**1.** What do you think applying this filter to a grayscale image will do? **0 / 1 point**

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

- Detect horizontal edges.
- Detect 45-degree edges.
- Detecting image contrast.
- (•) Detect vertical edges.

Expand

**Incorrect**
Incorrect. Notice that there is a different direction in which we can notice a high delta in the values.

**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**

- 9,000,001
- 27,000,001
- 9,000,100
- (•) 27,000,100

Expand

**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**

- (•) 7600

- 2501
- 2600
- 7500

Expand

**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 $127 \times 127 \times 16$, and convolve it with 32 filters of $5 \times 5$, using a stride of 2 and no padding. What is the output volume? **1 / 1 point**

- $123 \times 123 \times 16$
- $123 \times 123 \times 32$
- $62 \times 62 \times 16$
- $$62 \times 62 \times 32$$

Loading [MathJax]/jax/output/CommonHTML/jax.js

Expand

**Correct**
Correct, using the formula $$n_H^{[l]} = \frac{n_H^{[l-1]} + 2 \times p - f }{s} + 1$$ with $$n_H^{[l-1]} = 127$$, $$p = 0$$, $$f = 5$$, and $$s = 2$$ we get 62.

**5.** You have an input volume that is 15x15x8, and pad it using “pad=2”. What is the dimension of the resulting volume (after padding)? **1 / 1 point**

- 17x17x10
- 19x19x8
- 17x17x8
- 19x19x12

Expand

**Correct**
Correct, padding is applied over the height and the width of the input image. If the padding is two, you add 4 to the height dimension and 4 to the width dimension.

**6.** You have a volume that is $121 \times 121 \times 32$, and convolve it with 32 filters of $5 \times 5$, and a stride of 1. You want to use a "same" convolution. What is the padding? **0 / 1 point**

- 3
- 0
- 2
- 5 (selected)

Expand

**Incorrect**
No, remember that when using padding of 5 then 10 is added to each dimension.

**7.** You have an input volume that is 128x128x12, and apply max pooling with a stride of 4 and a filter size of 4. What is the output volume? **1 / 1 point**

- $128 \times 128 \times 3$
- $64 \times 64 \times 12$
- $$32 \times 32 \times 3$$
- $$32 \times 32 \times 12$$ (selected)

Expand

**Correct**
Yes, using the formula $$n_H^{[l]} = \frac{n_H^{[l-1]} + 2 \times p - f }{s} + 1$$ with $$p = 0$$, $$f = 4$$, $$s = 4$$ and $$n_H^{[l-1]} = 32$$.

**8.** Because pooling layers do not have parameters, they do not affect the backpropagation (derivatives) calculation. **1 / 1 point**

- True
- False (selected)

Expand

**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**

- [x] 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.

- [x] 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.

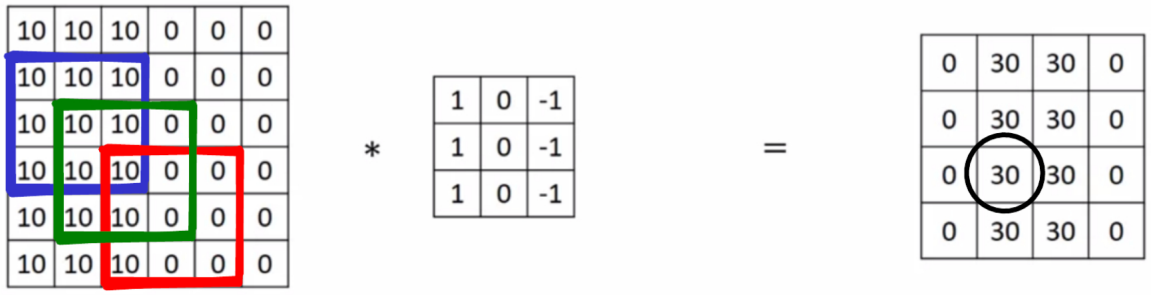
- [ ] It allows gradient descent to set many of the parameters to zero, thus making the connections sparse.
- [ ] It allows parameters learned for one task to be shared even for a different task (transfer learning).

Expand

✓ **Correct**

Great, you got all the right answers.

**10.** The following image depicts the result of a convolution at the right when using a stride of 1 and the filter is shown right next. **1 / 1 point**




On which pixels does the circled pixel of the activation at the right depend?

- ( ) It depends on the pixels enclosed by the blue square.
- (•) It depends on the pixels enclosed by the green square.