Exploratory Data Analysis (EDA) Report on Cityscapes dataset

Project title: AI-based Autonomous Driving

Prepared by: Shivanand Krishna Bhatte

Date: 16/10/2024

Problem Statement:

The goal of this project is to develop a comprehensive autonomous driving system that can perform semantic segmentation, object detection, lane detection, and traffic sign recognition. Using the Cityscapes dataset, we will train deep learning models to understand and interpret urban street scenes.

Introduction:

The autonomic driving technology development has presented relatively new solutions that appear to have the potential to change the mode of transportation completely. As focused on safety, effectiveness, and ease of use as possible, autonomous driving is designed to eliminate not only accidents but also supposed ‘human errors’ and improve the urban mobility. Autonomous driving can be said to be all about helping cars ‘see’, ‘understand’ and ‘drive’ in difficult situations. For doing this, understanding of road geometry, obstacles, lanes and their boundaries, and traffic signals and signs all in motion is needed.

In order to do so, it is necessary to apply computer vision techniques based on deep learning. Let’s say that semantic segmentation, object detection, lane and sign recognition are the crucial components of autonomous driving.

In this activity, we use Cityscapes dataset which is famous for annotated images containing street views of various cities. In building our case, we will use several deep learning approaches focusing on different aspects of the problem such that they integrate toward an autonomous driving system. The final target of the project is to connect the existing deep learning theories to practical driving applications.

Objectives:

- **Promote Safer and Effective Driving:**

  Design a system that autonomously navigates roads with the focus on the passengers and other vehicles safety. The focus will also be on obeying the road rules including keeping safe speeds, maintaining possible gaps and within lanes.

- **Obtain Accurate Environmental Perception:**

  The system should have its own sensors and deep learning models to perceive the environment with high accuracy. This includes vehicle, people, obstacles, traffic lights and road surface detection and recognition in the real time.

- **Make Decisions in Real Time:**

  The system should be able to make on-the-spot fast and dependable decisions in accordance with the state of the road. This refers to decisions on braking, acceleration, lane changing, more so, shock scenarios such as road blocks, accidents, pedestrians and the like.

- **Enhance the Traffic Efficiency and Minimize the Traffic Jam:**

  Implement Some Algorithms that will enable the system optimize route seeking, adjust traffic, and conserve energy. This will be aimed at minimizing the stagnation and increasing the effectiveness of the transport network as whole.

- **Ensure Fail-Safe in All Environments:**

  Incorporate self-driving technology that is capable of all driving profiles; urban, highways, country roads and under different weather and light conditions eg rain, paris, night driving and snow.

- **Aid the Vehicle's Lateral Navigation:**

  To design a suitable lane-change-resisting system with an effective lane-change system to allow the vehicle to remain in the lanes and make lane-changes whenever the need arises ensuring driving profiles are adhered to even in complex situations.

Dataset Analysis:

The Cityscapes dataset provides comprehensive, high-definition data for understanding scenes in urban contexts. It provides 5000 beautifully labeled pictures spread out across 50 cities taken in varying weather, ambient light and intricate street layouts. In the dataset, there are pixel-wise segmentation of 30 classes which also fall into higher level categories like road, people, cars, road signs and other items which helps in effective training and testing of computer vision algorithms.

The dataset can be categorized into three principal sets:

Training set: 2975 images

Validation set: 500 images

Test set: 1525 images

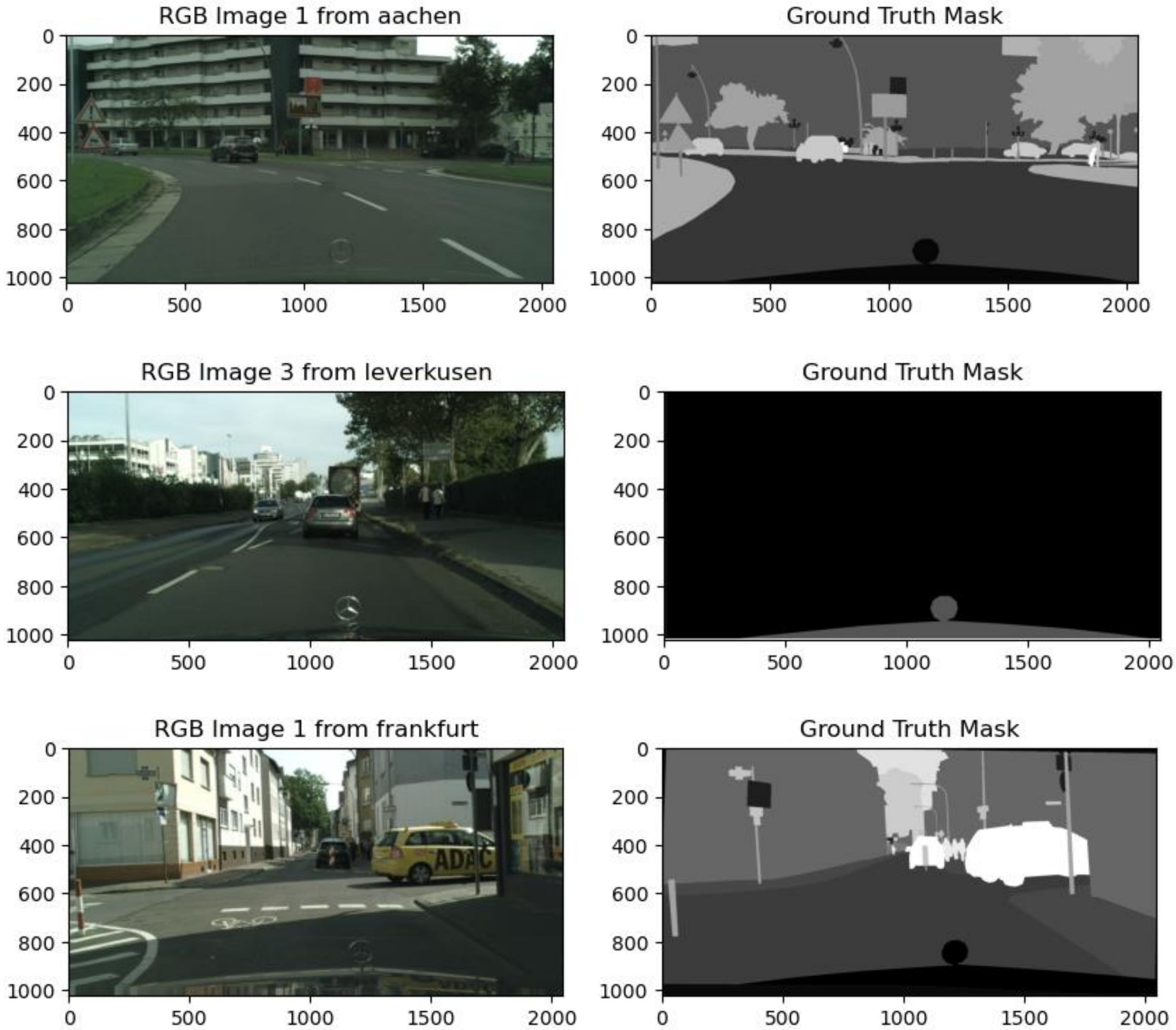
Each image has ground-truth annotations in the form of masks with pixel-level labels, which can be used for performing semantic segmentation tasks. Each of the object classes in the mask is assigned a unique label ID; for instance, ‘0’ corresponds to the background, ‘7’ refers to the road, ‘26’ represents a car and so forth.

The dataset also provides high quality-coloured images (RGB) of the dimensions 2048 by 1024 pixels which are rich in what they portray thus can be utilized in a wide scope of applications such as semantic segmentation, object detection as well as lane detection. Diversity is one of the major issues with the Cityscapes dataset. The images are taken in different seasons, at different times of the day and even under different traffic conditions which brings in some variance that the models have to learn to cope with.

Besides, the Cityscapes dataset is perfect for the application of driving systems, as there is not only the road structure but also the behaviour of pedestrians within it. Austin traffic captured all the components necessary for understanding the class distribution of the above dataset. It will be beneficial in creating strong models that will work in various driving situations.

In next page I shown some dataset samples which contains training, testing and

Validation image samples.



Pre-Processing:

The importance of data pre-processing cannot be stressed enough when it comes to learning deep learning models. This is even more relevant for datasets such as Cityscapes which are large in scale. The pre-processing allows for appropriate data formatting which enhances model training and overall system effectiveness.

Some steps were implemented in our pre-processing pipeline:

Resizing of Images and Masks: Input images and ground truth masks were zoomed into a size of 256 * 256 pixels. This assists in minimizing the computing training cost of the models developed without much degradation in the image quality.

Normalization: The RGB images were rescaled by normalizing each pixel between 0 and 1. This is an important step at the training stage of deep learning models since it allows the model to achieve fast and stable convergence.

Binary Mask Classification: For our project, we worked with the data in its binary format since it implemented a pair-wise classification of masks. Consequently, the multi-class masks that were present in the Cityscapes data set were reworked into polite binary masks where the object of interest (roads vehicles, pedestrians, etc) took the value of 1 and everything else background took the value of 0. This depiction allowed us to eliminate any biases towards any of the classes and simply concentrate on the detection of the class entities within the class regions.

Conclusion:

As a summary, the current undertaking is primarily to utilize deep learning to address the issues involved in self-driving cars. In other words, we are developing models such as semantic segmentation, object recognition, lane estimation, relaying traffic signs and building a system that will be able to understand the driving scenario quite well. To that end, training of these models would be best done using the Cityscapes dataset due to its careful annotations that cover a wide range of urban scenes. The duration of this project will include model performance evaluation, fine tuning as well as appropriate data pre-processing in order to increase the model performance. Last but not least, an integrating system would be constructed which is very close to the fully autonomy-oriented vehicles navigating through mazy cities streets.

References:

1. **Cordts, M., Omran, M., Ramos, S., et al.** "The Cityscapes Dataset for Semantic Urban Scene Understanding." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2016.
2. **He, K., Gkioxari, G., Dollár, P., and Girshick, R.** "Mask R-CNN." *Proceedings of the IEEE International Conference on Computer Vision (ICCV),* 2017.