

# Your grade: 100%

Your latest: **100%** • Your highest: **100%**

To pass you need at least 80%. We keep your highest score.

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



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

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

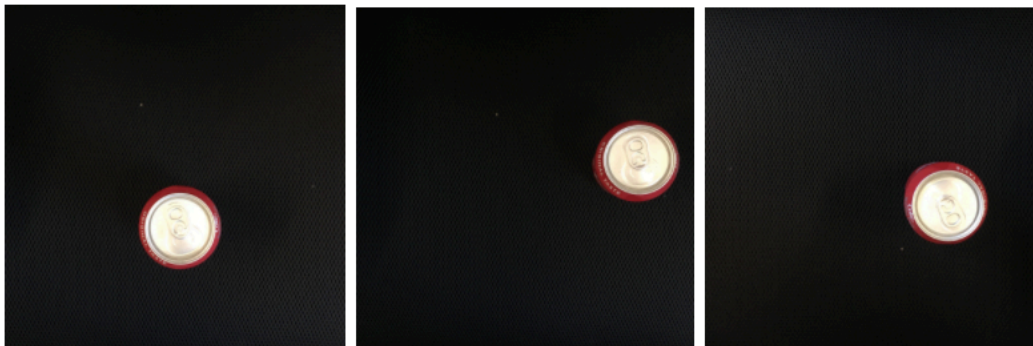
- ☐ //www.pexels.com/es-es/foto/mujer-vestida-con-falda-azul-y-blanca-caminando-cerca-de-la-hierba-verde-durante-el-dia-144474/
- ☒  $y = [1, 0.66, 0.5, 0.75, 0.16, 1, 0, 0]$
- ☐  $y = [1, ?, ?, ?, ?, 1, ?, ?]$

✓ **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.

✓ **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), which is true about  $\hat{y}^{(i)}$ ?

1 / 1 point

- ☐  $\hat{y}^{(i)}$  has shape  $(1, 2N)$
- ☒  $\hat{y}^{(i)}$  has shape  $(2N, 1)$
- ☐  $\hat{y}^{(i)}$  has shape  $(N, 1)$
- ☐  $\hat{y}^{(i)}$  stores the probability that a landmark is in a given position over the face.

✓ **Correct**

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

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?

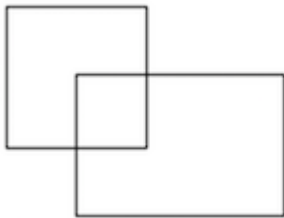
- ☒ We can't add the internet images unless they have bounding boxes.
- ☐ We should add the internet images (without the presence of bounding boxes in them) to the train set.
- ☐ We can't use internet images because it changes the distribution of the dataset.
- ☐ We should use the internet images in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

✓ **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



- ☐ None of the above
- ☐  $\frac{1}{6}$
- ☒  $\frac{1}{9}$
- ☐  $\frac{1}{10}$

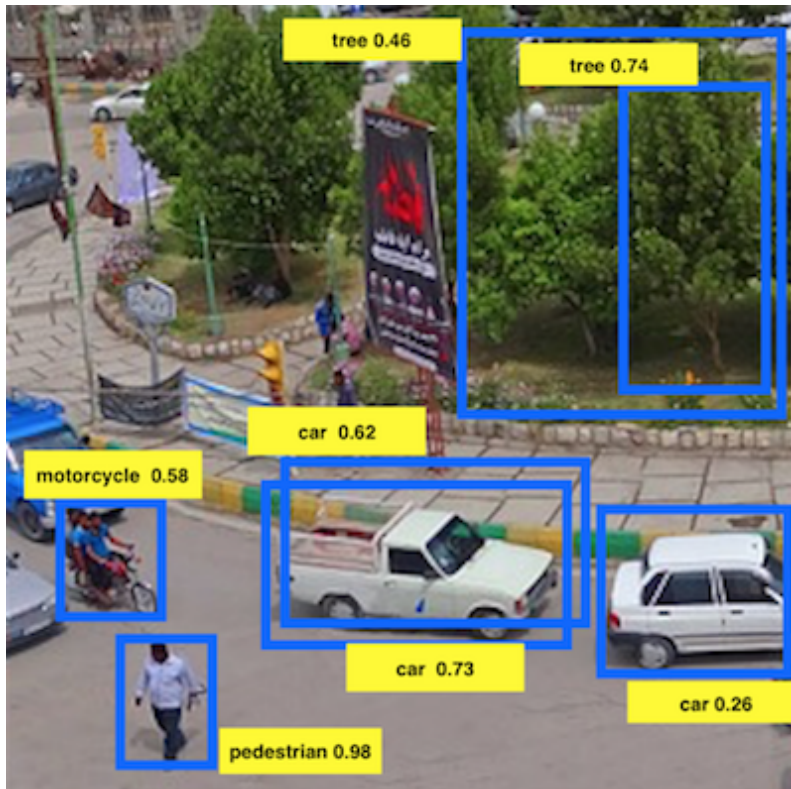
✓ **Correct**

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  $\leq 0.4$  are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5. How many boxes will remain after non-max suppression?

1 / 1 point



- ☐ 3
- ☐ 4
- ☒ 5
- ☐ 6
- ☐ 7

✔ **Correct**  
Correct!

7. Which of the following do you agree with about the use of anchor boxes in YOLO? Check all that apply.

1 / 1 point

- ☐ They prevent the bounding box from suffering from drifting.
- ☒ Each object is assigned to the grid cell that contains that object's midpoint.

✔ **Correct**  
Correct. This is the way we choose the corresponding cell.

- ☒ Each object is assigned to an anchor box with the highest IoU inside the assigned cell.

✔ **Correct**  
Correct. This is the way we choose the corresponding anchor box.

- ☐ Each object is assigned to any anchor box that contains that object's midpoint.

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

**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	<b>X</b>	
	<b>Y</b>	4	2	2	
	0	0	0	0	
	-3	<b>Z</b>	-4	-4	

- ☐  $X = 0, Y = -1, Z = -4$
- ☒  $X = 0, Y = 2, Z = -7$
- ☐  $X = 0, Y = 2, Z = -1$
- ☐  $X = 0, Y = -1, Z = -7$

✓ **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 ?

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

✓ **Correct**