**Report : Week 01**

**Infosys Springboard Internship**

**AI-based Autonomous Driving**

**Introduction:**

AI-based autonomous driving is transforming the future of transportation by allowing vehicles to operate without human input. These vehicles, also known as self-driving or driverless cars, utilize advanced systems to perceive their environment, make decisions, and execute driving tasks such as steering, acceleration, and braking. The goal is to improve road safety, reduce human error, and enhance transportation efficiency.

Autonomous vehicles rely on a combination of sensors like cameras, radar, LiDAR, and GPS to collect data about their surroundings. This data is then processed to perform key functions, including semantic segmentation, object detection, lane detection, and traffic sign recognition.

* Semantic Segmentation: This technique divides the road scene into distinct segments, such as roads, sidewalks, vehicles, and pedestrians. By understanding the environment at a pixel level, the vehicle can differentiate between various objects and road types, enabling more accurate decision-making.
* Object Detection: Using data from sensors, the vehicle identifies and tracks objects in its vicinity, such as other vehicles, pedestrians, and obstacles. Object detection is essential for avoiding collisions and safely navigating through dynamic environments.
* Lane Detection: The vehicle identifies lane markings on the road, ensuring it stays within the correct lane. Lane detection helps the vehicle maintain its path, even in challenging conditions like faded or obstructed road markings.
* Traffic Sign Recognition: The system recognizes and interprets road signs, allowing the vehicle to adjust its behaviour accordingly. Whether it’s recognizing a speed limit, stop sign, or a warning signal, this feature ensures compliance with traffic rules.

Autonomous driving systems are developed across levels of automation, from Level 0 (no automation) to Level 5 (full automation). While most current systems operate at Level 2 or 3, the goal is to achieve Level 5, where vehicles can drive independently in any environment.

By incorporating these capabilities, autonomous driving promises safer roads, reduced traffic, and improved accessibility, shaping the future of urban mobility and smart city development.

**Dataset:** Cityscapes Dataset

The Cityscapes dataset is a large-scale, high-quality dataset designed for urban scene understanding, specifically focusing on semantic segmentation, object detection, and other tasks relevant to autonomous driving. Collected from 50 cities in Germany, the dataset includes 5,000 finely annotated images and 20,000 coarsely annotated images, providing diverse scenes with varying weather, lighting, and urban structures. It is particularly valuable for training and evaluating AI models in tasks like semantic segmentation (e.g., roads, sidewalks, vehicles, pedestrians), lane detection, and traffic sign recognition.

Cityscapes is widely used for developing perception algorithms in autonomous driving systems due to its detailed annotations and focus on complex urban environments. The dataset helps researchers and developers enhance object recognition, environment perception, and decision-making processes, making it an essential resource for advancing AI-driven autonomous vehicle technology.

**Dataset Exploration:**

1. Dataset Overview: Begin by loading the dataset and examining its structure. The Cityscapes dataset contains two main sets of annotations: finely annotated images (5,000 images) and coarsely annotated images (20,000 images). These images are taken from 50 different cities, providing a wide range of urban environments.
2. Data Distribution: Analyse the distribution of data based on city, weather conditions, and object types. Check how many images are available for each category, such as roads, vehicles, pedestrians, and traffic signs. This helps in understanding the balance of the dataset and ensuring sufficient representation of all classes.
3. Annotation Types:

* Semantic Segmentation: Explore the segmentation masks, which provide pixel-level labels for various objects in the scene (e.g., road, building, pedestrian, car). Understanding these labels is critical for training segmentation models.
* Instance-level Annotations: Analyse the instance-level annotations, where individual objects like cars and pedestrians are labelled separately. This helps in detecting and tracking multiple objects in a scene.

1. Class Imbalance: Check for class imbalance issues. Certain objects, like roads or vehicles, may be more common in the dataset, while others like cyclists or pedestrians may appear less frequently. Addressing this imbalance is important for effective model training.
2. Visualization: Visualize sample images with their corresponding labels (semantic segmentation masks, bounding boxes). This provides insights into the complexity of scenes and the quality of annotations.
3. Resolution and Image Size: Explore the image resolutions to ensure compatibility with the chosen model architecture. Cityscapes provides high-resolution images (1024x2048), which may need resizing or cropping based on the requirements of the neural network.

My work:

* The exploratory data analysis (EDA) of the Cityscapes dataset begins by loading images using the os libraries to navigate directories.
* A sample image is displayed to understand the type of images in the dataset.
* Image properties, such as size and pixel values, are examined for further analysis.
* The image size distribution is analysed to check if the dataset consists of uniform or varying image dimensions.
* Label images, which contain segmentation masks, are loaded to explore the class distribution.
* Unique class IDs and their pixel counts are analysed to understand the representation of different classes in the dataset.
* A visualization is created by overlaying the segmentation mask onto the original image, providing a clearer view of the labels in real-world context.
* A correlation matrix is computed to examine relationships between the classes in the dataset.
* The matrix helps identify which classes tend to co-occur more frequently.
* The comprehensive EDA provides insights into the dataset structure and distribution, which are important for guiding model training decisions.

In Week 1, we focused on ensuring the successful acquisition and verification of the Cityscapes dataset. This involved downloading the dataset in its entirety and confirming the integrity of the files to guarantee accurate data for subsequent analysis. Additionally, a visual exploration was conducted, allowing for an examination of sample images alongside their corresponding annotations. This step provided valuable insights into the dataset's structure and the quality of the annotations.

**Results:**

* A sample image from the Cityscapes dataset was successfully loaded and displayed.
* Image size distribution showed whether the dataset consists of uniformly sized images or varies across the set.
* The label distribution analysis provided insights into the frequency of different semantic segmentation classes in the dataset.
* Overlaying the segmentation mask on the original image gave a clearer visualization of the dataset’s ground truth.
* The correlation matrix revealed relationships between the classes, showing how often certain classes occur together in the dataset.

**Conclusion:**

The EDA highlights key features of the Cityscapes dataset, such as image size consistency and the class distribution within segmentation labels. By understanding these factors, we can identify potential class imbalances and image size variations that may affect model performance. The correlation between classes provides further insights into class relationships, which could be useful when designing models to handle specific combinations of objects. This analysis sets the foundation for developing and refining machine learning models, ensuring they are appropriately trained with balanced and representative data.

# **Future Objectives (for the next two weeks)**

* Data Loading Functions: Data loading functions facilitate the efficient retrieval and management of image and annotation files from the dataset, ensuring smooth access for subsequent processing steps.
* Image Preprocessing: Image preprocessing involves operations such as resizing, normalization, and edge detection to enhance image quality and prepare it for analysis, ensuring that the data is suitable for model training.
* Annotation Preprocessing: Annotation preprocessing includes converting polygon-based annotations into bounding boxes or segmentation masks, ensuring compatibility with the chosen object detection framework and improving the accuracy of model predictions.
* Dataset Splitting & Preparation: This step might not be necessary because our dataset has already been divided into Train, validation and test. But to receive more precise output we can change the proportion of the already divided dataset.

**References:**

* https://www.researchgate.net/publication/301880609\_The\_Cityscapes\_Dataset\_for\_Semantic\_Urban\_Scene\_Understanding
* https://ieeexplore.ieee.org/document/7780719