Report

Team 15

Xiang Liu, Han Wang

1. Task 1

Average recall for the val-set:

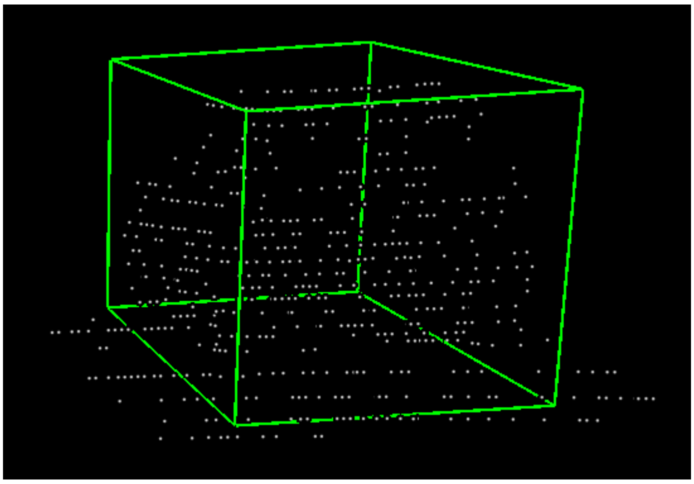
0.8134197351195016

Why is recall a good metric to assess the quality of the first stage proposals (as opposed to e.g. average precision)?

Firstly in terms of the average precision, it measures how many detections of the model correctly predicted over the amount of all the detection sets, while the average recall measures how many correct detections over the total amount of ground-truth detection. In the driving scenes, it is very important to reduce the amount of FN predictions because missing a car on the road can be dangerous. Here, recall measures how many cars/objects we might have missed. We are not too concerned with mistaking an object, which is not a safety hazard, for a car because safety always comes first. Actually, for the redundancy predictions, we could discard them in the following process.

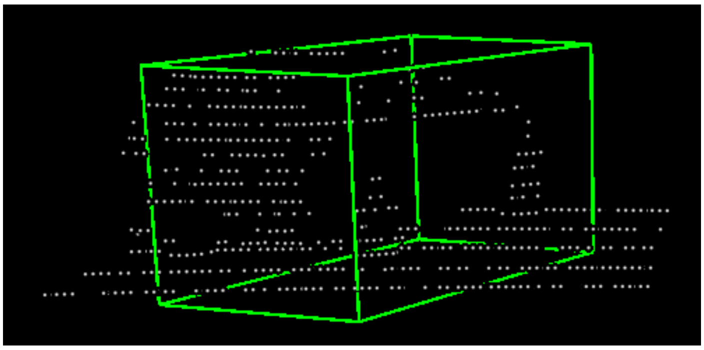
2. Task 2

Three examples of ROI figures:



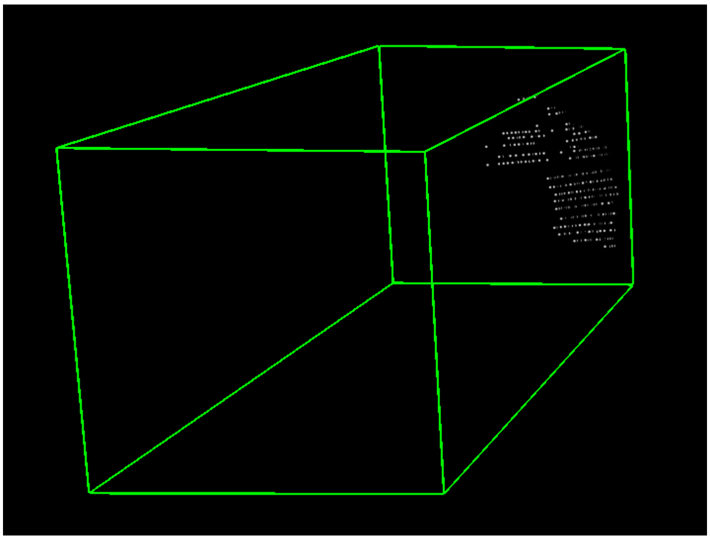
frame name: elpBlgmdIiJkAAVuYPB2

[3.4737 1.6519 10.9696 1.4765 1.5752 3.478 -1.5795]



frame name: rWvdEifFU9YpFeFyMroa

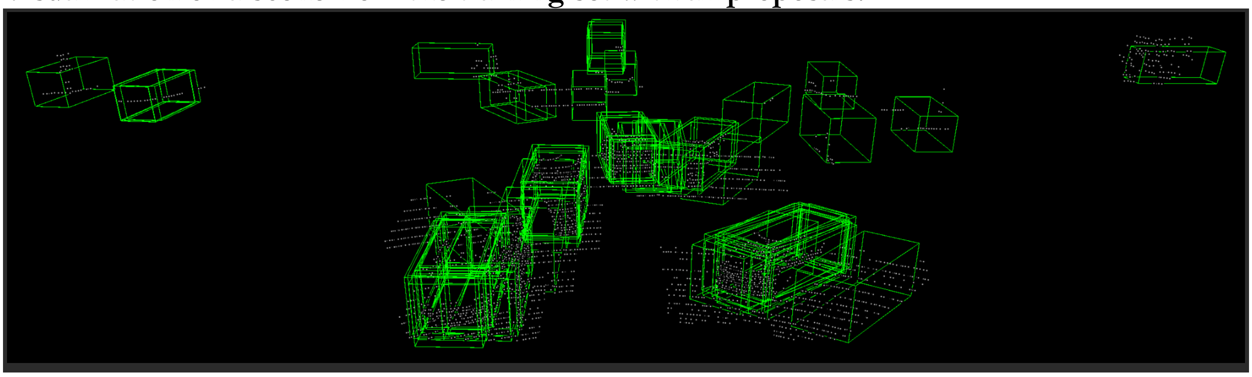
[-3.0632 1.9011 18.2856 1.6598 1.6711 4.0265 -1.208]



frame name: DXdxO5H7RXbfVtSqBVd8
[-4.0598, 1.7726, 3.364 , 1.6461, 1.6484, 4.1187, 1.5303]

3. Task 3

Visualization of a scene from the training set with all proposals:



frame name: 0lvf6yhDIQP6RikCciGf

Why is such a sampling scheme required:
In the input set, background samples dominate in number, so we have to hard sample in order to balance the class distribution of proposals.

What would happen if
1) randomly sample from all proposals for each scene?
Due to the imbalance of the input set, a lot of proposals will be background samples. The model trained with these proposals will be highly biased and predicts a lot of false negatives.

2) Don't sample an easy background proposal at all?
The easy negative samples help the network to converge quickly. Only sampling hard negatives will make it hard for the network to learn at an early stage.

3) consider a sample to be foreground if it's above 0.5 IoU and background otherwise?
As there's no margin between foreground and background, some foreground and

background samples may appear very similar, thus making it harder for the network to learn to classify and unstabilize the training, and slower the convergence.

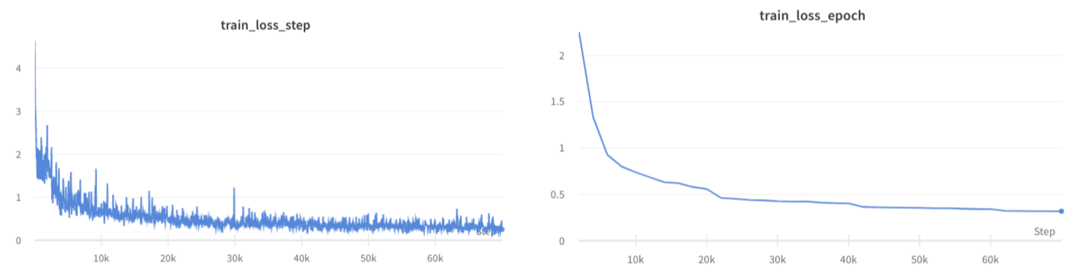
Why do we need to match the ground truth with its highest IoU proposal?
If there are no predictions matched with this ground truth with the highest IoU, then this ground truth won't be considered in the forward pass, and the network won't learn to generate predictions corresponding to this ground truth, thus this information in the ground truth is not surpervised in training, which damage the performance as the network generates incomplete predictions with respect to the ground truth. Matching the ground truth with its highest IoU proposal will ensure all ground truths are learned.

4. Task 5

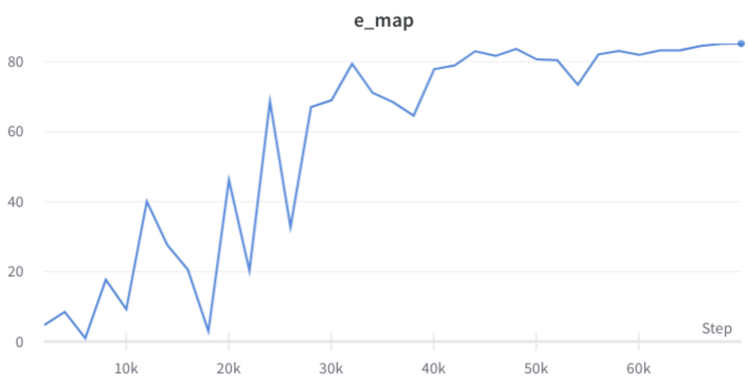
Must the IoU be calculated on BEV or can it also be calculated on 3D? What advantages does the 2D BEV IoU have over 3D (or vice versa) for NMS?
From a theoretical perspective, it can also be calculated on 3D, but with diminished effect. If a box is perfectly aligned with the proposal with the highest IoU in width and length, but only has 1/10 the height of the proposal, then this box will not be suppressed if we calculate IoU in 3D, and thus results in two highly overlapping bounding box for one object. However, if we calculate 2D IoU, it will be 1, and thus be suppressed, which is more catering to our needs in this operation, and result in cleaner predictions.

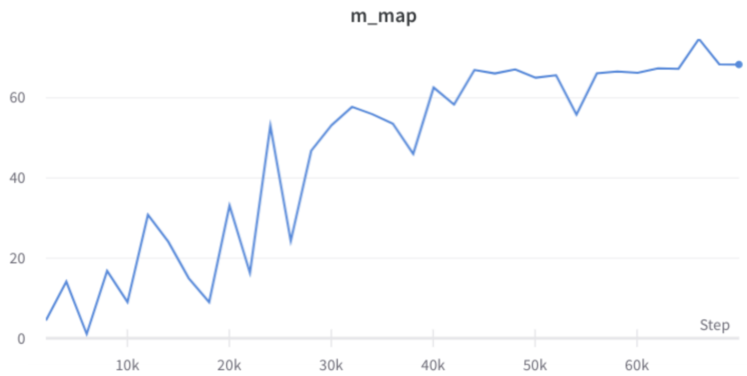
5. Training

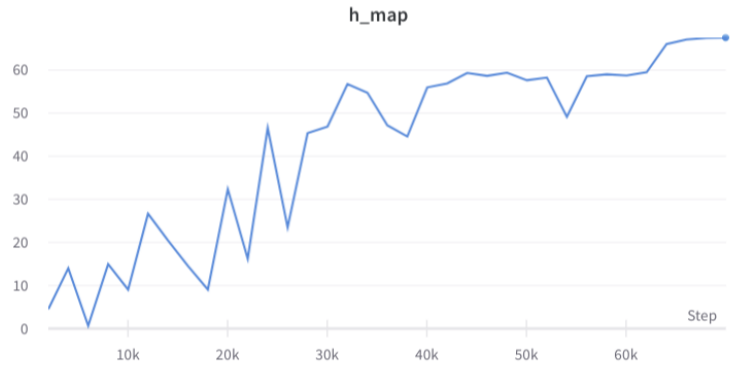
The loss- and precision curves along with 3 example scene visualizations from the beginning, mid-way, and end of the training cycle.



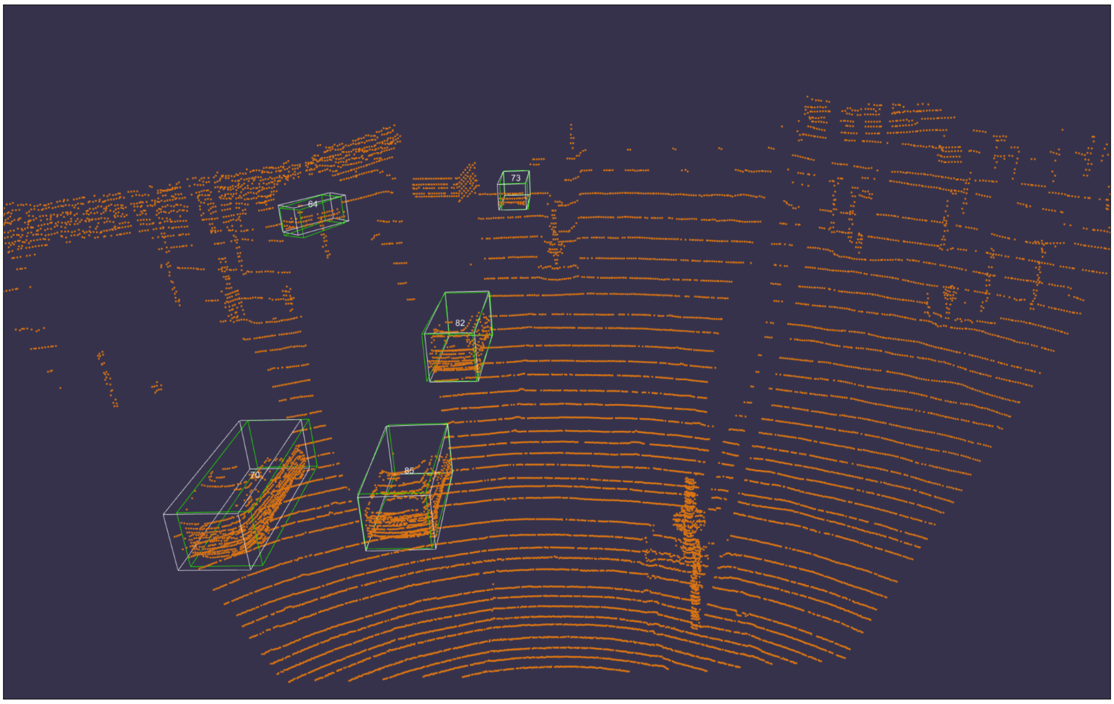
Training loss goes down gradually.





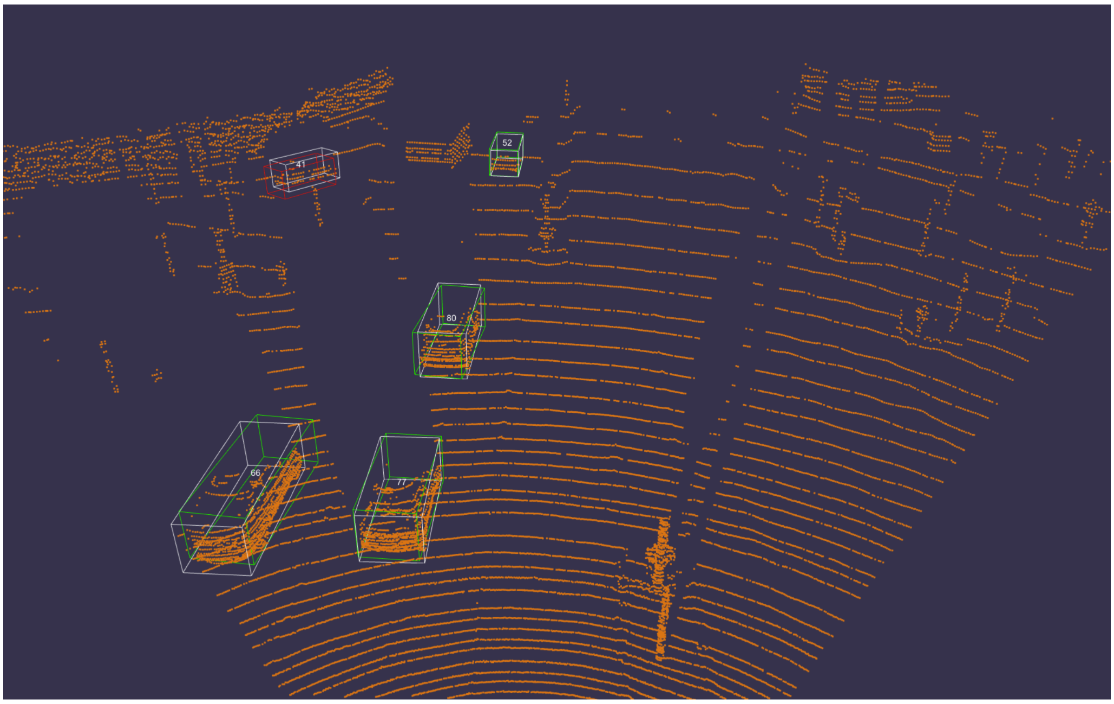


The mAP is unstable at the early stage of training and quickly ramps up as training progressing then saturates at the end.

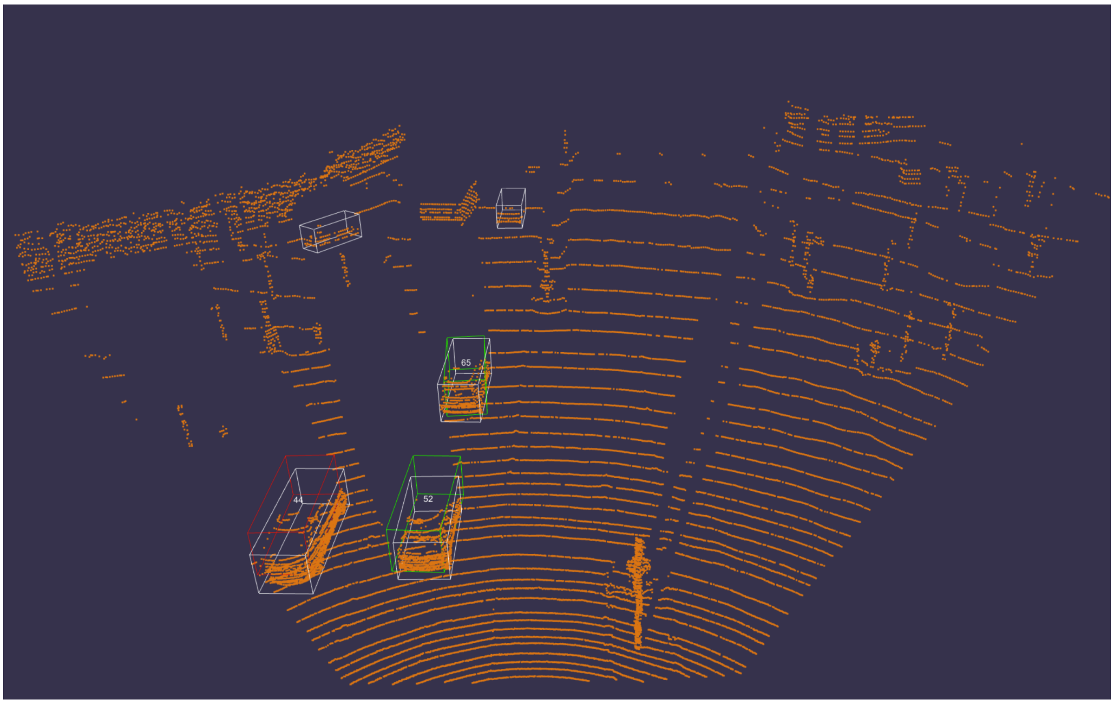


35th epoch

All the ground truths are predicted with high IoU.



20th epoch
Detect closer objects well with high IoU, though failed to predict with high IoU for farther objects.



4th epoch
Predict closer objects with low IoU, and didn't even propose farther predictions.

Submitted to CoadLab, the test metrics are Easy: 86.05, Moderate: 74.57, Hard: 67.41.