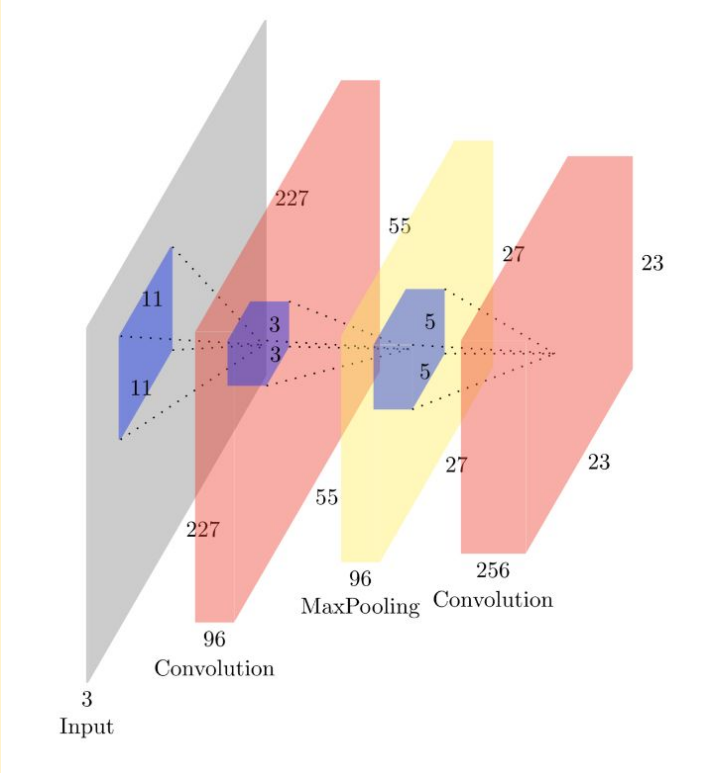
PadhAI: From Convolution Operation to Neural Network

One Fourth Labs

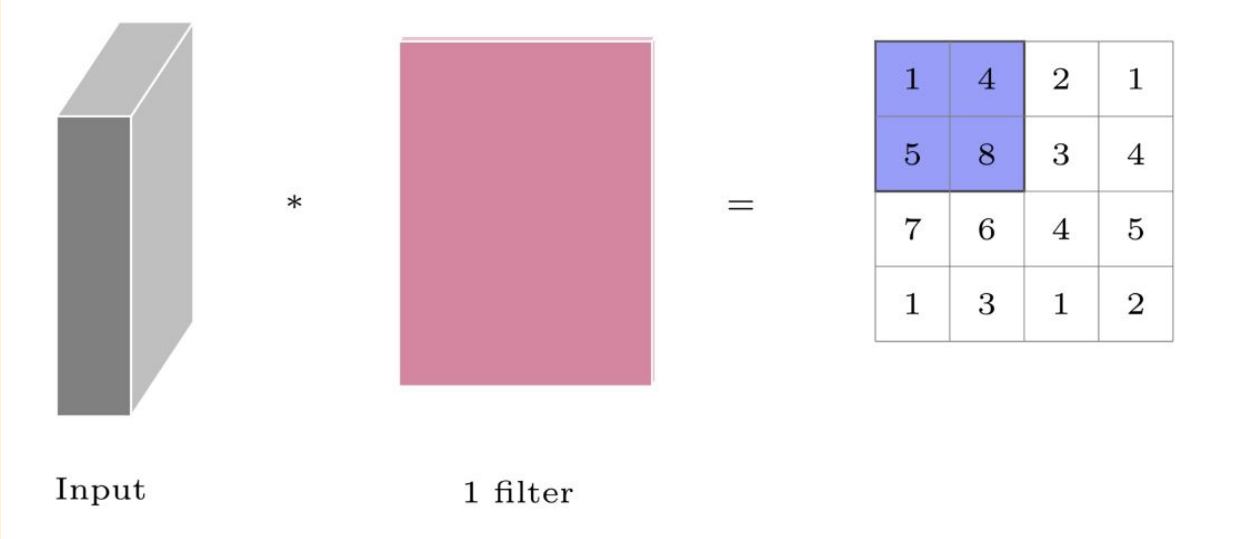
Max Pooling and Non-Linearities

What is the max pooling operation?

1. Let us look at a diagram of a CNN to better understand what max pooling does.

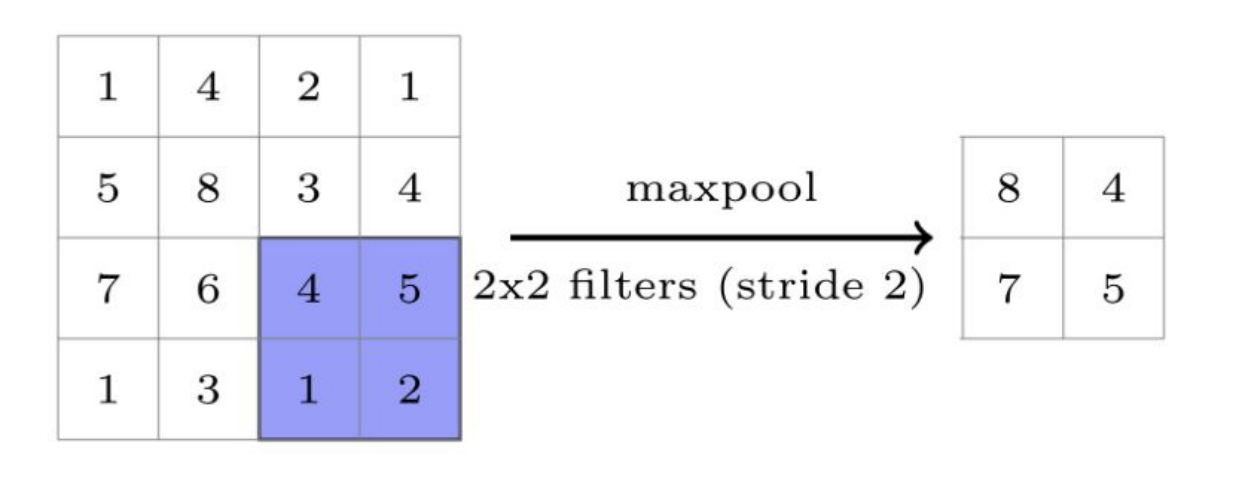


2. Here, as we have discussed earlier, the first operation performed is a convolutional transformation using a filter.



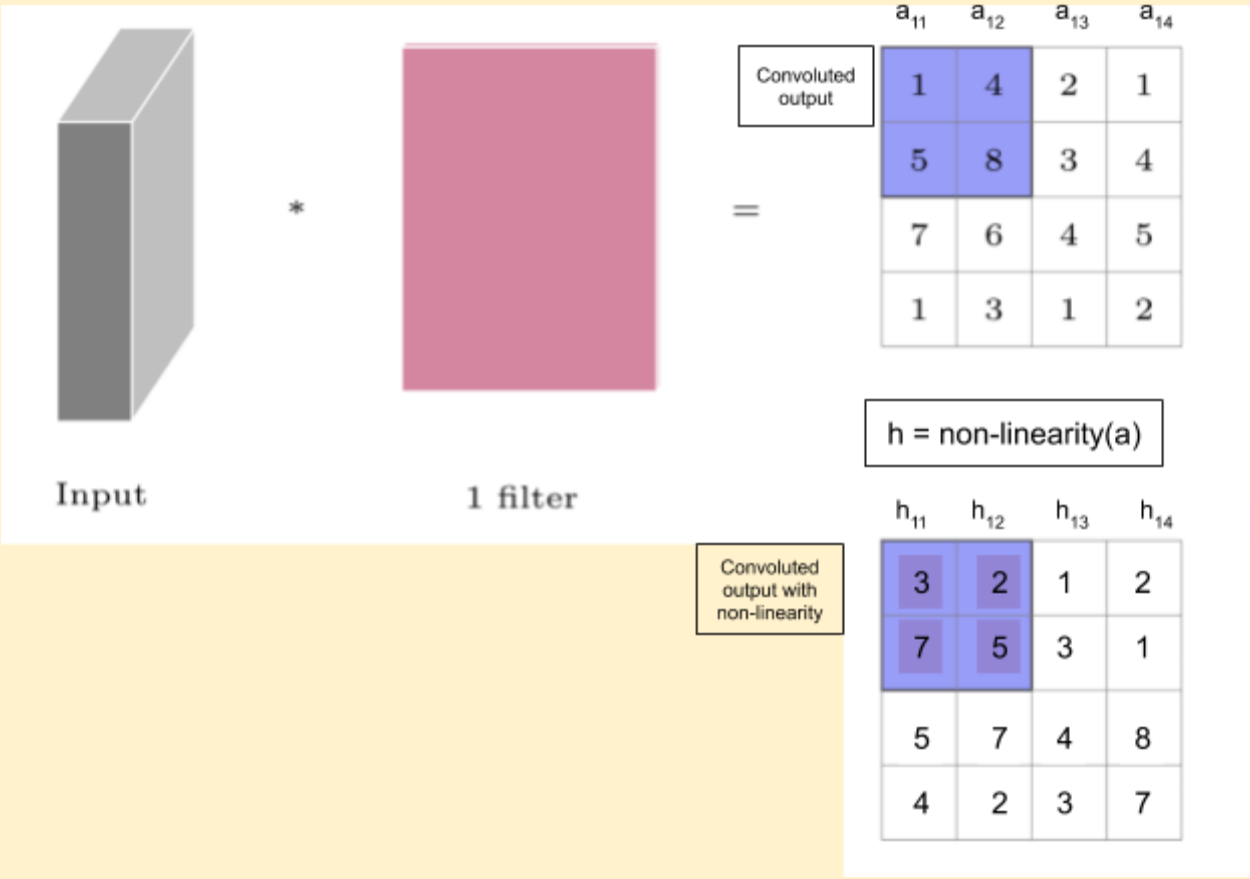
   a. Here, by passing the filter over an image (with or without padding), we get a transformed matrix of values

3. Now, we perform max-pooling over the convoluted input to select the max-value from each position of the kernel, as specified by stride length.



   a. Here, we select a stride length of 2 and a 2x2 filter, meaning the 4x4 convoluted output is split into 4 quadrants.
   b. The max value of each of these quadrants is taken and a 2x2 matrix is generated.
   c. Max pooling is done to select the most prominent or salient point within a neighborhood. It is also known as subsampling, as we are sampling just a single value from a region.
   d. Similar to Max pooling, average pooling is also done sometimes and it's carried out by taking the average value in a sampled neighborhood.
   e. The idea behind Max Pooling is to condense the convolutional input into a smaller size, thereby making it easier to manage.

4. Another point to consider is the application of nonlinearities to the convoluted output.



a. Here, we can consider the convoluted output to be similar to the pre-activation layer of a neuron.
b. By applying a non-linear transformation like sigmoid, tanh, ReLU etc, we are effectively transforming the convoluted output
c. The resultant transformed matrix is then passed through the subsequent stages in the CNN, such as Max Pooling etc.
d. Thus, we are creating a non-linear relationship between the input and the output, thereby allowing us to approximate more complex functions.
e. Yet at the same time, we avoid over-complexity by reducing the number of parameters by weight sharing and condensing the convoluted output using max pooling.