# DL Lab 4

# IT21271182

## Question 7

The reason there are 1783 boxes is that the YOLO model divides the input image into a grid, typically 19x19 cells in this case, with each cell predicting 5 bounding boxes due to the presence of 5 anchor boxes. Initially, this results in 19x19x5, which equals 1805 boxes. However, the yolo\_filter\_boxes function then filters these boxes based on a confidence threshold, discarding boxes with scores below the threshold. After this filtering process, 1783 boxes remain, each represented by four coordinates: width, height, x-center, and y-center.

The maximum number of boxes occurs when the confidence scores of all the initially predicted boxes exceed the threshold, resulting in the full 1805 boxes being retained. Conversely, the minimum number of boxes is achieved when none of the boxes meet the confidence threshold, leaving zero boxes. Therefore, the possible range of boxes after filtering can vary from 0 to 1805.

## Question 8

Anchor boxes offer a significant benefit by allowing the model to evaluate all potential object predictions simultaneously, rather than relying on a sliding window approach that would require making separate predictions for each possible location within the image. This efficiency is achieved because anchor boxes provide predefined shapes and sizes, enabling the model to predict objects of various dimensions in one pass through the image, which greatly speeds up the detection process.

The sizes of anchor boxes are typically determined through a combination of manual selection and analysis of the dataset. A common method involves applying a clustering algorithm, such as K-Means, to the dataset. This algorithm helps identify the most common bounding box shapes and sizes across the dataset, which then serve as the anchor boxes. By tailoring the anchor boxes to the specific characteristics of the data, the model can more accurately predict the objects it encounters.

## A crosswalk on a street Description automatically generated0114.jpgQuestion 10

For this, no bounding boxes were detected despite the presence of vehicles and traffic lights. This lack of detection could be attributed to the specific shape of the vehicle, which may not have aligned well with the anchor boxes or model predictions. Additionally, the traffic lights may have gone undetected due to lighting conditions in the image, which could have affected the model's ability to recognize them.

In the second image, both a vehicle and a traffic light were successfully detected. This observation supports the earlier assumptions, as the vehicle was detected even though only half of it was visible, suggesting that the shape or partial visibility in the previous image may have indeed influenced the detection results. Moreover, the detection score for the traffic light in this image was very close to the threshold, which may explain why traffic lights in other images, with less favorable conditions, were not detected.

## Question 11

max\_boxes = 30



Despite adjusting the max\_boxes value, there appears to be no noticeable improvement.

score\_threshold = 0.3

*A street with traffic signs

Description automatically generated*

By reducing the score threshold from 0.6 to 0.3, the model has detected more objects, including both traffic lights and vehicles. However, in the first image, it incorrectly labeled one object as a "bus" when it should have been identified as a "truck."

iou\_threshold = 0.8 and iou\_threshold = 0.2 on 2 separate runs

*A street with a green light

Description automatically generated*

I adjusted the IoU threshold values to 0.8 and 0.2 and tested the model in two separate runs, but there was no significant improvement in the results.