

EE267 Final Project Report (Option 1)

Semantic Segmentation for Autonomous Driving across Diverse Weather Conditions

Hefeifei Jiang, Kunyi Yu

December 7, 2025

Abstract

Semantic segmentation is essential for autonomous vehicle (AV) perception, providing pixel-level understanding of the driving environment. However, adverse weather conditions such as rain, fog, and snow can significantly impair segmentation accuracy, posing challenges to AV safety and reliability. This project investigates the robustness of state-of-the-art semantic segmentation models under diverse weather scenarios using a CARLA-simulated dataset with ground-truth annotations. We benchmark several popular architectures and introduce modifications aimed at improving performance in adverse conditions. Experimental results show consistent gains in challenging weather, demonstrating the effectiveness of our proposed enhancements. The GitHub repository is available here¹.

Keywords: Autonomous Driving, Semantic Segmentation, Diverse Weather Conditions

1 Introduction

Autonomous vehicles (AVs) are rapidly advancing, promising to reform transportation systems. Perception as the foundation of AVs plays a important role in capturing the surrounding environment accurately, so that planning and decision-making modules can make safe and efficient driving behaviors. Semantic segmentation provides pixel-level understanding of the scene, which is critical for AVs to understand the categories of objects and regions (e.g., drivable areas, sidewalks, pedestrians).

However, the accuracy and robustness of semantic segmentation models can be significantly affected by diverse weather conditions, such as rain, fog, and snow. These corner cases introduce huge challenges for AVs' perception systems as well as following planning and decision-making modules. The

performance decline in low-visibility scenarios can lead to higher risks of accidents and mistrust among society when deploying AVs in real-world environments.

In this project, we research the popular and state-of-the-art semantic segmentation models, and evaluate their performance across diverse weather conditions. Besides, we collect different weather perception datasets with ground truth annotations from CARLA simulator [1]. Based on the evaluation results, we modify and improve popular semantic segmentation models to improve their performance in diverse weather conditions. Finally, we conduct ablation studies to analyze the effectiveness of our proposed improvements.

The core contributions of this project are summarized as follows:

- Conduct a comprehensive research on existing semantic segmentation models.
- Collect and preprocess a diverse weather perception dataset with ground truth annotations from CARLA simulator.
- Evaluate the performance of selected models across diverse weather conditions.
- Propose and implement improvements to enhance model robustness.

The rest of the report is organized as follows: Section 2 reviews related work in semantic segmentation for autonomous driving under diverse weather conditions. Section 3 details the methodology, including dataset collection, model selection, and proposed improvements. Section 4 presents the experimental results and analysis. Finally, Section 5 concludes the report and discusses future work.

¹https://github.com/Rock3Yu/EE267_Project_SegScenarioAdaption

2 Related Work

Semantic segmentation has been widely studied in the context of autonomous driving, given its critical role in scene understanding. The models can be simply categorized into five categories: decoder-based methods, context information-based methods, GAN-based methods, RNN-based methods, and weak supervision-based methods [2].

Classic models based on decoder structures include U-Net, SegNet, and DeepLab series, where the DeepLabV3+ reaches 89.0% mIoU on the PASCAL VOC 2012 dataset [3]. Context information-based methods, such as PSPNet, EncNet and DenseASPP, leverage global context or multi-scale context to enhance segmentation performance, where EncNet achieves 85.9% mIoU of small objects on the PASCAL VOC 2012 dataset. In GAN-based methods, models like SegAN and MMANet outperform in the problem of small sample size and imbalanced class distribution but the training process is unstable. RNN-based methods, such as ReNet, ReSeg, LSTM-RNN, and Graph-LSTM, capture local spatial features and their correlations. Lastly, weak supervision-based methods like BoxSup and DEXTR reduce the reliance on time-consuming and labor-intensive data annotation [2].

Datasets for all-weather segmentation are critical for the training and evaluation of many learning-based semantic segmentation models. Conditional-specific datasets provide semantic segmentation data under certain weather conditions, for example, the Foggy Cityscapes dataset [4] simulates foggy conditions based on the Cityscapes dataset, and the Raincouver benchmark [5] focuses on rainy weather. On the other hand, unified all-weather datasets, such as the ACDC dataset [6] contain real-world images captured under four common adverse weather conditions (fog, nighttime, rain, and snow), and the WildDash2 dataset [7] includes dashcam footage from diverse weather, lighting, and camera conditions. The pros of those datasets are that they provide comprehensive coverage of various weather scenarios, however, the cons are the high cost and incomplete scenario coverage.

Data augmentation and Synthetic weather simulation are increasingly studied and utilized to generate diverse weather conditions due to the nature of low cost and high flexibility. Simple data augmentation techniques include random blurring, noise addition, and brightness/contrast adjustments. Moreover, GAN-based weather translation methods, such as CycleGAN [8], can successfully translate images from clear weather to adverse weather conditions by learning the underlying distribution of weather effects. Simulation-based approaches are also popular, where CARLA simulator

[1] is one of the most widely-used platforms that can simulate various weather conditions in provided urban environments with ground truth annotations.

Domain adaption and Curriculum learning are two effective techniques to improve the robustness of semantic segmentation models across diverse weather conditions. Adversarial feature alignment methods, like CyCADA [9], guide transfer between domains in an unsupervised manner. Pseudo-labeling self-training methods contain AdaptSegNet [10]. Those domain adaptation methods can realize sim2real transfer as well as weather adaptation, however the performance is still limited and unstable for AV systems. Meanwhile, curriculum learning is a widely-used training manner that gradually increases the complexity of training samples which can help models learn better and generalize well especially in complex scenarios like diverse weather conditions.

Weather-specific architectures and sensor fusion are also proposed to enhance segmentation performance in the context of extreme weather. Feature-level normalization techniques can mitigate the impact of weather-induced variations. Multi-sensor fusion approaches, combining inputs from cameras, LiDAR, radar, or other sensors, can provide a more comprehensive understanding of the environment, reduce the rise of sensor-specific failures under specific weather conditions, and improve overall segmentation accuracy. Methods like PointPainting [11] demonstrates a promising result.

To summarize, while significant progress and various techniques have been developed for semantic segmentation including the scenarios of autonomous driving under diverse weather conditions, challenges remain in achieving robust, reliable and generalizable performance. The gap between synthetic and real-world data persists. Some methods rely heavily on transformer backbones which are computationally expensive and too slow for real-time AV applications. Moreover, mixed extreme weather conditions are rarely studied. This limitation motivates our project to further explore the robustness of basic semantic segmentation models under diverse weather conditions by using a CARLA-simulated dataset with ground-truth annotations, trying to inspire future research in this important area.

3 Methodology

This section describes the overall methodology of our project, including dataset collection and pre-processing, model selection, training strategies, and evaluation metrics. The proposed training strategies “Mixed-Weather Training” and “Weather-

Aware Curriculum Learning” is one of our core contributions.

3.1 Dataset Collection and Preprocessing

We utilize the CARLA simulator version 0.9.15 to collect a diverse weather perception dataset with ground truth annotations. The simulator uses the provided docker image from EE267 lab0. After launching the CARLA server, we run a Python script to spawn an ego vehicle and other random vehicles, pedestrians with autopilot mode enabled. The camera sensor is set up at a height of 1.6 meters from the ground, with a field of view (FOV) of 90 degrees and a resolution of 512x512 pixels.

We define several weather conditions and time of day settings in CARLA. The weather conditions include *Clear*, *SoftRain*, and *HardRain*, while the time of day settings include *Noon*, *Sunset*, and *Night*, resulting in a total of 9 different weather-time combinations. A demonstration of different weather conditions is shown in Figure 1.



Figure 1: Dataset settings with different weather and time combinations.

The data annotation classes are defined based on the CARLA semantic segmentation labels, where we choose 19 classes (CityScapes Standard): *Road*, *Vehicle*, *Pedestrian*, *Building*, *Sidewalk*, *Vegetation*, etc.

As for data collection, we run the simulation for each weather-time combination, capturing images and their corresponding semantic segmentation masks at a frequency of 1 frame per 10 seconds to 20 seconds, depending on whether the scene is dynamic enough. We collect a total of 3,600 images where each weather-time combination has around 400 images. The dataset is then split into training and testing sets with an 80-20 ratio.

3.2 Model Selection

We select several popular and representative semantic segmentation models for evaluation and further

improvement, including: (1) DeepLab V3, (2) FCN, and (3) Lite R-ASPP. These models are widely used in the studies and as benchmarks for semantic segmentation tasks and their parameters are all pre-trained on a 20-class subset of the COCO dataset.

DeepLab V3 [12] utilizes atrous convolution and spatial pyramid pooling to capture multi-scale context information effectively. FCN [13] is one of the pioneering models that introduced the concept of fully convolutional networks for semantic segmentation, enabling end-to-end training and inference. Lite R-ASPP [14] is a lightweight model designed for real-time applications, employing a reduced Atrous Spatial Pyramid Pooling (ASPP) module to balance accuracy and efficiency.

At the time of writing, we mainly focus on DeepLab V3 for performance evaluation and other model implementations are in our future work plan. The structures overview of DeepLab V3 is shown in Figure 2.

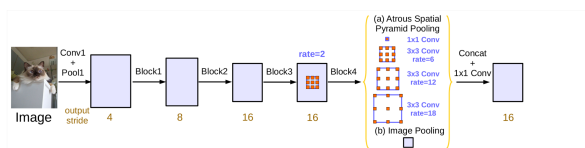


Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

Figure 2: DeepLab V3 structure overview [12].

3.3 Training Strategies

In the training process, we evaluate three different training strategies: (1) Clear-Only Training, (2) Mixed-Weather Training, and (3) Weather-Aware Curriculum Learning. These settings are designed to examine how different levels of weather diversity in training data affect model robustness and generalization.

Clear-Only Training (Setting 0). In this baseline setting, the model is trained exclusively on clear-weather images and evaluated across all weather conditions. This configuration measures the inherent domain gap caused by weather variations, as the model receives no exposure to rain or nighttime conditions during training. Although this strategy is computationally efficient and widely used in autonomous-driving perception benchmarks, it typically results in severe performance degradation when the model is deployed outside clear-weather environments.

Mixed-Weather Training (Setting 1). In this strategy, training images from all available weather conditions (clear, rain, and night) are randomly

mixed within each batch. This simple multi-domain training approach is commonly adopted for improving robustness and serves as a strong reference point. By exposing the model to diverse weather conditions during training, mixed-weather learning mitigates clear-weather bias; however, it does not explicitly address the imbalance among weather domains and may underfit rare or difficult conditions.

Weather-Aware Curriculum Learning (Setting 2). The third strategy adopts a progressive curriculum that gradually increases weather difficulty while balancing the sampling ratios across conditions. Training proceeds in three stages:

- **Stage 1:** Clear weather only.
- **Stage 2:** Clear + SoftRain with a 50:50 sampling ratio.
- **Stage 3:** Clear + SoftRain + HardRain with a sampling ratio of 10:20:70.

This curriculum design allows the model to first acquire stable feature representations under easy conditions before being exposed to progressively more challenging weather. The balanced sampling explicitly counteracts the natural imbalance between clear and adverse-weather images, ensuring sufficient training on rare but safety-critical scenarios such as heavy rain. Overall, this strategy aims to enhance generalization under diverse weather conditions while avoiding the instability or overfitting commonly observed in direct multi-weather training.

3.4 Metric

To be same with the related work, we use mean Intersection over Union (mIoU) as the sole evaluation metric for semantic segmentation performance across different weather conditions. The IoU for each class is calculated as the ratio of the intersection area to the union area between the predicted segmentation and the ground truth segmentation. The mIoU is then computed as the average IoU across all classes. formally, for a given class c , the IoU is defined as:

$$IoU_c = \frac{TP_c}{TP_c + FP_c + FN_c}$$

where TP_c is the number of true positive pixels for class c , FP_c is the number of false positive pixels, and FN_c is the number of false negative pixels. The mIoU is then calculated as:

$$mIoU = \frac{1}{N} \sum_{c=1}^N IoU_c$$

where N is the total number of classes. This metric provides a comprehensive measure of segmentation accuracy, accounting for both false positives and false negatives across all classes. A higher mIoU indicates better segmentation performance. The following Figure 3 illustrates the visual representation of IoU.

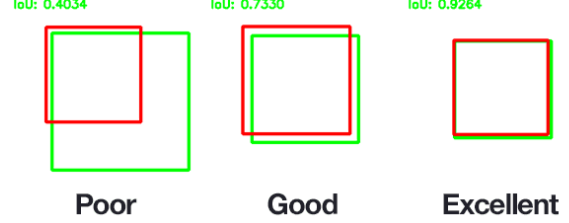


Figure 3: IoU visual representation (image from the internet).

4 Experiments

This section presents the experimental setup, results, and analysis of our proposed methodology.

4.1 Experimental Setup

All experiments are conducted on a machine with the following specifications:

- **GPU:** NVIDIA GeForce RTX 5080 with 16GB VRAM
- **CPU:** Intel Core Ultra 7 265K @ 5.50GHz
- **RAM:** 32GB DDR4
- **Operating System:** Ubuntu 22.04 LTS
- **Python Version:** Python 3.10
- **Framework:** PyTorch 2.7.1 (for CUDA 12.8)
- **CARLA Version:** 0.9.15
- **Epochs:** 100

For more information about the experimental setup, please refer to our GitHub repository (README.md, requirements.txt and other files).

4.2 Results and Analysis

4.3 Ablation Study

5 Conclusion

Acknowledgements

We would like to express our sincere gratitude to Professor Hang Qiu, our teaching assistants Shilpa Mukhopadhyay and all of the EE267 course staff for their invaluable support and guidance throughout this project.

Wishing everyone a great winter break and a happy new year!

References

- [1] 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.
- [2] Shijie Hao, Yuan Zhou, and Yanrong Guo. A brief survey on semantic segmentation with deep learning. *Neurocomputing*, 406:302–321, 2020.
- [3] Mark Everingham and John Winn. The pascal visual object classes challenge 2012 (voc2012) development kit. *Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep*, 8(5):2–5, 2011.
- [4] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018.
- [5] Frederick Tung, Jianhui Chen, Lili Meng, and James J Little. The raincouver scene parsing benchmark for self-driving in adverse weather and at night. *IEEE Robotics and Automation Letters*, 2(4):2188–2193, 2017.
- [6] Christos Sakaridis, Haoran Wang, Ke Li, René Zurbrugg, Arpit Jadon, Wim Abbeels, Daniel Olmeda Reino, Luc Van Gool, and Dengxin Dai. Acdc: The adverse conditions dataset with correspondences for robust semantic driving scene perception. *arXiv preprint arXiv:2104.13395*, 2021.
- [7] Oliver Zendel. *Test data for dependable computer vision applications*. PhD thesis, Technische Universität Wien, 2023.
- [8] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [9] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. Pmlr, 2018.
- [10] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7472–7481, 2018.
- [11] Sourabh Vora, Alex H Lang, Bassam Helou, and Oscar Beijbom. Pointpainting: Sequential fusion for 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4604–4612, 2020.
- [12] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [13] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [14] Sachin Mehta, Mohammad Rastegari, Linda Shapiro, and Hannaneh Hajishirzi. Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9190–9200, 2019.