

Your grade: 90%

[Next item →](#)

Your latest: 90% • Your highest: 90% • To pass you need at least 80%. We keep your highest score.

1. Which of the following do you typically see in ConvNet?

1 / 1 point

- ☐ Multiple FC layers followed by a CONV layer.
- ☒ Use of FC layers after flattening the volume to generate output classes.
- ☐ ConvNet makes exclusive use of CONV layers.
- ☐ Use of multiple POOL layers followed by a CONV layer.

✓ **Correct**

Yes, FC layers are typically used in the last few layers after flattening the volume to generate the output in classification.

2. 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?

1 / 1 point

- ☒ True
- ☐ False

✓ **Correct**

Correct, 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.

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

1 / 1 point

☒ False

☐ True

☒ **Correct**

Correct, very deep neural networks are hard to train and a deeper network does not always imply lower training error. Residual Networks allow us to train very deep neural networks.

4. Which of the following equations captures the computations in a ResNet block?

1 / 1 point

☐ $a^{[l+2]} = g(W^{[l+2]} g(W^{[l+1]} a^{[l]} + b^{[l+1]}) + b^{[l+2]})$

☐ $a^{[l+2]} = g(W^{[l+2]} g(W^{[l+1]} a^{[l]} + b^{[l+1]}) + b^{[l+2]} + a^{[l]}) + a^{[l+1]}$

☒ $a^{[l+2]} = g(W^{[l+2]} g(W^{[l+1]} a^{[l]} + b^{[l+1]}) + b^{[l+2]} + a^{[l]})$

☐ $a^{[l+2]} = g(W^{[l+2]} g(W^{[l+1]} a^{[l]} + b^{[l+1]}) + b^{[l+2]}) + a^{[l]}$

☒ **Correct**

Correct. This expresses the computations of a ResNet block, where the last term $a^{[l]}$ is the shortcut connection.

5. Adding a ResNet block to the end of a network makes it deeper. Which of the following is true?

1 / 1 point

☐ It shifts the behavior of the network to be more like the identity function.

☐ The number of parameters will decrease due to the shortcut connections.

☒ The performance of the networks doesn't get hurt since the ResNet block can easily approximate the

- ☐ The performance of the networks is hurt since we make the network harder to train.

✓ **Correct**

Yes, as noted in the lectures in a ResNet block the computations are given by $a^{[l+2]} = g(W^{[l+2]}a^{[l+1]} + b^{[l+2]} + a^{[l]})$ thus if $W^{[l+2]}$ and $b^{[l+2]}$ are zero then we get the identity function.

6. 1×1 convolutions are the same as multiplying by a single number. True/False?

1 / 1 point

☐ True

☒ False

✓ **Correct**

Yes, a 1×1 layer doesn't act as a single number because it makes a sum over the depth of the volume.

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

1 / 1 point

☒ The use of bottlenecks doesn't seem to hurt the performance of the network.

✓ **Correct**

Yes, although it reduces the computational cost significantly.

☐ Bottleneck layers help to compress the 1×1 , 3×3 , 5×5 convolutional layers in the inception network.

☒ By adding these layers we can reduce the computational cost in the inception modules.

✓ **Correct**

Yes, 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.

☐ The bottleneck layer has a more powerful regularization effect than Dropout layers.

8. When having a small training set to construct a classification model, which of the following is a strategy of transfer learning that you would use to build the model?

1 / 1 point

- ☐ Use an open-source network trained in a larger dataset, freeze the softmax layer, and re-train the rest of the layers.
- ☒ Use an open-source network trained in a larger dataset freezing the layers and re-train the softmax layer.
- ☐ It is always better to train a network from a random initialization to prevent bias in our model.
- ☐ Use an open-source network trained in a larger dataset. Use these weights as an initial point for the training of the whole network.

✓ **Correct**

Yes, this is a strategy that can provide a good result with small data.

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

1 point

- ☐ The pointwise convolution convolves the output volume with 1×1 filters.
- ☒ Depthwise-separable convolutions are composed of two different types of convolutions.

✓ **Correct**

Yes, it is composed of a depthwise convolution followed by a pointwise convolution.

- ☒ The depthwise convolution convolves each channel in the input volume with a separate filter.

✓ **Correct**

Yes, the output of this kind of convolution is the same as the input.

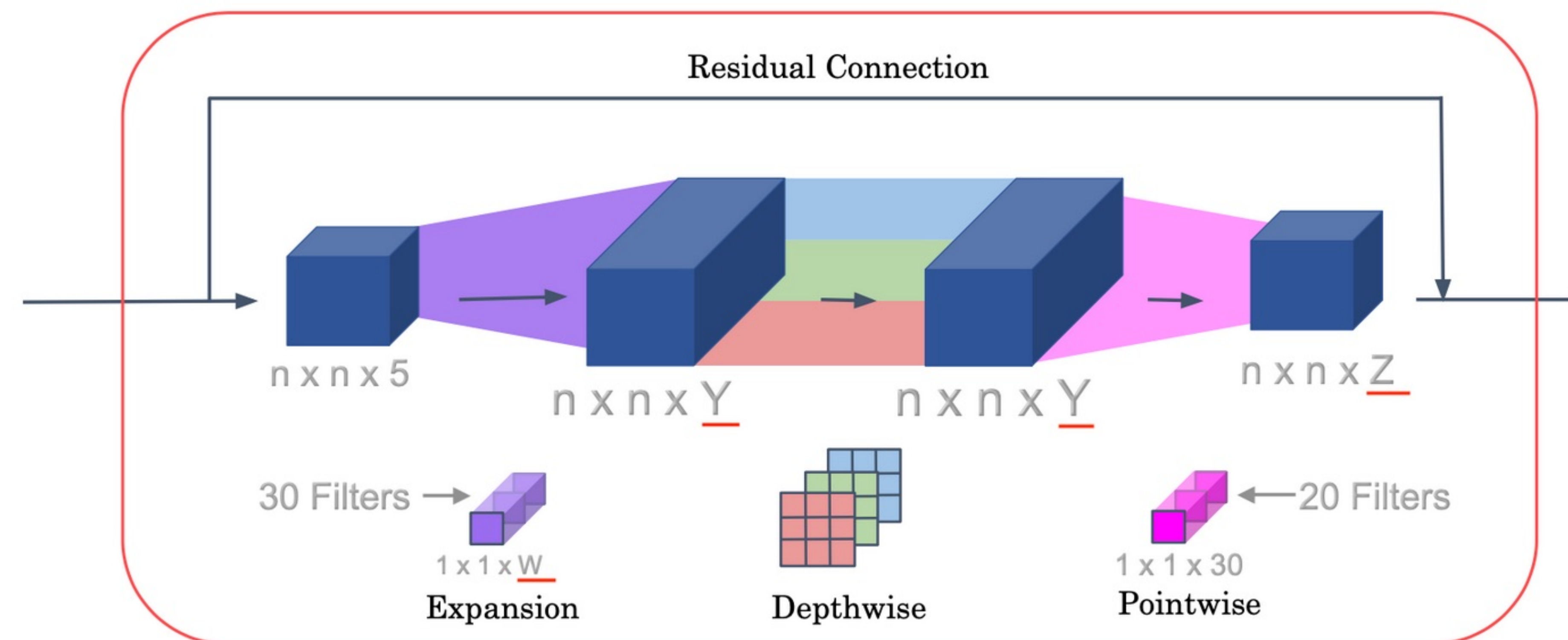
☐ The depthwise convolution convolves the input volume with 1×1 filters over the depth dimension.

You didn't select all the correct answers

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

1 / 1 point

MobileNet v2 Bottleneck

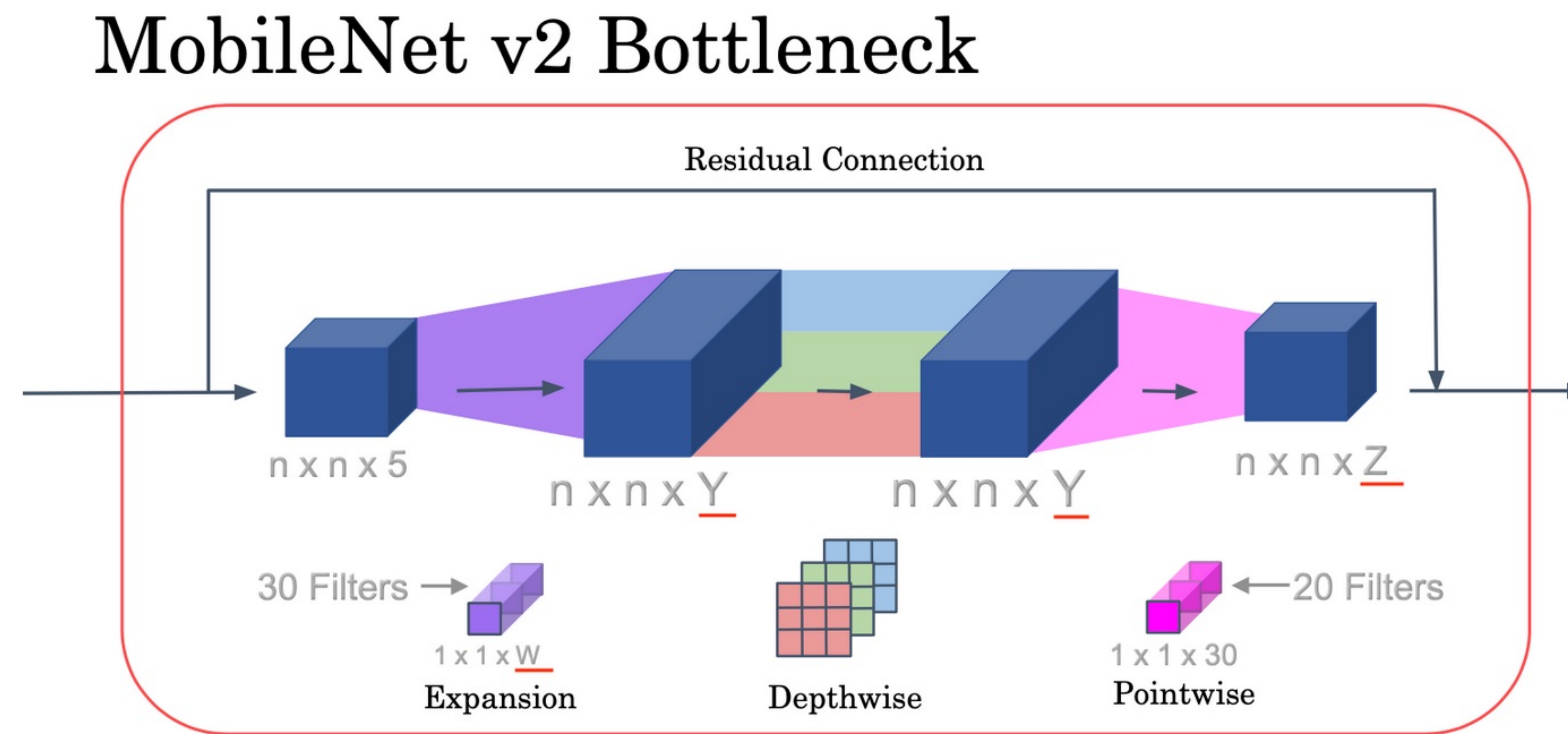


☒ W = 5, Y = 30, Z = 20

☐ W = 5, Y = 20, Z = 5

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

1 / 1 point



- ☒ $W = 5, Y = 30, Z = 20$
- ☐ $W = 5, Y = 20, Z = 5$
- ☐ $W = 30, Y = 20, Z = 20$
- ☐ $W = 30, Y = 30, Z = 5$

✓ **Correct**