Congratulations! You passed!

Grade received 100% | Latest Submission Grade 100% | 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/

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

Loading [MathJax]/jax/output/CommonHTML/jax.js

Expand

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're some typical images in your training set:

1 / 1 point

To solve this task it is necessary to divide the task into two: 1. Construct a system to detect if a can is present or not. 2. Construct a system that calculates the bounding box of the can when present. Which one of the following do you agree with the most?

- The two-step system is always a better option compared to an end-to-end solution.
- We can't solve the task as an image classification with a localization problem since all the bounding boxes have the same dimensions.
- An end-to-end solution is always superior to a two-step system.
- We can approach the task as an image classification with a localization problem.

Expand

Correct
Correct. We can use a network to combine the two tasks similar to that described in the lectures.

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

- (•) True
- ( ) False

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. You are working to create an object detection system, like the ones described in the lectures, to locate cats in a room. To have more data with which to train, you search on the internet and find a large number of cat photos. 1 / 1 point

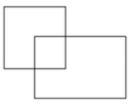
Which of the following is true about the system?

- ( ) We can't use internet images because it changes the distribution of the dataset.
- (•) We can't add the internet images unless they have bounding boxes.
- ( ) We should use the internet images in the dev and test set since we don't have bounding boxes.
- ( ) We should add the internet images (without the presence of bounding boxes in them) to the train set.

Expand

Correct
Correct. As this is a localization model, we also need the coordinates of the bounding boxes, not just the images.

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



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

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 ≤ 0.4 are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5. 1 / 1 point

Notice that there are three bounding boxes for cars. After running non-max suppression, only the bounding box of the car with 0.73 is kept from the three bounding boxes for cars. True/False? Choose the best answer.

- False. Two bounding boxes corresponding to cars are left since their IoU is zero.
- False. All the cars are eliminated since there is a pedestrian with a higher score of 0.98.
- True. The non-maximum suppression eliminates the bounding boxes with scores lower than the ones of the maximum.

Expand

Correct
Correct. The bounding box for the car on the right is eliminated because its probability is less than 0.4. Of the two bounding boxes in the middle, one is eliminated because their IoU is higher than 0.5. So, only one bounding box remains.

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. 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.

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

Expand

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.