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/

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

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 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, since this is a localization problem.
- False, we don't need b_h, b_w since the cans are all the same size.
- 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.

Expand

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. 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 $\hat{y}^{(i)}$? **1 / 1 point**

- $\hat{y}^{(i)}$ 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 (N, 1)
- $\hat{y}^{(i)}$ has shape (1, 2N)

Expand

Correct
Correct. Since we have two coordinates (x,y) for each landmark we have N of them.

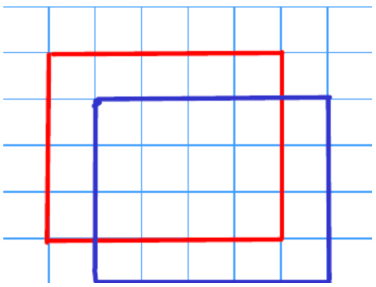
4. When training one of the object detection systems described in the lectures, you need a training set that contains many pictures of the object(s) you wish to detect. However, bounding boxes do not need to be provided in the training set, since the algorithm can learn to detect the objects by itself. **1 / 1 point**

- False
- True

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 the red box and the blue box in the following figure? Assume that all the squares have the same measurements. **1 / 1 point**



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

Expand

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 ≤ 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.

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

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. 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? **0 / 1 point**

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

Expand

Incorrect
How is a box defined for a problem with 20 classes?

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	1	1
0	0	0
-1	-1	-1

Result: 6x6

	0	0	0	X	
	Y	4	2	2	
	0	0	0	0	
	-3	Z	-4	-4	

- (•) X = 0, Y = 2, Z = -7
- ( ) X = 0, Y = -1, Z = -4
- ( ) X = 0, Y = 2, Z = -1
- ( ) X = 0, Y =-1, Z = -7

Expand

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 $h \times w$. 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.