Convolutional Neural network

* A convolutional neural network determines object detection, face recognition, image recognition and image classification.
* It classifies images into different categories.
* It takes pixel arrays as input into height, weight and length of RBG values (red, blue, green) components.
* A deep neural network usually consists of convolution layers. The training set passes through a series of these layers with filters known as kernels.

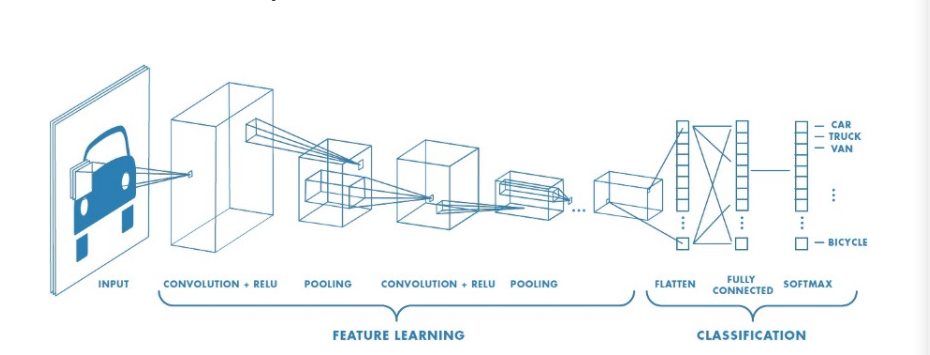


Figure 1: The layers of convolution

* Convolution is the foremost input layer for the purpose of feature extraction.
* It makes sure that the data is preserved while it learns the images in small squares of input data.
  1. The mathematical operation:
     + The volume of the image matrix’ s dimensions are given by: Height\* width\* length
     + A filter is given by fh \* fw \* length
     + Output is given by (h - fh +1) \* (w - fw +1) \* 1
  2. For example:

Assume a 5\*5 Image matrix and a 3\*3 Filter matrix

When the convolution of 5\*5 matrix is multiplied with 3\*3 filter it is called a feature map.

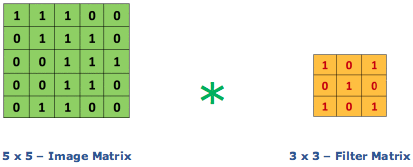


Figure 2: The multiplication of Input layer and filter layer

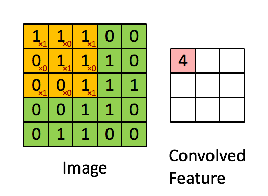
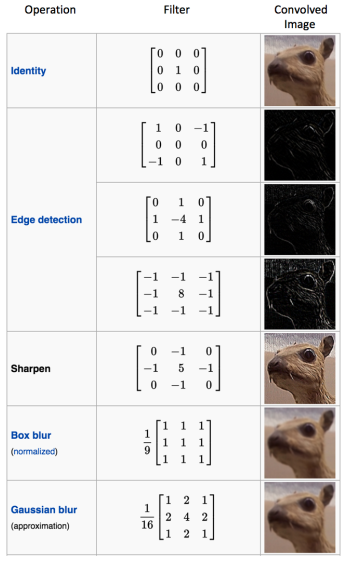


Figure 3: The output of the above multiplication

* Each layer performs a certain operation:
  1. Edge detection
  2. Blur
  3. Sharpen



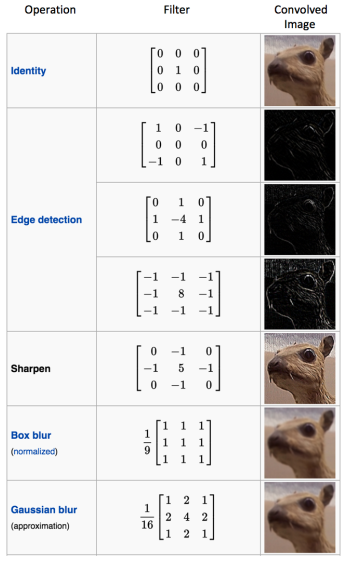


Figure 4: The various filters and their purpose

* **Strides**: Stride is the number of pixels shifts over the input matrix.
  1. When stride is 1 then move the filters to 1 pixel at a time.
  2. And we increment as the stride increments.

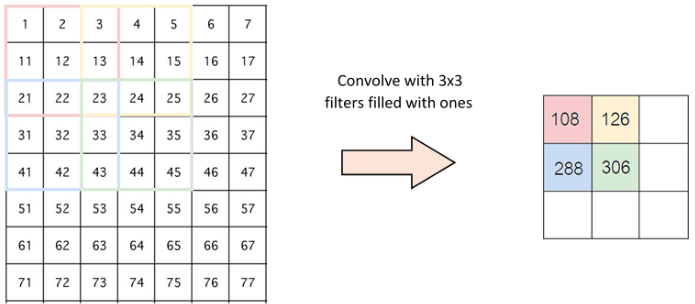
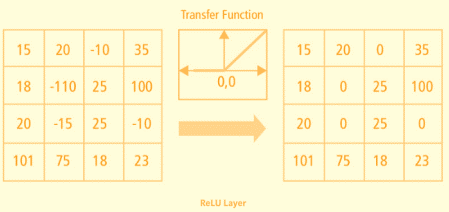


Figure 5: Stride of 2 pixels

* **Padding**: When the filter does not fit the input image perfectly, then there are two solutions
  1. Zero-padding: Pad the image with zeroes
  2. Valid Padding: Drop the parts of the image which does not fit the image. Thereafter keeping only, the valid portions of the image.
* **Non-Linearity (ReLU)**: Rectified Linear Unit for a non-linear operation.
  1. ReLU is used to introduce non-linearity in a convolution network. Therefore, making sure that there are non-negative linear values.
  2. This method is the most efficient among tanh and sigmoid functions.



* **Pooling Layer**: Reduces the number of parameters when the image is too large.
  + Spatial pooling or subsampling or down sampling reduce the dimensionality of each map.
    1. Max pooling: Takes the largest elements in a feature map.

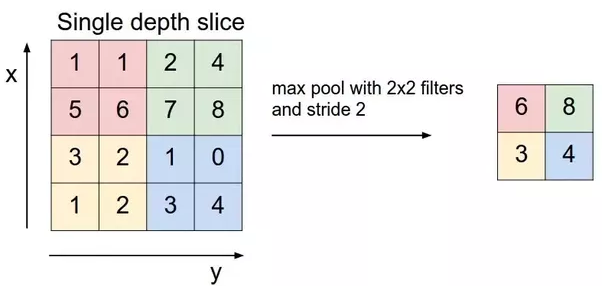


Figure 6: Example of max pooling

* + 1. Average pooling: Takes Average of all the elements in a feature map.
    2. Sum pooling: Takes sum of all the elements in a feature map.

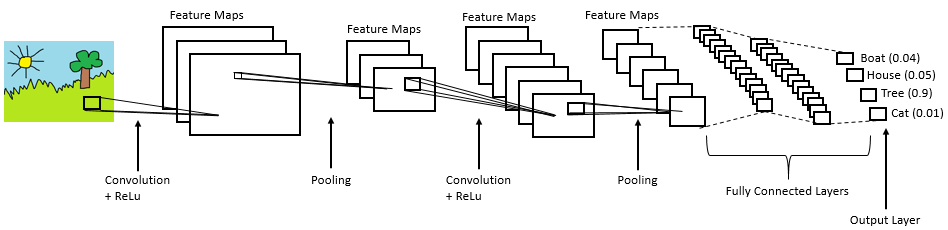


Figure 7: Architecture of a CNN

* We flatten the matrix into vectors and feed it into the Fully connected layer.
* Hence generating the desired classification of the images entered.