Analysis and Preparation of the Cityscapes Dataset for Self-Driving Car Applications

Adars T S

October 17, 2024

Abstract

This report presents a comprehensive analysis of the Cityscapes dataset in the context of self-driving car applications. We focus on four key modules: object detection, traffic light detection, lane detection, and semantic segmentation. Our analysis includes data exploration, preprocessing techniques, and dataset preparation strategies. The results provide insights into the dataset's characteristics and its suitability for training models for autonomous driving tasks.

1 Introduction

The Cityscapes dataset is a large-scale dataset for semantic urban scene understanding [1]. It comprises a diverse set of stereo video sequences recorded in street scenes from 50 different cities, primarily in Germany but also in neighboring countries. This dataset is particularly valuable for developing and benchmarking computer vision algorithms for self-driving cars.

In this project, we aim to leverage the Cityscapes dataset for four crucial modules in autonomous driving:

- Object Detection
- Traffic Light Detection
- Lane Detection
- Semantic Segmentation

Our objectives include implementing data loading functions, preprocessing images and annotations, and preparing the dataset for model training.

2 Dataset and Methodology

2.1 Dataset Overview

The Cityscapes dataset includes:

- 5,000 images with high-quality pixel-level annotations
- 20,000 additional images with coarse annotations
- Images from 50 cities, captured during different seasons
- 30 classes for annotations, grouped into 8 categories

2.2 Data Exploration

We performed an extensive analysis of the dataset to understand its characteristics:

2.2.1 Class Distribution

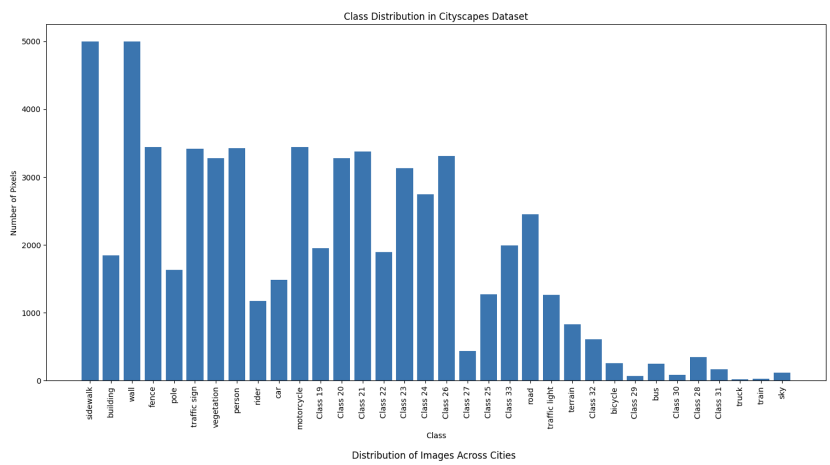

Figure 1: class distribution

The class distribution analysis revealed significant imbalances. Classes such as 'road', 'building', and 'vegetation' are well-represented, while classes like 'traffic light' and 'traffic sign' have fewer instances.

2.2.2 City Distribution

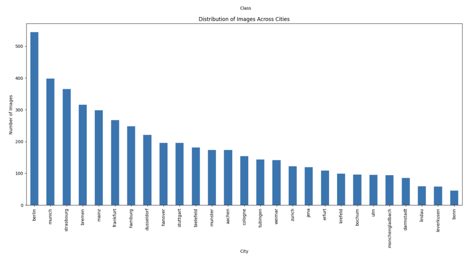

Figure 2: City Distribution

The analysis showed varying representation of different cities, with some cities having significantly more images than others.

2.2.3 Split Distribution

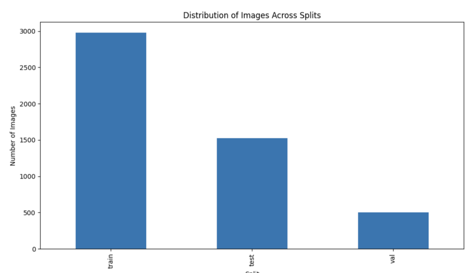

Figure 3: Split Distribution

The dataset is split into training, validation, and test sets, with the majority of images in the training set.

2.2.4 Annotation Quality

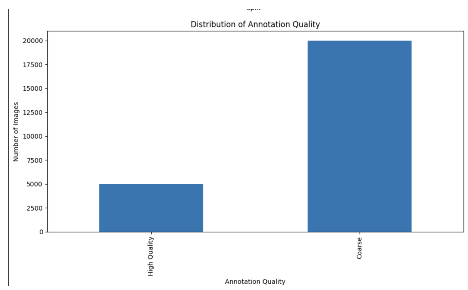

Figure 4: Annotation quality

I observed that while 5,000 images have high-quality pixel-level annotations, the majority (20,000) have coarse annotations.

2.3 Methodology

2.3.1 Data Loading

I implemented Python functions to load images and their corresponding annotations. These functions handle both pixel-level and coarse annotations.

2.3.2 Image Preprocessing

For image preprocessing, we applied the following techniques:

- Resizing to a consistent input size (e.g., 512x512 pixels)
- Normalization of pixel values to the range [0, 1]
- Data augmentation techniques such as random flipping and rotation

2.3.3 Annotation Preprocessing

Annotation preprocessing varied depending on the task:

- For semantic segmentation: One-hot encoding of pixel-level labels
- For object detection: Extraction of bounding boxes and class labels
- For lane detection: Extraction of lane markings and conversion to polylines
- For traffic light detection: Isolation of traffic light instances and their states

2.3.4 Dataset Splitting

We maintained the original Cityscapes split for high-quality annotations:

- Training set: 2,975 images
- Validation set: 500 images
- Test set: 1,525 images

The 20,000 coarsely annotated images were added to the training set for tasks where coarse annotations are sufficient.

3 Results

Our analysis and preprocessing efforts yielded several key findings:

- The class imbalance in the dataset necessitates careful consideration in model training, possibly requiring class weighting or oversampling techniques.
- The diversity of cities represented provides a good basis for generalization, but some cities may be underrepresented.
- The high number of coarsely annotated images provides additional training data, particularly useful for tasks like object detection and semantic segmentation at a coarse level.
- Traffic light and traffic sign classes, crucial for autonomous driving, are underrepresented and may require additional data collection or augmentation.

4 Conclusion

The Cityscapes dataset provides a rich source of data for training models for autonomous driving tasks. Our analysis reveals both strengths and challenges in using this dataset. The high-quality annotations for a subset of images provide excellent ground truth for model training, while the larger set of coarsely annotated images offers opportunities for semi-supervised learning approaches.

The class imbalance and varying representation of different urban environments present challenges that need to be addressed in model development. However, these challenges also reflect real-world conditions, potentially leading to more robust models.

5 Future Objectives

For the next two weeks, we plan to focus on:

- Implementing and fine-tuning models for each of the four tasks (object detection, traffic light detection, lane detection, and semantic segmentation)
- Developing strategies to address class imbalance, such as weighted loss functions or data augmentation techniques
- Exploring transfer learning approaches to leverage pre-trained models on larger datasets
- Implementing a unified data pipeline that can efficiently serve data to all four task-specific models
- Beginning preliminary model training and evaluation on the prepared dataset

References

[1] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3213-3223).

[2] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. International journal of computer vision, 128(2), 261-318.

[3] Arnold, E., Al-Jarrah, O. Y., Dianati, M., Fallah, S., Oxtoby, D., Mouzakitis, A. (2019). A survey on 3d object detection methods for autonomous driving applications. IEEE Transactions on Intelligent Transportation Systems, 20(10), 3782-3795.

[4] Wang, Z., Wu, Y., Niu, Q. (2019). Multi-sensor fusion in automated driving: A survey. IEEE Access, 8, 2847-2868.

[5] Grigorescu, S., Trasnea, B., Cocias, T., Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. Journal of Field Robotics, 37(3), 362-386.

[6] Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Glaeser, C., Timm, F., Wiesbeck, W., Dietmayer, K. (2021). Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. IEEE Transactions on Intelligent Transportation Systems, 22(3), 1341-1360.