

Your grade: **100%**

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

Next item →

1. What do you think applying this filter to a grayscale image will do?

1 / 1 point

$$\begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$$

- ☐ Detect vertical edges.
- ☒ Detect 45-degree edges.
- ☐ Detecting image contrast.
- ☐ Detect horizontal edges.

✓ Correct

Correct. Notice that there is a high delta between the values in the top left part and the ones in the bottom right part. When convolving this filter on a grayscale image, the edges forming a 45-degree angle with the horizontal will be detected.

2. Suppose your input is a 300 by 300 color (RGB) image, and you are not using a convolutional network. If the first hidden layer has 100 neurons, each one fully connected to the input, how many parameters does this hidden layer have (including the bias parameters)?

1 / 1 point

- ☐ 27,000,001
- ☒ 27,000,100
- ☐ 9,000,100
- ☐ 9,000,001

✓ Correct

Correct, the number of weights is $300 \times 300 \times 3 \times 100 = 27,000,000$, when you add the bias terms (one per neuron) you get 27,000,100.

3. Suppose your input is a 300 by 300 color (RGB) image, and you use a convolutional layer with 100 filters that are each 5x5. How many parameters does this hidden layer have (including the bias parameters)?

1 / 1 point

- ☐ 2501
- ☐ 7500
- ☐ 2600
- ☒ 7600

✓ Correct

Correct, you have $25 \times 3 = 75$ weights and 1 bias per filter. Given that you have 100 filters, you get 7,600 parameters for this layer.

4. You have an input volume that is 63x63x16, and convolve it with 32 filters that are each 7x7, using a stride of 2 and no padding. What is the output volume?

1 / 1 point

- ☐ 29x29x16
- ☒ 29x29x32
- ☐ 16x16x32
- ☐ 16x16x16

✓ Correct

Yes, $\frac{63-7+0 \times 2}{2} + 1 = 29$ and the number of channels should match the number of filters.

5. You have an input volume that is 31x31x32, and pad it using "pad=1". What is the dimension of the resulting volume (after padding)?

1 / 1 point

- ☐ 31x31x34
- ☐ 33x33x33
- ☒ 33x33x32

32x32x32

☐ 32x32x32

☒ Correct

Yes, if the padding is 1 you add 2 to the height dimension and 2 to the width dimension.

6. You have a volume that is $64 \times 64 \times 32$, and convolve it with 40 filters of 9×9 , and stride 1. You want to use a "same" convolution. What is the padding?

1 / 1 point

☒ 4

☐ 0

☐ 6

☐ 8

☒ Correct

Yes, when using a padding of 4 the output volume has $n_H = \frac{64-9+2 \times 4}{1} + 1$.

7. You have an input volume that is $66 \times 66 \times 21$, and apply max pooling with a stride of 3 and a filter size of 3. What is the output volume?

1 / 1 point

☐ $22 \times 22 \times 7$

☐ $21 \times 21 \times 21$

☒ $22 \times 22 \times 21$

☐ $66 \times 66 \times 7$

☒ Correct

Yes, using the formula $n_H^{[l]} = \frac{n_H^{[l-1]} + 2 \times p - f}{s} + 1$ with $p = 0$, $f = 3$, $s = 3$ and $n_H^{[l-1]} = 66$.

8. Because pooling layers do not have parameters, they do not affect the backpropagation (derivatives) calculation.

1 / 1 point

☐ True

☒ False

☒ Correct

Everything that influences the loss should appear in the backpropagation because we are computing derivatives. In fact, pooling layers modify the input by choosing one value out of several values in their input volume. Also, to compute derivatives for the layers that have parameters (Convolutions, Fully-Connected), we still need to backpropagate the gradient through the Pooling layers.

9. In lecture we talked about "parameter sharing" as a benefit of using convolutional networks. Which of the following statements about parameter sharing in ConvNets are true? (Check all that apply)

1 / 1 point

☐ It allows gradient descent to set many of the parameters to zero, thus making the connections sparse.

☐ It allows parameters learned for one task to be shared even for a different task (transfer learning).

☒ It reduces the total number of parameters, thus reducing overfitting.

☒ Correct

Yes, a convolutional layer uses parameter sharing and usually has a lot less parameters than a fully-connected layer.

☒ It allows a feature detector to be used in multiple locations throughout the whole input image/input volume.

☒ Correct

Yes, by sliding a filter of parameters over the entire input volume, we make sure a feature detector can be used in multiple locations.

10. In lecture we talked about "sparsity of connections" as a benefit of using convolutional layers. What does this mean?

1 / 1 point

☐ Each filter is connected to every channel in the previous layer.

☒ Each activation in the next layer depends on only a small number of activations from the previous layer.

☐ Regularization causes gradient descent to set many of the parameters to zero.

☐ Each layer in a convolutional network is connected only to two other layers

☒ Correct

Correct

Yes, each activation of the output volume is computed by multiplying the parameters from **only one filter** with a volumic slice of the input volume and then summing all these together.