Foundations of Machine Learning (CS 725)

FALL 2024

Lecture 16:

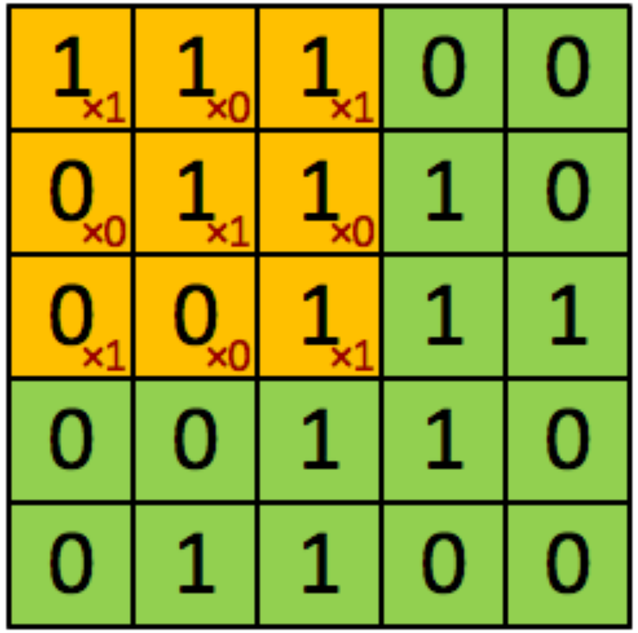
- Convolutional Neural Networks

Instructor: Preethi Jyothi

Convolutional Neural Networks (CNNs)

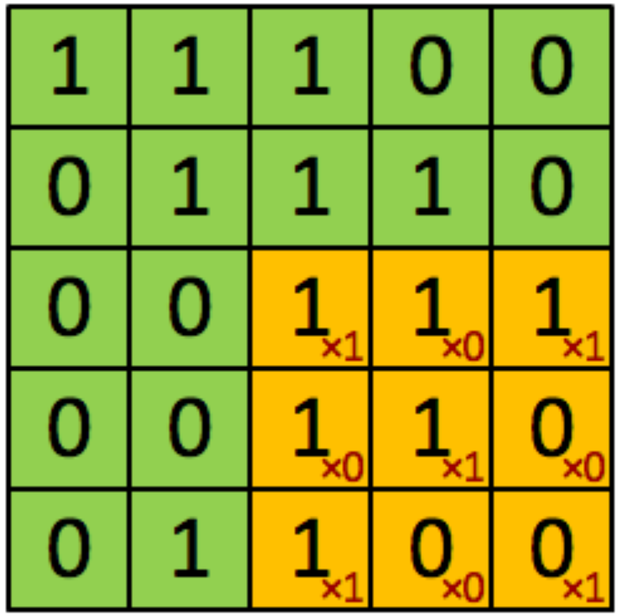
- Fully connected (dense) layers have no awareness of spatial information
- Key concept behind convolutional layers is that of ***kernels*** or ***filters***
- Filters slide across an input space to detect spatial patterns (translation invariance) in local regions (locality)

Convolution Layer



Animation from:https://hackernoon.com/visualizing-parts-of-convolutional-neural-networks-using-keras-and-cats-5cc01b214e59

Convolution Layer



Image

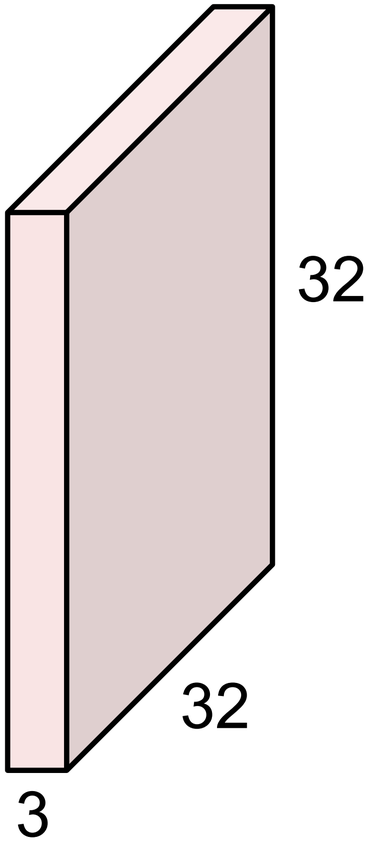
4	3	4
2	4	3
2	3	4

Convolved Feature

Animation from:https://hackernoon.com/visualizing-parts-of-convolutional-neural-networks-using-keras-and-cats-5cc01b214e59

Convolution Layer

32x32x3 image



5x5x3 filter

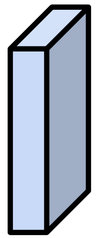
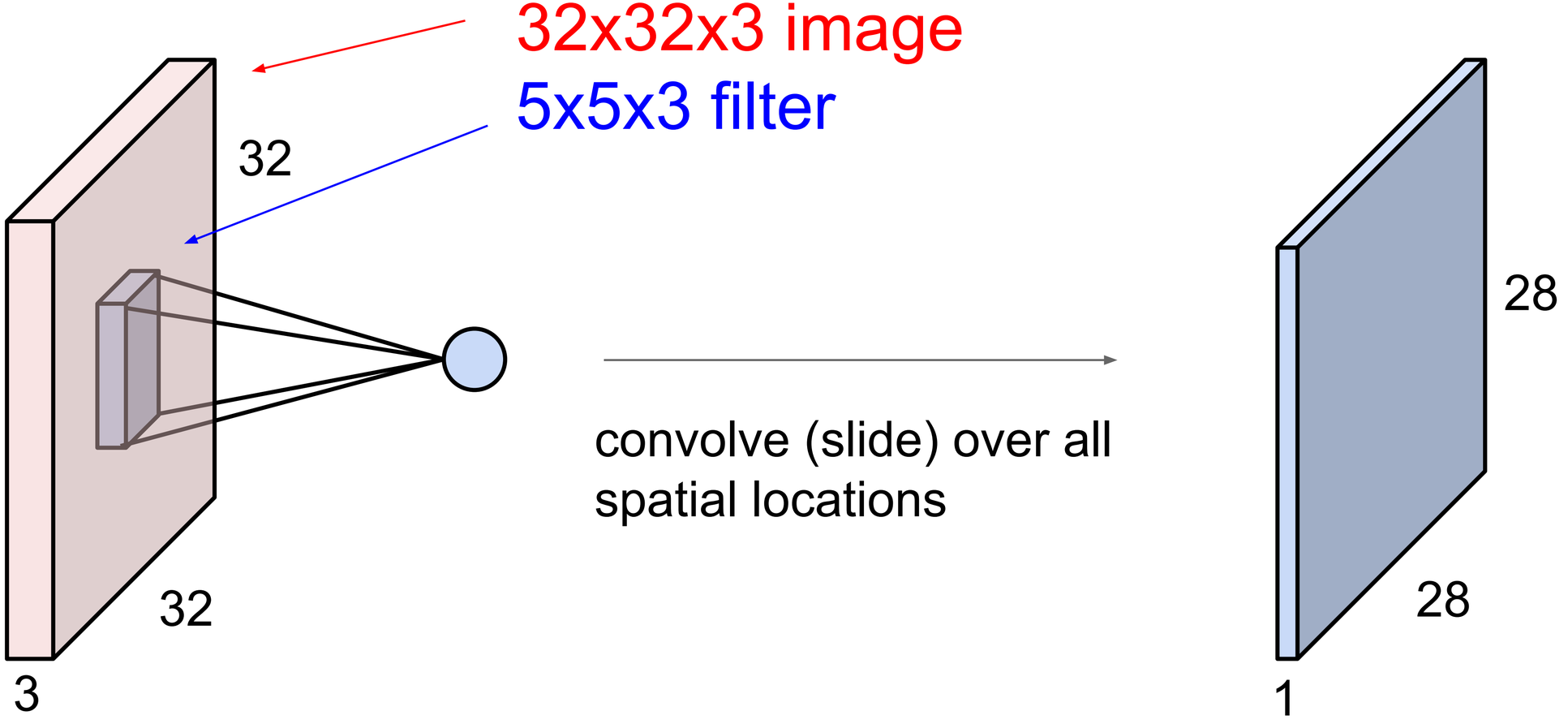


Image from:http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture05.pdf

Convolution Layer



Convolution Layer

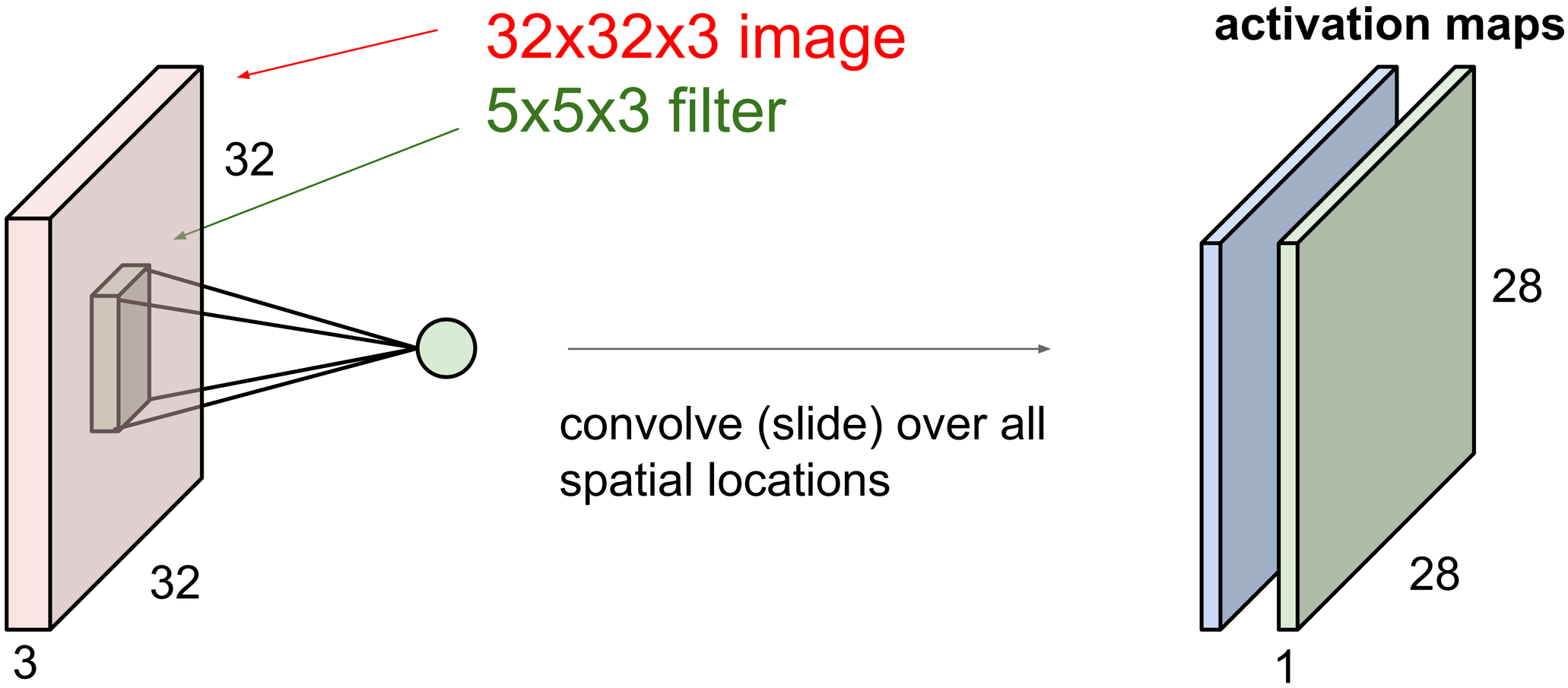


Image from:http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture05.pdf

Convolution Layer

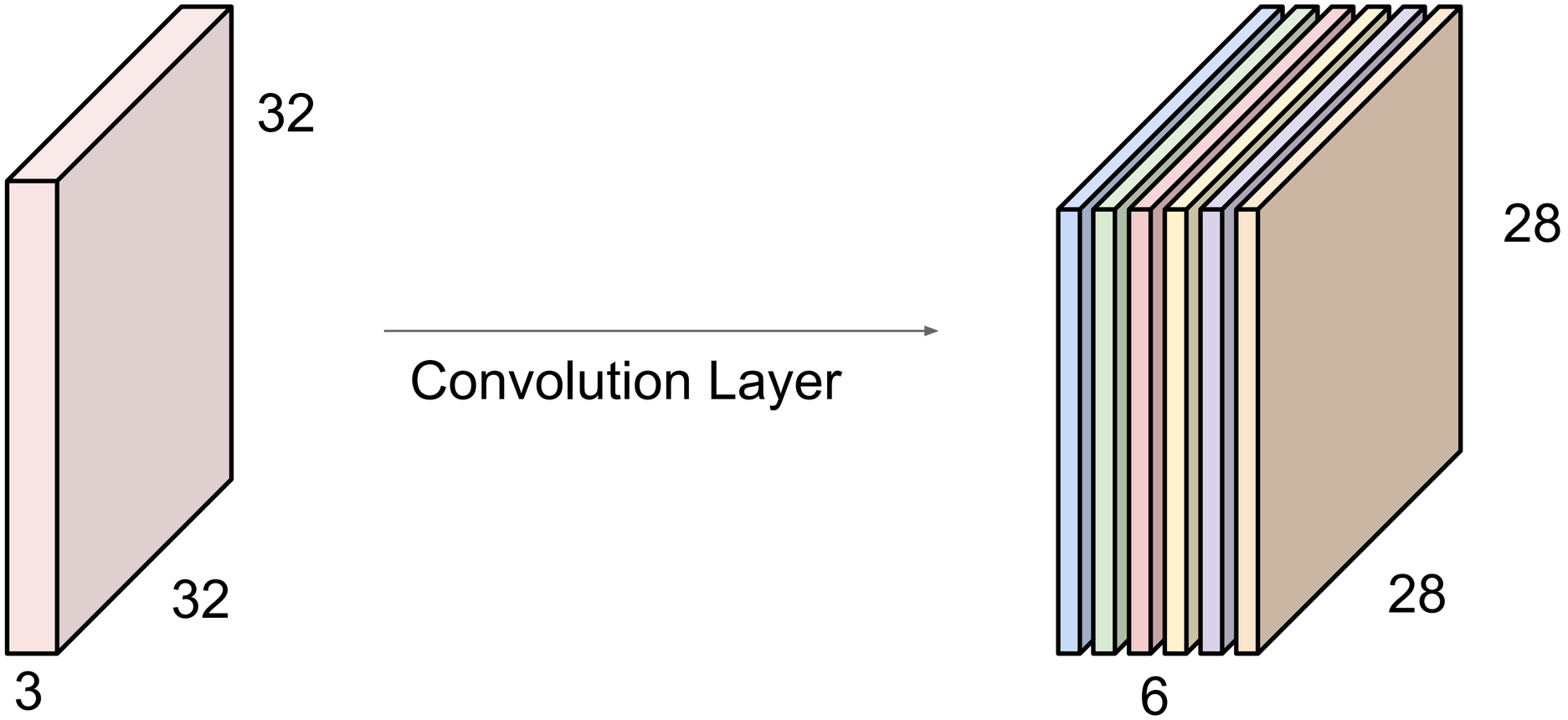


Image from:http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture05.pdf

Convolutional Neural Network

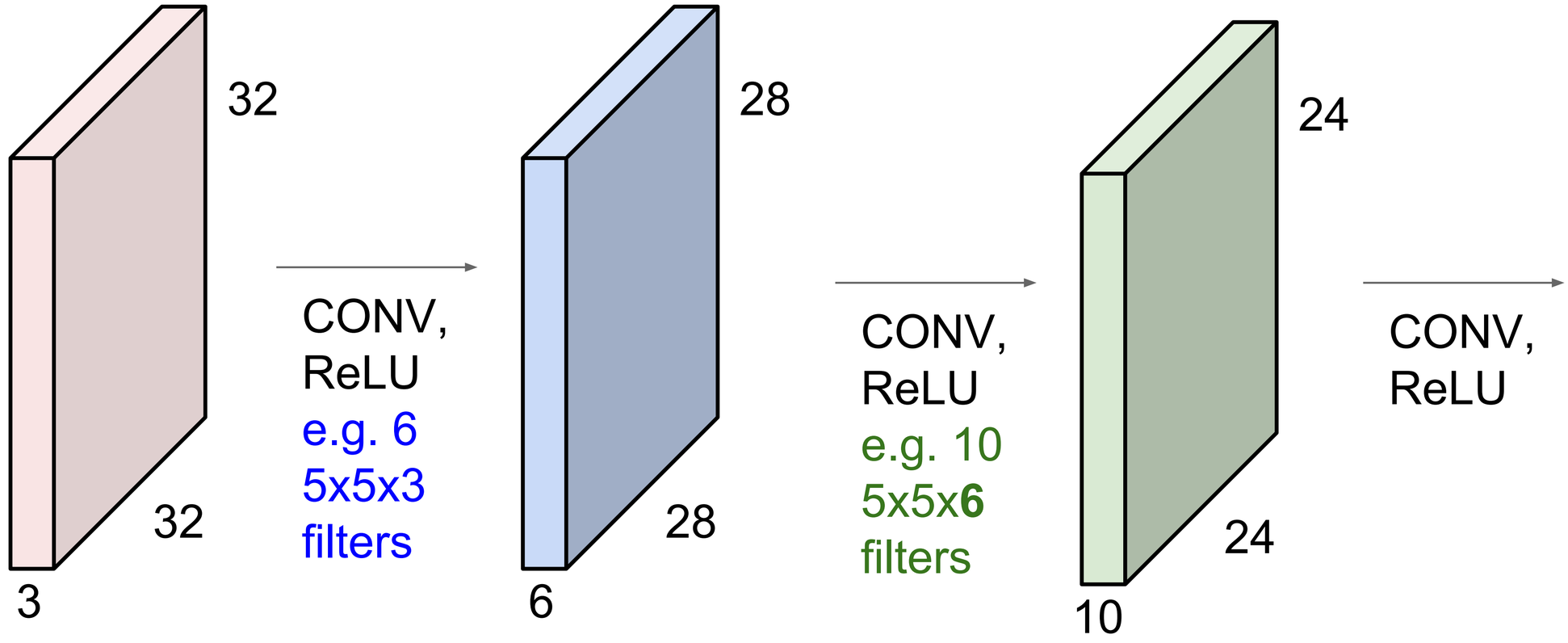


Image from:http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture05.pdf

What do these layers learn?

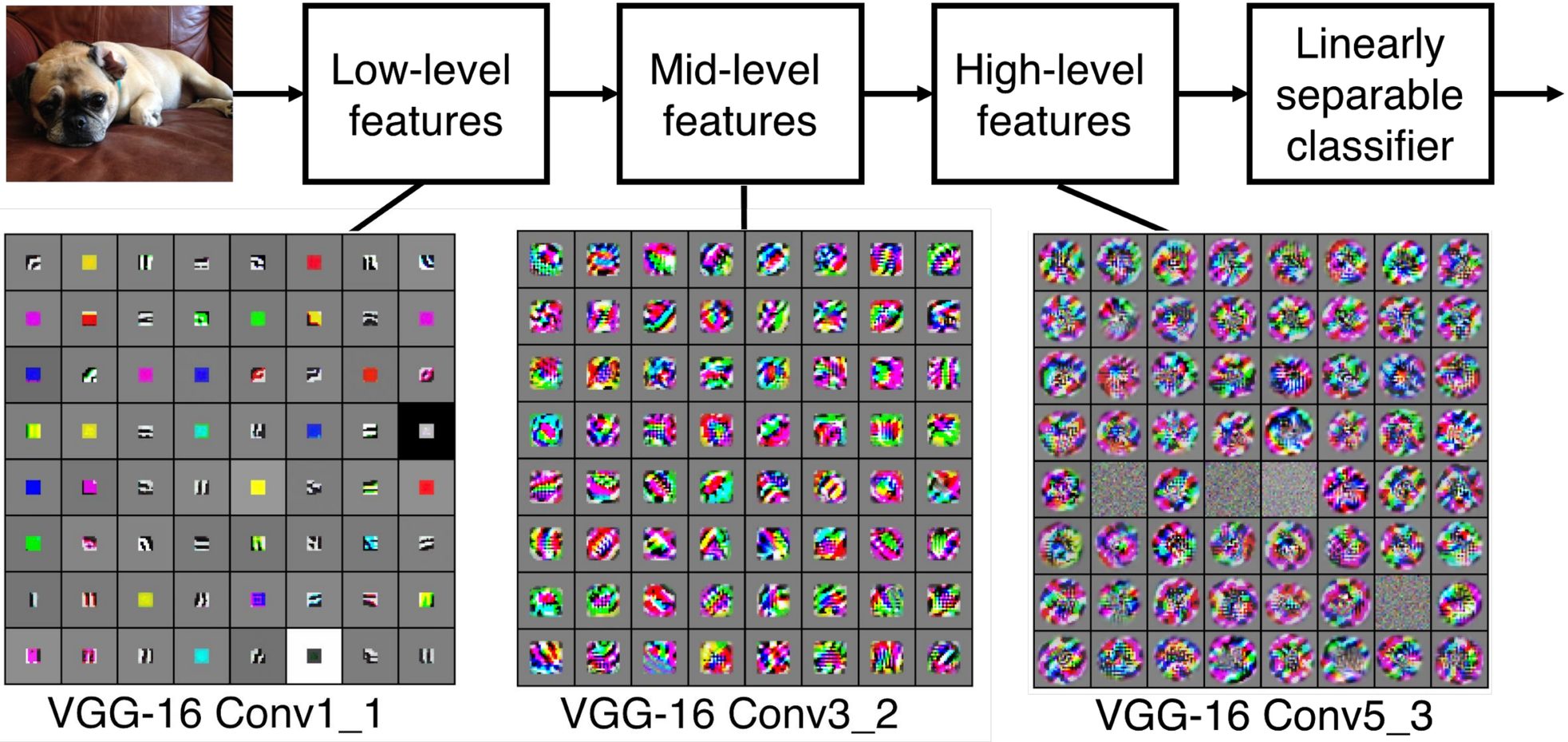
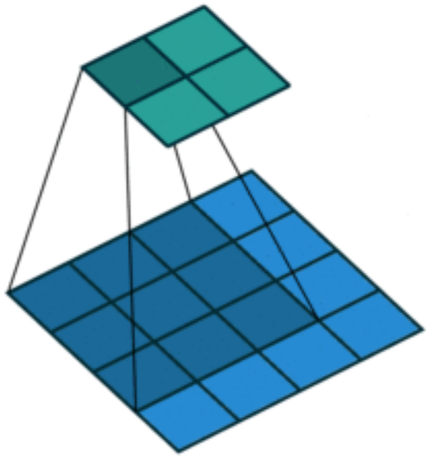


Image from: Simonyan and Zisserman, 2014

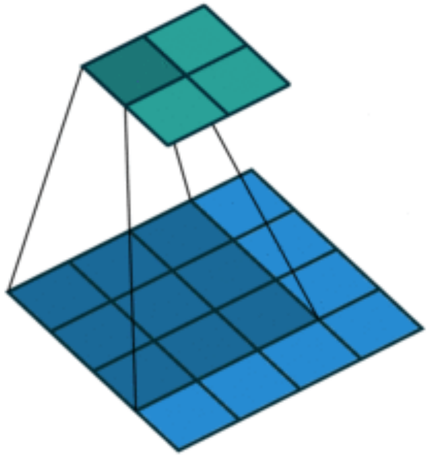
Convolutional Neural Networks (CNNs)



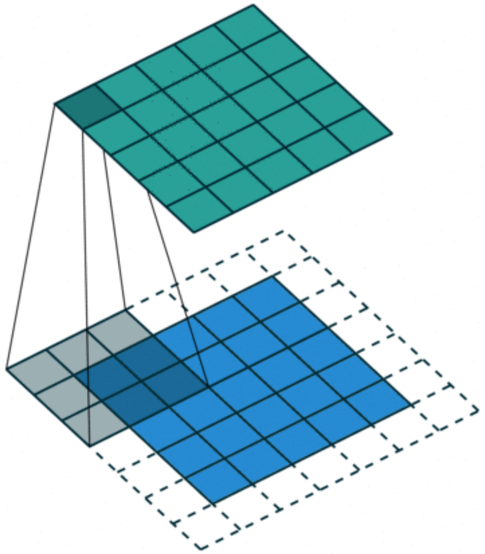
Stride=1, No padding

All animations are from: https://github.com/vdumoulin/conv_arithmetic

Convolutional Neural Networks (CNNs)



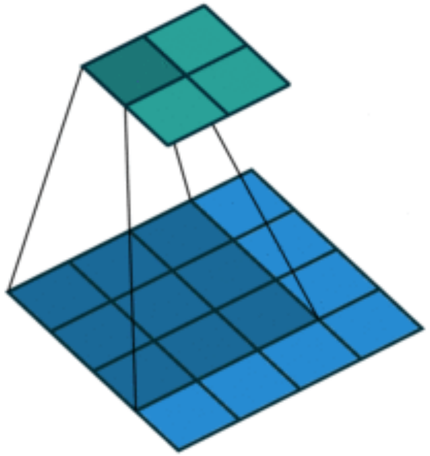
Stride=1, No padding



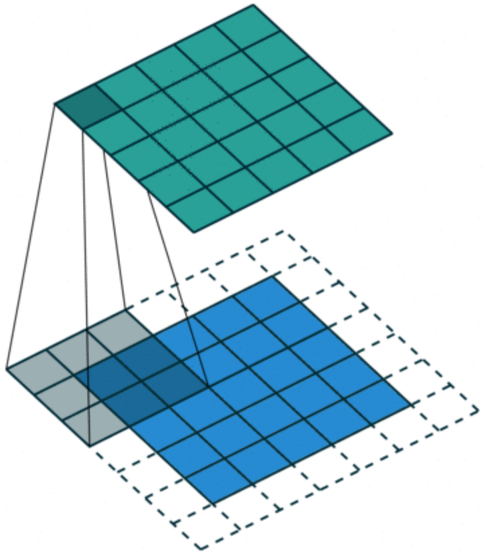
Stride=1, Padding, P=1

All animations are from: https://github.com/vdumoulin/conv_arithmetic

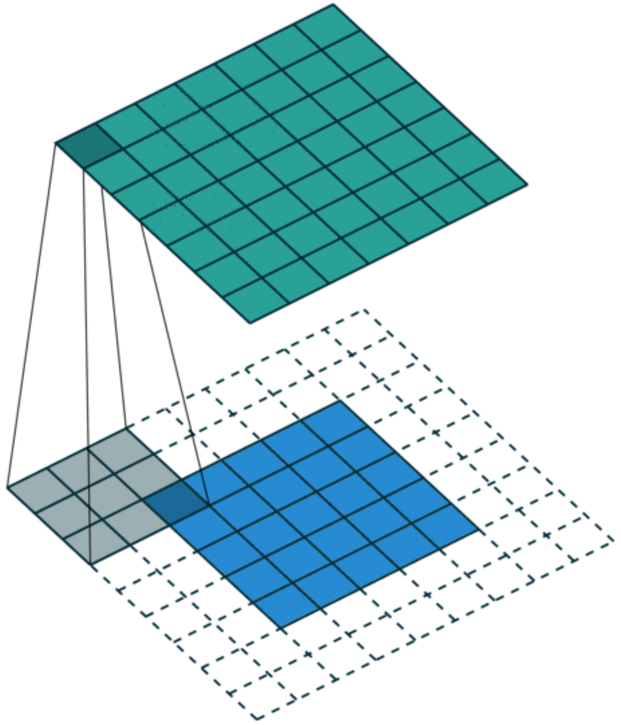
Convolutional Neural Networks (CNNs)



Stride=1, No padding



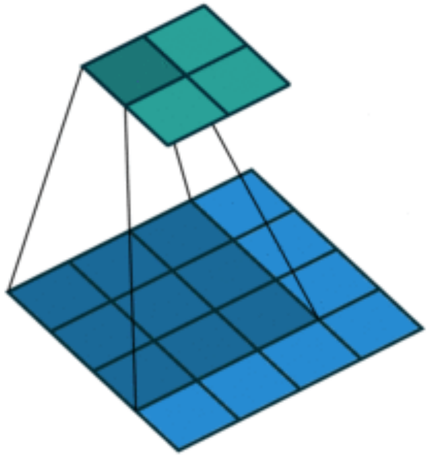
Stride=1, Padding, P=1



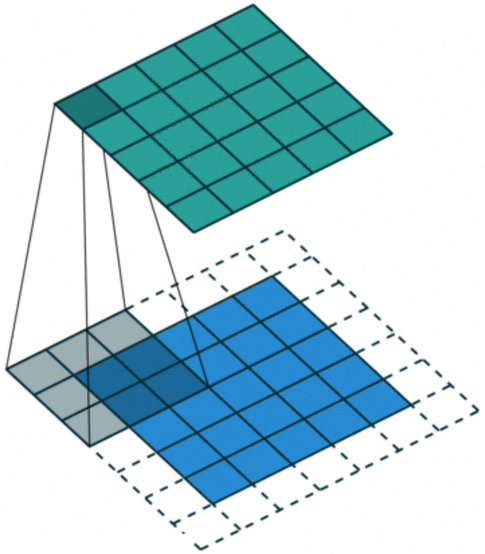
Stride=1, Padding, P=2

All animations are from: https://github.com/vdumoulin/conv_arithmetic

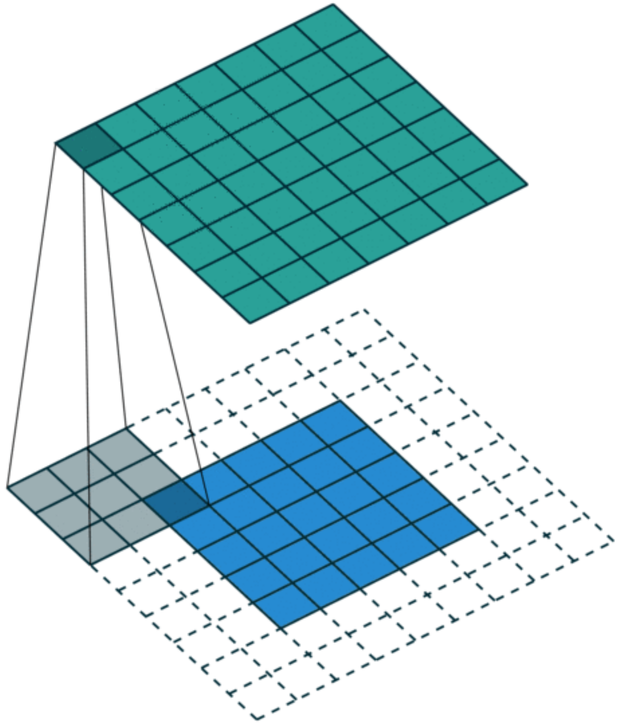
Convolutional Neural Networks (CNNs)



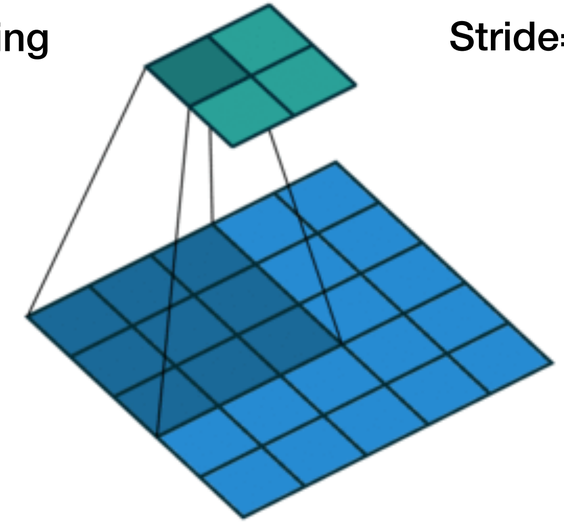
Stride=1, No padding



Stride=1, Padding, P=1



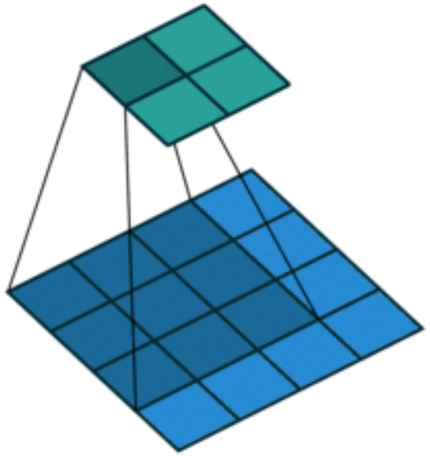
Stride=1, Padding, P=2



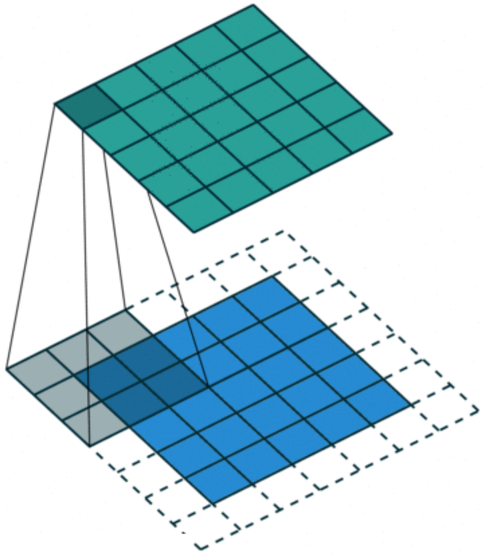
Stride=2, No padding

All animations are from: https://github.com/vdumoulin/conv_arithmetic

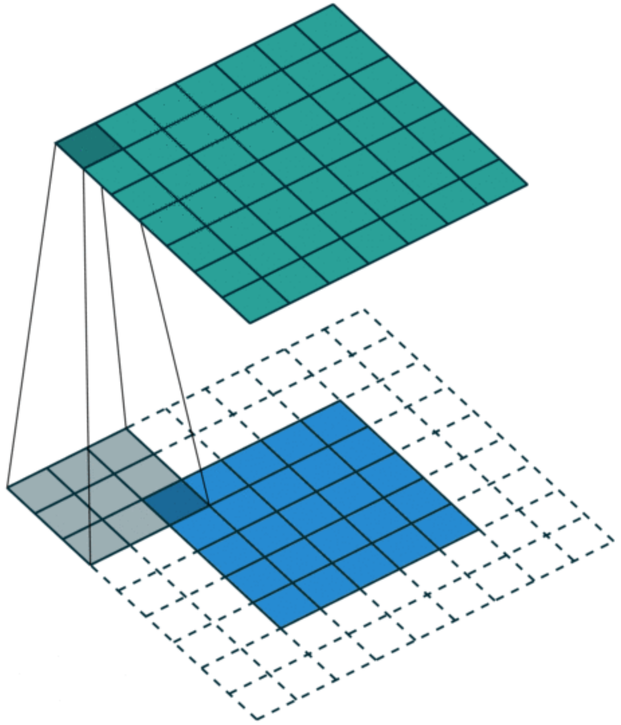
Convolutional Neural Networks (CNNs)



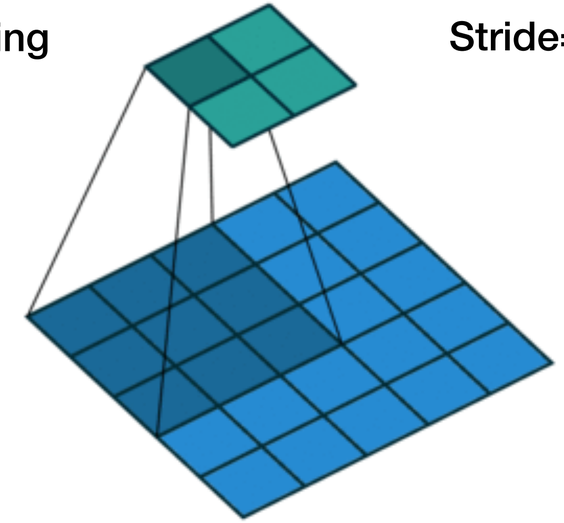
Stride=1, No padding



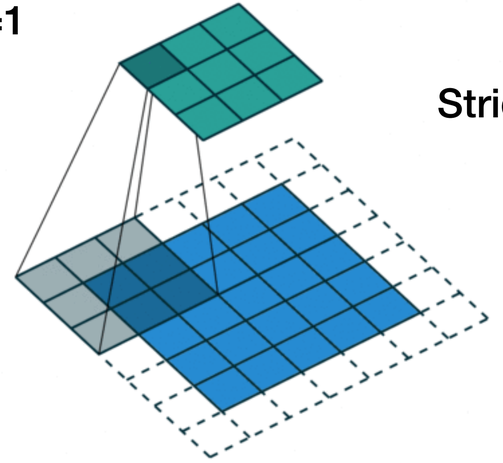
Stride=1, Padding, P=1



Stride=1, Padding, P=2



Stride=2, No padding



Stride=2, Padding, P=1

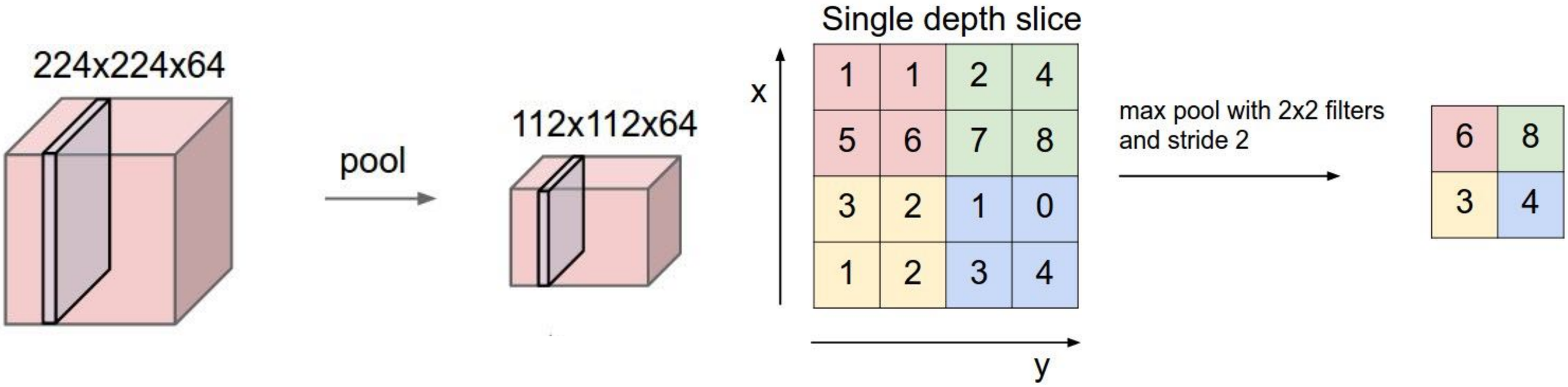
All animations are from: https://github.com/vdumoulin/conv_arithmetic

Convolution Layers: Summary

- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires four hyperparameters:
  - Number of filters K,
  - their spatial extent F,
  - the stride S,
  - the amount of zero padding P.
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
  - $W_2 = \lfloor (W_1 - F + 2P)/S \rfloor + 1$
  - $H_2 = \lfloor (H_1 - F + 2P)/S \rfloor + 1$ (i.e. width and height are computed equally by symmetry)
  - $D_2 = K$
- With parameter sharing, it introduces $F \cdot F \cdot D_1$ weights per filter, for a total of $(F \cdot F \cdot D_1) \cdot K$ weights and K biases.

Summary from: http://cs231n.github.io/convolutional-networks/

Pooling Layer



- Why pooling?
  Reduce the size of the representation, speed up the computations and make the features a little more robust.

- Max pooling is popularly used in CNNs.

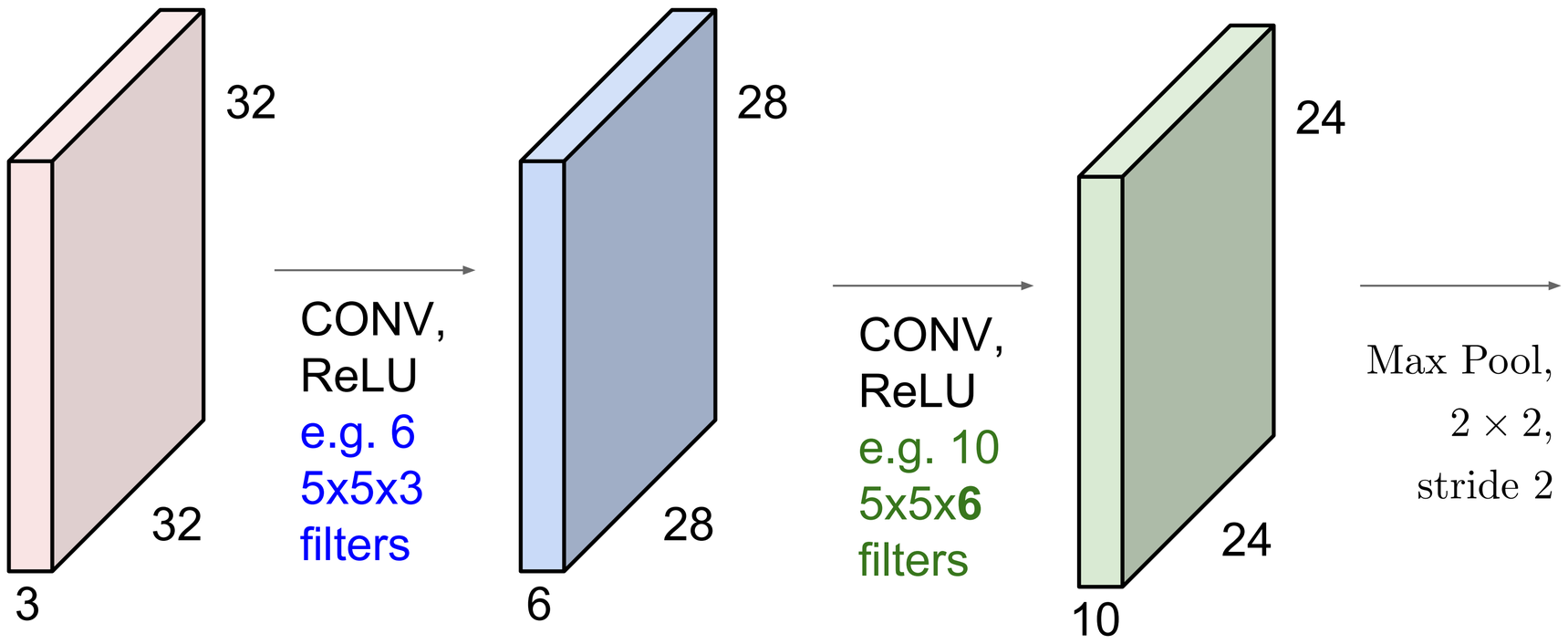
Image from: http://cs231n.github.io/convolutional-networks/

Pooling Layer

- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires two hyperparameters:
  - their spatial extent F,
  - the stride S,
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
  - $W_2 = \lfloor (W_1 - F)/S \rfloor + 1$
  - $H_2 = \lfloor (H_1 - F)/S \rfloor + 1$
  - $D_2 = D_1$

Summary from: http://cs231n.github.io/convolutional-networks/

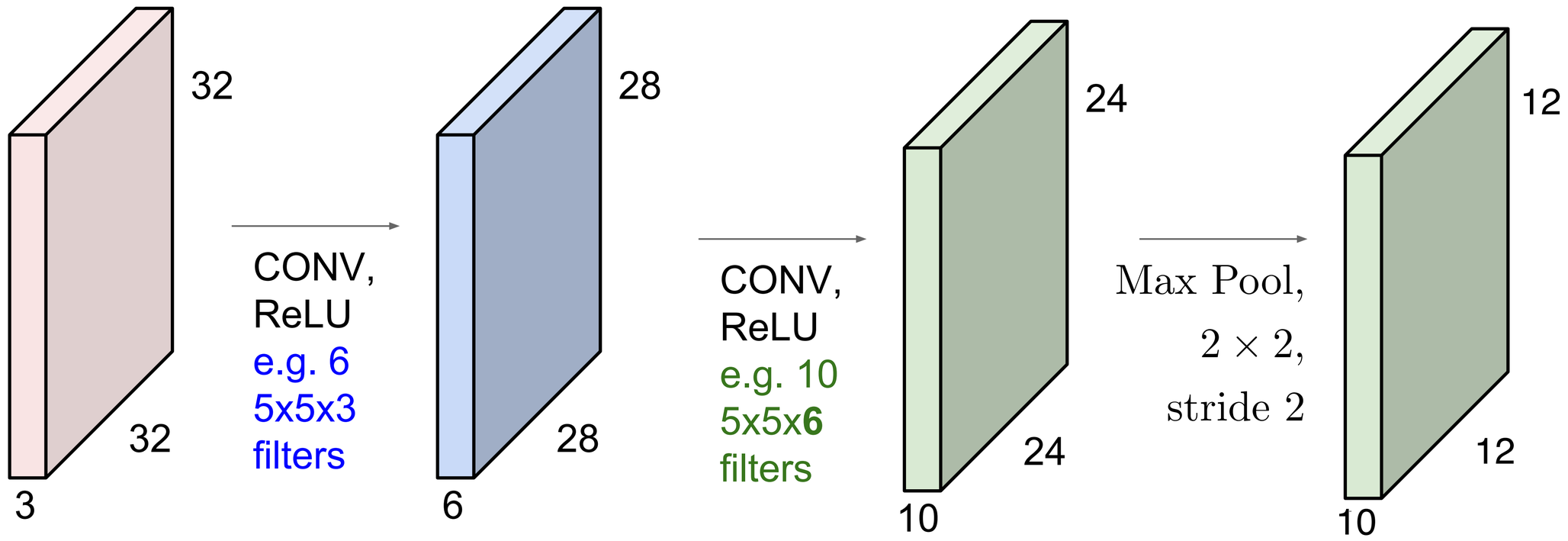
Convolutional NNs



Consider the convolutional neural network shown above with two convolutional layers and a max pooling layer with filters of size 2 x 2, stride 2. What are the dimensions of the output feature map?

(A) 12 x 12 x 10 (B) 23 x 23 x 5 (C) 11 x 11 x 10 (D) 12 x 12 x 5

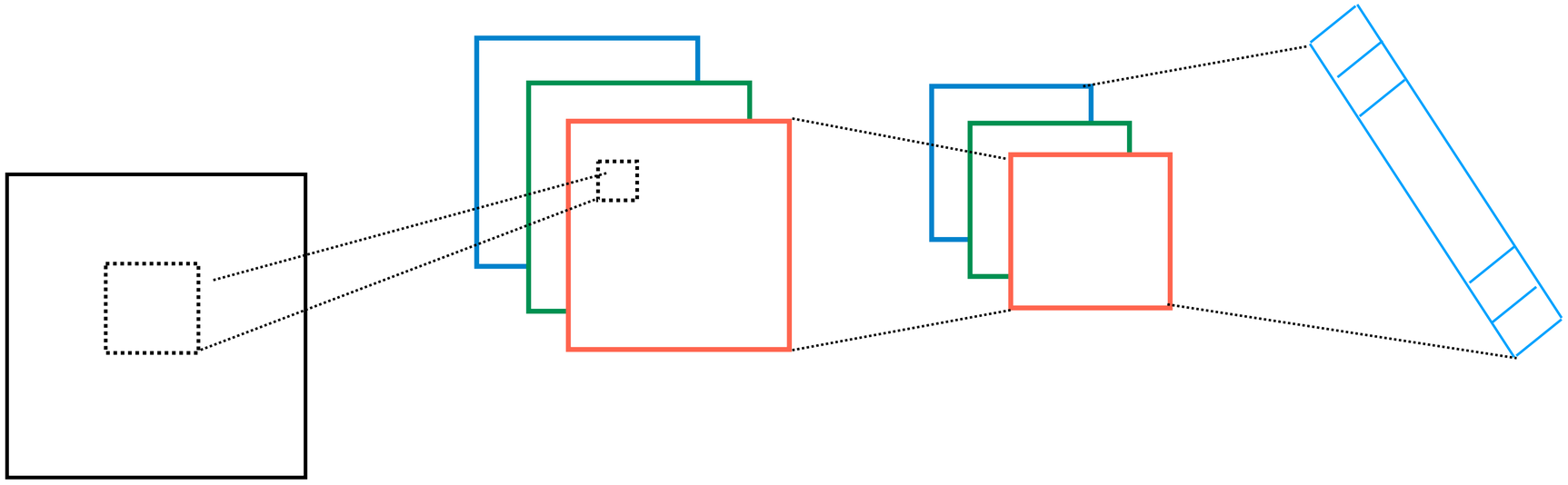
Convolutional NNs



Consider the convolutional neural network shown above with two convolutional layers. How many parameters need to be estimated overall for this network? (Ignore biases.)

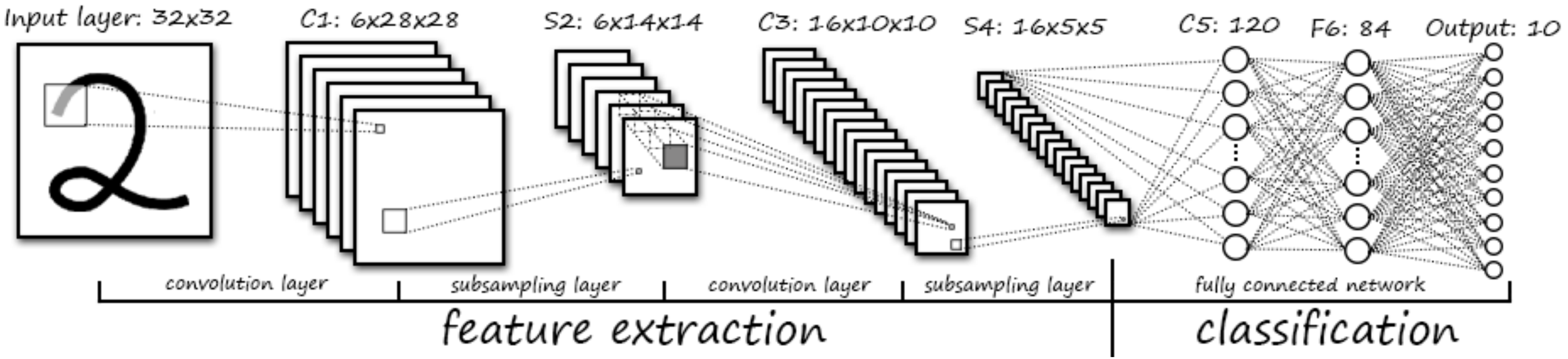
(A) 1500 (B) 1990 (C) 1950 (D) 1954

Convolutional Architectures



- Block that can be repeated: Convolutional layer, followed by non-linearity (e.g. ReLU) + Max pooling
- Fully connected layers before classification

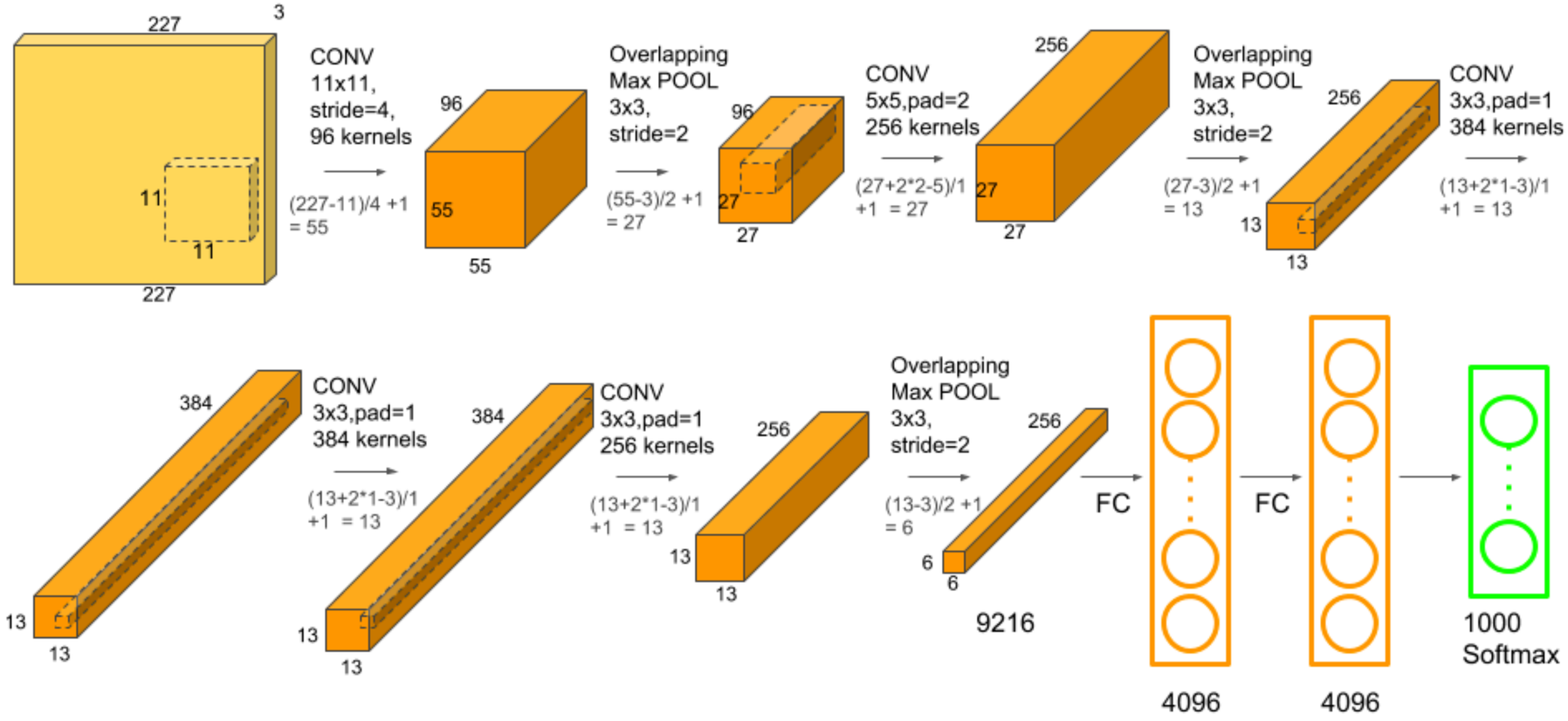
LeNet-5



- One of the first successful CNN architectures
- Used to classify images of hand-written digits

LeNet-5: LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. “Gradient-based learning applied to document recognition”. Proceedings of the IEEE, 86(11), pp.2278-2324.

AlexNet

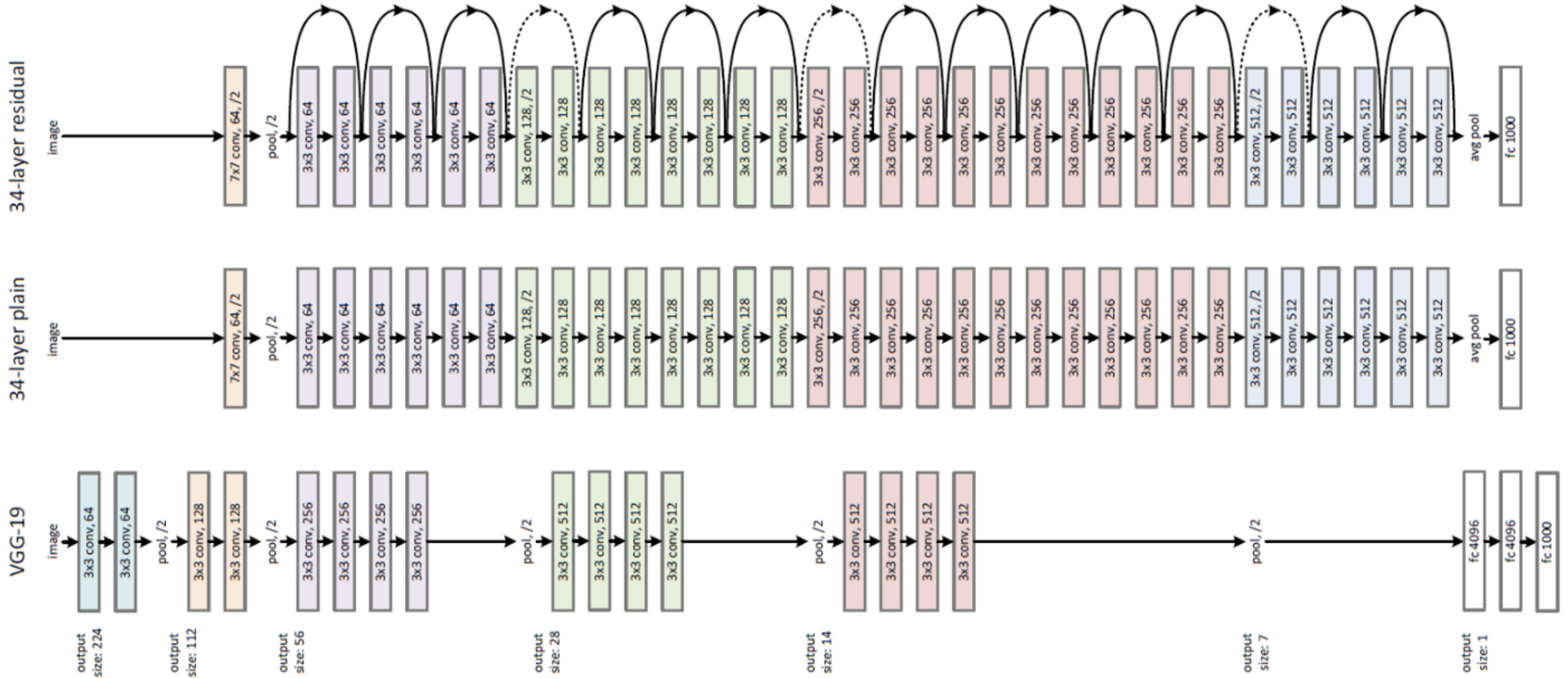


- Winner (by a large margin) of the ImageNet challenge in 2012.
- Much larger than previous architectures.

AlexNet: ImageNet Classification with Deep Convolutional Neural Networks by Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, NIPS 2012

Image from: https://www.learnopencv.com/understanding-alexnet/

ResNet



- Uses residual blocks

Image from: https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015-image-classification-localization-detection-e39402bfa5d8