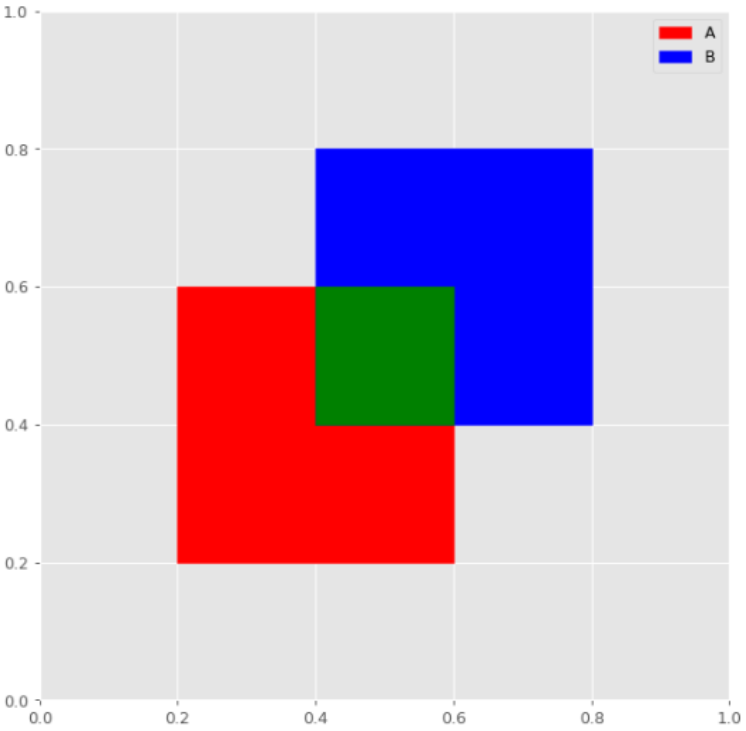
Quiz 3



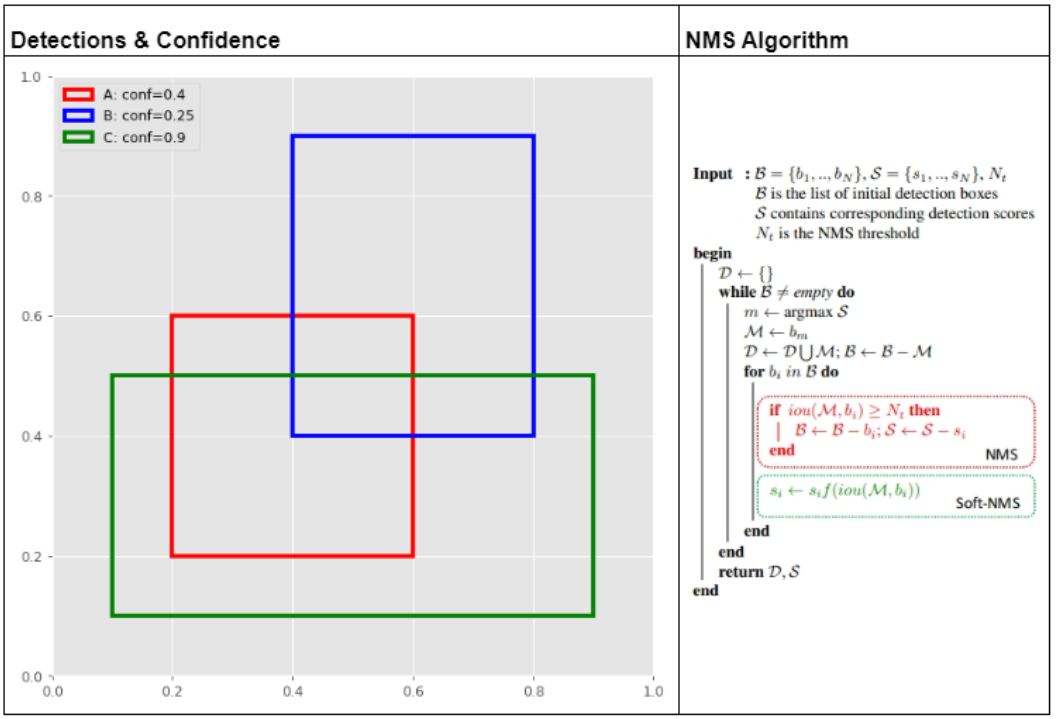
Question 1

1.0/1.0 point (graded)

Consider the above rectangles A and B. Overlap these two rectangles is shown in green. Intersection over Union (IoU) of A and B is:

- 1/8
- 1/7 ✔
- 1/6
- 1/5

Submit You have used 1 of 2 attempts



The linear scoring function of soft-NMS:

$$s_i = \begin{cases} s_i(1 - iou(\mathscr{M}, b_i)), & iou(\mathscr{M}, b_i) \geqslant N_t, \\ s_i, & iou(\mathscr{M}, b_i) < N_t. \end{cases}$$

where,

$\mathscr{M}$ is the detected box with a maximum score.

There are three bounding boxes in the above image:

- $A : confidence = 0.4$
- $B : confidence = 0.25$
- $C : confidence = 0.9$

Question 2

0.0/1.0 point (graded)

Apply the NMS algorithm with the above linear scoring function on the above detections. The IoU threshold is 0.5 ($N_t = 0.5$). Let's denote a score of A with $S(A)$, and so on. The updated score is:

$S(A) = 0.9$, $S(B) = 0.1$, and $S(C) = 0.2$

$S(A) = 0.9$, $S(B) = 0.1$, and $S(C) = 0.16$ ✖

$S(A) = 0.9$, $S(B) = 0.075$, and $S(C) = 0.2$

$S(A) = 0.9$, $S(B) = 0.4$, and $S(C) = 0.25$ ✔

Explanation

None of the combinations of two bounding boxes has IoU greater than 0.5. So there will be no change in their scores.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Question 3

0.0/1.0 point (graded)

Suppose we are training a single-stage detector with a single anchor size of (height=32, width=16). Which of the following statement is true about the trained model's predicted bounding box:

the size of the predicted bounding box always will be equal to the anchor size

the area of the predicted box will always be the same as the area of the anchor box, but aspect ratio may be different from the anchor box. ✖

the aspect ratio of the predicted box will always be the same as the anchor box, but the area may be different from the anchor box.

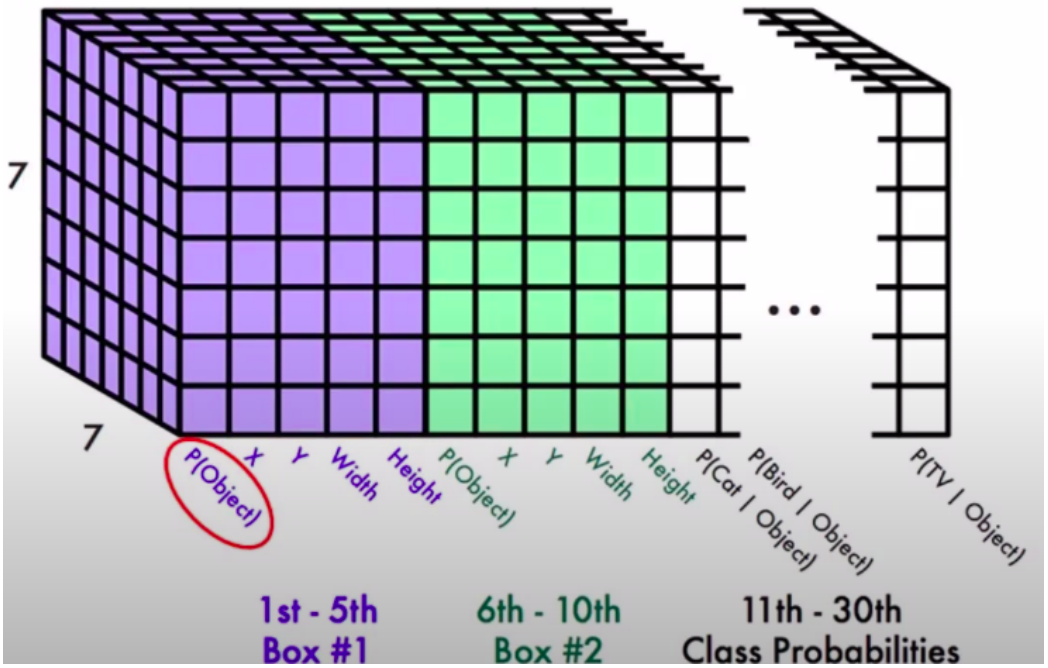
The area and the aspect ratio both may be different than the anchor box. ✔

Explanation

At the time of training, the ground truth uses to be encoded in terms of deviation from the anchor center and change in height and width relative to the anchor box height and width. So the predicted box may have different area and aspect ratio.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem



Recall the output of YOLO network. There is 7x7 grid with a depth of 30.

For each grid, there are two bounding boxes; for each bounding box, there are five numbers, one for the probability of being an object and the other four to represent the bounding box. The other twenty numbers are the conditional probability of twenty classes.

Assume the following information for an output grid:

For bounding box 1:

- Probability of being an object, $P_1(object) = 0.2$

For bounding box 2:

- Probability of being an object, $P_2(object) = 0.7$

The conditional probability of classes is as follows:

- $P(TV|object) = 0.52$
- $P(Bird|object) = 0.3$
- For the class in all classes except TV and Bird:
  - P(class | object) = 0.01

Before passing through NMS, we need to assign at max one class to each anchor box. Any bounding box will be assigned any class only if the class probability greater than a threshold.

Question 4

0.0/1.0 point (graded)

Suppose the class probability threshold is 0.2. Which of the following class mapping is true:

- bounding box 1 class = TV, bounding box 2 class = Bird ✖
- bounding box 1 class = Bird, bounding box 2 class = TV
- bounding box 1 class = None, bounding box 2 class = TV ✔
- bounding box 1 class = TV, bounding box 2 class = None

Explanation

$P\left(class\right) = P\left(class|object\right) P\left(object\right)$

For bounding box 1:

$P\left(TV\right) = 0.52 * 0.2 < 0.2$

$P\left(Bird\right) = 0.3 * 0.2 < 0.2$

All probabilities are less than the threshold (0.2). So no class will assign to bounding box 1

For bounding box 2:

$P\left(TV\right) = 0.52 * 0.7 = 0.364 > 0.2$

$P\left(Bird\right) = 0.3 * 0.7 = 0.21 > 0.2$

The probability of TV and Bird is greater than the threshold (0.2). One bounding box can be assigned only one class, so bounding box 2 will be assigned TV as it has a maximum probability.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Question 5

0.0/1.0 point (graded)

Imagine that you are training a deep neural network for Object Detection. Your current architecture is giving an mAP of 0.65. You decide to add a convolutional layer to the architecture. What is the mAP of this new network?

- Less than 0.65
- At least 0.65
- More than 0.65 ✖

Can not predict before the actual training ✔

Explanation

Adding Convolution does not guarantee to increase or decrease the model performance.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Question 6

0.0/1.0 point (graded)

Consider the same scenario as is mentioned in the above qestion. This time, instead of adding a convolutional layer, you add a residual block. What is the mAP of this new network?

Less than 0.65

At least 0.65 ✔

More than 0.65 ✖

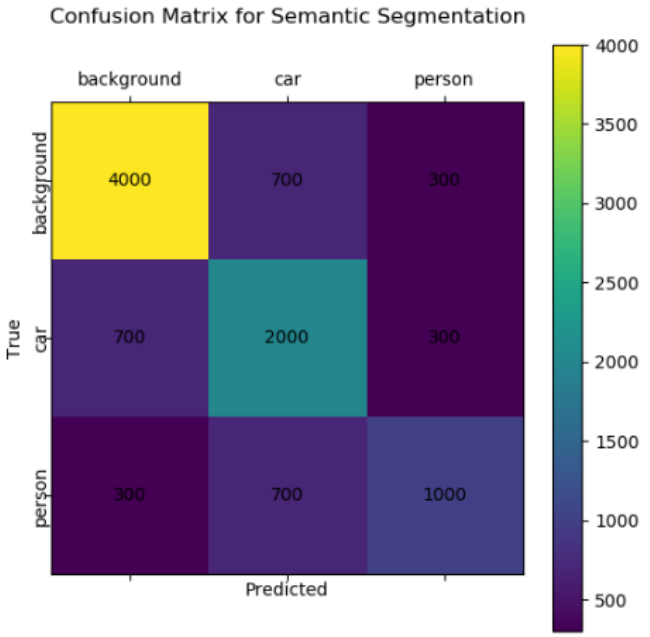
Can not predict before the actual training

Explanation

Adding a residual block does not harm the model performance.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem



`Sørensen-Dice coefficient` can be calculated by the following:

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN},$$

where:

- `true positives` (TP) - the cases when the model properly predicted positive class.
- `false positives` (FP)- the opposite to TP outcomes, when the positive class was incorrectly predicted.
- `false negative` (FN) - the cases of improperly predicted negative classes.

Question 7

1.0/1.0 point (graded)

The Sørensen-Dice coefficient for the above confusion matrix is:

$DSC_{background} = 0.8, DSC_{car} = 0.667, DSC_{person} = 0.5$

$DSC_{background} = 0.8, DSC_{car} = 0.625, DSC_{person} = 0.556$ ✔

$DSC_{background} = 0.8, DSC_{car} = 0.588, DSC_{person} = 0.625$

$DSC_{background} = 0.4, DSC_{car} = 0.2, DSC_{person} = 0.1$

Explanation

$$DSC_{background} = \frac{2 \cdot 4000}{2 \cdot 4000 + (700+300) + (700+300)} = 0.8$$
$$DSC_{car} = \frac{2 \cdot 2000}{2 \cdot 2000 + (700+300) + (700+700)} = 0.625$$
$$DSC_{person} = \frac{2 \cdot 1000}{2 \cdot 1000 + (300+700) + (300+300)} = 0.556$$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Given,

- detection and its confidence
- ground truth

Algorithm to calculate average precession (AP):

- sort detection in decreasing order of confidence
- from highest to lowest confidence detection:
  - calculate precession and recall
- interpolate precession using the following at different recall points:

$$p_{\text{interp}}(r) = \max_{r' : r' \geq r} p(r')$$

- calculate the area under the curve of interpolated precession.

Question 8

0.0/1.0 point (graded)

Which of the following statement(s) are true:

- For fixed detection with confidence and ground truth, AP is unique.
- For fixed detection with confidence and ground truth, AP may have more than one value. ✔
- For a fixed confidence threshold, an increase in the IoU threshold must lead to a decrease in AP.
- For a fixed confidence threshold, an increase in the IoU threshold must lead to an increase in AP.

✖

Explanation

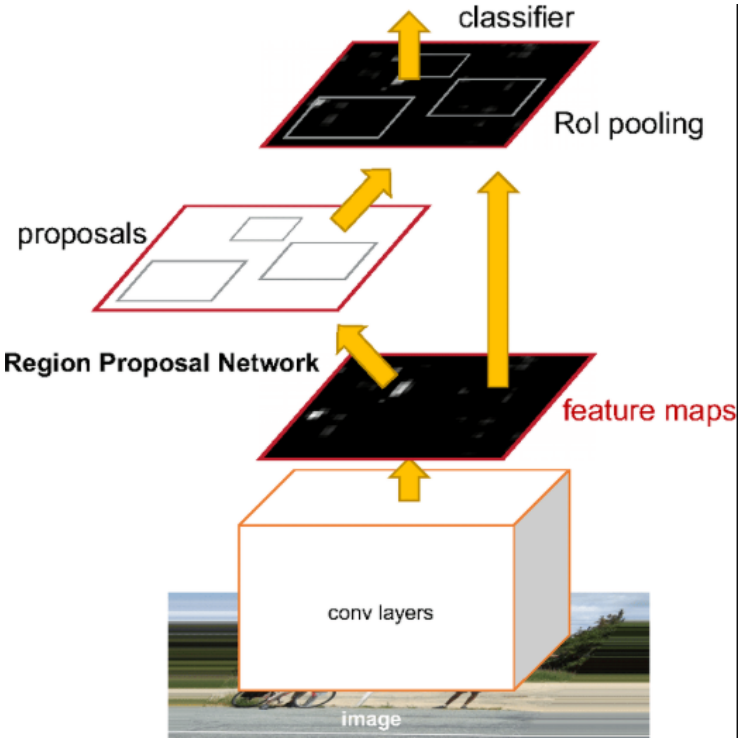
Consider the following four detections with 100% recall:

A: conf = 0.8, True positive

B: conf = 0.8, False Positive
C: conf = 0.7, True positive
D: conf = 0.6, True positive
In the above situation, if one starts with A will get more AP than start with B.
For a fixed confidence threshold, an increase in the IoU threshold most probably leads to a decrease in AP but can not be granted.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem



Question 9

0.75/1.0 point (graded)

Which of the following statement(s) are true about the above faster RCNN?

- Backbone (the feature extractor) must not be fully-connected layers. ✔
- Region Proposal Network gives bounding boxes of fixed size and its objectness score.
- RoI pooling layers take bounding boxes from Region Proposal Network, and for these bounding boxes, it takes features from features map and resizes to constant height and

width. ✔

The classifier classifies object proposals in different classes and also improves bounding boxes. ✔

✱

Explanation

Region Proposal Network does not give fixed size bounding boxes. But the fixed size is achieved by the RoI pooling layer.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Question 10

0.25/1.0 point (graded)

Which of the following statement(s) are true about nn.ConvTranspose2d and F.interpolate?

both can be used to upsample the input ✔

both have learning parameters

nn.ConvTranspose2d has learning parameters, and F.interpolate does not have ✔

F.interpolate has learning parameters, and nn.ConvTranspose2d does not have

✱

Explanation

F.interpolate just a function for interpolation; it does not have learning parameters.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem