← Back

Detection Algorithms

Graded Assignment • 50 min

English

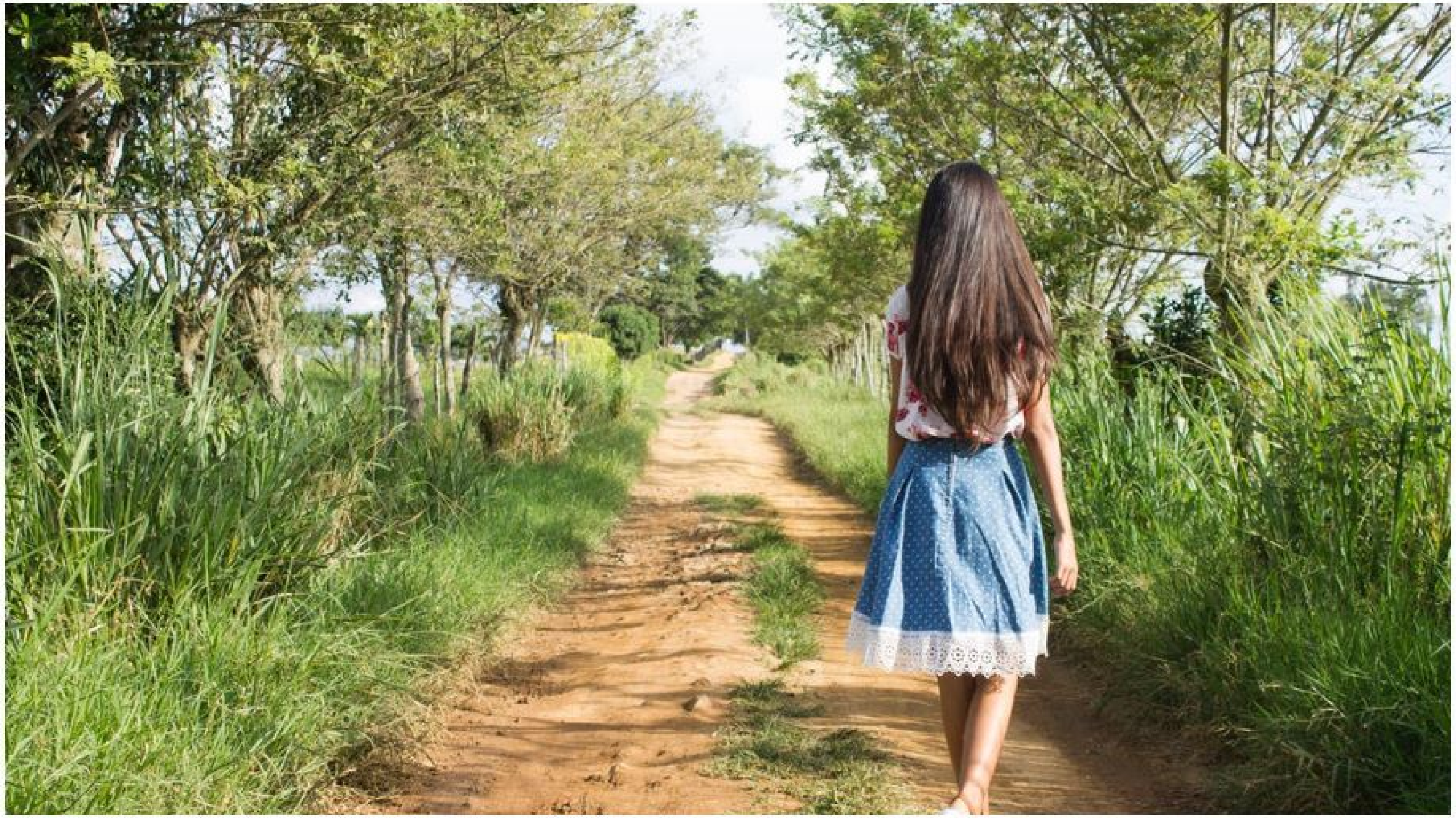
Your grade: 90%

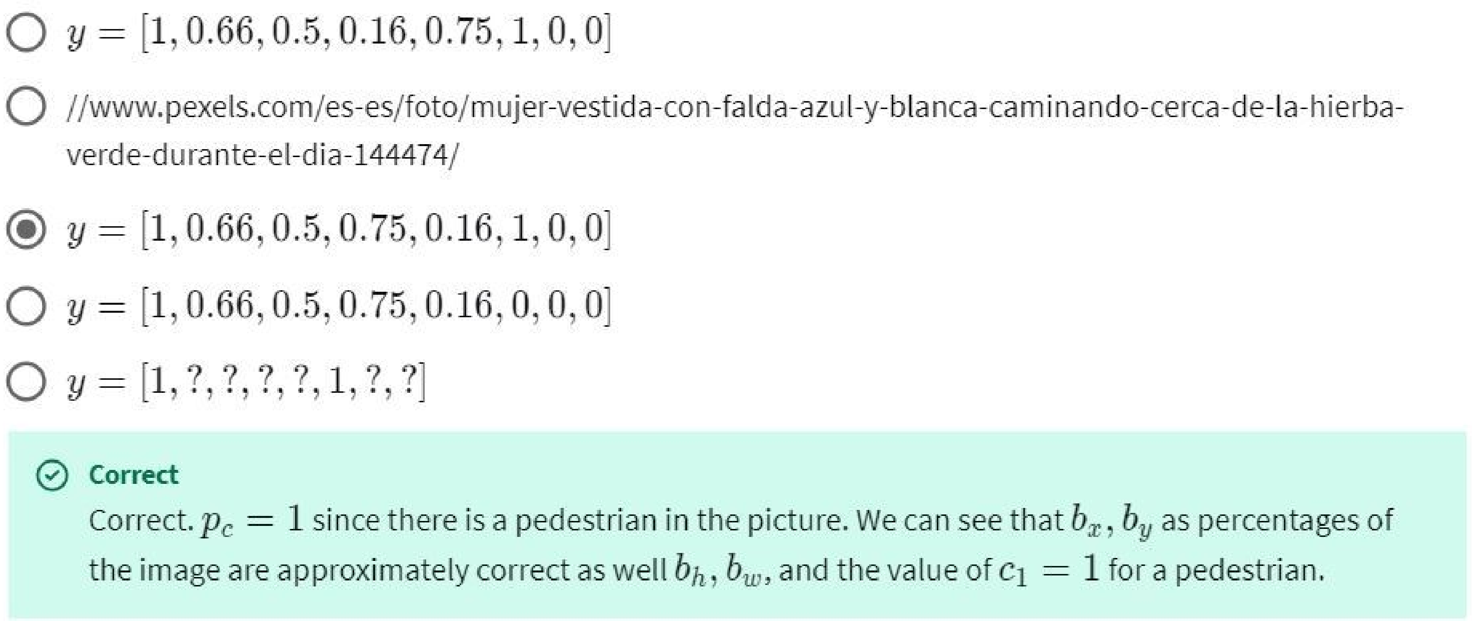
Your latest: **80%** • Your highest: **90%** • To pass you need at least 80%. We keep your highest score.

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





○ $y = [1, 0.66, 0.5, 0.16, 0.75, 1, 0, 0]$

○ //www.pexels.com/es-es/foto/mujer-vestida-con-falda-azul-y-blanca-caminando-cerca-de-la-hierba-verde-durante-el-dia-144474/

● $y = [1, 0.66, 0.5, 0.75, 0.16, 1, 0, 0]$

○ $y = [1, 0.66, 0.5, 0.75, 0.16, 0, 0, 0]$

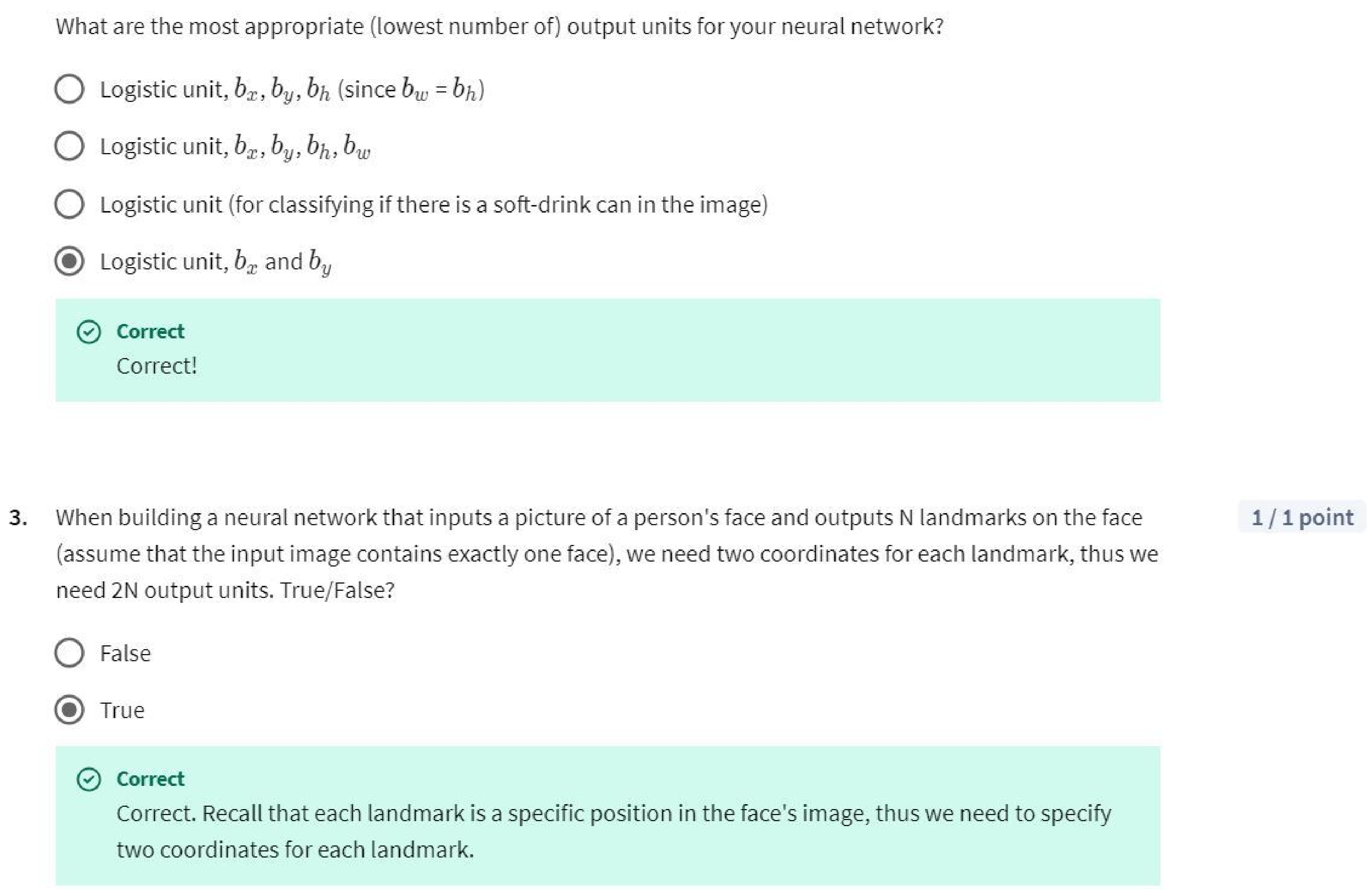
○ $y = [1, ?, ?, ?, ?, 1, ?, ?]$

Correct

Correct. $p_c = 1$ since there is a pedestrian in the picture. We can see that b_x, b_y as percentages of the image are approximately correct as well b_h, b_w, and the value of $c_1 = 1$ for a pedestrian.

1 / 1 point

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, b_x, b_y, b_h (since $b_w = b_h$)

○ Logistic unit, b_x, b_y, b_h, b_w

○ Logistic unit (for classifying if there is a soft-drink can in the image)

● Logistic unit, b_x and b_y

✓ **Correct**
Correct!

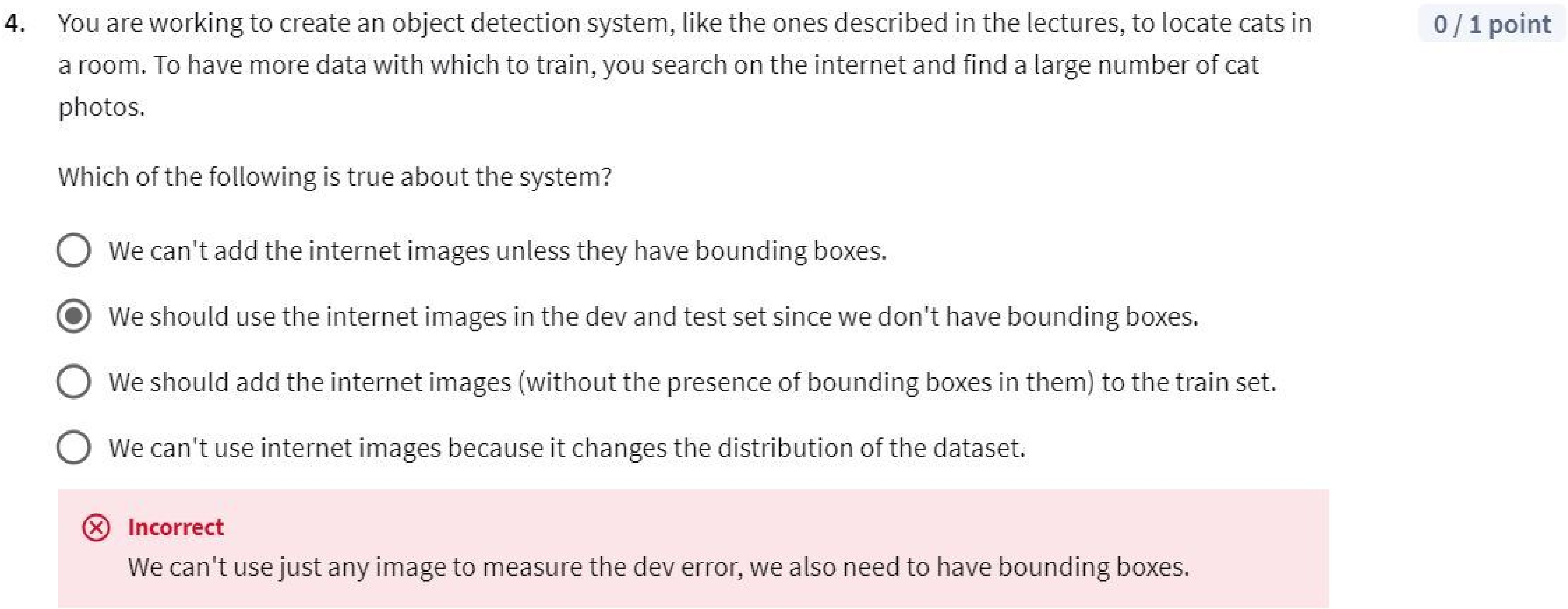
3. When building a neural network that inputs a picture of a person's face and outputs N landmarks on the face (assume that the input image contains exactly one face), we need two coordinates for each landmark, thus we need 2N output units. True/False?

1 / 1 point

○ False

● True

✓ **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. 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 use the internet images in the dev and test set since we don't have bounding boxes.

○ We should add the internet images (without the presence of bounding boxes in them) to the train set.

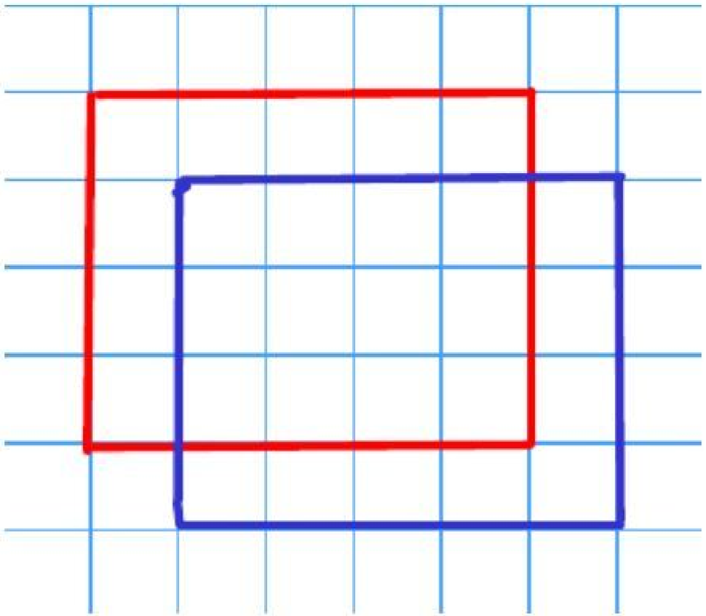
○ We can't use internet images because it changes the distribution of the dataset.

⊗ **Incorrect**

We can't use just any image to measure the dev error, we also need to have bounding boxes.

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



○ $\frac{2}{5}$

○ $\frac{4}{5}$

● $\frac{3}{7}$

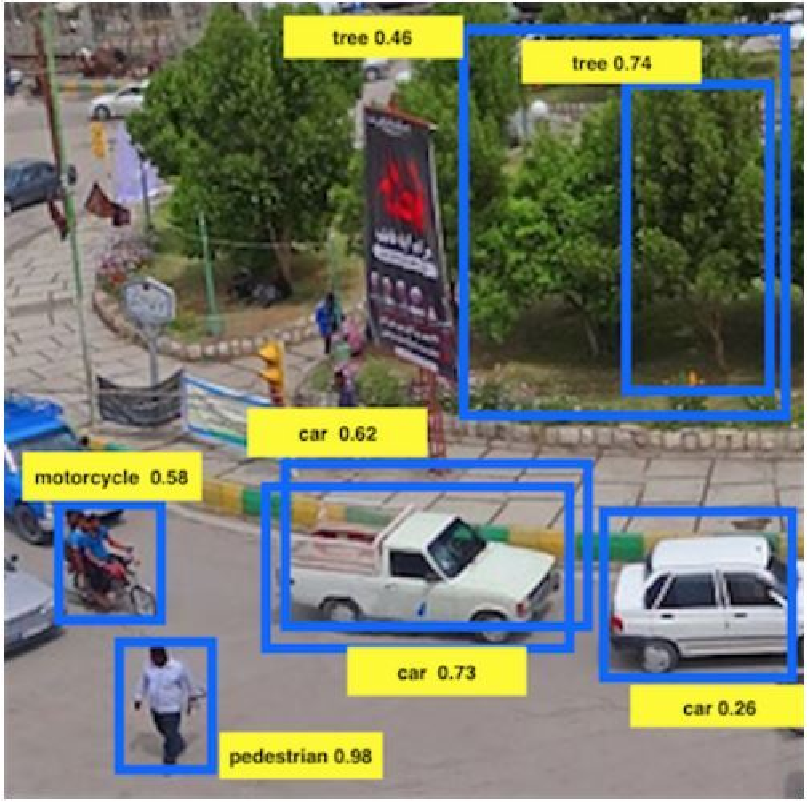
○ $\frac{1}{2}$

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.

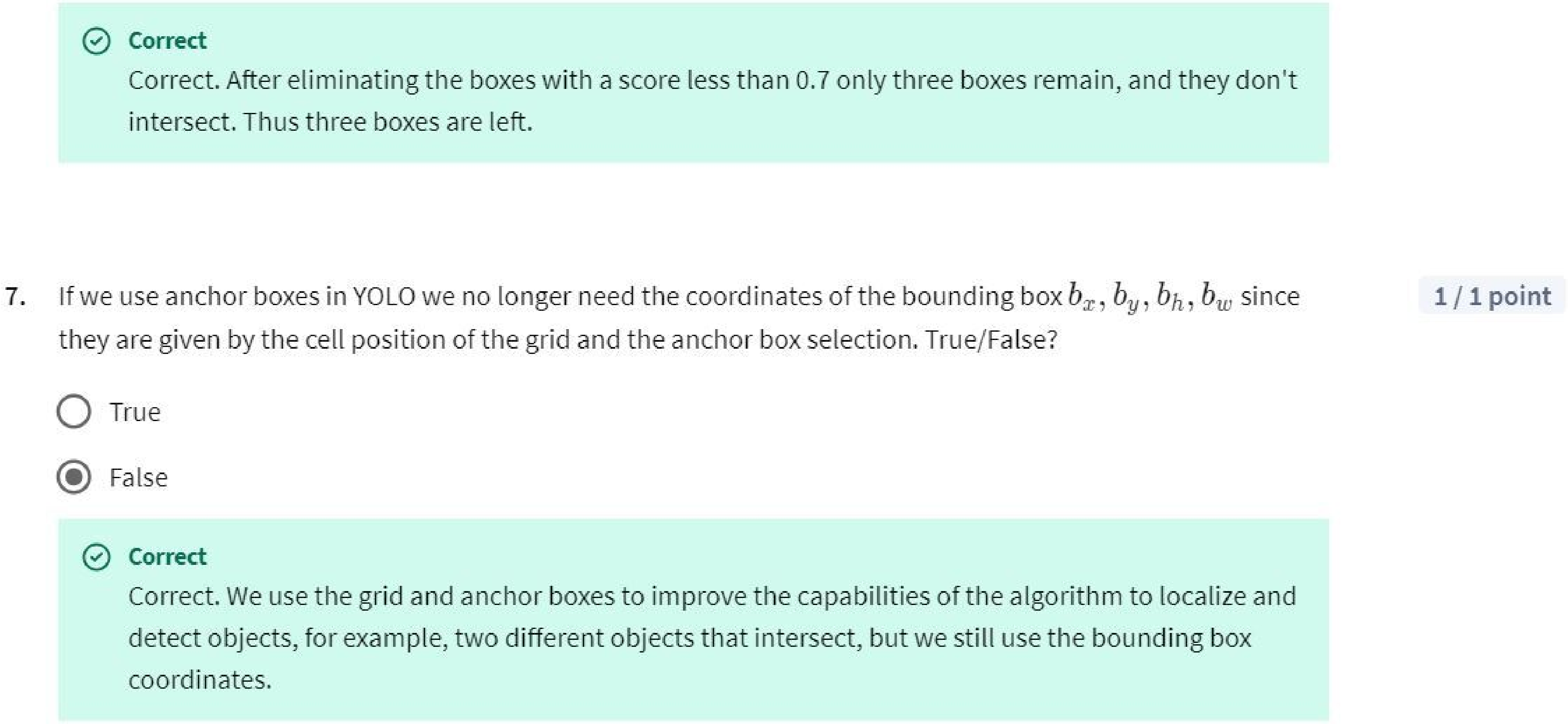
1 / 1 point



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

○ False

● True



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. If we use anchor boxes in YOLO we no longer need the coordinates of the bounding box b_x, b_y, b_h, b_w since they are given by the cell position of the grid and the anchor box selection. True/False?

1 / 1 point

○ True

● False

Correct

Correct. We use the grid and anchor boxes to improve the capabilities of the algorithm to localize and detect objects, for example, two different objects that intersect, but we still use the bounding box coordinates.

8. What is Semantic Segmentation?

1 / 1 point

○ Locating an object in an image belonging to a certain class by drawing a bounding box around it.

● Locating objects in an image by predicting each pixel as to which class it belongs to.

○ Locating objects in an image belonging to different classes by drawing bounding boxes around them.

✓ **Correct**

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

0 / 1 point

(padding = 1, stride = 2)

○ X = 0, Y = 2, Z = -7

○ 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	

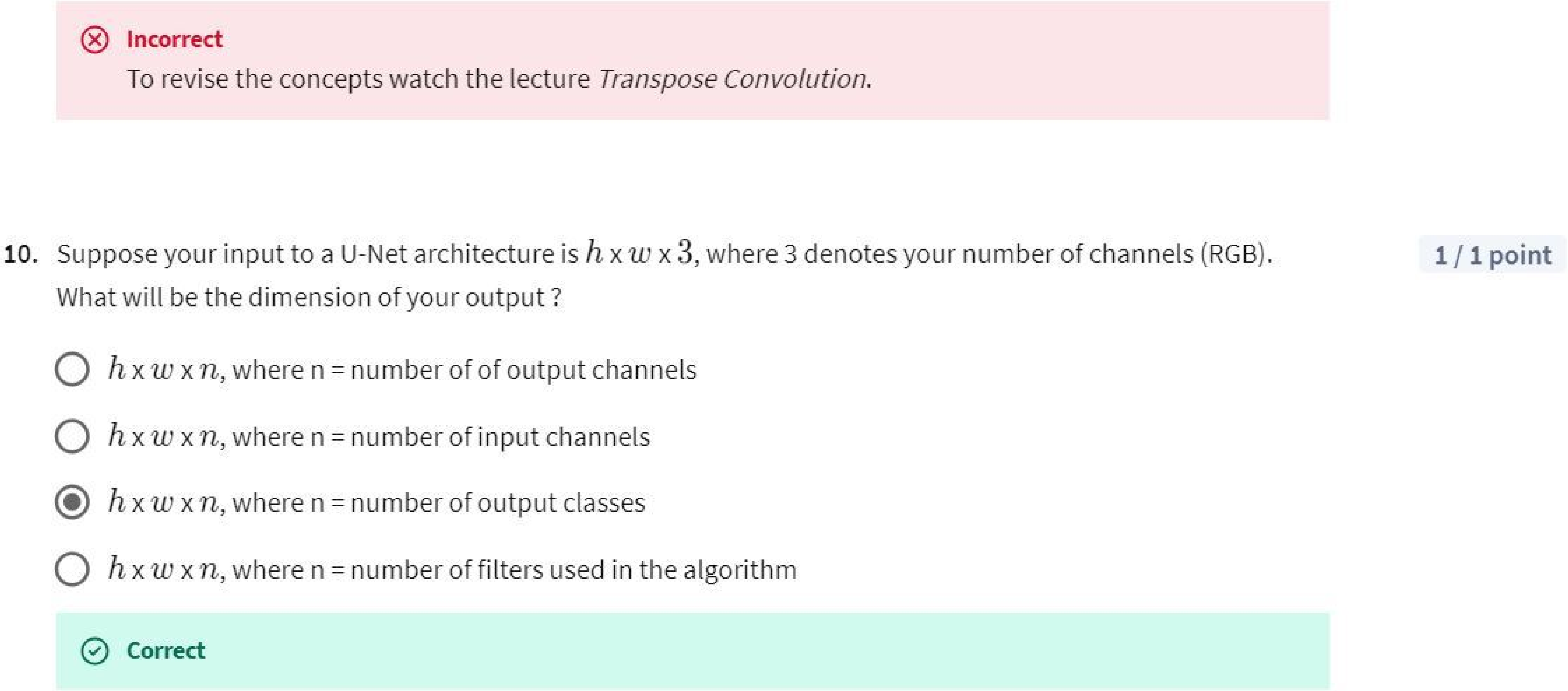
○ Input: 2x2

1	2
3	4

○ X = 0, Y = 2, Z = -1

● X = 0, Y =-1 , Z = -7

○ X = 0, Y = -1, Z = -4



⊗ **Incorrect**

To revise the concepts watch the lecture *Transpose Convolution.*

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 of output channels

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

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

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

✓ **Correct**