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 $y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$.

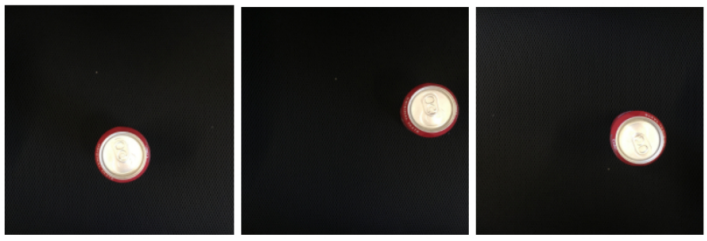
1 / 1 point

- (●) $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 1]$
- ( ) $y = [1, 0.22, 0.5, 0.2, 0.3, ?, ?, 1]$
- ( ) //www.pexels.com/es-es/foto/fotografia-de-motocicleta-clasica-en-carretera-995487/
- ( ) $y = [1, 0.22, 0.5, 0.2, 0.3, 1, 1, 1]$
- ( ) $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 0]$

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, b_x and b_y
- ( ) 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, b_y, b_h (since $b_w = b_h$)

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

1 / 1 point

(assume that the input image contains exactly one face), we need two coordinates for each landmark, thus we need 2N output units. True/False?

- (●) True
- ( ) False

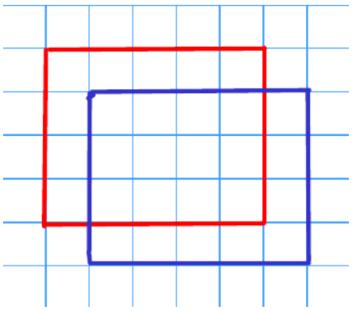
Correct
Correct. Recall that each landmark is a specific position in the face's image, thus we need to specify two coordinates for each landmark.

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

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**



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

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 $\leq$ 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
- ( ) 6
- ( ) 7

- 4

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(20x25)
- 19x19x(25x20)
- (•) 19x19x(5x25)
- 19x19x(5x20)

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 ($c_1, \ldots, c_{20}$).

8. What is Semantic Segmentation? **1 / 1 point**

- 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.
- (•) Locating objects in an image by predicting each pixel as to which class it belongs to.

Correct

9. Using the concept of Transpose Convolution, fill in the values of **X**, **Y** and **Z** below. **0 / 1 point**

(*padding = 1, stride = 2*)

- X = 4, Y = 3, Z = 2
- Filter: 3x3

1	0	1
0	0	0
1	0	1

- X = 10, Y = 0, Z = 0
- Result: 6x6

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

- X = 10, Y = 0, Z = 6
- (•) X = 3, Y = 0, Z = 4
- Input: 2x2

1	3
2	4

Incorrect
To revise the concepts watch the lecture *Transpose Convolution*.

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**

- True
- (•) False

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.