

Week 4: 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 person's faces.

Ans: True

2. Why do we learn a function $d(img1, img2)$ for face verification? (Select all that apply.)

Ans:

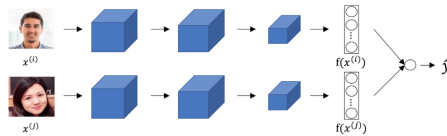
- This allows us to learn to recognize a new person given just a single image of that person.
 - We need to solve a one-shot learning problem.
3. In order to train the parameters of a face recognition system, it would be reasonable to use a training set comprising 100,000 pictures of 100,000 different persons.

Ans: False

4. Which of the following is a correct definition of the triplet loss? Consider that $\alpha > 0$

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

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



Ans: True

6. You train a CNN on a dataset with 100 different classes. You wonder if you can find a hidden unit which responds strongly to pictures of cats. You are more likely to find this unit in layer-4 of the network than in layer-1.

Ans: True

7. 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 .

Ans: False

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.

Ans: True

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

Ans: Pixel values of generated images.

10. You are working with 3D data. You are building a network layer whose input volume has size $32 \times 32 \times 32 \times 16$, and applies convolutions with 32 filters of dimension $3 \times 3 \times 3$ (no padding, stride of 1). What is the resulting output volume?

Ans: Using $O = \left\lfloor \frac{N+2p-f}{s} + 1 \right\rfloor$ and having 32 channels, we get $30 \times 30 \times 30 \times 32$