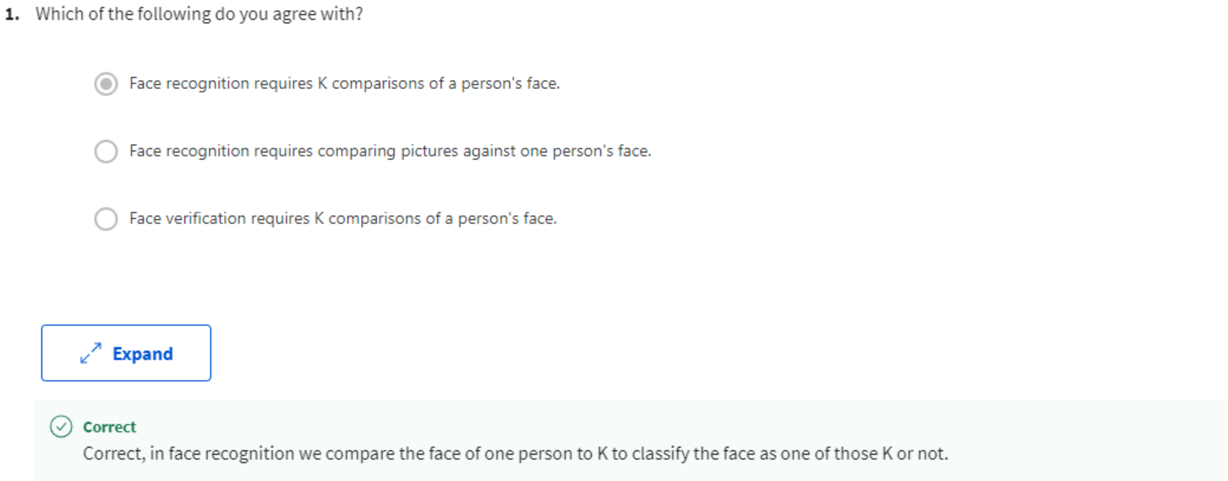


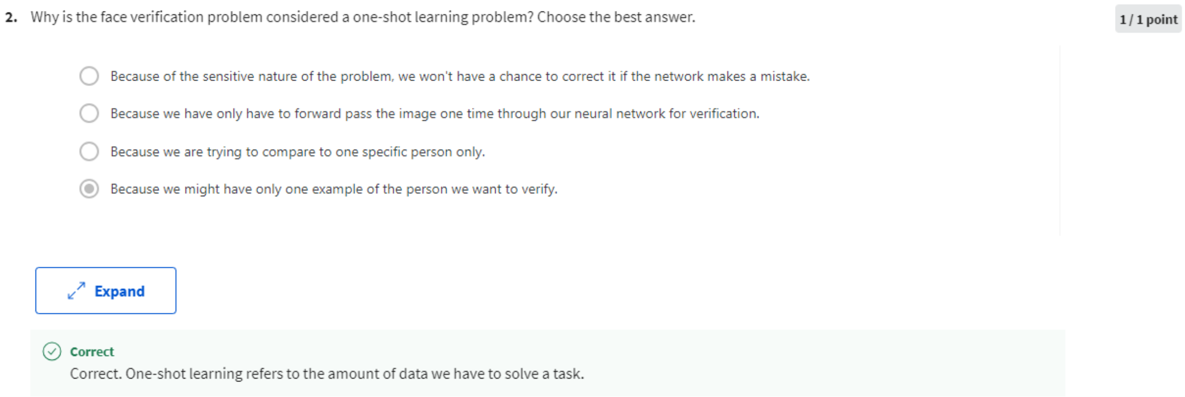
1. Which of the following do you agree with?

● Face recognition requires K comparisons of a person's face.

○ Face recognition requires comparing pictures against one person's face.

○ Face verification requires K comparisons of a person's face.

✓ **Correct**
Correct, in face recognition we compare the face of one person to K to classify the face as one of those K or not.



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

1 / 1 point

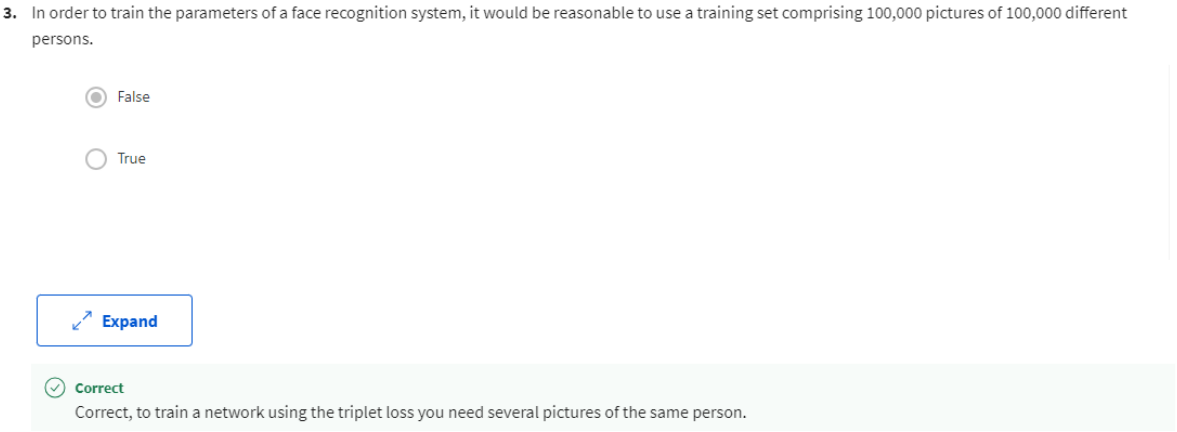
○ Because of the sensitive nature of the problem, we won't have a chance to correct it if the network makes a mistake.

○ Because we have only have to forward pass the image one time through our neural network for verification.

○ Because we are trying to compare to one specific person only.

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

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

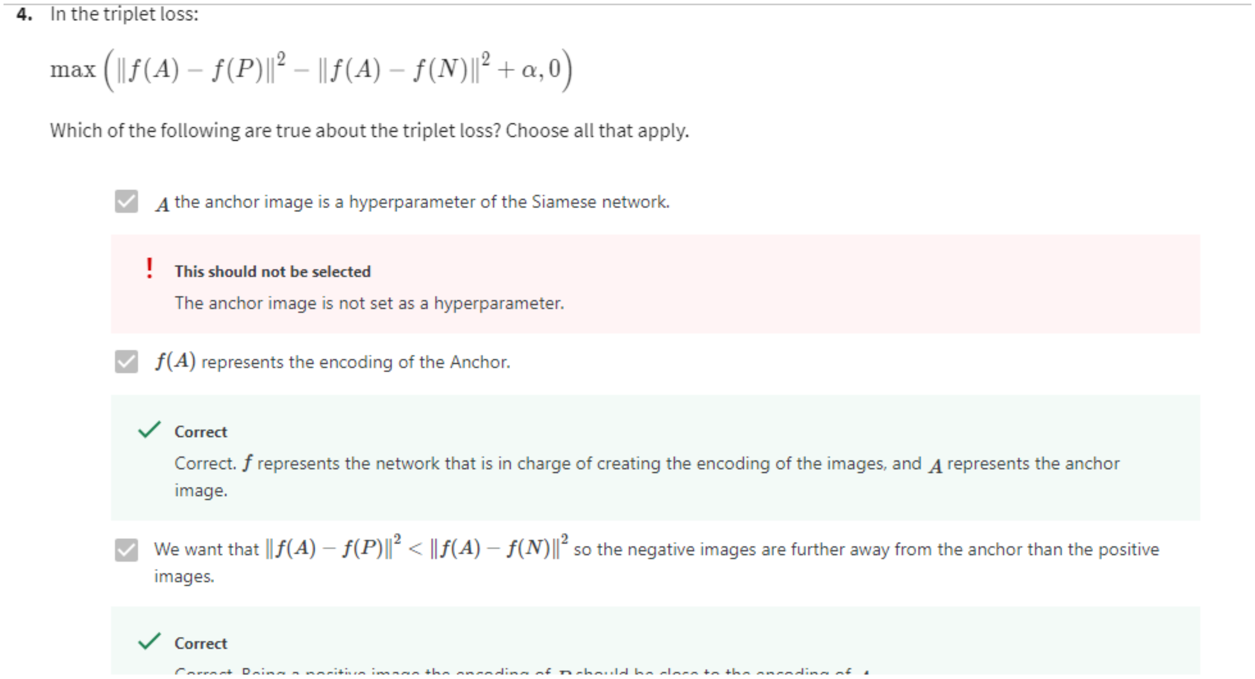


3. In order to train the parameters of a face recognition system, it would be reasonable to use a training set comprising 100,000 pictures of 100,000 different persons.

● False

○ True

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



4. In the triplet loss:

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

Which of the following are true about the triplet loss? Choose all that apply.

☑ A the anchor image is a hyperparameter of the Siamese network.

! **This should not be selected**

The anchor image is not set as a hyperparameter.

☑ $f(A)$ represents the encoding of the Anchor.

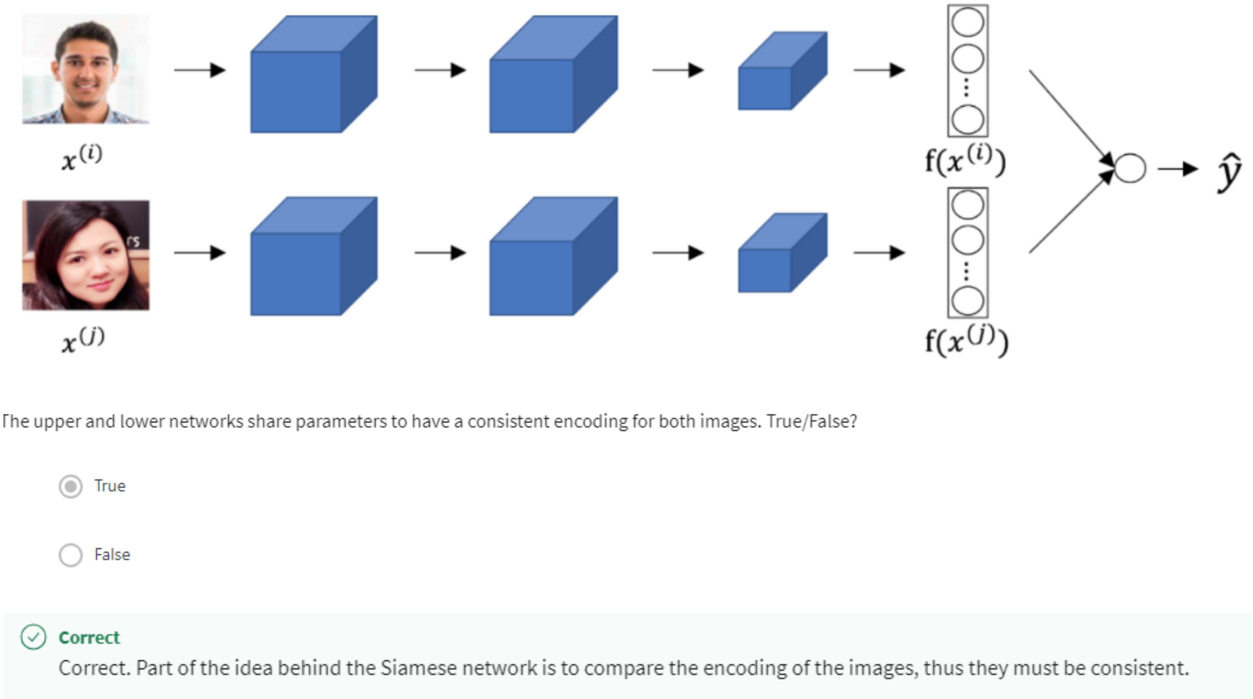
✓ **Correct**

Correct. f represents the network that is in charge of creating the encoding of the images, and A represents the anchor image.

☑ We want that $\|f(A)-f(P)\|^2<\|f(A)-f(N)\|^2$ so the negative images are further away from the anchor than the positive images.

✓ **Correct**

Correct. Being a positive image the encoding of P should be close to the encoding of A



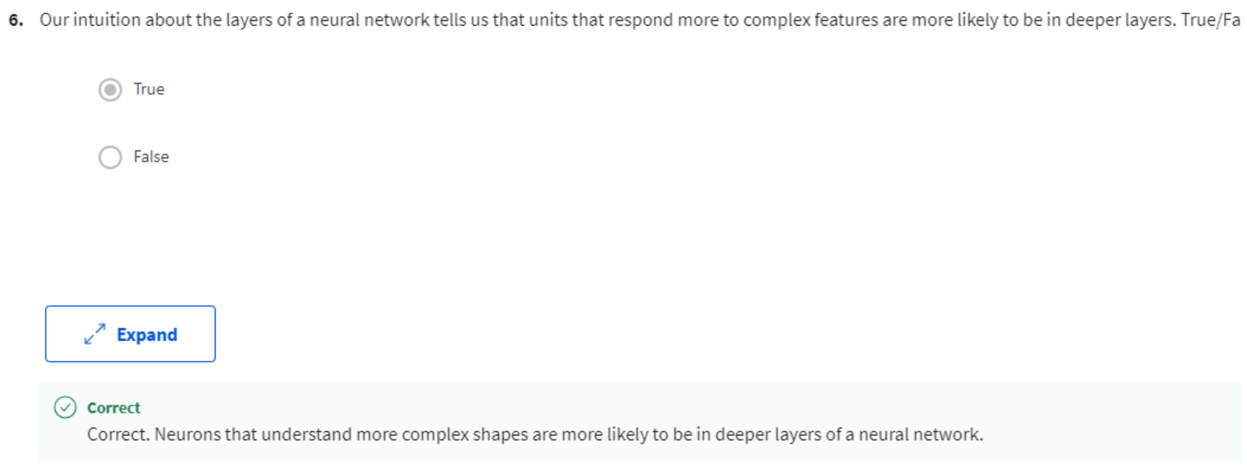
The upper and lower networks share parameters to have a consistent encoding for both images. True/False?

● True

○ False

✓ **Correct**

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



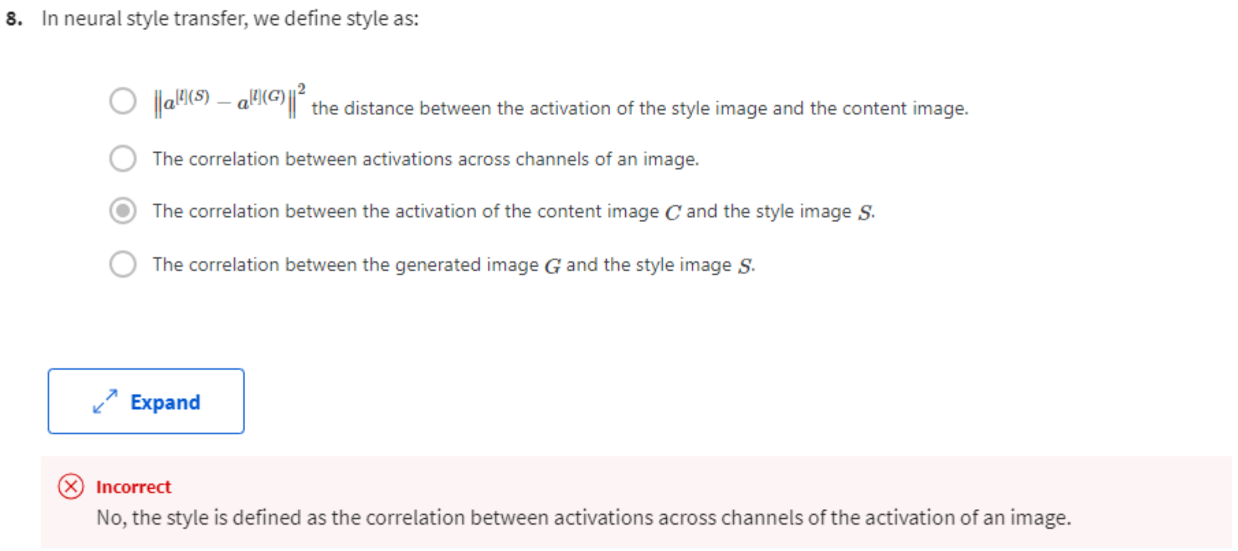
6. Our intuition about the layers of a neural network tells us that units that respond more to complex features are more likely to be in deeper layers. True/Fa

● True

○ False

Correct

Correct. Neurons that understand more complex shapes are more likely to be in deeper layers of a neural network.



8. In neural style transfer, we define style as:

○ $\left\|a^{[l](S)} - a^{[l](G)}\right\|^2$ the distance between the activation of the style image and the content image.

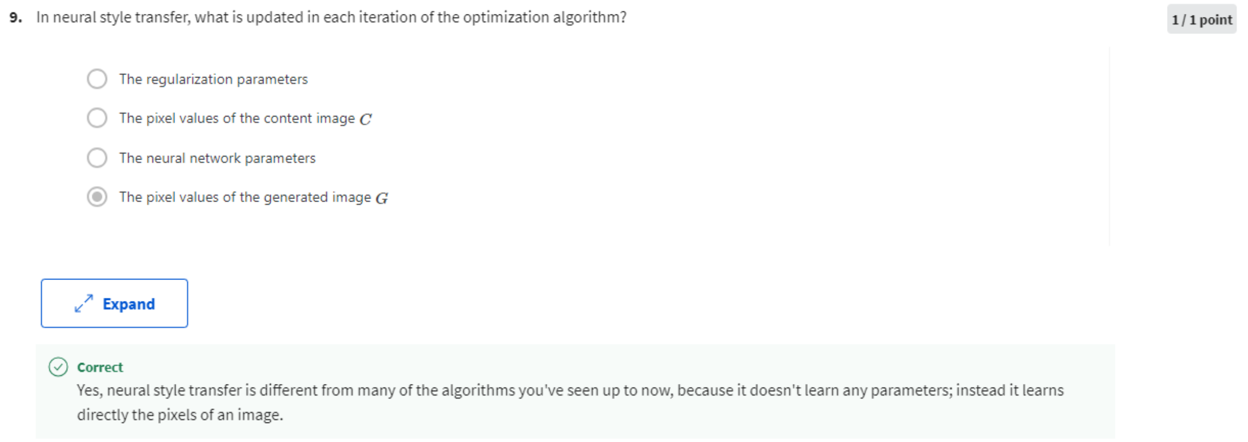
○ The correlation between activations across channels of an image.

● The correlation between the activation of the content image C and the style image S.

○ The correlation between the generated image G and the style image S.

Incorrect

No, the style is defined as the correlation between activations across channels of the activation of an image.



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

1 / 1 point

○ The regularization parameters

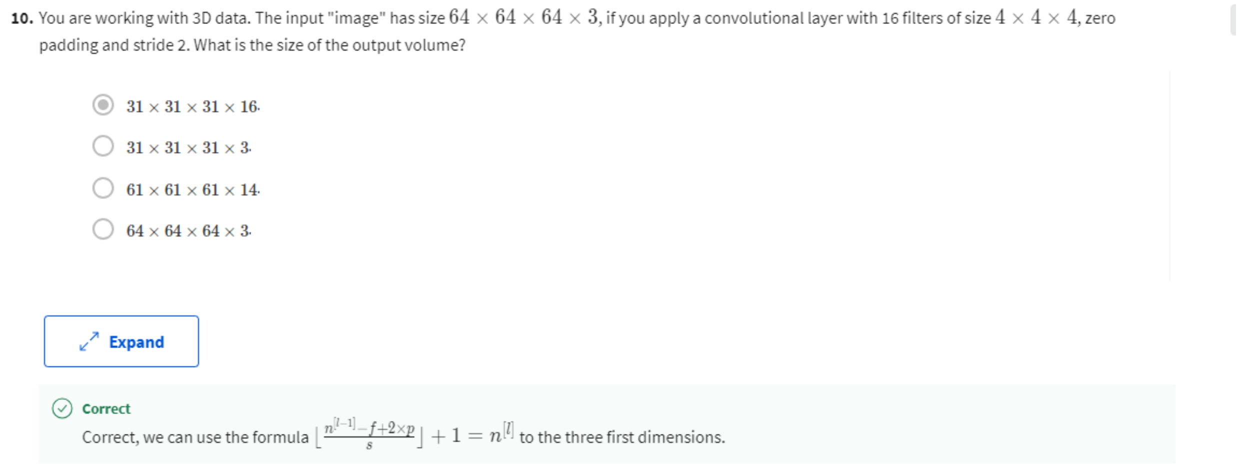
○ The pixel values of the content image C

○ The neural network parameters

● The pixel values of the generated image G

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. The input "image" has size $64 \times 64 \times 64 \times 3$, if you apply a convolutional layer with 16 filters of size $4 \times 4 \times 4$, zero padding and stride 2. What is the size of the output volume?

● $31 \times 31 \times 31 \times 16.$

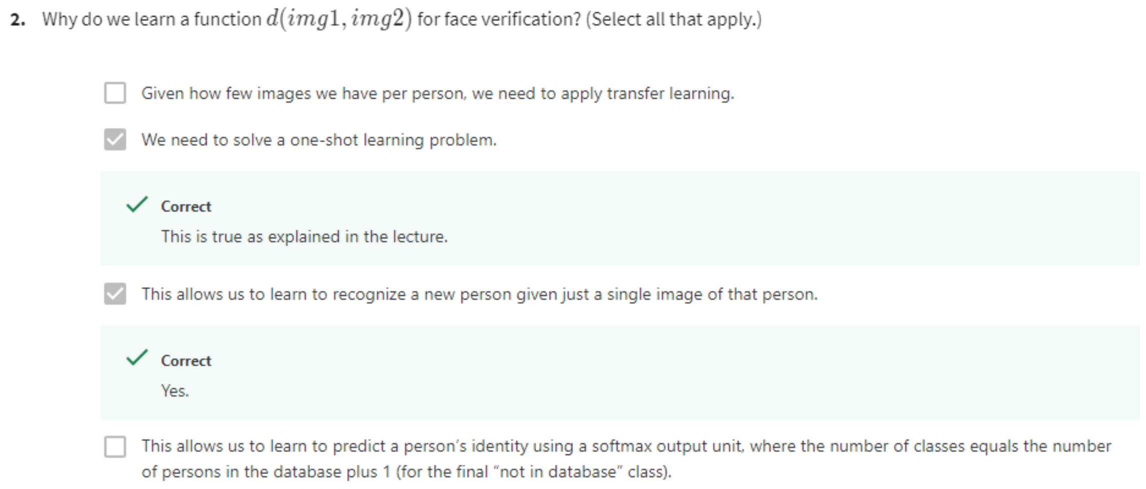
○ $31 \times 31 \times 31 \times 3.$

○ $61 \times 61 \times 61 \times 14.$

○ $64 \times 64 \times 64 \times 3.$

✓ **Correct**

Correct, we can use the formula $\lfloor \frac{n^{[l-1]} - f + 2 \times p}{s} \rfloor + 1 = n^{[l]}$ to the three first dimensions.



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

□ Given how few images we have per person, we need to apply transfer learning.

☑ We need to solve a one-shot learning problem.

✓ **Correct**

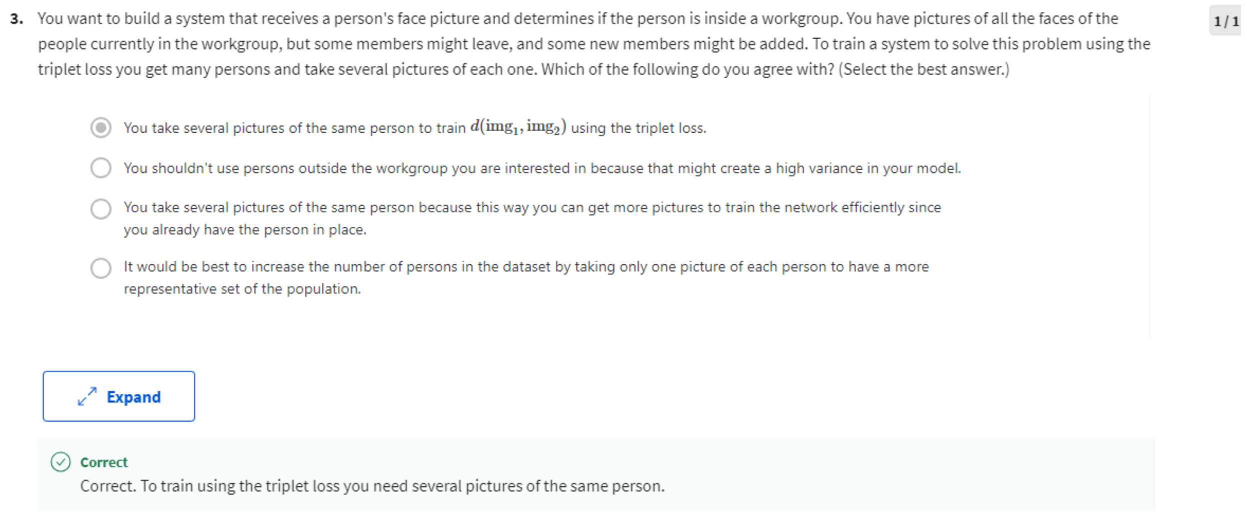
This is true as explained in the lecture.

☑ This allows us to learn to recognize a new person given just a single image of that person.

✓ **Correct**

Yes.

□ This allows us to learn to predict a person’s identity using a softmax output unit, where the number of classes equals the number of persons in the database plus 1 (for the final “not in database” class).



3. 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.)

1 / 1

● You take several pictures of the same person to train $d(\text{img}_1, \text{img}_2)$ using the triplet loss.

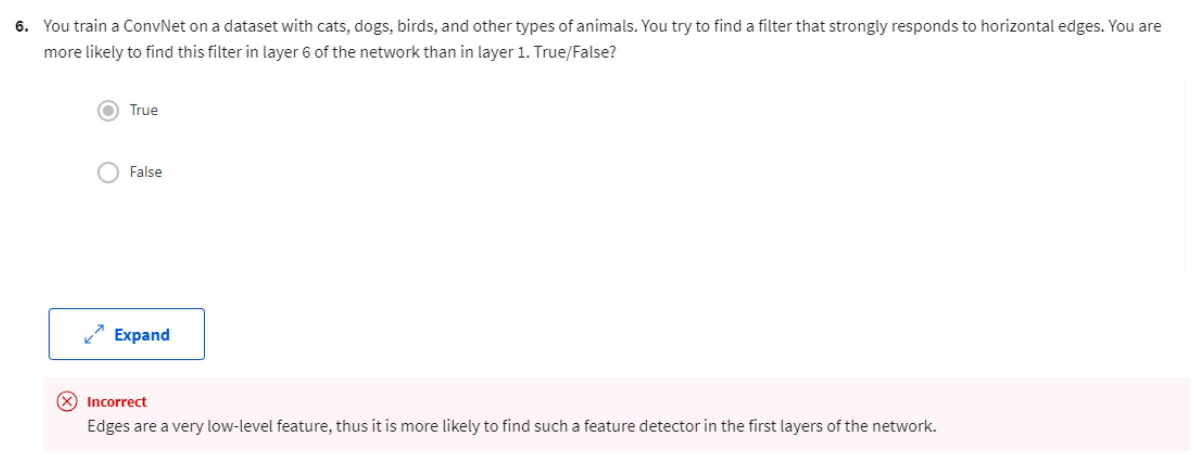
○ You shouldn't use persons outside the workgroup you are interested in because that might create a high variance in your model.

○ You take several pictures of the same person because this way you can get more pictures to train the network efficiently since you already have the person in place.

○ It would be best to increase the number of persons in the dataset by taking only one picture of each person to have a more representative set of the population.

✓ **Correct**

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



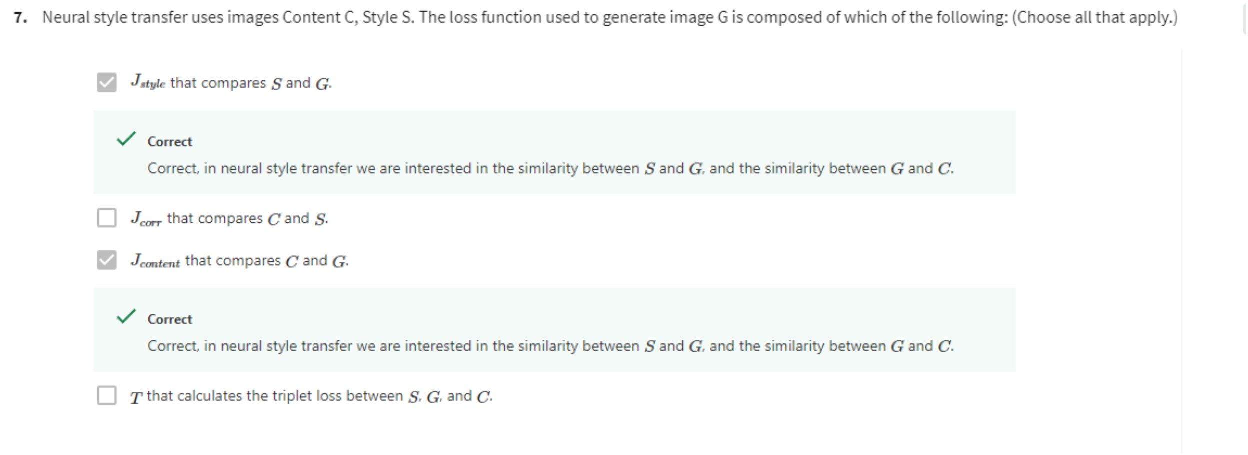
6. You train a ConvNet on a dataset with cats, dogs, birds, and other types of animals. You try to find a filter that strongly responds to horizontal edges. You are more likely to find this filter in layer 6 of the network than in layer 1. True/False?

● True

○ False

ⓧ **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.)

☑ 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.

□ J_{corr} that compares C and S.

☑ $J_{content}$ 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.

□ T that calculates the triplet loss between S, G, and C.

8. In neural style transfer the content loss J_{cont} is computed as:

$$J_{cont}(G, C) = \left\|a^{[l](C)} - a^{[l](G)}\right\|^2$$

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 abstrac activation of each image. True/False?

● True

○ False

Expand

Incorrect

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

10. You are working with 3D data. The input "image" has size $32 \times 32 \times 32 \times 3$, if you apply a convolutional layer with 16 filters of size $4 \times 4 \times 4$, zero padding and stride 1. What is the size of the output volume?

● $29 \times 29 \times 29 \times 16$.

○ $29 \times 29 \times 29 \times 3$.

○ $29 \times 29 \times 29 \times 13$.

○ $31 \times 31 \times 31 \times 16$.

Expand

Correct

Correct, we can use the formula $\lfloor \frac{n^{[l-1]} - f + 2 \times p}{s} \rfloor + 1 = n^{[l]}$ on the three first dimensions.