TDT4265 - Computer Vision & Deep Learning

Assignment 4 Report - Group 66

> Dionysios Rigatos
> dionysir@stud.ntnu.no

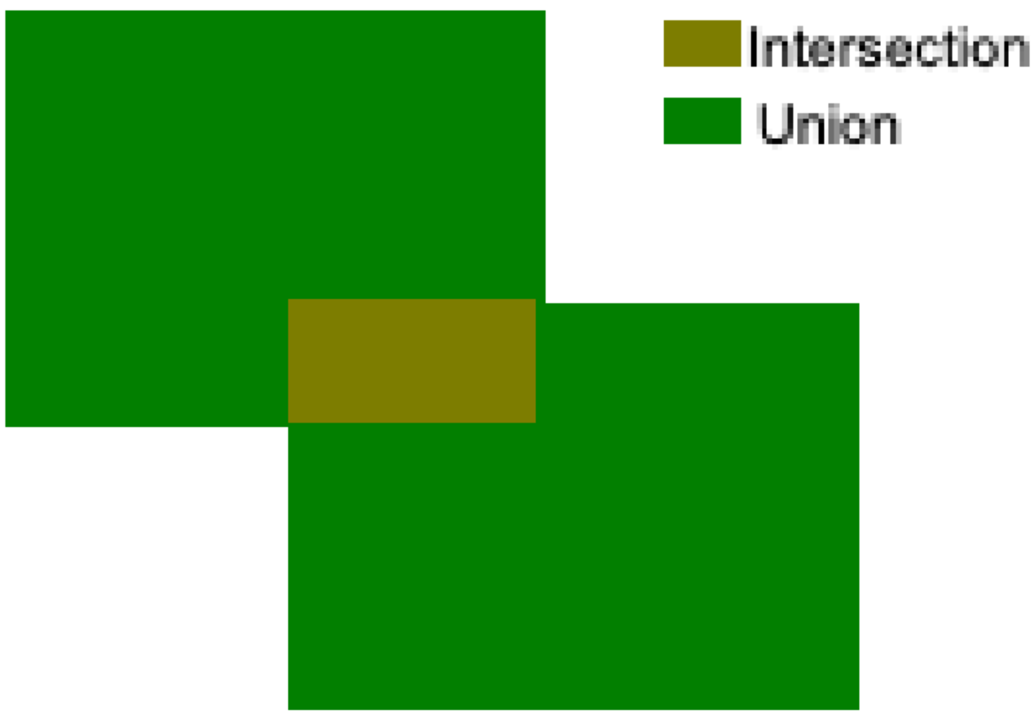
Task 1

Task 1a)

Intersection over Union is a metric that measures the overlap between two bounding boxes (usually between our prediction against a ground truth) in object detection models.

The formula for calculating IoU for two bounding boxes X and Y is:

$IoU = \frac{area(X \cap Y)}{area(X \cup Y)}$



We can see the area of intersection and the area of union. Of course, this would be a bad example of IoU, since the intersection is very small compared to the union.

Task 1b)

We have:

- $Precision = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+FN}$

Where TP is the amount of $TruePositives$, FP is the amount of $FalsePositives$ and FN is the amount of $FalseNegatives$.

A $TruePositive$ is a prediction that was correctly assigned a label/bounding box - in our case, a bounding box with IoU >= threshold.

A $FalsePositive$ is a prediction that was incorrectly assigned a label/bounding box - in our case, a bounding box with IoU < threshold.

Task 1c)

Given the following precision and recall curve for the two classes, what is the mean average precision? Precision and recall curve for class 1: Precision1 = [1.0, 1.0, 1.0, 0.5, 0.20] Recall1 = [0.05, 0.1, 0.4, 0.7, 1.0] Precision and recall curve for class 2: Precision2 = [1.0, 0.80, 0.60, 0.5, 0.20] Recall2 = [0.3, 0.4, 0.5, 0.7, 1.0] Hint: To calculate this, find the precision for the following recall levels: 0.0, 0.1, 0.2, ... 0.9, 1.0.

In order to find the mAP (mean Average Precision) we'll calculate the precision for the recall interval [0, 1] with step 0.1.

- For Class_1, we have the following precisions per interval:
  - Recall 0.0 - Precision 1.0
  - Recall 0.1 - Precision 1.0
  - Recall 0.2 - Precision 1.0
  - Recall 0.3 - Precision 1.0
  - Recall 0.4 - Precision 1.0
  - Recall 0.5 - Precision 0.5
  - Recall 0.6 - Precision 0.5
  - Recall 0.7 - Precision 0.5
  - Recall 0.8 - Precision 0.2
  - Recall 0.9 - Precision 0.2
  - Recall 1.0 - Precision 0.2
- So the AP for Class_1 is around 0.65.

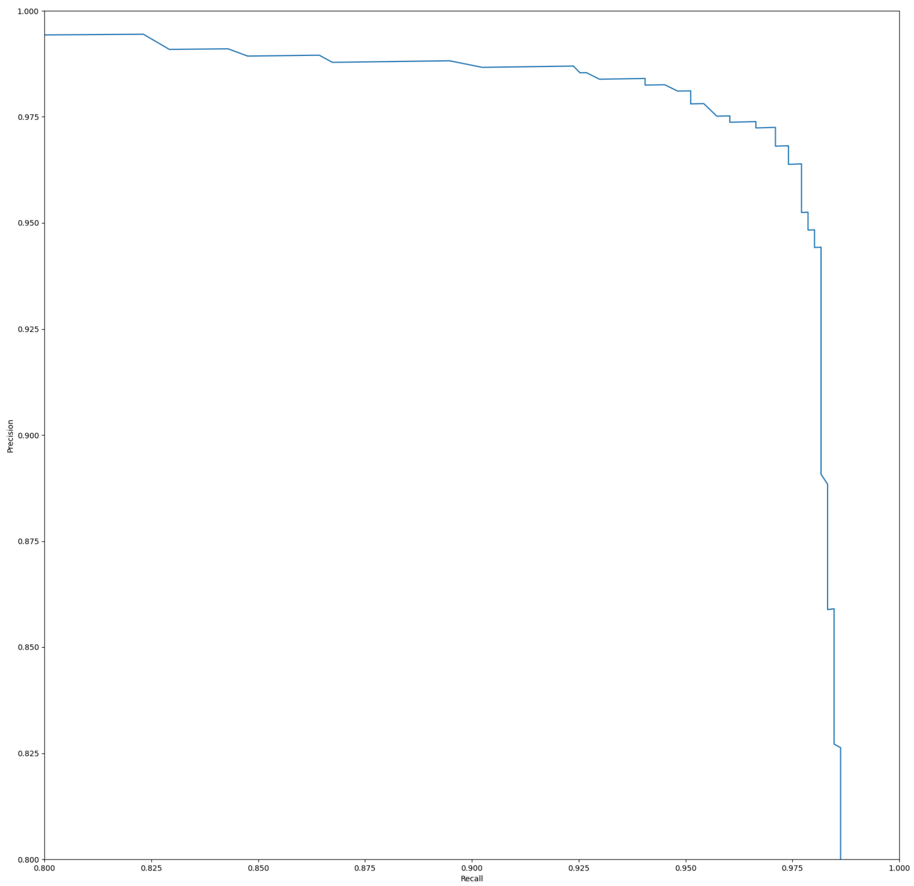
- For Class_2, we have the following precisions per interval:
  - Recall 0.0 - Precision 1.0
  - Recall 0.1 - Precision 1.0
  - Recall 0.2 - Precision 1.0

    - Recall 0.3 - Precision 1.0
    - Recall 0.4 - Precision 0.8
    - Recall 0.5 - Precision 0.6
    - Recall 0.6 - Precision 0.6
    - Recall 0.7 - Precision 0.5
    - Recall 0.8 - Precision 0.2
    - Recall 0.9 - Precision 0.2
    - Recall 1.0 - Precision 0.2
- So the AP for Class_2 is around 0.71.

So the mAP is the average of the APs for each class, which is around 0.68.

Task 2

Task 2f)



Task 3

Task 3a)

Picking the best box for our ground-truth label requires a matching strategy that will maximize the IoU between the predicted bounding box and the ground-truth bounding box. The best box is the one that maximizes the IoU.

Task 3b)

False. The input image's resolution is higher in earlier layers, thus the bounding boxes capture a smaller area of the picture - thus detecting objects of smaller sizes. As the layers progress, the resolution decreases and one bounding box is able to capture more information - thus detecting larger objects.

Task 3c)

By using different bounding box aspect ratios we allow the model to capture a larger variety of objects. Using a single aspect ratio, such as a square, would inhibit the model's ability to detect objects that are wider or taller - such as cars or people.

Task 3d)

SSD eliminates region proposal networks by using a fixed set of bounding boxes at different scales and aspect ratios unlike YOLO which uses a single bounding box for each grid cell. YOLO also works on a single scale, while SSD uses multiple scales to detect objects of different sizes.

Task 3e)

If the feature map is of size 38x38 and the number of default boxes is 6, then the total number boxes is 38x38x6 = 8664 for this feature map.

Task 3f)

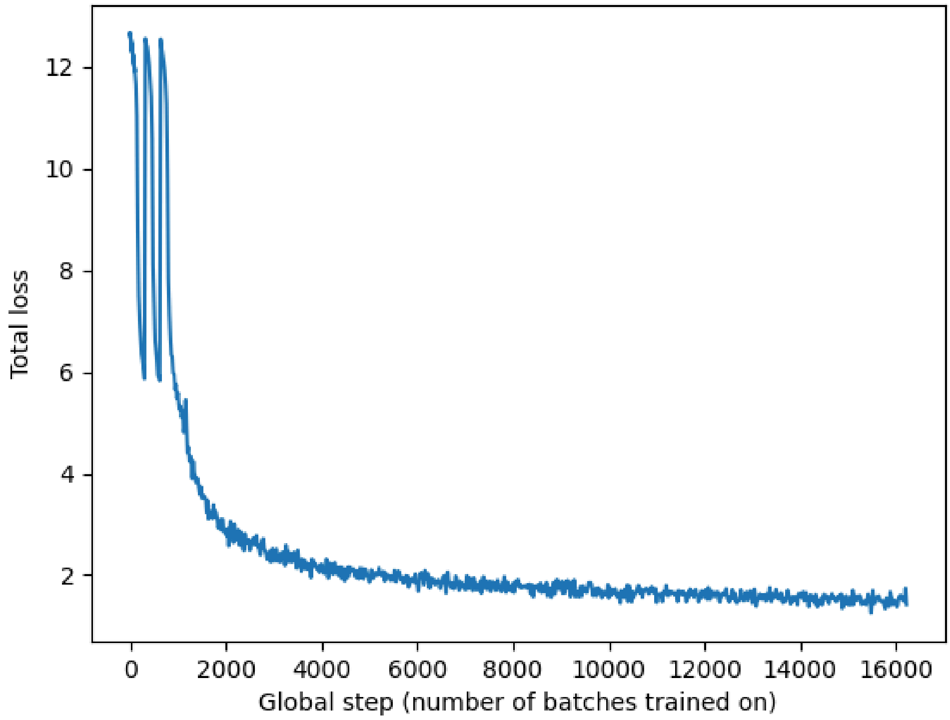
For each feature map we have:

- 38x38x6 = 8664 boxes
- 19x19x6 = 2166 boxes
- 10x10x6 = 600 boxes
- 5x5x6 = 150 boxes
- 3x3x6 = 54 boxes
- 1x1x6 = 6 boxes

So the total number of boxes is 8664 + 2166 + 600 + 150 + 54 + 6 = 11640 boxes.

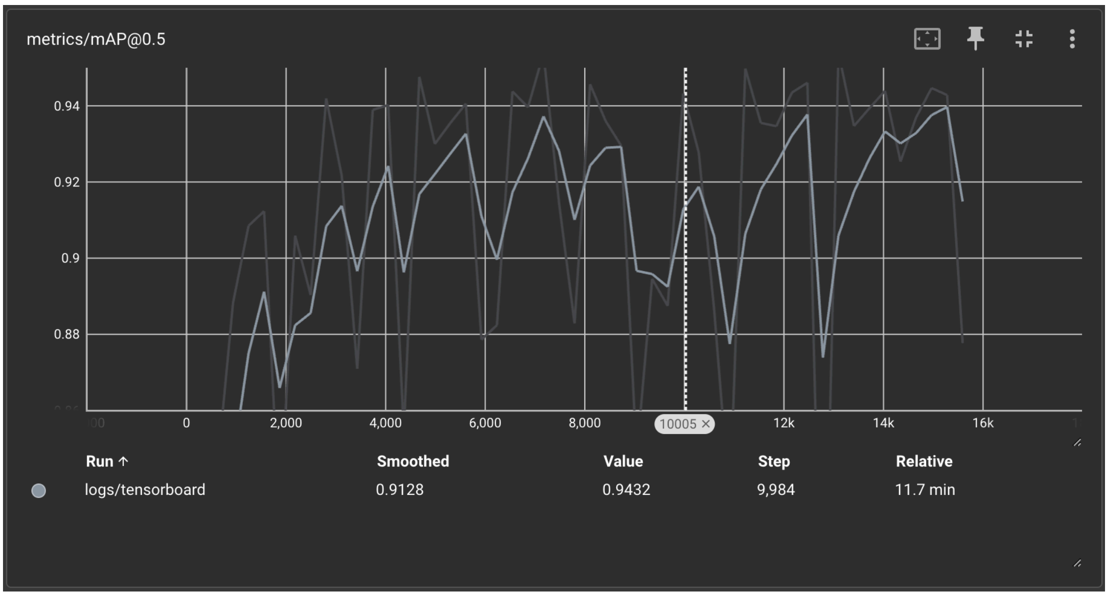
Task 4

Task 4b)



We achieved a $mAP@0.5$ of 0.791.

Task 4c)



The final achieved $mAP@0.5$ was approximately 0.9128.

Improvements done were:

- Used batch normalization.
- Added another, larger feature map (76x76) for detecting smaller objects.

- Used PReLU activation function instead of ReLU.
- Adam optimizer with a learning rate half of the original one.

No augmentation was used.

Task 4d)

There was no time to implement the extra task. However, I have already completed all the mandatory assignments with the required grade (75% on 3/4) so it should not be an issue.

Task 4e)

There was no time to implement the extra task. However, I have already completed all the mandatory assignments with the required grade (75% on 3/4) so it should not be an issue.