← Back **Detection Algorithms**
Graded Quiz • 50 min

Due Jul 2, 11:59 PM EDT

Congratulations! You passed!

Grade received 80%

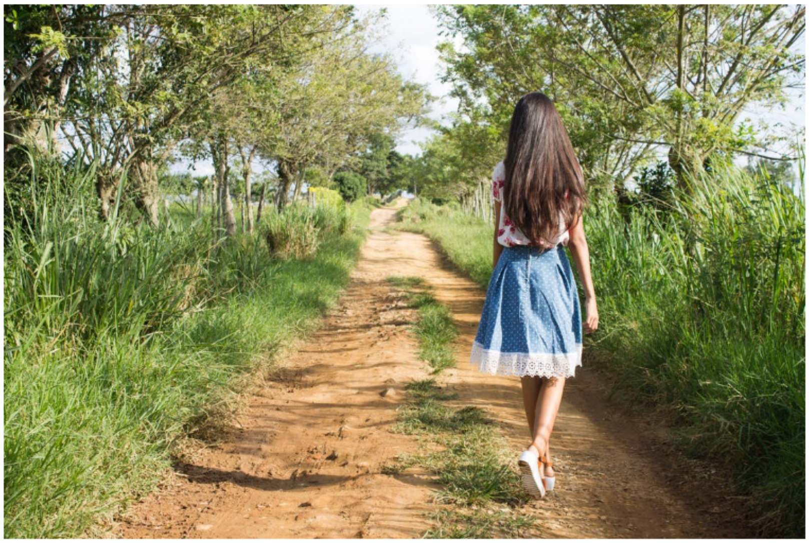
Latest Submission Grade 80%

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 $y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$.

0 / 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/

○ $y = [1, ?, ?, ?, ?, 1, ?, ?]$

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

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

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

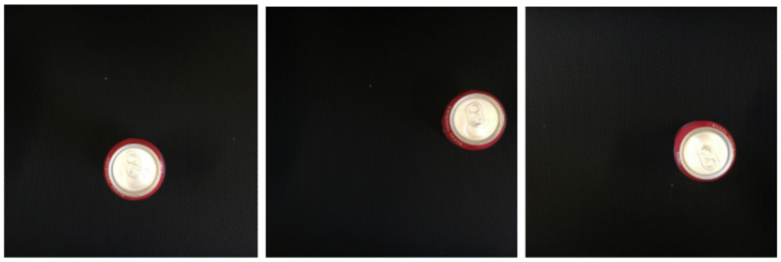
Expand

Incorrect

Notice that here $b_w > b_h$, and that doesn't correspond to the proportions of the bounding box for the 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:

0 / 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, 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.

● True, since this is a localization problem.

Expand

Incorrect

Although it is a localization problem, it has characteristics that differ from others where all these outputs might be necessary.

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?

1 / 1 point

○ 3N

○ N

○ N^2

● 2N

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

Expand

Correct

Correct

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

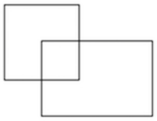
Expand

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 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

○ None of the above

○ $\frac{1}{6}$

○ $\frac{1}{10}$

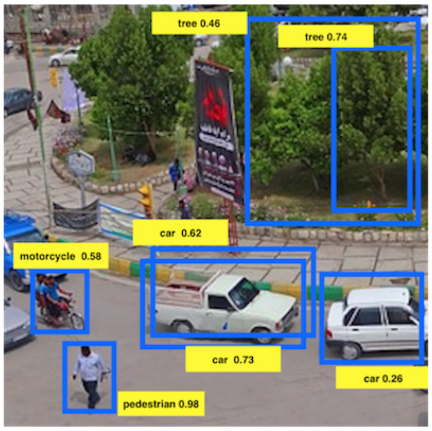
Expand

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 nelow. The parameters you use for non-max suppression are that boxes with probability ≤ 0.7 are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5.

1 / 1 point



After non-max suppression, only three boxes remain. True/False?

● True

○ False

Expand

Correct

Correct. After eliminating the boxes with a score less than 0.7 only three boxes remain, and they don't intersect. Thus three boxes are left.

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 an object in an image belonging to a certain class by drawing a bounding box around it.

○ Locating objects in an image belonging to different classes by drawing bounding boxes around them.

● 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.