Semantic Segmentation under Adverse Conditions: A Weather and Nighttime-aware Synthetic Data-based Approach

Abdulrahman Kerim[1]
a.kerim@lancaster.ac.uk
Felipe Chamone[2]
cadar@dcc.ufmg.br
Washington Ramos[2]
washington.ramos@dcc.ufmg.br
Leandro Soriano Marcolino[1]
l.marcolino@lancaster.ac.uk
Erickson R. Nascimento[2]
erickson@dcc.ufmg.br
Richard Jiang[1]
r.jiang2@lancaster.ac.uk

[1] School of Computing and Communications
Lancaster University
Lancaster, UK

[2] Computer Science Department
Universidade Federal de Minas Gerais,
Minas Gerais, Brazil

Abstract

Recent semantic segmentation models perform well under standard weather conditions and sufficient illumination but struggle with adverse weather conditions and nighttime. Collecting and annotating training data under these conditions is expensive, time-consuming, error-prone, and not always practical. Usually, synthetic data is used as a feasible data source to increase the amount of training data. However, just directly using synthetic data may actually harm the model's performance under normal weather conditions while getting only small gains in adverse situations. Therefore, we present a novel architecture specifically designed for using synthetic training data for domain adaptation. We propose a simple yet powerful addition to DeepLabV3+ by using weather and time-of-the-day supervisors trained with multi-task learning, making it both weather and nighttime aware, which improves its mIoU accuracy by 14 percentage points on the ACDC dataset while maintaining a score of 75% mIoU on the Cityscapes dataset. Our code is available at https://github.com/lsmcolab/Semantic-Segmentation-under-Adverse-Conditions.

1 Introduction

Understanding the environment using visual data has been an active research problem since the early beginning of computer vision. It started to attract even more researchers with the great advancement in autonomous cars [20, 38, 42], human-computer-interaction [23, 27, 29], and augmented reality [3, 7, 10]. Semantic segmentation is at the core of these

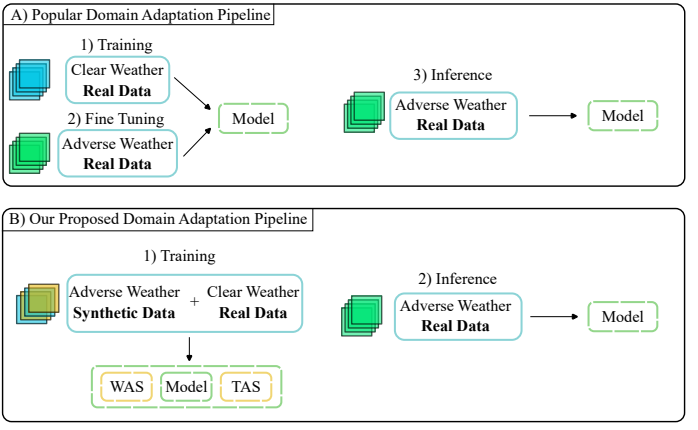

Figure 1: **Existing domain adaptation vs. our proposed pipeline.** Unlike other approaches, our pipeline utilizes synthetic data, Weather-Aware-Supervisor (WAS), and Time-Aware-Supervisor (TAS) to handle standard-to-adverse domain adaptation. Leveraging our synthetic-aware training procedure, we train our weather and daytime-nighttime aware architecture, simultaneously, on synthetic adverse weather and real normal weather data.

applications, with the data-driven supervised learning methods dominating this field, achieving state-of-the-art results [5, 14, 31, 48, 49]. Training these models on real data requires large-scale human annotated images, which is expensive and time-consuming, especially for images taken under challenging weather and illumination conditions such as fog and nighttime. For instance, a person takes about 90 minutes to annotate an image from the Cityscapes dataset [9], which contains only daylight and clear weather conditions, while it exceeds three hours for the Adverse Conditions Dataset with Correspondences (ACDC) [33] dataset.

Despite the success of recent semantic segmentation models in clear weather and standard illumination conditions, these methods struggle with adverse conditions (*e.g.*, rainy, foggy, snowy, and nighttime), which degrade the feature extraction process. Falling rain and snow particles change the visual appearance of objects, partially occlude them, and cause distortion on the camera sensor, while fog works as a low-pass filter, removing high-frequency components. Nighttime is even more problematic because of the dramatic change in the light distribution and other severe artifacts, such as lens flare, bright spots, and chromatic aberration. Yet, few works have tried to investigate the effect of weather conditions and nighttime in semantic segmentation [1, 17, 21, 24, 46]. Although they achieve remarkable results, they are limited to one weather condition only and are too narrow in their scope.

In this paper, we propose a novel training procedure to address the issues in the semantic segmentation under adverse conditions and in the annotation efforts, simultaneously. We leverage synthetic data to produce ground-truth images at no human annotation effort and create a new dataset, the AWSS, which is composed of images specially generated by a modified version of the *Silver* [19] simulator. To reduce the gap between synthetic and real, our approach combines synthetic and real images by alternating their batches at training time as illustrated in Fig. 1. We also propose the Weather-Aware Supervisor (WAS) and the Time-Aware Supervisor (TAS), which are trained jointly with the main module to improve the feature extraction. Our main module derives from the DeepLabV3+ which contains the powerful atrous convolutions that increase the receptive field while not increasing the dimensions of feature maps and computation cost. Thus, better performance at low computation.

Table 1: **Comparison among synthetic semantic segmentation datasets.** Our dataset, named AWSS, is composed of photo-realistic pixel-wise annotated images under standard and adverse conditions.

	Weather Conditions				Times-of-Day		Photo-realism	Public Availability
	Normal	Rain	Fog	Snow	Daytime	Nighttime	/	/
GTA-V [30]	✔	✔	-	-	✔	-	✔	✔
Synscapes [39, 42]	✔	-	-	-	✔	-	✔	✔
Virtual KITTI [18]	✔	✔	✔	-	✔	-	-	✔
Synthia [32]	✔	✔	-	✔	✔	✔	-	-
SHIFT [37]	✔	✔	✔	-	✔	✔	✔	✔
AWSS (Ours)	✔	✔	✔	✔	✔	✔	✔	✔

Unlike the current methods that work only with a single weather condition, our approach can handle the three main ones, *i.e.*, rainy, foggy, and snowy, as well as nighttime images. The results show that our novel model achieves state-of-the-art results under adverse weather conditions (0.49 mIoU on ACDC) while it maintains adequate performance under standard conditions (0.75 mIoU on Cityscapes).

In summary, our contributions are three-fold: *i)* a novel synthetic-aware training procedure that can be used to train on both synthetic and real data simultaneously. In particular, we significantly improve DeepLabV3+ [5] robustness on adverse conditions by making its encoder both weather and nighttime aware;[1] *ii)* We extend the *Silver* [19] simulator to generate more photo-realistic and diverse adverse weather conditions and increase the supported semantic segmentation classes; *iii)* leveraging our modified version of *Silver*, we generate a new synthetic semantic segmentation dataset, the AWSS, composed of photo-realistic annotated images spanning foggy, rainy, and snowy weather conditions and nighttime attributes.

2 Related Work

Synthetic data for semantic segmentation. The high performance of recent semantic segmentation models is associated with the ability to train deep models on large-scale training data. The early real semantic segmentation datasets like CamVid [4], Stanford Background [22, 35, 36], and KITTI-Layout [2] are limited in terms of the number of training samples, classes, resolution, and diversity. The problem is partially alleviated with the recent availability of datasets like Cityscapes [9], ACDC [33], ADE20K [50], and Mapillary Vistas [28]. Nevertheless, annotating large-scale datasets of high-resolution images is still the bottleneck. At the same time, ensuring diverse training data under challenging attributes like adverse weather conditions is not only dangerous, time-consuming, and hard to collect but also cumbersome and subjective to human errors in the annotation process.

Synthetic data comes as a resort to handle all the above issues. Their success in computer vision is specifically seen in semantic segmentation. Goyal *et al*. [16] demonstrate that augmenting synthetic data with weakly annotated data can improve the performance on the PASCAL VOC dataset [13]. Similarly, Richter *et al*. [30] generate synthetic training data by utilizing the Grand Theft Auto V game. They show that training semantic segmentation models on one third of the training split of CamVid [4] dataset along with their generated

[1] The synthetic data, code, and our modified version of the *Silver* [19] simulator are all publicly available under the paper's GitHub repository.

synthetic data achieves superior results compared to training on the full CamVid [4]. In parallel, Ivanovs *et al*. [18] augment the Cityscapes [9] dataset with synthetic images generated using the CARLA [11] simulator. They show that the performance improves when compared to training only on Cityscapes [9]. Similar to these works, we use synthetic data to boost the performance of semantic segmentation models. However, we tackle the domain shift problem using synthetic data and a synthetic-aware training procedure.

Domain adaptation in semantic segmentation. A major limitation of synthetic data is the domain shift: models trained on synthetic data do not perform well on real-world data [12, 34, 47]. Sankaranarayanan *et al*. [34] propose a Generative Adversarial Network (GAN) based approach that minimizes the distance between the encodings of both domains. They show that their approach can boost the performance of synthetic-to-real domain adaptation tasks. Our work is similar to theirs as we use synthetic data for domain adaptation and propose a synthetic-aware training procedure. However, our work tackles this problem under harder set-ups utilizing synthetic data to mitigate standard-to-adverse domain shifts. In the same context, Alshammari *et al*. [1] address standard to foggy weather domain shift by using an adversarial training strategy that guides the model to produce outputs close to the target domain. Similarly, Ma *et al*. [24] tackle standard weather to foggy weather domain adaptation using both fog and style variations by adopting a Cumulative style-fog-dual disentanglement Domain Adaptation method (CuDA-Net). Alternatively, Xu *et al*. [46] address the daytime to nighttime domain shift. They utilize a novel Curriculum Domain Adaptation method (CDAda) that uses labeled synthetic nighttime images. Our method is closely related to these works. However, we tackle domain adaptation from a standard domain (*i.e.,* daytime and normal weather condition) to an adverse domain (*i.e.*, nighttime and adverse weather conditions such as rain, fog, and snow).

3 The AWSS Dataset

There have been many synthetic datasets proposed for the semantic segmentation problem. However, they are usually non-photo-realistic such as Synthia [32] and Virtual KITTI [15], limited in diversity such as GTA-V [30] and Synscapes [39, 44] as clearly demonstrated in Table 1. Recently, SHIFT [37] dataset was introduced, which is photo-realistic and diverse similar to our generated synthetic dataset but does not cover the snowy weather.

We extend *Silver*, proposed by Kerim *et al*. [19], to generate adverse weather photo-realistic images along with their corresponding ground-truth for the semantic segmentation task. We generate the Adverse Weather Synthetic Segmentation (AWSS) dataset, which comprises 1,250 images with a resolution of $1{,}200 \times 780$ pixels and spans normal, rainy, foggy, and snowy weather conditions at daytime and nighttime. It follows the same conventions, *i.e.*, classes definitions and color encoding, as Cityscapes [9] and ACDC [33] datasets. However, we limit the number of classes to 10, namely *Road*, *Sidewalk*, *Building*, *Pole*, *Traffic Light*, *Traffic Sign*, *Vegetation*, *Sky*, *Person*, and *Car*. Figure 2 shows sample images from the AWSS dataset spanning various standard and challenging attributes.

Extensions to the Silver framework. *Silver* is based on the Unity game engine [41]. It allows users to create 3D virtual worlds by only specifying a set of scene descriptive parameters like the weather condition, time-of-the-day, number of cars and humans, camera type, and lens artifacts. The simulator achieves photo-realism by using the recent High Definition

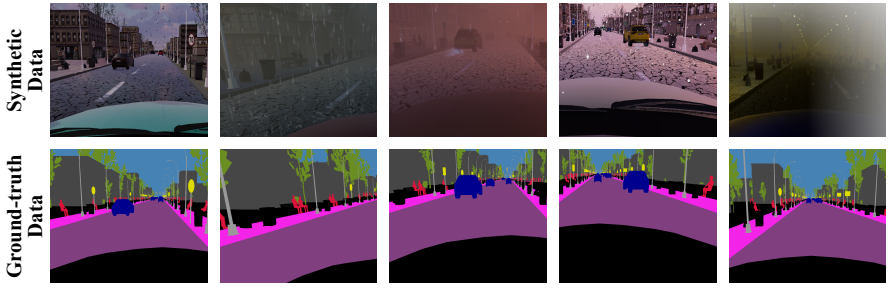

Figure 2: **Samples from AWSS dataset.** Our generated AWSS synthetic dataset spans normal, rainy, foggy, snowy, and nighttime attributes.

Rendering Pipeline (HDRP). In addition, the simulator applies a set of Procedural Content Generation (PCG) concepts to generate, populate, and control the scenes [19].

i) Adverse conditions. The original simulator can simulate standard and adverse weather conditions at daytime and nighttime but with a limited photo-realism and diversity. For each weather condition, we diversify weather severeness, time-of-the-day, and other scene elements if not specified. Based on the environment being simulated, scene elements materials, shaders and textures are selected from a predefined large set. We customize and integrate Procedural Terrain [43] with Adobe Substance materials [25] to simulate photo-realistic snow accumulation on ground, mud, mold, wet surfaces, and water puddles. Water drops splashes on the ground are simulated by customizing the Unity particle system. Rain splash intensity is controlled by the rain weather severeness which is sampled from a uniform distribution. Additionally, we simulate slightly foggy weather condition once heavy rain is simulated. For nighttime simulation, street lights are turned on and their intensity is randomized. Some of these lights are flickered or turned off to increase diversity.

ii) Dash camera mode. Initially *Silver* simulates Unmanned Aerial Vehicle (UAV) and first-person view cameras. However, most existing semantic segmentation datasets like Cityscapes [9] and ACDC [33] datasets are recorded using a dash camera mounted on a car. To generate our AWSS dataset, we develop the dash camera mode to facilitate this task. Furthermore, to increase view angle diversity, we simulate vertical and horizontal lens shifts.

iii) Semantic segmentation automatic ground-truth. The simulator supports semantic segmentation automatic ground-truth generation. However, the number of semantic classes was limited to 4: humans, ground, buildings, and trees. We extend the number of supported classes by adding new elements to the scene like traffic signs and modify the road mesh into road and sidewalk. At the same time, we customize the ground-truth generation pipeline to match Cityscapes [9] color codes and conventions. With our extension, *Silver* now can provide semantic segmentation ground-truth for 10 classes, as specified earlier in this section.

4 Methodology

We aim to reduce the domain shift in adverse weather conditions while not acquiring additional real data. Hence, we propose a novel training approach that leverages synthetic data, while making the architecture aware of the weather condition and nighttime. Our architecture is trained on both synthetic and real data simultaneously (see Figure 3). Our methodology is based on three components: i) adding two simple networks WAS and TAS that work as

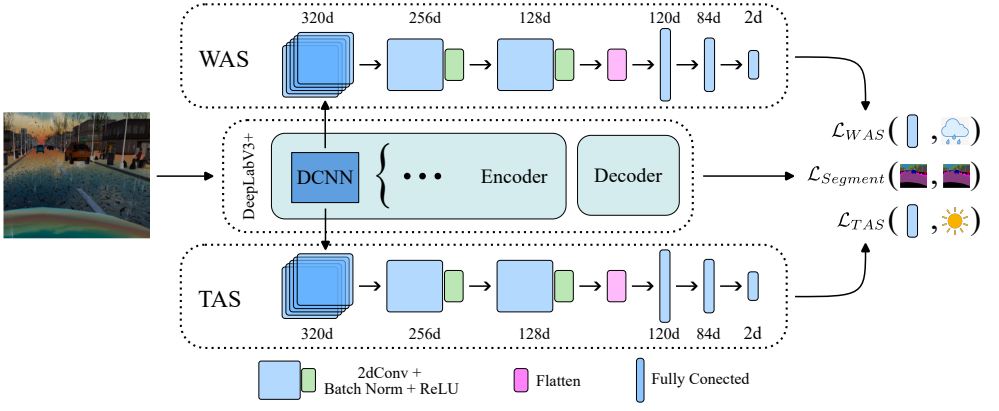

Figure 3: **An overview of our proposed architecture.** DCNN of DeepLabV3+ [5] is forced to learn weather and daytime-nighttime specific and roboust features by the means of multi-task learning. WAS and TAS branches learn to predict weather and daytime-nighttime, respectively. At the same time, they guide the encoder and specifically DCNN to learn extracting robust features under adverse and standard conditions.

supervisors to teach the model to learn weather and nighttime specific features; ii) the full-model is trained using multi-task learning where the baseline learn semantic segmentation and WAS and TAS learns to predict weather condition and day-night, respectively; iii) the model is trained on images from synthetic domain $\mathcal{D}_{adv-synth}$ and real domain $\mathcal{D}_{stand-real}$ in alternating fashion to ensure that the model learn to extract adverse weather features only from synthetic data which presents a proxy of the adverse real domain $\mathcal{D}_{adv-real}$. At the same time, it does not overfit to synthetic data and still ensure that the architecture other components leverage real data. Throughout the paper, $\mathcal{D}_{stand-real}$, $\mathcal{D}_{adv-real}$, and $\mathcal{D}_{adv-synth}$ are represented by Cityscapes [9], ACDC [33], and AWSS datasets, respectively.

Weather and nighttime aware encoder. We use the DeepLabV3+ [5] architecture because of its powerful encoder-decoder architecture. Originally, it is assumed that the encoder will learn how to extract low-level and high-level features independent of weather and illumination conditions. This prevents the model from learning how to extract weather-specific features, resulting in low-quality features being fed to the decoder. The problem becomes even harder without training samples under these conditions.

To alleviate this problem, we focus on the Deep Convolutional Neural Network (DCNN) which is a modified version of Xception [8]. We leverage multi-task learning to enforce the encoder to learn weather and time specific features. We add two simple identical models Weather-Aware-Supervisor (WAS) and Time-Aware-Supervisor (TAS). Each model is composed of two 3×3 atrous 2D convolutions with a rate of 2 and padding of 6. Each convolution is followed by a batch normalization and a rectified linear unit (ReLU). After this, the feature map is flattened and fed to 3 fully connected layers. The last layer predicts the weather for WAS and the daytime-nighttime for TAS. It is worth noting that WAS and TAS are only activated in the training process to guide the feature extraction learning process.

Multi-task learning to improve semantic segmentation. In the original implementation of DeepLabV3+ [5], the output of DCNN is passed to the remaining part of the encoder and to the decoder. In our implementation, we also feed the output of DCNN to WAS and TAS.

Table 2: **mIoU results for our approach Vs. standard domain adaptation methods.** Training our weather and nighttime-aware architecture on both Cityscapes [9] and AWSS, improves the performance on ACDC [33] dataset and achieves adequate peformance on Cityscapes [9]. Best results are **bolded**. Fnt stands for Fine-Tuned.

		ACDC					Cityscapes
		Rain	Fog	Snow	Night	Overall	Overall
DeepLabV3+ [5]	Baseline	0.41	0.46	0.36	0.17	0.35	0.78
	FnT on AWSS	0.44	0.48	0.47	0.19	0.39	0.59
HRNet [28]	Baseline	0.46	0.42	0.41	0.09	0.35	0.75
	FnT on AWSS	0.47	0.49	0.35	0.14	0.36	0.51
DANet [14]	Baseline	0.47	0.57	0.44	0.21	0.42	0.82
	FnT on AWSS	0.48	0.58	0.48	0.26	0.45	0.74
PSPNet [29]	Baseline	0.49	0.54	0.43	0.20	0.41	**0.86**
	FnT on AWSS	0.52	0.56	0.46	0.18	0.43	0.86
Ours	Full-Model	**0.57**	**0.60**	**0.50**	**0.27**	**0.49**	0.75

The total objective to train the new architecture is defined as:

$$\min_{\theta} \mathcal{L} = \mathcal{L}_{Segment} + \alpha \times \mathcal{L}_{WAS} + \beta \times \mathcal{L}_{TAS}, \tag{1}$$

where $\mathcal{L}_{Segment}$ is the original loss used to train DeepLabV3+ [5], $\mathcal{L}_{WAS}$ and $\mathcal{L}_{TAS}$ are the cross-entropy losses utilized to train WAS and TAS, respectively. α and β are scalars to ensure numerical stability during the training and to give more emphasis to the main loss, i.e., $\mathcal{L}_{Segment}$. It should be noted that each loss is back-propagated separately. $\mathcal{L}_{Segment}$ is back-propagated over all the architecture except WAS and TAS. On the other hand, $\mathcal{L}_{WAS}$ and $\mathcal{L}_{TAS}$ are back-propagated only to DCNN.

Synthetic-aware training procedure. Training on source domain and fine-tuning on the target domain is a well-known approach to mitigate the domain gap [40]. However, it is not practical as it requires annotated real data from the target domain which may not be always affordable. Furthermore, training the model on data from one distribution and then forcing the model to learn a new distribution limits the ability of the network to learn and may not converge to a global minima.

Thus, we propose training our modified DeepLabV3+ [5] on data from both synthetic and real distributions simultaneously and from scratch. For that aim, we train in alternating fashion: one batch from $\mathcal{D}_{stand-real}$ and next batch from $\mathcal{D}_{adv-synth}$. At the same time, since the aim is to learn how to extract useful features under adverse conditions, we freeze DCNN weights when training on a batch from $\mathcal{D}_{stand-real}$ and update them for a batch from $\mathcal{D}_{adv-synth}$. It is worth noting that all other weights are updated for data from both domains. This strategy encourages the encoder to leverage synthetic data to better learn feature extraction for the target domain while it ensures that the decoder is learning how to interpret both features to perform segmentation task under standard and adverse conditions.

5 Experiments

Datasets. For training experiments, we use two datasets: AWSS dataset and the training split of Cityscapes [9] dataset. For evaluation, we use validation splits of Cityscapes and ACDC [33] datasets. The three datasets follow the same conventions and color codes. Cityscapes comprises 2975 training images and 500 validation images. It is captured in urban scenes under normal weather conditions in the daytime. ACDC validation split comprises 506 images spanning rainy, foggy, snowy weather conditions and nighttime attributes.
Implementation details. Experiments are conducted on a Tesla V100 GPU. For all experiments, we keep the default parameters of the authors. For our adopted DeepLabV3+ architectures, we use a batch size of 4 while we keep all other parameters same as DeepLabV3+. For DeepLabV3+ baseline, our architecture, and all ablation study experiments, we train for 30K iterations. We set $\alpha = \beta = 10^{-5}$, as these values achieved the best results.
Baselines. To analyse the robustness of recent semantic segmentation methods under adverse conditions, we use DeepLabV3+ [6], HRNet [48], DANet [14], and PSPNet [49].
Evaluation metric. We use the common Mean Intersection over Union (mIoU) [6, 14, 48, 49] on the validation sets of Cityscapes and ACDC similar to [6, 26, 45].

5.1 Results

Before discussing our architecture results, we will discuss how the domain shift degrades the state-of-the-art, and the improvements achieved by fine-tuning on synthetic data.
Standard-Adverse domain shift. As shown by our results in Table 2, the performance of recent methods clearly degrade under adverse weather conditions and at nighttime (see rows *Baseline*). Additionally, it seems that snow and nighttime represent a clear challenge for recent models. Snow causes a drastic change in scene appearance: falling snow particles, snow on pavements and other scene elements makes these objects considerably different compared to what the model learned in the training phase. Thus, the model struggles to segment these elements. Similarly, nighttime scenes with the radical decrease in illumination presents a major challenge for segmentation methods.
Domain adaptation using synthetic data. Transfer learning is usually applied to handle a domain shift. However, although it improves the performance on the target domain, it generally degrades the performance on the source domain. As shown in Table 2 (FnT on AWSS), we can improve the performance of each semantic segmentation model. For some attributes like night and snow, the improvement was more than 50% (e.g. HRNet [48] under night). Generally, each semantic segmentation model was able to leverage AWSS to improve its performance for each adverse attribute. However, when evaluating these fine-tuned models on the original domain (Cityscapes), we see a clear degradation in performance. This degradation was more severe for some models like HRNet [48] while it was slight for PSPNet [49]
Weather and night aware architecture. While the previous solution is simple, the improvement on the target domain was limited, and the performance on the source domain was sharply degraded. As a remedy, our architecture based solution achieves the best results on the target domain and it maintains an adequate performance on the source domain. As reported in Table 2, making the model aware of the weather condition and daytime-nighttime attributes of the images in the training phase helps the model to learn how to extract more efficient features under both standard and challenging scenarios. Qualitative results are shown in Figure 4. Furthermore, per-class results are demonstrated in Table 3, our model achieves the best results on *Road*, *Sidewalk*, *Building*, and *Person* semantic classes. The largest improve-

Table 3: **Per-class mIoU results on ACDC [33] dataset.** Our model achieves the best overall results on ACDC [33]. It maintains the best results on *Road*, *Sidewalk*, *Building*, and *Person* classes. Best and second best results are **bolded** and underlined, respectively.

	Road	Sidewalk	Building	Pole	Tr. Light	Tr. Sign	Vegetation	Sky	Person	Car	Overall
DeepLabV3+	0.71	0.22	0.31	0.18	0.22	0.29	**0.72**	0.38	0.24	0.23	0.35
HRNet	0.55	0.16	0.44	0.14	0.28	0.24	0.66	**0.72**	0.07	0.19	0.35
DANet	0.68	0.11	0.19	0.28	**0.54**	**0.67**	0.26	0.65	0.29	**0.53**	0.42
PSPNet	0.63	0.12	0.60	**0.30**	0.48	0.41	0.62	0.61	0.21	0.17	0.41
Ours	**0.79**	**0.40**	**0.63**	0.25	0.26	0.33	0.69	0.66	**0.32**	0.52	**0.49**

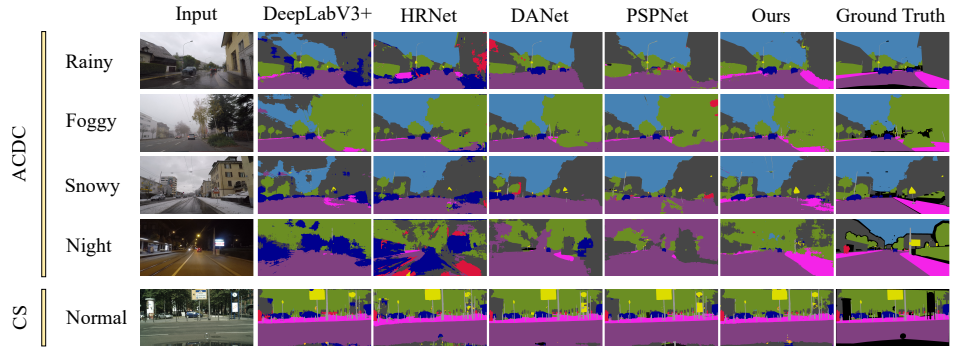

Figure 4: **Visual comparison between baselines and our approach.** Segmentation results are shown on ACDC [33] and Cityscapes [9] dataset, respectively.

ment was on the *Sidewalk* which is around 82% improvement compared to DeepLabV3+, the best performing baseline on this class. As expected, snow and rain changes the visual appearance of this class significantly. This is because of snow accumulation, footsteps on snow, rain splash and mud, in addition to light reflection due to wet surface when raining.

5.2 Ablation Study

To understand the effect of each design decision, we perform several experiments.

Training data type. We train the baseline model on AWSS from scratch (Table 4 first row). As expected, training on synthetic data alone does not achieve satisfactory results due to domain gap between synthetic and real data. Thus, this suggests that AWSS can be used as complementary to the real data and not as an alternative. On the other hand, training the model from scratch on standard weather will perform well on these conditions but will fail under challenging conditions (Table 4 second row).

Training strategy. As shown in Table 4 third row, the standard method of transfer learning (fine-tuning the last layers on the target domain) improves the performance on the target domain but degrades the performance on the source domain.

Weather-Time awareness. Our approach achieves the best results under adverse conditions while still maintaining a satisfactory performance under standard conditions. Making the model synthetic aware and training the model without weather and nighttime-awareness achieve better results on the source domain but low performance on the target domain, compared to fine-tuning experiment. Adding the weather awareness to the model, i.e WAS, improves the performance at standard and adverse conditions. All adverse weather attributes

Table 4: **Ablation analysis of weather and time awareness on performance.** Making the DeepLabV3+ weather and time aware improved the performance significantly at both normal weather, i.e. Cityscapes [9] (CS), and adverse weather, i.e. ACDC [33], scenarios. Best and second best results are **bolded** and underlined, respectively.

	Training Mode	ACDC					Cityscapes
		Rain	Fog	Snow	Night	Overall	Overall
Baseline	scratch on AWSS	0.24	0.25	0.26	0.11	0.22	0.27
	scratch on CS	0.41	0.46	0.36	0.17	0.35	**0.78**
	scratch on CS and fine-tuned on AWSS	0.44	0.48	0.47	0.19	0.39	0.59
Ours	scratch on CS and AWSS	0.41	0.43	0.38	0.19	0.35	0.69
	scratch on CS and AWSS + Weather Aware	0.49	0.55	0.47	0.20	0.43	*0.73*
	scratch on CS and AWSS + Weather and Nighttime Aware	**0.57**	**0.60**	**0.50**	**0.27**	**0.49**	0.75

were improved clearly as expected but the night attribute maintained a slight improvement. Finally, making the model aware of nighttime too, boosts significantly the performance under nighttime. Interestingly, it improves also the performance of the other weather conditions too. This is expected as TAS and WAS teachers allow the model to learn weather specific and nighttime-specific robust features which enables the model to achieve better results under these challenging conditions.

6 Conclusions

We introduce a novel synthetic dataset, the AWSS, that covers various adverse conditions. We show that fine-tuning four state-of-the-art semantic segmentation models improve performance under adverse conditions but degrades the performance under standard conditions. Our proposed solution shows that making the model aware of the synthetic data and utilizing weather-aware-supervisor and time-aware-supervisor achieves the best results under adverse weather conditions while maintaining an adequate performance under standard conditions.

Acknowledgement

This work was funded by the Faculty of Science and Technology of Lancaster University. We thank the High End Computing facility of Lancaster University for the computing resources. The authors would also like to thank CAPES, CNPq, and FAPEMIG for funding different parts of this work.

References

[1] Naif Alshammari, Samet Akcay, and Toby P Breckon. Competitive simplicity for multi-task learning for real-time foggy scene understanding via domain adaptation. *arXiv preprint arXiv:2012.05304*, 2020.

[2] Jose M Alvarez, Theo Gevers, Yann LeCun, and Antonio M Lopez. Road scene segmentation from a single image. In *European Conference on Computer Vision*, pages 376–389. Springer, 2012.

[3] Dawi Karomati Baroroh, Chih-Hsing Chu, and Lihui Wang. Systematic literature review on augmented reality in smart manufacturing: Collaboration between human and computational intelligence. *Journal of Manufacturing Systems*, 61:696–711, 2021.

[4] Gabriel J Brostow, Julien Fauqueur, and Roberto Cipolla. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2): 88–97, 2009.

[5] 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)*, pages 801–818, 2018.

[6] Ping-Rong Chen, Hsueh-Ming Hang, Sheng-Wei Chan, and Jing-Jhih Lin. Dsnet: An efficient cnn for road scene segmentation. *APSIPA Transactions on Signal and Information Processing*, 9, 2020.

[7] Feng-Kuang Chiang, Xiaojing Shang, and Lu Qiao. Augmented reality in vocational training: A systematic review of research and applications. *Computers in Human Behavior*, 129:107125, 2022.

[8] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[9] 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. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[10] Gabriel de Moura Costa, Marcelo Roberto Petry, and António Paulo Moreira. Augmented reality for human–robot collaboration and cooperation in industrial applications: A systematic literature review. *Sensors*, 22(7):2725, 2022.

[11] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017.

[12] Aysegul Dundar, Ming-Yu Liu, Zhiding Yu, Ting-Chun Wang, John Zedlewski, and Jan Kautz. Domain stylization: A fast covariance matching framework towards domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7): 2360–2372, 2020.

[13] Mark Everingham, SM Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111(1):98–136, 2015.

[14] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3146–3154, 2019.

[15] Adrien Gaidon, Qiao Wang, Yohann Cabon, and Eleonora Vig. Virtual Worlds as Proxy for Multi-Object Tracking Analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[16] Manik Goyal, Param Rajpura, Hristo Bojinov, and Ravi Hegde. Dataset augmentation with synthetic images improves semantic segmentation. In *National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics*, pages 348–359. Springer, 2017.

[17] Mengxi Guo, Mingtao Chen, Cong Ma, Yuan Li, Xianfeng Li, and Xiaodong Xie. High-level task-driven single image deraining: Segmentation in rainy days. In *International Conference on Neural Information Processing*, pages 350–362. Springer, 2020.

[18] Maksims Ivanovs, Kaspars Ozols, Artis Dobrajs, and Roberts Kadikis. Improving semantic segmentation of urban scenes for self-driving cars with synthetic images. *Sensors*, 22(6):2252, 2022.

[19] Abdulrahman Kerim, Leandro Soriano Marcolino, and Richard Jiang. Silver: Novel rendering engine for data hungry computer vision models. In *2nd International Workshop on Data Quality Assessment for Machine Learning*, 2021.

[20] Varun Ravi Kumar, Marvin Klingner, Senthil Yogamani, Stefan Milz, Tim Fingscheidt, and Patrick Mader. Syndistnet: Self-supervised monocular fisheye camera distance estimation synergized with semantic segmentation for autonomous driving. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 61–71, 2021.

[21] Yayun Lei, Takanori Emaru, Ankit A Ravankar, Yukinori Kobayashi, and Su Wang. Semantic image segmentation on snow driving scenarios. In *2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 1094–1100. IEEE, 2020.

[22] Beyang Liu, Stephen Gould, and Daphne Koller. Single image depth estimation from predicted semantic labels. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 1253–1260. IEEE, 2010.

[23] Yubin Liu, CB Sivaparthipan, and Achyut Shankar. Human–computer interaction based visual feedback system for augmentative and alternative communication. *International Journal of Speech Technology*, 25(2):305–314, 2022.

[24] Xianzheng Ma, Zhixiang Wang, Yacheng Zhan, Yinqiang Zheng, Zheng Wang, Dengxin Dai, and Chia-Wen Lin. Both style and fog matter: Cumulative domain adaptation for semantic foggy scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18922–18931, 2022.

[25] Adobe Substance Material. https://substance3d.adobe.com/assets, 2022. Online; accessed: 2022-07-26.

[26] Valentina Musat, Ivan Fursa, Paul Newman, Fabio Cuzzolin, and Andrew Bradley. Multi-weather city: Adverse weather stacking for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2906–2915, 2021.

[27] Mobeen Nazar, Muhammad Mansoor Alam, Eiad Yafi, and MS Mazliham. A systematic review of human-computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques. *IEEE Access*, 2021.

[28] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999, 2017.

[29] Fuji Ren and Yanwei Bao. A review on human-computer interaction and intelligent robots. *International Journal of Information Technology & Decision Making*, 19(01): 5–47, 2020.

[30] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for Data: Ground Truth from Computer Games. In *European conference on computer vision*, 2016.

[31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[32] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[33] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10765–10775, 2021.

[34] Swami Sankaranarayanan, Yogesh Balaji, Arpit Jain, Ser Nam Lim, and Rama Chellappa. Learning from synthetic data: Addressing domain shift for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3752–3761, 2018.

[35] Ashutosh Saxena, Sung Chung, and Andrew Ng. Learning depth from single monocular images. *Advances in neural information processing systems*, 18, 2005.

[36] Ashutosh Saxena, Min Sun, and Andrew Y Ng. Make3d: Learning 3d scene structure from a single still image. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):824–840, 2008.

[37] 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.

[38] Marvin Teichmann, Michael Weber, Marius Zoellner, Roberto Cipolla, and Raquel Urtasun. Multinet: Real-time joint semantic reasoning for autonomous driving. In *2018 IEEE intelligent vehicles symposium (IV)*, pages 1013–1020. IEEE, 2018.

[39] Apostolia Tsirikoglou, Joel Kronander, Magnus Wrenninge, and Jonas Unger. Procedural Modeling and Physically Based Rendering for Synthetic Data Generation in Automotive Applications. *arXiv:1710.06270*, 2017.

[40] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017.

[41] Unity. https://unity.com/, 2022. Online; accessed: 2022-07-26.

[42] Yair Wiseman. Autonomous vehicles. In *Research Anthology on Cross-Disciplinary Designs and Applications of Automation*, pages 878–889. IGI Global, 2022.

[43] 4K Procedural Terrain with Rocks and Mold Substance Material (HDRP). https://assetstore.unity.com/, 2022. Online; accessed: 2022-07-26.

[44] Magnus Wrenninge and Jonas Unger. Synscapes: A Photorealistic Synthetic Dataset for Street Scene Parsing. *arXiv:1810.08705*, 2018.

[45] Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, and Xinjing Cheng. Towards fewer annotations: Active learning via region impurity and prediction uncertainty for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8068–8078, 2022.

[46] Qi Xu, Yinan Ma, Jing Wu, Chengnian Long, and Xiaolin Huang. Cdada: A curriculum domain adaptation for nighttime semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2962–2971, 2021.

[47] Yonghao Xu, Bo Du, Lefei Zhang, Qian Zhang, Guoli Wang, and Liangpei Zhang. Self-ensembling attention networks: Addressing domain shift for semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5581–5588, 2019.

[48] Yuhui Yuan, Xiaokang Chen, Xilin Chen, and Jingdong Wang. Segmentation transformer: Object-contextual representations for semantic segmentation. *arXiv preprint arXiv:1909.11065*, 2019.

[49] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[50] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017.