, we assign an importance weight

[4] note that we do not assume the availability of source data as our method does not need any training for the chosen backbone. However, we do assume the availability of the statistics for the source dataset on which a given model is pre-trained on.

to a sample k in the mini-batch based on the entropy, e^k, of the sample. We compute the entropy of a sample as:

$$e^k = -\frac{1}{w \cdot h} \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} \sum_{p=0}^{c-1} \mathbf{S}^k[i][j][p] \ \log\left(\mathbf{S}^k[i][j][p]\right), \tag{7}$$

where $\mathbf{S}^k$ denotes the segmentation output for sample k, and w, and h denotes width and height of the image respectively. If the entropy of a sample is high, we give more weight to the `WL-MNC` loss of the corresponding sample thus increasing the importance of edge consistency for the sample. Importance weighting is based on the softmax normalization of unsupervised Shannon entropy [42] loss. Importance weight w^k_{imp} of k^{th} sample in the mini-batch of size b is defined as:

$$w^k_{\text{imp}} = \frac{\exp(e^k)}{\sum_{i=0}^{b-1} \exp(e^i)}. \tag{8}$$

Overall loss for a mini-batch is computed as:

$$\mathcal{L}_{\text{total}} = \sum_{i=0}^{b-1} w^k_{\text{imp}} \mathcal{L}_{\texttt{WL-MNC}}(\mathbf{S}^k) \tag{9}$$

For a pixel (i, j) in an image, we define affinity matrix $\mathbf{A}$ using a Gaussian kernel in 5D (`RGBXY`) space. where `RGB` corresponds to R, G, and B color values of the pixel, and `XY` corresponds to the pixel location (i, j) in the image. For two pixels with similar color and location, similarity or affinity is high. In order to facilitate efficient calculation of loss and its gradient, we use a fast bilateral filtering based technique [1].

3.4 MALL-Sample

In the `MALL-sample` method, a pre-trained `DNN` model f_θ, with parameters, θ is adapted to a single image. Given an image x_t, we create a mini-batch of size b, and batch size b, we generate a batch of images by applying $(b-1)$ data augmentation transformations to x_t. During forward pass, the generated batch of images is passed into the network, and loss is back-propagated. We conduct multiple forward and back-propagation iterations over the same batch (and update the weights) until a pre-defined termination criterion (explained below) is met. Once termination criteria are met, we save segmentation label predictions of the image, and weights of the network revert back to the initial model parameters θ. The advantage of reverting back weights of the network for each image is that we do not suffer performance degradation on the source domain (e.g. daylight images). $\mathbf{S}$ denotes segmentation predictions from the network.

$$\min_{\theta} \mathcal{L}\left(x_t, \theta, \mathbf{S}\right). \tag{10}$$

For the `MALL-sample` setting, as a termination criterion, we consider the number of iterations per image as 5, with an early stopping criteria when the difference of loss between two consecutive iterations is less than a particular threshold.

3.5 MALL-domain

In `MALL-domain` method, a pre-trained model f_θ is adapted to a different target domain directly during inference. Similar to `MALL-sample`, for an unseen target domain image x_t, we pass a batch of images to the network in the forward pass, and `WL-MNC` loss is similarly back-propagated. We also conduct multiple iterations in the same fashion. However, in this case, we do not revert back the weights of network parameters at the end of iterations. Further, no augmentations of the input image are performed to generate a mini-batch in the `MALL-domain` method. The compilation of a batch of images X_t in `MALL-domain` is based on maintaining a buffer of size b; once the buffer is full, we forward pass the batch of images into the network and empty the buffer. We use X_t to denote the batch of images at time t, The benefit of the `MALL-domain` method over unsupervised domain adaptation methods is the ability to adapt on-the-fly to unseen adverse weather images (e.g. rainy night images) directly during inference.

$$\min_{\theta} \mathcal{L}\left(X_t, \theta, \mathbf{S}\right). \tag{11}$$

For the `MALL-domain` setting, we consider the early stopping criteria such that the difference of loss of the network between two consecutive iterations is less than the threshold. The threshold is set to 10^{-10} based on empirical experiments.

4 Experiments

4.1 Datasets and Evaluation criteria

Nighttime Driving (ND)[10]: The Nighttime Driving-test dataset consists of 50 nighttime images and their corresponding pixel-level coarse ground truth annotations. Each image is of resolution 1920 $\times$ 1080 labeled with 19 classes.

Dark Zurich (DZ)[34]: It is a collection of 8779 images with a resolution of 1920 $\times$ 1080 captured during the daytime, twilight, and nighttime. For each image, corresponding `GPS` coordinates of the camera are also provided. Dark Zurich-val corresponds to 50 nighttime images used for validation. Dark Zurich-test corresponds to 151 nighttime images used for testing. Dark Zurich-test does not provide ground truth pixel label information to users, and an online evaluation server has been provided to evaluate the performance.

ACDC[39]: The dataset is a collection of 4006 adverse visual condition images with a resolution of 1920$\times$1080 pixels. It contains images from adverse visual conditions of fog, night, snow, and rain. Each visual domain (fog, night, snow, and rain) contains 400 train images along with ground truth semantic labels, 100 validation images (106 validation images for night domain) along with ground truth semantic labels, and 500 unlabeled test images. ACDC Night-val corresponds to results reported on ACDC night validation images.

Foggy Driving (FDD, FD)[37]: It is a collection of 101 real-world fog images with a resolution of 960$\times$1280 pixels. It contains images captured with fog ranging from moderate to dense fog. Foggy Driving dense is a subset of Foggy Driving dataset consisting of 21 images with dense fog.

Foggy Zurich (FZ)[36]: It contains 3808 real-world fog images collected in Zurich city with a resolution of 1920×1080 pixels. Foggy Zurich-test consists of 40 real fog images with corresponding ground truth semantic labels used for evaluation.

C-Driving[24]: The dataset is collection of four adverse weather conditions compiled from BDD100K dataset[56] consisting of four weather conditions such as cloudy, rain, snow and overcast conditions.

Cityscapes (Day)[7]: It is a collection of 5000 broad daylight images of resolution 2048 × 1024 pixels. For evaluation,

Evaluation Criteria: Mean Intersection over Union (`mIoU`) is considered as an evaluation criterion. Higher `mIoU` indicates better segmentation label predictions.

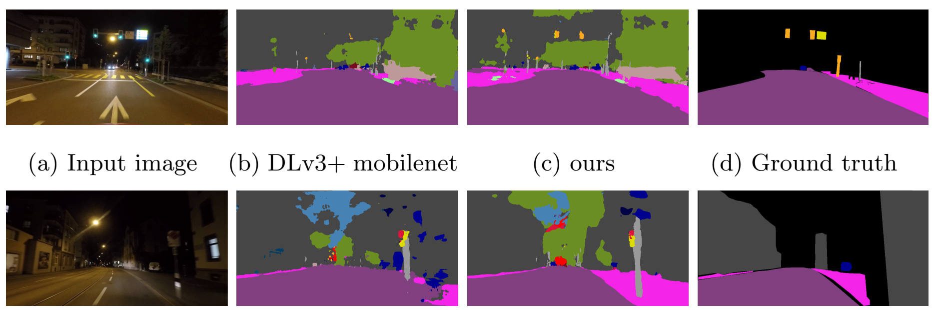
Implementation details: We implement the proposed `MALL` technique using Pytorch [29], and train it on an NVIDIA 32GB V100 GPU. During training, the network parameters are updated using Stochastic Gradient Descent(SGD) optimizer with a learning rate of 1×10^{-3}, momentum of 0.9, and weight decay of 5×10^{-4}. σ_{rgb} and σ_{xy} is set to 15, 80 respectively when reporting our results. Augmentation used for `MALL` is horizontal flip, gaussian blur and color jitter. For `MALL-sample` and `MALL-domain` methods, the mini-batch size is set to 12, images are resized to the resolution of 1024 × 512 pixels during test-time adaptation.

Comparison with TENT[52]: `TENT` is designed for single image adaptation during inference. We perform semantic segmentation task on the adverse weather datasets discussed in Sec. 4.1. For a fair comparison, we report the results of `MALL-sample` using three pre-trained models namely, DeepLabv3+ mobilenet [4], DeepLabv3+ resnet101 [4] and RefineNet [23]. The mIOU of pre-trained models on the cityscapes-val dataset is 61.6%, 78.5% and 71.4% respectively. Results are presented in Tab. 2. The `MALL` framework significantly improves the pre-trained DeepLabv3+ mobilenet, DeepLabv3+ resnet101 and RefineNet models to adapt to unseen adverse weather images directly during inference. We have observed that `MALL-sample` and `MALL-domain` methods achieve 29%, 35%, 66% and 82% better mIOU over pretrained DeepLabv3+ mobilenet model on the Night-time Driving test and Dark Zurich validation dataset respectively. Similarly `MALL-sample` and `MALL-domain` methods achieve 10.2%, 11.8%, 10.4% and 12.8% better mIOU over pretrained DeepLabv3+ resnet101 model on the Night-time Driving test and Dark Zurich validation dataset respectively. Results signify that `MALL-sample` outperforms `TENT`. `MALL` is able to adapt to unseen nighttime images using 50 images in night test datasets without any additional target domain data like `GPS` information.

Impact on Daytime Performance: For `MALL-sample` there is no drop in day-time performance as we revert to initial weights after adapting the model to a single image. For `MALL-domain` on nighttime driving dataset, we observed that a decrease in daytime mIOU performance for DeepLabv3+ mobilenet, DeepLabv3+ resnet101 and RefineNet by 3.2%, 3.6% and 3.1% respectively. Adapting to a

Model Daytime (Cityscapes) →	ND	DZ	FDD	FD	FZ	ACDC Fog	Rain	Snow	Night	C-Driving Cloud	Rain	Snow	Overcast	Day
DeepLabv3+ mobilenet[4]	28.5	11.9	25.2	36.1	26.4	47.4	37.8	30.3	15.8	34.1	27.2	27.0	37.1	61.6
with TENT[52]	33.3	17.4	23.7	39.3	31.4	52.3	42.0	41.3	21.9	36.6	29.5	30.2	38.6	61.6
with `MALL-sample`	36.8	19.8	26.6	40.5	32.6	54.9	43.4	42.7	23.9	37.9	30.8	31.4	39.7	61.6
with `MALL-domain`	38.4	21.7	27.6	40.3	32.3	55.3	42.8	41.9	24.4	37.7	30.4	30.9	39.8	59.6
DeepLabv3+ Resnet101[4]	38.2	20.2	37.9	44.4	31.2	64.1	48.3	44.0	23.5	42.1	34.6	35.7	44.9	78.5
with TENT[52]	39.3	21.6	37.4	45.8	33.6	63.5	49.4	47.6	25.2	42.6	35.9	36.9	45.9	78.5
with `MALL-sample`	42.1	22.3	38.8	47.2	34.3	64.8	49.9	48.8	26.8	43.8	36.6	38.4	47.4	78.5
with `MALL-domain`	42.7	22.8	38.7	46.7	34.1	65.4	49.7	49.5	26.9	43.7	37.4	38.9	47.7	75.7
RefineNet[23]	33.5	17.1	25.1	35.6	24.9	55.9	42.6	44.2	21.5	41.1	34.6	35.9	44.7	71.4
with TENT[52]	34.0	18.3	26.8	37.3	29.0	57.5	43.1	45.8	22.8	41.0	34.2	37.3	44.8	71.4
with `MALL-sample`	35.5	19.7	30.5	39.0	30.8	59.3	44.5	47.3	23.9	42.5	35.7	38.3	46.2	71.4
with `MALL-domain`	36.3	21.9	31.8	39.9	30.4	59.6	44.2	47.4	23.6	42.8	36.4	38.9	47.3	69.2

Table 2: Results of `MALL` on pre-trained daytime models. Dataset descriptions are provided in subsection 4.1.



(a) Input image (b) DLv3+ mobilenet (c) ours (d) Ground truth

(e) Input image (f) DLv3+ r101 (g) ours (h) Ground truth

Fig. 2: Qualitative visual comparison of our proposed `MALL` framework on pre-trained daylight models: Deeplabv3+ mobilenet, Deeplabv3+ resnet101

single image using `MALL` directly during inference can enhance generalization capability without any impact on daytime performance.

Inference Time: For a batch size of 12, with each image resolution of 1024×512, `MALL-sample` takes 1312 ms, 1564 ms and 1578 ms per iteration on DeepLabv3+ mobilenet, DeepLabv3+ resnet101 and RefineNet respectively. `MALL-domain` takes 1180 ms, 1219 ms and 1282 ms per iteration on DeepLabv3+ mobilenet, DeepLabv3+ resnet101 and RefineNet respectively.

4.2 Improvement on SOTA Daytime Models

We apply the proposed `MALL` framework on the `SOTA` daytime pre-trained models. We consider pre-trained models namely, BiseNetV2 [55], STDC[11], ISANet[17], LRASPP[16], SegFormer-B0[54], GCNet[2] and Mobilenet V2[40]. Results are

Model Daytime (Cityscapes) →	ND	DZ	FDD	FD	FZ	ACDC Fog	Rain	Snow	Night	C-Driving Cloud	Rain	Snow	Overcast	Avg
BiseNetV2 (IJCV 2021)[55]	21.8	11.2	23.8	33.5	25.3	48.3	38.3	35.7	13.7	36.6	29.4	26.6	38.6	23.7
with `MALL-domain`	27.7	16.8	33.6	35.6	26.4	52.3	40.2	36.1	19.8	38.8	31.6	31.0	41.2	**30.4**
ISANet (IJCV 2021)[17]	28.1	15.8	37.3	38.6	39.7	62.2	46.8	44.9	18.8	42.2	32.6	35.0	43.9	37.3
with `MALL-domain`	38.3	24.7	38.7	44.9	41.5	63.2	52.8	50.4	27.6	46.5	38.6	39.5	48.3	**42.7**
STDC (CVPR 2021)[11]	26.3	14.7	35.5	41.7	35.3	62.7	46.4	45.3	18.7	40.9	33.2	33.1	44.6	36.8
with `MALL-domain`	43.6	24.1	42.4	44.7	42.8	67.0	48.9	50.4	23.5	44.4	36.4	37.5	47.3	**42.5**
SegFormer (NeurIPS 2021)[54]	32.3	18.9	35.1	40.8	31.4	63.4	48.2	46.4	21.5	39.7	32.7	32.6	42.6	37.3
with `MALL-domain`	34.8	20.2	36.0	41.5	33.8	65.3	49.2	48.6	23.2	40.6	33.0	33.2	43.4	**38.7**
GCNet (TPAMI 2020)[2]	26.5	16.8	40.7	45.8	35.3	62.7	47.9	48.1	19.4	42.6	33.3	35.0	46.0	38.5
with `MALL-domain`	40.2	25.3	42.1	46.8	40.0	64.8	52.0	52.0	25.8	46.4	39.3	39.1	49.7	**43.4**
LRASPP (ICCV 2019)[16]	24.1	12.7	28.6	30.5	21.4	51.3	37.4	38.2	15.4	33.3	26.4	24.9	35.4	29.2
with `MALL-domain`	33.4	16.9	31.3	36.0	28.5	54.1	41.6	39.7	18.9	36.5	29.5	29.3	39.6	**33.5**
Mobilenet V2 (CVPR 2019)[40]	9.7	3.9	23.4	30.1	20.4	41.2	38.3	26.1	3.7	34.6	25.3	22.0	36.6	24.2
with `MALL-domain`	33.8	20.3	30.0	43.9	32.8	61.2	49.0	43.6	26.5	42.9	35.9	36.2	45.1	**38.6**

Table 3: Results of `MALL` on `SOTA` pre-trained daytime models. Dataset descriptions are provided in subsection 4.1.

presented in Tab. 3. Average `mIoU` is reported to demonstrate overall improvement in segmentation performance. We observed significant improvements with `MALL` across multiple weather conditions. We report that `MALL-domain` improves STDC [11] performance on Nighttime driving by 66% in terms of `mIoU`. `MALL` also performs well with transformer-based models such as Segformer [54]. `MALL-domain` improves Mobilenet V2[40] performance on Dark Zurich by 419%. In terms of average `mIoU` we report an improvement of 60% for Mobilenet V2. Performance gains by the `MALL` technique on various baselines validates the generalization ability of proposed approach for adverse weather and visual conditions.

4.3 Improvement on Domain Generalization Models

In Tab. 4, we demonstrate the effectiveness of the `MALL` framework over the state-of-the-art domain generalization methods for semantic segmentation IBNNet[27], Switchable whitening (SW)[28], and RobustNet[6]. We consider pre-trained models of the above-mentioned methods and apply our `MALL-domain` methods. We report that `MALL-domain` on IBNNet, RobustNet-Resnet101 shows an increase of 6.1%, 5.2% `mIoU` on the Nighttime Driving dataset. This shows that `MALL` can complement domain generalization approaches to adapt to unseen adverse weather images without altering training or requiring unlabeled target data during training. Visual results are shown in Fig. 3.

4.4 Improvement on Unsupervised Domain Adaptation Models

In this section, we consider `SOTA` `UDA` methods for night image segmentation.

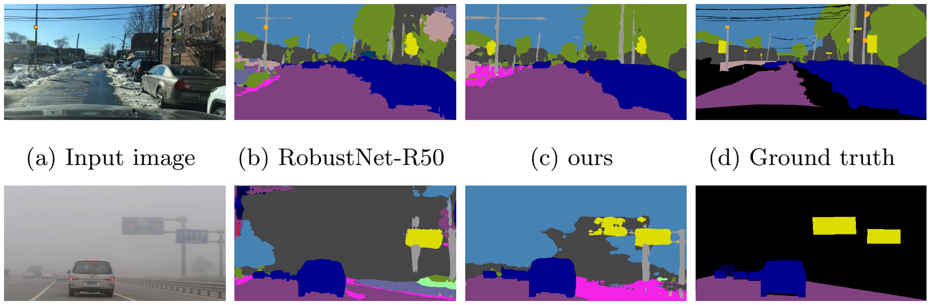
Model Daytime (Cityscapes) →	ND	DZ	FDD	FD	FZ	ACDC Fog	 Rain	 Snow	 Night	C-Driving Cloud	 Rain	 Snow	 Overcast	Avg
IBNNet[27]	33.4	21.7	31.7	41.9	34.5	63.6	50.4	50.2	25.9	43.8	35.8	35.0	48.0	39.6
with `MALL-domain`	39.5	24.8	32.5	42.7	35.2	64.7	51.5	51.2	28.9	44.9	36.4	36.2	48.9	**41.3**
SW[28]	34.0	15.6	22.2	38.2	28.6	54.8	40.2	45.6	20.7	39.2	30.9	31.6	41.7	34.1
with `MALL-domain`	37.2	21.1	30.6	39.8	30.5	59.8	44.5	49.1	26.2	40.5	33.6	34.8	44.4	**37.9**
RobustNet-Resnet50[6]	35.8	20.9	37.0	43.0	34.5	60.6	46.9	50.1	26.1	42.7	36.5	34.7	45.7	39.5
with `MALL-domain`	41.0	22.5	39.3	43.8	35.7	61.9	47.7	51.2	28.7	43.5	37.6	36.1	45.9	**41.2**
RobustNet-Resnet101[6]	37.7	21.8	38.8	43.5	33.5	59.3	46.5	45.2	26.8	41.6	34.4	33.8	42.7	38.8
with `MALL-domain`	42.9	24.8	39.7	44.7	37.4	64.2	49.9	51.1	30.7	42.0	35.8	36.4	47.3	**42.0**

Table 4: Results of `MALL` framework on `SOTA` Domain Generalization methods. Dataset descriptions are provided in subsection 4.1.

We evaluate the performance in two scenarios: (1) where the pre-trained model has not seen night images during training, and (2) where the model is trained on nighttime images. For the first category of models trained on source data (daytime) only, we experiment with Zeroshot-DayNight[20]. For the second category, when the models are trained on source data (daytime) and target data (nighttime), we consider DANNet (PSPNet)[53]. Tab. 5, Tab. 6 shows the results. We observe that `MALL-domain` improves Zeroshot-DayNight[20] by 2.6% and 1.7% in terms of `mIoU` on Nighttime Driving-test and Dark Zurich-test respectively. `MALL-domain` method on DANNet[53] and MGCDA[38] increases mIOU by 1.1% and 0.3%, 0.5% and 0.7% on the Nighttime Driving test, Dark Zurich test datasets, respectively. Furthermore, `MALL` can significantly improve DANNet performance in other weather conditions, as demonstrated in Tab. 7. For instance `MALL-domain` on DANNet improves `mIoU` on Foggy Zurich dataset by 3.6%. Significant improvement in the performance using `MALL-domain` demonstrates the ability of our framework to enhance the generalization ability of `SOTA` nighttime image segmentation methods in both the scenarios: with and without access to unlabeled target data during training. Fig. 2 shows qualitative results for improvement over models trained on daytime images. Whereas in Fig. 4

Method	**ND-test**	**DZ-test**
Trained on source data only		
Zeroshot-DN [20]	41.2	34.5
with `MALL-domain`	**43.8**	**36.2**
Trained on source and target data		
ADVENT[51]	34.7	29.7
BDL[22]	34.7	30.8
AdaptSegNet[48]	34.5	30.4
DMAda[9]	41.6	32.1
Day2Night[44]	45.1	–
GCMA[33]	45.6	42.0
MGCDA[38]	49.4	42.5
with `MALL-domain`	**49.9**	43.2
DANNet[53]	47.7	45.2
with `MALL-sample`	48.3	45.3
with `MALL-domain`	48.8	**45.5**

Table 5: Results of `MALL` framework on state-of-the-art methods in night image segmentation, zeroshot-DN: Zeroshot DayNight[20].



(a) Input image (b) RobustNet-R50 (c) ours (d) Ground truth

(e) Input image (f) RobustNet-R101 (g) ours (h) Ground truth

Fig. 3: Qualitative visual comparison of our proposed `MALL` framework on pre-trained Domain generalization models: Robustnet-Resnet50, Robustnet-Resnet101 ours: `MALL-domain` method.

Model						ACDC				C-Driving				
Nighttime (DZ) →	ND	DZ	FDD	FD	FZ	Fog	Rain	Snow	Night	Cloud	Rain	Snow	Overcast	Avg
DANNet[53]	47.7	36.7	33.9	35.4	31.6	52.5	44.3	47.0	39.8	42.0	34.1	35.5	42.7	40.2
with `MALL-domain`	48.8	37.3	36.4	37.8	35.2	53.8	45.8	48.9	41.5	43.5	36.7	37.8	43.9	**42.1**

Table 7: Results of `MALL` framework on DANNet[53]. Dataset descriptions are provided in subsection 4.1.

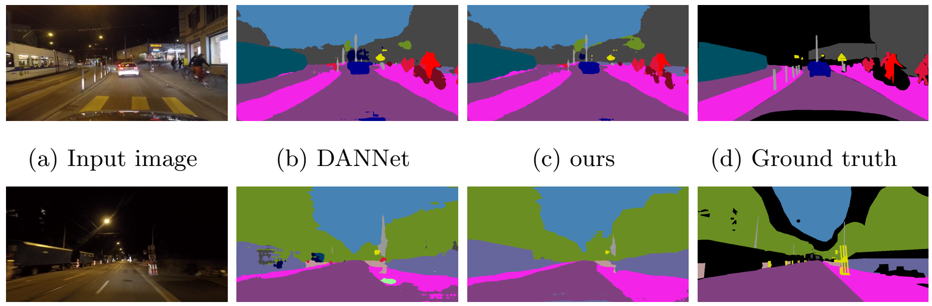
we show qualitative results for models which are trained on both daytime and nighttime images.

`MALL` qualitative visual results for the night image datasets for the DANNet pre-trained model with `MALL-domain` are reported in Fig. 4. In Fig. 4 we see that DANNet fails to label moving object (car), or detect traffic light, whereas by using `MALL` during inference it can detect traffic light, and car correctly. Results on Dark Zurich-val, `MGCDA` segmentation outputs, class wise `mIoU` values on Dark Zurich-test dataset are reported in the supplementary material.

Method	**DZ-val**
GCMA [33]	26.6
MGCDA [38]	26.1
with `MALL-domain` (ours)	**26.8**
DANNet [53]	36.7
with `MALL-sample` (ours)	**37.1**
with `MALL-domain` (ours)	**37.3**

Table 6: Results of `MALL` framework on state-of-the-art methods in night image segmentation

Ablation Study: In order to demonstrate the effectiveness of class imbalance re-weighting and sample importance weighting, we perform an ablation study on loss formulation considering a pre-trained model, DeepLabv3+ mobilenet[4]. Results are shown in Tab. 8.



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

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

Fig. 4: Qualitative visual comparison of our proposed `MALL` framework on DANNet pre-trained model on two images from night image datasets. DANNet with `MALL-domain` further improve the results.

Method	**ND-test**	**DZ-val**
DeepLabv3+ mobilenet[4]	28.5	11.9
with `MALL-domain` (ours)	38.4	21.7
w/o class imbalance re-weighting	37.4	20.6
w/o sample importance weighting	36.1	19.2

Table 8: Ablation study on several loss variants using DeepLabv3+ mobilenet[4]

.

Conclusion: This paper introduces a novel weighted log multi-class normalized cut loss to enforce edge consistency prior for improving semantic segmentation predictions in adverse weather and visual conditions. The proposed framework improves segmentation performance of `SOTA` models trained on daytime images as well as the ones which have seen adverse weather images during training. We show that our framework can be used in conjunction with `SOTA` domain generalization approaches to further improve their performance for adverse weather images. Our experiments indicate that edge consistency prior could also be effective in multiple adverse weather conditions, such as rain, snow, night, cloudy, overcast and fog.

Acknowledgment: We thank UQ-IIT Delhi Research Academy (UQIDAR) for providing contingency grant. This work has also been partly supported by the funding received from DST through the IMPRINT program (IMP/2019/000250). We acknowledge National Super-computing Mission (NSM) for providing computing resources of 'PARAM Siddhi-AI', under National PARAM Super-computing Facility (NPSF), C-DAC, Pune, and supported by the Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology (DST), Government of India.

References

1. Adams, A., Baek, J., Davis, M.A.: Fast high-dimensional filtering using the permutohedral lattice. In: Computer graphics forum. vol. 29, pp. 753–762. Wiley Online Library (2010) 7
2. Cao, Y., Xu, J., Lin, S., Wei, F., Hu, H.: Global context networks. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–1 (2020). https://doi.org/10.1109/TPAMI.2020.3047209 10, 11
3. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017) 2
4. 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) 2, 9, 10, 13, 14
5. Chen, Y., Li, W., Van Gool, L.: Road: Reality oriented adaptation for semantic segmentation of urban scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7892–7901 (2018) 2
6. 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, 4, 11, 12
7. 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: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3213–3223 (2016) 6, 9
8. Dai, D., Sakaridis, C., Hecker, S., Van Gool, L.: Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. International Journal of Computer Vision **128**(5), 1182–1204 (May 2020), `https://doi.org/10.1007/s11263-019-01182-4` 2
9. 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) 4, 12
10. 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) 8
11. Fan, M., Lai, S., Huang, J., Wei, X., Chai, Z., Luo, J., Wei, X.: Rethinking bisenet for real-time semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9716–9725 (2021) 10, 11
12. Geusebroek, J.M., Van den Boomgaard, R., Smeulders, A.W.M., Geerts, H.: Color invariance. IEEE Transactions on Pattern analysis and machine intelligence **23**(12), 1338–1350 (2001) 4
13. Girshick, R.: Fast r-cnn in proceedings of the ieee international conference on computer vision (pp. 1440–1448). Piscataway, NJ: IEEE.[Google Scholar] (2015) 2
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 2
15. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: International conference on machine learning. pp. 1989–1998. PMLR (2018) 2

16. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., Adam, H.: Searching for mobilenetv3. In: The IEEE International Conference on Computer Vision (ICCV). pp. 1314–1324 (October 2019). https://doi.org/10.1109/ICCV.2019.00140 10, 11
17. Huang, L., Yuan, Y., Guo, J., Zhang, C., Chen, X., Wang, J.: Interlaced sparse self-attention for semantic segmentation. arXiv preprint arXiv:1907.12273 (2019) 10, 11
18. Kim, M., Byun, H.: Learning texture invariant representation for domain adaptation of semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12975–12984 (2020) 2
19. Kundu, J.N., Kulkarni, A., Singh, A., Jampani, V., Babu, R.V.: Generalize then adapt: Source-free domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7046–7056 (October 2021) 2
20. Lengyel, A., Garg, S., Milford, M., van Gemert, J.C.: Zero-shot day-night domain adaptation with a physics prior. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4399–4409 (2021) 4, 12
21. 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) 2
22. 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) 12
23. 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) 9, 10
24. 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) 9
25. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015) 2
26. Luo, Y., Zheng, L., Guan, T., Yu, J., Yang, Y.: Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2507–2516 (2019) 2
27. 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) 4, 11, 12
28. 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) 4, 11, 12
29. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32**, 8026–8037 (2019) 9
30. Quiñonero-Candela, J., Sugiyama, M., Lawrence, N.D., Schwaighofer, A.: Dataset shift in machine learning. Mit Press (2009) 2
31. Robey, A., Pappas, G.J., Hassani, H.: Model-based domain generalization. arXiv preprint arXiv:2102.11436 (2021) 4

32. Romera, E., Bergasa, L.M., Yang, K., Alvarez, J.M., Barea, R.: Bridging the day and night domain gap for semantic segmentation. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp. 1312–1318. IEEE (2019) 4
33. 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) 2, 4, 12, 13
34. 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) 8
35. Sakaridis, C., Dai, D., Hecker, S., Van Gool, L.: Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In: European Conference on Computer Vision (ECCV). pp. 707–724 (2018) 2
36. Sakaridis, C., Dai, D., Hecker, S., Van Gool, L.: Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 687–704 (2018) 9
37. Sakaridis, C., Dai, D., Van Gool, L.: Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision **126**(9), 973–992 (2018) 8
38. 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) 2, 4, 12, 13
39. 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) 8
40. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520 (2018) 10, 11
41. Sankaranarayanan, S., Balaji, Y., Jain, A., Lim, S.N., Chellappa, R.: Learning from synthetic data: Addressing domain shift for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3752–3761 (2018) 2
42. Shannon, C.E.: A mathematical theory of communication. The Bell system technical journal **27**(3), 379–423 (1948) 7
43. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on pattern analysis and machine intelligence **22**(8), 888–905 (2000) 3, 5
44. Sun, L., Wang, K., Yang, K., Xiang, K.: See clearer at night: towards robust nighttime semantic segmentation through day-night image conversion. In: Artificial Intelligence and Machine Learning in Defense Applications. vol. 11169, p. 111690A. International Society for Optics and Photonics (2019) 4, 12
45. Sun, Y., Wang, X., Zhuang, L., Miller, J., Hardt, M., Efros, A.A.: Test-time training with self-supervision for generalization under distribution shifts. In: ICML (2020) 2, 3
46. 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) 5, 6
47. 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: Pro-

ceedings of the IEEE conference on computer vision and pattern recognition. pp. 7472–7481 (2018) 2
48. 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) 12
49. Tsai, Y.H., Sohn, K., Schulter, S., Chandraker, M.: Domain adaptation for structured output via discriminative patch representations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1456–1465 (2019) 2
50. 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) 2
51. 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) 12
52. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. arXiv preprint arXiv:2006.10726 (2020) 2, 3, 4, 9, 10
53. 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) 2, 4, 12, 13
54. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. arXiv preprint arXiv:2105.15203 (2021) 10, 11
55. Yu, C., Gao, C., Wang, J., Yu, G., Shen, C., Sang, N.: Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. International Journal of Computer Vision **129**(11), 3051–3068 (2021) 10, 11
56. 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: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020) 9
57. Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F.: Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. arXiv preprint arXiv:2101.10979 (2021) 2
58. Zhang, Y., Qiu, Z., Yao, T., Liu, D., Mei, T.: Fully convolutional adaptation networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6810–6818 (2018) 2
59. Zheng, Z., Yang, Y.: Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. International Journal of Computer Vision **129**(4), 1106–1120 (2021) 2
60. Zhou, K., Yang, Y., Hospedales, T., Xiang, T.: Deep domain-adversarial image generation for domain generalisation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 13025–13032 (2020) 4
61. Zhou, K., Yang, Y., Hospedales, T., Xiang, T.: Learning to generate novel domains for domain generalization. In: European Conference on Computer Vision. pp. 561–578. Springer (2020) 4

62. Zou, Y., Yu, Z., Kumar, B., Wang, J.: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proceedings of the European conference on computer vision (ECCV). pp. 289–305 (2018) 2
63. Zou, Y., Yu, Z., Kumar, B., Wang, J.: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proceedings of the European conference on computer vision (ECCV). pp. 289–305 (2018) 2
64. Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J.: Confidence regularized self-training. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5982–5991 (2019) 2