Enhancing Road Segmentation with U-Net Architecture for Autonomous Driving

Bong Hong Jun,
School of Engineering and Technology, Sunway University,
Selangor, Malaysia,
22018519@imail.sunway.edu.my

***Abstract*—The rise in the form of self-driving cars clearly points to the fact that there is a need for top-notch computer vision systems to classify roads properly. In this paper, the authors propose an optimized U-Net model which has its application in segmenting road scenes, which are paramount for an autonomous driving system.**

Keywords— Autonomous Driving, Road Segmentation, Computer Vision, U-Net Architecture

I. Introduction

With the recent advancements in self-driving car technology, it has become apparent that appropriate computer vision systems that are capable of precisely interpreting road sceneries are required. Despite the numerous challenges in this field, the assessment of road segmentations is still crucial. Division of the accessible portions of the road area from vehicles, pedestrians, and other nearby structures is known as road segmentation. As a result, this procedure is regarded as a fundamental part of the autonomous vehicle as its efficacy and safety are closely correlated with it.

Most classical image segmentation approaches fail to effectively address the issues associated with the variability and complexity of road scenes which include inadequate lighting, and changeable environmental conditions, and varying objects on roads. This has resulted in the usage of deep learning models, especially CNN since they have enhanced excellent accuracy and flexibility. The primary factor contributing to U-Net architecture's rise in popularity is its superiority in medical image segmentation. Given that it achieved NSD 0.0077 in pixel-level prediction-intensive medical applications like retinal blood vessel segmentation, it is plausible that it might likewise demonstrate strong performance in road segmentation tasks [1].

II. Methodology and Proposed Algorithm

A. Overview

The approach used for this project involves proposing a computer vision system that employs the U-Net model optimized for road segmentation. The U-Net model originated for the biomedical image segmentation problem is efficient for distinguishing edges within images. Using this architecture, the proposed system will be able to delineate road boundaries and obstacles hence helping the autonomy of vehicles on the road.

B. Development Steps

Data Acquisition and Preparation The model was trained and validated using a diverse set of road images and corresponding masks from the 'RoadSegmentationDataset_TrainingData' folder. All images were resized to 512x512 pixels and normalized to ensure pixel values ranged from 0 to 1 before mask application. This normalization is crucial as it helps stabilize numerical computations and accelerates the learning process.

C. Model Design

The U-Net architecture was selected for its effective encoding and decoding pathways, essential for capturing contextual details necessary for precise localization. The decoder has numerous up sampling layers to rebuild the segmentation map from encoded features, while the encoder consists of multiple convolutional blocks followed by max pooling for down sampling. Adjustments made to the standard U-Net architecture include tuning the filter sizes and integrating dropout layers to prevent overfitting.

The Adam optimizer, known for its adaptive learning rate capabilities, was chosen to facilitate efficient model training. Because binary cross-entropy handles segmentation tasks well, it was chosen as the loss function.

D. Validation And Testing

To ensure the model's effectiveness and mitigate overfitting, the dataset was divided into training and testing subsets. Model performance, including accuracy and loss, was meticulously recorded during each training epoch on both subsets.

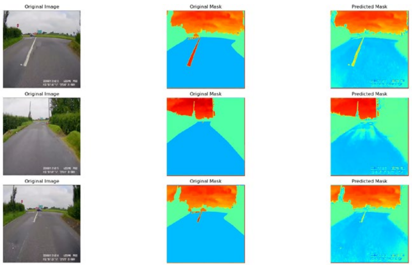
E. Enhancements For Road Scene Adaptation

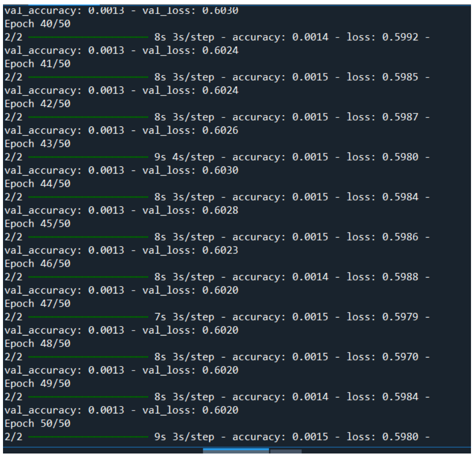
The U-Net model was specifically tailored to enhance its utility in road scene segmentation. To improve the model, changes like the dropout approach and data augmentation tactics were put into practice. These modifications were aimed at addressing the variability and complexities encountered in outdoor scenes, which are typical in road environments. These enhancements help improve the model's resilience and segmentation accuracy, offering substantial benefits for autonomous driving application

III. Results and Discussions

This section delves into the experimental setup and thoroughly analyses the outcomes derived from the process of training a U-Net model tailored for the task of road segmentation.

A. Results





```
val_accuracy: 0.0013 - val_loss: 0.6030
Epoch 40/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 8s 3s/step - accuracy: 0.0014 - loss: 0.5992 -
val_accuracy: 0.0013 - val_loss: 0.6024
Epoch 41/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 8s 3s/step - accuracy: 0.0015 - loss: 0.5985 -
val_accuracy: 0.0013 - val_loss: 0.6024
Epoch 42/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 8s 3s/step - accuracy: 0.0015 - loss: 0.5987 -
val_accuracy: 0.0013 - val_loss: 0.6026
Epoch 43/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 9s 4s/step - accuracy: 0.0015 - loss: 0.5980 -
val_accuracy: 0.0013 - val_loss: 0.6030
Epoch 44/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 8s 3s/step - accuracy: 0.0015 - loss: 0.5984 -
val_accuracy: 0.0013 - val_loss: 0.6028
Epoch 45/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 8s 3s/step - accuracy: 0.0015 - loss: 0.5986 -
val_accuracy: 0.0013 - val_loss: 0.6023
Epoch 46/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 8s 3s/step - accuracy: 0.0014 - loss: 0.5988 -
val_accuracy: 0.0013 - val_loss: 0.6020
Epoch 47/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 7s 3s/step - accuracy: 0.0015 - loss: 0.5979 -
val_accuracy: 0.0013 - val_loss: 0.6020
Epoch 48/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 8s 3s/step - accuracy: 0.0015 - loss: 0.5970 -
val_accuracy: 0.0013 - val_loss: 0.6020
Epoch 49/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 8s 3s/step - accuracy: 0.0014 - loss: 0.5984 -
val_accuracy: 0.0013 - val_loss: 0.6020
Epoch 50/50
2/2 ━━━━━━━━━━━━━━━━━━━━ 9s 3s/step - accuracy: 0.0015 - loss: 0.5980 -
```

Throughout the duration of 50 epochs, the model demonstrated exceedingly low accuracy, consistently hovering around 0.0015. The loss displayed a marginal decrease, commencing at 0.693 and concluding at 0.602 by the end of the training period. This pattern was evident across both the training and validation datasets, underscoring the model's substantial challenges in learning from the provided data.

B. Discussion

The analytical review of the results pinpoints several plausible factors that could be detrimentally affecting the model's performance. Firstly, the low accuracy and elevated loss values could be indicative of either overfitting or underfitting, suggesting that the model may lack the requisite complexity to adeptly capture the intricacies of road segmentation. This could potentially stem from issues related to the way the training data was processed or presented to the model.

Moreover, the architecture of the U-Net model, while generally robust for various segmentation tasks may not have been optimally configured for this specific application. Variables such as filter sizes, the number of layers, and the optimizer's learning rate could have been suboptimal, thereby not aligning well with the unique demands of road scene segmentation.

The quality and quantity of the dataset also warrant consideration. The adequacy of how the images and masks were processed and aligned plays a critical role, as does the diversity within the dataset to encompass a wide array of road types and conditions. These form the basics that are needed in an endeavor to train a model good enough to perform in any given situation.

[1] V. S. Solanki, A. Dewan, H. Singh, A. Kumar, and P. Kaur, "U-Net Based Semantic Segmentation for Road Scene Understanding," presented at the 2023 IEEE International Conference on Smart Technologies, 2023, doi: 10.1109/smart59791.2023.10428246.