Try again once you are ready

Grade received 70%
Latest Submission Grade 70%
To pass 80% or higher

Try again

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

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

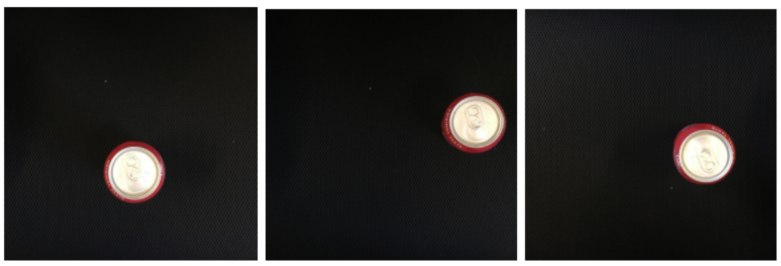
- $y = [1, 0.22, 0.5, 0.2, 0.3, 1, 1, 1]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 1]$
- $y = [1, 0.22, 0.5, 0.2, 0.3, ?, ?, 1]$
- $$y = [1, 0.22, 0.5, 0.2, 0.3, 0, 0, 0]$$

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. 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.
- True, since this is a localization problem.
- False, since we only need two values

  c_1

  for no soft-drink can and

Expand

Incorrect
Since there is only one object we can use only $$c_1$$ or simply $$p_c$$.

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), we need two coordinates for each landmark, thus we need 2N output units. True/False? **1 / 1 point**

- True
- False

Expand

Correct
Correct. Recall that each landmark is a specific position in the face's image, thus we need to specify two coordinates for each landmark.

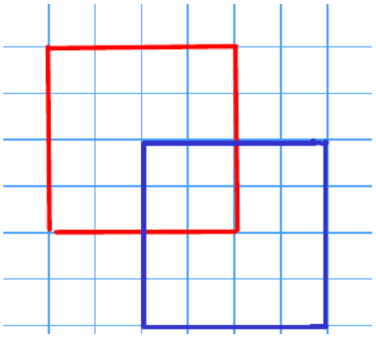
4. When training one of the object detection systems described in the lectures, you need a training set that contains many pictures of the object(s) you wish to detect. However, bounding boxes do not need to be provided in the training set, since the algorithm can learn to detect the objects by itself. **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**



- $\frac{1}{4}$
- $\frac{1}{8}$
- $\frac{1}{7}$
- 1

Expand

Correct
Correct. IoU is calculated as the quotient of the area of the intersection (4) over the area of the union (28).

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. **1 / 1 point**

Notice that there are three bounding boxes for cars. After running non-max suppression, only the bounding box of the car with 0.73 is kept from the three bounding boxes for cars. True/False? Choose the best answer.

- False. Two bounding boxes corresponding to cars are left since their IoU is zero.
- False. All the cars are eliminated since there is a pedestrian with a higher score of 0.98.
- True. The non-maximum suppression eliminates the bounding boxes with scores lower than the ones of the maximum.

Expand

Correct
Correct. The bounding box for the car on the right is eliminated because its probability is less than 0.4. Of the two bounding boxes in the middle, one is eliminated because their IoU is higher than 0.5. So, only one bounding box remains.

7. Which of the following do you agree with about the use of anchor boxes in YOLO? Check all that apply. **0 / 1 point**

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

This should not be selected
There is more than just that to assign anchor boxes.

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

This should not be selected
There were no drifting phenomena discussed in the lectures.

Expand

Incorrect
You didn't select all the correct answers

8. What is Semantic Segmentation? **1 / 1 point**

- (•) Locating objects in an image by predicting each pixel as to which class it belongs to.
- ( ) 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.

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?

0 / 1 point

- ○ False
- ◉ True

Expand

Incorrect
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.