Convolutional Neural Networks

Week 4: Special Applications: Face Recognition and Neural Style Transfer

1 Face verification requires comparing a new picture against one person's face, whereas face recognition requires comparing a new picture against K persons' faces.

Answer: True.

2 Why is the face verification problem considered a one-shot learning problem? Choose the best answer.

Answer: Because we might have only one example of the person we want to verify.

Comment: One-shot learning refers to the amount of data we have to solve a task.

3 Version 1: You want to build a system that receives a person's face picture and determines if the person is inside a workgroup. You have pictures of all the faces of the people currently in the workgroup, but some members might leave, and some new members might be added. To train a system to solve this problem using the triplet loss you get many persons and take several pictures of each one. Which of the following do you agree with? (Select the best answer.)

Answer: You take several pictures of the same person of train $d(img_1, img_2)$ using the triplet loss.

Comment: To train using the triplet loss you need several pictures of the same person.

3 Version 2: In order to train the parameters of a face recognition system, it would be reasonable to use a training set comprising 100,000 pictures of 100,000 different persons.

Answer: False.

4 Which of the following is a correct definition of the triplet loss? Consider that $\alpha > 0$. (We encourage you to figure out the answer from first principles, rather than just refer to the lecture.)

Answer: $max(||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + \alpha, 0)$.

5 Consider the following Siamese network architecture: The upper and lower neural networks have different input images, but have exactly the same parameters.

Answer: True.

Comment: Parameters are shared among these two networks.

6 You train a ConvNet on a dataset with cats, dogs, birds, and other types of animals. You try to find a filter that strongly responds to horizontal edges. You are more likely to find this filter in layer 6 of the network than in layer 1. True/False?

Answer: False.

Comment: Edges are a very low-level feature, thus it is more likely to find such a feature detector in the first layers of the network.

7 Version 1: In neural style transfer, we train the pixels of an image, and not the parameters of a network.

Answer: True.

7 Version 2: Neural style transfer is trained as a supervised learning task in which the goal is to input two images (x), and train a network to output a new, synthesized image (y).

Answer: False.

8 Version 1: In neural style transfer the content loss J_{cont} is computed as: $J_{cont}(G, C) = ||a^{[l](C)} - a^{[l](G)}||^2$ where $a^{[l](k)}$ is the activation of the l-th layer of a ConvNet trained for classification. We choose l to be a very high value to use compared to the more abstract activation of each image. True/False?

Answer: False.

Comment: We don't use a very deep layer since this will only compare if the two images belong to the same category.

8 Version 2: In the deeper layers of a ConvNet, each channel corresponds to a different feature detector. The style matrix $G^{[l]}$ measures the degree to which the activations of different feature detectors in layer l vary (or correlate) together with each other.

Answer: True.

9 Version 1: In neural style transfer, we can't use gradient descent since there are no trainable parameters. True/False?

Answer: False.

Comment: We use gradient descent on the cost function J(G) and we update the pixel values of the generated image G.

9 Version 2: In neural style transfer, what is updated in each iteration of the optimization algorithm?

Answer: The pixel values of the generated image G.

10 You are working with 3D data. The input image has size 64 by 64 by 64 by 3, if you apply a convolutional layer with 16 filters of size 4 by 4 by 4, zero padding and stride 2. What is the size of the output volume?

Answer: 31 by 31 by 31 by 16.

Comment: We can use the formula $[(n^{[l-1]} - f + 2 \times p)/s] + 1 = n^{[l]}$ to the three first dimensions.