CIS 680: Advanced Topics in Machine Perception — Fall 2020

Homework 2b

Due: Oct 4

Name: SHUBHAM GUPTA

PennKey: shubi

PennID: 89380662

7.5

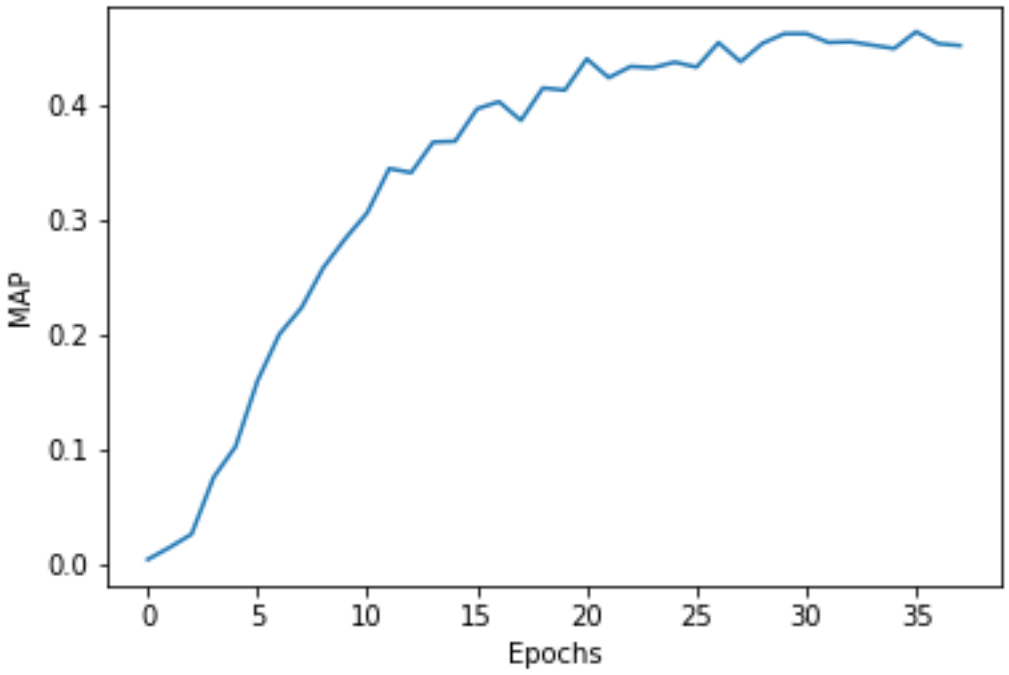

Figure 1: Mean Average Precision over training epochs

7.6

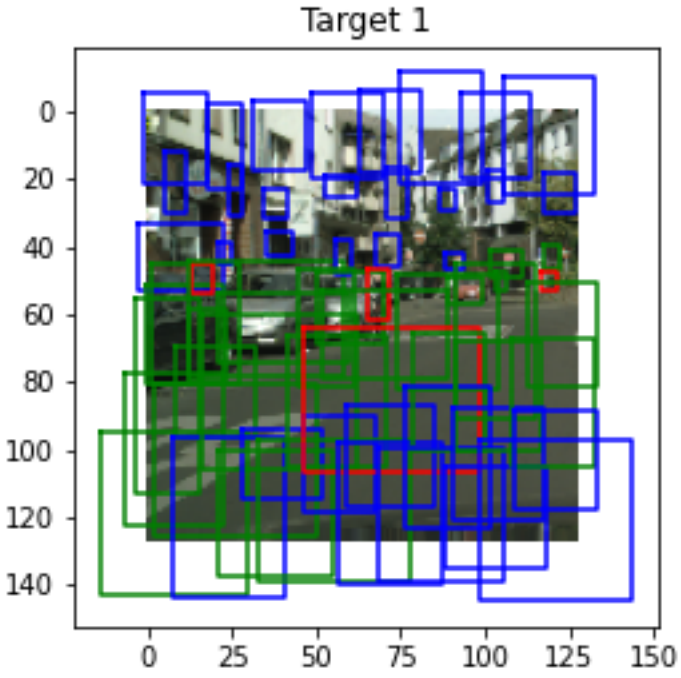

Figure 2: Bounding boxes before any elimination

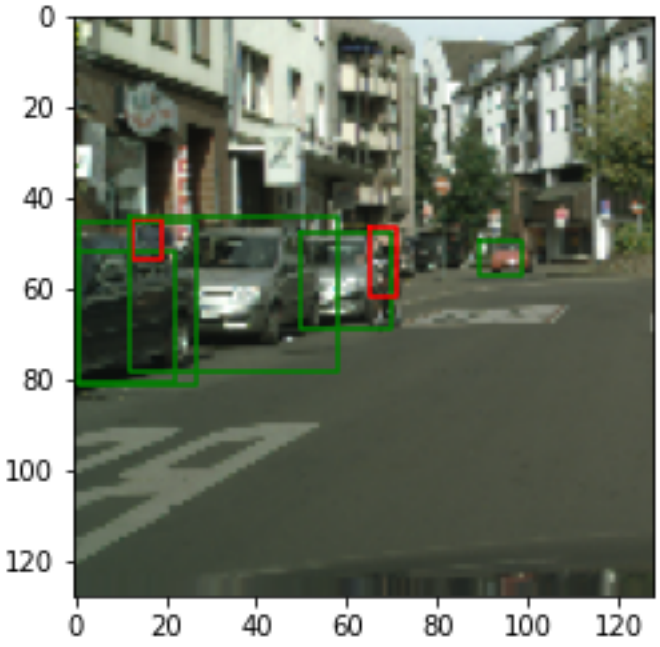

Figure 3: Bounding boxes after suppressing low confidence boxes

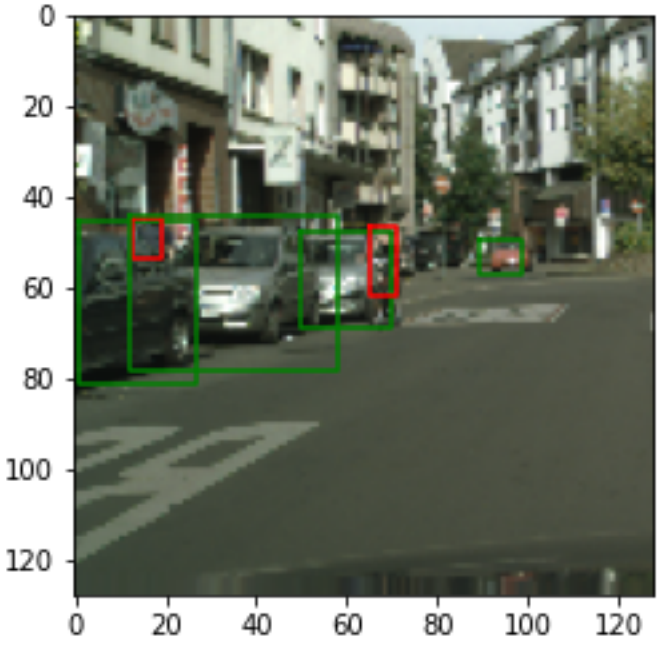

Figure 4: Bounding boxes after non-maximum suppression

7.7

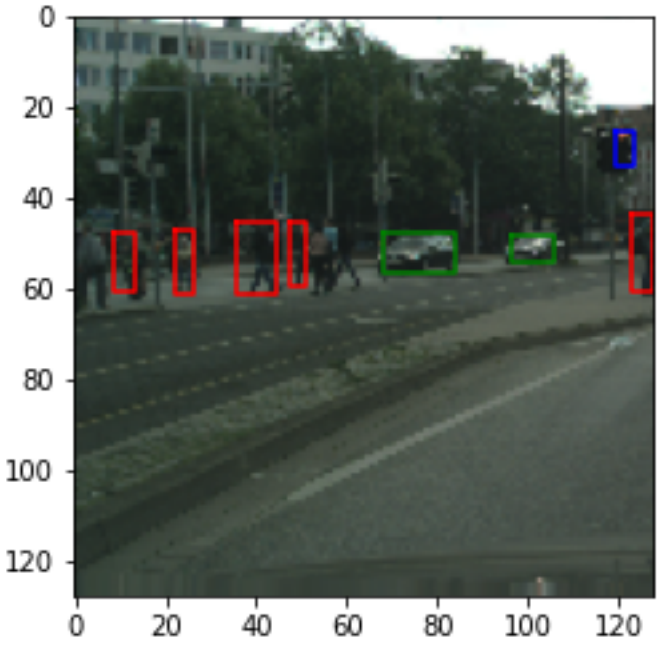

Figure 5: Bounding box visualized

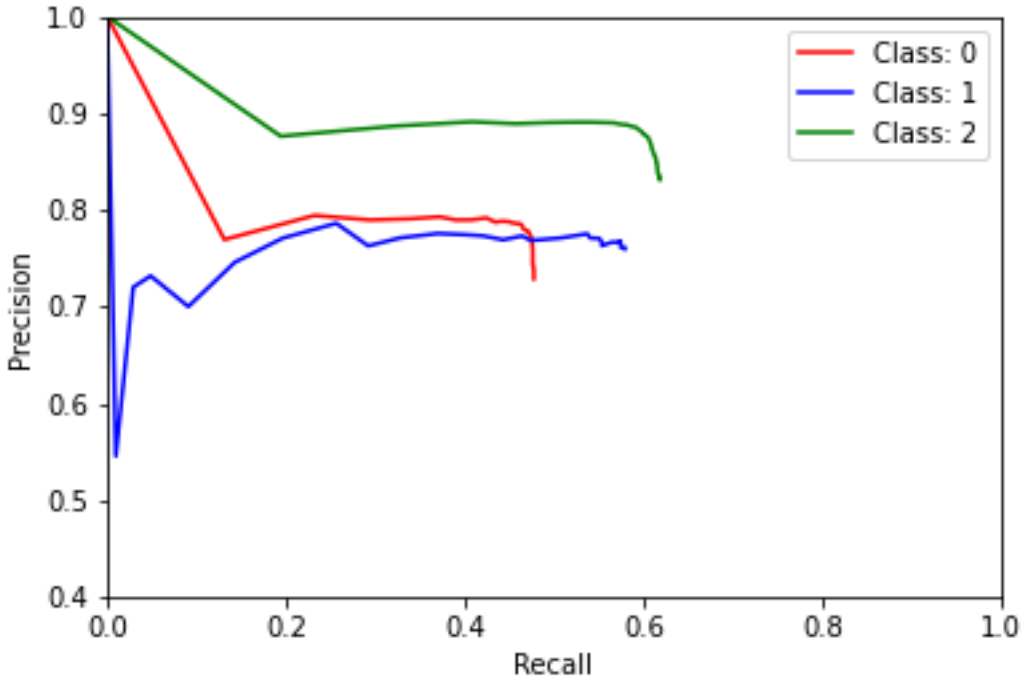

Figure 6: Precision-recall curves for each class

Reported Mean Average Precision = 0.462

7.8

One major issue I faced while implementing this project was to get the model to detect traffic lights.The dataset is imbalanced with 37000 instances of cars, 16000 instances of pedestrians and only 2800 instances of traffic lights. On top of that, the bounding boxes for traffic lights are quite small in dataset.

In my understanding, in the initial stages it is hard for the predicted boxes to have high IoU with the ground truth boxes. The confidence loss aims to reduce the gap between the confidence and IoU. So in the initial stages, maybe the network learns that if it's a traffic light, it should have low confidence.

Ideally, this should be fixed by the localization optimization as the network trains on. But in this case, the dataset for traffic lights is quite small so the network localization optimization may not be getting enough training instances to correct the localization to an extent where the IoU is greater than 0.6.

The failure of network to detect traffic lights is also shown by the average precision values of the 3 classes:

1. Pedestrian: 0.289
2. Traffic light: 0.034
3. Car: 0.533

The network works much better when the rather than using IoU, a target of 1 is used in confidence loss. The precision values for this case are:

1. Pedestrian: 0.389
2. Traffic light: 0.438
3. Car: 0.558

Another observation: Different losses converge at different rate. This is shown in Figure 7 and 8 on next page. The confidence loss decrease slowly and results in a lower level than other losses. Maybe the network will work better if rather than using a single optimizer on the sum, four different optimizers are used to optimize the losses individually.

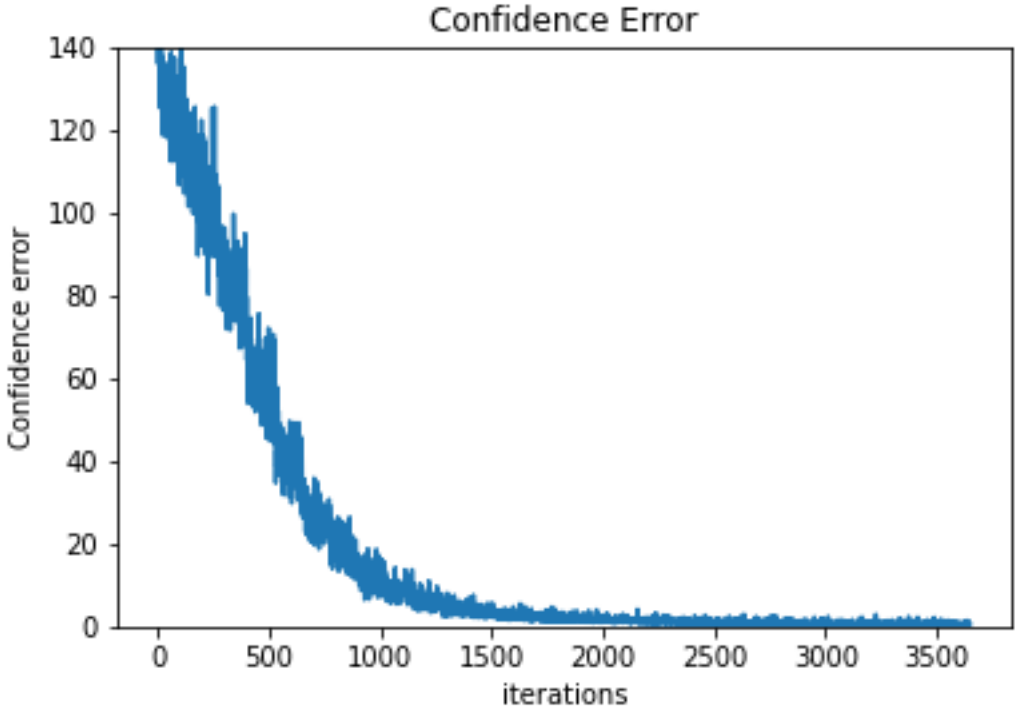

Figure 7

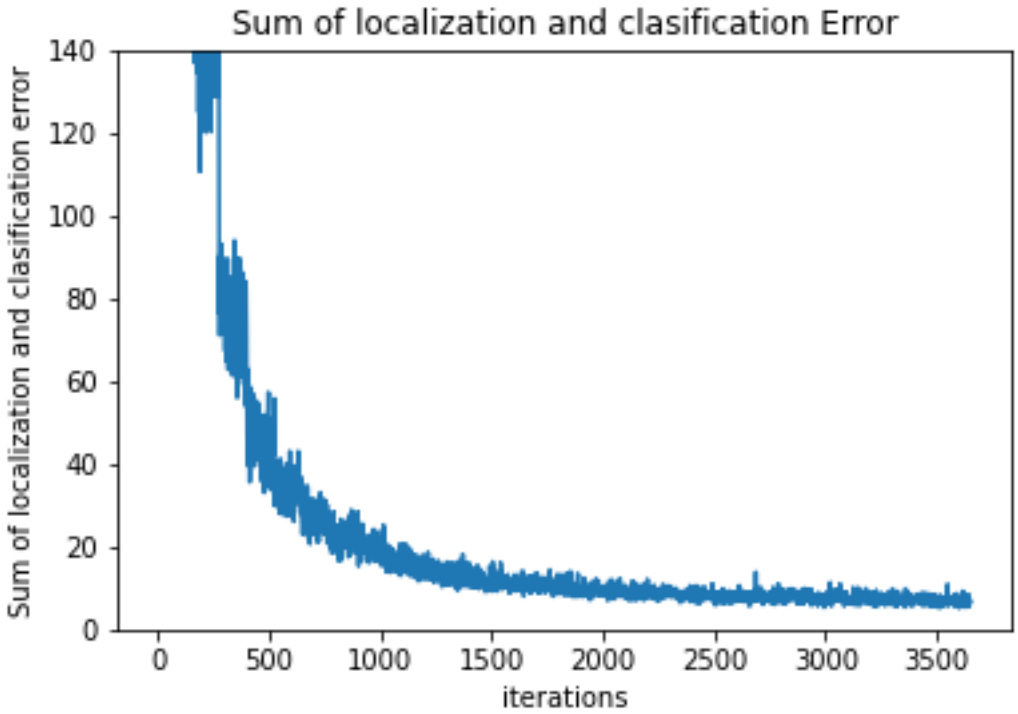

Figure 8

Hypothesis to improve further:

- For the case when IoU is used in confidence loss, we want the network to correct localization for traffic lights quickly. So a higher penalty can be added in localization loss for the case when the object is traffic light.
- Another workaround can be to ignore the IoU for confidence loss for initial stages and use target as 1 instead. After a few epochs once the localization is better, IoU can be put back.
- Lastly, the network will learn much better if the dataset is balanced and has enough instances of every class. So just expanding the dataset to include more images of traffic lights and pedestrians should improve the model.