# Explaining A Neural Network Model for Segmenting Point Clouds

$Group\ 12$

**Barbara Noemi Szabo**
s3263371

## Abstract

In the context of self-driving cars, LiDAR-generated point cloud data plays a crucial role. This project investigates the decision-making processes of point cloud segmentation models, focusing on the PolarNet architecture applied to the SemanticKITTI dataset. The research explores three key questions: the reliance of the model on non-target points within point clouds and the impact of overlapping and missing points and the examination of how other classes are confused with cars and the road. To enhance the interpretability of the segmentation model, Grad-CAM (Gradient-weighted Class Activation Mapping) is adapted to visualize and quantify the importance of individual points in the decision-making process. Analysis reveals that points closer to the LiDAR sensor and those representing critical features such as roads and vehicles have high influence on model predictions. Furthermore, the project highlights instances of misclassification among similar classes and the role of non-object points in influencing predictions. Insights gained from this research contribute to improving the transparency, reliability, and performance of point cloud segmentation models crucial for autonomous driving applications. The code of the project is available on GitHub.

## 1 Introduction

Point clouds are used for the 3D representation of real-world environments that can be analyzed by computers, particularly in applications such as autonomous systems. These data sets, captured using scanning techniques like LiDAR (Light Detection and Ranging), consist of millions of spatial data points that collectively form detailed models.

In the context of self-driving cars, vehicles rely heavily on accurate perception and interpretation of their surroundings to navigate safely and effectively. Lidar-generated data is essential for tasks such as object detection, classification, and segmentation. This enables the vehicle to recognize and understand various elements, ensuring safe and efficient navigation.

Given the critical role of perception in autonomous systems, it is important that the models used are not only accurate but also explainable and transparent. Safety and reliability in self-driving cars are essential, making explainability a key requirement. Understanding how and why a model makes certain decisions can help in debugging and improving model performance. Moreover, it also builds trust in autonomous systems by providing insights into the model's decision-making process.

In this project, the focus will be on analyzing how the model makes decisions on the dataset. The model used in this project is PolarNet (Zhang et al., 2020), pairing it with the SemanticKITTI dataset (Behley et al., 2019).

The research will address the following questions:

- *How much does the network rely on the other points in the point cloud data that are not part of the object?*
- *What happens if there is significant overlap resulting in missing points for one of the objects?*
- *What are the classes that are often confused with the target classes of the project?*

By exploring these questions, the project seeks to enhance the understanding and explainability of point cloud models in the context of autonomous driving.

To achieve this, the Grad-CAM (Gradient-weighted Class Activation Mapping) (Selvaraju et al., 2017) method will be applied. It is a popular Explainable AI (XAI) technique in image processing, which provides visual explanations of the model's decisions by highlighting important regions in the input data. This project aims to adapt Grad-CAM for point cloud data.

## 2 Background

### 2.1 Semantic Segmentation of LiDAR Point Clouds

Point clouds are collections of points in a 3D coordinate system, each point representing a specific location in space. These points collectively represent the surface of an object or scene, providing a detailed spatial representation that can be used for various analytical and visualization purposes.

LiDAR is a remote sensing technology that uses laser light to measure distances. By emitting laser pulses and measuring the time it takes for the light to return after hitting an object, it creates point clouds as a detailed representation of the environment.

3D semantic segmentation is a process in computer vision and machine learning that involves classifying each point in the space into predefined categories or classes. This task is essential for understanding and interpreting complex environments. Examples of categories include buildings, vegetation, roads, and vehicles.

There are several challenges are associated with semantic segmentation. Processing large volumes of point cloud data requires significant computational resources. Furthermore, real-world environments are highly variable, complex and contains a lot of noise, requiring robust algorithms that can adapt to different scenarios and conditions.

Recent technological advances have significantly improved the capabilities for this task. The advent of deep learning models has revolutionized the field, providing powerful tools. These models can learn complex patterns and features, enhancing the accuracy and efficiency of the segmentation.

### 2.2 Dataset: SemanticKITTI

The SemanticKITTI dataset is a widely-used benchmark for evaluating the performance of 3D semantic segmentation algorithms. It provides a rich set of annotated LiDAR data captured from a moving vehicle in urban, suburban, and rural areas. This dataset includes over 43,000 densely annotated scans. SemanticKITTI categorizes each point into one of 34 classes, including vehicles, pedestrians, buildings, and vegetation.

### 2.3 The Model: PolarNet

PolarNet is designed to enhance the semantic segmentation of LiDAR point clouds through the use of a novel grid representation based on polar coordinates. Traditional approaches that rely on Cartesian grid representations often suffer from imbalanced point distribution across grid cells, leading to poor segmentation performance, particularly in distant and sparse regions. PolarNet addresses this issue by utilizing a polar grid representation.

The key innovation lies in its polar quantization technique, which converts the data from Cartesian to polar coordinates and then quantizes it into grid cells. To process the points within each grid cell, they apply a simplified version of PointNet. This module uses a series of fully connected layers, batch normalization, and ReLU activations to extract local features from the point cloud data.

The network architecture includes a feature aggregation step, where max-pooling is used to aggregate features within each grid cell, forming the feature matrix. This matrix serves as the input to a convolutional neural network (CNN), which captures spatial relationships and produces the final segmentation output.

## 3 Methodology

In this project, I chose to focus on two categories out of the 34 available in the dataset: cars and roads. These two categories were selected for several reasons. First, they have significantly different shapes, allowing for an analysis of two distinctly different classes. Second, both are crucial for the objective of autonomous driving, as accurate detection and segmentation of cars and roads are vital for safe and effective navigation. Additionally, these categories appear in most of the scenes and have the best Intersection over Union (IoU) scores among the classes, indicating strong model performance in these areas.

To conduct this analysis, I decided to examine the first 500 point clouds of scene 8. My initial step was to observe general statistics about the predictions for the two classes. The accuracy for the car class was found to be 0.96, while the accuracy for the road class was 0.98. While the accuracy for the class road is a little better they are both really high which suggest that the model performs well in identifying these points.

### 3.1 False Positives

Then, I analyzed the false positives for each class. The average number of false positives for the car class was 1451.4, and for the road class, it was 748.4. These are quite small numbers given that the average size of a point cloud was 123984.1. To gain a deeper understanding, I examined the distribution of these false positive predictions among the other classes. This distribution provides insights into which classes are most frequently confused with cars and roads.

### 3.2 False Negatives

Similarly, I analyzed the cases of false negatives, where the model failed to predict the correct class for the target label. The average number of false negatives for the car class was 146.7, and for the road class, it was 405.7 for a point cloud. Then again, I investigated the distribution of these cases to gain deeper insights.

### 3.3 Grad-Cam

The importance of individual points for model predictions was assessed using Grad-CAM, a widely-used algorithm in image classification for explaining model decisions. Grad-CAM provides insights into which points in the input data are most influential in making predictions. Specifically, I employed Grad-CAM to analyze the average importance scores across the dataset, distinguishing between instances where the predicted label matched the target and where it did not. This global analysis allowed for understanding how consistently the model assigned importance to points across different prediction scenarios and for answering my first research question: *"How much does the network rely on the other points in the point cloud data that are not part of the object?"*.

Additionally, I examined the importance scores locally for individual instances, providing insights into specific cases where certain points significantly influenced the model's decisions or the importances are surprisingly low. This dual approach—global and local analysis of Grad-CAM importance scores—helped in comprehensively assessing the model's behavior and understanding its decision-making process for point cloud segmentation.

I point is considered to be important if it has an importance score higher than 0.

## 4 Results and Discussion

### 4.1 Cars

### Global Explanations

As described in 3.1 I analyzed the distribution of true labels for false positive predictions.

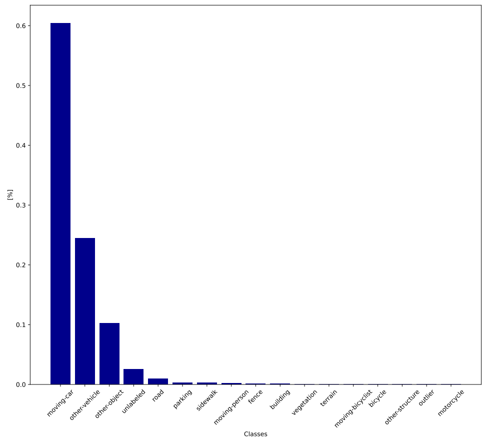

Figure 1: Distribution of false positives for the class car.

In Figure 1, we observe that when the model predicted the class "car" instead of the true label, it frequently confused it with other vehicle classes, such as "moving-car" and "other-vehicle." The "moving-car" category was omitted from the predictions by the model because segmentation is based on a single scan, making it difficult to differentiate moving cars from stationary ones. However, the model still correctly identifies these objects as cars. Misclassifications also commonly occur with the "other object" and "unlabeled" classes. This is because the "other object" category was mapped to the "unlabeled" category in the learning map, leading to random, unstructured data points that are difficult for the model to learn from. Additionally, some elements of the ground, such as roads or sidewalks, were sometimes misclassified, likely due to labeling errors. It can be challenging to determine even for a human whether a point belongs to the ground or the bottom of a car, which can result in inconsistencies in the dataset. Therefore, reviewing and refining the dataset could potentially improve model performance. Of course it is also possible that the cause of miscalassification is the discretization with voxels.

Regarding the false positives, on Figure 2 we can see that the model most commonly predicted the "road" class instead of "car," which aligns with observations from false negative cases. "Sidewalk" also ranks highly among these misclassifications. "Other vehicles" and "trucks" are frequently confused with cars, suggesting that the size or shape of the object may play a crucial role in the model's predictions. Further investigation is necessary to fully understand these dynamics. An intriguing finding is that the "vegetation" class is often predicted instead of the correct "car" label. This could be because the size and appearance of for example bushes can resemble cars, especially for vehicles that are partially occluded or obscured from view.

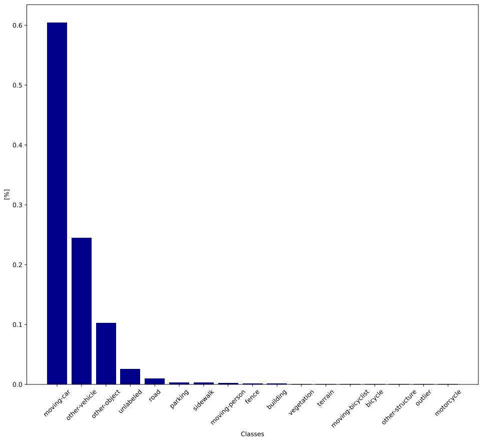

Figure 2: Distribution of false negatives for the class car.

Analyzing the importance scores as detailed in section 3.3, I initially focused on points classified by the model as "car". Figure 15 illustrates an error bar plot depicting the importance scores for these points. The average importance score generally aligns closely with the overall average of 0.68, but notable dips in importance are observed around the 360th and 460th points. These instances warrant further investigation and comparison with scans that exhibit higher scores. The average importances of neighboring point clouds are similar, which correlates with the chronological continuity of adjacent point cloud datasets.

Figure 16 displays the same plot but for points not classified as "car". Here, the standard deviations of these values are significantly lower due to most scores being zero or near-zero. Moreover, the average importances are relatively modest, suggesting that some points are still crucial for predicting cars, despite they are not classified as such.

Additionally, Figure 6 compares the maximum importance scores of points within the "car" set against those outside it. Naturally, points within the set exhibit higher importances, but other points often show remarkably high scores, typically close to 1. These points are most likely neighboring ones.

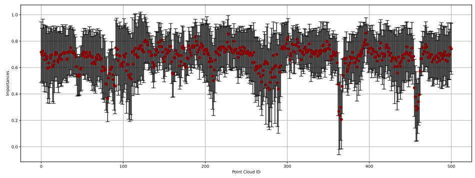

Figure 3: Points with label "car".

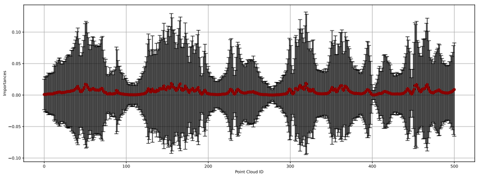

Figure 4: Points with label other than "car".

Figure 5: Importance scores

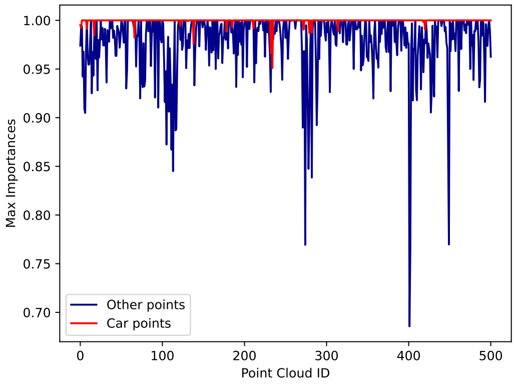

Figure 6: Maximum importance scores for each point cloud for class car.

### Local Explanations

In Figure 9, we see the point cloud with the lowest average importance for points classified as cars. The first subfigure presents the Grad-CAM heatmap of the point cloud, with the predicted labels for each class shown below on the second one. The heatmap reveals that the points deemed important are mostly classified as cars. In Figure 10 there is the relative frequency of classes for important points. It is evident that nearly 70% of the important points for the car class were classified as cars. The figure also shows that points from neighboring classes like road, vegetation, and sidewalk were also considered important, this might be only because of discretizing the point cloud with polar voxels. In this instance, the model failed to fully identify the "other-vehicle" in the center, but intriguingly, the points belonging to the "other-vehicle" were still important for the car label. This suggests the significance of neighboring points. Another noteworthy observation is that around 7% of points that were not important at all were still classified as cars, likely based on their surroundings.

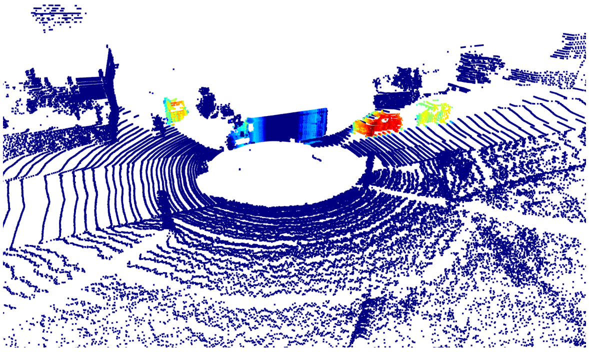

Figure 7: Grad-CAM heatmap.

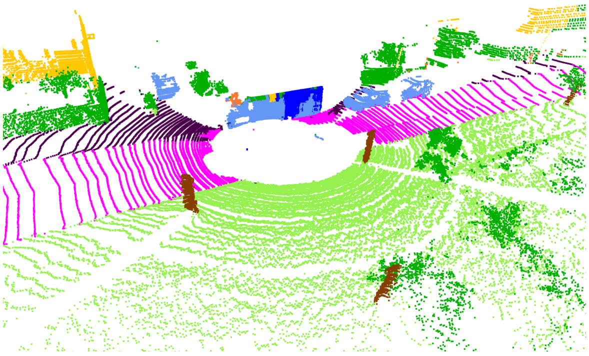

Figure 8: The predicted labels.

Figure 9: Point cloud 366 which has the lowest average importance score for predicting label "car".

In Figure 7, we observe that the highest importance scores are attributed to the second car from the right. This is likely due to the car being fully visible and intact. In contrast, the car on the left is missing many points and only its rear is visible, resulting in lower importance scores, although the points are still correctly classified. This overlap, along with the false positive cases, likely contributes to the significantly low average importance score.

Regarding the second research question, we can observe that when only half of the car is visible—whether the front is obscured or, as with the car on the far right, only one long side is visible—the importance of the points decreases but the classification is still correct. Additionally, it is noteworthy that the part of the car closer to the LiDAR tends to have higher importance.

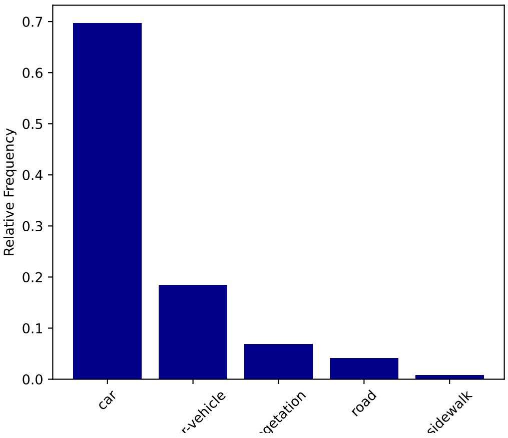

Figure 10: Relative frequency of classes for important points of point cloud 366.

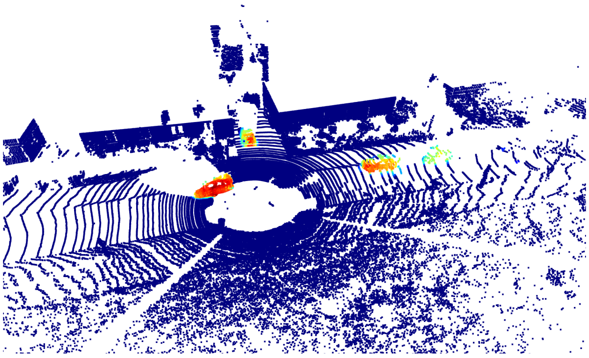

Figure 11: Grad-CAM heatmap for point cloud 189.

Point cloud 189 has one of the highest average importance scores for points identified as cars. In Figure 11, we can further support the assumption from the previous example that the parts of the cars closest to the LiDAR are the most important. One possible reason for this is that the closer the points are to the LiDAR, the more densely packed they are. This density might be important for the model for making predictions. This is particularly evident when observing that the second car from the right has lower importance than the one in the middle, even though it is nearly completely visible from two sides and the top, which is the maximum visibility achievable with a LiDAR.

However, as illustrated in Figure 12, even after removing the points identified as cars, there are still significant points present. Since the predicted label is consistent for each point within a voxel, it indi-

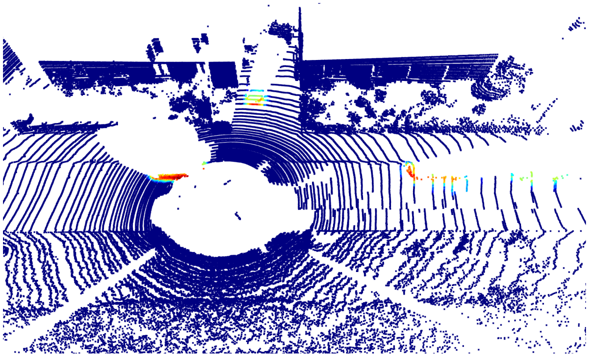

Figure 12: Grad-CAM heatmap for point cloud 189 without points identified as cars.

cates that neighboring points, such as road points near the car, are important for the prediction. This could be a distinguishing factor from "other objects" or "vegetation," although this would require further investigation.

## 4.2 Roads

### Global Explanations

Analyzing the false positive values for the class "road" in Figure 13, we observe that it is mostly confused with similar classes such as sidewalk or terrain, as well as with vehicles that are typically in close proximity to the road. A similar trend can be seen in the false negative cases, as shown in Figure 14.

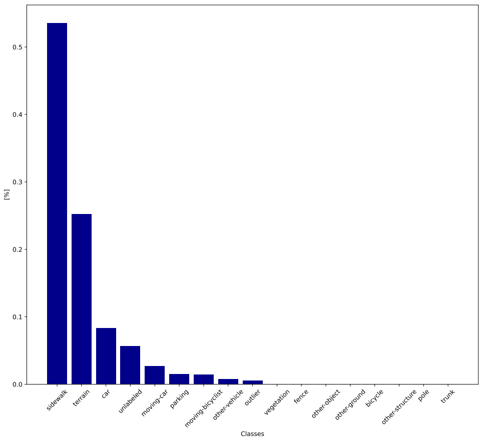

Figure 13: Distribution of false positives for the class road.

Examining the error bars for the road points in Figure 17, we find that the standard deviation is more consistent across the dataset compared to the car class. Additionally, the standard deviation is higher, and the average importance score is much lower at 0.24. However, Figure 18 reveals that the maximum scores for the road points are almost always 1, similar to the cars. This higher standard deviation can be attributed to the fact that the road class encompasses more points than a car and also covers a larger surface area.

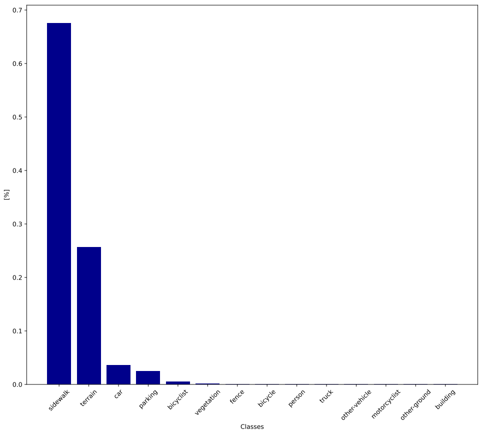

Figure 14: Distribution of false negatives for the class road.

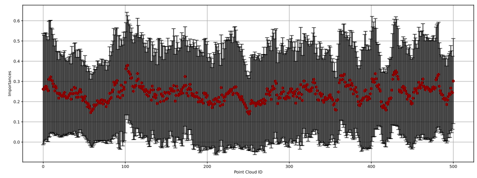

Figure 15: Points with label "road".

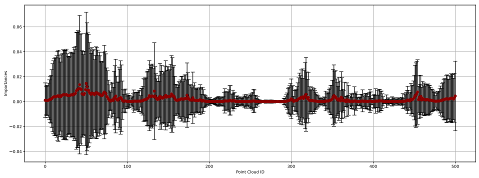

Figure 16: Points with label other than "road".

Figure 17: Importance scores

For points that weren't identified as road, the standard deviation is much lower than for the cars. As illustrated in Figure 18, the maximum values for these points are also significantly lower. This indicates that for the road class, the contribution of points outside of the class is much less than it was for the cars.

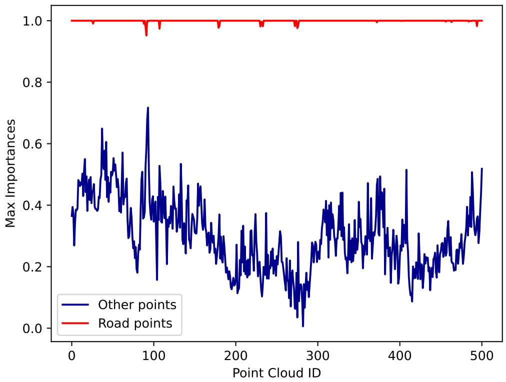

Figure 18: Maximum importance scores for each point cloud for class road.

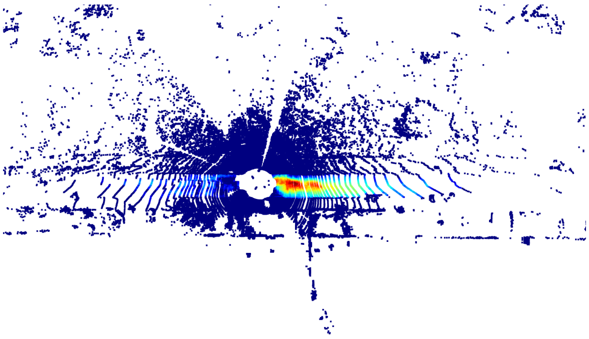

Figure 19: Importances for road class on point cloud 95.

### Local Explanations

Examining multiple examples of the importance heatmap for the road, we observe a consistent pattern, as shown in Figure 19: the road points are primarily significant, with the most crucial part being the center of the road in front of the car.

Another interesting phenomenon is that if the road branches off in multiple directions and is not obscured by the corner of a building, that part also becomes more important, as seen in Figure 20. To draw more definitive conclusions, we would need to analyze additional point clouds with crossroads. Unfortunately, the subset of the database I analyzed did not include more examples.

## 5 Conclusions

In conclusion, point clouds are indispensable for 3D representation and analysis in autonomous systems, facilitating critical tasks like object detection, classification, and segmentation. Lidar-generated data plays a pivotal role in enhancing the perception capabilities of self-driving cars, enabling them to navigate safely through complex environments. The imperative of ensuring model explainability and transparency cannot be overstated, as it underpins safety, reliability, and trust in autonomous systems.

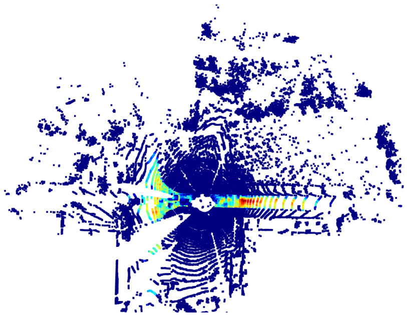

Figure 20: Importances for road class on point cloud 103.

This project focused on delving into the decision-making process of models using PolarNet with the SemanticKITTI dataset. The study explored the extent to which the network relies on non-object points and the implications of significant overlap that may cause missing points for objects. By adapting Grad-CAM, an Explainable AI technique, visual insights were gained into the model's decisions on point cloud data.

The importance scores derived from Grad-CAM highlighted that points closer to the LiDAR or representing critical features such as roads or vehicles exerted substantial influence on model predictions. The analysis revealed instances where the model misclassified similar classes (e.g., cars versus other vehicles), and notably, points not directly part of the main object (e.g., road points for car classification) sometimes played significant roles in predictions. The classes that are usually confused with cars are not only similar in size but usually similar in real-life function too: for example "other-vehicles" or "moving cars".

Moving forward, potential improvements in model performance could involve reviewing labels for points at the intersection of cars and the ground. It's essential to consider whether misclassifications might be attributable to voxel size limitations. For roads, the most crucial points were consistently those directly in front of the car, emphasizing their importance for accurate navigation and scene understanding in autonomous driving scenarios.

## References

J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. 2019. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. In *Proc. of the IEEE/CVF International Conf. on Computer Vision (ICCV)*.

Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Yang Zhang, Zixiang Zhou, Philip David, Xiangyu Yue, Zerong Xi, Boqing Gong, and Hassan Foroosh. 2020. Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.