TDT4265 - Computer Vision and Deep Learning

Assignment 4

Written Spring 2020 By Dinosshan Thiagarajah and Kyrre S. Haugland

Task 1: Object Detection Metrics (0.5 point)

Task 1a:

IoU measures the overlap between 2 images by dividing the intersecting area by union of both images. It is used to measure how much our predicted image boundary overlaps with the ground truth image.

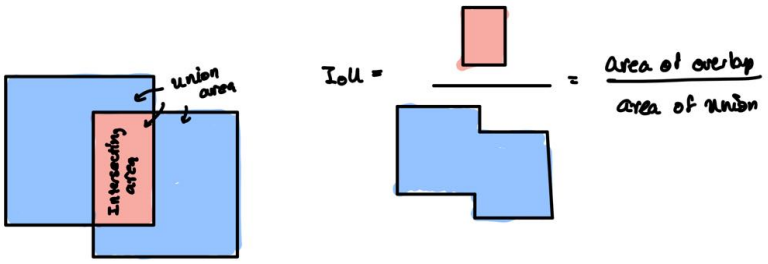


Figure 1: IoU illustration by hand

Task 1b:

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

where TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative.

True Positive in object detection is when there is an object, and one predicts that there is an object. False Positive on the other hand is when we predict that there is an object, but it's actually no object. Examples of these two are illustrated below.

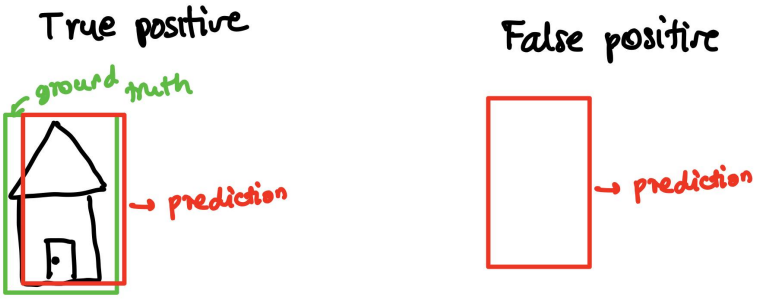


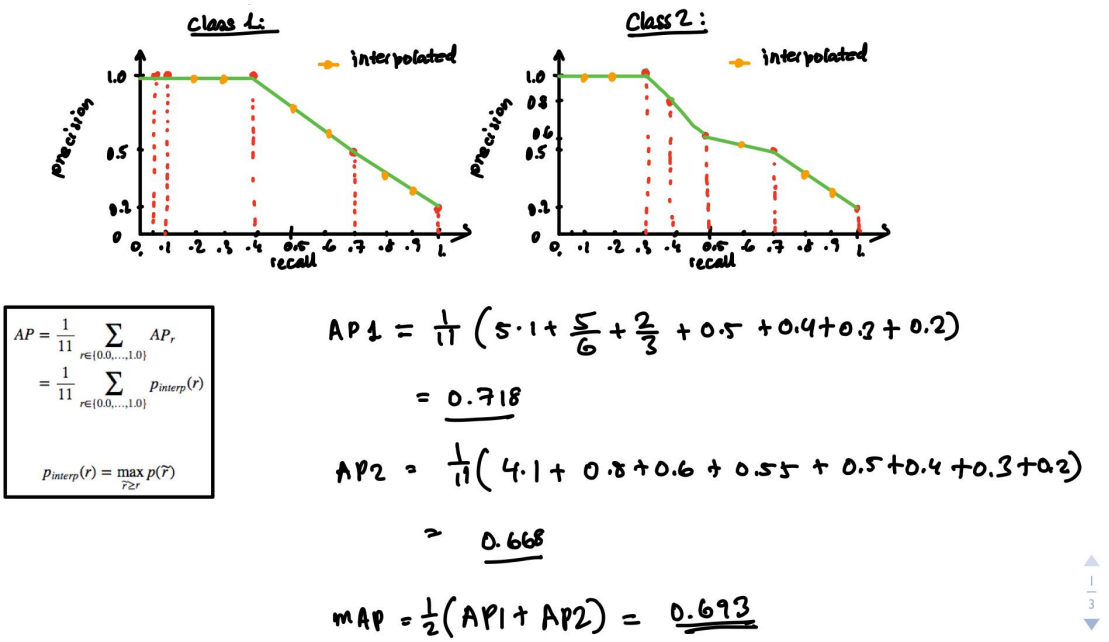
Figure 2: True Positive vs. False Positive illustration by hand

Task 1c:

In figure 3 we illustrate how to calculate the mAP. Here we first start of by interpolating the recall/precision-curve according to the formula for ρ_{interp}, and then calculate the *average precision*

according to the AP formula. Both of these formulas are shown in figure 3. Equation 1 is then used to calculate mAP, which is found to be 0.693.

$$mAP = \frac{1}{|classes|} \sum_{c \epsilon classes} \frac{TP(c)}{TP(c) + FP(c)} \tag{1}$$



$$AP = \frac{1}{11} \sum_{r \in \{0.0, \ldots, 1.0\}} AP_r$$
$$= \frac{1}{11} \sum_{r \in \{0.0, \ldots, 1.0\}} p_{interp}(r)$$
$$p_{interp}(r) = \max_{\tilde{r} \geq r} p(\tilde{r})$$

$AP1 = \frac{1}{11}\left(5 \cdot 1 + \frac{5}{6} + \frac{2}{3} + 0.5 + 0.4 + 0.3 + 0.2\right)$

$= 0.718$

$AP2 = \frac{1}{11}\left(4 \cdot 1 + 0.8 + 0.6 + 0.55 + 0.5 + 0.4 + 0.3 + 0.2\right)$

$= 0.668$

$mAP = \frac{1}{2}(AP1 + AP2) = 0.693$

Figure 3: AP and mAP calculation using 11-point interpolated AP

Task 2: Implementing Mean Average Precision (2 points)

Task 2:

Mean average precision: 0.9066

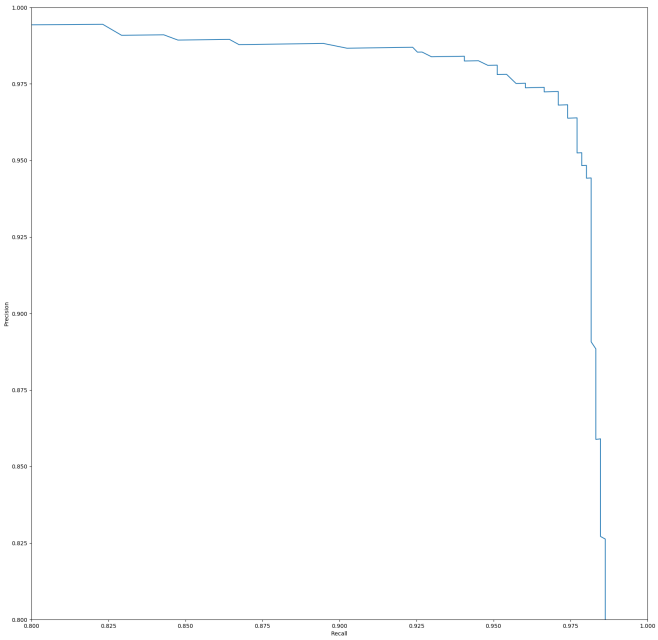


Figure 4: Precision-recall curve

Task 3: Theory (0.5 point)

Task 3a:

The final step used to filter out overlapping boxes is called **non-maxima suppression**.

Task 3b:

The statement: "Predictions from the deeper layers in SSD are responsible to detect small objects" is **false**.

Task 3c:

Using different aspect ratios for the bounding boxes will allow the model to detect more kinds of objects. Different objects have significantly different aspect ratios. A pedestrian for instance is captured by a tall and thin box, while a car is captured by a very wide box. This will in essence make the training much easier and more stable.

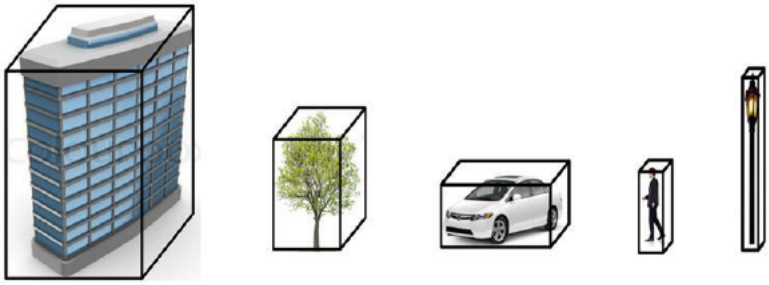


Figure 5: Bounding boxes for buildings, trees, cars, pedestrians and poles

Task 3d:

The difference between SSD and YOLOv1/v2 is that SSD are able to locate objects at different levels of feature maps meaning each layer is specialized at different scales. YOLO on the other hand is not able to do this. It operates on a single scale feature map.

Task 4: Single Shot Detector (3 points)

Task 4b:

The resulting is mAP = 0.7744 after 6000 iterations of training.

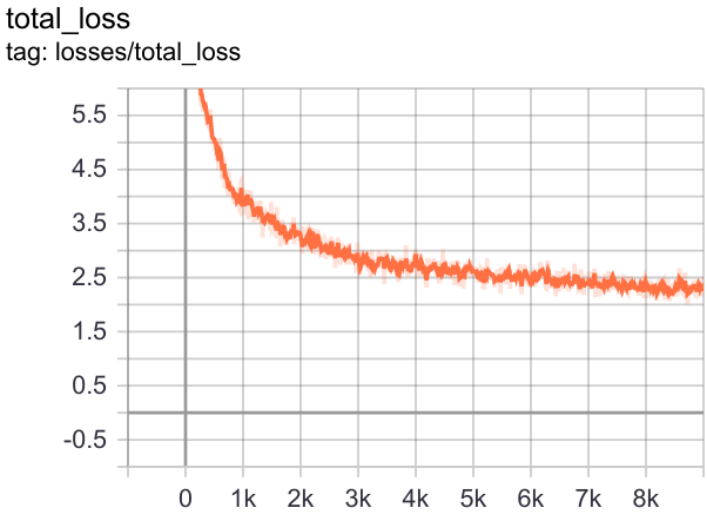


Figure 6: Total loss plot

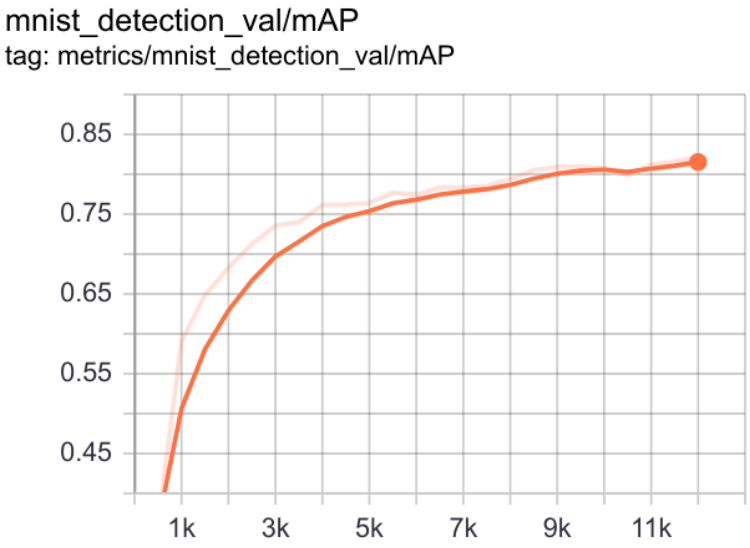


Figure 7: mAP plot

Task 4c

Implementing following modifications/improvements:

- Increased learning rate from $1 * 10^{-3}$ to $3 * 10^{-3}$
- Added batch normalization at the end of each convolutional layer.
- Activation function changed from ReLU to LeakyReLU
- Increased batch size from 16 to 32

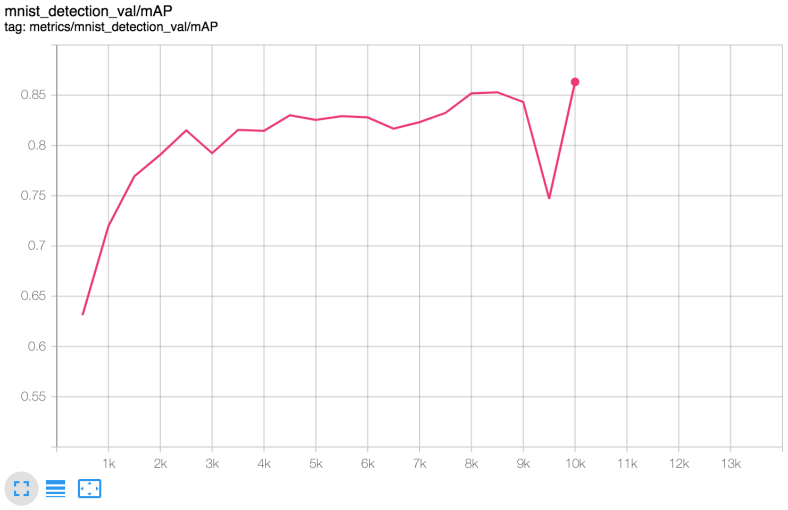


Figure 8: mAP plot with implemented modifications described above

Task 4d

Task 4e

Was there any digits your model was not able to detect? Yes, as we can see from the results in figure 9,10 and 11 that the SSD is having problem detecting very small numbers, as expected. In addition, we can from figure 10 also see that it is struggling dividing the numbers 8 and 9 which is in a cluster.

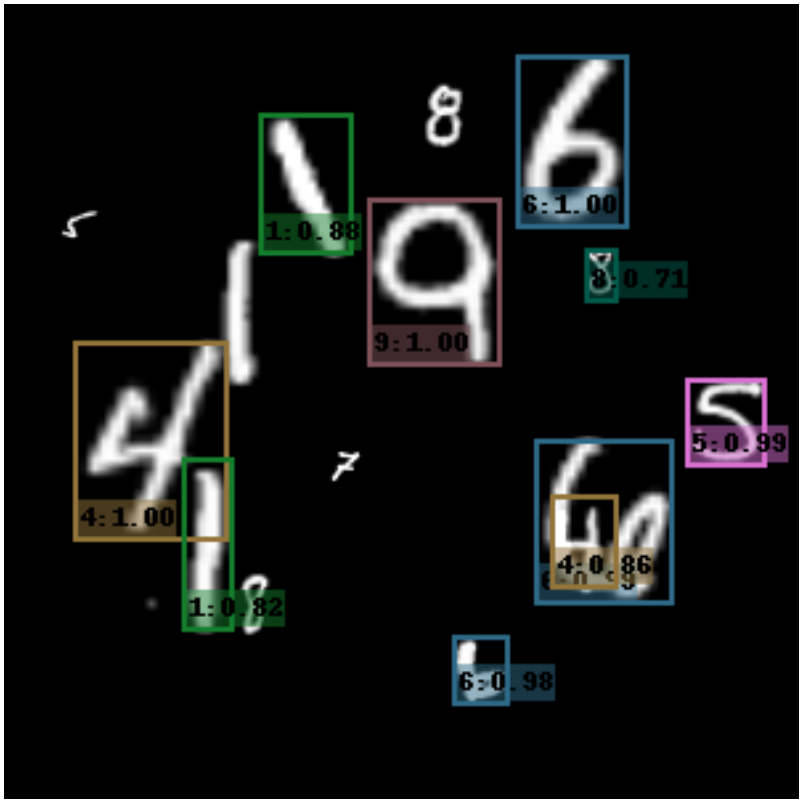


Figure 9: Digits detection, classification and localization of image 6

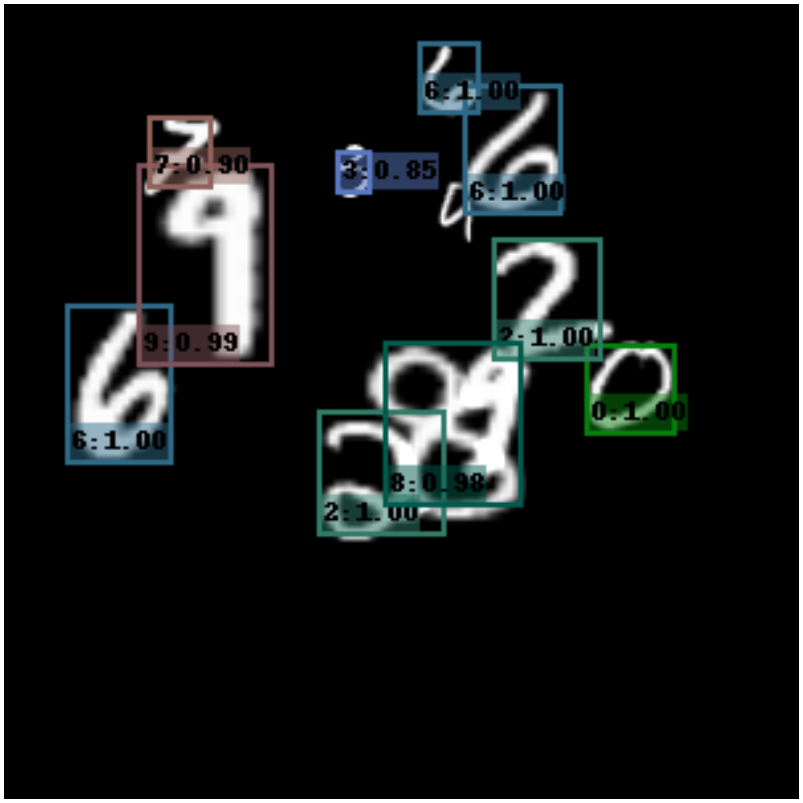


Figure 10: Digits detection, classification and localization of image 8

Figure 11: Digits detection, classification and localization of image 12

Task 4f

The final mAP is 0.5603 after 5000 iterations.

```
100%|
2020-03-21 14:28:56,572 SSD.inference INFO: mAP: 0.5603
aeroplane       : 0.6093
bicycle         : 0.6826
bird            : 0.5766
boat            : 0.3474
bottle          : 0.2361
bus             : 0.6237
car             : 0.7310
cat             : 0.7495
chair           : 0.3480
cow             : 0.5511
diningtable     : 0.4776
dog             : 0.7149
horse           : 0.6872
```

Figure 12: Final mAP

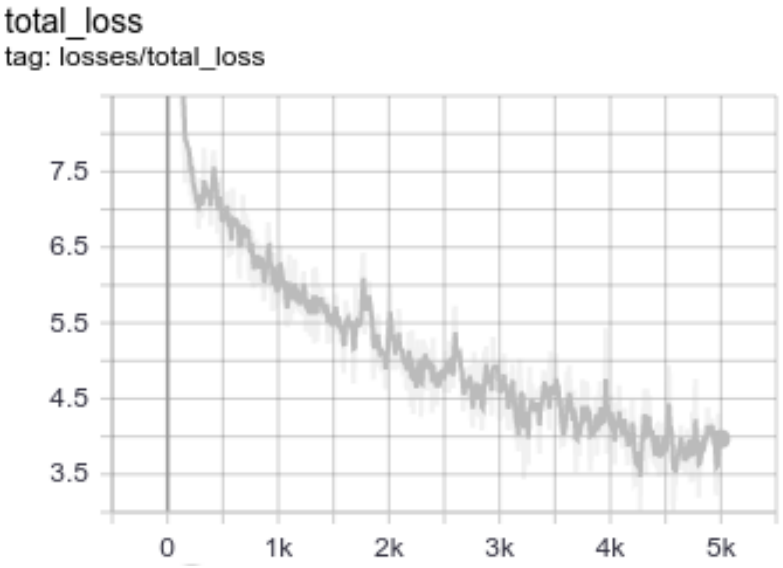


Figure 13: Total loss

Figure 14: Image classified by SSD network with VGG16 as the backbone

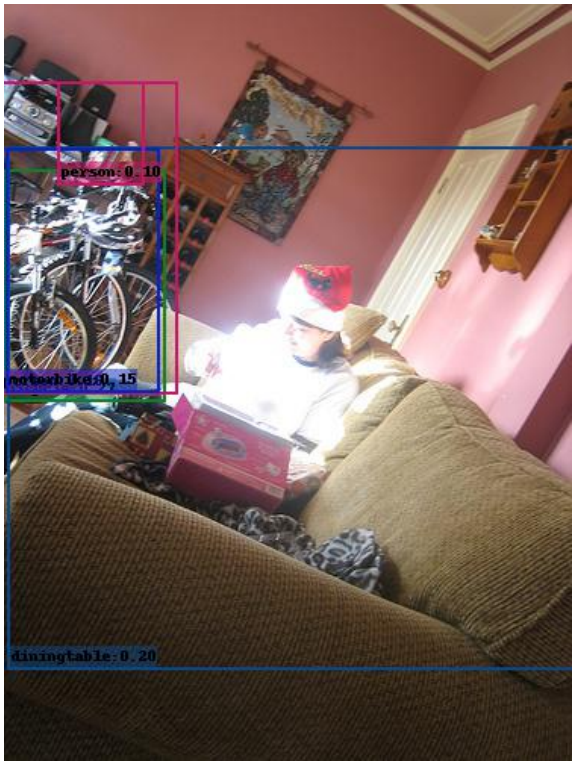


Figure 15: Image classified by SSD network with VGG16 as the backbone