

Deep Convolutional Models

TOTAL POINTS 10

1. Which of the following do you typically see in a ConvNet? (Check all that apply.)

1 point

- Multiple POOL layers followed by a CONV layer
- Multiple CONV layers followed by a POOL layer
- FC layers in the first few layers
- FC layers in the last few layers

2. In order to be able to build very deep networks, we usually only use pooling layers to downsize the height/width of the activation volumes while convolutions are used with “valid” padding. Otherwise, we would downsize the input of the model too quickly.

1 point

- True
- False

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 following equation captures the computation in a ResNet block. What goes into the two blanks above?

1 point

$$a^{[l+2]} = g(W^{[l+2]}g(W^{[l+1]}a^{[l]} + b^{[l+1]}) + b^{l+2} + \underline{\hspace{2cm}}) + \underline{\hspace{2cm}}$$

- 0 and $a^{[l]}$, respectively
- $z^{[l]}$ and $a^{[l]}$, respectively
- $a^{[l]}$ and 0, respectively

0 and $z^{[l+1]}$, respectively

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

1 point

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

6. Suppose you have an input volume of dimension $n_H \times n_W \times n_C$. Which of the following statements you agree with? (Assume that “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 2D pooling layer to reduce n_H , n_W , and n_C .
- You can use a 2D pooling layer to reduce n_H , n_W , but not n_C .
- You can use a 1x1 convolutional layer to reduce n_C but not n_H , n_W .

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

1 point

- Inception networks incorporate a variety of network architectures (similar to dropout, which randomly chooses a network architecture on each step) and thus has a similar regularizing effect as dropout.
- Making an inception network deeper (by stacking more inception blocks together) *might* not hurt training set performance.

- Inception blocks usually use 1×1 convolutions to reduce the input data volume's size before applying 3×3 and 5×5 convolutions.
- A single inception block allows the network to use a combination of 1×1 , 3×3 , 5×5 convolutions and pooling.

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

- A model trained for one computer vision task can usually be used to perform data augmentation even 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.
- Parameters trained for one computer vision task are often useful as pretraining for other computer vision tasks.
- It is a convenient way to get working with an implementation of a complex ConvNet architecture.

9. In Depthwise Separable Convolution you:

1 point

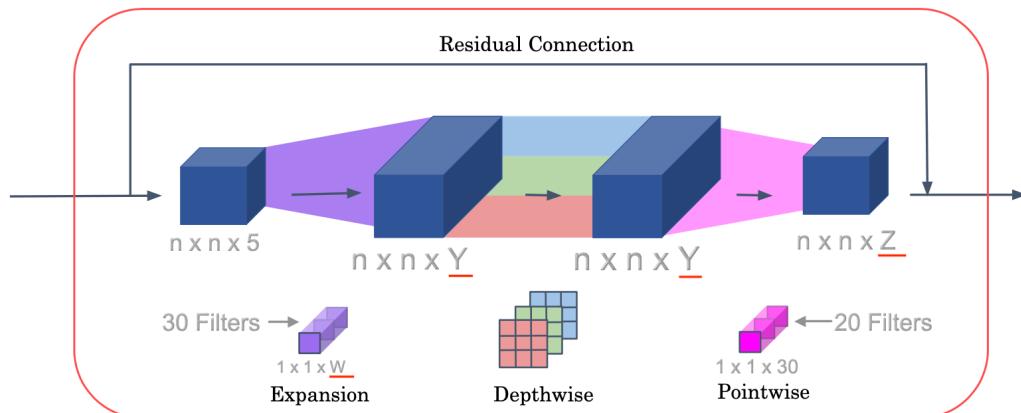
- For the "Depthwise" computations each filter convolves with only one corresponding color channel of the input image.
- For the "Depthwise" computations each filter convolves with all of the color channels of the input image.
- You convolve the input image with a filter of $n_f \times n_f \times n_c$ where n_c acts as the depth of the filter (n_c is the number of color channels of the input image).
- Perform one step of convolution.
- The final output is of the dimension $n_{out} \times n_{out} \times n'_c$ (where n'_c is the number of filters used in the previous convolution step).
- The final output is of the dimension $n_{out} \times n_{out} \times n_c$ (where n_c is the number of color channels of the input image).

- You convolve the input image with n_c number of $n_f \times n_f$ filters (n_c is the number of color channels of the input image).
- Perform two steps of convolution.

10. Fill in the missing dimensions shown in the image below (marked W, Y, Z).

1 point

MobileNet v2 Bottleneck



- W = 30, Y = 20, Z = 20
- W = 30, Y = 30, Z = 5
- W = 5, Y = 20, Z = 5
- W = 5, Y = 30, Z = 20