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

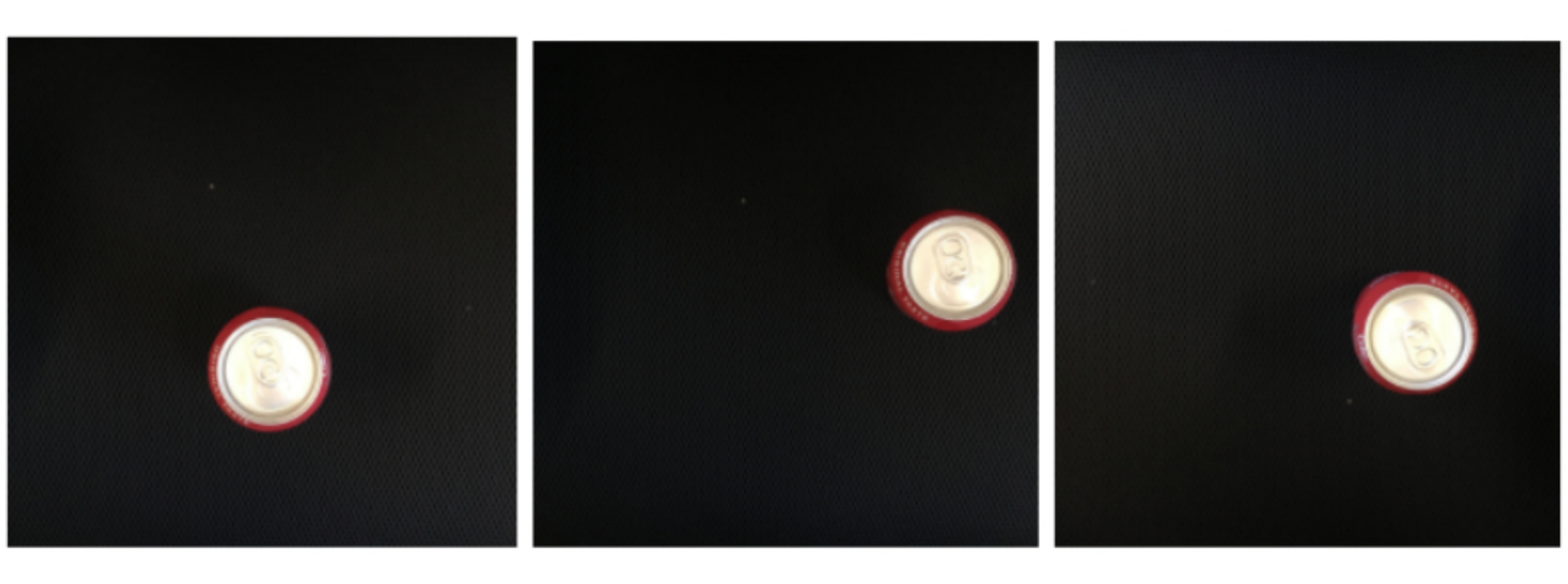
https://www.pexels.com/es-es/foto/fotografia-de-motocicleta-clasica-en-carretera-995487/

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

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

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

- ○ False
- ◉ True

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.

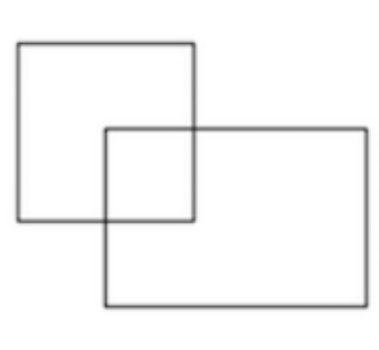
4. When training one of the object detection systems described in the lectures, you need a training set that contains many pictures of the object(s) you wish to detect. However, bounding boxes do not need to be provided in the training set, since the algorithm can learn to detect the objects by itself. — 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 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
- ◉ $\frac{1}{9}$
- ○ $\frac{1}{10}$
- ○ $\frac{1}{6}$

Expand

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 ≤ 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.

- ◉ True. The non-maximum suppression eliminates the bounding boxes with scores lower than the ones of the maximum.
- ○ False. Two bounding boxes corresponding to cars are left since their IoU is zero.

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

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. 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

- ○ True
- ◉ False

Expand

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	X	
	Y	4	2	2	
	0	0	0	0	
	-3	Z	-4	-4	

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

Expand

Correct
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 $h \times w$. True/False? — 1 / 1 point

- ○ True
- ◉ False

Expand

Correct
Correct. The output of the U-Net architecture can be $h \times w \times k$ where k is the number of classes.