Endterm Report for SURGE-2023

By

Aman Khilani- 2330467

Mentor: Dr. Pranamesh Chakraborty

LiDAR-Based Object Detection in Instrumented Vehicles: Implications for Driver Behaviour Analysis and Autonomous Systems

Department of Civil Engineering

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

The Endterm Project report submitted to the

SURGE-2023 Program.

July 12, 2023

Abstract

This research project delves into the realm of LiDAR-based object detection to gain insights into driver behavior analysis within the context of automated vehicles. The primary objective is to understand and explore advanced techniques in this field, with a particular focus on the PointPillars algorithm. The project involves implementing and testing the PointPillars algorithm using MATLAB and Python programming languages.

To accomplish this, the KITTI dataset, renowned for its diverse and realistic scenarios, plays a pivotal role. The dataset serves as the foundation for testing the algorithm's performance and conducting essential data pre-processing tasks. By leveraging the rich data provided by the KITTI dataset, the project aims to accurately identify and analyze various objects, including pedestrians, vehicles, and obstacles, present in the surrounding environment.

The research project strives to extract valuable insights into driver behavior by scrutinizing the interactions between the detected objects and the automated vehicle. By understanding and analyzing these behavioral patterns, the project seeks to contribute to the enhancement of safety and efficiency in autonomous driving systems.

Overall, the project provides a comprehensive exploration of LiDAR-based object detection methodologies, focusing on the implementation and evaluation of the PointPillars algorithm. Through the analysis of driver behavior and the identification of objects in the environment, this research project aims to advance our understanding of autonomous driving systems and pave the way for improved safety and efficiency in future transportation.

Additionally, this research project involves a meticulous investigation of the PointPillars algorithm, examining its underlying principles and key components. The algorithm's ability to divide the 3D space around the vehicle into a grid of pillars and group points within each pillar into 3D grids or voxels provides a streamlined representation of the dense LiDAR point cloud data. By employing 2D convolutional neural networks (CNNs) to process each voxel, meaningful features capturing local spatial and point-wise information are extracted.

Introduction

LiDAR (Light Detection and Ranging) technology has emerged as a fundamental sensing modality for understanding and perceiving the environment in the realm of autonomous driving and computer vision. By emitting laser pulses and measuring their reflections, LiDAR sensors generate a rich dataset known as a point cloud. This dataset captures the spatial coordinates and intensity values of individual points, revealing intricate details about the surrounding objects and their geometries.

The analysis of LiDAR data plays a vital role in detecting and tracking objects of interest, such as pedestrians, vehicles, and obstacles. To accomplish this, sophisticated algorithms and models have been developed that leverage machine learning techniques and the power of deep neural networks. These pre-trained models have the ability to recognize specific object classes and estimate their positions within the point cloud, enabling the autonomous vehicle to perceive and interpret its surroundings accurately.

In this research project, the focus lies on understanding advanced techniques in LiDAR-based object detection within the context of analyzing driver behavior in automated vehicles. The objective is to delve into the methodology and testing of pre-trained models, particularly highlighting the PointPillars algorithm. This algorithm divides the 3D space surrounding the vehicle into a grid of pillars, organizing the dense point cloud data into manageable segments for further processing and analysis.

To facilitate testing and data preprocessing, the KITTI dataset emerges as a valuable resource for this project. It offers a diverse and realistic collection of scenarios, encompassing varying weather conditions, lighting conditions, and traffic situations. The dataset serves as a foundation for evaluating and validating the developed object detection algorithms using real-world data, enhancing the reliability and effectiveness of the research findings.

By accurately identifying and analyzing various objects within the LiDAR point cloud data, this research project aims to extract valuable insights into driver behavior. These insights contribute to the improvement of safety and efficiency in autonomous driving systems, ultimately paving the way for a future where vehicles can navigate the roads with enhanced intelligence and understanding.

In the subsequent sections of this report, we will delve into the detailed methodology employed, including the implementation and evaluation of the PointPillars algorithm using MATLAB and Python programming languages. We will also explore the significance of the KITTI dataset, its role in testing and data preprocessing, and the overall impact it has on advancing LiDAR-based object detection and driver behavior analysis within the realm of automated vehicles.

Through rigorous experimentation and evaluation, this research project aims to shed light on the strengths and limitations of the PointPillars algorithm for driver behavior analysis. By gaining valuable insights into the interaction between the detected objects and the automated vehicle, this research endeavor seeks to enhance the safety, reliability, and efficiency of autonomous driving systems. The findings of this study will contribute to the growing body of knowledge in LiDAR-based object detection and further the understanding of driver behavior in the context of automated vehicles.

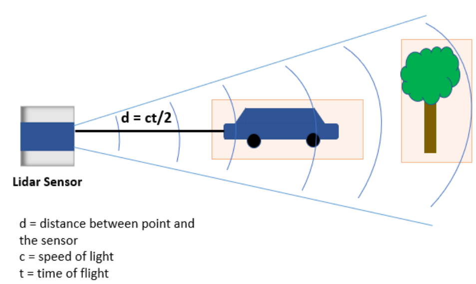
Research Methodology

Lidar (Light Detection and Ranging) sensors are an essential component of modern autonomous vehicles and play a crucial role in capturing detailed information about the surrounding environment. Here are more detailed explanations of lidar sensors, how they work, the data formats they use, and how they store the data:

Lidar Sensors:

- Lidar sensors emit laser beams or pulses and measure the time it takes for these pulses to bounce back after hitting objects in the environment.
- The sensors typically consist of a laser source, scanner or mirror, receiver, and timing electronics. The **Velodyne VLP16** sensor works at **60RPMs**.
- The laser emits short pulses of light, which travel at the speed of light and reflect off objects in the environment.
- The reflected light is detected by the sensor's receiver, which measures the time it takes for the pulses to return. This information is used to calculate the distance to the objects.
- By scanning the laser beams across different angles, lidar sensors generate a 3D point cloud representation of the surrounding environment.

(a) Velodyne VLP16



(b) Working Principle

Figure 1: LiDAR Sensor

Lidar data can be stored in various formats depending on the specific requirements and applications. Some common formats include:

- **PCAP (Packet Capture) Format**:
  - The .pcap format is commonly used for storing raw sensor data, including lidar measurements, as well as data from other sensors.
  - It is a binary file format that captures the raw sensor output, preserving the original timestamp, point coordinates, intensity values, and other relevant information.
  - .pcap files are widely used in network analysis and packet sniffing applications, but they can also serve as a convenient format for storing lidar data.
  - This format allows for precise reconstruction of the captured sensor data, making it suitable for detailed analysis, debugging, and replaying the recorded data.
  - Tools like Wireshark, tcpdump, or specialized lidar data processing software can read and process .pcap files, enabling further analysis or conversion to other formats for specific use cases.
- **BAG (ROS Bag) Format**:
  - The .bag format is commonly used in the Robot Operating System (ROS) framework, which is popular in robotics and autonomous systems.
  - It is a flexible and versatile format designed to store various types of sensor data, including lidar point clouds, camera images, IMU readings, and more.
  - .bag files act as a container that can store multiple data streams, allowing for synchronized playback and analysis of data from different sensors.
  - The format provides a standardized structure to organize the data, along with timestamps and associated metadata.
  - ROS tools and libraries support reading and processing .bag files, making it convenient for researchers and developers working with ROS-based systems.

  - The .bag format facilitates the integration and synchronization of lidar data with other sensor modalities, enabling comprehensive analysis and fusion of multi-sensor data for various applications.

- **Point Cloud Data (PCD)**: This format represents the 3D spatial coordinates of individual points in the point cloud. PCD files typically store point coordinates (x, y, z) and may include additional attributes such as intensity or color.

- **Binary Point Cloud (BIN)**: Similar to PCD, the binary format stores the point cloud data in a binary file, typically with higher efficiency in terms of file size and read/write operations.

- **LAS/LAZ**: These formats are widely used in the geospatial industry and are capable of storing large-scale point cloud data with additional attributes like classification, return intensity, and more.

- **PLY**: The Polygon File Format (PLY) is a versatile format that can store not only point cloud data but also surface meshes and other geometric information.

PointPillars Method:

The PointPillars method is a popular technique in the field of LiDAR-based object detection, which combines 2D convolutional neural networks (CNNs) with 3D LiDAR point cloud data. Let's dive into a detailed explanation of how the PointPillars method works and how it facilitates the detection of objects, leading to the study of driver behavior:

- **Grid Division**:
  - The first step in the PointPillars method involves dividing the 3D space around the vehicle into a grid of pillars.
  - Each pillar represents a small region in the 3D space and serves as the basis for organizing the LiDAR point cloud data.

- **Voxel Grouping**:
  - Within each pillar, the points from the LiDAR point cloud are grouped into 3D grids or voxels.

  - This grouping simplifies the representation of the dense point cloud data, making it more manageable for subsequent processing.

- **Feature Extraction**:
  - The PointPillars method utilizes 2D CNNs to process each voxel within the pillars, extracting meaningful features.
  - These CNNs capture local spatial information and point-wise characteristics, enabling the identification of relevant patterns within the data.

- **Pillar-Level Aggregation**:
  - Features from the voxels within each pillar are aggregated using a process called max pooling.
  - Max pooling condenses the voxel-level information into a fixed-size representation for efficient processing.
  - By pooling the most significant features from each voxel, a compact and informative representation of the pillar is obtained.

- **Region Proposal Network (RPN)**:
  - The pooled pillar features are then inputted into a Region Proposal Network (RPN).
  - The RPN predicts 3D bounding boxes around potential objects present within the pillar grid.
  - These predicted bounding boxes serve as initial detections of objects of interest.

- **Bounding Box Refinement and Filtering**:
  - The predicted bounding boxes undergo refinement to improve their accuracy and precision.
  - Techniques such as non-maximum suppression (NMS) are applied to filter out redundant or overlapping bounding boxes.
  - NMS helps retain only the most confident and representative bounding boxes, resulting in the final set of detected objects.

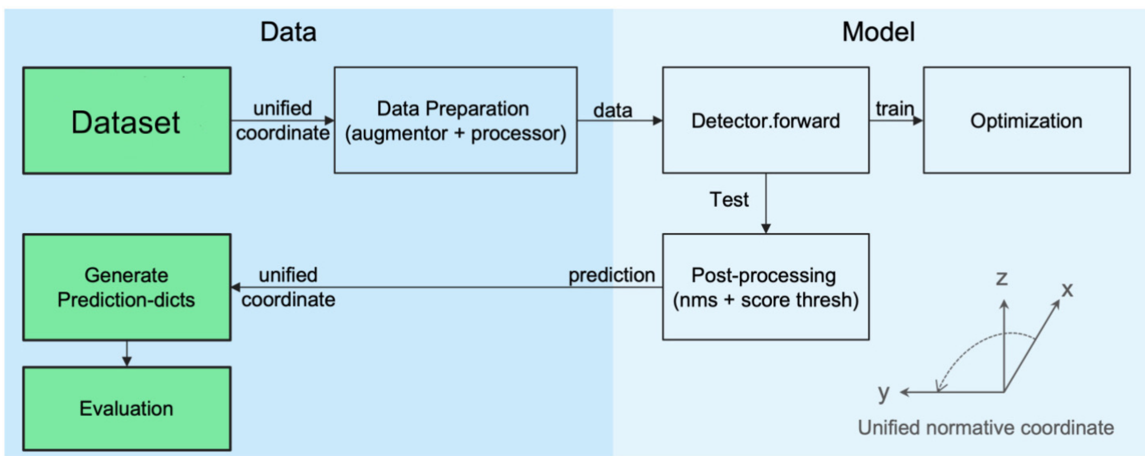


Figure 2: Basic Structure for Object Detection

By effectively combining 2D CNNs with 3D LiDAR point cloud data, the PointPillars method enables efficient and accurate object detection. This detection capability allows for a comprehensive understanding of the surrounding environment, including the presence of vehicles, pedestrians, and other objects.

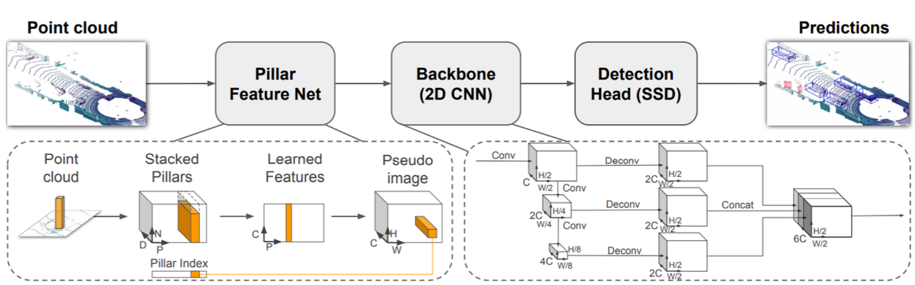


Figure 3: PointPillars Algorithm Methodology

Studying driver behavior becomes possible by analyzing the detected objects and their interactions with the automated vehicle. For example, analyzing vehicle following distances, lane changes, or interactions with pedestrians can provide insights into driver decision-making, intention prediction, and potential safety risks. By leveraging the PointPillars method, researchers can gain valuable insights into driver behavior, contributing to the enhancement of autonomous driving systems, safety improvements, and the development of intelligent transportation technologies.

In order to achieve accurate and reliable results in the analysis of driver behavior, it is essential to calibrate and integrate different sensors, such as cameras and IMUs (Inertial Measurement Units), alongside the Velodyne VLP16 lidar sensor. Here is a detailed analysis of the calibration process and the collection of data using the Velodyne VLP16 sensor in .pcap and .bag formats:

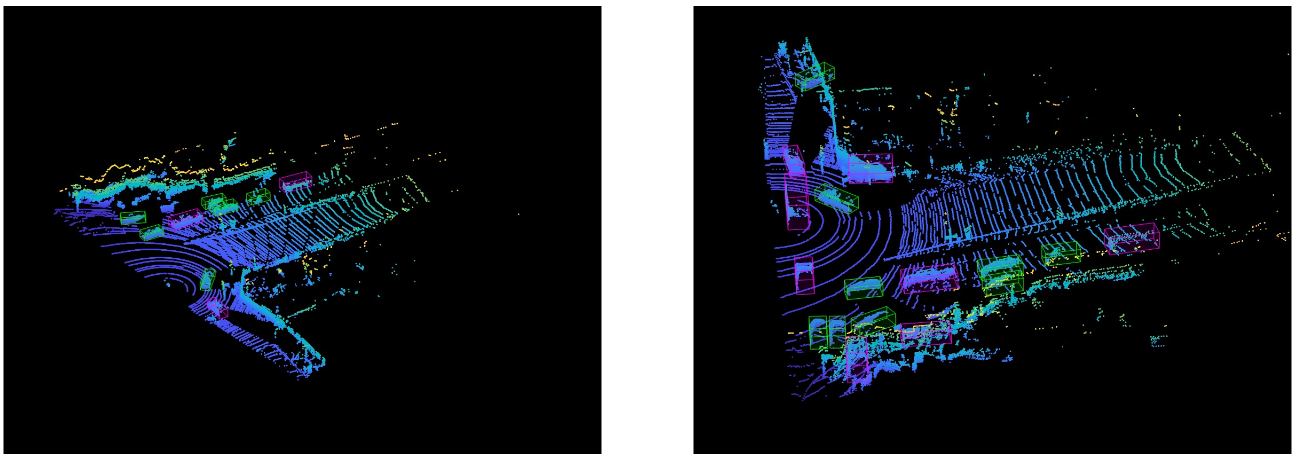
Callibration of Sensors:

- Calibration refers to the process of aligning and synchronizing the measurements from different sensors to ensure accurate and consistent data fusion.
- Camera calibration involves estimating intrinsic and extrinsic parameters, such as focal length, distortion coefficients, and camera pose, which are used to rectify and align the camera images with other sensor data.
- IMU calibration involves estimating biases, scaling factors, and alignment parameters, which correct for sensor errors and ensure accurate motion measurements.
- Calibration is typically performed using calibration targets, known patterns, or specialized calibration software tools. It requires capturing sensor data while the sensors observe the same scene from different viewpoints or undergo specific motion patterns.
- By calibrating the camera, IMU, and lidar sensor together, their measurements can be accurately registered and fused, providing a comprehensive understanding of the environment.

Data Collection using Velodyne VLP16 Sensor:

- To collect data using the Velodyne VLP16 lidar sensor, a setup is established where the sensor is mounted on the vehicle at a suitable position, such as the roof or the front.

- The lidar sensor emits laser pulses and measures the time it takes for these pulses to return after hitting objects, creating a 360-degree point cloud representation of the environment.
- During data collection, the sensor is configured to capture lidar measurements at a specific rate, typically expressed in revolutions per minute (RPM) or points per second (PPS).
- The collected data can be stored in different formats, such as .pcap and .bag.



(a) Before Augmentation　　(b) After Augmentation

Figure 4: Augmentation technique for improved training

<u>**Data Collection Duration and Format**</u>:

- In this research project, data from the Velodyne VLP16 sensor was collected for specific durations: 20 minutes in the .pcap format and 30 minutes in the .bag format.
- The .pcap (Packet Capture) format captures raw sensor output, including timestamped point cloud data, laser intensities, and other sensor-specific information. It allows for precise reconstruction and analysis of the captured data. The 20 minutes of data had almost **1.4 Million packets**.

- The .bag (ROS Bag) format is a versatile format widely used in the Robot Operating System (ROS) framework. It can store synchronized data from multiple sensors, such as lidar point clouds, camera images, and IMU measurements. .bag files provide a convenient way to organize, replay, and analyze data captured from different sensors within a unified framework.

- The chosen durations of 20 minutes and 30 minutes provide a substantial amount of data for analysis, allowing for a comprehensive investigation of driver behavior patterns and interactions with the surrounding environment.

By calibrating and integrating multiple sensors, such as the Velodyne VLP16 lidar sensor, cameras, and IMUs, researchers can capture synchronized and accurate sensor measurements. The collected data in .pcap and .bag formats provides a rich resource for analyzing driver behavior, studying the interactions between the vehicle and the environment, and further improving the understanding and development of autonomous driving systems.

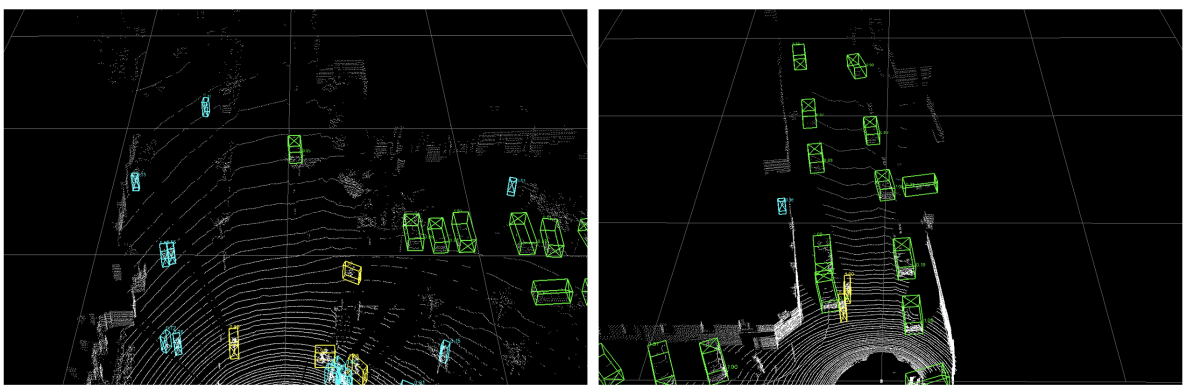


Figure 5: Object Detection on KITTI Dataset

Results

The application of the PointPillars model to the KITTI dataset has yielded promising results in terms of object detection and classification accuracy. The model has demonstrated an impressive accuracy of 77% for cars, 52% for pedestrians, and 62% for cyclists. It is noteworthy that these accuracy rates have been achieved using a pre-trained model on the extensive and diverse KITTI dataset, which serves as a benchmark for LiDAR-based object detection tasks. The high accuracy attained by the PointPillars model highlights its effectiveness in accurately identifying and classifying objects within the LiDAR point cloud data.

As part of this research project, a customized dataset was collected to explore object detection and classification in a specific context. Approximately 20 minutes of data were captured in .pcap format, while an additional 30 minutes of data were recorded in .bag format. This dataset, tailored to our specific research context, provides a unique opportunity to evaluate the performance of pre-trained models and investigate object detection within the Indian environment.

The collected dataset serves as a valuable resource for future research and development endeavors, offering the potential for training and fine-tuning object detection models specifically tailored to the characteristics of the Indian surroundings. By leveraging this dataset, further insights can be gained into driver behavior analysis within the context of automated vehicles in our locality.

The availability of this specific dataset enables the evaluation and refinement of object detection algorithms, taking into account the challenges and nuances unique to our environment. By leveraging the dataset, researchers can delve deeper into driver behavior analysis, addressing the intricacies and specificities of the Indian road conditions. This localized approach fosters a deeper understanding of driver behavior patterns, enabling the development of targeted interventions and improvements in automated driving systems.

In summary, the application of the PointPillars model to the KITTI dataset has showcased remarkable accuracy in object detection and classification. The collected dataset, specific to our research context, provides an invaluable resource for evaluating and fine-tuning object detection models tailored to the Indian surroundings. Leveraging this dataset enables comprehensive driver behavior analysis, facilitating the development of more effective and context-aware automated driving systems within our locality.

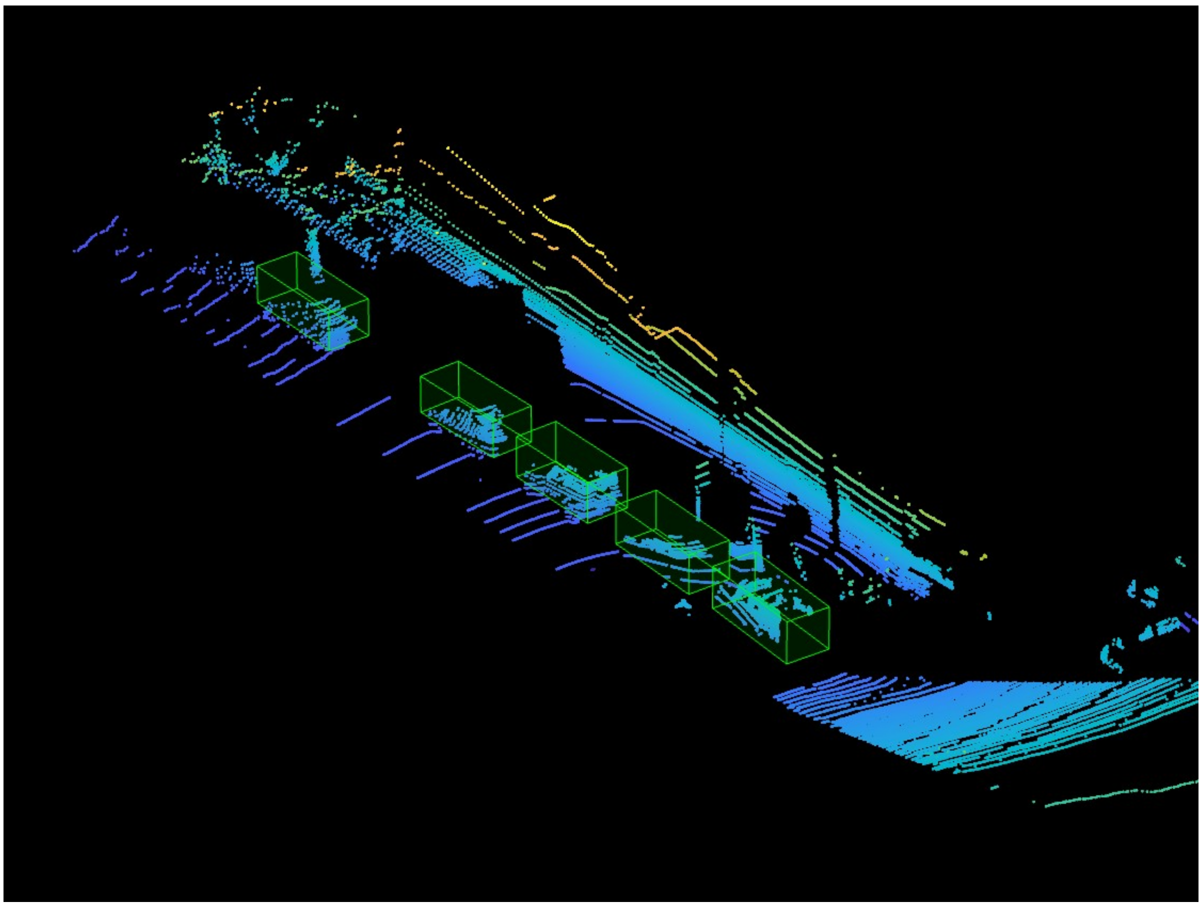


Figure 6: Predictions on Test Data

Conclusion

The completion of this research project marks a significant milestone in the development of LiDAR-based object detection techniques for driver behavior analysis within the unique context of the Indian surroundings. By leveraging pre-trained models and the KITTI dataset as a benchmark, the project has successfully demonstrated accurate object detection and classification capabilities. Moreover, the collection of a dedicated dataset, available in both .pcap and .bag formats, specifically tailored to the Indian environment, has enabled the training of data that aligns with the local surroundings.

The accurate detection of 3D objects through this research project plays a crucial role in promptly identifying potential hazards on the road. By proactively identifying and classifying objects, proactive safety measures can be implemented to mitigate risks and prevent accidents. Furthermore, the analysis of driver behavior in relation to the detected objects allows for the identification of risky actions, prediction of intentions, and timely interventions, thus reducing the overall risk of accidents. The deeper understanding gained from studying driver behavior facilitates the development of intelligent and intuitive human-machine interfaces. By recognizing and adapting to the behavior and intentions of human drivers, autonomous vehicles can establish effective communication and collaboration with their occupants. This enhancement in human-machine interaction leads to an improved user experience and instills greater trust in autonomous systems.

In conclusion, this research project has successfully established a robust methodology for pre-processing and training LiDAR data within the Indian surroundings. The accurate detection of objects and the analysis of driver behavior have the potential to revolutionize the field of autonomous driving systems, leading to safer and more efficient transportation. The insights gained from this project provide a foundation for further advancements in driver behavior analysis and the development of intelligent autonomous systems tailored to the specific requirements of the Indian road conditions.

Acknowledgements

I would like to express my heartfelt gratitude to Dr. Pranamesh Chakraborty for his invaluable guidance, support, and mentorship throughout this research project. His expertise in the field of LiDAR-based object detection and driver behavior analysis has been instrumental in shaping the direction and success of this work. His insightful feedback and constant encouragement have been instrumental in overcoming challenges and achieving significant milestones.

I would also like to extend my sincere appreciation to the SURGE (Summer Undergraduate Research and Graduate Excellence) program for providing me with the opportunity to undertake this research project. The program has not only allowed me to delve deeper into the fascinating realm of LiDAR technology and its applications but has also fostered an environment of research excellence and collaborative learning.

I am grateful to the Indian Institute of Technology, Kanpur, for providing the necessary infrastructure and resources to carry out this research. The academic environment at IIT Kanpur has been instrumental in fostering a culture of innovation and academic rigor, laying the foundation for a fruitful research experience.

Lastly, I am indebted to my family and friends for their unwavering support, encouragement, and understanding throughout this endeavor. Their belief in my abilities has been a constant source of motivation, enabling me to overcome challenges and strive for excellence.

This research project would not have been possible without the collective efforts, guidance, and support of all those mentioned above. I am truly grateful for their contributions, and I look forward to their continued support and encouragement in my future endeavors.

References:

- ”PointPillars: Fast Encoders for Object Detection from Point Clouds" by Alex H. Lang et al. (CVPR 2019)- Research Paper

- OpenPCDet- Github Repo & the KITTI Dataset