Detection Algorithms

1.
Question 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 $y=[p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$.

https://www.pexels.com/es-es/foto/fotografia-de-motocicleta-clasica-en-carretera-995487/

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

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.

Question 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:

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

- ( ) 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
- ( ) Logistic unit, b_x, b_y, b_h (since $b_w = b_h$)

Expand

Correct
Correct!

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

Expand

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.

0 / 1 point

Which of the following is true about the system?

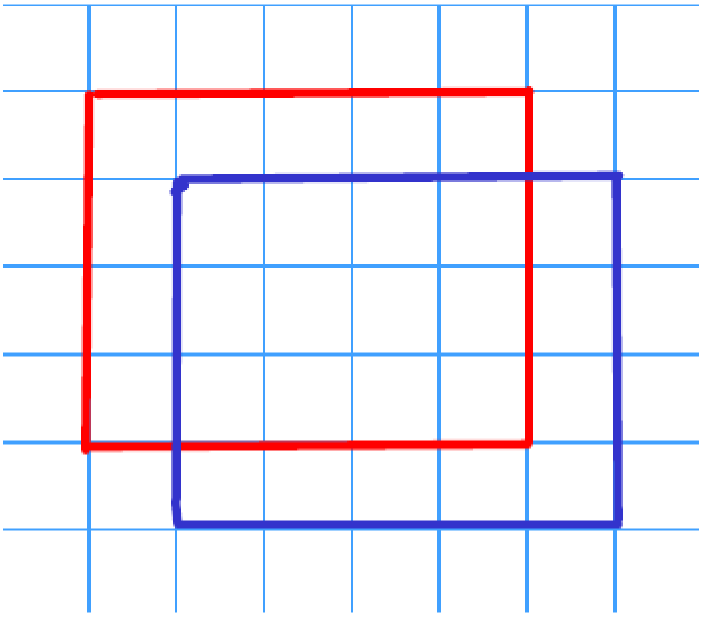
- ( ) We can't add the internet images unless they have bounding boxes.
- (•) We should add the internet images (without the presence of bounding boxes in them) to the train set.
- ( ) We can't use internet images because it changes the distribution of the dataset.
- ( ) We should use the internet images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

Expand

Incorrect
As this is a localization model, we also need the coordinates of the bounding boxes, not just the images.

5.
Question 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.



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

Expand

Incorrect
Recall that IoU is calculated as the quotient of the area of the intersection over the area of the union, not the area of only one of the boxes.

6.
Question 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 $\leq 0.7 \leq 0.7$ are discarded, and the IoU threshold for deciding if two boxes overlap is $0.5 0.5$.

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

- False
- True

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

1 / 1 point

- [ ] Each object is assigned to any anchor box that contains that object's midpoint.
- [x] 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.

- [ ] They prevent the bounding box from suffering from drifting.
- [x] 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.

Expand

Correct
Great, you got all the right answers.

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.
Question 9

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

(*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	

1 / 1 poin

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

Expand

Correct

Correct.

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 / 1 point

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

Expand

Correct