Quiz: Detection Algorithms

Congratulations! You passed!

Grade received 90%	Latest Submission Grade 90%	To pass 80% or higher

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/mujer-vestida-con-falda-azul-y-blanca-caminando-cerca-de-la-hierba-verde-durante-el-dia-144474/

- $y = [1, 0.66, 0.5, 0.75, 0.16, 0, 0, 0]$
- $y = [1, ?, ?, ?, ?, 1, ?, ?]$
- $y = [1, 0.66, 0.5, 0.75, 0.16, 1, 0, 0]$
- $y = [1, 0.66, 0.5, 0.16, 0.75, 1, 0, 0]$

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

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), which is true about $\hat{y}^{(i)}$? **0 / 1 point**

- $\hat{y}^{(i)}$ has shape (N, 1)
- $\hat{y}^{(i)}$ stores the probability that a landmark is in a given position over the face.
- $\hat{y}^{(i)}$ has shape (2N, 1)
- $\hat{y}^{(i)}$ has shape (1, 2N)

Expand

Incorrect
Remember that we use column vectors and not row vectors for the output.

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**

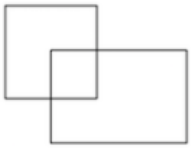
- False
- True

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

1 / 1 point



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

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

- 6
- 4
- 5
- 3
- 7

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. 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 $h \times w \times c$, 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