Will make it Segmenting Success: A Nighttime Driving Environment Deep Learning Experiments

Hong Sujeong
AiFFEL
South Korea
sjhong1007@hanmail.net

September 21, 2023

Abstract

Nighttime driving challenges due to limited visibility and the importance of road illumination. Our research aims to develop a deep learning model for reliable nighttime road recognition, enhancing safety and autonomous vehicle performance. We curate a specialized nighttime driving dataset, using images and radar data for training, applying U-net, SegFormer and GPS-GLASS architectures for better object segmentation. These models are capable of recognizing and segmenting road lanes, moving objects, and other obstacles from CCTV images. They hold great promise for advanced driver-assistance systems, effectively enhancing safety and mitigating accidents

1 Introduction

Driving heavily relies on visual information, and traffic accidents represent a significant societal issue. Nighttime driving, in particular, presents challenges due to limited visibility, compounded by the difficulty of obtaining visual information from CCTV environments. Furthermore, accidents involving two-wheelers often escalate into major incidents, carrying inherent risks.

In light of these circumstances, improving the recognition rate of nighttime CCTV images has become an urgent task, drawing attention to the application of deep learning techniques. This research aims to develop and implement deep learning models to more accurately recognize and interpret visual information in nighttime driving environments.

To achieve this, we have employed state-of-the-art deep learning architectures, including U-Net and SegFormer, both of which excel at image segmentation tasks. U-Net enhances our ability to identify and segment road lanes, moving objects, and other obstacles in CCTV images, providing drivers with more reliable visual information for safe driving.

In addition to U-Net, we have leveraged the SegFormer architecture, which is specifically designed for efficient image segmentation. SegFormer further enhances the accuracy and efficiency of our model, ensuring that it performs optimally in both daytime and nighttime driving scenarios.

These deep learning models are trained using driving data collected in nighttime road conditions, with GPS information aiding in the integration of training data from both daytime and nighttime environments. By combining these advanced architectures and methodologies, we anticipate not only improving the recognition rate of nighttime CCTV images but also promoting safer driving practices and contributing to accident prevention.

Consequently, this research strives to provide drivers with accurate and visually interpretable information, thereby fostering safer driving practices and contributing to accident prevention in both nighttime and daytime driving environments.

2 Methods

2.1 Dataset and Preprocessing

The "Motorcycle Ride Dataset" is comprised of 200 frames sourced from open-access YouTube videos, with the primary purpose of facilitating testing for object detection and mobility-focused AI applications, particularly within the context of computer vision-equipped motorcycle helmets. The dataset encompasses six standard classes: Undrivable, Road, Lanemark, My bike, Rider, and Movable. Here, "Movable" refers to objects in motion, "Undrivable" designates impassable areas, and the remaining classes are self-explanatory, covering road, lane markings (including reflectors), the rider, and the motorcycle itself.

As for data preprocessing, the following steps were applied:

- **Data Split:** The dataset of 200 Input and Mask images was divided into a 75:15:10 ratio for the Train, Validation, and Test sets.
- **Augmentation:** Augmentation techniques were applied to the dataset. These include:
  - **Horizontal Flip** with a 50% probability (only applied to the Train set).
  - **RandomSizedCrop** with a 50% probability (only applied to the Train set).
  - **Resizing** all images to 224x224 pixels (applied to the entire dataset).
  - **Input Image Normalization:** Normalization was applied to the Input images.

These preprocessing steps were implemented to effectively prepare the dataset for training and testing, ensuring compatibility with object detection and AI applications related to mobility and computer vision-equipped motorcycle.

2.2 Models

For our experiment, we have carefully selected two well-researched segmentation architectures [1-3] (one pretrained and one not pretrained) as well as a pretrained model that aligns with the current scenario. These architectures include:

- **U-Net**

- **SegFormer**

- **GPS-GLASS**

These models will play a crucial role in our research, addressing various aspects of image segmentation and object recognition in the context of our project's objectives.

2.3 Metrics

Within our experiments, we utilized two primary evaluation metrics. Firstly, 'Pixel Accuracy (PA)' measures the ratio of correctly predicted pixels to the total number of pixels in the dataset. PA serves as a crucial metric that quantifies the overall accuracy of pixel classification.

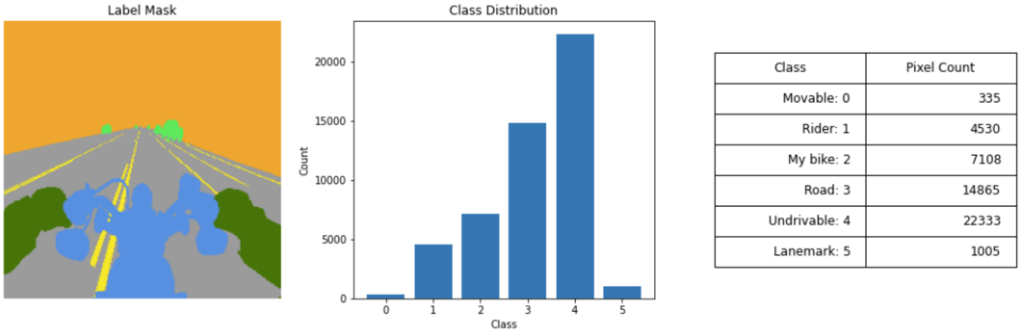
Secondly, 'Mean Intersection over Union (mIOU)' calculates the Intersection over Union (IoU) for each class and computes their mean. mIOU provides a comprehensive evaluation of segmentation accuracy across various classes and serves as an indicator of overall segmentation performance. These two metrics play a pivotal role in quantitatively assessing our experimental results.

3 Results

In order to assess the extent of class imbalance within the images, an analysis of pixel distribution across classes was conducted, and the results are visually depicted in Figure 1. This analysis revealed that regions falling under the 'undrivable' class, such as the sky and trees, collectively accounted for approximately 30% of the image content. Conversely, classes critical for road navigation, such as 'lane' and 'movable,' exhibited notably limited representation within the images.

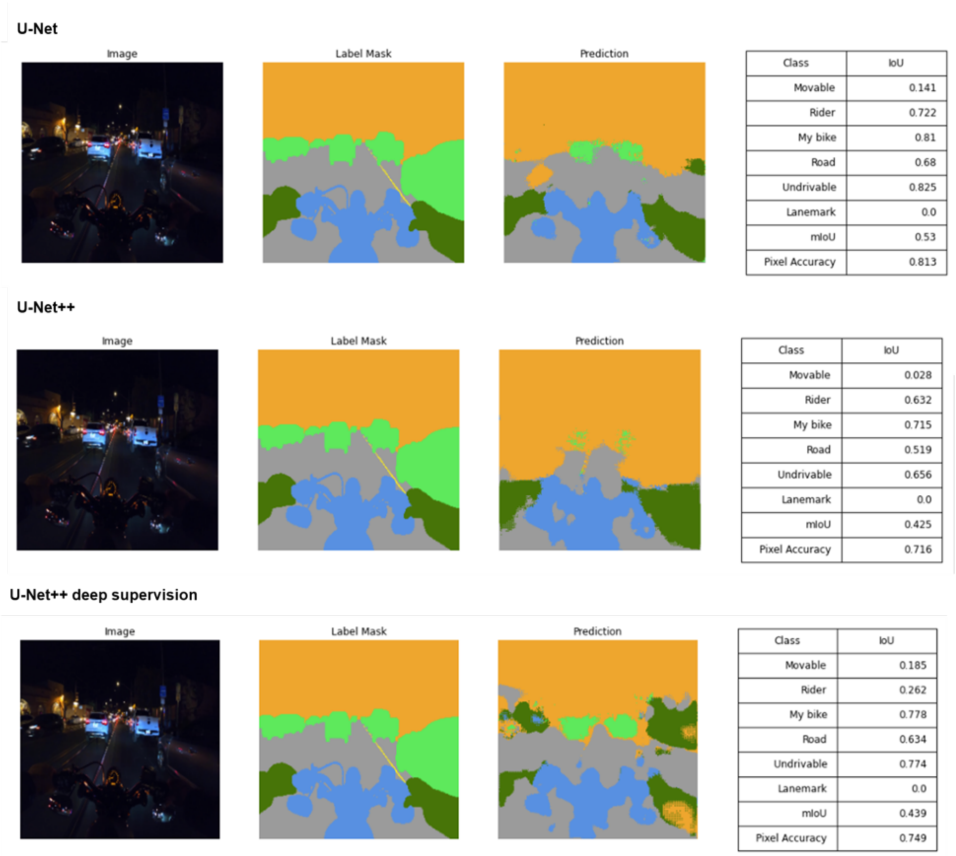
Unet and Its Variants [4]: In this study, we compare Unet , Unet++, and Unet++ with Deep Supervision, three different models designed for image segmentation. Our findings indicate that the basic Unet model outperforms the Unet++ and Unet++ with Deep Supervision models, which utilize images at various resolutions. Additionally, none of the models demonstrated improved performance in lane line detection. Figure 2 provides visual reference to the experimental results.

We conducted the following experiments using SegFormer with pretrained models provided by Hugging Face. We experimented with six different model sizes, and observed that larger model



Class	Pixel Count
Movable: 0	335
Rider: 1	4530
My bike: 2	7108
Road: 3	14865
Undrivable: 4	22333
Lanemark: 5	1005

Figure 1: Distribution of segmentation data classes



Class	IoU
Movable	0.141
Rider	0.722
My bike	0.81
Road	0.68
Undrivable	0.825
Lanemark	0.0
mIoU	0.53
Pixel Accuracy	0.813

Class	IoU
Movable	0.028
Rider	0.632
My bike	0.715
Road	0.519
Undrivable	0.656
Lanemark	0.0
mIoU	0.425
Pixel Accuracy	0.716

Class	IoU
Movable	0.185
Rider	0.262
My bike	0.778
Road	0.634
Undrivable	0.774
Lanemark	0.0
mIoU	0.439
Pixel Accuracy	0.749

Figure 2: U-Net-based Image Segmentation

sizes tend to result in clearer class distinctions. Notably, the largest model, b5, was the only one able to detect lane markings accurately.

Lastly, we also experimented with GPS-GLASS. However, due to differences in the pretrained classes, it struggled to achieve clear class distinctions. We have summarized the performance of these experiments in Table 1, based on various performance metrics.

Model	**mIOU**	**PA**
U-Net	0.53	0.813
U-Net++	0.439	0.749
SegFormer-b0	0.455	0.647
SegFormer-b1	0.442	0.643
SegFormer-b2	0.596	0.686
SegFormer-b3	**0.619**	0.673
SegFormer-b4	0.590	0.724
SegFormer-b5	0.600	0.713
GPS-GLASS	0.2432	0.5778

Table 1: mIOU and PA Scores for Different Models

4 Discussion

We conducted experiments using various models, but encountered challenges in achieving satisfactory results for the segmentation task. Notably, the issue of class imbalance posed a significant hurdle, particularly in failing to recognize the ’load lane,’ which we consider a priority issue to address.

While the results may not directly reflect it, the performance of the model dedicated to isolating the ’load lane’ appeared promising. We believe that implementing this approach could potentially lead to performance improvements. However, given that it involves the use of two models concurrently, there are hardware resource implications. Therefore, we prioritize the adjustment of class weights as a more efficient method to explore, as it holds the potential to yield favorable outcomes.

To address this concern, we propose leveraging a dedicated lane detection model to supplement lane class information within the segmentation process. Additionally, regarding the utilization of pretrained models with different class structures, we anticipate that proper fine-tuning could have resulted in improved segmentation performance. This assumption arises from the understanding that these models have limited recognition capabilities due to their predefined class sets.

In summary, our discussion highlights the potential solutions of incorporating specialized lane detection models and the importance of fine-tuning pretrained models to address class-specific segmentation challenges.

References

[1] Ronneberger, O. & Fischer, P. & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597.

[2] Xie, E. & Wang, W. & Yu, Z. & Anandkumar & Alvarez, J. M. & Luo, P. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. arXiv:2105.15203.

[3] Lee, H. & Han, C. & Jung, S. (2023). GPS-GLASS: Learning Nighttime Semantic Segmentation Using Daytime Video and GPS Data. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops.

[4] Zhou, Z. & Siddiquee, M. M. R. & Tajbakhsh, N. & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. arXiv:1807.10165.