

Detection Algorithms

✓ Congratulations! You passed!

Grade
received **100%**

Latest Submission
Grade 100%

To pass 80% or
higher

Retake the
assignment in **23h
52m**

Go to
next
item

1. You are building a 3-class object classification and localization algorithm. The classes are: pedestrian (c=1), car (c=2), motorcycle (c=3). What should y be for the image below? Remember that “?” means “don’t care”, which means that the neural network loss function won’t care what the neural network gives for that component of the output. Recall $y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$.

1 / 1 point



<https://www.pexels.com/es-es/foto/fotografia-de-motocicleta-clasica-en-carretera-995487/>

- $y = [1, 0.22, 0.5, 0.2, 0.3, 1, 1, 1]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 0]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 1]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, ?, ?, 1]$

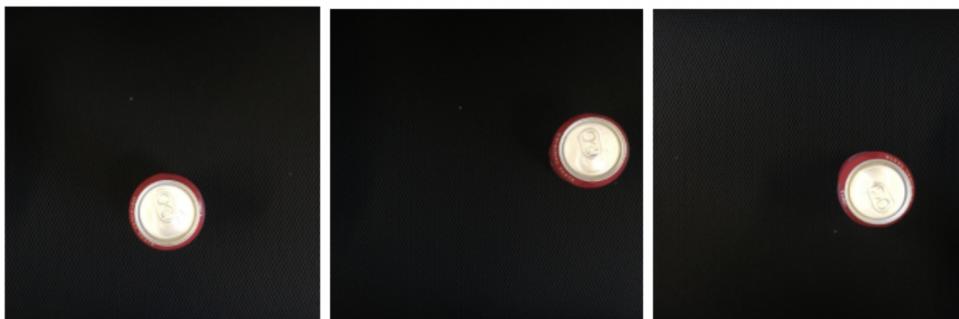
 Expand

 Correct

Correct. $p_c = 1$ since there is a motorcycle in the picture. We can also see that b_x, b_y as percentages of the image are adequate. They look approximately correct as well as b_h, b_w , and the value of $c_3 = 1$ for the motorcycle.

2. You are working on a factory automation task. Your system will see a can of soft-drink coming down a conveyor belt, and you want it to take a picture and decide whether (i) there is a soft-drink can in the image, and if so (ii) its bounding box. Since the soft-drink can is round, the bounding box is always square, and the soft drink can always appear the same size in the image. There is at most one soft drink can in each image. Here're some typical images in your training set:

1 / 1 point



To solve this task it is necessary to divide the task into two: 1. Construct a system to detect if a can is present or not. 2. Construct a system that calculates the bounding box of the can when present. Which one of the following do you agree with the most?

- We can't solve the task as an image classification with a localization problem since all the bounding boxes have the same dimensions.
- An end-to-end solution is always superior to a two-step system.
- The two-step system is always a better option compared to an end-to-end solution.
- We can approach the task as an image classification with a localization problem.

 Expand

 Correct

Correct. We can use a network to combine the two tasks similar to that described in the lectures.

3. When building a neural network that inputs a picture of a person's face and outputs N landmarks on the face (assume that the input image contains exactly one face), we need two coordinates for each landmark, thus we need $2N$ output units. True/False?

1 / 1 point

True

False

 Expand

 Correct

Correct. Recall that each landmark is a specific position in the face's image, thus we need to specify two coordinates for each landmark.

4. When training one of the object detection systems described in the lectures, you need a training set that contains many pictures of the object(s) you wish to detect. However, bounding boxes do not need to be provided in the training set, since the algorithm can learn to detect the objects by itself.

1 / 1 point

True

False

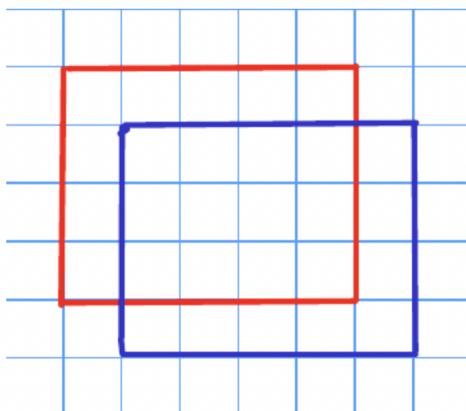
 Expand

 Correct

Correct, you need bounding boxes in the training set. Your loss function should try to match the predictions for the bounding boxes to the true bounding boxes from the training set.

5. What is the IoU between the red box and the blue box in the following figure? Assume that all the squares have the same measurements.

1/1 point



- $\frac{3}{7}$
- $\frac{1}{2}$
- $\frac{2}{5}$
- $\frac{4}{5}$

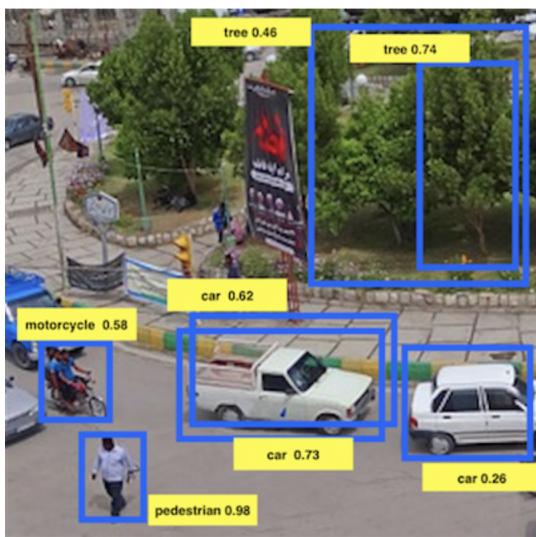
 Expand

 Correct

Correct. IoU is calculated as the quotient of the area of the intersection (16) over the area of the union (28).

6. Suppose you run non-max suppression on the predicted boxes below. The parameters you use for non-max suppression are that boxes with probability ≤ 0.7 are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5.

1 / 1 point



After non-max suppression, only three boxes remain. True/False?

True

False

 Expand

 Correct

Correct. After eliminating the boxes with a score less than 0.7 only three boxes remain, and they don't intersect. Thus three boxes are left.

7. If we use anchor boxes in YOLO we no longer need the coordinates of the bounding box b_x, b_y, b_h, b_w since they are given by the cell position of the grid and the anchor box selection. True/False?

1 / 1 point

False

True

 Expand

 Correct

Correct. We use the grid and anchor boxes to improve the capabilities of the algorithm to localize and detect objects, for example, two different objects that intersect, but we still use the bounding box coordinates.

8. Semantic segmentation can only be applied to classify pixels of images in a binary way as 1 or 0, according to whether they belong to a certain class or not. True/False?

1 / 1 point

False

True

 Expand

 Correct

Correct. The same ideas used for multi-class classification can be applied to semantic segmentation.

9. Using the concept of Transpose Convolution, fill in the values of **X**, **Y** and **Z** below.

1 / 1 point

(padding = 1, stride = 2)

Input: 2x2

1		2
3		4

Filter: 3x3

1	1	1
0	0	0
-1	-1	-1

Result: 6x6

	0	0	0	X	
	Y	4	2	2	
	0	0	0	0	
	-3	Z	-4	-4	

X = 0, Y = 2, Z = -7

X = 0, Y = -1, Z = -7

X = 0, Y = -1, Z = -4

X = 0, Y = 2, Z = -1

 Expand

 Correct

Correct.

10. When using the U-Net architecture with an input $h \times w \times c$, where c denotes the number of channels, the output will always have the shape $h \times w$. True/False?

1/1 point

False

True

 Expand

 Correct

Correct. The output of the U-Net architecture can be $h \times w \times k$ where k is the number of classes.