Convolutional Neural Networks

Week 2: Deep Convolutional Models

1 Version 1: When building a ConvNet, typically you start with some POOL layers followed by some CONV layers. True/False?

Answer: False.

Comment: It is typical for ConvNets to use a POOL layer after some CONV layers; sometimes even one POOL layer after each CONV layer; but is not common to start with POOL layers.

1 Version 2: Which of the following do you typically see as you move to deeper layers in a ConvNet?

Answer: n_H and n_W decrease, while n_C increases.

1 Version 3: Which of the following do you typically see in ConvNet? (Check all that apply).

Answer: Use of FC layer after flattening the volume to output classes.

2 Version 1: In LeNet - 5 we can see that as we get into deeper networks the number of channels increases while the height and width of the volume decreases. True/False?

Answer : True.

Comment: Since in its implementation only valid convolutions were used, without padding, the height and width of the volume were reduced at each convolution. These were also reduced by the POOL layers, whereas the number of channels was increased from 6 to 16.

2 Version 2: Which of the following do you typically see in a ConvNet? (Check all that apply.)

Answer: Multiple CONV layers followed by a POOL layer; FC layers in the last few layers.

2 Version 3: In order to be bale 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.

Answer: False.

3 Version 1: Based on the lectures, in the following picture (blue curve on higher than green curve), which curve corresponds to the expected behavior in theory, and which one corresponds to the behavior we get in practice? This when using plain neural networks.

Answer: The green one depicts the results in theory, and the blue on the reality.

Comment: In theory, we expect that as we increase the number of layers the training error decreases; but in practice after a certain number of layers the error increases.

3 Version 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.

Answer: False.

3 Version 3: The motivation of Residual Networks is that very deep networks are so good at fitting conmplex functions that when training them we almost always overfit the training data. True/False?

Answer: False.

4 Version 1: The following equation ($a^{[l+2]}$)captures the computation in a ResNet block. What goes into the two blanks above?

Answer: $a^{[l]}$ and 0, respectively.

4 Version 2: 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. True/False. For this question, assume we're referring to "plain" networks.

Answer: False.

4 Version 3: The computation of a ResNet block is expressed in the equation: which part corresponds to the skip connection?

Answer: The term in the orange box, marked as B ($a^{[l]}$).

5 In the best scenario when adding a ResNet block it will learn to approximate the identity function after a lot of training, helping improve the overall performance of the network. True/False?

Answer: False.

Comment: When adding a ResNet block it can easily learn to approximate the identity function, thus in a worst-case scenario, it will not affect the performance of the network at all.

6 Version 1: 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 1×1 convolutional layer below always uses a stride of 1 and no padding.)

Answer: You can use a 2D polling layer to reduce n_H, n_W but not n_C; You can use a 1×1 convolutional layer to reduce n_C but not n_H and n_W.

Comment: A 1×1 convolutional layer with a small number of filter is going to reduce n_C but will keep the dimensions n_H and n_W.

6 Version 2: Which ones of the following statements on Residual Networks are true? (Check all that apply.)

Answer: 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 Version 3: 1 by 1 convolutions are the same as multiplying by a single number. True/False.

Answer: False.

7 Version 1: Which of the following are true about bottleneck layers? (Check all that apply)

Answer: The use of bottlenecks does not seem to hurt the performance of the network; By adding these layers we can reduce the computational cost in the inception modules.

Comment: Although it reduces the computational cost significantly; By using the 1×1 convolutional layers we can reduce the depth of the volume and help reduce the computational cost of applying other convolutional layers with different filter sizes.

7 Version 2: Suppose you have an input volume of dimension 64 by 64 by 16. How many parameters would a single 1 by 1 convolutional filter have (including the bias)?

Answer: 17.

8 Parameters trained for one computer vision task cannot be used directly in another task. In most cases, we must change the softmax layer, or the last layers of the model and re-train for the new task. True/False?

Answer: True.

Comment: This is a good way to take advantage of open-source models trained more or less for the task you want to do. This may also help you save a great number of computational resources and data.

9 Version 1: Which of the following are true about Depthwise-separable convolutions? (Choose all that apply)

Answer: Depthwise separable convolutions are composed of two different types of convolutions; The depthwise convolution convolves each channel in the input volume with a separate filter; The pointwise convolution convolves the output volume with 1×1 filters.

Comment: It is composed of a depthwise convolution followed by a pointwise convolution; The output of this kind of convolution is the same as the input; The number of filters for the output of the depthwise separable convolution is determined by the number of 1×1 filters used.

9 Version 2: Which ones of the following statements on inception Networks are true? (Check all that apply).

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

9 Version 3: Which of the following are common reasons for using open source implementations of ConvNets (both the model and/or weights)? (Check all that apply)

Answer: Parameters trained for one computer vision task are often useful as pre training for other computer vision tasks; It is a convenient way to get working with an implementation of a complex ConvMet architecture.

10 Version 1: Suppose that in a MobileNet v2 Bottleneck block we have an n by n by 5 input volume, we use 30 filters for the expansion, in the depthwise convolutions we use 3 by 3 filters, and 20 filters for the projection. How many parameters are used in the complete block, suppose we do not use bias?

Answer: 1020.

Comment: The expansion filters use 5×30 parameters, the depthwise convolutions need $3 \times 3 \times 30$ parameters, and the projection part 30×20 parameters.

10 Version 2: Which of the following are common reasons for using open-source implementations of ConvNets (both the model and/or weights)? Check all that apply.

Answer: It is a convenient way to get working an implementation of a complex ConvNet architecture; Parameters trained for one computer vision task are often useful as pretraining for other computer vision tasks.