DLAD Project 3 Report

Kexin Shi, Zheng Shen

June 2021

1 Problem 1

1. The average recall of the first stage is 0.801.

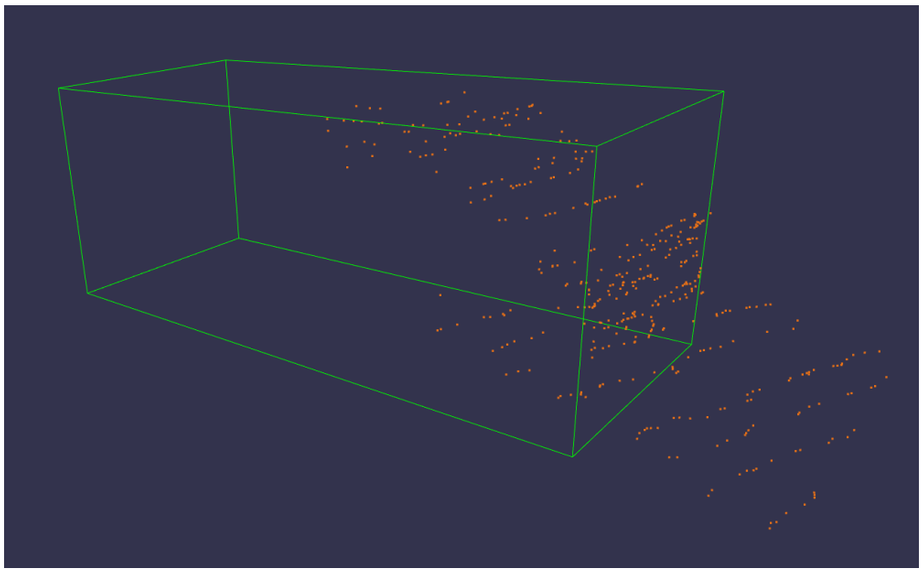
   We know recall = TP / (TP+FN), precision = TP / (TP+FP). Recall is a better metric comparing to precision for the first stage because we want to select all proposals for objects as much as possible. Typically, increasing recall will results in a drop in precision and vice versa. We prefer recall over precision here because we can still refine precision at later stage, however, if we miss some objects at this stage, we cannot recover it anymore.

2. Fig.1 shows the three examples of ROI figures. The name of the figure corresponds to the frame name. The corresponding box parameters are:

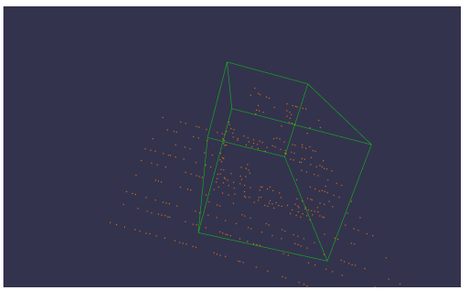
   (a): $[-0.5345, 1.6490, 25.9508, 1.6750, 1.7026, 4.2879, -1.5607]$;

   (b): $[-3.1509, 1.6907, 13.6435, 1.5311, 1.6099, 3.8902, 1.6171]$;

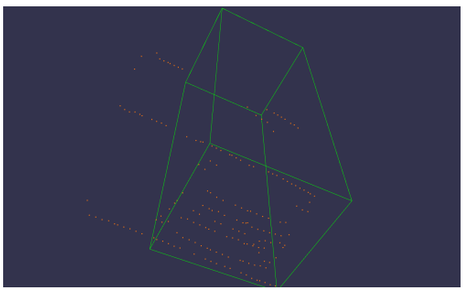
   (c): $[4.7585, 1.5570, 8.7796, 1.5255, 1.6306, 3.9068, -0.7152]$.



(a) nNZgRsbRAPZCUvPJa4yQ



(b) EcIn8U1MgsrE1rt4kW9d



(c) 0lvf6yhDIQP6RikCciGf

Figure 1: Task 1.2

3. Fig.2 shows a visualization of a scene(nNZgRsbRAPZCUvPJa4yQ) from the training set with all proposals.

   (a) Since there are usually only a handful of objects in a scene, if we randomly sample from all proposals, the percentage of foreground proposals will be low, and the loss from foreground proposals would only have a small weight. As the result, the information gained from foreground proposals would be obscured by the other proposals, and the trained NN may not be able to find accurate boxes.

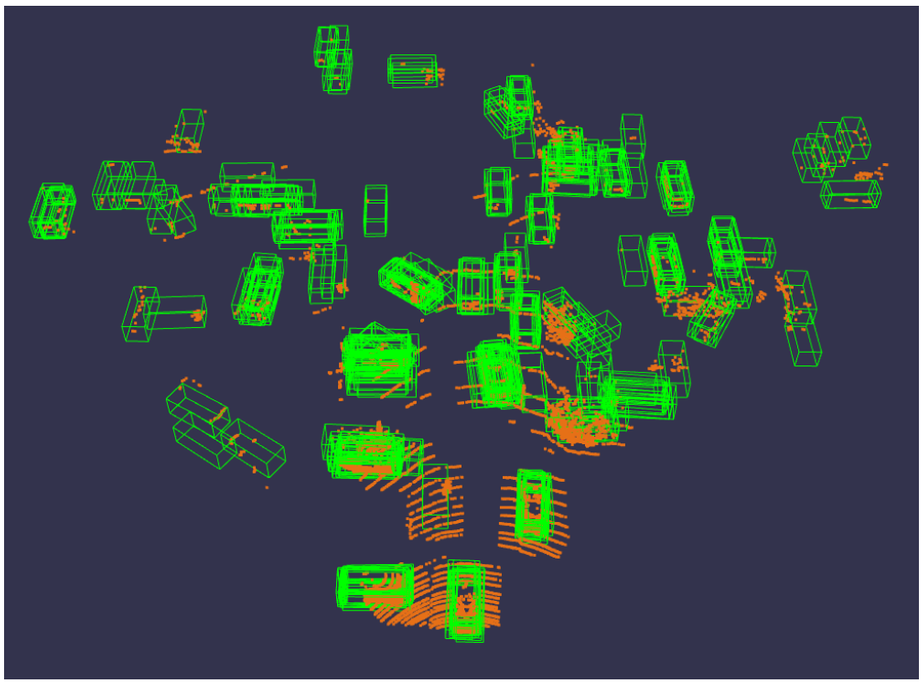


Figure 2: Task 1.3

(b) If we don't sample easy background proposal at all, we lose the information of the background of the scene that we could have gained easily. The trained NN may not be able to distinguish the contour of the objects from the background accurately, which would result in degeneration of the prediction.

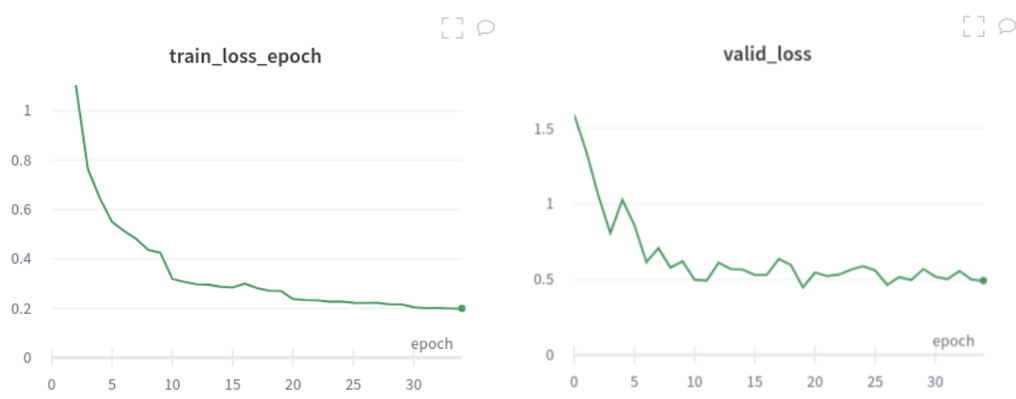
(c) Using 0.5 IoU will easily cause misclassification. As the result, the proposals will be imbalanced. In addition, the proposals that have IoU close to 0.5 are difficult to classify by themselves. Involving them might actually deteriorate the learning ability of the network. Therefore, restricting the threshold will be better actually, because there might have some cases that a fore ground sample has IoU lower than 0.5 and a background sample has IoU higher than 0.5.

First, it will make sense and we are more confident with using the highest IoU proposal. Second, learning and mapping from highest IoU proposal will be much faster and easier than doing the same thing for lower IoU proposals.

4. Please refer to the code implementation.

5. If all the bounding box have the same height, then it is no difference of using between BEV or 3D. However, if the heights are different, BEV approach will usually returns higher IoU values than 3D approach. This means that in the NMS algorithm, we will get more final predictions than we expected, which means 2D BEV IoU reduces the computational cost compared to 3D IoU, and its predictions could be almost as good as that of 3D when the road is flat.

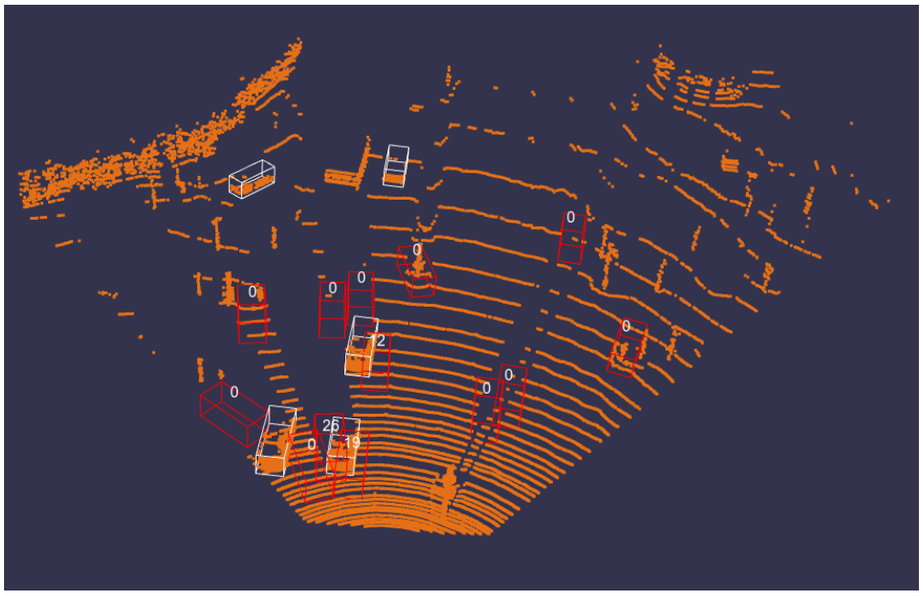
6. The validation at the end of training is 0.492. A visualization of the scene at the end of training is shown in Fig.4.



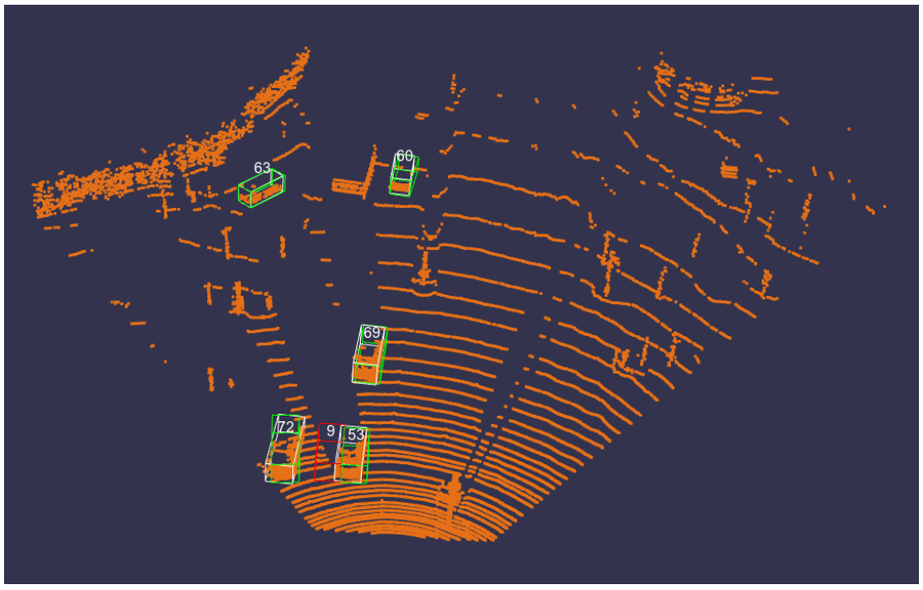
(a) Training loss

(b) Validation loss

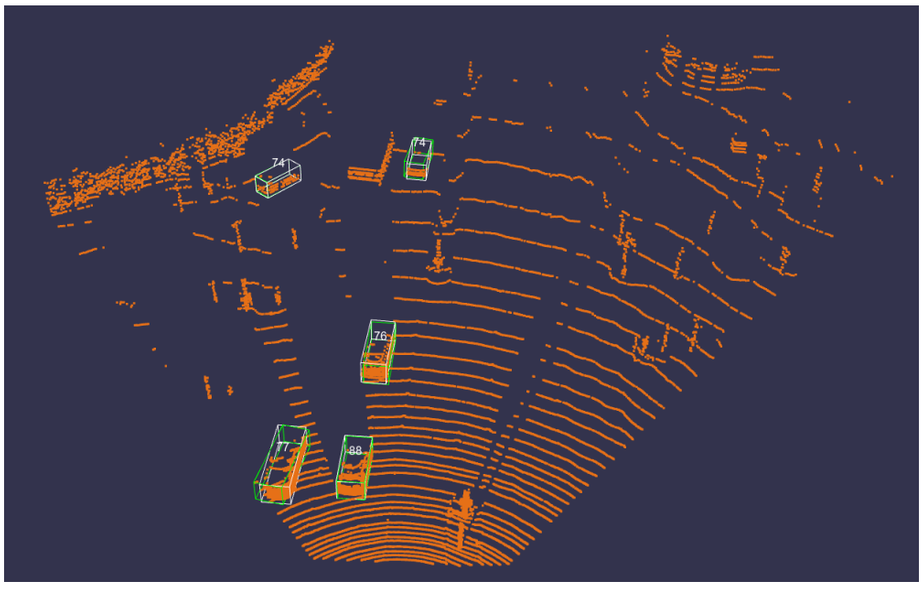
Figure 3: Task 1.6 Loss curve



(a) Scene visualization at the beginning



(b) Scene visualization in the middle



(c) Scene visualization at the end

Figure 4: Task 1.6 Scene visualization