Question 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 y^(i)?

The output y^(i) of a neural network that outputs N landmarks on a face would have shape (2N, 1).

The reason for this is that each landmark on the face will have two coordinates: an x-coordinate and a y-coordinate. Hence, for N landmarks, there would be 2N coordinates. The shape of the output y^(i) thus would be (2N, 1), where each pair of values corresponds to the x and y coordinates of a landmark.

Note that y^(i) does not store the probability that a landmark is in a given position over the face. Instead, it directly predicts the coordinates of the landmarks. 1 represents the column

Question 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?

When using YOLO on a 19x19 grid with 20 classes and 5 anchor boxes, the output volume dimension can be computed as follows:

For each anchor box, we have:

20 class probabilities

5 bounding box values (x, y, h, w, and objectness score)

So, for each anchor box, we have a total of 20 + 5 = 25 values.

Since we have 5 anchor boxes, the total values per grid cell are 5 \* 25 = 125.

As the grid is 19x19, the output volume's dimension will be 19x19x125.

When using the U-Net architecture with an input

ℎ

×

�

×

�

h×w×c, where

�

c denotes the number of channels, the output will always have the shape

ℎ

×

�

×

�

h×w×c. True/False?

Which of the following do you agree with about the use of anchor boxes in YOLO? Check all that apply?

how to choose correct anchor box among that 5 anchor boxes:

During the inference process, each anchor box will predict an object and the associated bounding box. Each prediction will be associated with a confidence score, which represents how confident the model is that the predicted bounding box actually encloses an object, and how accurate it thinks the box is.

To choose the correct anchor box among the 5 (or however many you have), you typically pick the one with the highest confidence score. However, to prevent multiple boxes for the same object, non-max suppression is used.

In non-max suppression, you:

Sort the predictions by the confidence scores.

Start from the top of this list, select that prediction and remove any remaining predictions that have high Intersection over Union (IoU) with the one you just picked.

Repeat step 2 until no more predictions remain.

At the end of this process, you will be left with the predictions that the model believes to be the most accurate.