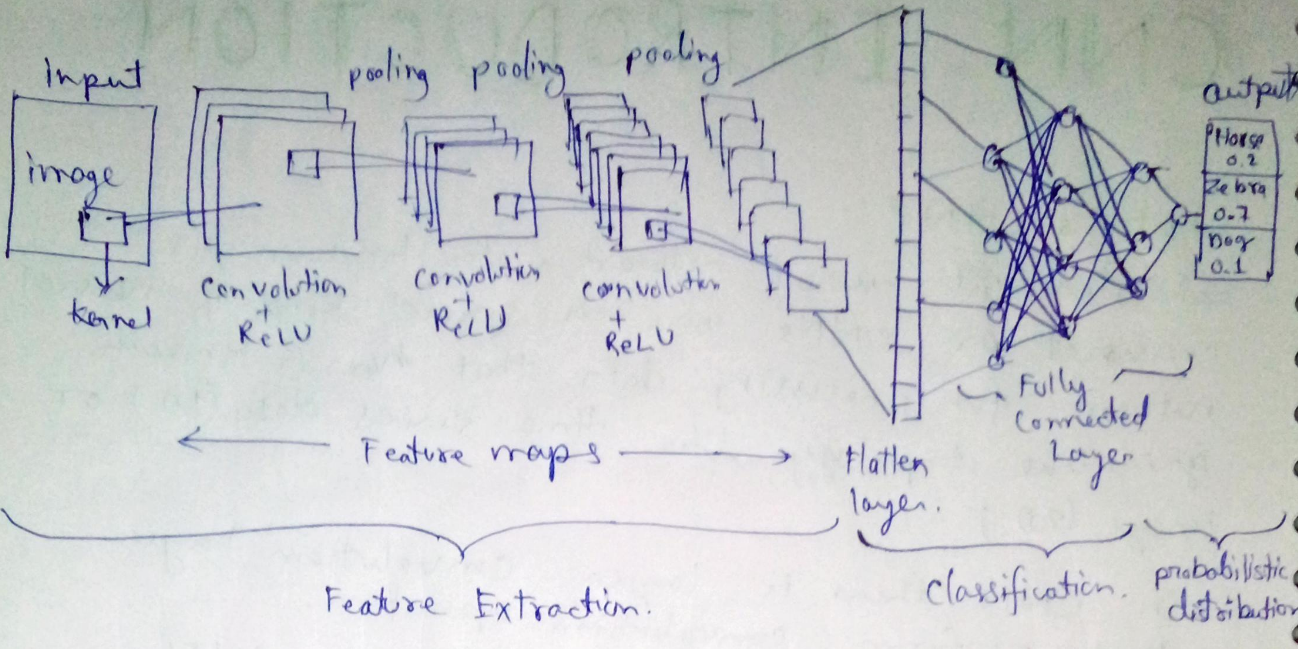
10 Oct 2024



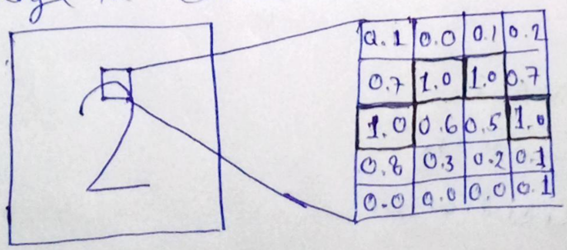
Convolutional Neural Network
(CNN) Architecture.

~~gray~~

gray scale image:

- only one channel.
- values are 0-255 but in normalized 0-1
- if image. 28x28 pixel. in MNIST dataset

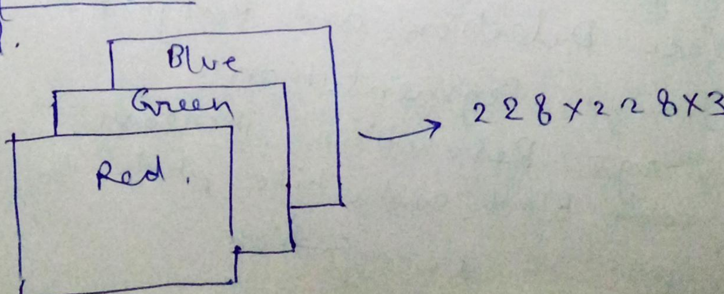
image in converted in 2D-array



0.1	0.0	0.1	0.2
0.7	1.0	1.0	0.7
1.0	0.6	0.5	1.0
0.8	0.3	0.2	0.1
0.0	0.0	0.0	0.1

color image. (RGB):

here 3 channel.



Edge Detection (Convolution Operation)

- edge where the intensity is changes

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 255 & 255 & 255 & 255 & 255 & 255 \\ 255 & 255 & 255 & 255 & 255 & 255 \\ 255 & 255 & 255 & 255 & 255 & 255 \end{bmatrix} * \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 255 & 255 & 255 & 255 \\ 255 & 255 & 255 & 255 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

image (6x6) horizontal filter edge detector (3x3) feature map (4x4)

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} = -1\times0 + -1\times0 + -1\times0 + 0\times0 + 0\times0 + 0\times0 + 0\times1 + 0\times1 + 0\times1 = 0$$

- similarly for other boxes also to fill the feature map.
- based on filter type got different type of edge detector. example.

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

vertical edge detector.

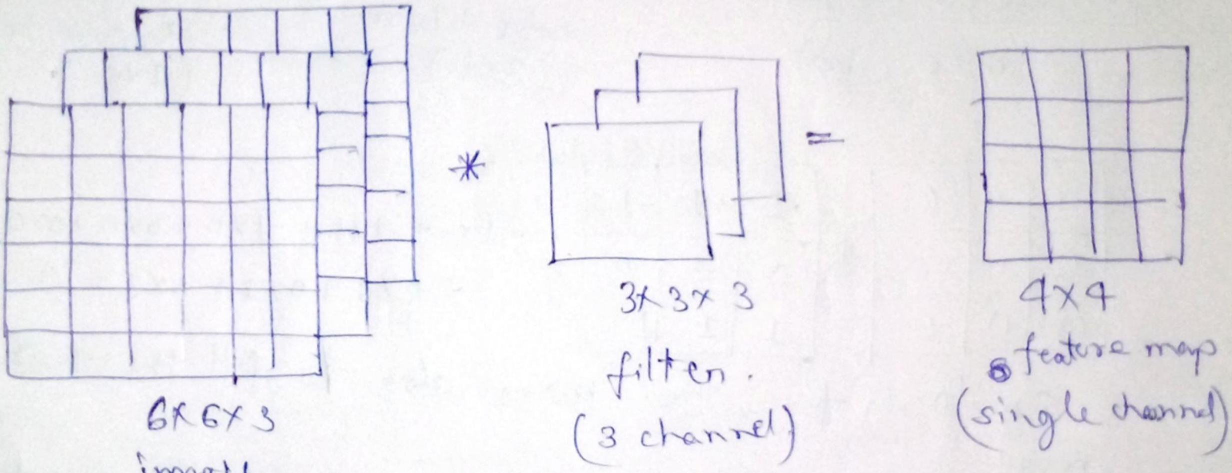
- the values inside the filter are randomly initialize/decided ~~by CNN~~ then ~~then~~ CNN using backpropagation decide what values are present in the filter. here the value we can consider as a ~~as~~ similar in weights in ANN

Ques if the image is (28×28) and filter size is (3×3) then what is the size of feature map.

Answer – (26×26)

$$\underset{(n\times n)}{\text{image}} * \underset{(m\times m)}{\text{filter}} = \underset{\text{feature map}}{(n-m+1)\times(n-m+1)}$$

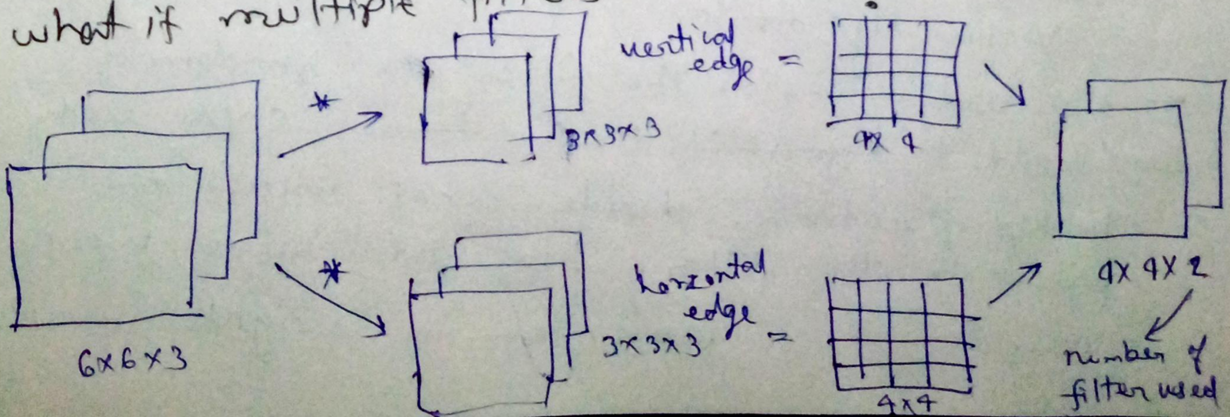
working with color image:



if

$$\underset{\text{input image}}{(m\times m\times c)} * \underset{\text{filter}}{(n\times n\times c)} = \underset{\text{single channel feature map}}{(m-n+1)(m-n+1)}$$

what if multiple filters are used?



PADDING AND STRIDES

why padding is required?

- if our image is small then we are not able to use the multiple convolutional operation (filtering)
- in convolutional operation the corners are less contribute in the operation as compared to centre one values.

Padding operation:

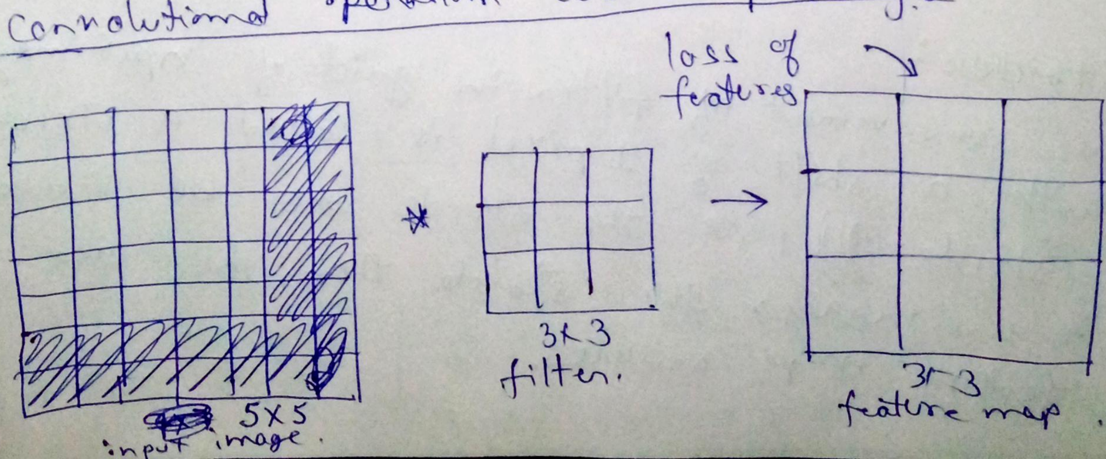
- the above two problem is solve by padding
- in padding keeping all the features of the input image we add column and row in the input image before operation.

7	2	3	3	8
4	5	3	8	4
3	3	2	8	4
2	8	7	2	7
5	4	4	5	4

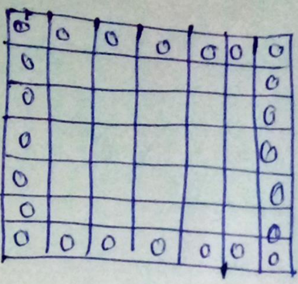
$\longrightarrow$ padding or zero padding

0	0	0	0	0	0	0
0	7	2	3	3	8	0
0	4	5	3	8	4	0
0	3	3	2	8	4	0
0	2	8	7	2	7	0
0	5	4	4	5	4	0
0	0	0	0	0	0	0

Convolutional operation without padding:



convolutional operation with padding:



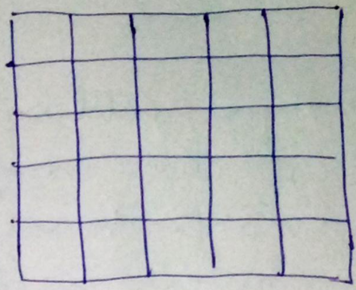
input image with zero padding
(7x7) input image

$*$

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

3X3 filter (kernal)

$\longrightarrow$



5X5 feature map.

* without padding the size of feature map is $(n-f+1) \times (n-f+1)$
* with padding, the size of feature map is $(n+2p-f+1) \times (n+2p-f+1)$

In keras two type of padding operation

- Valid (no padding is apply)
- Same (padding is apply)

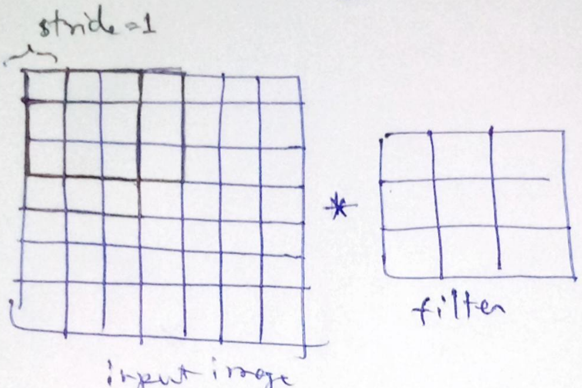
Strides:

- The convolvement of filter in grid of inputimage with a shift of 1 pixel is called a stride (stride = (1,1)), here the stride can be stride = (2,2)
- if we increas the stride then the size of feature map reduces.

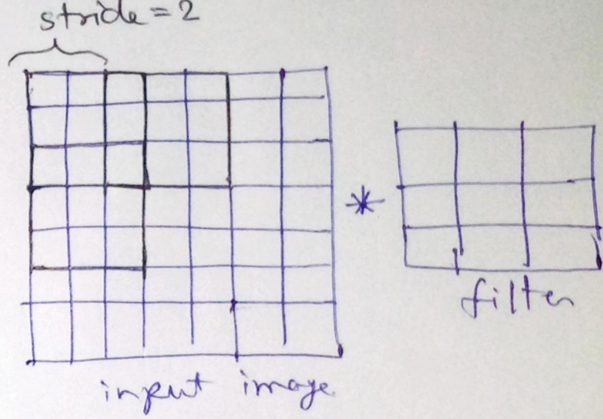
- if we apply also then the size of feature map is $\left(\frac{n-f}{S} + 1\right)$

- if we consider the padding also then size of feature map is

$$\left[\frac{n+2p-f}{2} + 1\right]$$



output feature map size is (5×5)



output feature size is (3×3)

* if the stride is greater than 1 then it is called the strided convolution.

Ques if input image size is 6×7 and filter size is 3×3 with stride = 2 then what is the size of output image (feature map)

using row

$$\left[\frac{n-f}{S} + 1\right] = \left[\frac{6-3}{2} + 1\right] = 1.5+1 = 2$$

using column

$$\left[\frac{n-f}{S} + 1\right] = \left[\frac{7-3}{2} + 1\right] = 2+1 = 3$$

Hence feature map size is (2×3)

why strides are required?

- High level features requirement.
- computing power. (for fast training)