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

- ○ $y = [?, ?, ?, ?, ?, ?, ?, ?]$
- ◉ $y = [0, ?, ?, ?, ?, ?, ?, ?]$
- ○ $y = [1, ?, ?, ?, ?, 0, 0, 0]$
- ○ $y = [1, ?, ?, ?, ?, ?, ?, ?]$

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

What are the most appropriate (lowest number of) output units for your neural network?

- ○ Logistic unit, b_x, b_y, b_h (since $b_w = b_h$)
- ○ Logistic unit (for classifying if there is a soft-drink can in the image)
- ○ Logistic unit, b_x, b_y, b_h, b_w
- ◉ Logistic unit, b_x and b_y

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

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

Expand

Correct

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

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

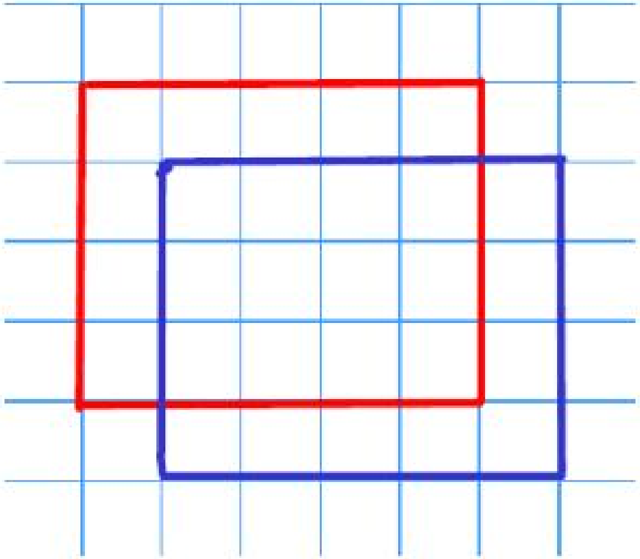
Expand

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

1 / 1 point



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

Expand

Correct

Correct. IoU is calculated as the quotient of the area of the intersection (16) over the area of the union (28).

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.

0 / 1 point

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

- (selected) False
- True

Expand

Incorrect

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. 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?

1 / 1 point

- ○ 19x19x(5x20)
- ○ 19x19x(25x20)
- ◉ 19x19x(5x25)
- ○ 19x19x(20x25)

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 (p_c), 4 coordinates (b_x, b_y, b_h, b_w) and classes ($c_1, \ldots, c_{20}$).

8. Semantic segmentation can only be applied to classify pixels of images in a binary way as 1 or 0, according to whether they belong to a certain class or not. True/False?

1 / 1 point

- [x] False
- [ ] True

Expand

Correct

Correct. The same ideas used for multi-class classification can be applied to semantic segmentation.

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	

- (selected) X = 2, Y = -6, Z = -4
- X = 2, Y = 6, Z = 4
- X = -2, Y = -6, Z = -4
- X = 2, Y = -6, Z = 4

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

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

Expand

Incorrect

To revise, watch the lecture .