

## Detection Algorithms

Latest Submission Grade 90%

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



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

Expand

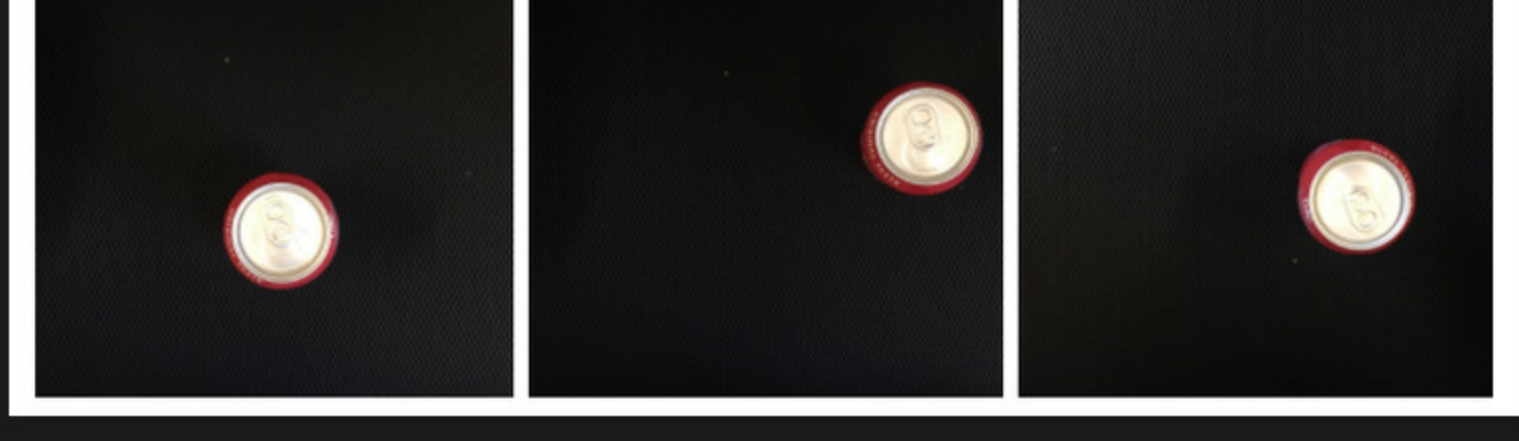
Correct

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 are some typical images in your training set:

1 / 1 point



What are the most appropriate (lowest number of) output units for your neural network?

- ☐ Logistic unit (for classifying if there is a soft-drink can in the image)
- ☐ Logistic unit:  $b_x, b_y, b_h, b_w$
- ☒ Logistic unit:  $b_x$  and  $b_y$
- ☐ Logistic unit:  $b_x, b_y, b_h$  (since  $b_w = b_h$ )

Expand

Correct

Correct!

3.

If you build a neural network that inputs a picture of a person’s face and outputs N landmarks on the face (assume the input image always contains exactly one face), how many output units will the network have?

1 / 1 point

- ☒ 2N
- ☐  $N^2$
- ☐ N
- ☐ 3N

Expand

Correct

Correct

4.

When training one of the object detection systems described in the lectures, each image must have zero or exactly one bounding box. True/False?

1 / 1 point

- ☒ False
- ☐ True

Expand

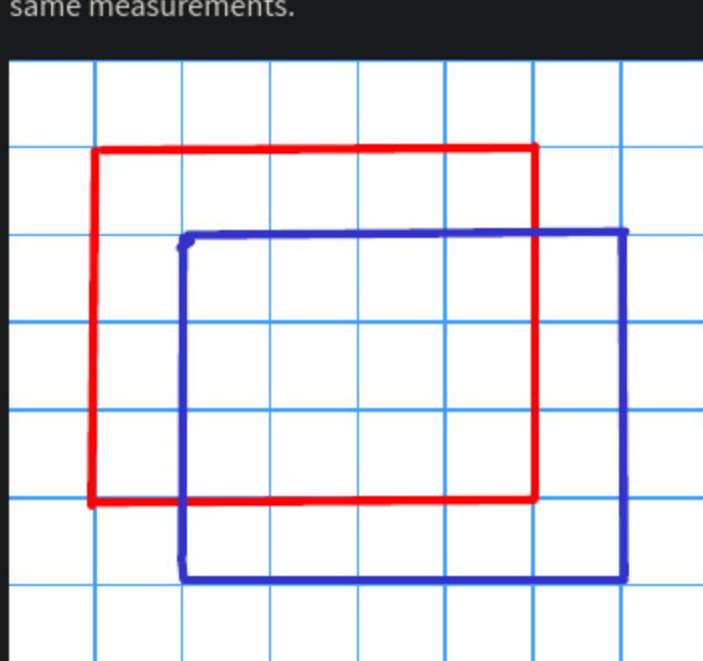
Correct

Correct. In a single image, there might be more than only one instance of the object we are trying to localize, so it must have several bounding boxes.

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.

0 / 1 point



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

Expand

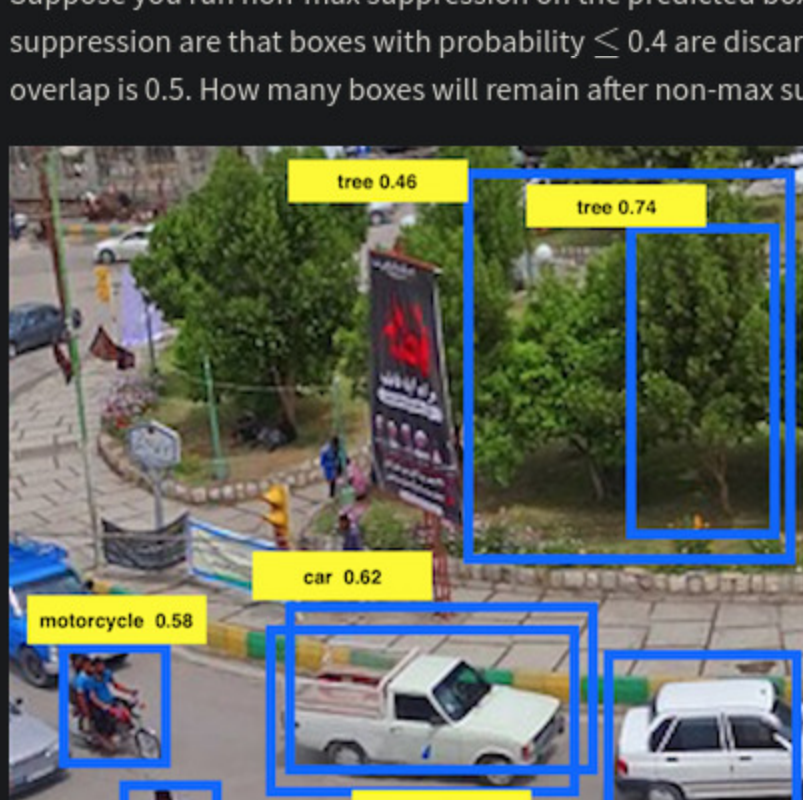
Incorrect

Recall that IoU is calculated as the quotient of the area of the intersection over the area of the union, and that the area of the union is not the sum of the areas.

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  $\leq 0.4$  are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5. How many boxes will remain after non-max suppression?

1 / 1 point



- ☐ 7
- ☐ 3
- ☐ 4
- ☒ 5
- ☐ 6

Expand

Correct

Correct!

7.

Suppose you are using YOLO on a 19x19 grid, on a detection problem with 20 classes, and with 5 anchor boxes. During training, for each image you will need to construct an output volume  $y$  as the target value for the neural network; this corresponds to the last layer of the neural network. ( $y$  may include some “?” or “don’t cares”). What is the dimension of this output volume?

1 / 1 point

- ☒ 19x19x(5x25)
- ☐ 19x19x(5x20)
- ☐ 19x19x(25x20)
- ☐ 19x19x(20x25)

Expand

Correct

Correct, you get a 19x19 grid where each cell encodes information about 5 boxes and each box is defined by a confidence probability ( $p_c$ ), 4 coordinates ( $b_x, b_y, b_h, b_w$ ) and classes ( $c_1, \dots, c_{20}$ ).

8.

What is Semantic Segmentation?

1 / 1 point

- ☒ Locating objects in an image by predicting each pixel as to which class it belongs to.
- ☐ Locating objects in an image belonging to different classes by drawing bounding boxes around them.
- ☐ Locating an object in an image belonging to a certain class by drawing a bounding box around it.

Expand

Correct

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	0	-1
1	0	-1
1	0	-1

Result: 6x6

	0	1	0	-2	
	0	X	0	Y	
	0	1	0	Z	
	0	1	0	-4	

- ☐ X = 2, Y = 6, Z = 4
- ☒ X = 2, Y = -6, Z = -4
- ☐ X = -2, Y = -6, Z = -4
- ☐ X = 2, Y = -6, Z = 4

Expand

Correct

10.

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

1 / 1 point

- ☐ True
- ☒ False

Expand

Correct

Correct. The output of the U-Net architecture can be  $\hat{h} \times w \times k$  where  $k$  is the number of classes. The number of channels doesn't have to match between input and output.