

When building a ConvNet, typically you start with some POOL layers followed by some CONV layers. True/False?

1 point

- ☐ False
- ☐ True

2. LeNet - 5 made extensive use of padding to create valid convolutions, to avoid increasing the number of channels after every convolutional layer. True/False?

1 point

- ☐ False
- ☐ True

3. Training a deeper network (for example, adding additional layers to the network) allows the network to fit more complex functions and thus almost always results in lower training error. For this question, assume we're referring to "plain" networks.

1 point

- ☐ True
- ☐ False

4. The computation of a ResNet block is expressed in the equation:

1 point

$$a^{[l+2]} = g \left( \underbrace{W^{[l+2]}}_C g \left( \underbrace{W^{[l+1]}}_A a^{[l]} + \underbrace{b^{[l+1]}}_B \right) + b^{[l+2]} + \underbrace{a^{[l]}}_B \right)$$

Which part corresponds to the skip connection?

- ☐ The term in the orange box, marked as  $B$ .
- ☐ The term in the red box, marked as  $C$ .
- ☐ The equation of ResNet.
- ☐ The term in the blue box, marked as  $A$ .

5. Which ones of the following statements on Residual Networks are true? (Check all that apply.)

1 point

- ☐ The skip-connections compute a complex non-linear function of the input to pass to a deeper layer in the network.
- ☐ The skip-connection makes it easy for the network to learn an identity mapping between the input and the output within the ResNet block.
- ☐ Using a skip-connection helps the gradient to backpropagate and thus helps you to train deeper networks
- ☐ A ResNet with  $L$  layers would have on the order of  $L^2$  skip connections in total.

6. Suppose you have an input volume of dimension  $n_H \times n_W \times n_C$ . Which of the following statements do you agree with? (Assume that the "1x1 convolutional layer" below always uses a stride of 1 and no padding.)

1 point

- ☐ You can use a 1x1 convolutional layer to reduce  $n_H$ ,  $n_W$ , and  $n_C$ .
- ☐ You can use a 1x1 convolutional layer to reduce  $n_C$  but not  $n_H$  and  $n_W$ .

☐ You can use a 2D pooling layer to reduce  $n_H$ ,  $n_W$ , but not  $n_C$ .

☐ You can use a 2D pooling layer to reduce  $n_H$ ,  $n_W$ , and  $n_C$ .

7. Which of the following are true about the inception Network? (Check all that apply)

1 point

☐ Inception blocks allow the use of a combination of 1x1, 3x3, 5x5 convolutions and pooling by stacking up all the activations resulting from each type of layer.

☐ Making an inception network deeper won't hurt the training set performance.

☐ One problem with simply stacking up several layers is the computational cost of it.

☐ Inception blocks allow the use of a combination of 1x1, 3x3, 5x5 convolutions, and pooling by applying one layer after the other.

8. Which of the following are common reasons for using open-source implementations of ConvNets (both the model and/or weights)? Check all that apply.

1 point

☐ It is a convenient way to get working with an implementation of a complex ConvNet architecture.

☐ Parameters trained for one computer vision task are often useful as pre-training for other computer vision tasks.

☐ A model trained for one computer vision task can usually be used to perform data augmentation for a different computer vision task.

☐ The same techniques for winning computer vision competitions, such as using multiple crops at test time, are widely used in practical deployments (or production system deployments) of ConvNets.

9. Which of the following are true about Depth wise-separable convolutions? (Choose all that apply)

1 point

☐ They have a lower computational cost than normal convolutions.

☐ The result has always the same number of channels  $n_c$  as the input.

☐ They are just a combination of a normal convolution and a bottleneck layer.

☐ They combine depthwise convolutions with pointwise convolutions.

10. Suppose that in a MobileNet v2 Bottleneck block we have an  $n \times n \times 5$  input volume, we use 30 filters for the expansion, in the depthwise convolutions we use  $3 \times 3$  filters, and 20 filters for the projection. How many parameters are used in the complete block, suppose we don't use bias?

1 point

☐ 1101

☐ 80

☐ 1020

☒ 8250