Congratulations! You passed!

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

Retake the assignment in **23h 44m**

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

- (●) $y = [0, ?, ?, ?, ?, ?, ?, ?]$
- ( ) $y = [?, ?, ?, ?, ?, ?, ?, ?]$
- ( ) $y = [1, ?, ?, ?, ?, 0, 0, 0]$
- ( ) $y = [1, ?, ?, ?, ?, ?, ?, ?]$

Expand

Correct
Correct.

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're some typical images in your training set:

1 / 1 point

To solve this task it is necessary to divide the task into two: 1. Construct a system to detect if a can is present or not. 2. Construct a system that calculates the bounding box of the can when present. Which one of the following do you agree with the most?

- (●) We can approach the task as an image classification with a localization problem.
- ( ) The two-step system is always a better option compared to an end-to-end solution.
- ( ) We can't solve the task as an image classification with a localization problem since all the bounding boxes have the same dimensions.
- ( ) An end-to-end solution is always superior to a two-step system.

Expand

Correct
Correct. We can use a network to combine the two tasks similar to that described in the lectures.

3. If you build a neural network that inputs a picture of a person's face and outputs N landmarks on the face (assume the input image always contains exactly one face), how many output units will the network have? **1 / 1 point**

- 3N
- 2N
- N
- N^2

Expand

Correct
Correct

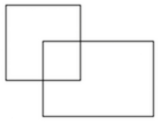
4. When training one of the object detection systems described in the lectures, each image must have zero or exactly one bounding box. True/False? **1 / 1 point**

- True
- False

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**

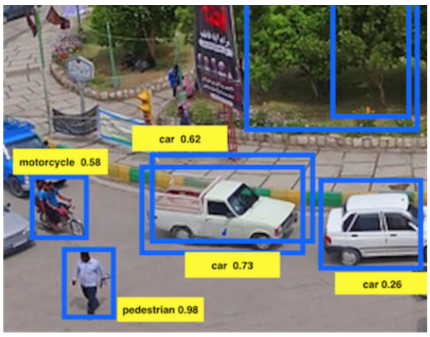


- $\frac{1}{9}$
- None of the above
- $\frac{1}{10}$
- $\frac{1}{6}$

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. How many boxes will remain after non-max suppression? **1 / 1 point**



- 7
- 4
- 6
- 5
- 3

Expand

Correct
Correct!

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 the grid cell that contains that object's midpoint.

Correct
Correct. This is the way we choose the corresponding cell.

- [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.

Expand

Correct
Great, you got all the right 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? **0 / 1 point**

- False
- True

Expand

Incorrect
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	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? **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. The number of channels doesn't have to match between input and output.