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 point

- [ ] //www.pexels.com/es-es/foto/fotografia-de-motocicleta-clasica-en-carretera-995487/
- [x] $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 1]$
- [ ] $y = [1, 0.22, 0.5, 0.2, 0.3, ?, ?, 1]$
- [ ] $y = [1, 0.22, 0.5, 0.2, 0.3, 1, 1, 1]$
- [ ] $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 0]$

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 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.
- ○ False, since we only need two values c_1 for no soft-drink can and c_2 for soft-drink can.
- ○ 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.
- ◉ True, since this is a localization problem.

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 point**

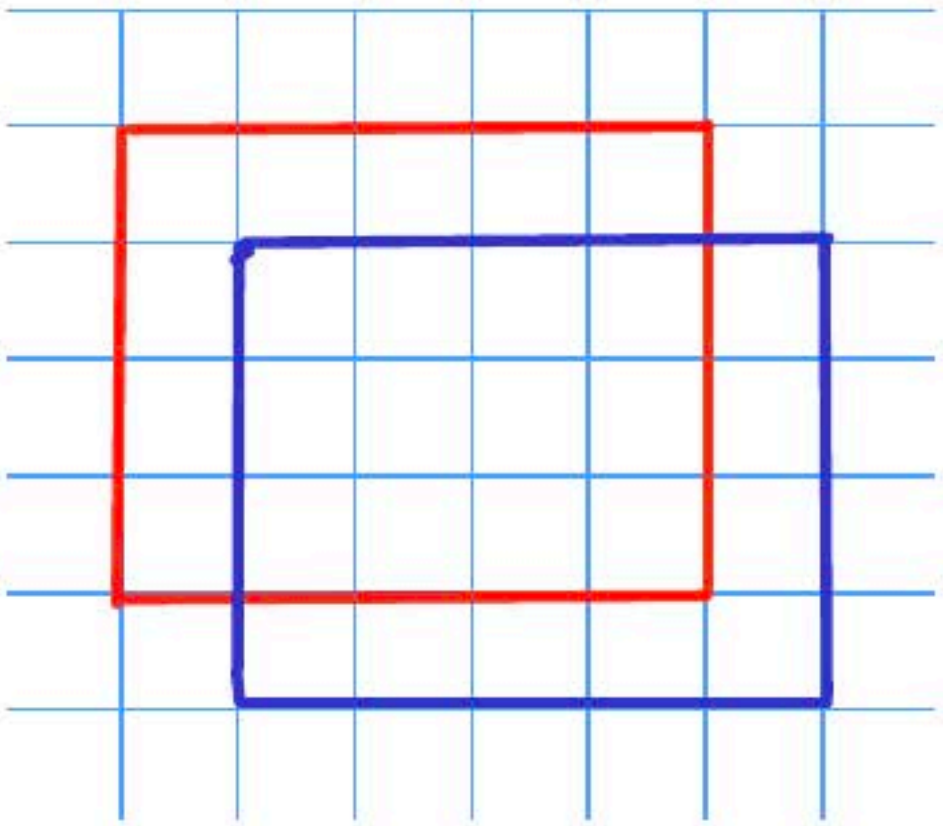
- ○ $\hat{y}^{(i)}$ has shape (N, 1)
- ○ $\hat{y}^{(i)}$ has shape (2N, 1)
- ○ $\hat{y}^{(i)}$ stores the probability that a landmark is in a given position over the face.
- ◉ $\hat{y}^{(i)}$ has shape (1, 2N)

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 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.

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 point



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

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. How many boxes will remain after non-max suppression? 1 point

- ○ 3
- ○ 6
- ○ 7
- ○ 5
- ◉ 4

7. Which of the following do you agree with about the use of anchor boxes in YOLO? Check all that apply. 1 point

- ☐ Each object is assigned to any anchor box that contains that object's midpoint.
- ☐ They prevent the bounding box from suffering from drifting.
- ☑ Each object is assigned to an anchor box with the highest IoU inside the assigned cell.
- ☑ Each object is assigned to the grid cell 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 point

This is a problem of localization? True/False

- ◉ False
- ○ True

9. Using the concept of Transpose Convolution, fill in the values of **X**, **Y** and **Z** below. 1 point

(*padding = 1, stride = 2*)

- ○ X = 0, Y = 2, Z = -1
- ○ Filter: 3x3

1	1	1
0	0	0
-1	-1	-1

- ○ Result: 6x6

	0	0	0	X	

○ 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 =-1 , Z = -7

○ Input: 2x2

1	2
3	4

10. Suppose your input to a U-Net architecture is h x w x 3, where 3 denotes your number of channels (RGB). What will be the dimension of your output ? 1 point

○ h x w x n, where n = number of of output channels

◉ h x w x n, where n = number of output classes

○ h x w x n, where n = number of input channels

○ h x w x n, where n = number of filters used in the algorithm