DANNet Modifications for Adverse Weather Semantic Segmentation

Ho Chiu Hanlu Li Haoru Xue Allen Zeng
University of California San Diego
9500 Gilman Dr, La Jolla, CA 92093
{ h8chiu, hal037, hxue, azeng }@ucsd.edu

Abstract

Being able to function in adverse condition is essential for autonomous driving vehicles. ACDC, the Adverse Conditions Dataset with Correspondences is a dataset that contains a large set of 4006 images which are equally distributed between four common adverse conditions: fog, night-time, rain, and snow. This work is trying to adapt DANNet, a robust domain adaptation network to the ACDC dataset, especially towards extreme weather condition such as rain and snow. Furthermore, we develop a novel pre-processing method for images. This pre-processing model aims to resolve misalignment in the images between normal and adverse condition, which leads to wrong predictions for semantic segmentation. This proposed modification has allowed DANNet to perform competitively compared to other models trained in adverse conditions.

1. Introduction

Self-driving cars require robust perception systems that can work in a variety of weather conditions and times of day. However, most autonomous vehicle research is currently focused on enabling vehicles to understand their environments and navigate safely in daytime conditions. Some vision models for image semantic segmentation, such as DANNet [7] extend research into nighttime conditions. Only recently has larger datasets containing nighttime conditions for the task of semantic segmentation been available, such as the Dark Zurich dataset [4]. And prior to the Adverse Conditions Dataset with Correspondences (ACDC) [5], there has not been any large dataset available specifically for semantic segmentation which contains the adverse conditions fog, rain, and snow, in addition to night. There have been previous datasets that contain some examples of adverse conditions, but they either lack dense pixel-level annotations or contain severe annotation errors [5].

In this paper, we propose a modification to DANNet such that it is able to perform semantic segmentation on ACDC. In the original DANNet paper, it is assumed that ground truth annotations only exist for the daytime set of images, and that there exists coarsely-aligned daytime and nighttime images. In our case, we are given images of adverse road conditions (fog, nighttime, rain, snow), their corresponding densely-annotated ground truth images, and for each adverse image a coarsely aligned normal-condition daytime image. There is no direct ground truth for the normal images. So, our training pipeline and architecture is similar to that of DANNet, except the inputs, ground truth, and loss functions are correspondingly swapped.

2. Related Research

The DANNet [7] was originally trained using the CityScapes [1] and Dark Zurich [4] datasets. DANNet addresses the problem of nighttime semantic segmentation via adversarial learning with a multi-target domain adaption network. In its one-stage architecture, it uses an image relighting network, a semantic segmentation network, and two discriminators. The CityScapes dataset, containing both daytime condition color images and ground truth annotations, is used to train the model through the *light loss* and *segmentation loss*. The Dark Zurich day and night coarsely-aligned images were used to train the model through the *light loss* and *static loss*, by providing weak supervision. Since the day and night images may have moving objects within them, the static loss only incorporates static object classes.

The Dark Zurich dataset was introduced along with the Guided Curriculum Model Adaptation (GCMA) [4] and Map-Guided Curriculum Domain Adaptation models (MGCDA) [6]. GCMA adapts semantic segmentation models from the daytime domain to the nighttime domain by using unlabeled real images and labeled synthetic images. The images become progressively darker, corresponding to darker times of day, and exploits these *cross-time-of-day* correspondences to guide model inference. MGCDA uses three separate models to predict labels: a daytime model, a twilight model, and a nighttime model. Additionally, it trains two CycleGANs for converting between the day and twilight domains, and for converting between the day and

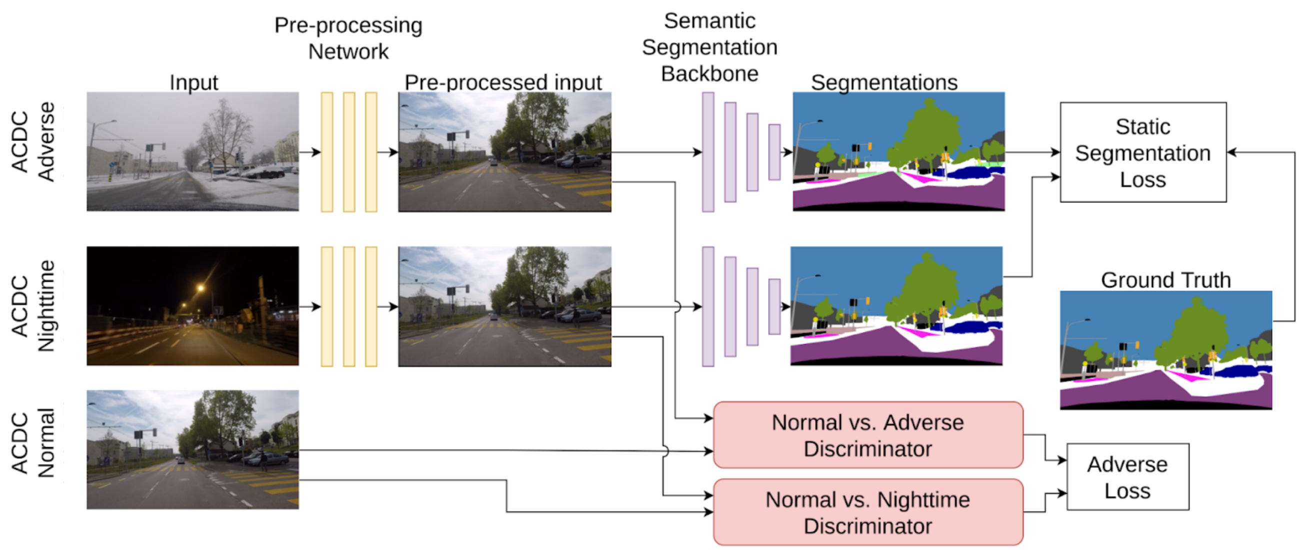

Figure 1. The architecture of our proposed model, which is a modification of DANNet [7]. The adverse condition images (fog, night, rain, snow) are first input into a pre-processing network which generates a new image that resembles a normal condition image. These processed images then pass through a traditional convolutional neural network, the *Semantic Segmentation Backbone*, to produce predicted segmentation labels. A static segmentation loss is computed by comparing the predictions with ground truth labels. Additionally, the input images are compared against their normal condition correspondences and an adverse loss is produced.

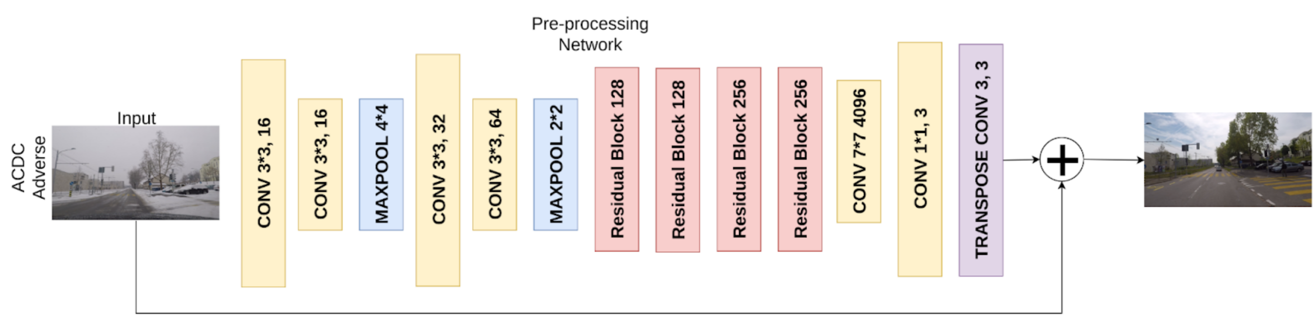

Figure 2. The structure of our new pre-processing network. The output of this network is added to the input image to produce a processed image that is in the normal condition domain. This output is subsequently passed into a semantic segmentation network.

night domains.

The ACDC paper [5] lists GCMA and MGCDA as some of the top performing pre-trained models on the ACDC dataset for each individual condition as well as all conditions jointly. However, DANNet is the best performing pre-trained model for the night condition. However, it does not list any results for the fog, rain, snow, or all categories. This is because of DANNet's relighting pre-processing network. It does not make sense to apply this relighting to the other adverse conditions. So in this project, we aim to modify DANNet such that it produces the appropriate predictions and metrics for the other adverse conditions.

To the best of the authors' knowledge, the ACDC [5] dataset is the largest adverse-condition dataset to date for the task of semantic segmentation (and uncertainty-aware semantic segmentation). ACDC contains the same classes as Cityscapes, which enables easy adaptation of trained models from one dataset to the other. ACDC contains accurate, densely annotated ground truth labels. It consists of 4006 adverse condition images with ground truth, and each adverse image has a corresponding "reference" normal condition daytime image. This normal image is coarsely aligned to the adverse image via matching the GPS sequences of each data collection. There are 1000 fog, 1006 night, 1000 rain, and 1000 snow condition images. Additionally, "invalid" regions corresponding to uncertain ground truth labels are provided for the task of uncertainty-aware semantic segmentation.

Model	Trained on	Fog	Night	Rain	Snow	All
RefineNet	CS	46.4	29.0	52.6	43.3	43.7
DeepLabv3+	CS	45.7	25.0	50.0	42.0	41.6
SFSU	FC	45.6	29.5	51.6	41.4	42.9
CMAda	FC-DBF+FZ	51.2	32.0	53.4	47.6	47.1
DMAda	ND	50.7	32.7	54.9	48.9	47.9
GCMA	CS+DZ	52.4	42.9	58.0	53.8	53.4
MGCDA	CS+DZ	45.9	40.8	54.2	50.5	48.9
DANNet	CS+DZ	42.8	47.6	38.4	35.2	40.1
DANNet (Ours)	ACDC	51.7	27.3	41.9	42.6	40.9

Table 1. Comparison with other models on ACDC, in terms of mIoU values. Note that unlike in the ACDC Table 4 [5], our results on the DANNet are not directly comparable because we use ACDC as our training set. The "trained on" column uses the same dataset acronyms as in ACDC.

3. Methodology

3.1. Framework overview

We base our model and training pipeline on that of DANNet [7]. Our network architecture, shown in Figure 1, consists of a pre-processing network and a semantic segmentation backbone. The network also uses discriminators that determine if the input image is an adverse condition image or a normal condition image.

3.2. Network architecture

The modules of our modified DANNet are described in detail below:

Image pre-processing network For refining our model in adverse condition, An image pre-processing network was designed to remove snow, haze and rain. The detailed network architecture is shown in Figure 2. The network consists of convolutional layers, maxpool layers residual blocks and one transposed convolutional layer. This pre-processing network takes the adverse condition image input and output a image closer to its normal condition pair. The result is shown in Figure 3, 4, 5.

Semantic segmentation network Similar to DANNet, we use a semantic segmentation model that was pre-trained on Cityscapes. The segmentation backbone we used was PSPNet [8], which incorporates ResNet-101 [2]. We continue training these models using the ACDC dataset, minimizing the static segmentation loss shown in Figure 1.

Discriminators Similar to DANNet, we utilize discriminators which distinguish between input images of the adverse condition domains or the normal daytime condition domain. Adversarial learning is performed on the output space. This adversarial loss ensures that the pre-processing network, which can be interpreted as a generator, produces the best possible results. The pre-processing network takes adverse condition images as input and attempts to transform those images to the daytime domain. Subsequently, a daytime semantic segmentation model can be used to predict pixel-level class labels.

3.3. ACDC dataset

Images in ACDC dataset are evenly distributed into four sets of the examined conditions. 1000 foggy, 1006 night-time, 1000 rainy and 1000 snowy images are captured from the recordings for dense pixel-level semantic annotation. Each adverse condition image set are also split into 400 training, 100 validation and 500 test images, with the exception of the night-time set that includes 106 validation images. This results in a total of 1600 training and 406 validation images with publicly available annotations and 2000 test images with annotations withheld for benchmarking purposes. The only inferiority of ACDC is that ground truth of normal condition is missing. However, it does not affect our model result as the ground truth of other condition is provided.

4. Experimental Evaluation

Due to limited GPU resources, we trained the pre-processing model and the semantic segmentation model separately. To produce our final results, we save the intermediate images produced by the pre-processing model, and then separately feed them into the segmentation model afterwards to produce the desired label prediction images.

We show results of the pre-processing network in Figures 3, 4, and 5. Results of the adverse condition semantic segmentation are shown in Figure 6, 7, 8, and 9. Quantitative results of our new model are shown in Table 1. For completeness, we also evaluate the baseline DANNet, trained only on Cityscapes and Dark Zurich, on the other adverse weather conditions.

We focus less on the nighttime domain which was the strength of the original DANNet architecture. In training the prepossessing network, we deliberately only trained it with daytime adverse scenarios. Reflected on the segmentation results, we see a consistent increase in mIOU across all the daytime adverse scenarios, while a decrease in the nighttime condition.

As shown in Table 1, our results in the Fog domain are comparable to the other given top scores. However, our results in the Night domain are worse than in the version of DANNet pretrained on Cityscapes and Dark Zurich only. It is possible that some discrepancy in score is due to our limited time constraints leading to less hyperparameter tuning. Additionally, since the DANNet-based model shares weights of its segmentation network between the different input domains, it is possible that the increased number of domains in our case surpasses the models' capability to perform well on ACDC. In other words, perhaps the results can be improved by adding more parameters to our preprocessing network and semantic segmentation backbone.

For the rain category, our model performs worse than the other pre-trained models. This is possibly due to the preprocessing model not being able to represent the rain artifacts sufficiently well. For the snow category, our model is competitive with RefineNet, DeepLabv3+, and SFSU. But, our model does not achieve top-tier performance. Overall for all conditions combined, our model is competitive but the score is reduced due to poor nighttime performance.

Given more time for parameter tuning of our model, we believe we could attain better results for all of the categories, and especially for the nighttime domain.

5. Concluding Remarks

Overall, our new model performed competitively in predicting labels of the ACDC dataset. Given more time, we would tune our hyperparameters and training pipeline to obtain better results. Additionally, we would investigate how we can improve our results for the nighttime segmentation case.

In future work, we would like to investigate a better way to process the coarsely-aligned adverse condition and normal condition images. A simple approach would be to detect visual feature points in each image, and warp the normal condition image to align more closely to the adverse condition image. A more robust but computationally intensive approach would be to use a technique such as PatchGAN [3]. The PatchGAN is a type of discriminator which only penalizes structures in local image patches, so misalignments do not affect the overall network training as much.

References

[1] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding, 2016. 1

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 3

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2018. 4

[4] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation, 2019. 1

[5] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding, 2021. 1, 2, 3

[6] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation, 2021. 1

[7] Xinyi Wu, Zhenyao Wu, Hao Guo, Lili Ju, and Song Wang. Dannet: A one-stage domain adaptation network for unsupervised nighttime semantic segmentation, 2021. 1, 2, 3

[8] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network, 2017. 3

Figure 3. Results of the pre-processing network on fog condition input images. Top is the adverse condition input, middle is the processed image, and bottom is the corresponding normal condition image.

Figure 4. Results of the pre-processing network on rain condition input images. Top is the adverse condition input, middle is the processed image, and bottom is the corresponding normal condition image.

Figure 5. Results of the pre-processing network on snow condition input images. Top is the adverse condition input, middle is the processed image, and bottom is the corresponding normal condition image.

Figure 6. Semantic segmentation of nighttime images (left) and the predicted labels (right).

Figure 7. Semantic segmentation of fog images (left) and the predicted labels (right).

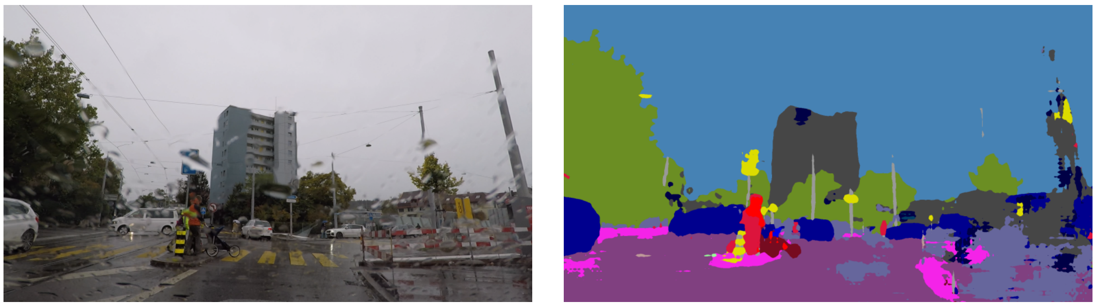

Figure 8. Semantic segmentation of rain images (left) and the predicted labels (right).

Figure 9. Semantic segmentation of snow images (left) and the predicted labels (right).