MU-Net: Semantic Segmentation for Motorcycle Night Road with Residual Connection

Junkyu Cho
Aiffel Online 5th
Gyeonggi-do, Republic of Korea
wnsk0427@gmail.com

Abstract

Code is available at https://github.com/wnsk0427/2023-AIFFEL-QUEST/tree/master/MainQUEST04

1 Introduction

Advances in deep learning have led to many advances in self-driving cars. As object detection has become an essential part of self-driving cars, deep learning has become a necessity. However, this is not only limited to cars, but also to other modes of transportation. Examples include drones, ships, and motorcycles.

1.1 Autonomous Driving for Unique Vehicles

Drones are the airborne transportation of the future. Autonomous driving in the air is very different from driving a car. Basically, there are no lanes, signs, or traffic lights, and you're traveling in three dimensions instead of a plane. In particular, obstacles such as birds can appear out of nowhere and require extra attention.

Ships are similar. They don't have lanes and traffic lights, and they have to avoid obstacles like fish and reefs. It's hard to navigate the vastness of the ocean with only a compass to guide you.

The case of motorcycles is a little different. At first glance, the environment is similar to that of a car, but it has a fatal weakness. Speed and direction. On average, motorcycles travel at more than twice the speed of cars. They also turn more freely than cars. This affects the sensors needed for autonomous driving: LiDAR and cameras. The sensors' recognition rates range from a basic 30 frames per second (FPS) up to 60 FPS, because motorcycles travel a lot of distance between those short frames. Eventually, the sensors can't keep up with the speed of the motorcycle, resulting in inaccurate recognition. Add to that the fact that motorcycles are traveling at night and autonomous driving will be even more difficult.

For a motorcycle like this, we need a model that can accurately detect objects even with inaccurate data. Our goal is to perform semantic segmentation to help with object recognition, and we used the **U-Net** model as a backbone with skip connections added to the most basic Autoencoder method. In addition, we used U-Net++ and added a dataset refinement method to suit the task.

1.2 Motorcycle Night Road

The **motorcycle night road** dataset, available on Kaggle, was created by capturing images from a webcam on a motorcyclist's helmet. The environment is mostly nighttime data, so there are many images of dark environments. The dataset is formatted as a COCO dataset. For one scene, there are three images, each representing an origin image, semantic segmentation ground truth, and segmentation mask.

The number of scenes totaled 200, and we split them 8:2 to form a train and test dataset. Validation was done with the test dataset, and evaluation for the final evaluation was also done with the test dataset.

2 Related Works

2.1 ResNet

When CNN first came out for image recognition through deep learning, it was a form of stacking multiple hidden layers and shaping the output with the number of classes of the desired task. A model that exemplifies this is **VGG-16** [3]. However, VGG-16 has the problem of a vanishing gradient as the model gets deeper, so we had to retrieve the information from the previous layer so that the past information could be remembered again and the vanishing gradient could be minimized no matter how many layers were built. This technique of retrieving and using information from the previous layer is called **residual connection**, and the model that applied it is **ResNet** [1].

Since ResNet's inception, there have been many advances in the field of computer vision, including the segmentation task. The segmentation task requires a relatively deep and heavy model because it performs pixel-by-pixel classification of the image, and there is a lot of information that is lost during convolution, so it is important to be able to organize the network so that there is as little information loss as possible.

2.2 U-Net & U-Net++

In a normal segmentation task, the shape of the auto-encoder is often used because it needs to be restored to the original image shape. However, there was a lot of information loss during the encoding process, which often resulted in poor segmentation results. **U-Net** [2] minimizes the amount of information lost during the encoding process by adding a residual connection between the encoder and decoder in the form of an auto-encoder. As a result, segmentation was performed in small parts without crushing, and another main axis model was born in the segmentation task.

U-Net++ [4] is a model that increases the number of skip connections in U-Net so that skip connections are applied per layer. Since it uses more historical information than U-Net, we expected it to produce better results in segmenting more difficult situations.

2.3 Augmentation

Data augmentation is a method used to broaden the distribution of data when the dataset is small. We used it as an experiment to see if augmentation has a significant impact on performance.

2.4 Histogram

A histogram is a plot of the distribution of brightness values ($0 \sim 255$) for each pixel in an image. For example, in a dark image, the brightness values are skewed toward zero and need to be adjusted to be more evenly distributed. A technique used to do this is histogram normalization. Histogram normalization normalizes the brightness of each pixel in an image so that it is distributed between a maximum and minimum value rather than ($0 \sim 255$). This has the effect of making a dark image look relatively bright.

In the case of the motorcycle night road dataset, there are a lot of dark images because it is a nighttime driving image. You might think that normalization should be applied, but there is a problem. Since it is driving data, most people drive at night with their headlights on, and the lights from streetlights, buildings, etc. are bright, so the maximum and minimum values will be near 0 and 255. In this case, we need to flatten the values so that they are evenly distributed, which is called histogram equalization. By using histogram equalization, you can see how dark images with lights can be made relatively brighter.

Since there are many dark images in this dataset, we ran an experiment to compare the results of training on a dataset with histogram equalization.

3 MU-Net

4 Experiments

5 Results

6 Discussion

References

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 2016 (pp. 770-778).

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing.

[3] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[4] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4 (pp. 3-11)*. Springer International Publishing.

[5] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging, 39*(6), 1856-1867.