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/fotografia-de-motocicleta-clasica-en-carretera-995487/

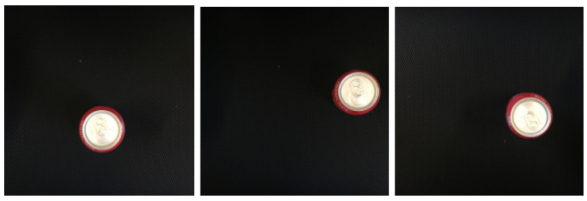
- $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 0]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, ?, ?, 1]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 1]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, 1, 1, 1]$

Expand

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



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, we don't need b_h, b_w since the cans are all the same size.
- True, since this is a localization problem.
- 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.
- False, since we only need two values $$c_1$$ for no soft-drink can and $$c_2$$ for soft-drink can

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

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), 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. **0 / 1 point**

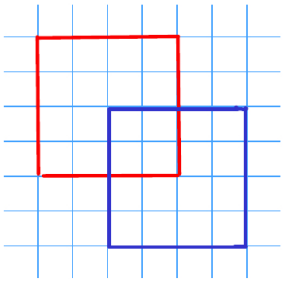
Which of the following is true about the system?

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

Expand

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

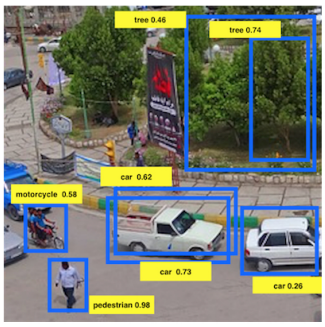


- $\frac{1}{4}$
- $\frac{1}{8}$
- $\frac{1}{7}$
- 1

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. **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. 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. 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. **1 / 1 point**

This is a problem of localization? True/False

- False
- True

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. **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 = -1
- X = 0, Y = 2, Z = -7
- X = 0, Y = -1, Z = -4
- X = 0, Y =-1, Z = -7

Expand

Correct
Correct.

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 ? **1 / 1 point**

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

Expand

Correct