

[Store](#)[Features](#)[Docs](#)[Share](#)[Support](#)[News](#)[Contact Us!](#)[About](#)[Sign Up](#)[Sign In](#)

Real-time collaboration for Jupyter Notebooks,
Linux Terminals, LaTeX, VS Code, R IDE, and more,
all in one place.

[leechanwoo-kor](#)[Edit your own copy...](#)[GitHub Repository: leechanwoo-kor/coursera](#)[☆ Star](#)

[Path: blob/main/deep-learning-specialization/course-4-convolutional-neural-network/Week 3 Quiz - Detection Algorithms.md](#)

Views: 23928

Week 3 Quiz - Detection Algorithms

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 $y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$.



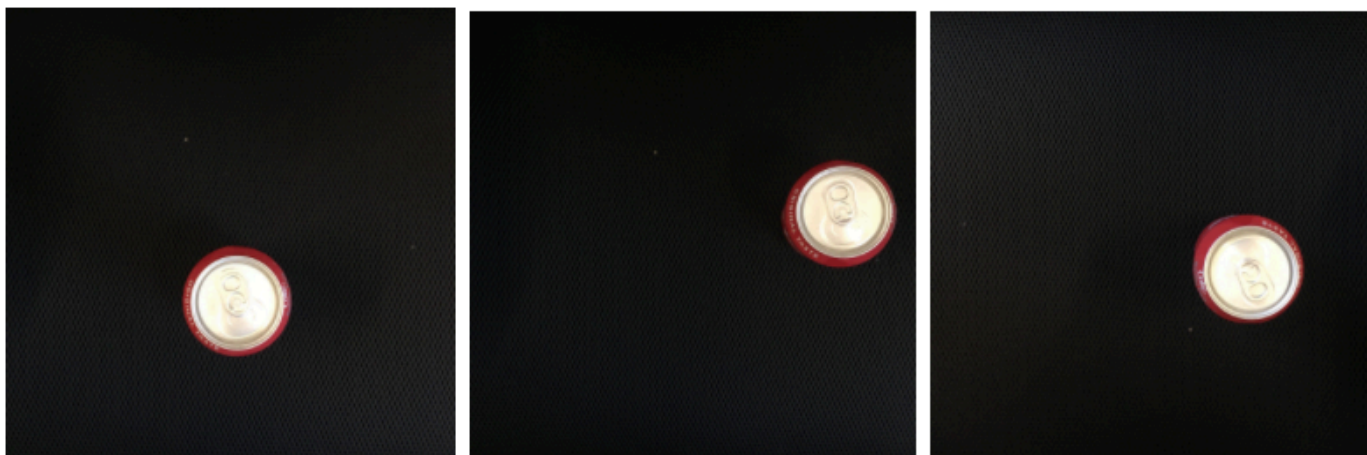
<https://www.pexels.com/es-es/foto/mujer-vestida-con-falda-azul-y-blanca-caminando-cerca-de-la-hierba-verde-durante-el-dia-144474/>

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

☐ ...

✶ $p_c = 1$ 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're some typical images in your training set:



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

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 $\hat{y}^{(i)}$?

- ☐ $\hat{y}^{(i)}$ has shape (1, 2N)
- ☐ $\hat{y}^{(i)}$ has shape (N, 1)
- ☒ $\hat{y}^{(i)}$ has shape (2N, 1)
- ☐ $\hat{y}^{(i)}$ stores the probability that a landmark is in a given position over the face.

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

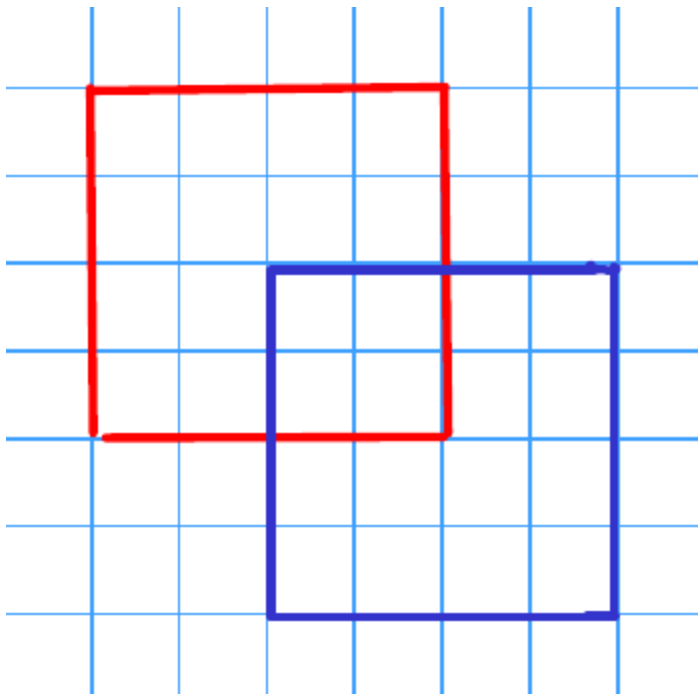
4. When training one of the object detection systems described in the lectures, you need a training set that contains many pictures of the object(s) you wish to detect. However, bounding boxes do not need to be provided in the training set, since the algorithm can learn to detect the objects by itself.

- ☐ True

☒ False

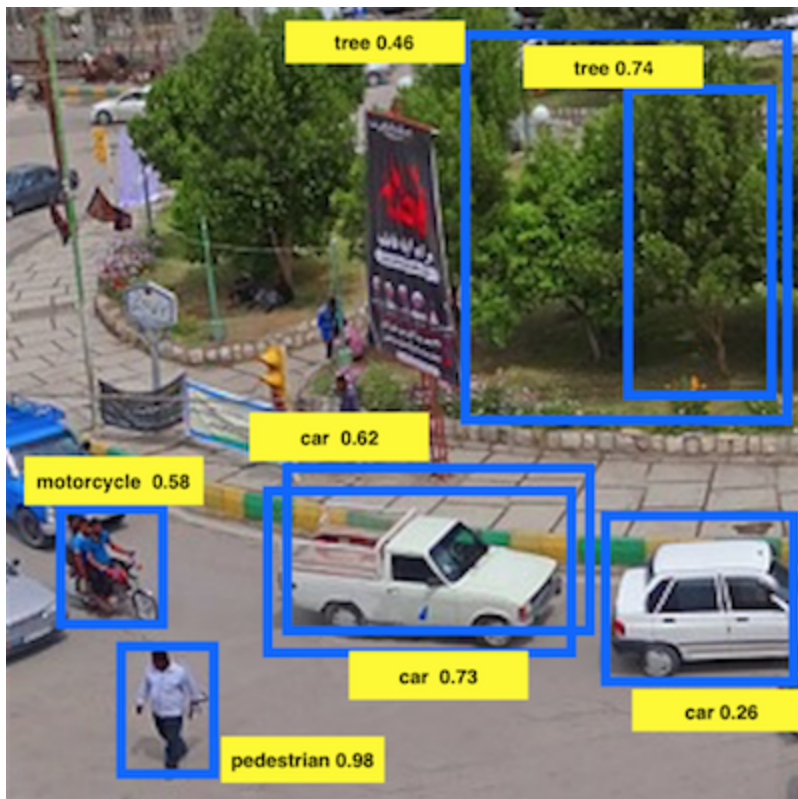
✦ You need bounding boxes in the training set. Your loss function should try to match the predictions for the bounding boxes to the true bounding boxes from the training set.

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.

☒ $\frac{1}{7}$
☐ ...

✦ 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. How many boxes will remain after non-max suppression?


☒ 5

☐ ...

7. Suppose you are using YOLO on a 19×19 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?

☒ $19 \times 19 \times (5 \times 25)$
☐ ...

✦ You get a 19×19 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, \dots, c_{20}).

8. What is Semantic Segmentation?

☒ 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.

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

9. Using the concept of Transpose Convolution, fill in the values of X, Y and Z below. (padding = 1, stride = 2)

Input: 2×2

12

34

Filter: 3×3

10-1

10-1

10-1

Result: 6×6

010-2

0X0Y

010Z

010-4

☒ X = 2, Y = -6, Z = -4

☐ ...

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?

☐ True

☒ False

✦ 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.



Product

Store

Resources

Documentation

Company

About

- Features
- Licenses
- Pricing
- On Premises
- System Activity
- Status

- Compute Servers
- Public Share
- Software
- Support
- Contact
- Events
- Team
- News
- Policies