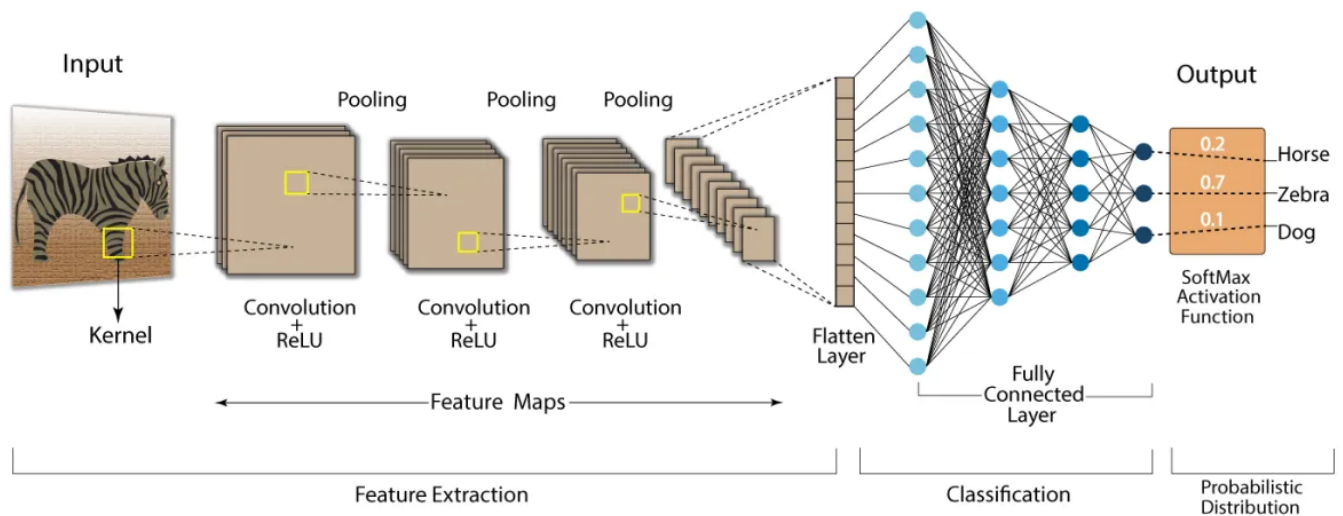


## Convolution Neural Network (CNN)



Credits: Ahlad Kumar [Deep Learning Series] &

<https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/>

### ▼ 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.

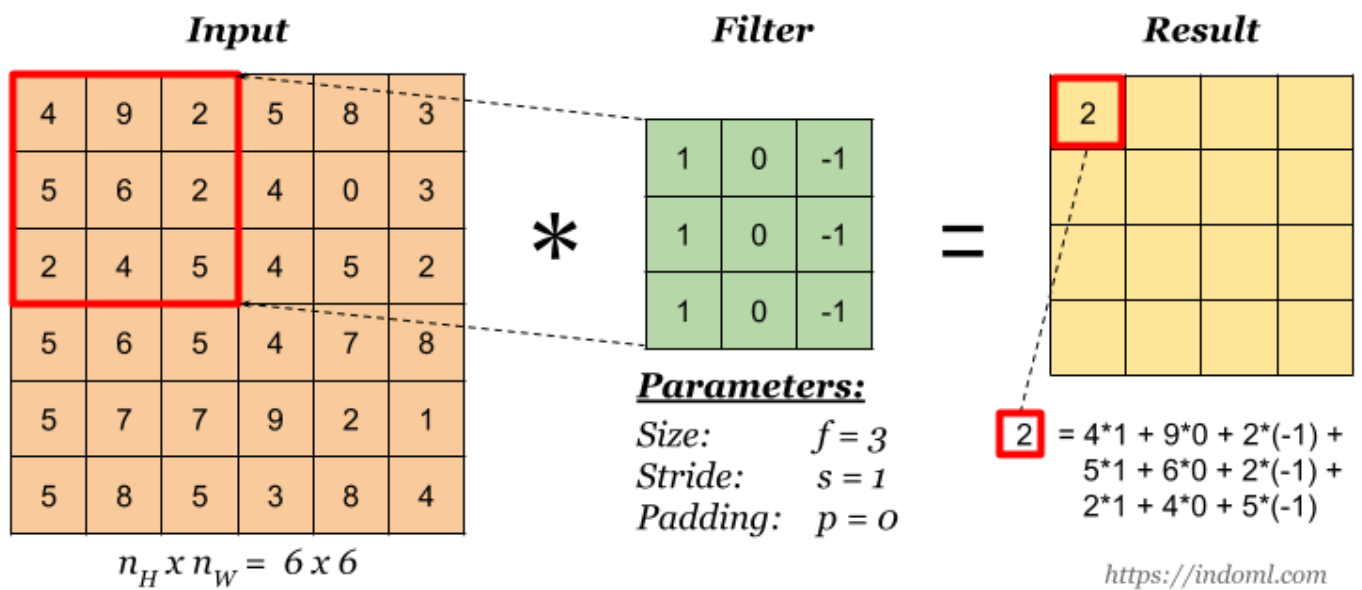
**Sparsity of connections:** in each layer, each output value depends only on small number of inputs.

Three inputs in Convolutional Neural Networks are

- Filter Size
- Stride
- Padding

### ▼ Convolution Operation

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



The Dimension of output matrix is given by

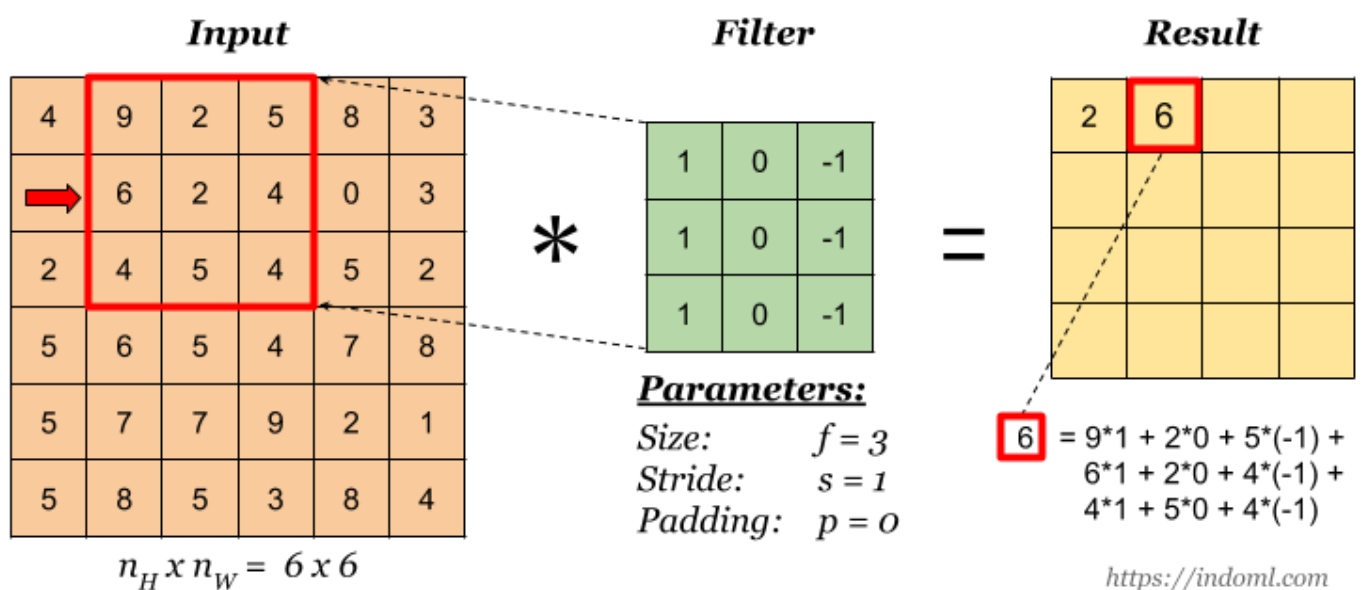
$$(nh-f+1) * (nw-f+1)$$

$$(6-3+1) * (6-3+1)$$

$$4 * 4$$

## ▼ Step 2:

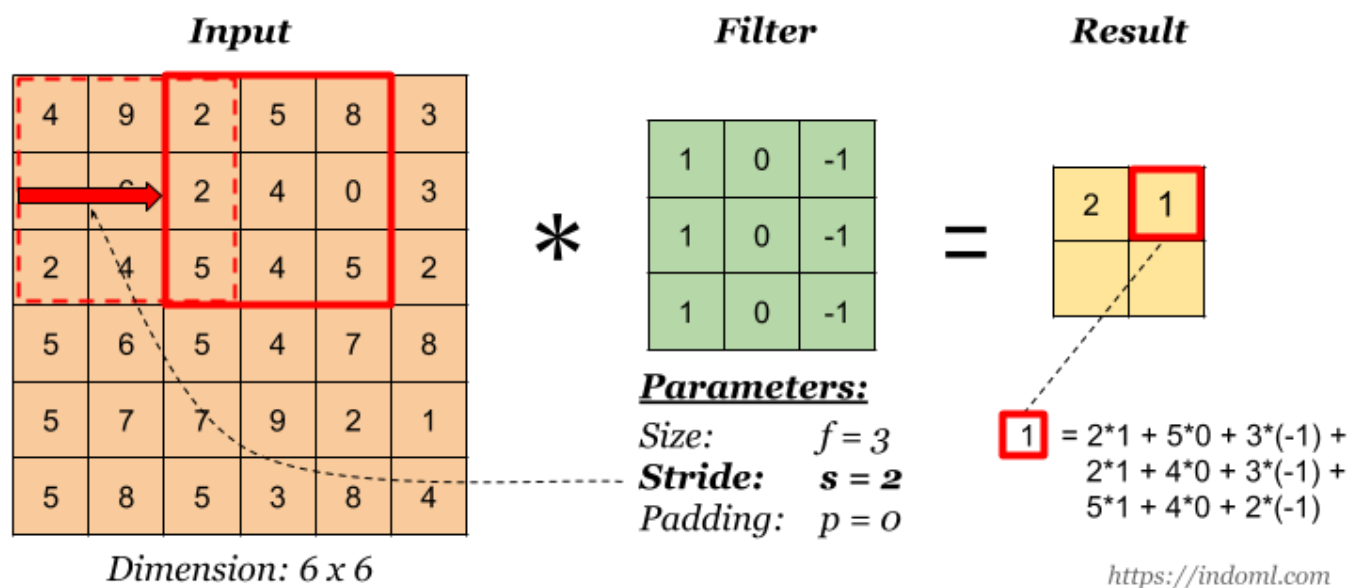
Move the overlay right one position (or according to the stride setting), and do the same calculation above to get the next result. And so on.



The total number of multiplications to calculate the result above is  $(4 \times 4) \times (3 \times 3) = 144$ .

## ▼ Stride

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



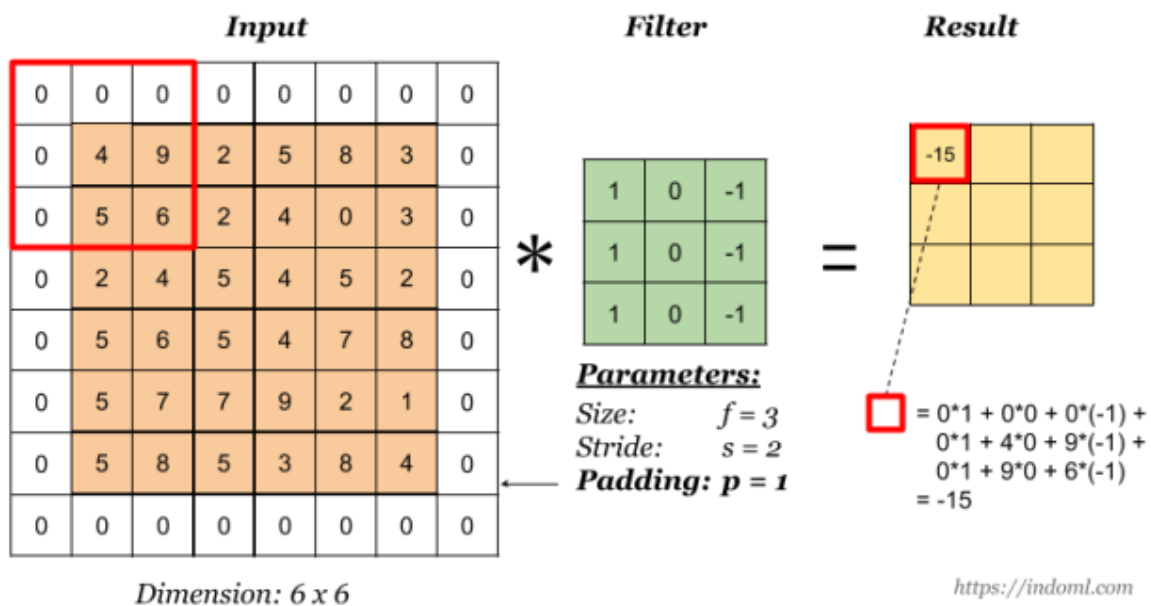
The total number of multiplications to calculate the result above is  $(2 \times 2) \times (3 \times 3) = 36$

## ▼ Padding

Padding has the following benefits:

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.

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.



Notice the the dimension of the result has changed due to padding. See the following section on how to calculate output dimension.

Some **padding terminologies**:

“**valid**” padding: **no padding**

“**same**” padding: padding so that the output dimension is the same as the input

## ▼ Calculating the Output Dimension

$$n^{[l]} = \left\lfloor \frac{n^{[l-1]} + 2p^{[l-1]} - f^{[l]}}{s^{[l]}} + 1 \right\rfloor$$

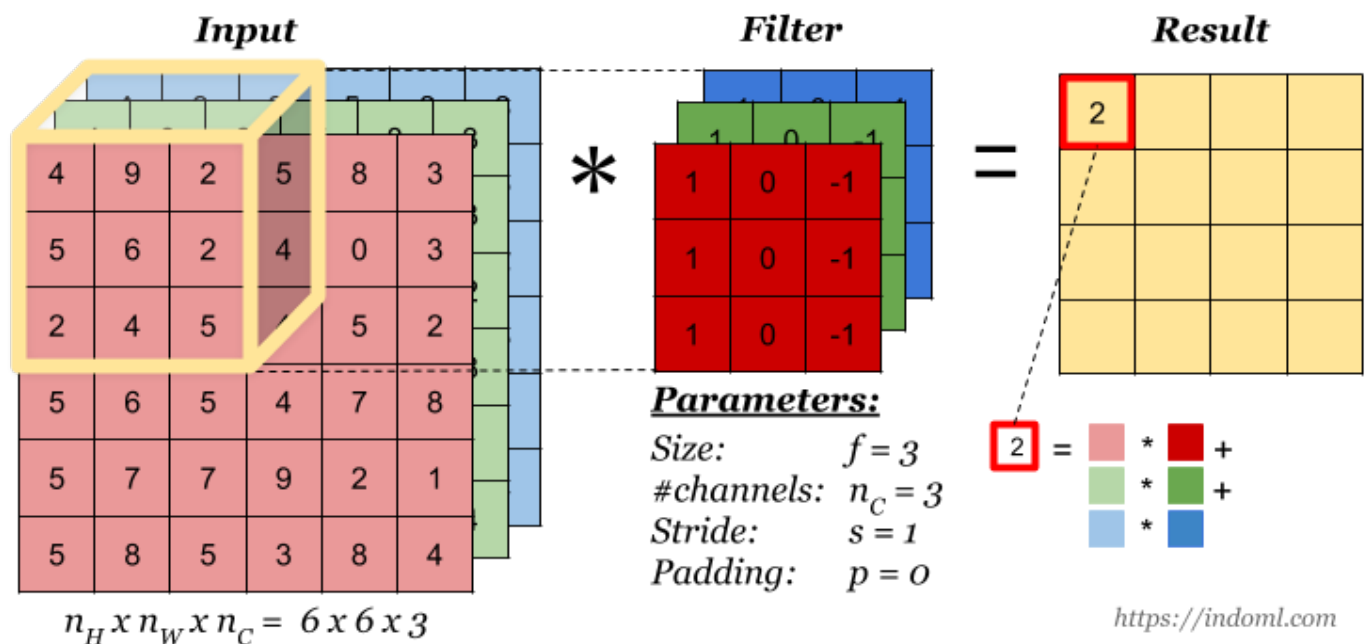
output\_image\_dimension=

[[6+2\*1-3/2]]+1]]

3

## ▼ Convolution Operation on Volume

When the **input has more than one channels** (e.g. an RGB image), the **filter should have matching number of channels**. To calculate one output cell, perform convolution on each matching channel, then add the result together.

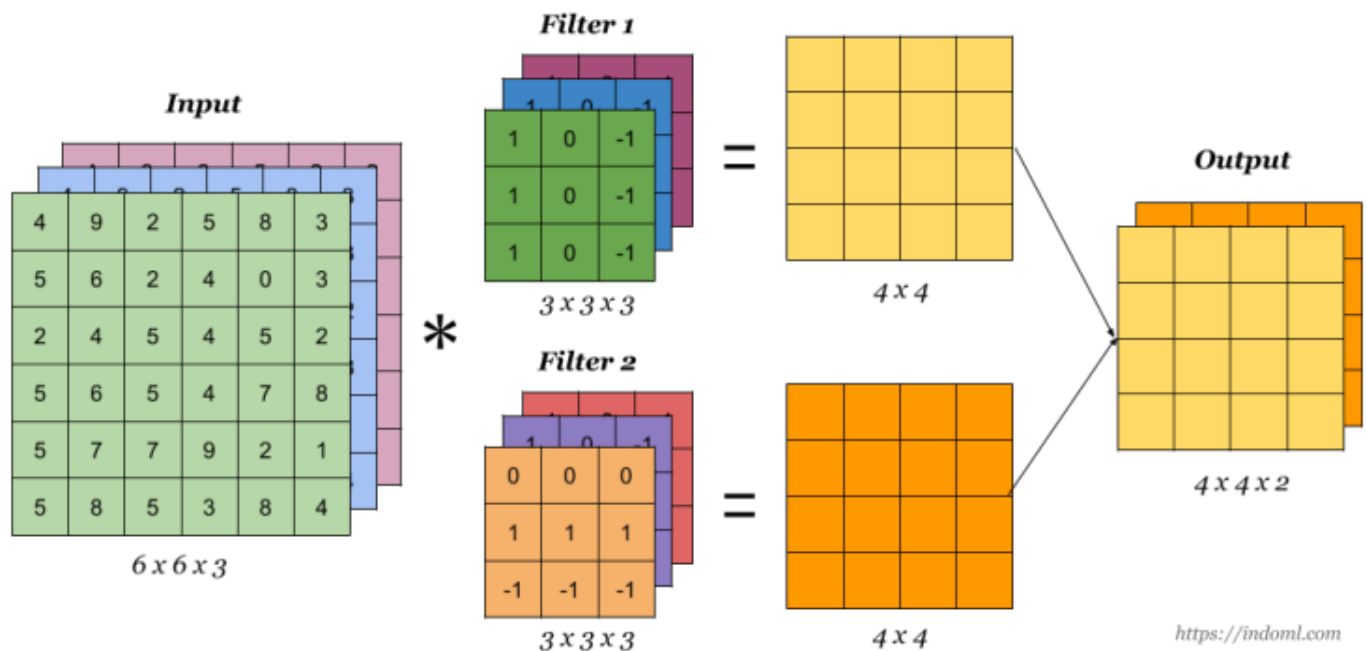


The total number of multiplications to calculate the result is  $(4 \times 4) \times (3 \times 3 \times 3) = 432$ .

## Convolution Operation with Multiple Filters

Multiple filters can be used in a convolution layer to detect multiple features.

The output of the layer then will have the same number of channels as the number of filters in the layer.



The total number of multiplications to calculate the result is  $(4 \times 4 \times 2) \times (3 \times 3 \times 3) = 864$ .

## 1 x 1 Convolution

This is **convolution with 1 x 1 filter**.

The **effect is to flatten or “merge” channels together**, which can save computations later in the **network**:

