Congratulations! You passed!

Grade received 90% To pass 80% or higher

Go to next item

Special Applications: Face Recognition & Neural Style Transfer

Latest Submission Grade 90%

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. **1 / 1 point**

- False
- True

Expand

Correct
Correct.

2. Why is the face verification problem considered a one-shot learning problem? Choose the best answer. **1 / 1 point**

- Because we have only have to forward pass the image one time through our neural network for verification.
- Because we might have only one example of the person we want to verify.
- Because of the sensitive nature of the problem, we won't have a chance to correct it if the network makes a mistake.
- Because we are trying to compare to one specific person only.

Expand

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

3. 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. **1 / 1 point**

- True
- False

Expand

Correct
Correct, to train a network using the triplet loss you need several pictures of the same person.

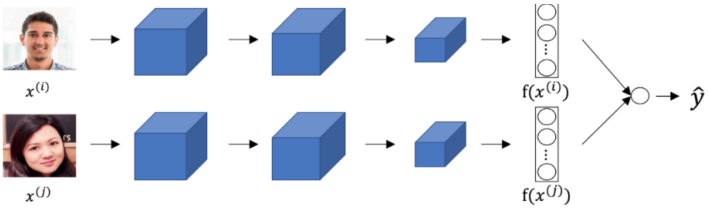
4. Which of the following is a correct definition of the triplet loss? Consider that $\alpha > 0$. (We encourage you to figure out the answer from first principles, rather than just refer to the lecture.) **1 / 1 point**

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

Expand

Correct
Correct

5. Consider the following Siamese network architecture: **1 / 1 point**



The upper and lower networks share parameters to have a consistent encoding for both images. True/False?

- (●) True
- ( ) False

Expand

Correct
Correct. Part of the idea behind the Siamese network is to compare the encoding of the images, thus they must be consistent.

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? **0 / 1 point**

- (●) True
- ( ) False

Expand

Incorrect
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. Neural style transfer uses images Content C, Style S. The loss function used to generate image G is composed of which of the following: (Choose all that apply.) **1 / 1 point**

- [x] J_{style} that compares S and G.

Correct
Correct, in neural style transfer we are interested in the similarity between S and G, and the similarity between G and C.

- [ ] T that calculates the triplet loss between S, G, and C.
- [x] $J_{content}$ that compares C and G.

Correct
Correct, in neural style transfer we are interested in the similarity between S and G, and the similarity between G and C.

- [ ] J_{corr} that compares C and S.

Expand

Correct
Great, you got all the right answers.

8. 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. **1 / 1 point**

- ( ) False
- (●) True

Expand

Correct
Yes, the style matrix $G^{[l]}$ can be seen as a matrix of cross-correlations between the different feature detectors.

9. In neural style transfer, what is updated in each iteration of the optimization algorithm? **1 / 1 point**

- The neural network parameters
- The regularization parameters
- The pixel values of the content image C
- The pixel values of the generated image G

Expand

Correct
Yes, neural style transfer is different from many of the algorithms you've seen up to now, because it doesn't learn any parameters; instead it learns directly the pixels of an image.

10. You are working with 3D data. You are building a network layer whose input volume has size 32x32x32x16 (this volume has 16 channels), and applies convolutions with 32 filters of dimension 3x3x3x16 (no padding, stride 1). What is the resulting output volume? **1 / 1 point**

- 30x30x30x32
- 30x30x30x16
- Undefined: This convolution step is impossible and cannot be performed because the dimensions specified don't match up.

Expand

Correct
Correct, you have used the formula $\lfloor \frac{n^{[l-1]} - f + 2 \times p}{s} \rfloor + 1 = n^{[l]}$ over the three first dimensions of the input data.