**Convolutional Neural Network(CNN)**

ANN/FCs (Fully connected layers) is unable to process image and videos so CNN is introduced which process image and videos generally used for image detection, classification, and other image related works.

In FCs, one input as a whole entity passes through all the activation units whereas Conv layers work on the principle of using a floating window that takes into account a specific number of pixels at a time. Therefore, in terms of computation time or memory usage FCs cannot be the first choice.

Another downside of FCs is their same approach of using the whole input that might not work well for all kinds of images. Or, we can say that FC's become dependent on the shape of the train images which might not be a good thing for the overall model.

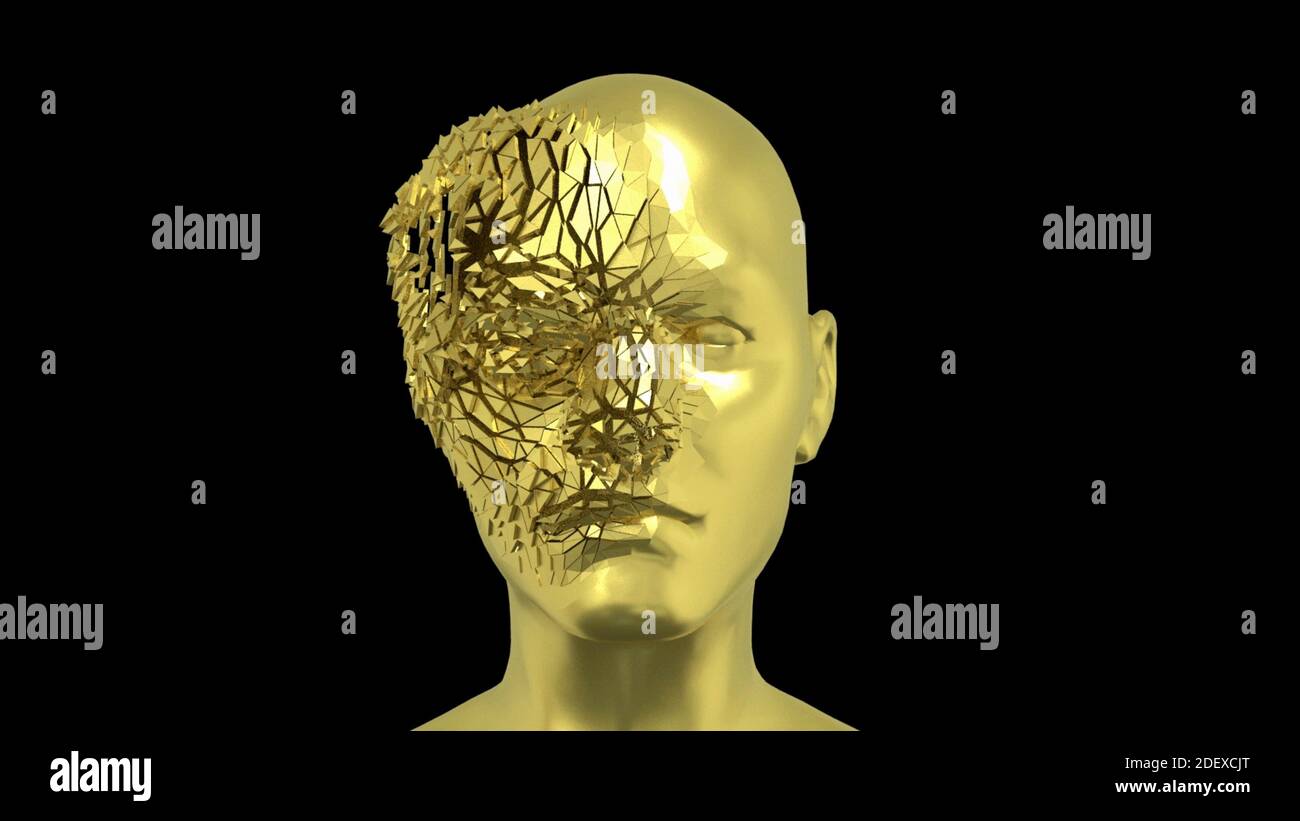
Another problem with FCs is that they have a larger number of weights or parameters thus highly prone to overfitting whereas a single convolution operation reduces the number of parameters quite significantly which makes it less prone to overfitting.

Convolutional layers are better suited for image data because they are able to exploit the spatial structure of the image. In a convolutional layer, a set of filters are applied to the input data, and each filter is able to detect a specific feature in the image. These features can be edges, textures, or patterns. Because the filters are only looking for specific features, they are able to learn representations that are more robust and generalizable than a fully connected layer. Additionally, convolutional layers have the ability to reduce the dimensionality of the input data by using pooling layers, which makes the network more efficient and reduces the risk of overfitting.

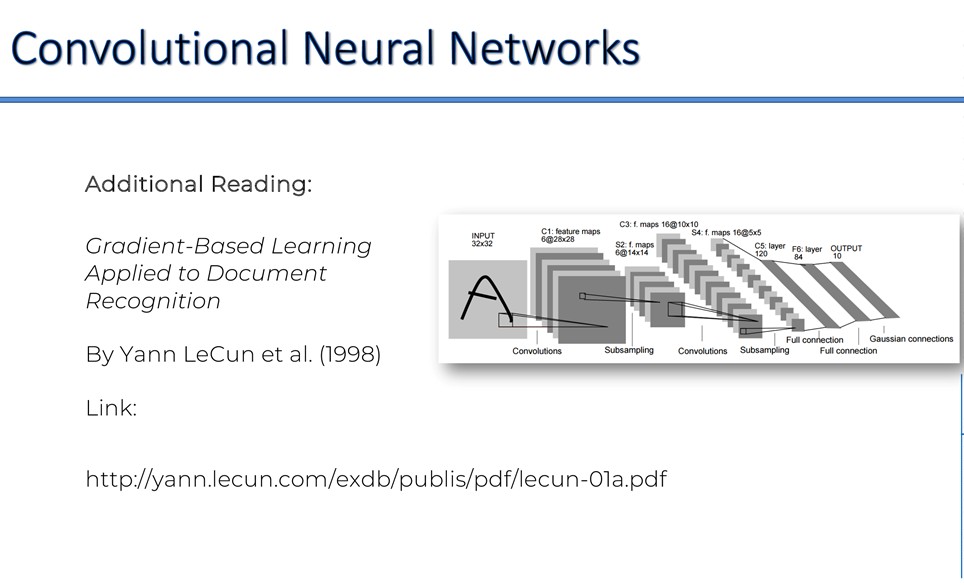
**Disadvantage of CNN**

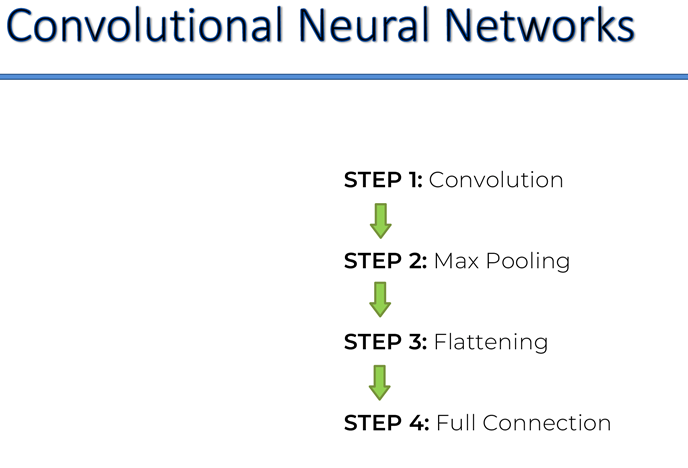
But, there is one major downside of the Conv layer is that it really didn’t preserve the relationship among the features which means an image with scattered human face components can be wrongly labeled as a human face.

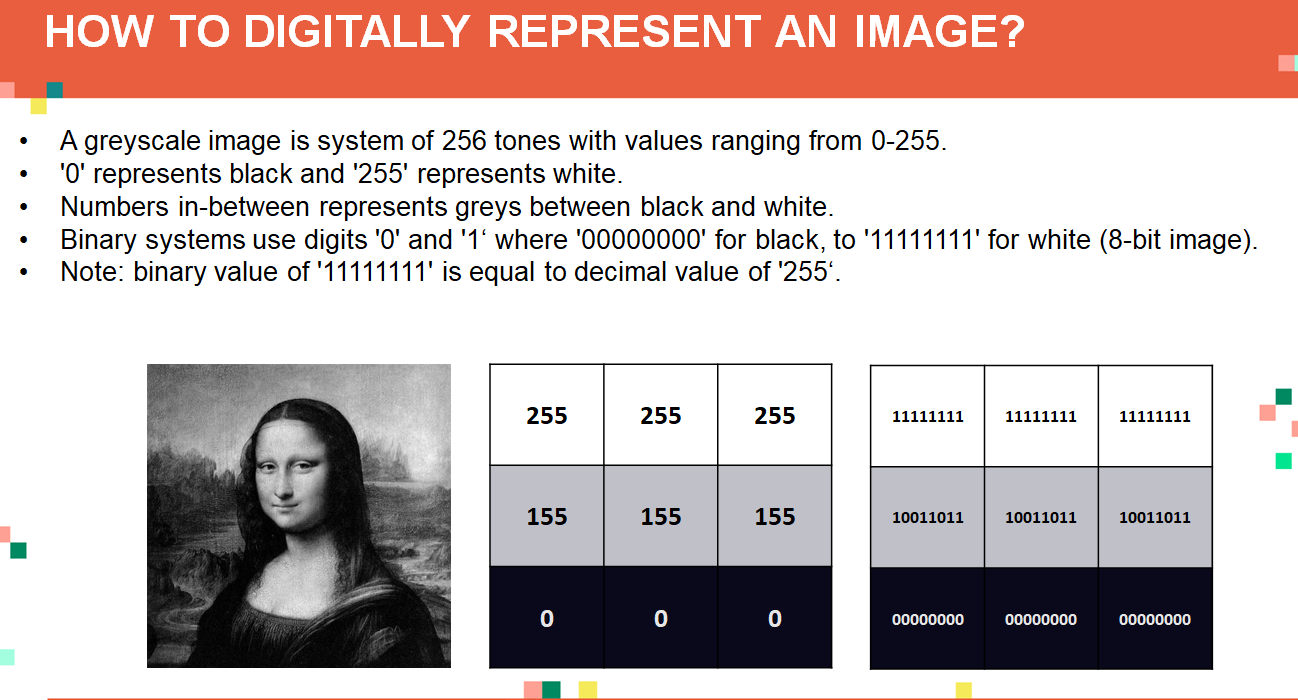
Eg.



**Yann LeCun**, director of [Facebook’s AI Research Group](https://ai.facebook.com/research/), is the pioneer of convolutional neural networks. He built the first convolutional neural network called LeNet in 1988. LeNet was used for character recognition tasks like reading zip codes and digits.

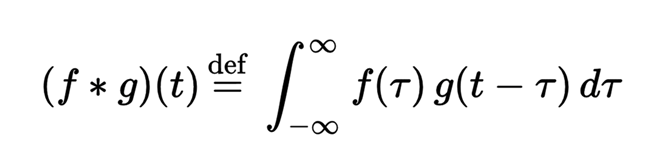




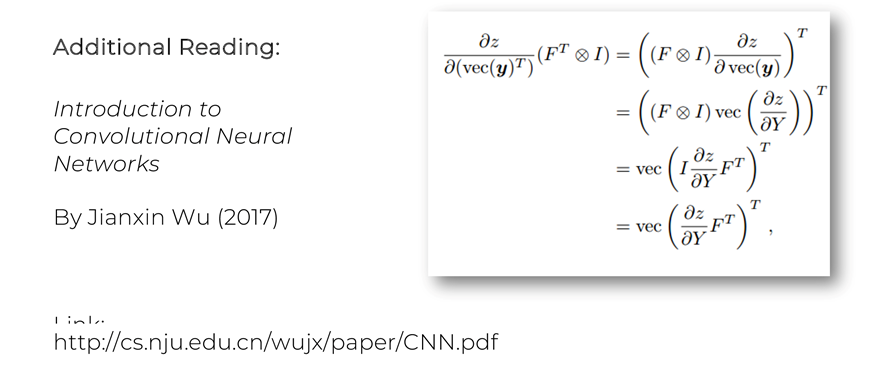


**Convolution**

This is a formula of convolution

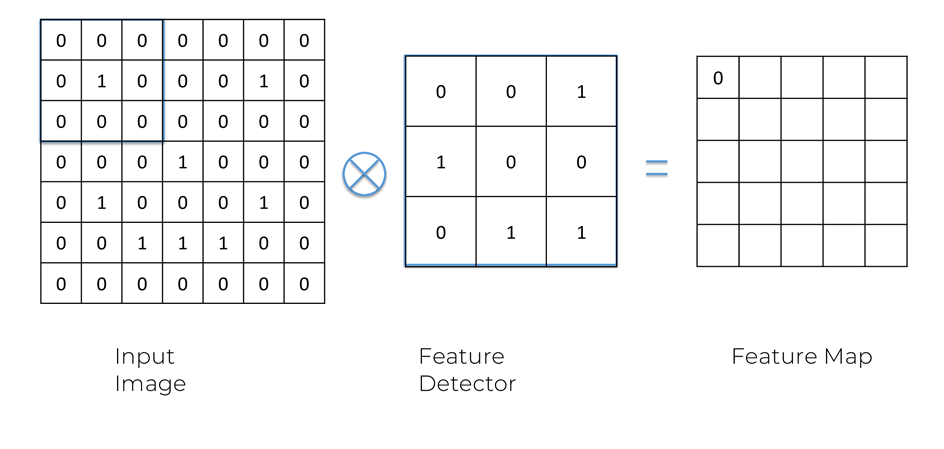


To read a best research paper for convolution network is



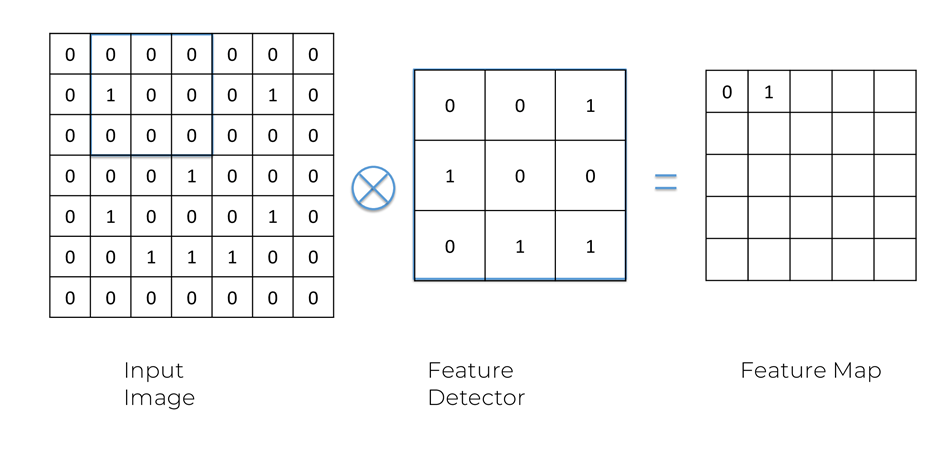
He has other materials also …remove two slashes data in url and read.

Convolution follows following rules to filter out the data as



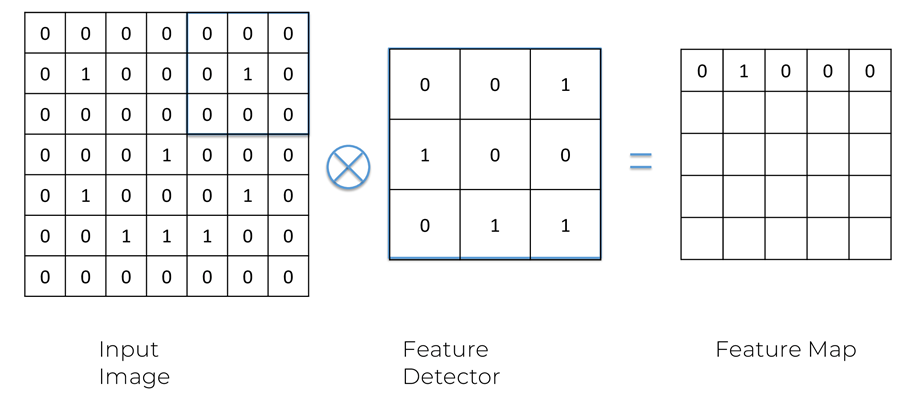
We all know that every Image can be represented in form of 0 or 1. So, we have to define one feature detector which will contain 3\*3(commonly preferred) or any square form of 0s and 1s which is contained by every image holding that feature.

By deciding that feature detector we takes the images 3\*3 row from top left and get that how many 1s is matching there. We count the matching 1s and write it in first row first column of output feature map.



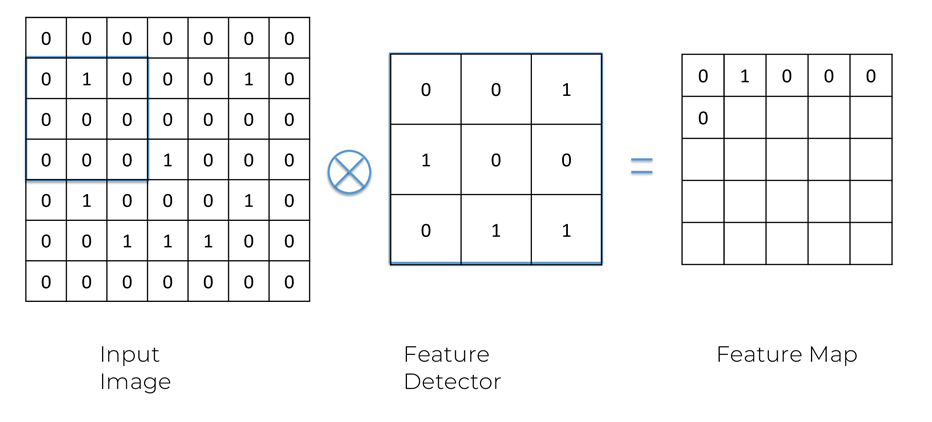
Here one 1 matches so output is 1.

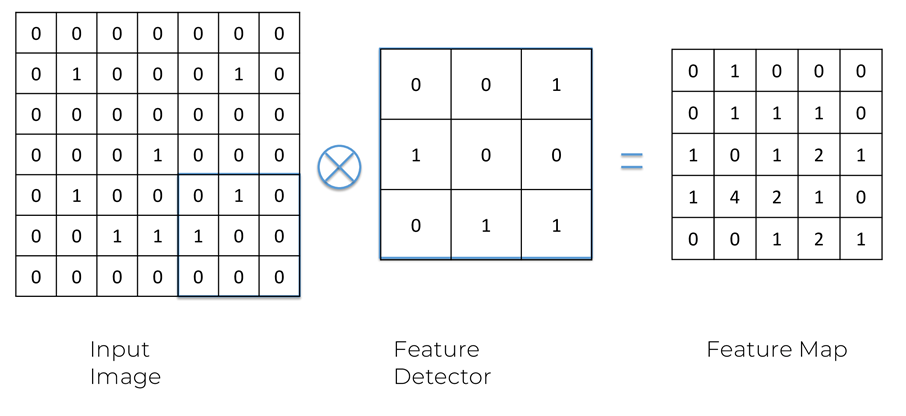
Like this we go on doing this till the last row last column.



One row completed.

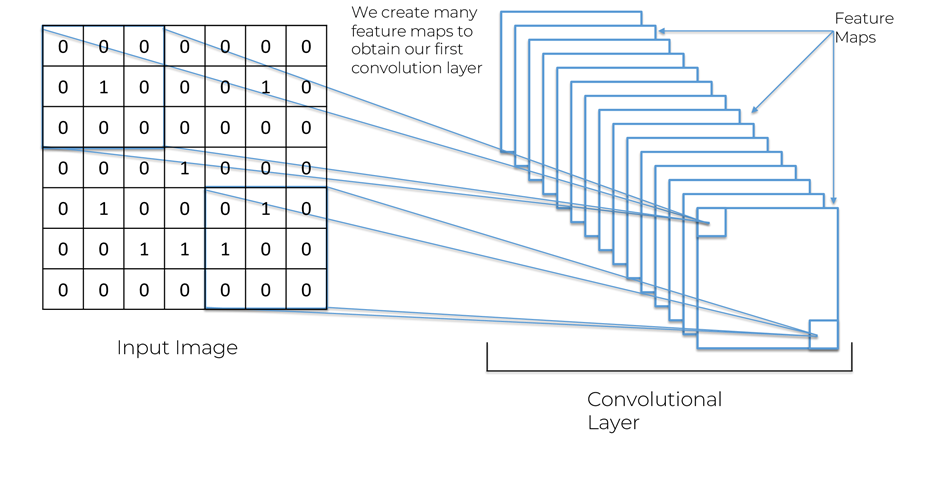
Now second row targeted.



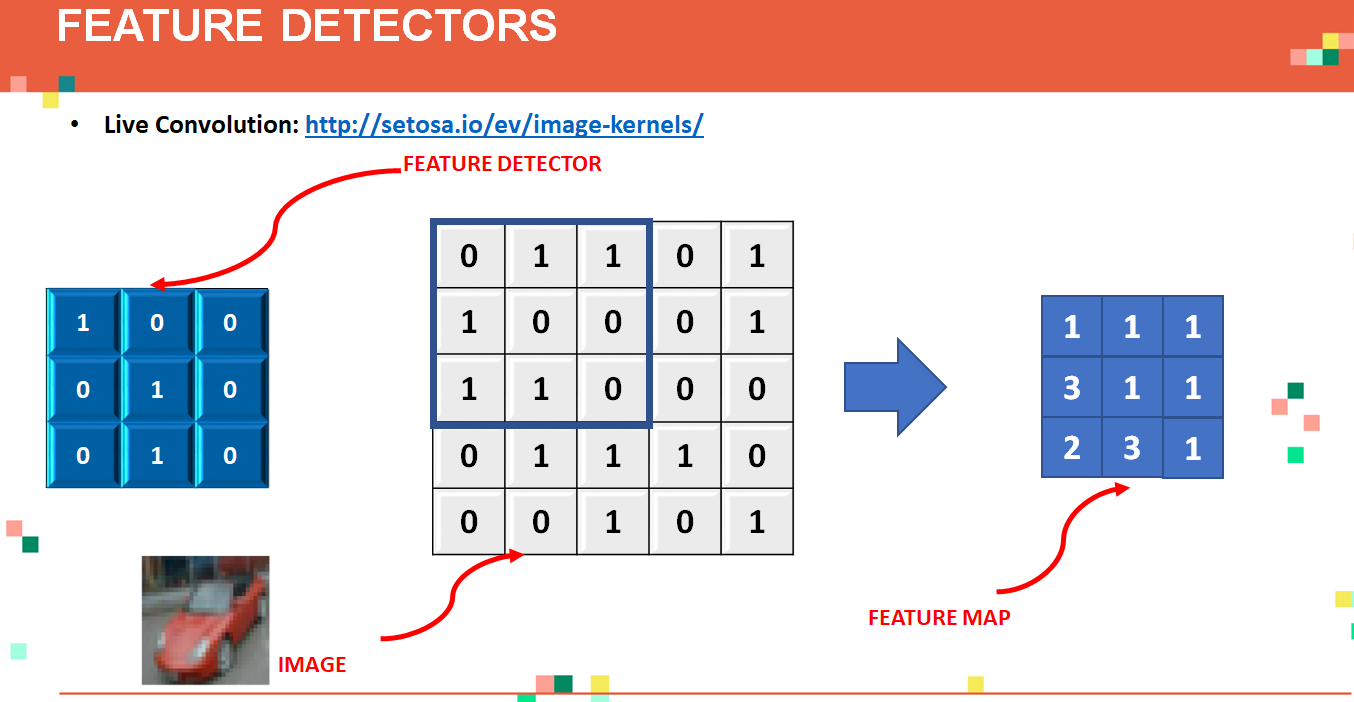


This is the resultant matrix of feature map.

Similarly if you have several feature detector and try to extract several feature maps then like this type of image is created.

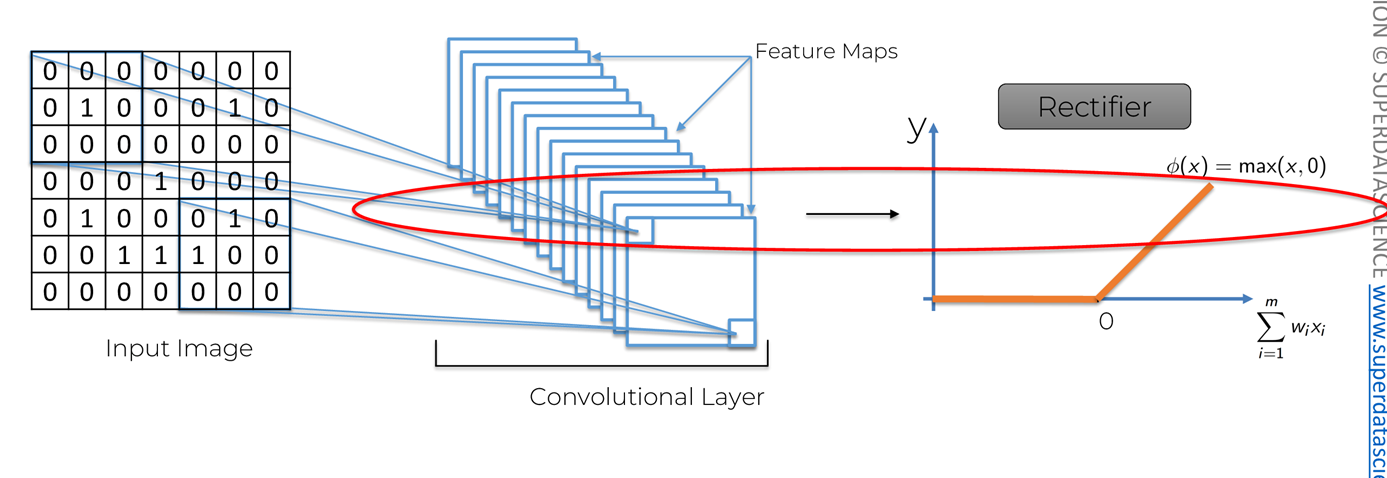


**Another example of convolutional layer**

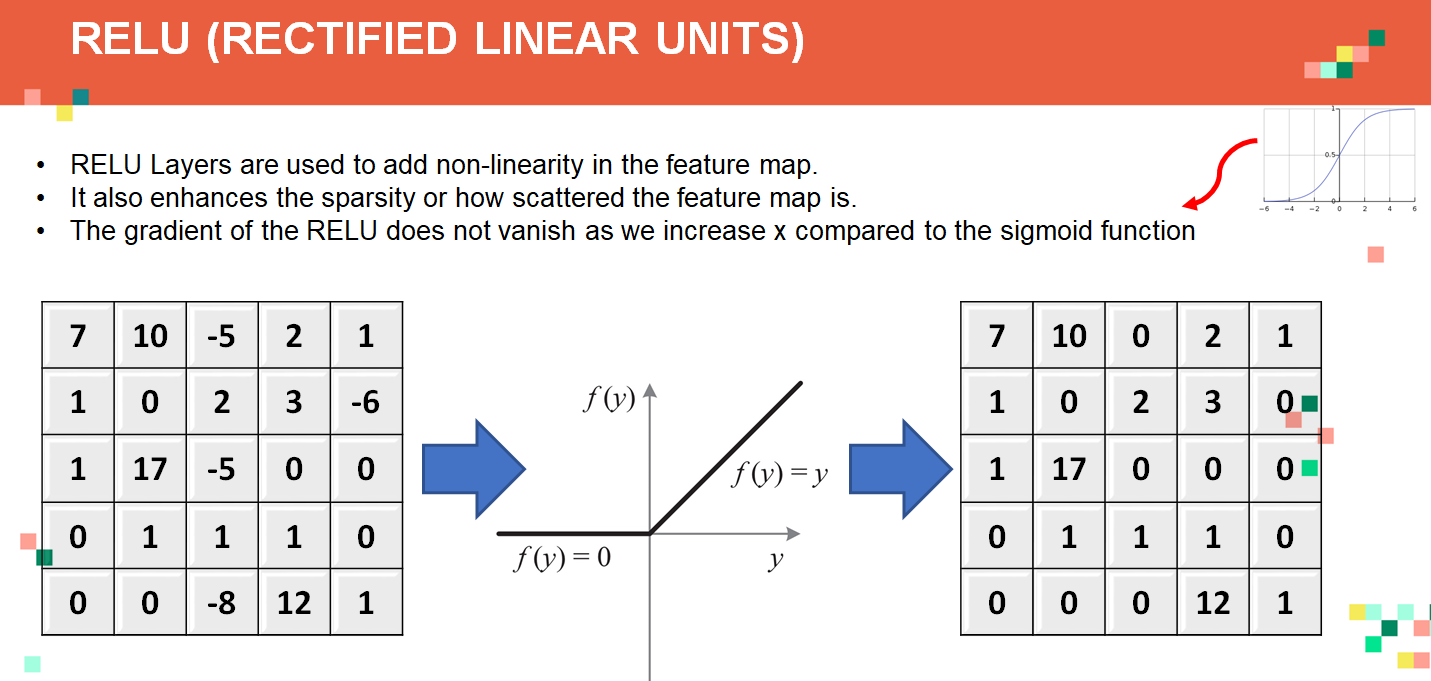
****

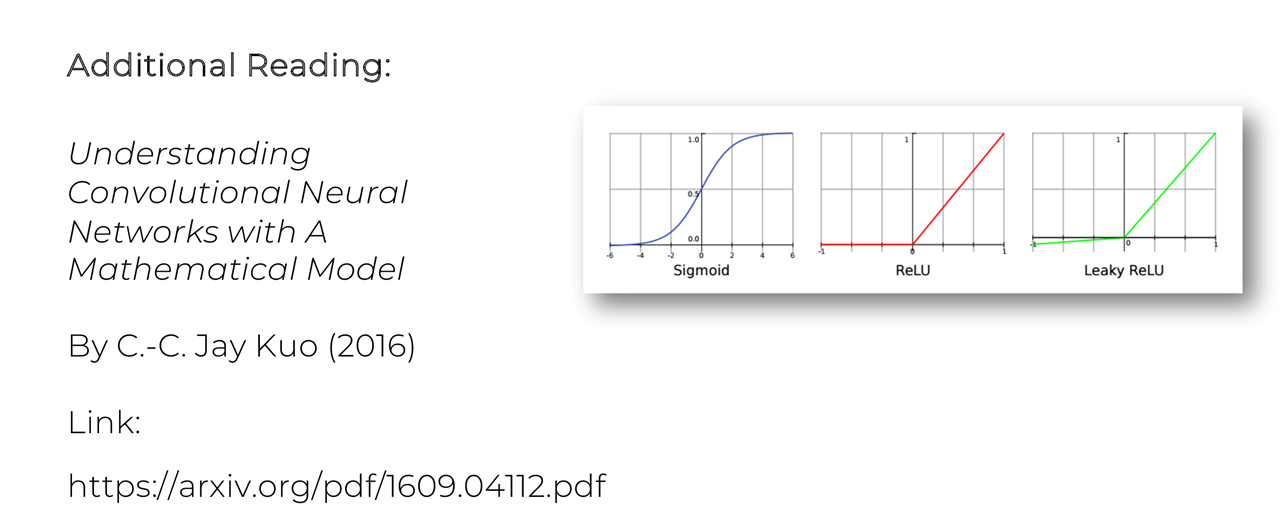
Here we can check how to sharpen any image , blur any image ,and many more.

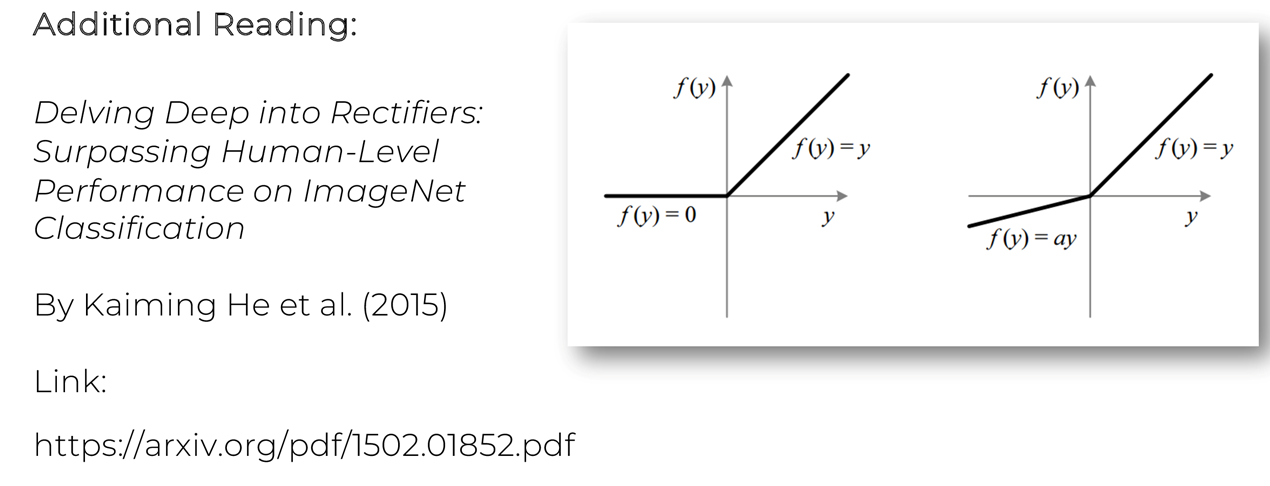
**ReLU(Rectifier linear Units)**



This is an additional step on the top of our Convolution. We apply rectifier function on the convolution layer got. The ReLU work is to break the linearity in the image. The rectifier function is applied to increase non-linearity in the image.

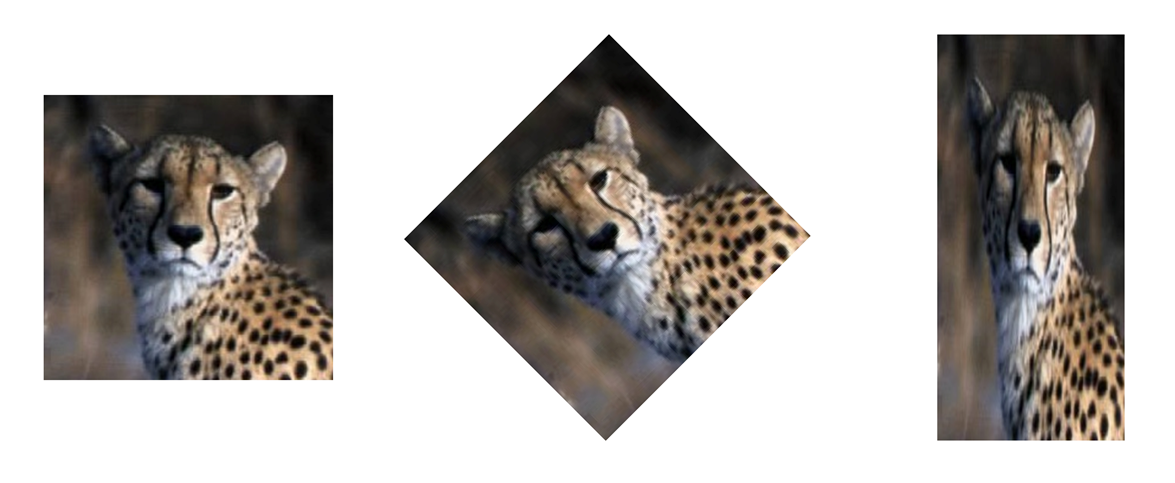






**Max-POOLING (Down-sampling)**

Suppose our image is little tilted or stretched So it will be hard to find out the feature , In that case pooling helps as it extract that feature if present in the image and categorize the image on the basis of that.



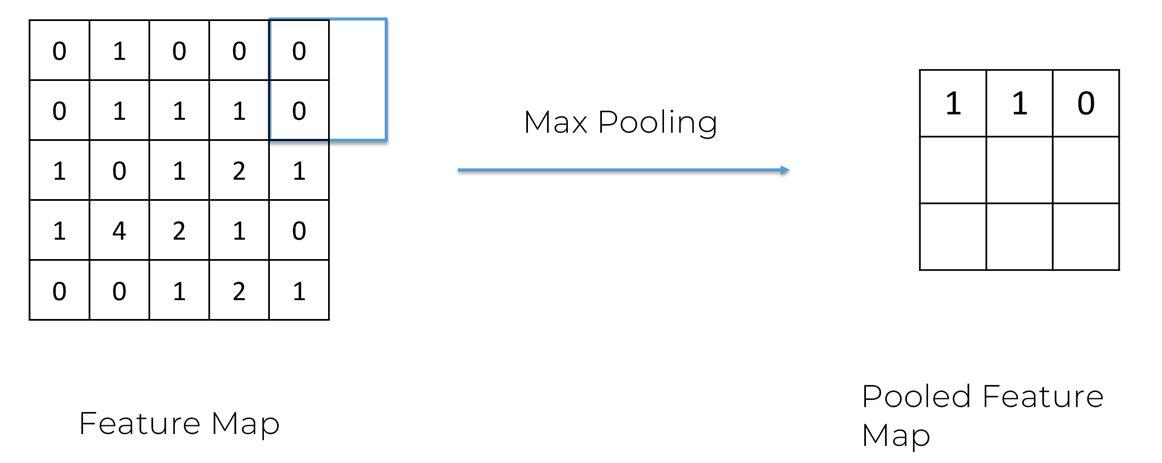
So pooling is done in this way

For every feature Map we select 2 \*2 square and for max-pooling we take out the max value in the box.

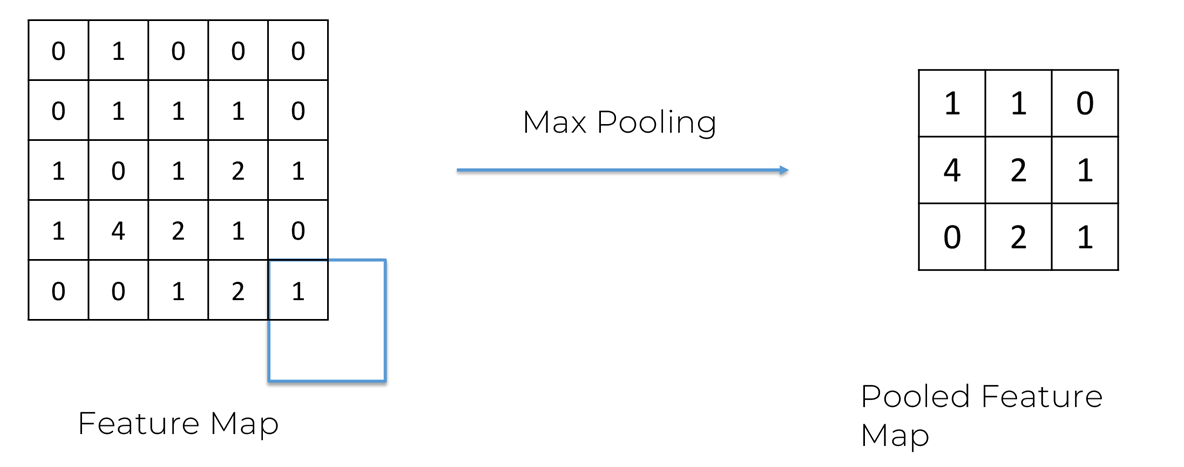
Then in next step we leave all the values covered previously unlike convolution.

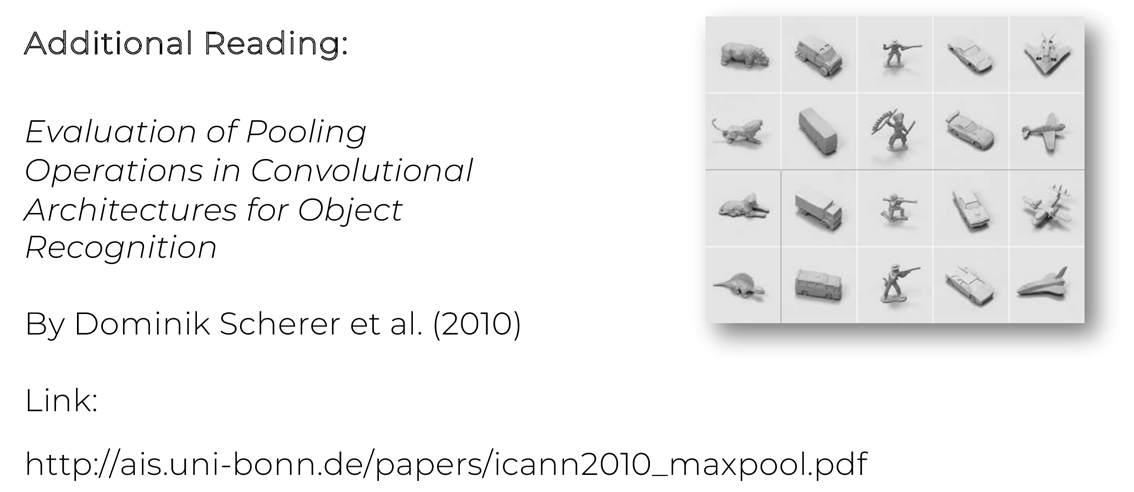




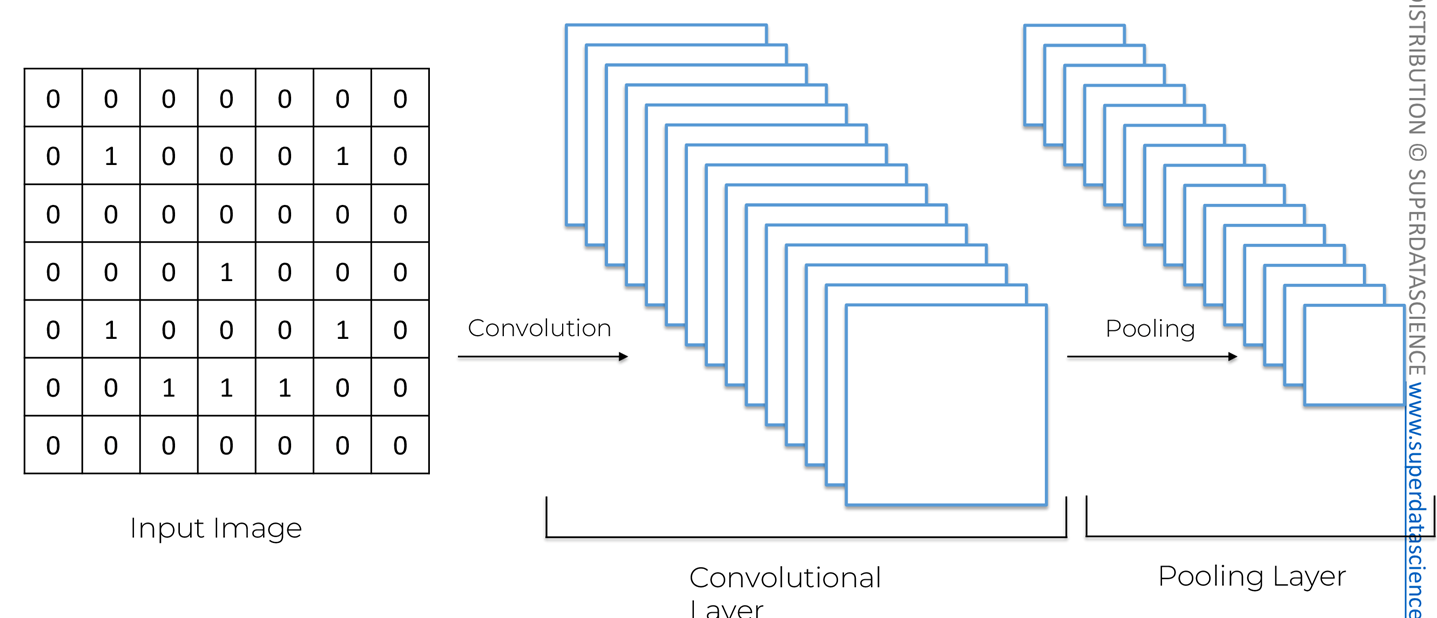






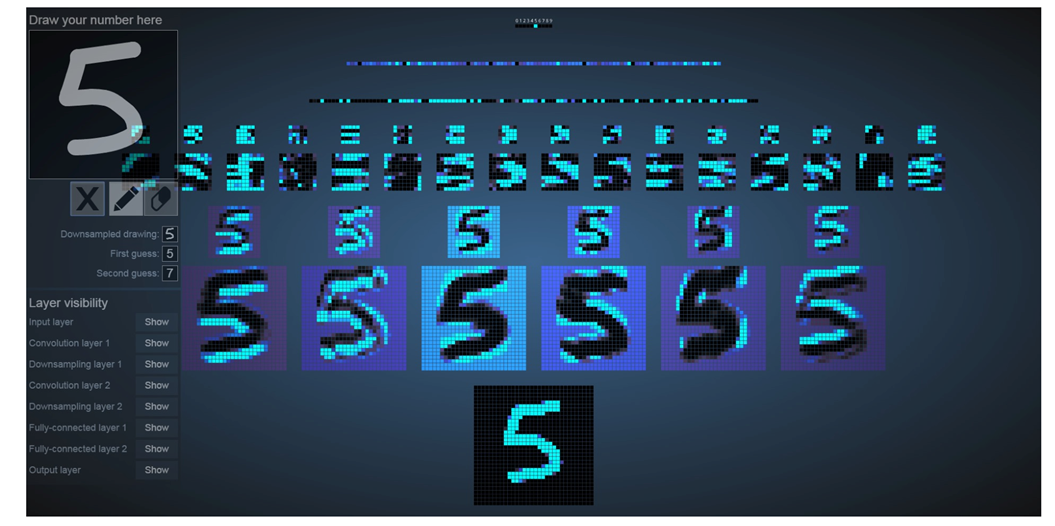


Use of max-pooling is reduction in the matrix.

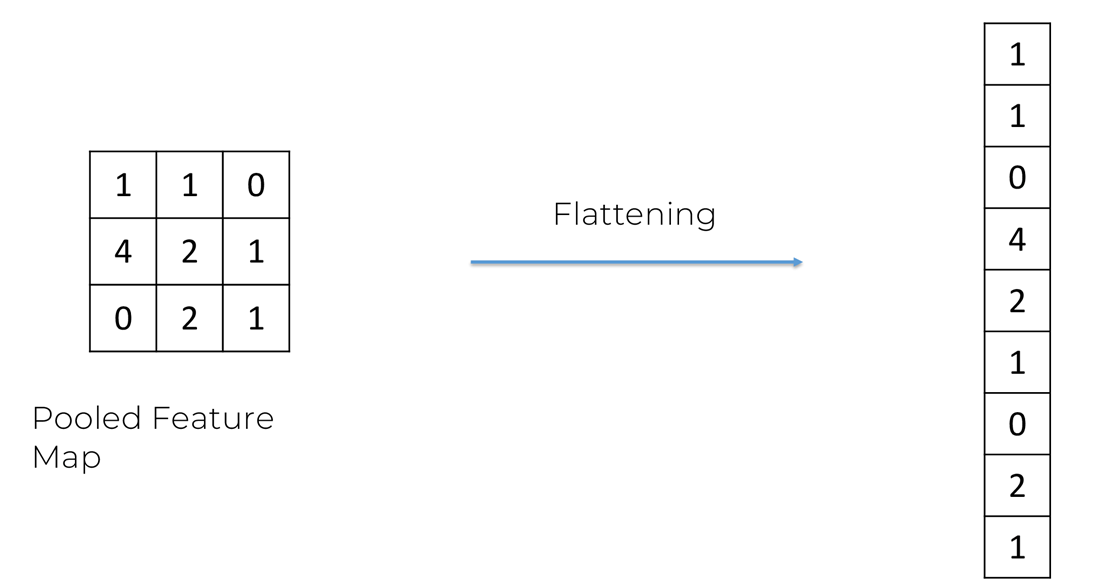


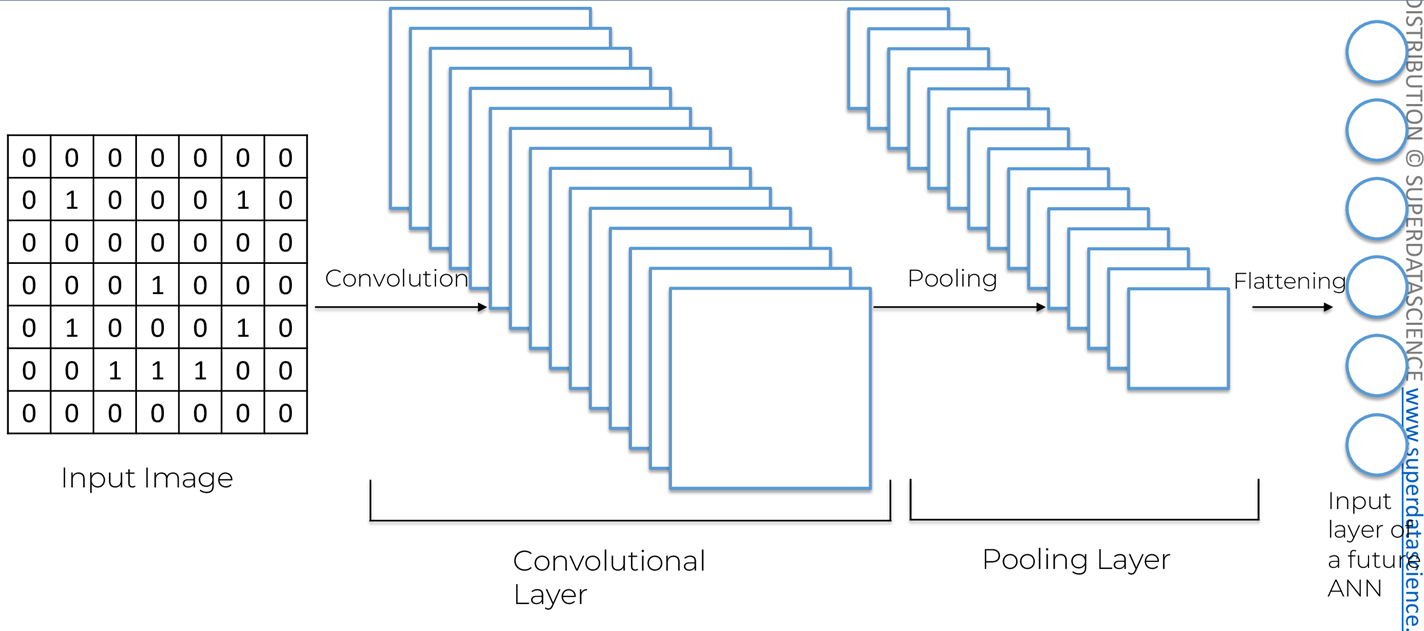
For Fun to recognize the number implements convolution and pooling check out website

scs.ryerson.ca/~aharley/vis/conv/flat.html



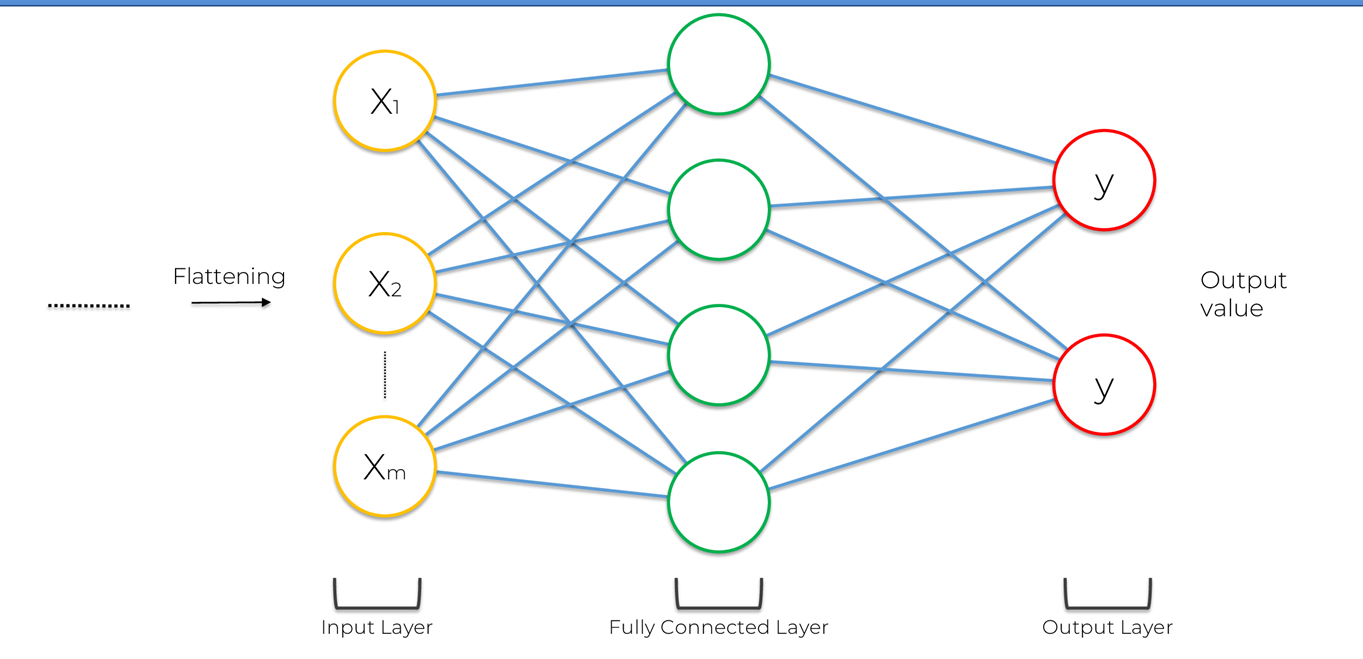
**Flattening**

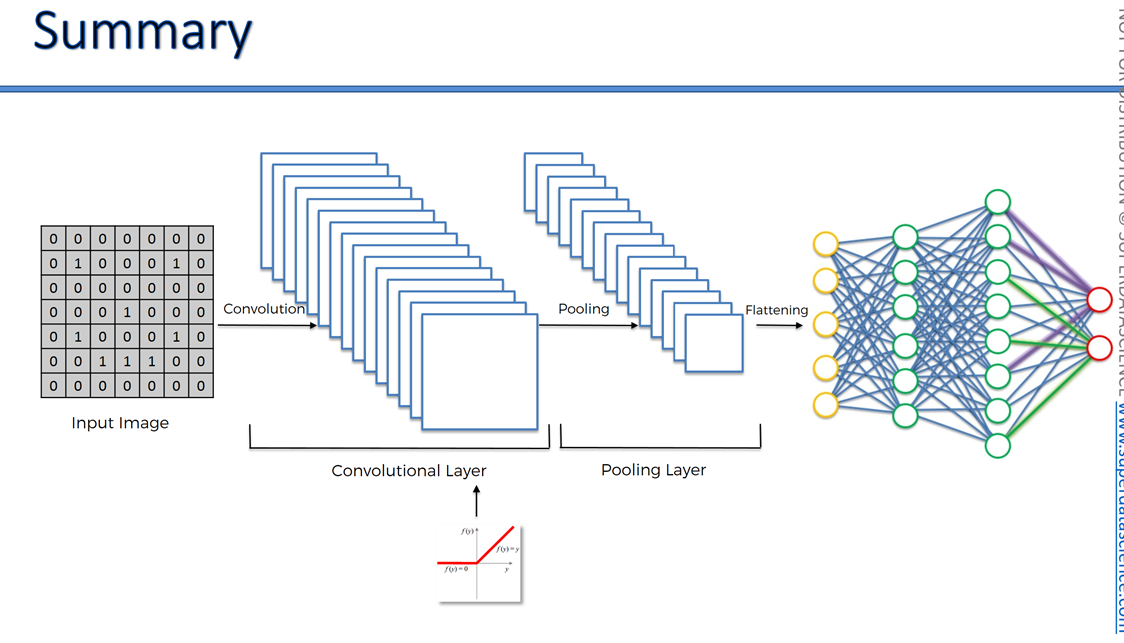
****

****

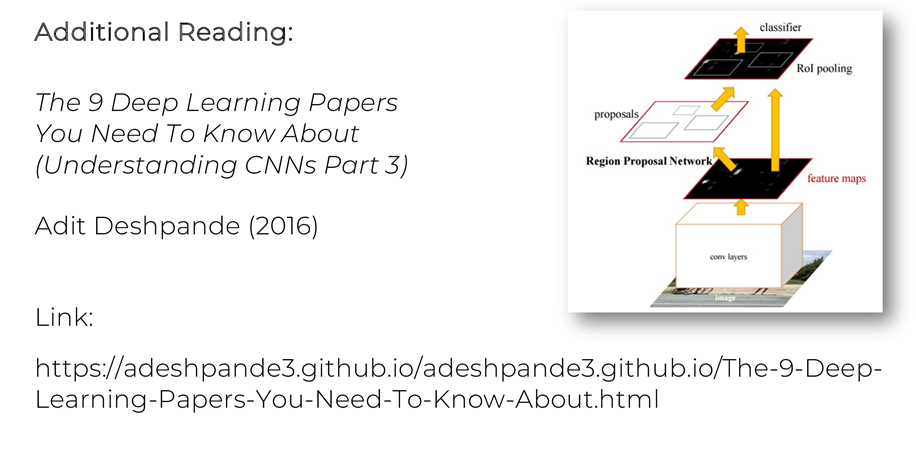
**Full-Connection**

After the flattening we generally transfer the pixels like in ANN but here the hidden layers will be fully connected with input layers.

****

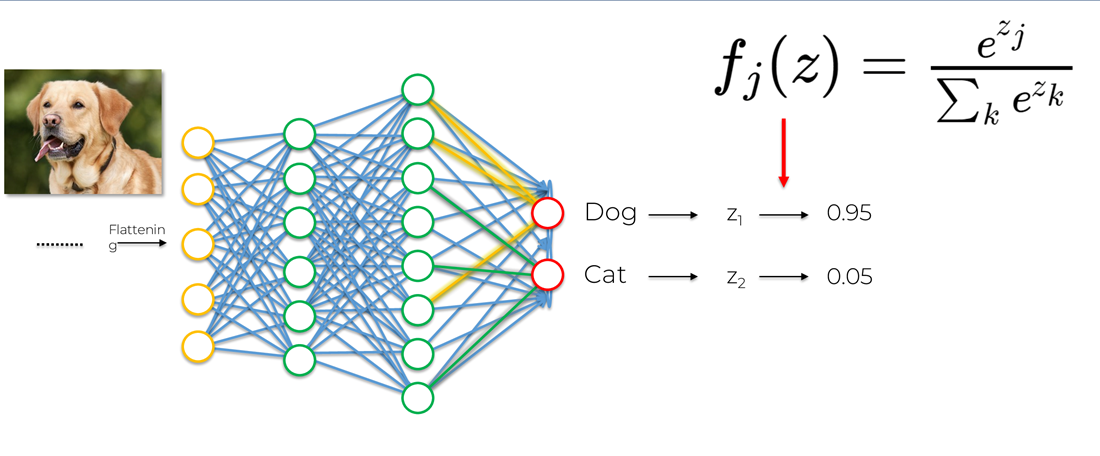
****

**Sigmoid(2 output) / softmax(more than 2 output) used here**

****

**Softmax(used for more than 2 output)**

So in Soft max function it applies a function which makes the predictions sum to 1.



Suppose if results here sum would not be 1 then it would have created a lot of confusions. 0.95 and 0.05.

To do this softmax function is used as shown above.