12. Convolutional Neural Networks (CNNs)

- CNNs exploit the spatial structure of the input.
  - inspired by the human visual system
  - primarily suited for images / video
- CNNs have much fewer parameters in comparison to standard neural nets. Hence, very efficient to train deep CNNs. in contrast to traditional fully-connected nets.
- Convolution between 2 signals is defined as

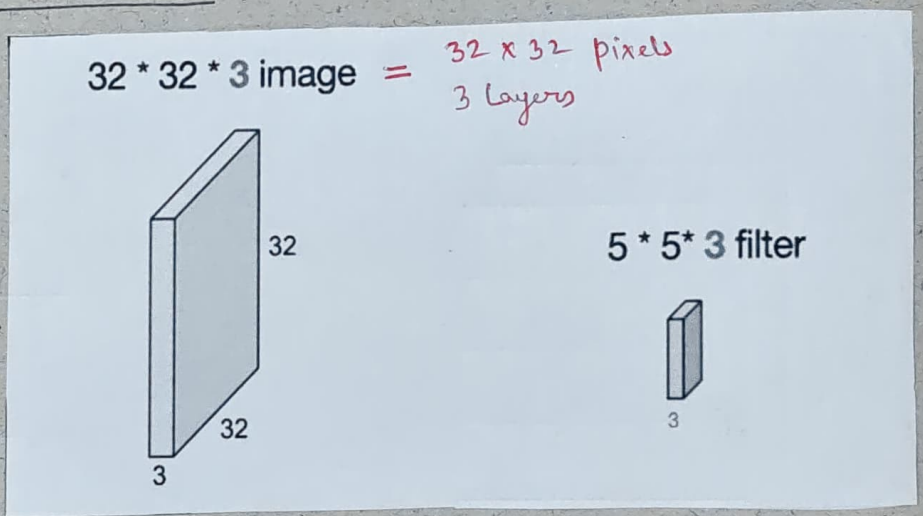
$$(f * g)(n) = \sum_{m=-\infty}^{\infty} f(m)\, g(n-m)$$

input, impulse response

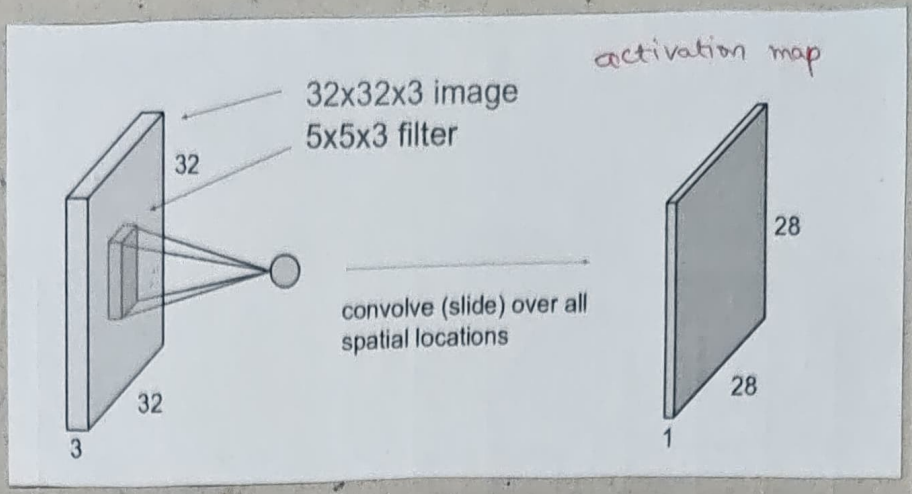
$$= \sum_{m=-\infty}^{\infty} f(n-m)\, g(m)$$

Images 2D convolution

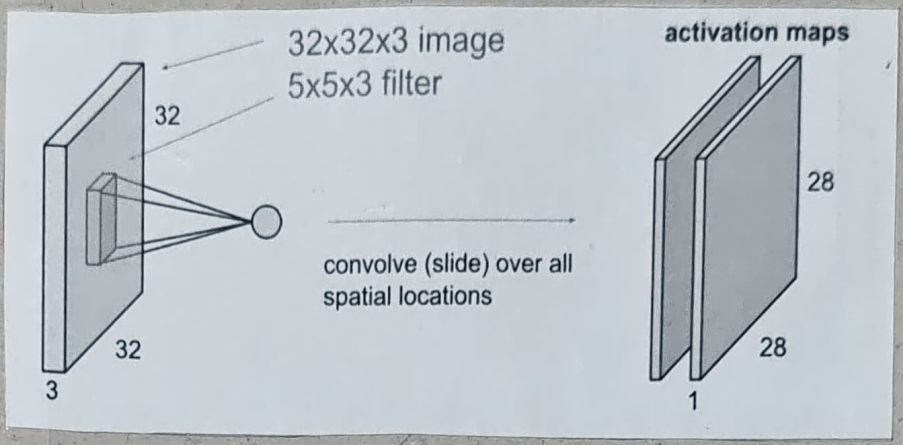
- Slide 5x5x3 filter over the image.
- Compute the Spatial dot products.



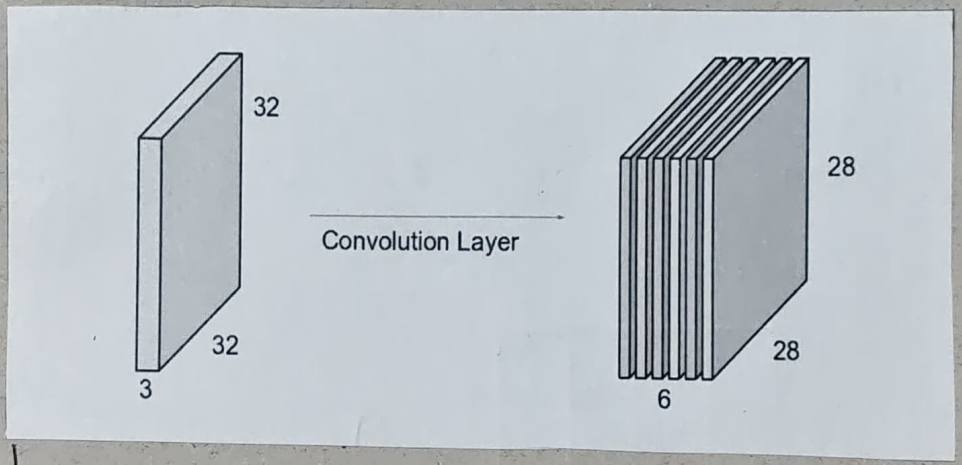
⊙ The convolution produces an activation map.



⊙ Different filters produce different activation maps.

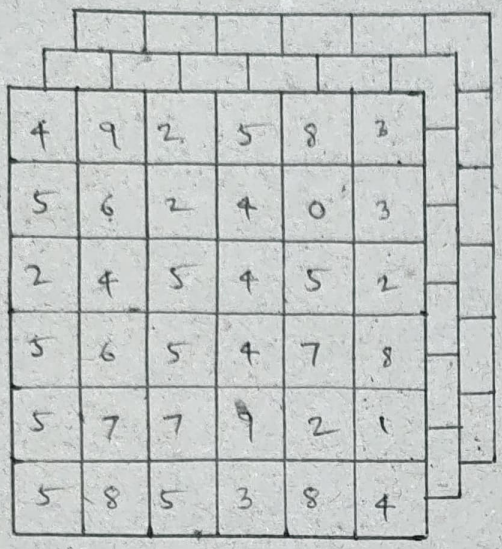


⊙ Use 6 different filters, we get a new image of size 28 × 28 × 6 !



Problem 1: What is the size of the output after convolution below ?

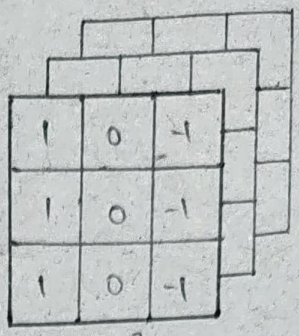
Input



6 × 6 × 3

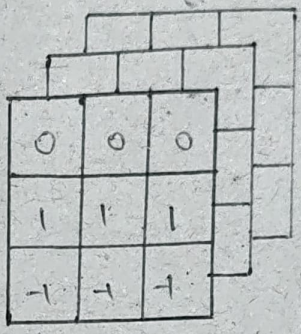
*

Filter 1



3 × 3 × 3

Filter 2

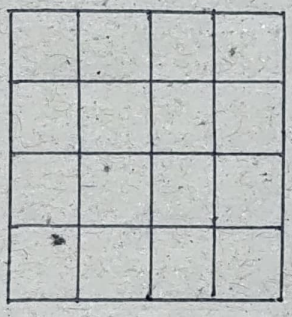


3 × 3 × 3

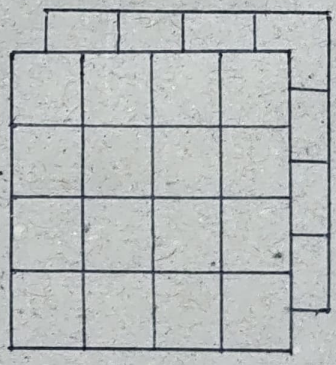
Solution:

Each filter produces a 4 × 4 output.

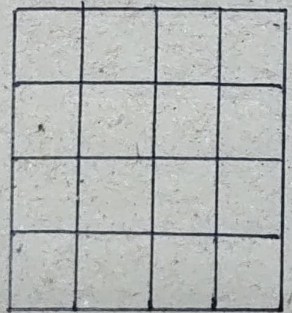
Net output is 4 × 4 × 2



4 × 4

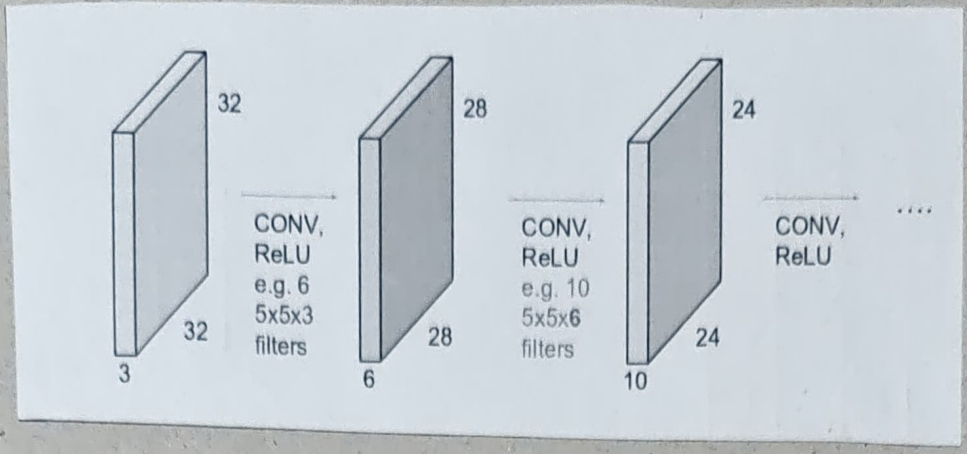


4 × 4 × 2



4 × 4

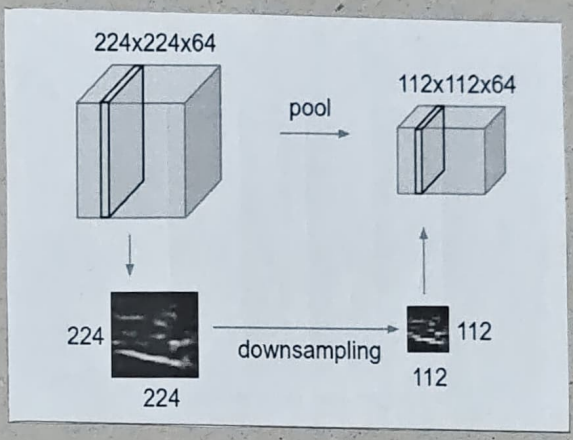
CNN Structure



CNN comprises of a sequence of alternating convolutional layers and activation functions

Pooling

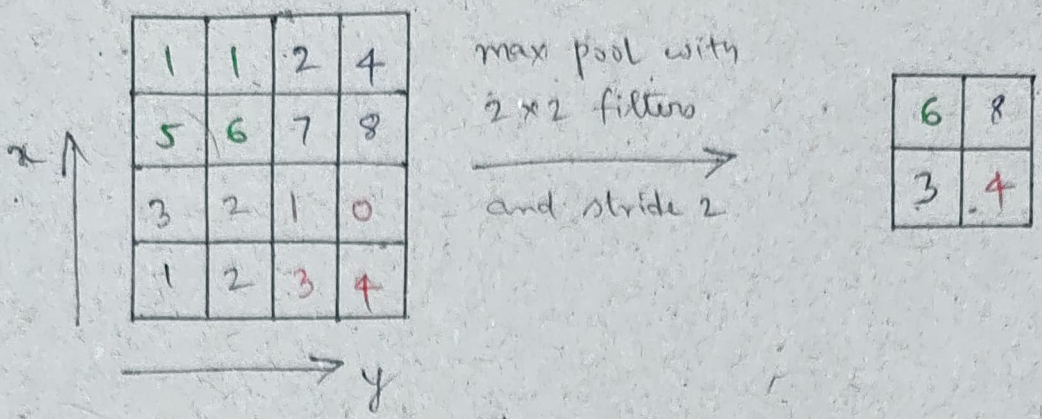
- Pooling compresses the activation maps, to make them more manageable



- It operates independently over each activation map. It is basically a downsampling operation!

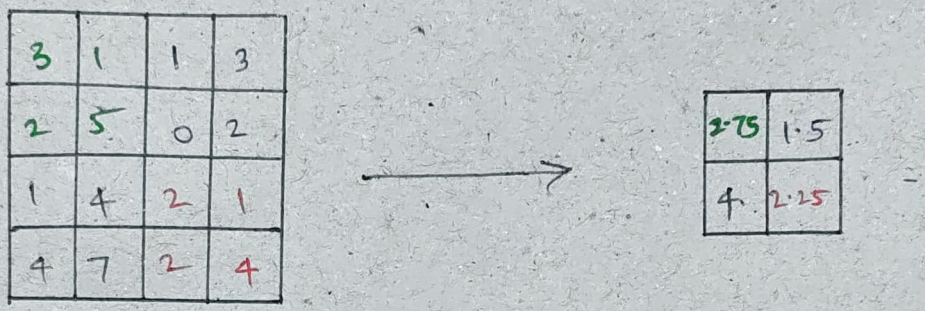
Max Pooling

Single depth slice



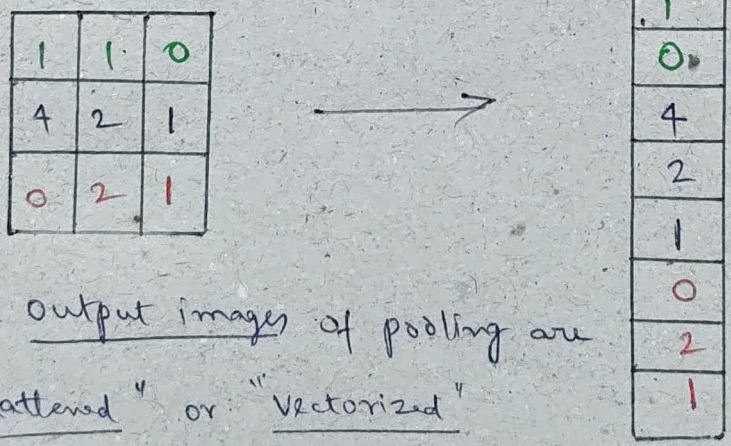
⊙ Max pooling chooses the max element in each block

Average Pooling



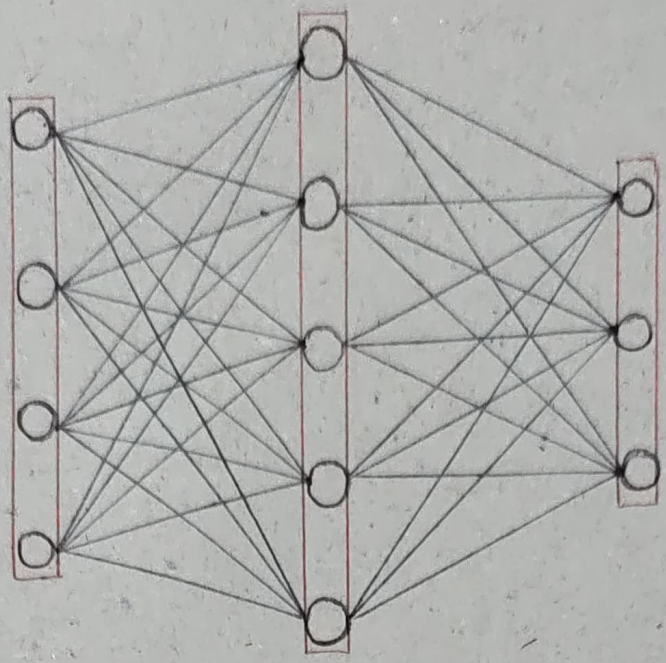
⊙ Average pooling retains the average of the elements in each block.

Flatten Layer



⊙ The output images of pooling are "flattened" or "vectorized"

Flatten Layer



Finally, the flattened images are followed by fully connected layers

Structure of CNN