Convolutional Neural Networks (CNN)

A Convolutional Neural Network (CNN) consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers.

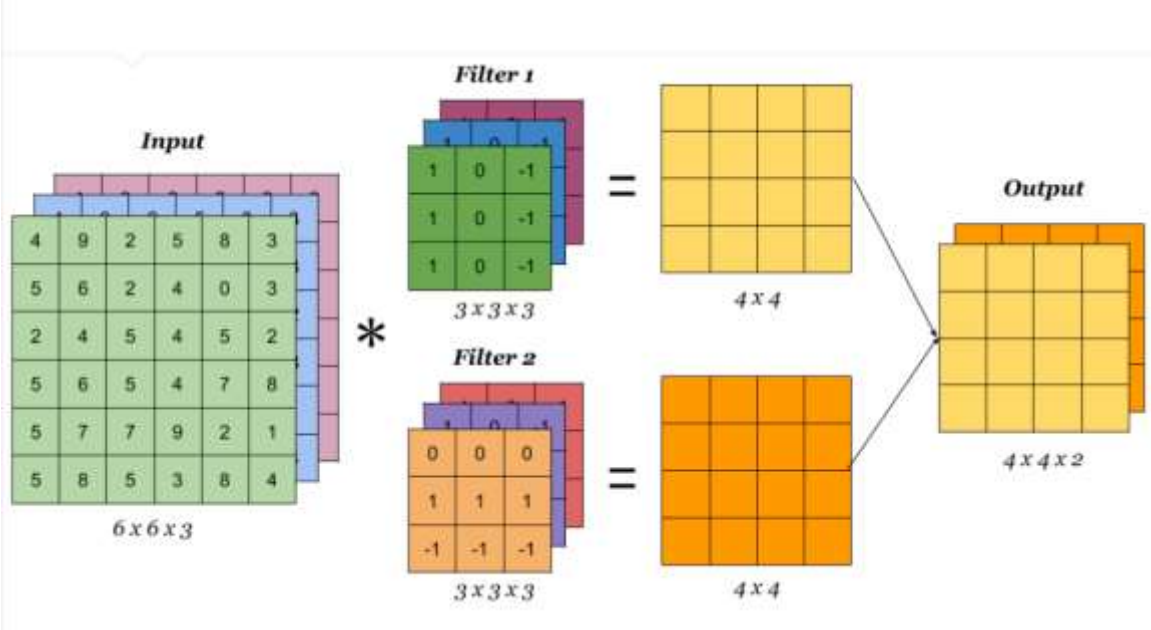

Figure 1

Why Convolutions

Parameter sharing: a feature detector (such as a vertical edge detector) that's useful in one part of the image is probably useful in another part of the image.

Sparsity of connections: in each layer, each output value depends only on small number of inputs.

Convolution Operation

Step 1: overlay the filter to the input, perform element wise multiplication, and add the result.

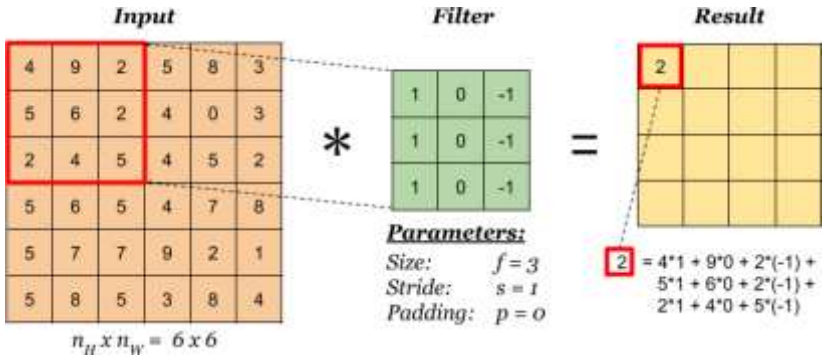

Figure 2

Step 2: move the overlay right one position (or according to the stride setting), and do the same calculation above to get the next result. And so on.

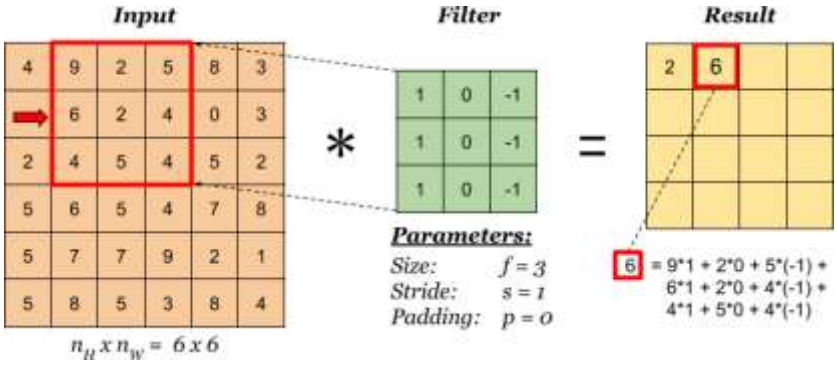

Figure 3

The total number of multiplications to calculate the result above is (4*x*4)*x*(3*x*3) = 144.

Stride

Stride governs how many cells the filter is moved in the input to calculate the next cell in the result.

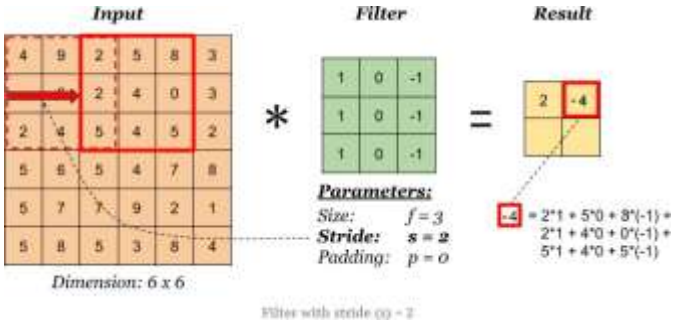

Figure 4

The total number of multiplications to calculate the result above is $(2x2)x(3x3) = 36$.

Padding

Padding has the following benefits:

1. It allows us to use a CONV layer without necessarily shrinking the height and width of the volumes. This is important for building deeper networks, since otherwise the height/width would shrink as we go to deeper layers.
2. It helps us keep more of the information at the border of an image. Without padding, very few values at the next layer would be affected by pixels as the edges of an image.

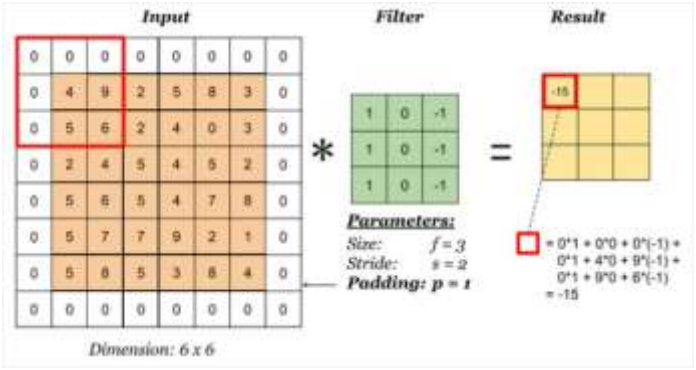

Figure 5

References:

For more details on CNN can refer below:
http://colah.github.io/posts/2014-07-Conv-Nets-Modular/
https://towardsdatascience.com/convolutional-neural-networks-from-the-ground-up-c67bb41454e1