Your grade: 90%

Your latest: **90%** • Your highest: **90%** • To pass you need at least 80%. We keep your highest score.

Next item →

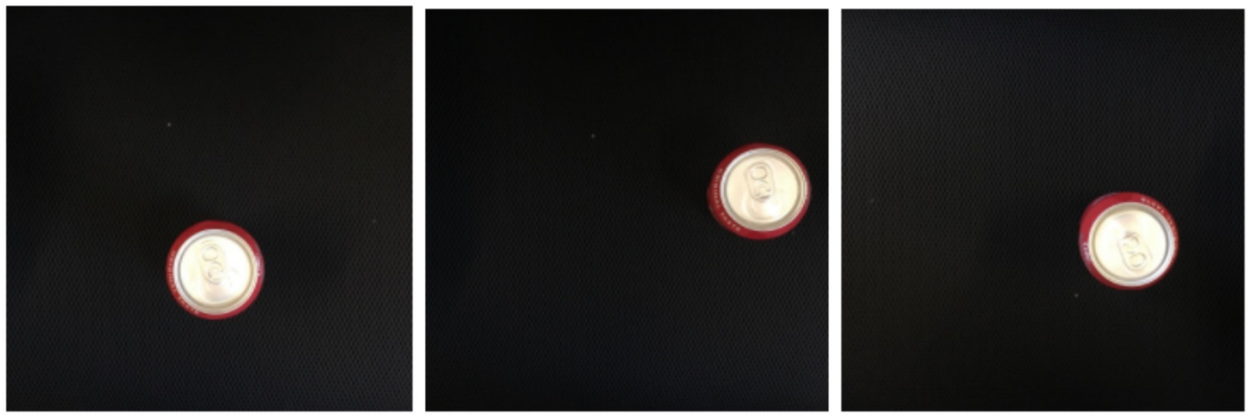
1. You are building a 3-class object classification and localization algorithm. The classes are: pedestrian (c=1), car (c=2), motorcycle (c=3). What should y be for the image below? Remember that "?" means "don't care", which means that the neural network loss function won't care what the neural network 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

- ○ $y = [1, 0.66, 0.5, 0.75, 0.16, 0, 0, 0]$
- ◉ $y = [1, 0.66, 0.5, 0.75, 0.16, 1, 0, 0]$
- ○ $y = [1, ?, ?, ?, ?, 1, ?, ?]$
- ○ //www.pexels.com/es-es/foto/mujer-vestida-con-falda-azul-y-blanca-caminando-cerca-de-la-hierba-verde-durante-el-dia-144474/
- ○ $y = [1, 0.66, 0.5, 0.16, 0.75, 1, 0, 0]$

Correct
Correct. $p_c = 1$ since there is a pedestrian in the picture. We can see that b_x, b_y as percentages of the image are approximately correct as well b_h, b_w, and the value of $c_1 = 1$ for a pedestrian.

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



The most adequate output for a network to do the required task is $y = [p_c, b_x, b_y, b_h, b_w, c_1]$. (Which of the following do you agree with the most?)

- False, since we only need two values c_1 for no soft-drink can and c_2 for soft-drink can.
- True, p_c indicates the presence of an object of interest, b_x, b_y, b_h, b_w indicate the position of the object and its bounding box, and c_1 indicates the probability of there being a can of soft-drink.
- (selected) False, we don't need b_h, b_w since the cans are all the same size.
- True, since this is a localization problem.

Correct
Correct. With the position b_x, b_y we can completely characterize the position of the object if it is present. We should use only one additional logistic unit to indicate if the object is present or not.

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

- (selected) 2N
- N
- 3N
- N^2

Correct
Correct

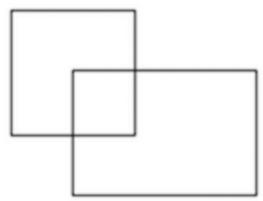
4. When training one of the object detection systems described in the lectures, each image must have zero or 1 / 1 point

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

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 these two boxes? The upper-left box is 2x2, and the lower-right box is 2x3. The overlapping region is 1x1. 1 / 1 point



- (•) 1/9
- ( ) 1/10
- ( ) None of the above
- ( ) ⅙

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 nelow. 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?

- False
- True

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

- ○ True
- ◉ False

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 point

- ◉ True
- ○ False

Incorrect
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	3
2	4

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	3
2	4

Filter: 3x3

1	0	1
0	0	0
1	0	1

Result: 6x6

	0	0	0	0	
	0	**X**	0	7	
	0	0	0	**Y**	
	0	**Z**	0	4	

	0	0	0	**Y**	
	0	**Z**	0	4	

- ○ X = 3, Y = 0, Z = 4
- ○ X = 4, Y = 3, Z = 2
- ◉ X = 10, Y = 0, Z = 6
- ○ X = 10, Y = 0, Z = 0

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 $h \times w \times c$. True/False? 1 / 1 point

- ◉ False
- ○ True

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