

## Congratulations! You passed!

[Go to next item](#)

Grade received 100% Latest Submission Grade 100% To pass 80% or higher

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



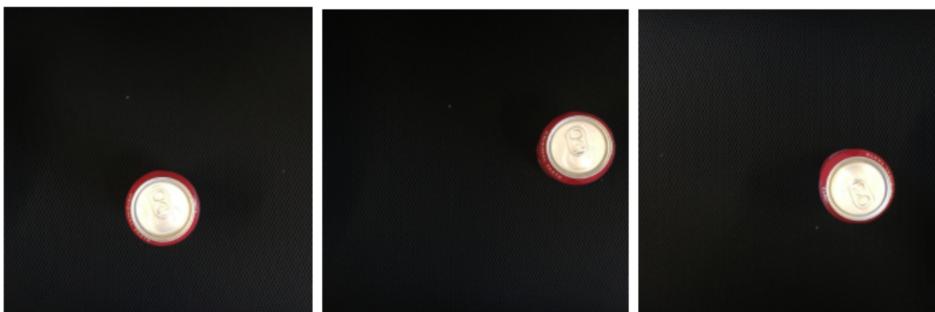
- $y = [0, ?, ?, ?, ?, ?, ?, ?]$
- $y = [1, ?, ?, ?, ?, 0, 0, 0]$
- $y = [1, ?, ?, ?, ?, ?, ?, ?]$
- $y = [?, ?, ?, ?, ?, ?, ?, ?]$

[Expand](#)

Correct.

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,  $b_x, b_y, b_h, b_w$
- Logistic unit,  $b_x, b_y, b_h$  (since  $b_w = b_h$ )

- Logistic unit (for classifying if there is a soft-drink can in the image)
- Logistic unit,  $b_x$  and  $b_y$

 Expand

 Correct

Correct!

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

- False
- True

 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

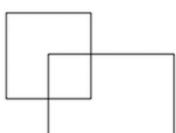
 Expand

 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 these two boxes? The upper-left box is  $2 \times 2$ , and the lower-right box is  $2 \times 3$ . The overlapping region is  $1 \times 1$ .

1 / 1 point



- $\frac{1}{10}$
- $\frac{1}{9}$
- $\frac{1}{6}$
- None of the above

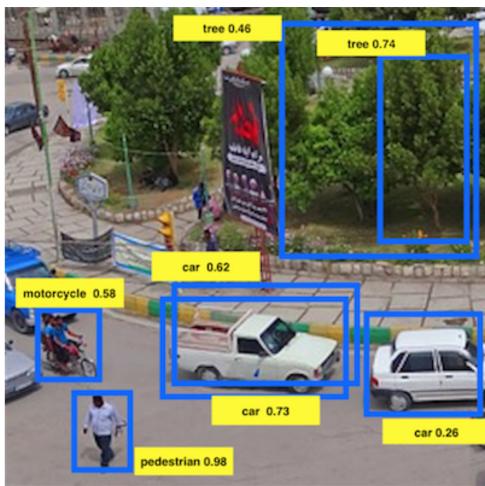
 Expand

 Correct

Correct. The left box's area is 4 while the right box's is 6. Their intersection's area is 1. So their union's area is  $4 + 6 - 1 = 9$  which leads to an intersection over union of  $1/9$ .

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



- 3
- 5
- 4
- 7
- 6

 Expand

 Correct

Correct!

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

- True
- False

 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	0	-1
1	0	-1
1	0	-1

Result: 6x6

	0	1	0	-2	
	0	<b>X</b>	0	<b>Y</b>	
	0	1	0	<b>Z</b>	
	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  $h \times w \times c$ , where  $c$  denotes the number of channels, the output will always have the shape  $h \times w$ .  
True/False?

1 / 1 point

True

False

 Expand

 Correct

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