Lidar and Radar Systems
Task 2
Evaluation of an Object Detector

Pravin Tulshidas Palve, Matriculation Number-13541620
Tanmay Tushar Totala, Matriculation Number-27941616

January 7, 2024

1 Introduction

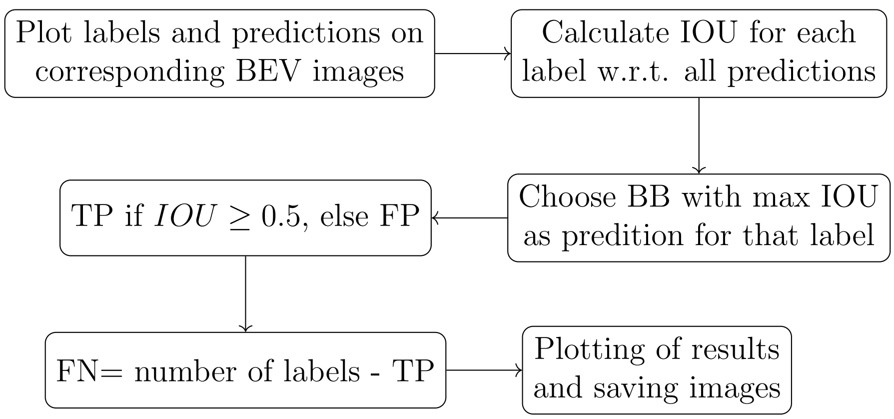
Throughout this report, we will extensively use the following terminologies to keep things clean and convenient:

`IOU= Intersection area/Union area; TP= True Positive; FP= False Positive; BEV= Bird's Eye View; BB= Bounding Boxes; FN= False Negative.`

In this task, we aim to assess the accuracy of an object detector built using complex YOLO by comparing the predictions with ground truths. In simple terms, it is pretty similar to checking an answer sheet (Predictions) of a Student (Object Detector) and scoring it by comparing with answer key (Ground truth labels). Aim of the task also includes calculating precision and recall values for each scene, which will be defining factors for evaluation of object detector. To consider a detection to be TP, the condition is $IOU \geq 0.5$. In this report, we will discuss the steps involved in data handling, plotting and calculations followed by presentation of results.

2 Methodology

The programming was done in *Python* where we used standard libraries like *NumPy, Matplotlib, shapely* and *OS* to read out, combine, draw, slice, process and plot our results. The flow of the program is shown below. The first problem we faced was of multiple detections for one ground-truth object. The solution to this problem can also be very easily understood from the flowchart below...



The second part of the task is to calculate Recall and Precision values for each scene. We will now discuss the significance of each one along with relations[1] to calculate them.

- **Recall**: It answers the question: "Out of all the actual positive instances, how many did the detector successfully find?". Recall focuses on minimizing false negatives, ensuring that as few relevant instances as possible are missed.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **Precision**: It answers the question: "Out of all the instances predicted as positive, how many were correctly identified?". Precision focuses on minimizing false positives, ensuring that the predicted positive instances are as accurate as possible.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

3 Results

In this section, we will look at the generated BEV images with BB's and try to conclude our findings.

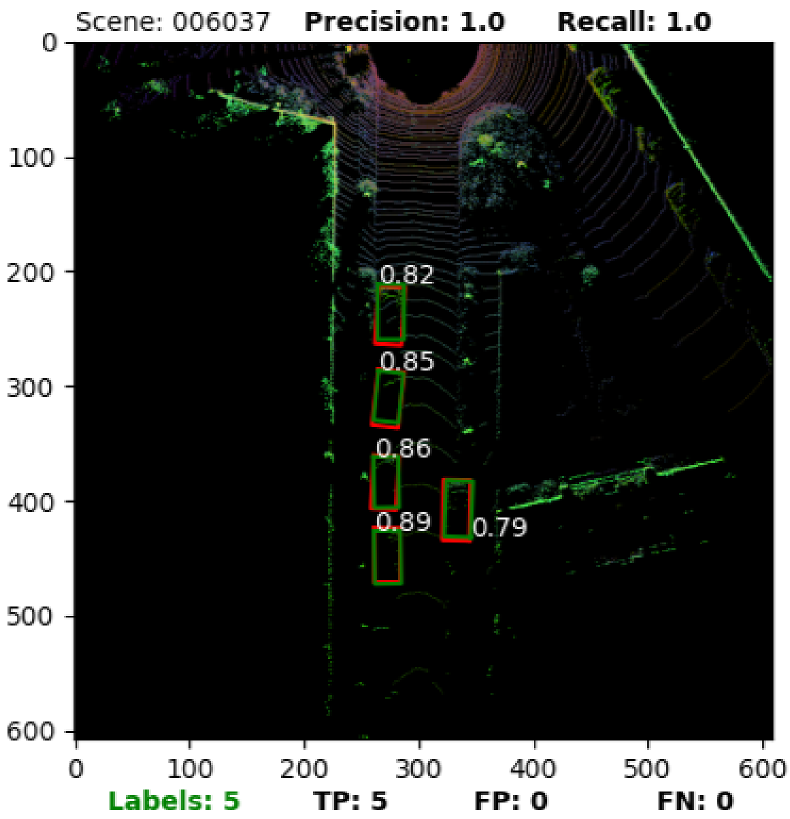


Figure 1: 006037

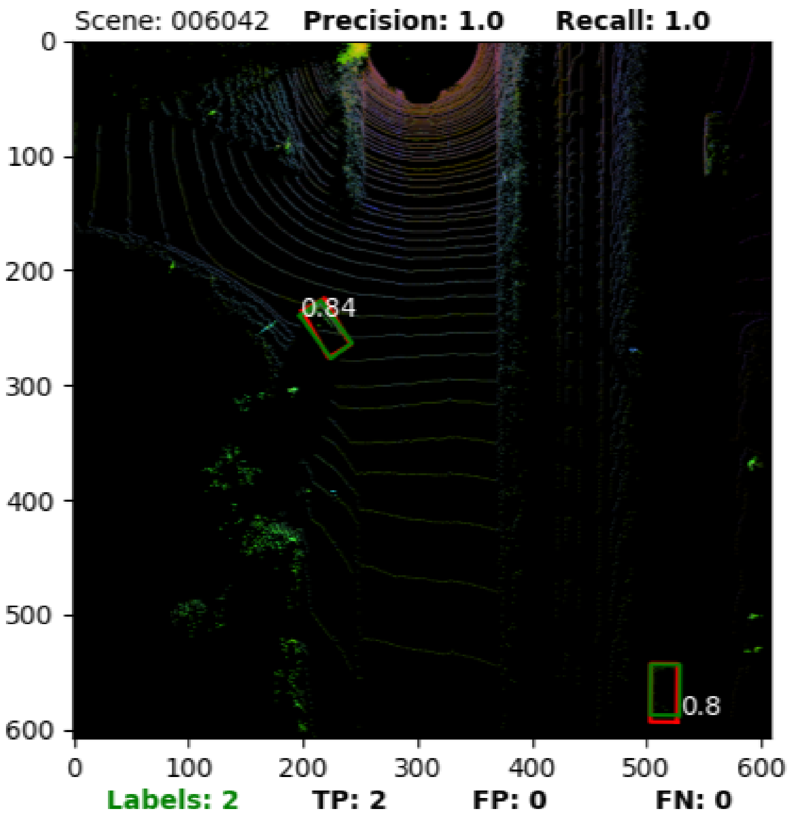


Figure 2: 006042

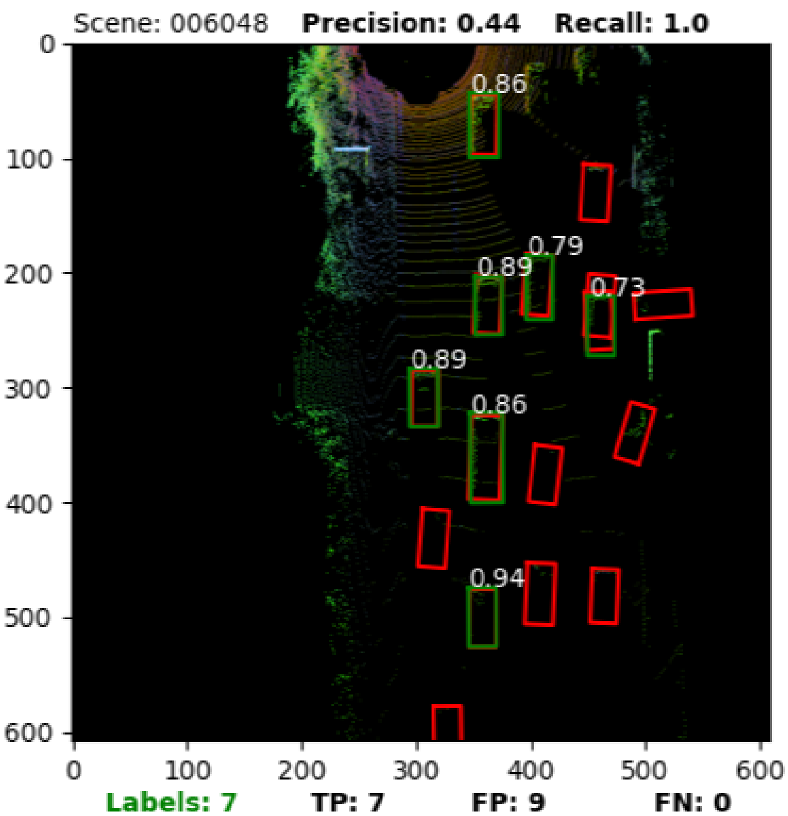


Figure 3: 006048

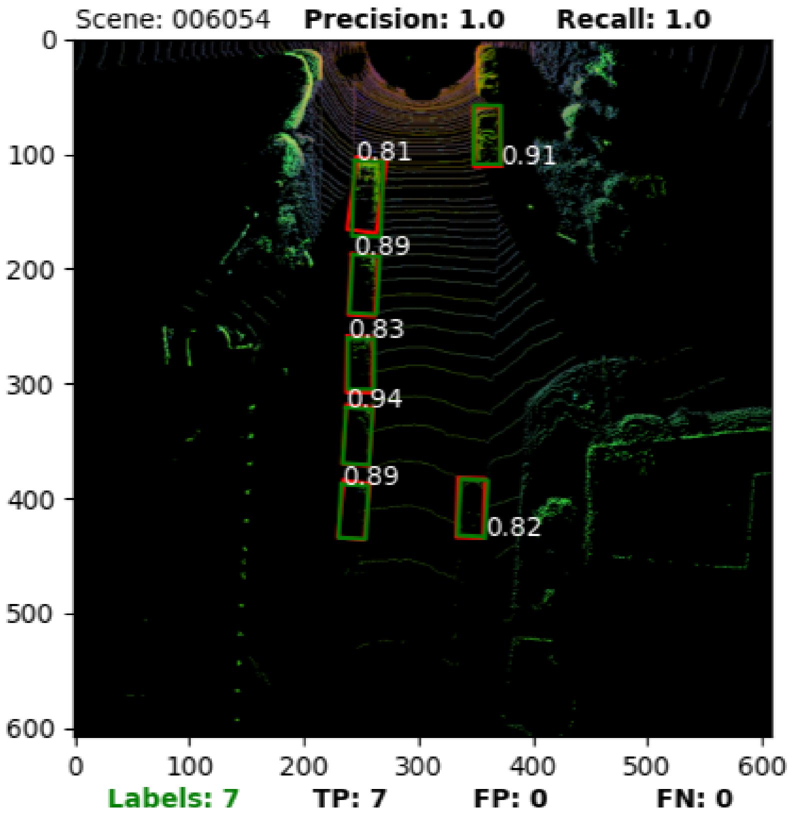


Figure 4: 006054

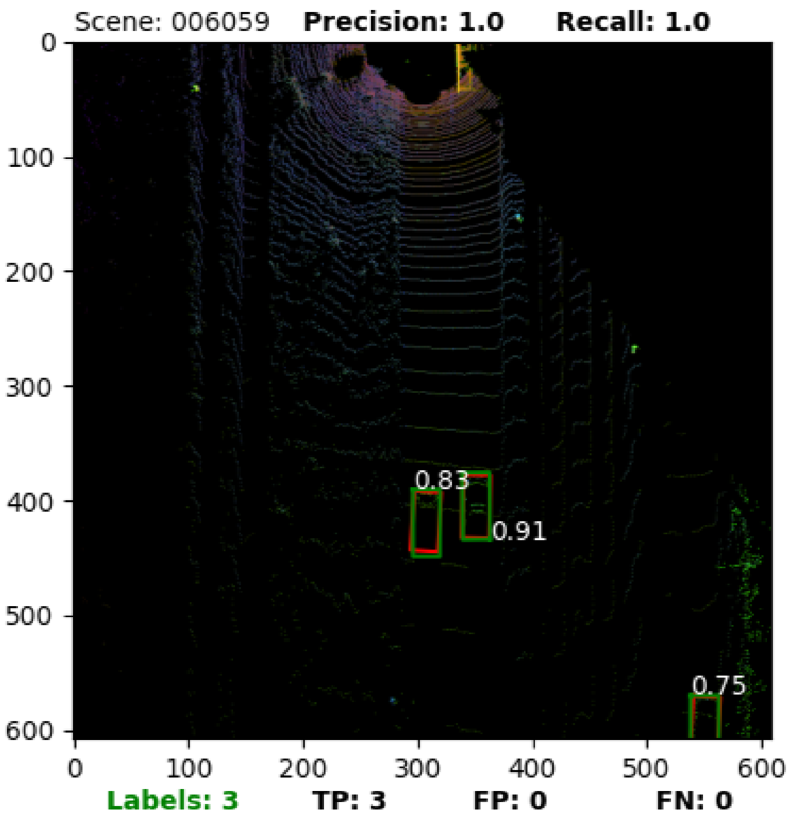


Figure 5: 006059

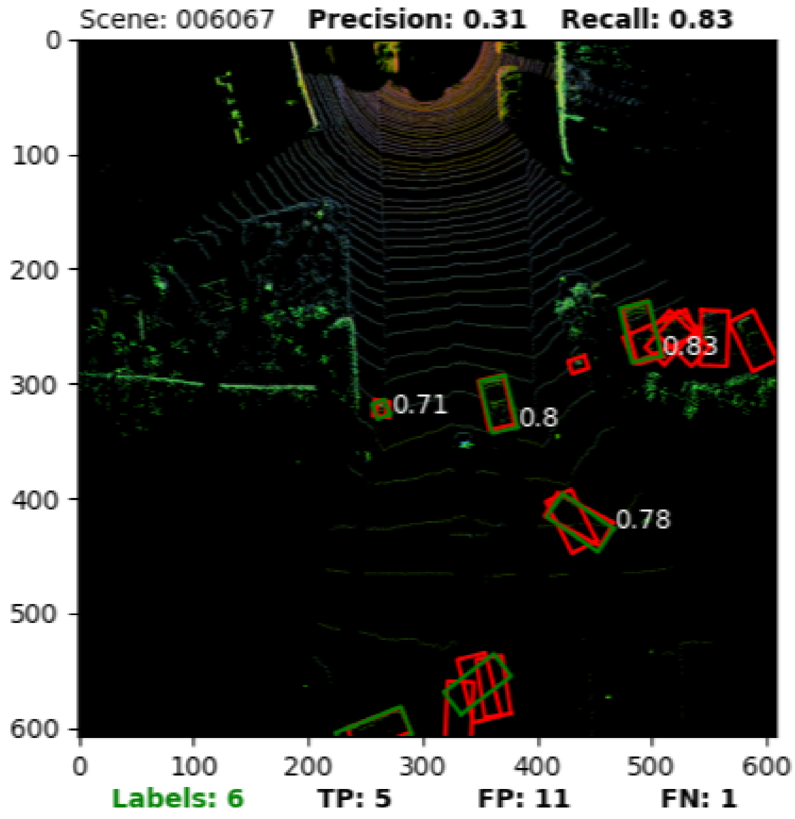


Figure 6: 006067

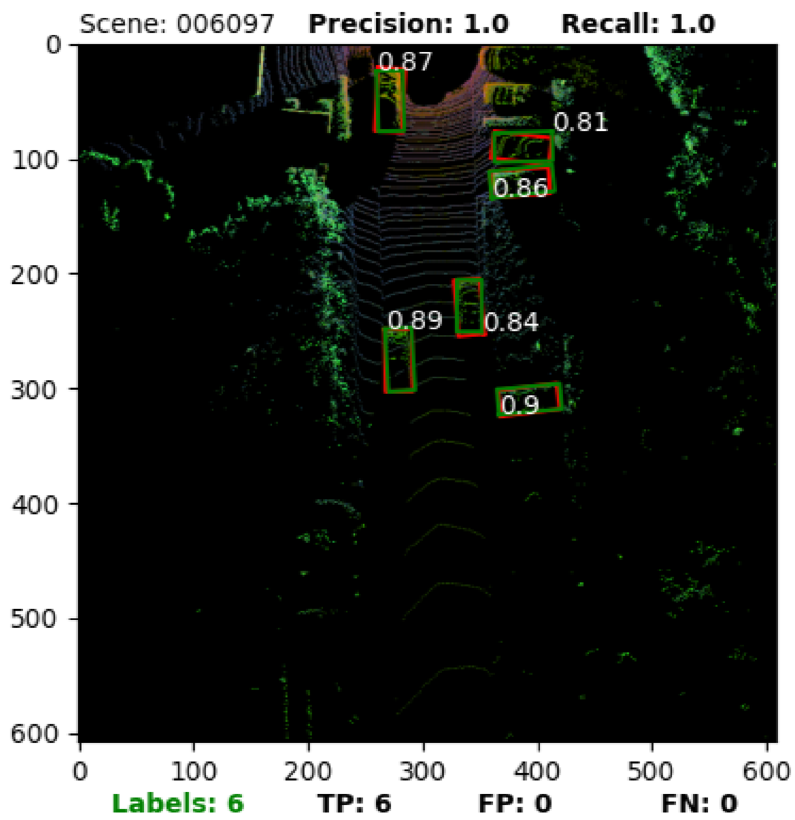


Figure 7: 006097

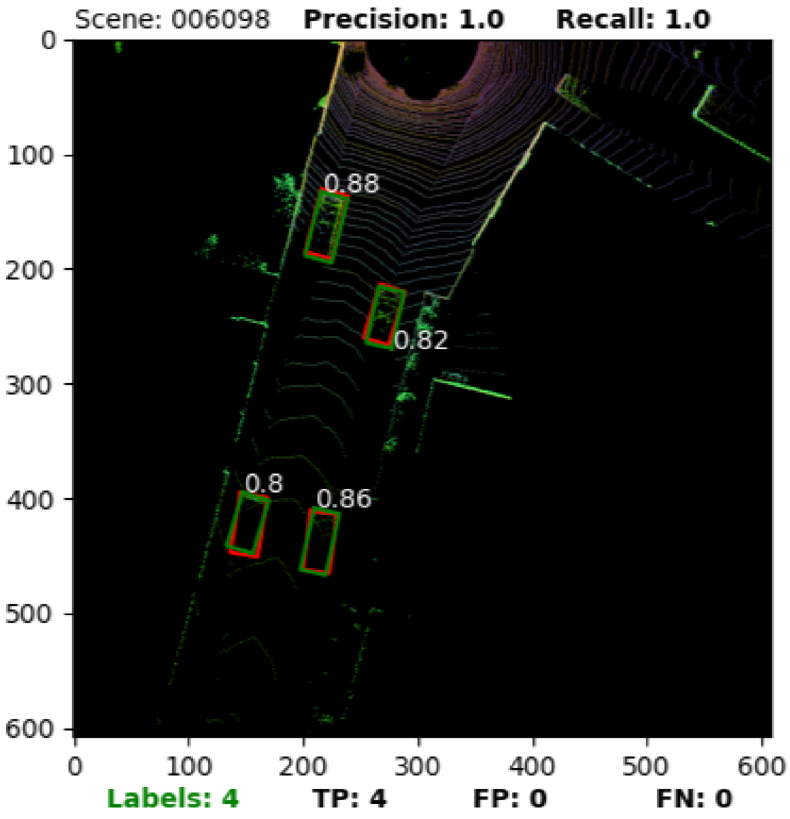


Figure 8: 006098

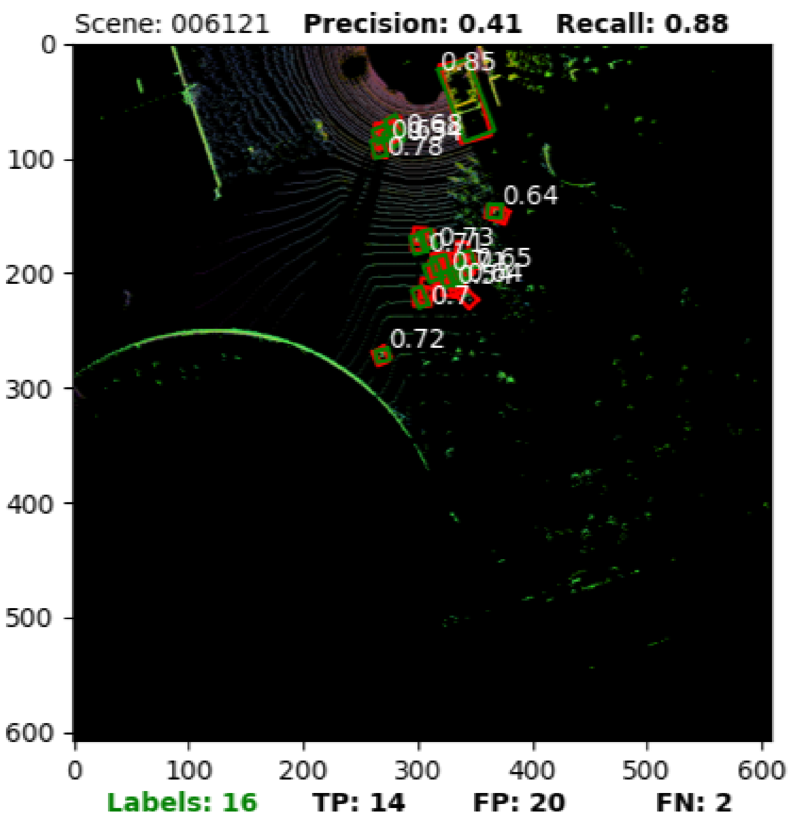


Figure 9: 006121

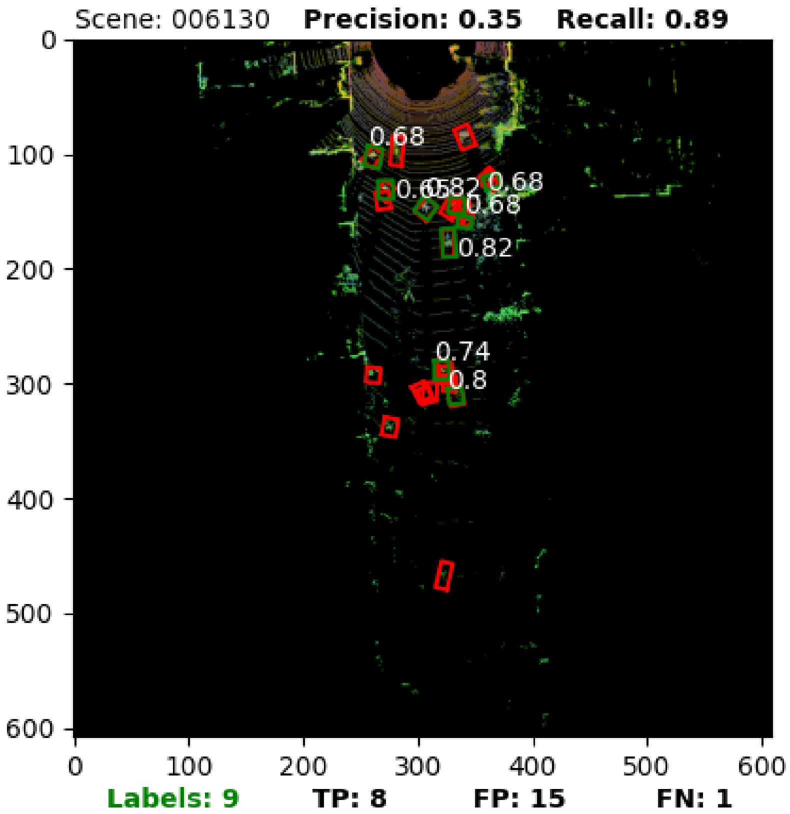


Figure 10: 006130

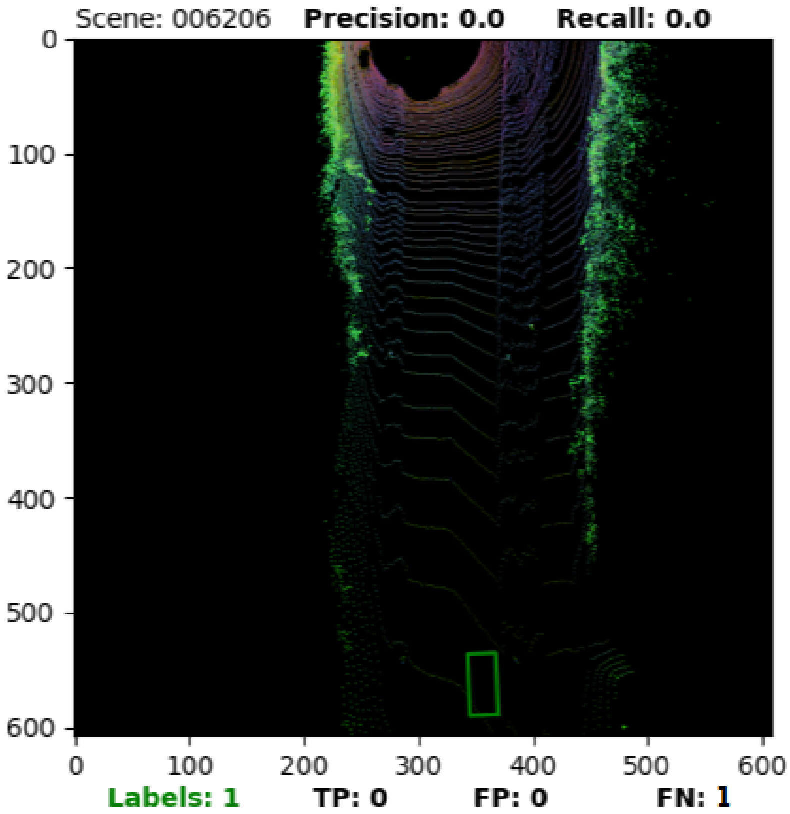


Figure 11: 006206

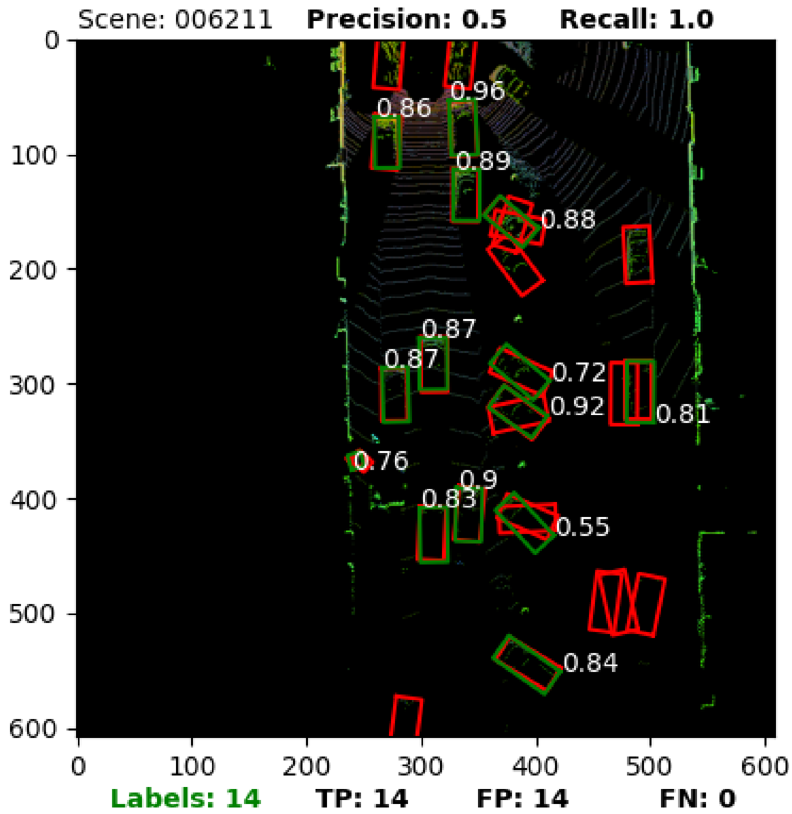


Figure 12: 006211

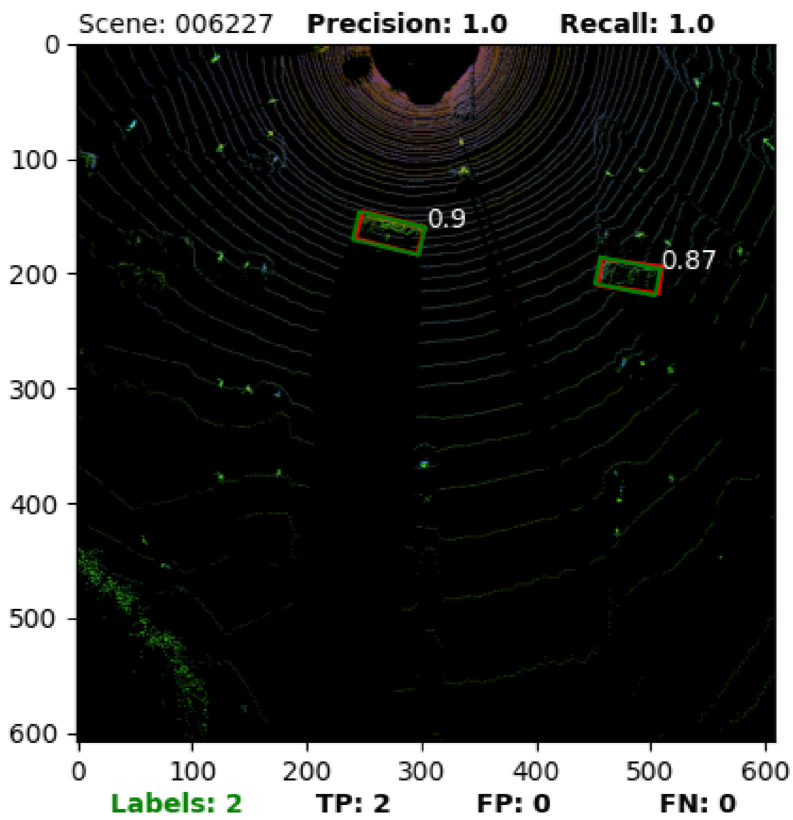


Figure 13: 006227

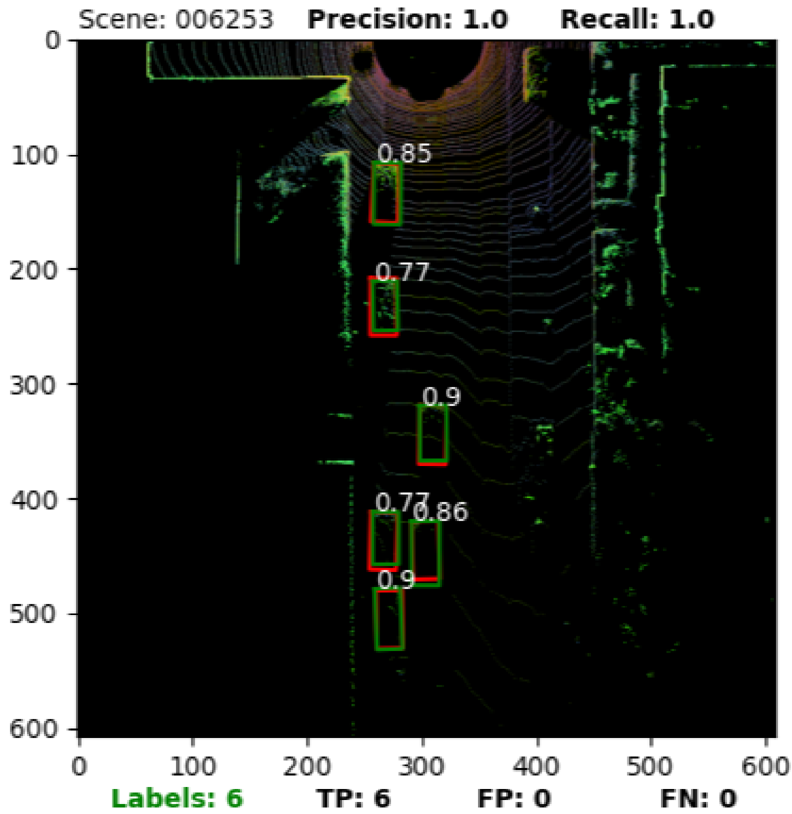


Figure 14: 006253

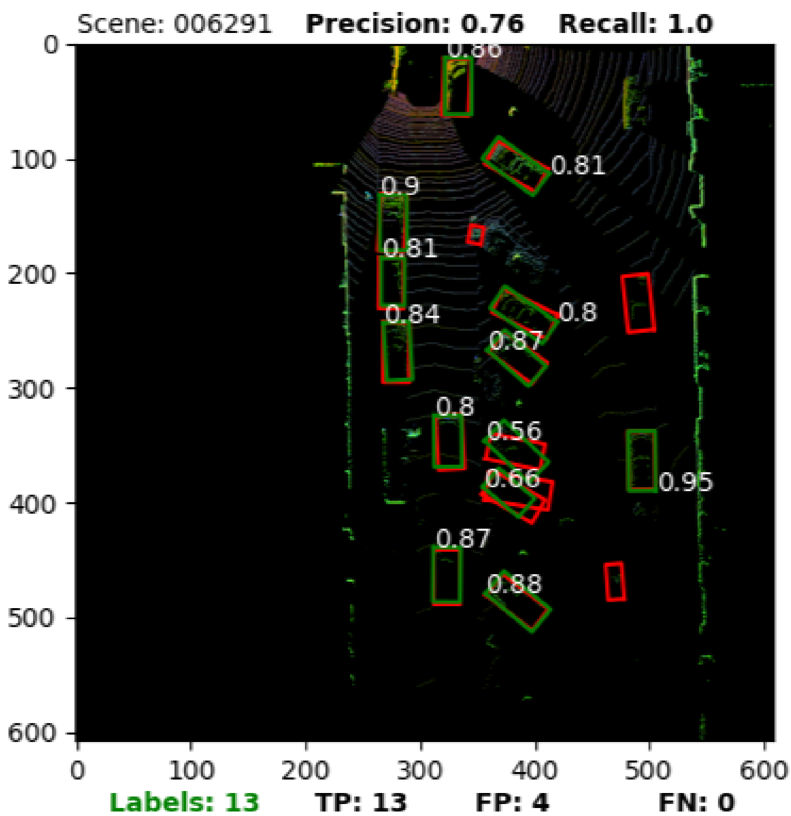


Figure 15: 006291

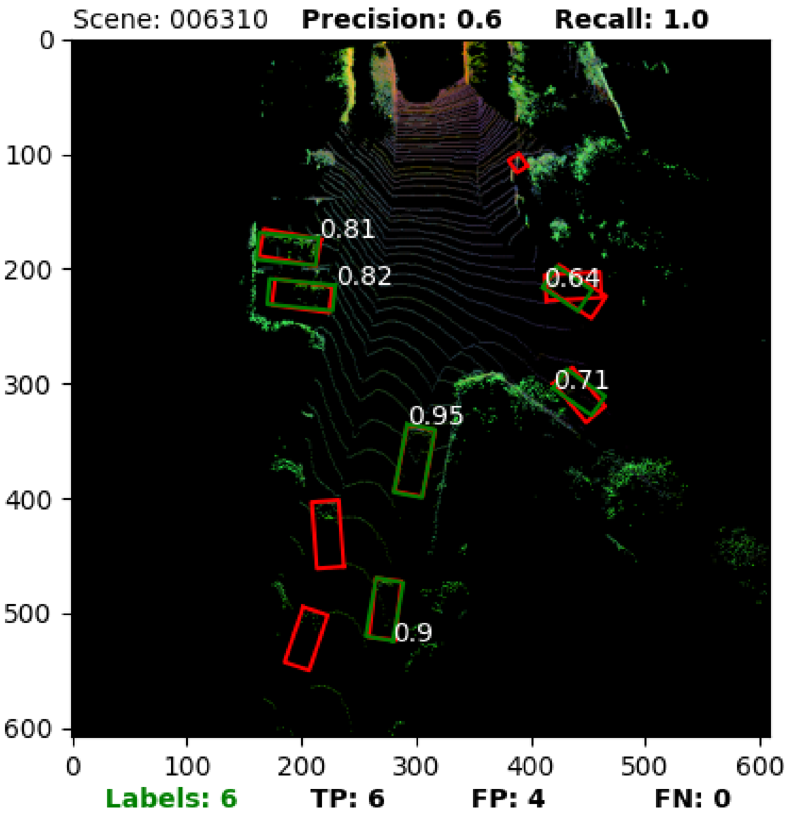


Figure 16: 006310

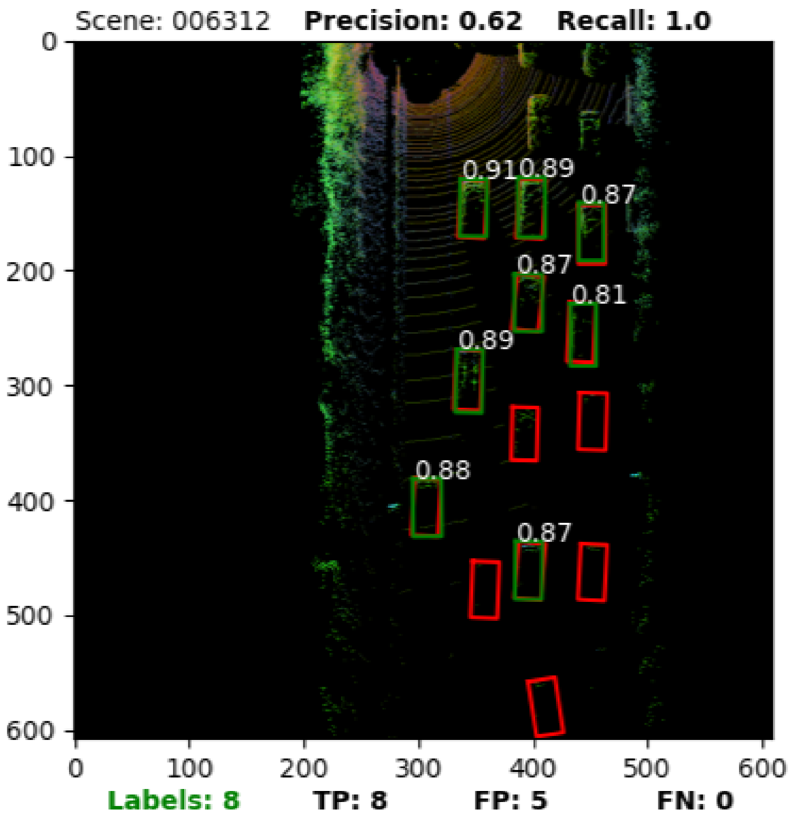


Figure 17: 006312

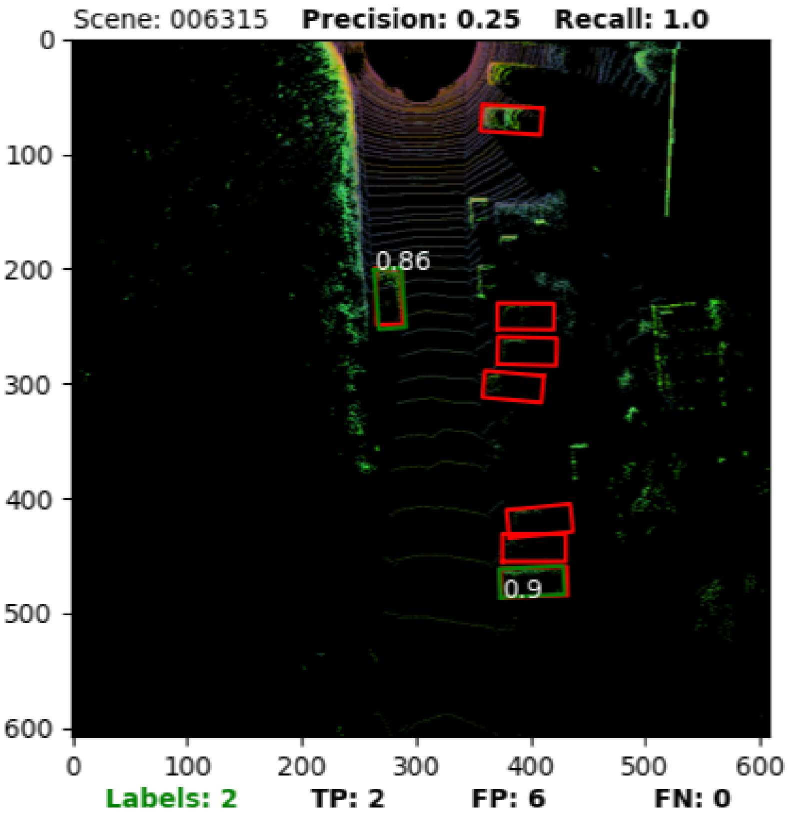


Figure 18: 006315

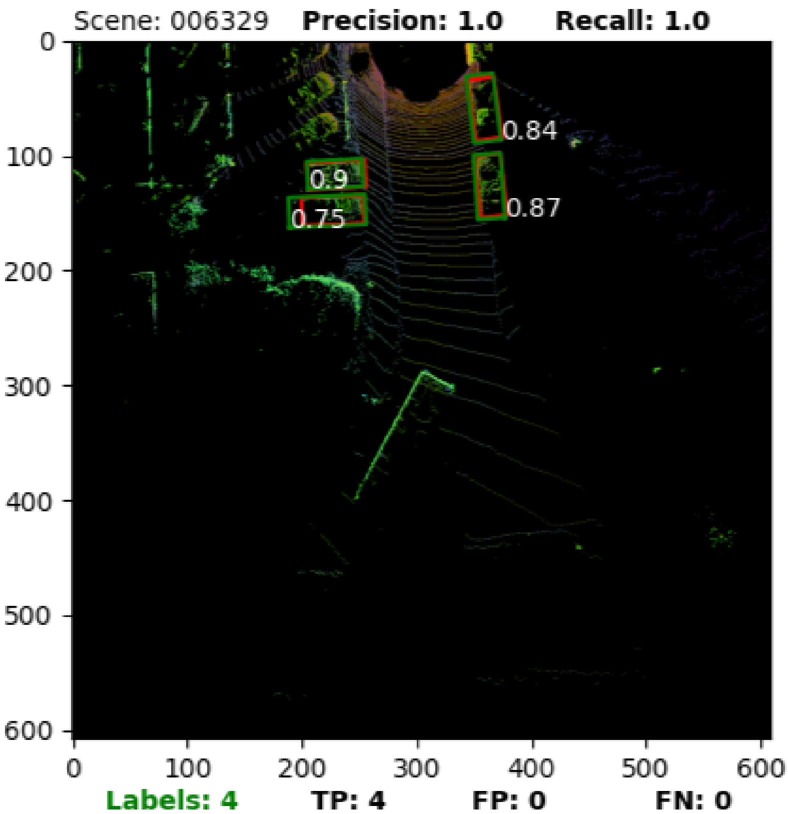


Figure 19: 006329

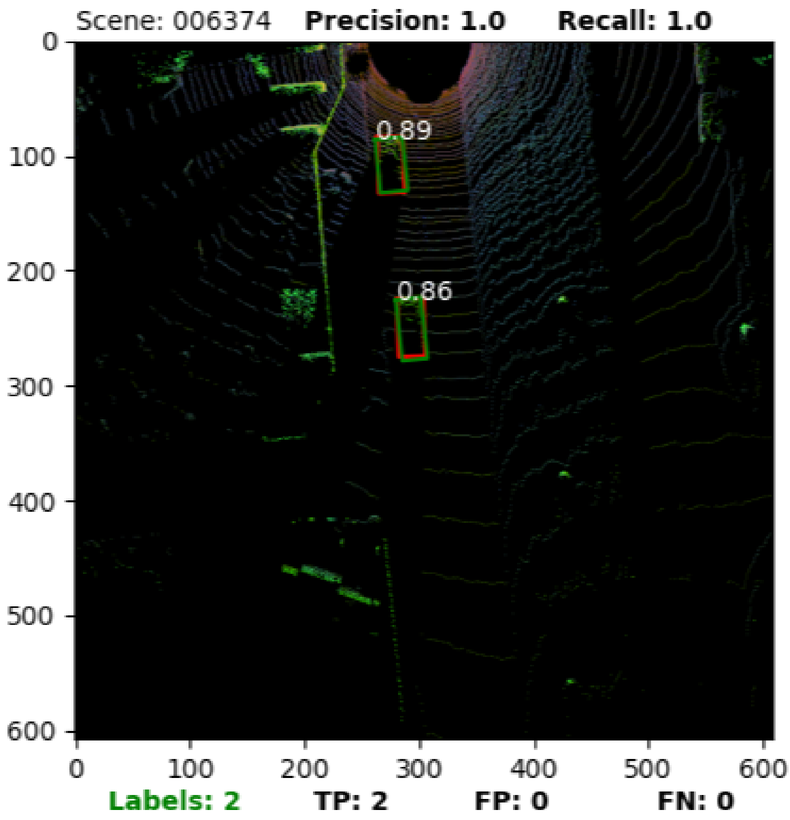


Figure 20: 006374

4 Conclusion

From above figures it is quite evident how the recall and precision varies depending upon complexity of the scenes. For example, in simple scenes like figure 20, 13 and 2 where cars are away from each other and there are not much features in the driving environment, the object detector works pretty well.
As soon as the complexity of the scene increases, like in scenes 6, 9, 10 and 12, where there are multiple objects along with other un-classified features, object detector starts picking up false positives.

5 Further Scope

The performance evaluation of an object detector based on only Precision and Recall is not enough. For example, a object detector with low confidence threshold would detect multiple ground truth BB's so the recall value would be near 1 but precision would be very small. This can be done by the Average Precision[1] (AP) of an object detector. It is the Area Under the Curve (AUC) of the Precision-Recall curve. The formula for Average Precision (AP) at a given recall level is given by:

$$AP = \sum_i P(r_i) \cdot \Delta\text{Recall}_i$$

where $P(r_i)$ is the precision at recall level r_i, and ΔRecall_i is the change in recall from the previous recall level. The summation is performed over all recall levels.

Additionally, the behavior of object classification program can also be studied by calculating the F1 score[1]. The F1 score is the harmonic mean of precision and recall and provides a balanced measure of a detector's performance.A high F1 score indicates a good balance between precision and recall. The F1 score is calculated as:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

6 References

[1] Flach, P. and Kull, M. (2015). Precision–recall–gain curves: PR analysis done right. In Advances in Neural Information Processing Systems, Montreal (Vol. 28, pp. 838–846). Neural Information Processing Systems Foundation