INFOSYS SPRINGBOARD INTERNSHIP 5.0

PROJECT REPORT

Submitted By: N.P.N.V.SAI

Submitted To: Nitig Singh

(Spring Board Mentor)

AI-based Autonomous Driving Model Using Semantic Segmentation

Objective of the Project

This project aims to explore and implement semantic segmentation techniques on the Cityscapes dataset to advance urban scene understanding, which is essential for autonomous driving systems. By categorizing pixels in images into classes such as roads, buildings, and vehicles, the project seeks to enhance image analysis for autonomous driving and smart city planning. The methodology involves loading, preprocessing, and visualizing the data, followed by using semantic segmentation models to classify each pixel accurately. This work aims to provide insights into urban environments and improve machine learning techniques in visual tasks.

Abstract

We present a comprehensive approach to preprocessing the Cityscapes dataset for semantic segmentation tasks relevant to autonomous driving. Key preprocessing techniques include data loading, normalization, and visualization using OpenCV and Python libraries, ensuring data readiness for deep learning applications. Our method effectively prepares images and annotations, improving model training and performance for semantic segmentation. Initial results indicate successful data normalization and provide insights into pixel intensity distribution, setting the foundation for accurate segmentation in urban driving contexts.

Introduction

Autonomous driving has significantly advanced, with accurate image segmentation being essential to detect various elements in the driving environment. The Cityscapes dataset, a benchmark for semantic segmentation, includes high-quality urban scene annotations across diverse settings, making it ideal for developing autonomous driving models. This study emphasizes the importance of preprocessing, including data loading, normalization, and visualization, to ensure clean, structured data for training deep learning models in urban environments.

Literature Review

Preprocessing in semantic segmentation is crucial for data consistency and model accuracy. Research highlights the role of data augmentation and normalization for improved model generalization, with common techniques like pixel intensity normalization and histogram equalization. Studies also stress the importance of data annotation quality and sensor fusion. For example, Ros et al. (2016) found that combining semantic data with depth and thermal imaging enhances accuracy in complex environments. Efficient data loading using OpenCV and other libraries is now standard, enabling faster processing for large datasets like Cityscapes.

Dataset Overview

The Cityscapes dataset consists of high-resolution images of urban environments captured across 50 European cities, annotated for 30 segmentation classes. For this study, we utilized the "gtFine" and "leftImg8bit" packages, focusing on fine annotations to segment semantic classes like roads, vehicles, and pedestrians. The data was preprocessed for a subset of cities in the training set.

Methodology

The project consists of four modules to enable urban environment understanding for autonomous driving:

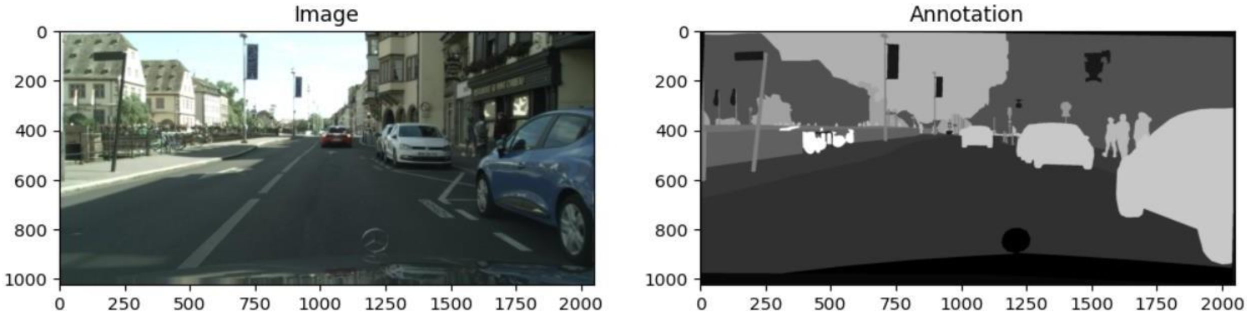
1. **Semantic Segmentation**
    - Goal: Classify each pixel into categories such as roads, vehicles, and pedestrians.
    - Dataset: Cityscapes, ADE20K, CamVid
    - Models: U-Net, DeepLabV3+, Fully Convolutional Networks (FCNs)
    - Libraries: TensorFlow, PyTorch
2. **Object Detection**
    - Goal: Real-time detection of vehicles, pedestrians, and road objects.
    - Datasets: Cityscapes, COCO, KITTI
    - Models: YOLO (fast detection), SSD, Faster R-CNN (high accuracy)
    - Libraries: TensorFlow Object Detection API, PyTorch with pre-trained models
3. **Lane Detection**
    - Goal: Identify and highlight road lane markings.
    - Datasets: Cityscapes, Tusimple Lane Detection, CULane
    - Models: SCNN (Spatial CNN) for deep learning
    - Libraries: OpenCV, PyTorch/TensorFlow
4. **Traffic Sign Recognition**
    - Goal: Classify traffic signs, enabling decisions such as speed limits and stop signs.
    - Dataset: GTSRB (German Traffic Sign Recognition Benchmark)
    - Models: CNN-based models like ResNet and MobileNet
    - Libraries: TensorFlow, PyTorch

Data Exploration and Preprocessing

The data was successfully downloaded and verified for completeness. Key preprocessing steps included:

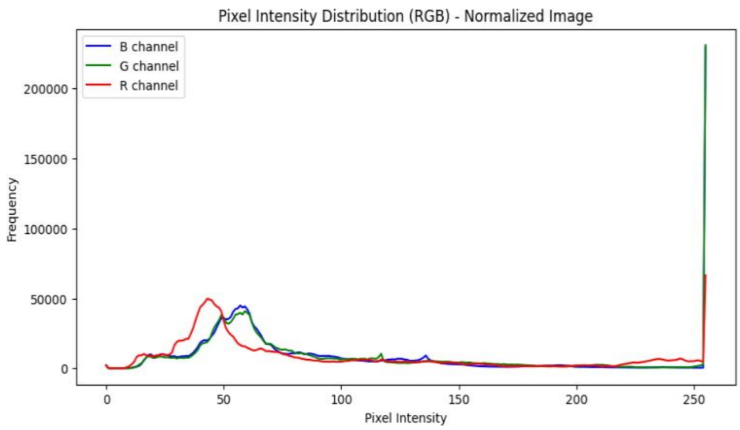
1. **Data Loading**
   Images and corresponding annotations were loaded from the Cityscapes dataset using

OpenCV and glob libraries. Images from the "leftImg8bit" directory and corresponding segmentation masks from the "gtFine" directory were parsed.



2. **Normalization**
   Pixel values were normalized to the [0, 1] range to ensure uniformity, preventing any pixel dominance and facilitating balanced feature treatment during model training.

3. **Visualization**
   Histograms were generated to visualize pixel intensity distributions for each color channel, using OpenCV's calcHist() and visualized with matplotlib. This analysis confirmed uniform intensity across channels, ensuring consistent input data.



4. **Image Segmentation**
   Semantic segmentation, crucial for pixel-wise classification, was applied using deep learning models like U-Net and DeepLab. This enabled contextual understanding of images by classifying each pixel into relevant categories, enhancing the autonomous system's perception of urban scenes.

Results

The preprocessing successfully loaded and normalized the Cityscapes dataset images, with uniform pixel intensity distribution across channels, as confirmed by histograms. All images and masks were aligned correctly, ensuring that each image had its corresponding annotation for accurate training. Additionally, the data exploration verified dataset structure, quality, and dimensional consistency.

Conclusion

The study outlines essential preprocessing steps for the Cityscapes dataset to prepare for autonomous driving applications. Ensuring consistent, clean data enables higher model accuracy and faster convergence during training. The processes described offer a robust framework for future work in semantic segmentation and urban scene understanding, particularly in autonomous vehicle development.

Future Scope

Future work will focus on advanced techniques such as data augmentation, histogram equalization, and contrast enhancement to further improve segmentation performance. Sensor fusion with LiDAR or radar data could be explored to improve accuracy in complex scenarios. Optimizing the preprocessing pipeline for real-time applications on edge devices or embedded systems could advance autonomous driving technology deployment.

References

1. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). "The Cityscapes Dataset for Semantic Urban Scene Understanding." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

2. Geiger, A., Lenz, P., & Urtasun, R. (2012). "Are we ready for autonomous driving? The KITTI vision benchmark suite." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

3. Long, J., Shelhamer, E., & Darrell, T. (2015). "Fully Convolutional Networks for Semantic Segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).