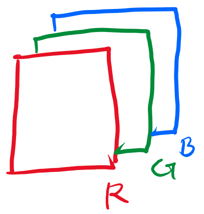
Lecture 14: Convolutional Neural Network

09 March 2024 11:01

Consider a coloured image. This image has three different channels (R, G, B)

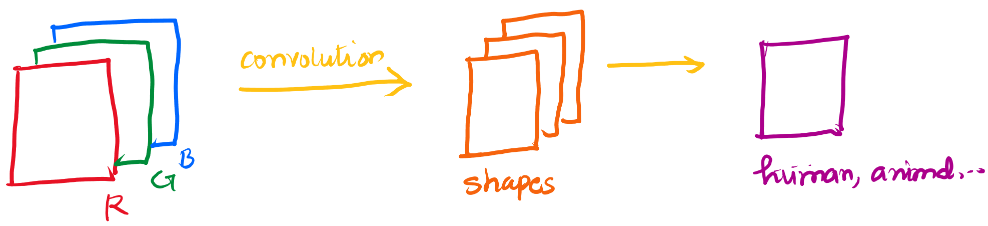


This is given by $1024 \times 1024 \times 3$ matrix.

For a NN using this image as input, the input layer will have $\sim 3\times10^6$ input nodes.
If hidden layer has $\sim 10^3$ nodes, $w \sim 3\times10^9$!

Convolution is used to reduce the image size and retain only useful information, without compromising the information content.

For an image we pass each channel through a convolution layer:



Convolution:

For any two functions $f(x)$, $g(x)$, the convolution is given by:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(t)\, g(x-t)\, dt$$

Refer Wiki page of convolution:
https://en.wikipedia.org/wiki/Convolution

Consider a 5x5 grayscal image (we've nos b/w 0 to 255)
(Black) to (white)

image

1	20	20	20	1
20	1	1	1	20
20	1	1	1	20
1	20	1	20	1
1	1	20	1	1

$n \times n$

$*$

filter

1	0	-1
1	0	-1
1	0	-1

$f \times f$

$=$

filtered image

21	0	-21
38	0	-38
0	0	0

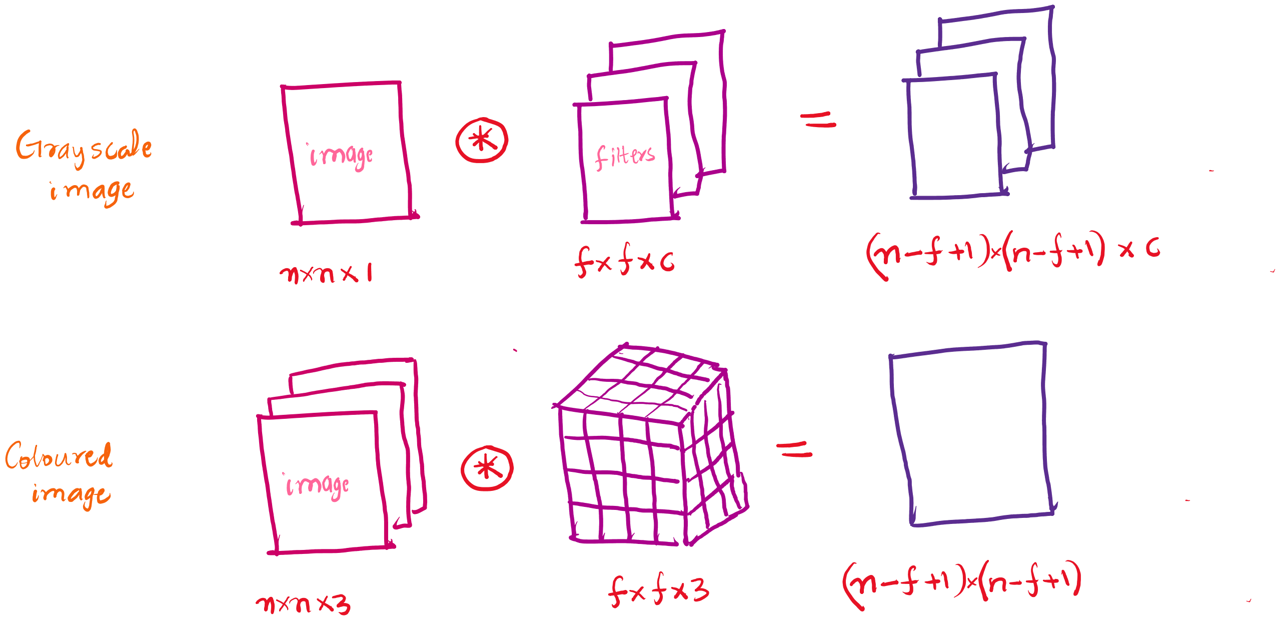
$(n-f+1)(n-f+1)$

This filter acts as a vertical edge detector.

If the filter is: $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$, it acts as a horizontal edge detector.

Filters are used to find shapes or useful information via convolution.

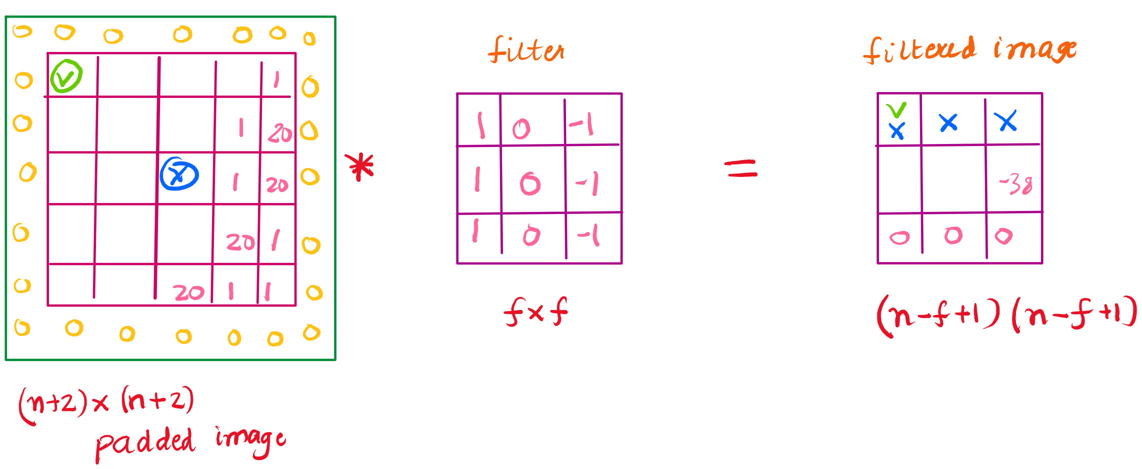
Filters are used to find shapes or useful information via convolution.



How will we find these filters? We need to "train" the weights of filters

Padding

Consider our example of convolution of 5×5 grayscale image with 3×3 filter.



"VALID Convolution" – without padding.

"SAME Convolution" – Padding such that final (filtered image size is same as original, given the filter

eg: [6×6] with padding and 3×3 filter gives [6×6] output

ie, $n' = n + 2p$, $(n' - f + 1) = n$

ie, $n' = n + 2p$, $(n' - f + 1) = n$

$\Rightarrow p = \frac{f-1}{2}$, f is odd

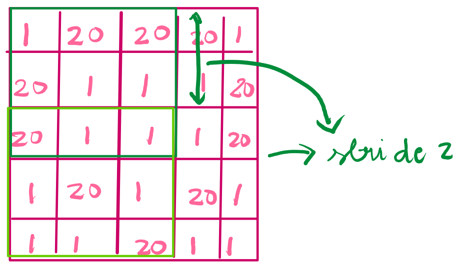
We use padding so that we don't loose information around the borders of the image.

Stride:

If we use a filter with a stride, gives a smaller image. For all that we discussed till now, stride is 1.

$$(5 \times 5) \circledast (3 \times 3) = (2 \times 2)$$

stride 2.

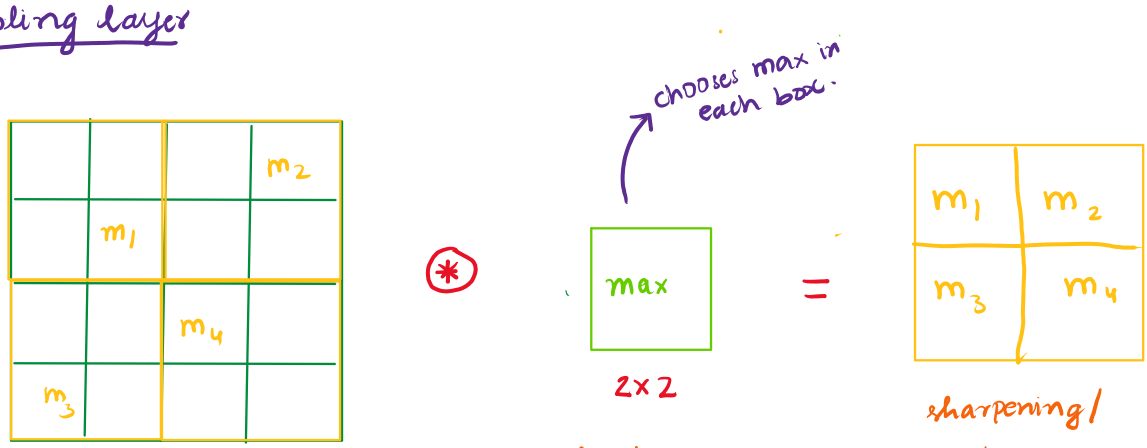


Stride ignores the part of the image where convolution can't be entirely within image.

$$\underset{n \times n}{\text{Image}} \circledast \underset{f \times f}{\text{filter}} = \text{filtered image } \left\lfloor \frac{n-f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n-f}{s} + 1 \right\rfloor$$

stride = s.

Max pooling layer



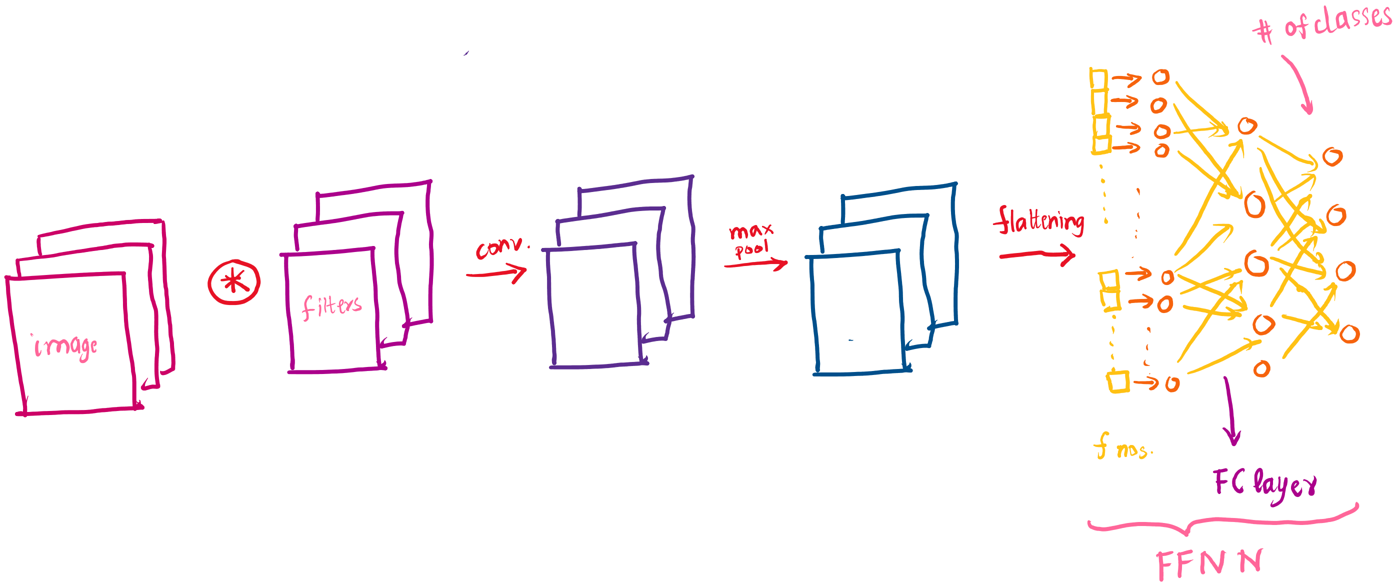
4×4 $\circledast$ 2×2 =

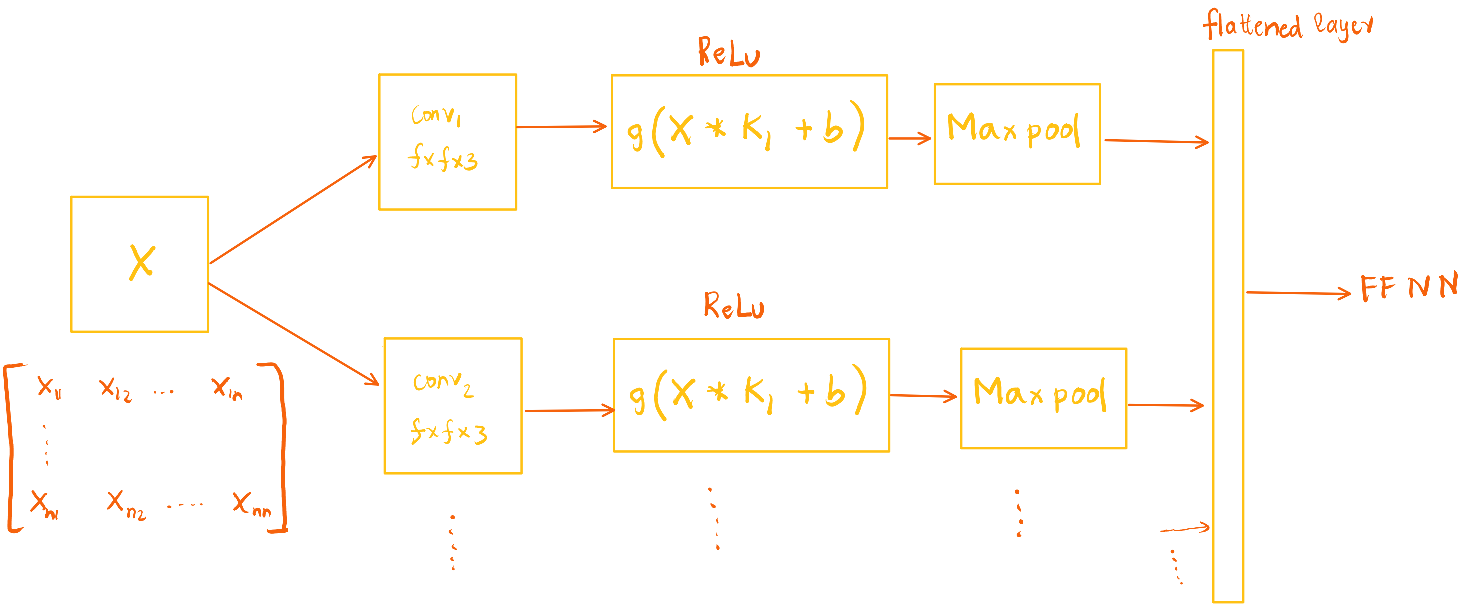
stride = 2 → typically strides = size of the filter.

sharpening / contrast increase

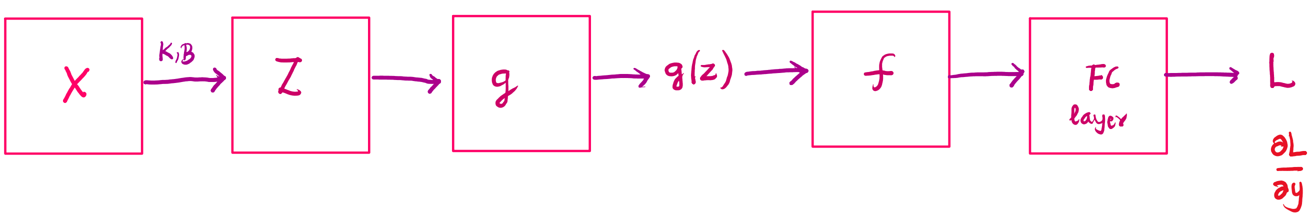
Other alternatives: Average pooling.

> Convolutional layers extract features from images.
> How will we use these features to classify?
> By using a fully connected layer (FC layer) : FFNN.





Backpropagation in CNN:



$$K^{(t+1)} \leftarrow K^{(t)} - \eta \left.\frac{\partial L}{\partial K}\right|_t$$

$$B^{(t+1)} \leftarrow B^{(t)} - \eta \left.\frac{\partial L}{\partial B}\right|_t$$

Let $X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$

$K = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}$ $\quad\therefore\quad X \circledast K + B = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \quad \therefore \quad X \circledast K + B = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

such that $Z_{11} = X_{11}K_{11} + X_{12}K_{12} + X_{21}K_{21} + X_{22}K_{22} + B$

$\vdots$

already known via back propagation

Consider $$\frac{\partial L}{\partial K_{mn}} = \sum_{i,j} \frac{\partial L}{\partial Z_{ij}} \cdot \frac{\partial Z_{ij}}{\partial K_{mn}}$$

Eg: $$\frac{\partial L}{\partial K_{11}} = \frac{\partial L}{\partial Z_{11}} \cdot \underbrace{\frac{\partial Z_{11}}{\partial K_{11}}}_{x_{11}} + \frac{\partial L}{\partial Z_{12}} \cdot \underbrace{\frac{\partial Z_{12}}{\partial K_{11}}}_{x_{12}} + \frac{\partial L}{\partial Z_{21}} \underbrace{\frac{\partial Z_{21}}{\partial K_{11}}}_{x_{21}} + \frac{\partial L}{\partial Z_{22}} \underbrace{\frac{\partial Z_{22}}{\partial K_{11}}}_{x_{22}}$$ (See the expression of Z_{11} above)

Verify that $\frac{\partial L}{\partial K} = X \circledast \frac{\partial L}{\partial Z}$.

Now, $$\frac{\partial L}{\partial B} = \sum_{i,j} \frac{\partial L}{\partial Z_{i,j}}$$

$$\frac{\partial L}{\partial X_{mn}} = \sum_{i,j} \frac{\partial L}{\partial Z_{i,j}} \cdot \frac{\partial Z_{ij}}{\partial X_{mn}}$$

Verify that $\frac{\partial L}{\partial X} = \left(\text{zero padded } \frac{\partial L}{\partial Z}\right) \circledast \tilde{K}$

where, $\tilde{K} = K$ rotated by 180°
$$= \begin{bmatrix} K_{22} & K_{21} \\ K_{12} & K_{11} \end{bmatrix}$$