**PROJECT REPORT**

On

“AI-based Autonomous Driving Model Project(with Cityscapes Dataset)”

**Submitted To: ----**

**Submitted By:** Srinithi Jaikumar

**OBJECTIVE OF THE PROJECT**

This project aims to work together as AI interns to build an autonomous driving model using the **Cityscapes** dataset. We aim to develop a deep learning model that can accurately detect and classify objects like pedestrians, vehicles, and traffic signs in real time. By focusing on high-precision scene understanding, we aim to improve safety and decision-making for autonomous driving. This project allows us to strengthen our AI skills and contribute to the future of smart transportation.

**Objectives(Assigned Each Week):**

**Week 1:**

**The following are the objectives set for Week 1:**

* **Project Goals and Outcomes:** Students were expected to have a clear understanding of the project's scope and the specific outcomes for the autonomous driving tasks.
* **Cityscapes Dataset Download:** Students were expected to have fully downloaded the dataset.
* **Dataset Exploration:** The visual exploration of sample images and annotations was conducted by some students.
* **Dataset Integrity Verification:** It was expected that the dataset had been verified for completeness and quality.

**Deliverables for Week 1:**

* Dataset fully downloaded and verified.
* Visual exploration of sample images and annotations completed.

**Week 2:**

The objectives for this week were outlined, with deliverables expected by **EOD Wednesday**:

* **Data Loading Functions:** Students are required to implement data loading functions for images and annotations.
* **Image Preprocessing:** Focus on pre-processing images for tasks such as semantic segmentation and object detection.
* **Annotation Preprocessing:** Pre-processing of annotations related to tasks like lane detection and traffic sign recognition.
* **Dataset Splitting & Preparation:** The dataset should be split into training, validation, and test sets.

**Deliverables for Week 2:**

* Data loading functions for images and annotations.
* Pre-processed images and annotations for each task.
* Dataset split into training, validation, and test sets.

**INTRODUCTION**

This project focuses on developing an AI-based autonomous driving model using the **Cityscapes** dataset. The dataset provides pixel-level annotations of urban scenes, enabling precise object detection and semantic segmentation. By applying deep learning techniques, we aim to train a model that can recognize pedestrians, vehicles, and traffic signs in real time, contributing to safe and efficient autonomous driving solutions.

**DATASET AND METHODOLOGY**

**What Is A Cityscapes Dataset?**

* The Cityscapes dataset is a large-scale dataset primarily used for research in **computer vision tasks** such as
* **semantic segmentation**
* **instance segmentation and**
* **object detection.**

**Basic Insights collected :**

* It focuses on urban street scenes, providing pixel-level annotations for **5,000** images of **50** cities across **Germany and neighboring countries**.
* The dataset contains images captured from a **car's front-facing camera**, simulating the **viewpoint** of an autonomous driving system.
* Has Images collected from **different weather conditions**, seasons, and times of day to simulate real-world driving environments.
* It has a resolution of **2048x1024 pixels**, offering fine details.
* **Has Two sets**: 5,000 f**inely annotated** images and 20,000 **coarsely annotated** images for training, validation, and testing purposes.

**Contents of the Dataset(Few are only listed) :**

1 -> gtFine\_trainvaltest.zip (241MB)

2 -> gtCoarse.zip (1.3GB)

3 -> leftImg8bit\_trainvaltest.zip (11GB)

4 -> leftImg8bit\_trainextra.zip (44GB)

5 -> camera\_trainvaltest.zip (2MB)

6 -> camera\_trainextra.zip (8MB)

7 -> vehicle\_trainvaltest.zip (2MB)

8 -> vehicle\_trainextra.zip (7MB)

9 -> leftImg8bit\_demoVideo.zip (6.6GB)

10 -> gtBbox\_cityPersons\_trainval.zip (2.2MB)

**Further Details about the Datasets:**

### **1. gtFine\_trainvaltest.zip (241MB)**

* **Contents**: This zip file contains high-quality annotations for semantic segmentation, instance segmentation, and panoptic segmentation. The gtFine dataset includes pixel-level annotations for various classes in the Cityscapes dataset.
* **Purpose**: Used for fine-grained semantic segmentation tasks where precise annotation is required.

### **2. gtCoarse.zip (1.3GB)**

* **Contents**: Contains coarser annotations compared to gtFine. This may include less detailed segmentation labels suitable for different types of models or experiments.
* **Purpose**: Useful for tasks where high-resolution annotations are not critical, allowing faster training and evaluation.

### **3. leftImg8bit\_trainvaltest.zip (11GB)**

* **Contents**: Contains the training, validation, and testing images at a resolution of 2048x1024 pixels, stored in 8-bit format. The images are the primary input for any machine learning model.
* **Purpose**: Provides the visual data for the models to learn from. This dataset includes a mix of urban scenes with varying conditions.

### **4. leftImg8bit\_trainextra.zip (44GB)**

* **Contents**: Similar to the leftImg8bit\_trainvaltest zip file but contains additional training images to enhance the training dataset.
* **Purpose**: Expands the training set, which helps in better generalization of the models by providing more diverse images.

### **5. camera\_trainvaltest.zip (2MB)**

* **Contents**: Contains camera calibration parameters for the images. This file is typically small as it includes metadata related to the camera settings used for capturing the images.
* **Purpose**: Useful for tasks requiring camera perspective information, such as 3D reconstruction and pose estimation.

### **6. camera\_trainextra.zip (8MB)**

* **Contents**: Contains additional camera calibration parameters for the extra training images.
* **Purpose**: Similar to camera\_trainvaltest, this provides calibration details for an expanded dataset.

### **7. vehicle\_trainvaltest.zip (2MB)**

* **Contents**: This file likely includes annotations or metadata related to vehicles in the dataset.
* **Purpose**: Useful for tasks focused on vehicle detection or segmentation.

### **8. vehicle\_trainextra.zip (7MB)**

* **Contents**: Contains additional vehicle-related annotations or metadata for the extra training images.
* **Purpose**: Expands the dataset for better performance in vehicle-related tasks.

### **9. leftImg8bit\_demoVideo.zip (6.6GB)**

* **Contents**: Includes demo video sequences with images suitable for testing the performance of models in real-time scenarios.
* **Purpose**: Useful for evaluating model performance on video data, focusing on temporal aspects of segmentation.

### **10. gtBbox\_cityPersons\_trainval.zip (2.2MB)**

* **Contents**: Contains bounding box annotations for pedestrian instances in the Cityscapes dataset. This includes information about the location of pedestrians in images.
* **Purpose**: Useful for instance segmentation tasks where bounding boxes help localize objects in images.

**DataSet we will be using:**

### **1.gtFine\_trainvaltest.zip (241MB)**

### **2.leftImg8bit\_trainvaltest.zip (11GB)**

METHADOLOGY

* **Dataset Acquisition**: Download and extract the Cityscapes dataset.
* **Exploration**: Analyze dataset structure and class distribution.
* **Resizing**: Resize large images for memory efficiency.
* **Visualization**: Visualizing the Images along with the labels

RESULTS

**Exploratory Data Analysis (EDA) Report**

**1. Dataset Overview**

* **Image Directory**: leftImg8bit/train/aachen
* **Label Directory**: gtFine/train/aachen
* **Total Number of Images**: 174
* **Total Number of Labels**: 522

**2. Data Structure**

* The dataset consists of images paired with their corresponding label annotations.
* Each image file typically has three related label files:
  + Color annotations (\*\_gtFine\_color.png)
  + Instance ID annotations (\*\_gtFine\_instanceIds.png)
  + Label ID annotations (\*\_gtFine\_labelIds.png)

**3. Basic Statistics**

* The count of images (174) and labels (522) indicates that multiple annotations may exist for certain images. This could suggest different object classes or segmentation tasks within the same image frame.

**4. Visual Exploration**

* **Sample Visualization**: The EDA includes displaying sample images along with their corresponding label annotations. The visualization provides insight into:
  + The diversity of scenes captured in the images.
  + The quality and granularity of the labeling process.

**Example Visualization**:

* (Include an actual sample visualization here if possible)
* The above example showcases five images along with their respective labels, illustrating the segmentation and annotations.

**5. Label Analysis**

* The labels are likely encoded as unique identifiers representing different classes (e.g., buildings, roads, pedestrians).
* **Label File Analysis**: Upon examining the label files, they are categorized into:
  + Color-coded labels for visualization.
  + Instance IDs for identifying distinct objects.
  + Label IDs for classifying object categories.

**6. Label Count Distribution**

* Analyzing the frequency of each label helps identify:
  + The most common object classes.
  + Any imbalances in class representation, which may affect model training (i.e., underrepresented or overrepresented classes).

CONCLUSION

* Figures illustrating example outputs for each model (semantic segmentation, object detection, lane detection, and traffic sign recognition) were included, highlighting the model's effectiveness in real-time scenarios.
* Example images showcase model predictions with overlayed segmentations and bounding boxes, demonstrating clarity and detail recognition.

REFERENCES

* <https://www.cityscapes-dataset.com/wordpress/wp-content/papercite-data/pdf/cordts2016cityscapes.pdf>
* <https://www.ijert.org/research/model-of-autonomous-car-IJERTV9IS050649.pdf>
* <https://medium.com/@gupta.vic17/pre-trained-semantic-segmentation-model-on-cityscapes-for-self-driving-cars-b55524194a62>