Convolutional Neural Network (CNN)

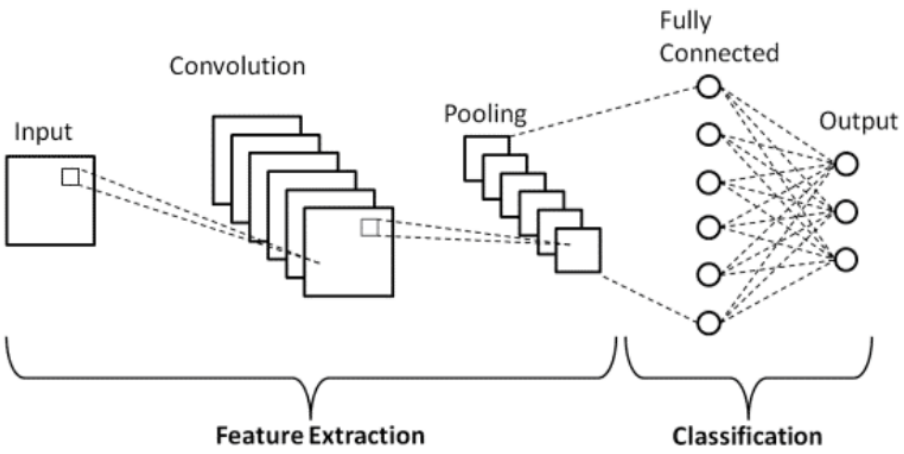
- **CNN** is a type of Deep Learning neural network.
- CNN is the extended version of ANN.
- CNN is used to extract the feature from the grid-like matrix dataset.
- CNN is a type Neural Network.
- CNN performs well for an **image or pixel-based data**.
- CNNs can capture the **spatial information** much better than other models.
- CNN performs much better than Feed Forward Network by a huge margin for Images and Videos.

- CNN used in
  Image/Video Analysis
  Classification tasks,
  Computer Vision
  Object detection,
  Image segmentation
  Face recognition problem.

CNN architecture:

There are mainly three types of layers:

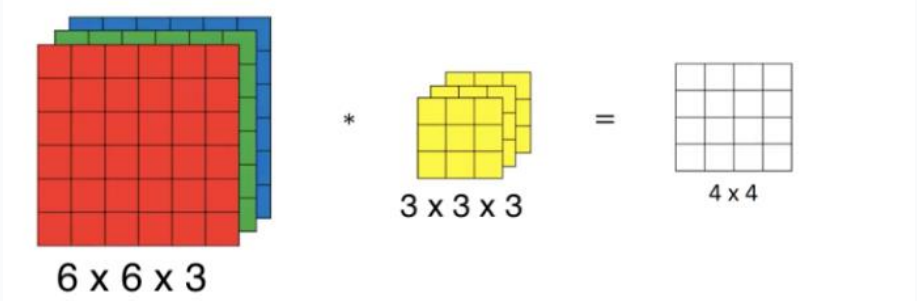
1. **Convolution Layer**
2. **Pooling Layer**
3. **Fully Connected Layer**



1. Convolution Layer

- it is main building block of CNN.
- CNN layer is followed by a **Non-linear activation** function
- Non-linear activation functions can be used(Tanh, Sigmoid, ReLU).
- **ReLU** is usually used in CNN models as they deliver the best results without the vanishing/exploding gradient problem.
- A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction.

In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size MxM. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter (MxM).



- the numbers of channels must be the same for the input image and the filter. See example below for an image with 3 channels (RGB) of dimensions 6×6×36×6×3. The filter has dimensions 3×3×33×3×3, where the last 33 is the number of channels. The resulting image has dimension 4×44×4, assuming s=1$\blacksquare$=1 and p=0$\blacksquare$=0.

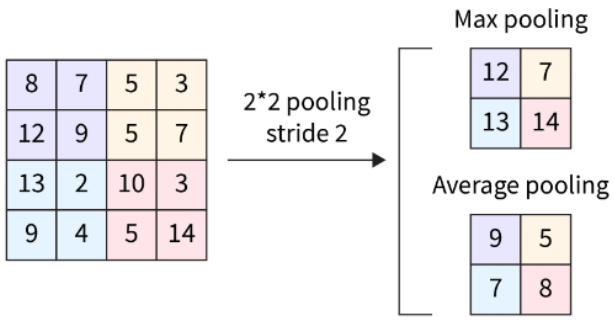
$$\left(\frac{n+2p-f}{s}+1\right)\times\left(\frac{n+2p-f}{s}+1\right)$$

2. Pooling Layer

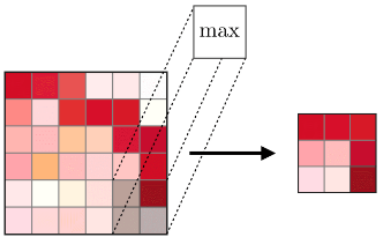
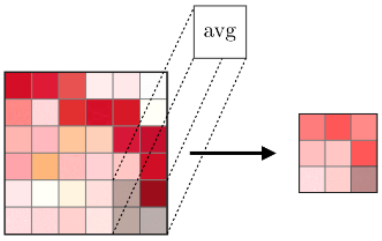
- The Pooling layer is used to reduce the dimension of the image.
- It is also known as **Downsampling**.
- The pooling Layer also uses a kernel and moves across the image but performs the pooling operation.
- Pooling operation reduces the data within the kernel into a single pixel data.
- It makes the computation fast, reduces memory and also prevents overfitting.

There are two types of pooling operation.

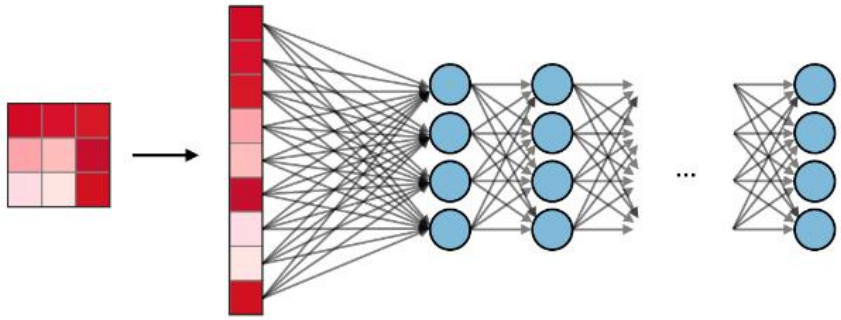
1. **Max Pooling** – This operation outputs the maximum value contained in the kernel. In addition, this pooling performs as a **"Noise Suppressant"**.
2. **Average Pooling** – The average value of the pixels contained in the kernel is returned as the output.



- **Flattening:** The resulting feature maps are flattened into a one-dimensional vector after the convolution and pooling layers so they can be passed into a completely linked layer for categorization or regression.

Type	Max pooling	Average pooling
Purpose	Each pooling operation selects the maximum value of the current view	Each pooling operation averages the values of the current view
Illustration		
Comments	• Preserves detected features • Most commonly used	• Downsamples feature map • Used in LeNet

3. **Fully Connected Layer**
    - A Fully connected layer(FC layer) is the last stage of the CNN.
    - The input FC layer **flattens** all the images into an array of values.
    - It takes the input from the previous layer and computes the final classification or regression task.
    - The fully connected layer (FC) operates on a flattened input where each input is connected to all neurons.



Applications of CNN

- Image classification:
- Object detection:
- Image segmentation:
- Image generation
- Medical image analysis:

Advantages of Convolutional Neural Networks (CNNs):

1. Good at detecting patterns and features in images, videos, and audio signals.
2. Robust to translation, rotation, and scaling invariance.
3. End-to-end training, no need for manual feature extraction.
4. Can handle large amounts of data and achieve high accuracy.

Disadvantages of Convolutional Neural Networks (CNNs):

1. Computationally expensive to train and require a lot of memory.
2. Can be prone to overfitting if not enough data or proper regularization is used.
3. Requires large amounts of labeled data.
4. Interpretability is limited, it's hard to understand what the network has learned.

Convolution Filters / Filters in CNN

- Filters in CNN are also known as Convolution Filters.
- It helps in extracting specific features from input data.

There are different types of Filters.

1) Edge Detection (Prewitt filter)

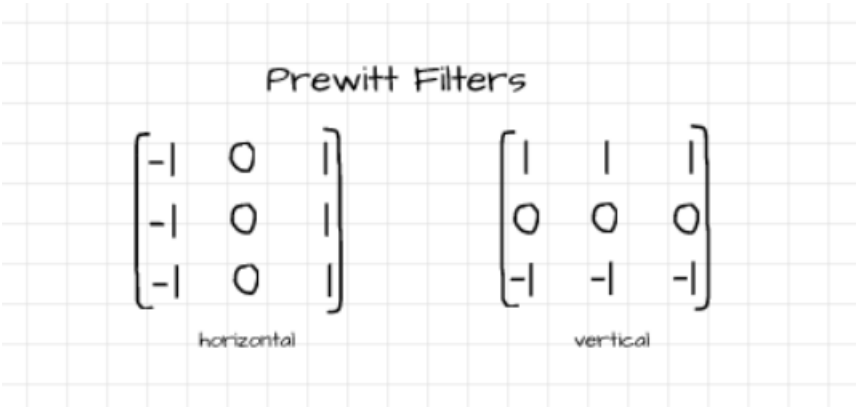
2) Sobel Filter

3) Laplacian Filter

1) Edge Detection (Prewitt filter)

The Prewitt operator is comprised of two filters which help to detect vertical and horizontal edges.

The horizontal (x-direction) filter helps to detect edges in the image which cut perpendicularly through the horizontal axis and vise versa for the vertical (y-direction) filter.



utilizing the horizontal filter

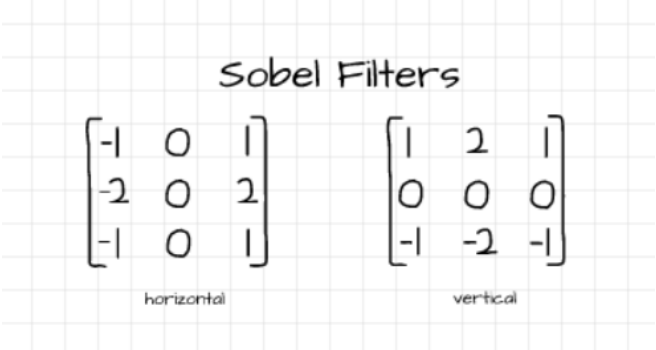
convolve('image.jpg', horizontal)

utilizing the vertical filter

convolve('image.jpg', vertical)

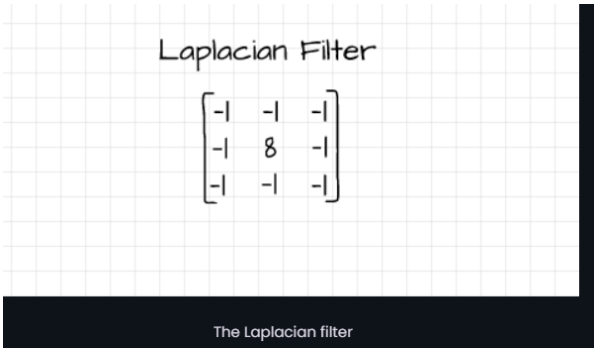
2) Sobel Filter

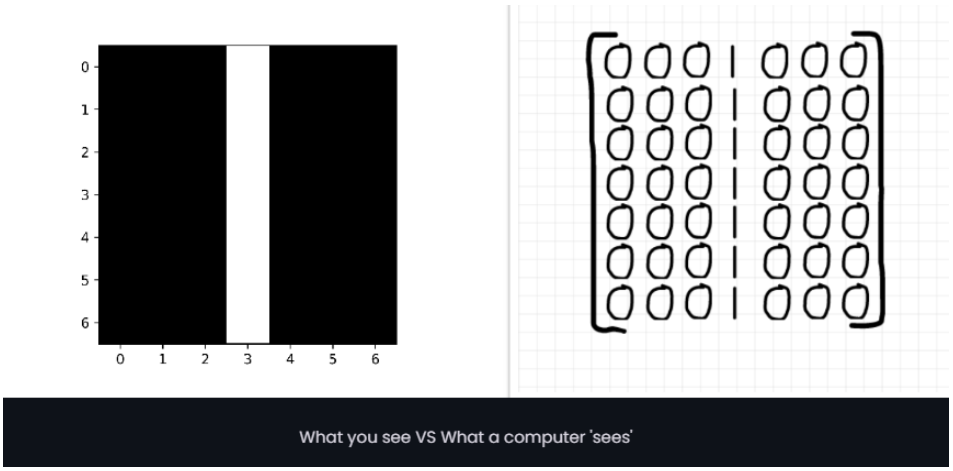
- It is Just like the Prewitt operator,
- the Sobel operator is also made up of a vertical and horizontal edge detection filter. Detected edges are quite similar to results obtained using Prewitt filters but with a distinction of higher edge pixel intensity.
- In other words, edges detected using the Sobel filters are sharper in comparison to Prewitt filters.



3) Laplacian Filter

- Opposite to the Prewitt and Sobel filters,
- the Laplacian filter is a single filter which detects edges of different orientation.
- From a mathematical standpoint, it computes second order derivatives of pixel values unlike the Prewitt and Sobel filters which compute first order derivatives.





Parameter Sharing

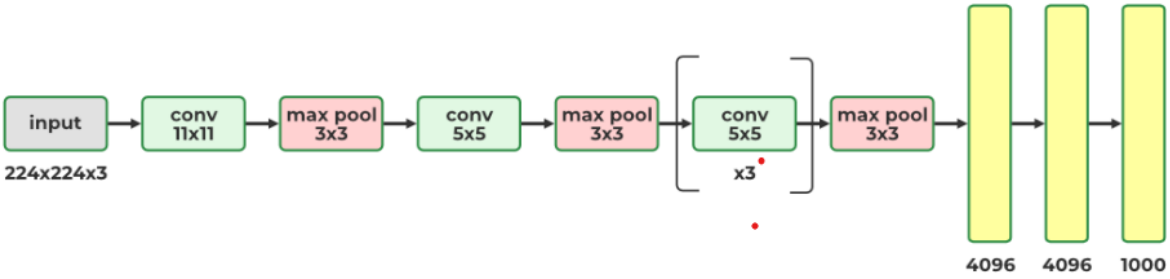
- Share parameters over multiple image locations
- Parameter sharing in CNN makes it translation invariant: i.e., we can find a cat with the same cat detector whether the cat appears at column i or column i+1 in the image
- Parameter sharing in CNN also dramatically lowers the model parameters, and significantly increases network sizes without requiring a corresponding increase in training data
- **Example:** The most extensive use of parameter sharing is in convolutional neural networks. Natural images have specific statistical properties that are robust to translation. For example photo of a cat remains a photo of a cat if it is translated one pixel to the right. Convolution Neural Networks consider this property by sharing parameters across multiple image locations. Thus we can find a cat with the same cat detector in column i or i+1 in the image.

CNN Architectures: LeNet, AlexNet, ZFNet, GoogleNet, VGG and ResNet

AlexNet

- The AlexNet CNN architecture won the 2012 ImageNet ILSVRC challenges of **deep learning algorithm**.
- It was introduced by Alex Krizhevsky,
- It has 8 layers with learnable parameters.
- The input to the Model is RGB images.
- It has 5 convolution layers with a combination of max-pooling layers.
- Then it has 3 fully connected layers.
- The activation function used in all layers is Relu.
- It used two Dropout layers.
- The activation function used in the output layer is Softmax.
- The total number of parameters in this architecture is 62.3 million.

- it was the first CNN architecture that **uses GPU to improve** the performance.



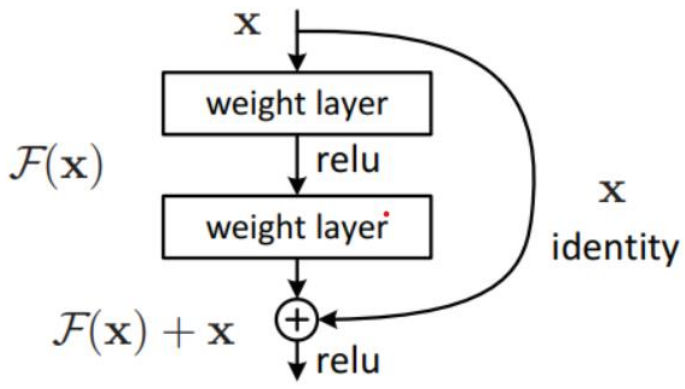
ALexNNet

ResNet (Residual Network)

- There are 152 layers in the Microsoft ResNet.
- The authors showed empirically that if you keep on adding layers the error rate should keep on decreasing in contrast to “plain nets”.
- where adding a few layers resulted in higher training and test errors.
- It took two to three weeks to train it on an 8 GPU machine.

    - When training a neural network, the goal is to make it replicate a target function h(x). By adding the input x to the output of the network (a skip connection), the network is made to model f(x) = h(x) – x, a technique known as residual learning.

```
F(x) = H(x) - x which gives H(x) := F(x) + x.
```



Year	CNN	Developed by	Place	Top-5 error rate	No. of parameters
1998	LeNet(8)	Yann LeCun et al			60 thousand
2012	AlexNet(7)	Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever	1st	15.3%	60 million
2013	ZFNet()	Matthew Zeiler and Rob Fergus	1st	14.8%	
2014	GoogLeNet(19)	Google	1st	6.67%	4 million
2014	VGG Net(16)	Simonyan, Zisserman	2nd	7.3%	138 million
2015	ResNet(152)	Kaiming He	1st	3.6%	