Intelligent Roadway Analysis: Automated Object Identification and Scene Description Retrieval for Highway Features

Abhinash Anilkumar[1], Arindam Barman[1], Arohan Hazarika[2], Vamshika Sutar[2]

[1]IIT Kharagpur, India, [2]IIT Bombay, India

UNIVERSITY OF ALBERTA

INTRODUCTION

- The objective of this research is to improve the skills of future autonomous vehicles, crash environment reproduction for improved highway safety scene understanding, big data analysis, maintenance planning, and asset management using advanced deep learning methods.
- The project aims to enhance the performance of the Point Transformer V2 model for semantic segmentation of road infrastructure elements' point cloud data by utilizing the Pointcept and CloudComp 310 libraries.
- Our present research aims to improve semantic segmentation model and integrate a text-based scene description tool that uses LiDAR-acquired data. This task will require the participation of human annotators and the utilization of natural language processing (NLP) tools to accurately depict the objects, structures, and context inside the LiDAR point cloud.
- Pretrained NLP models are utilized to produce informative textual descriptions based on road feature data. This comprehensive method attempts to establish a strong framework for evaluating the intricacy of the driving environment, which is vital for enhancing the safety and effectiveness of autonomous vehicles.

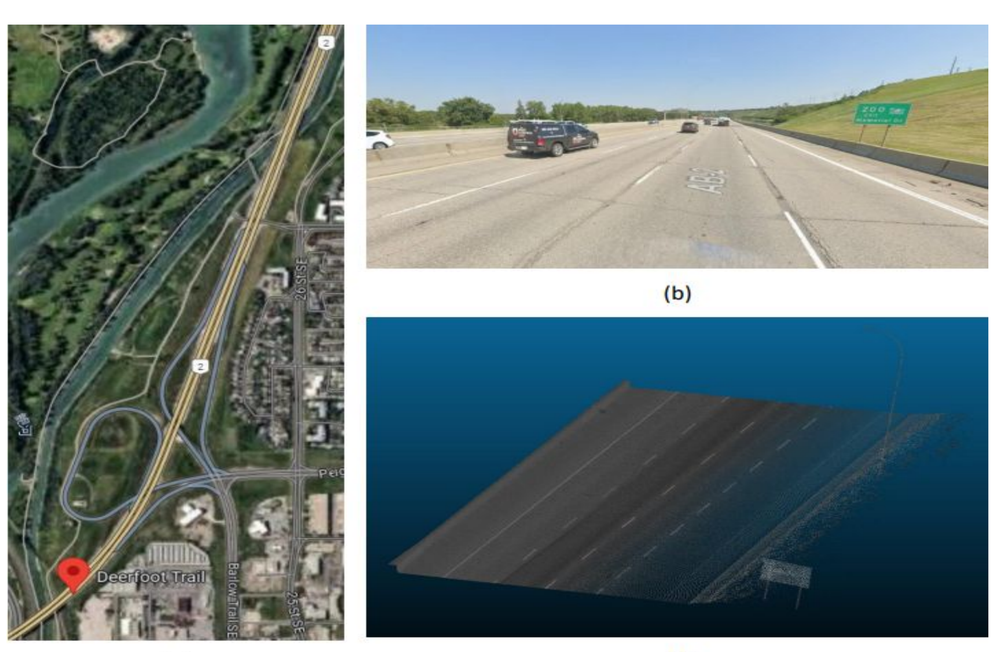


Figure 1. Highway AB-2 (Deerfoot Trail) situated in Alberta, Canada. (a) a satellite image, (b) a street view, and (c) a point cloud sample representing a 50-meter segment

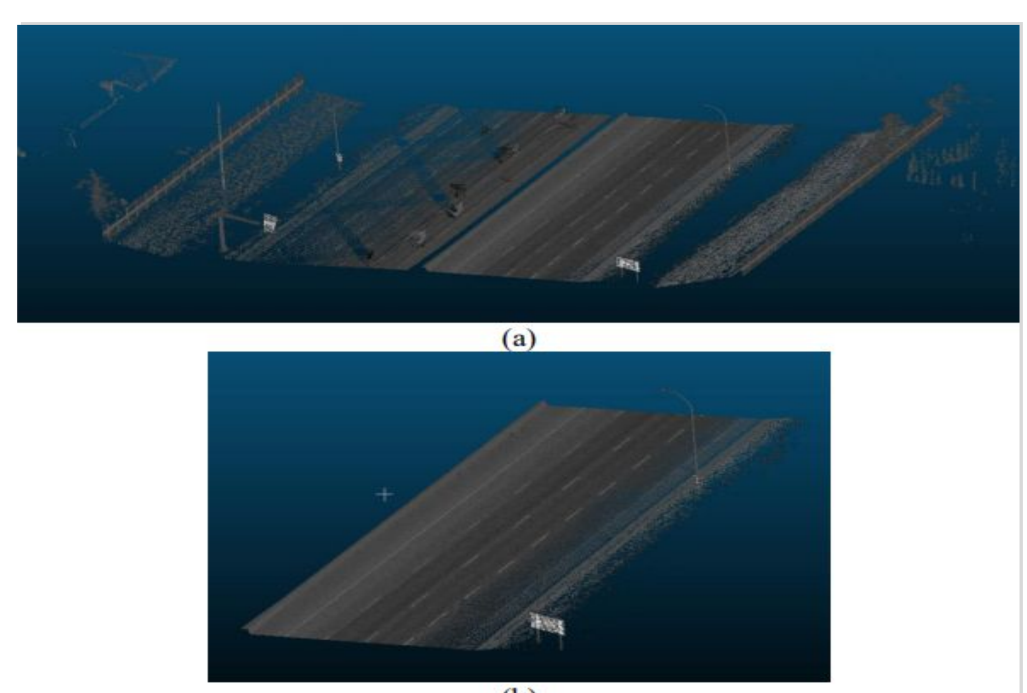


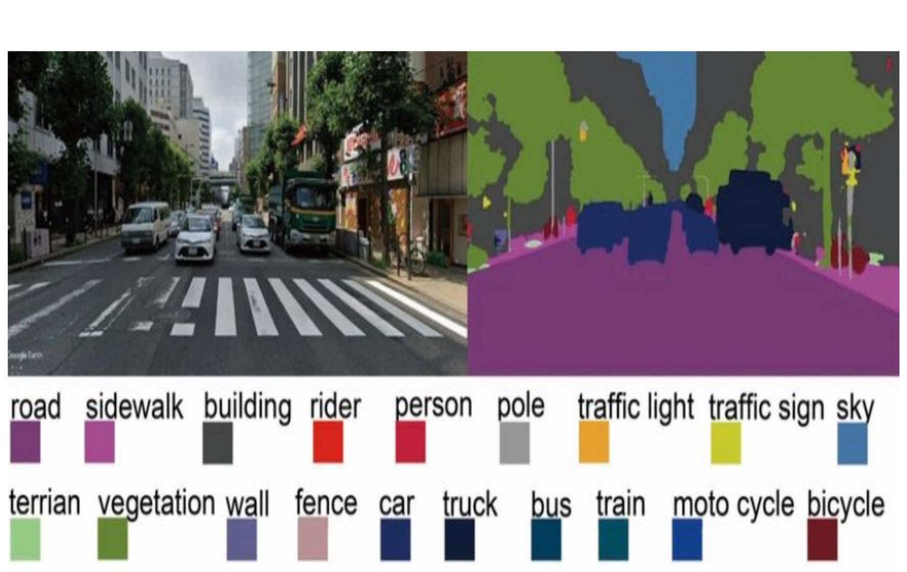
Figure 2. A sample of 50-meter segment raw data (a) before and (b) after data cleaning

SEMANTIC SEGMENTATION

WHY SEMANTIC SEGMENTATION?

Self-driving cars use semantic segmentation to see the world around them and react in real-time. This process separates visual regions into categories like lanes, other cars, and intersections.

To achieve this, we have employed the **Point Transformer V2** model. This advanced model significantly enhances the car's ability to accurately identify and classify various elements in its environment, providing a more detailed and precise understanding of the surrounding road features.



NATURAL LANGUAGE PROCESSING (NLP)

TEXT DESCRIPTION OF HIGHWAY FEATURES

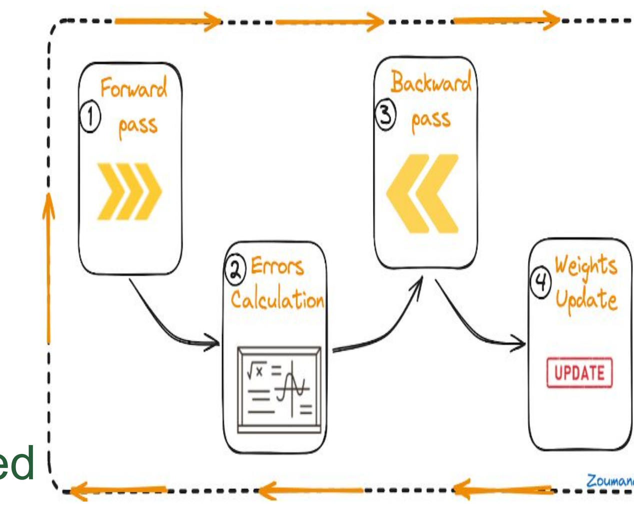
- We propose two methods:
  1.Point cloud data directly to text using clustering method.
  2.Point cloud data is converted to image before generating text description using APIs.
- We are currently using APIs like **Google Vision Pro, BERT**, and **GPT2** to build a text description model for describing highway features.

> "The highway point cloud data depicts a multi-lane highway with a total of four lanes, indicated by three broken lines and two solid lines. A concrete barrier median separates the opposing traffic flows. Additionally, there is a light pole on the side of the road and an overhead traffic sign to guide drivers "

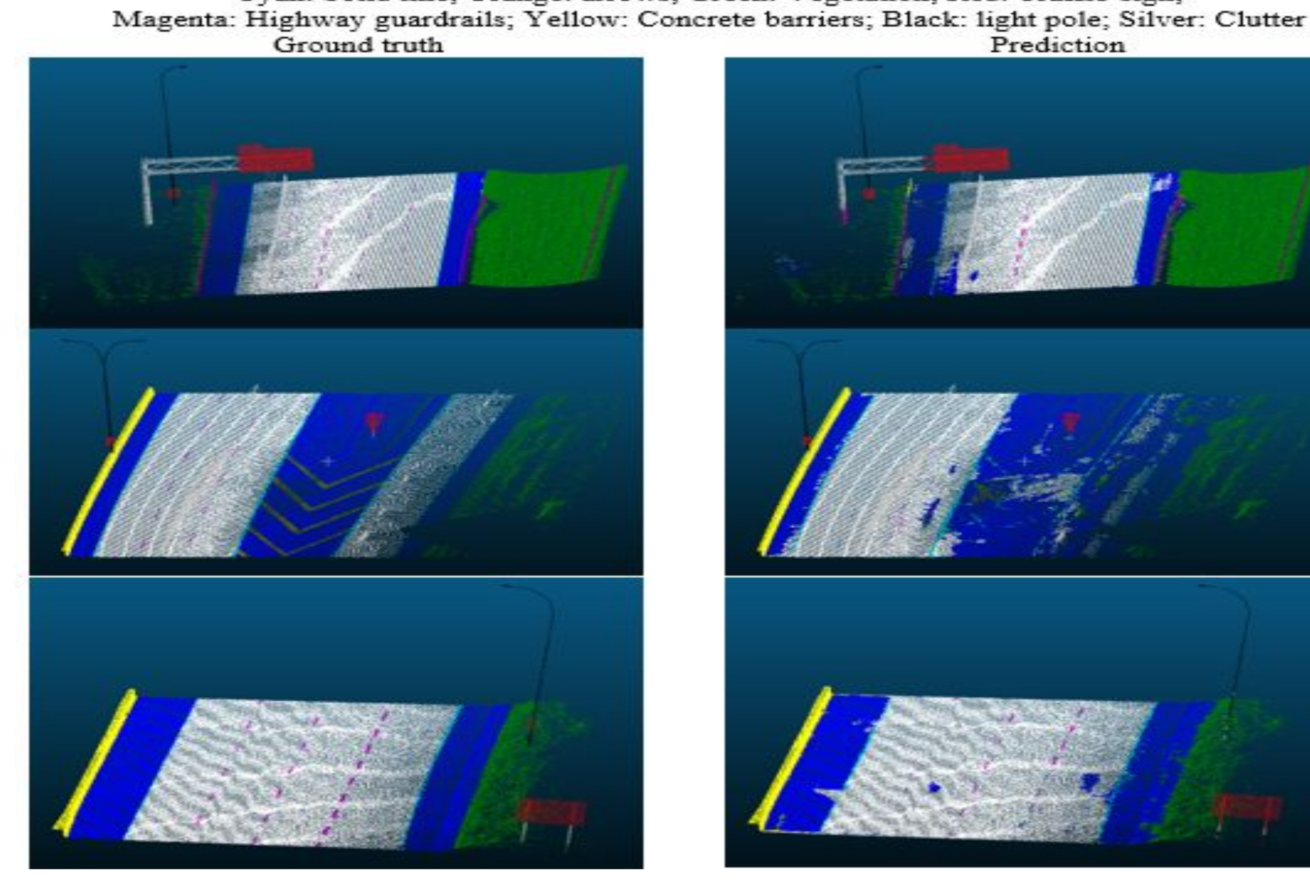
MODEL EVALUATION

FOR 12 CLASSES

The model was trained for 12 classes. Different loss functions were implemented to increase model's mean IoU (Intersection over Union) and mean Accuracy. A combination of **Cross Entropy, Focal Loss, Jaccard Loss** yielded the best output **(mIoU:64.00%, mAcc:71.82%)**.



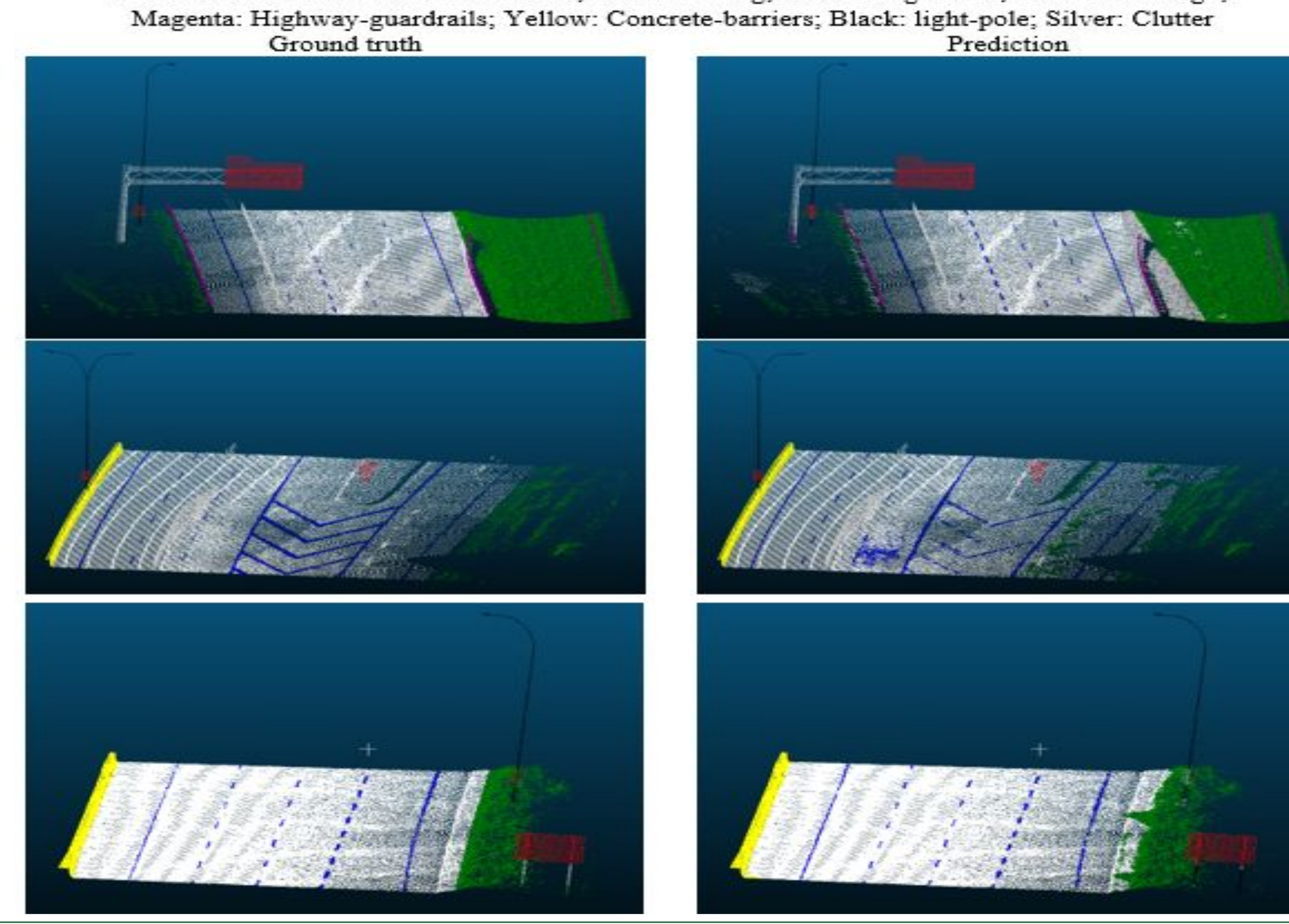
Ground Truth Color: White: Lane; Blue: Shoulder; Olive: Chevrons; Purple: Broken line; Cyan: Solid line; Orange: arrows; Green: Vegetation; Red: Traffic-sign; Magenta: Highway guardrails; Yellow: Concrete barriers; Black: light pole; Silver: Clutter



FOR 8 CLASSES

A second reduced version of the model was trained for 8 classes as shown in the below figure. Here, **Cross Entropy** loss yielded the best results**(mlou:78.92%, mAcc:90.53%)**.

Ground Truth Color: White: Pavement; Blue: Marking; Green: Vegetation; Red: Traffic-sign; Magenta: Highway-guardrails; Yellow: Concrete-barriers; Black: light-pole; Silver: Clutter



RESULTS