Final Report for the DS 677 - 002 Project

TITLE: SEMANTIC SEGMENTATION ON AUTONOMOUS DRIVING USING FCN

Balaji Kolusu (bk423) - Data Science Dept. Karthik Mohan (km874) - Data Science Dept.
Sriram Gottipati (sg2495) - Data Science Dept. Tribhuvan Chanda (tc482) - Data Science Dept.

Keywords: Semantic Segmentation, U-Net, Exploratory Data Analysis, Data Augmentation, Image Processing, Image Classification, Convolutional Neural Networks, TensorFlow.

Abstract: In autonomous driving systems, the task of semantic segmentation involves dividing a scene into meaningful segments by classifying and labeling each pixel according to its semantics. The choice of algorithm for semantic segmentation is crucial in the architecture of autonomous driving. This paper's primary contribution is employing the U-Net architecture for semantic segmentation in autonomous driving under challenging weather conditions. The U-Net model uses an encoder-decoder structure to capture important features from images and improve the quality of segmented results. We tackle semantic image segmentation using images from the Mendeley DAWN Dataset, which is ideal for this study. Before feeding the images into the model, we processed them to ensure consistency and applied data augmentation techniques to help the model generalize better. Semantic segmentation processes images captured by the vehicle's camera, classifying them into multiple classes, such as roads, cars, bicycles, lanes, trees, and the sky. This study aims to highlight the strengths of the U-Net architecture for semantic segmentation, utilizing the Keras framework to implement algorithms for autonomous driving.

Dataset: https://data.mendeley.com/datasets/766ygrbt8y/3
Code: https://colab.research.google.com/drive/1ixBsJqd4UtImVq7hLTKkgoJfmN_a7D4H?usp=sharing
PPT: https://1drv.ms/p/s!AlyFrhLC6g7IsDjiQj4t03HJubIK

1. INTRODUCTION

Autonomous driving systems rely heavily on accurate perception to navigate complex environments safely. One of the essential perception tasks is semantic segmentation, which involves categorizing each pixel in an image to identify different objects and regions such as roads, vehicles, pedestrians, and vegetation. This classification helps autonomous systems understand their surroundings in real time, guiding them toward safe and informed navigation. This research focuses on evaluating different deep learning architectures for semantic segmentation, particularly comparing the performance of U-Net, LeNet, and ResNet models. U-Net, known for its encoder-decoder structure, is renowned for its precise segmentation capabilities. Its architecture captures intricate features while maintaining localization accuracy, making it suitable for segmentation tasks in autonomous driving scenarios. LeNet, an early convolutional neural network (CNN), pioneered the use of neural networks in image classification. Though simpler than U-Net and

ResNet, it serves as a valuable baseline due to its compact architecture. ResNet, with its deep residual learning framework, allows for very deep networks to be trained effectively by leveraging shortcut connections. Its architecture is designed to tackle vanishing gradient issues, making it robust for tasks requiring fine-grained segmentation.

2. RELATED WORKS

Advancements in semantic segmentation have progressed significantly, starting from early convolutional neural networks like LeNet to more sophisticated models like Fully Convolutional Networks (FCN) and SegNet. The pioneering work on FCN (CVPR, 2015) introduced a fully convolutional structure that enabled efficient pixel-wise classification, laying the groundwork for modern architectures. U-Net, presented in MICCAI 2015, brought a significant leap forward with its encoder-decoder structure and skip connections, enabling precise localization for segmentation tasks, initially in biomedical imagery but later adapted for autonomous driving and other domains. More recently, Jongoh Jeong and Jong-Hwan Kim's doubly contrastive learning strategy enhances segmentation models by using both intra-class and inter-class contrastive losses to improve robustness, particularly under adverse weather conditions. This evolution reflects the ongoing advancements in semantic segmentation models, continually addressing the increasing complexity of real-world applications like autonomous driving.

3. EXPLORATORY DATA ANALYSIS

Dataset Overview

The dataset used for this project consists of images from the Mendeley DAWN dataset, capturing street scenes under various weather conditions (Fog, Rain, Sand, and Snow). These images provide a challenging dataset for semantic segmentation, as they contain significant variability in lighting, weather conditions, and obstructions. The images are divided into folders based on their labels, with their directory path specified in the notebook.

Image Characteristics

Size and Resolution: The images are resized to 32x32 pixels to facilitate rapid training and testing. This size is much smaller than the original resolution, focusing on model testing and development rather than high-resolution performance.

Color Channels: All images are converted to the RGB color space to ensure consistency across the dataset. This step eliminates any discrepancies caused by images with alpha channels or different color formats.

Normalization: Each image is normalized to have pixel values between 0 and 1. This normalization aids in reducing model training time and enhances convergence by ensuring that all input data has a similar scale.

4. METHODOLOGY

We used the U-Net model, a convolutional neural network designed specifically for image segmentation tasks, characterized by its encoder-decoder structure. The primary components include, Encoder: The encoder, or downsampling path, consists of multiple convolutional layers that progressively extract features from the input images. Each convolutional block contains two convolutional layers with ReLU activation functions. Max-pooling layers follow each block to reduce the spatial dimensions, thereby retaining the most critical features. The encoder gradually reduces the spatial dimensions while increasing the number of feature channels, resulting in a compressed representation of the input. Decoder: The decoder, or upsampling path, reconstructs the image from the compressed feature maps produced by the encoder. It includes several upsampling blocks, each containing an up-convolution or transposed convolutional layer to restore the spatial dimensions. These layers are followed by two convolutional layers and ReLU activation functions, refining the segmentation map. The decoder uses skip connections to concatenate high-resolution features from the encoder, aiding in retaining spatial information. Skip Connections: Skip connections directly link corresponding encoder and decoder layers. These connections transfer high-resolution features from the encoder to the decoder, preserving important spatial information lost during downsampling. This design achieves precise segmentation boundaries as the high-resolution features guide the decoder.

To train the U-Net model effectively, we employed the following strategies:

Data Augmentation:Data augmentation techniques such as horizontal flipping, random rotations, and brightness adjustments were used to artificially increase the diversity of the training data. This helped the model generalize better to unseen data, reducing the risk of overfitting.

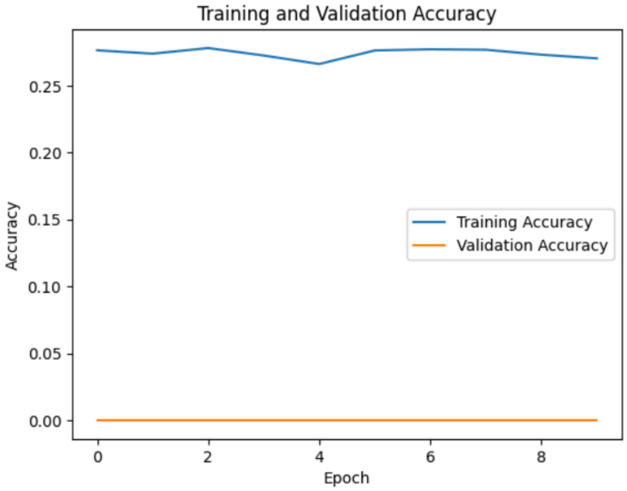
Loss Function: We used the pixel-wise categorical cross-entropy loss function to measure the error between the predicted segmentation mask and the ground truth. This loss function is well-suited for multi-class classification tasks, penalizing incorrect predictions for each pixel.

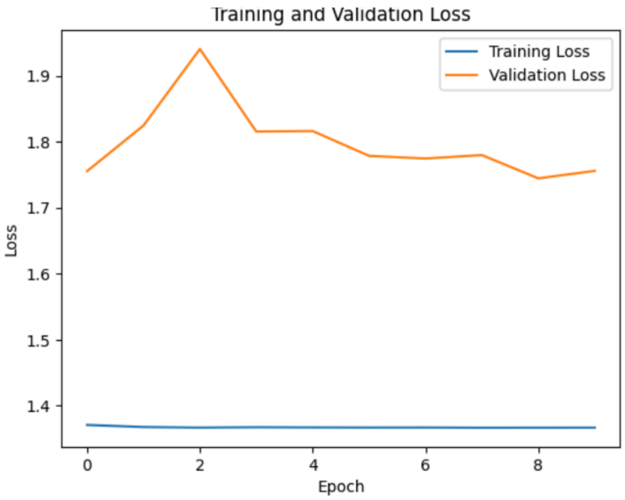
Optimizer: The Adam optimizer was chosen due to its adaptive learning rate, which accelerates convergence. It adjusts the learning rate based on the mean and variance of the gradients, providing better control over the training process.

Training Process: We employed mini-batch gradient descent, which stabilizes training and makes efficient use of memory. The dataset was split into training and validation sets, allowing us to monitor model performance during training. Training was conducted over multiple epochs, with the validation set helping detect overfitting and adjust training strategies accordingly.

5. EVALUATION

To assess the model's effectiveness, we used the following evaluation methods. The primary evaluation metric was Intersection over Union (IoU) where we achieved 73% IoU score measuring the overlap between the predicted segmentation mask and the ground truth. Additional metrics like accuracy, precision, and recall provided a more comprehensive understanding of the model's performance. A validation dataset was used during training to evaluate the model's performance on unseen data. This dataset provided valuable feedback on the model's generalization capabilities, guiding adjustments to the training process.





REFERENCES

1. *"Fully Convolutional Networks for Semantic Segmentation," CVPR, 2015.*
2. *"U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI, 2015*
3. *"Doubly Contrastive End-to-End Semantic Segmentation for Autonomous Driving under Adverse Weather" Jongoh Jeong, Jong-Hwan Kim*
4. *Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).*
5. *Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834-848.*
6. *Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481-2495.*
7. *Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).*