Back Detection Algorithms
Graded Quiz • 50 min

Due Mar 12, 11:59 PM IST

Congratulations! You passed!

Grade received 90% | Latest Submission Grade 90% | To pass 80% or higher

Go to 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/mujer-vestida-con-falda-azul-y-blanca-caminando-cerca-de-la-hierba-verde-durante-el-dia-144474/

- $y = [1, ?, ?, ?, ?, 1, ?, ?]$
- $y = [1, 0.66, 0.5, 0.16, 0.75, 1, 0, 0]$
- $y = [1, 0.66, 0.5, 0.75, 0.16, 1, 0, 0]$ (selected)
- $y = [1, 0.66, 0.5, 0.75, 0.16, 0, 0, 0]$

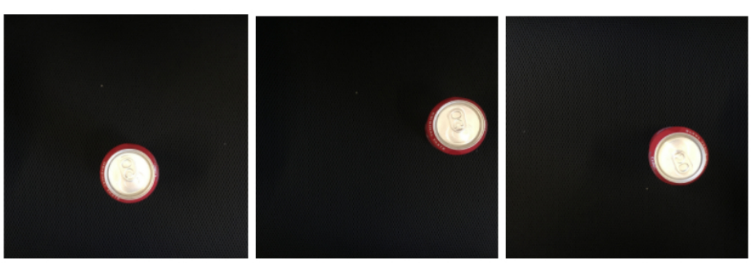
Expand

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.

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)
- (selected) Logistic unit, b_x and b_y
- Logistic unit, b_x, b_y, b_h, b_w

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 (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)
- (selected) $\hat{y}^{(i)}$ has shape (2N, 1)

Expand

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

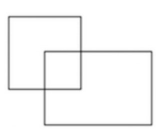
4. When training one of the object detection systems described in the lectures, each image must have zero or exactly one bounding box. True/False? **1 / 1 point**

- (selected) False
- True

Expand

Correct
Correct. In a single image, there might be more than only one instance of the object we are trying to localize, so it must have several bounding boxes.

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**



- $\frac{1}{6}$
- None of the above
- $\frac{1}{10}$
- (selected) $\frac{1}{9}$

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 **1 / 1 point**

suppression are that boxes with probability ≤ 0.4 are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5.

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.

- (selected) True. The non-maximum suppression eliminates the bounding boxes with scores lower than the ones of the maximum.
- False. All the cars are eliminated since there is a pedestrian with a higher score of 0.98.
- 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. Which of the following do you agree with about the use of anchor boxes in YOLO? Check all that apply.

0 / 1 point

- [ ] Each object is assigned to an anchor box with the highest IoU inside the assigned cell.
- [ ] They prevent the bounding box from suffering from drifting.
- [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 the grid cell that contains that object's midpoint.

Expand

Incorrect
You didn't select all the correct answers

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

- True
- (selected) False

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

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 $h \times w \times c$. True/False?

1 / 1 point

- ◉ False
- ○ True

Expand

Correct
Correct. 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.