Congratulations! you passed!

Grade received 80% Latest Submission Grade 80% 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 $y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$.

3 / 3 point

https://www.pexels.com/es-es/foto/fotografia-de-motocicleta-clasica-en-carretera-995687/

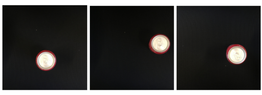
- $y = [1, 0.22, 0.5, 0.2, 0.3, 1, 1, 1]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, ?, ?, ?]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 1, 1]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 0]$

Expand

Correct
Correct. $SS_c = 1SS$ since there is a motorcycle in the picture. We can also see that SS_b_x,b_ySS as percentages of the image are adequate. They look approximately correct as well as SS_h, b_wSS, and the value of $SS_c_3 = 1SS$ 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're some typical images in your training set:

3 / 3 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?

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

Expand

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

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?

3 / 3 point

- N
- N^2
- 3N
- 2N

Expand

Correct
Correct

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?

3 / 3 point

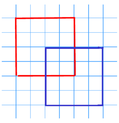
- False
- True

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

3 / 3 point



- $\frac{1}{7}$
- $\frac{1}{2}$
- $\frac{1}{6}$
- $\frac{1}{4}$

Expand

Correct
Correct. IoU is calculated as the quotient of the area of the intersection (4) 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.

3 / 3 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. 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.

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?

3 / 3 point

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

Expand

Correct
Correct, you get a 19x19 grid where each cell encodes information about 5 boxes and each box is defined by a confidence probability (SS_p_cSS), 4 coordinates (SS_b_x, b_y, b_h, b_wSS) and classes ($SS_c_1, ..., c_{20}$SS).

8. We are trying to build a system that assigns a value of 1 to each pixel that is part of a tumor from a medical image taken from a patient.

3 / 3 point

This is a problem of localization? True/False

- True
- False

Expand

Correct
Correct. This is a problem of semantic segmentation since we need to classify each pixel from the image.

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

0 / 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 = -1, Z = -4
- X = 0, Y = 2, Z = -7
- X = 0, Y = 2, Z = -1
- X = 0, Y = -1, Z = -7

Expand

Incorrect
To revise the concepts watch the lecture Transpose Convolution.

10. Suppose your input to a U-Net architecture is $h \times w \times 3$, where 3 denotes your number of channels (RGB). What will be the dimension of your output ?

0 / 1 point

- $h \times w \times n_c$ where n = number of filters used in the algorithm
- $h \times w \times n_c$ where n = number of of output channels
- $h \times w \times n_c$ where n = number of input channels
- $h \times w \times n_c$ where n = number of output classes

Expand

Incorrect
To revise, watch the lecture U-Net Architecture.