**Convolutional Neural Network:**

In a standard neural system, the info is changed through a progression of concealed layers having various neurons. Every neuron is associated with every one of the neurons in the past and the accompanying layers. This course of action is known as a completely associated layer and the last layer is the yield layer. In Computer Vision applications where the information is a picture, we utilize convolutional neural system in light of the fact that the standard completely associated neural systems don't function admirably. This is on the grounds that if every pixel of the picture is an information then as we include more layers the measure of parameters increments exponentially.

Consider a model where we are utilizing a three shading channel picture with size 1 megapixel (1000 stature X 1000 width) at that point our info will have 1000 X 1000 X (3 Million) highlights. On the off chance that we utilize a completely associated concealed layer with 1000 shrouded units then the weight grid will have 3 Billion (3 Million X 1000) parameters. Along these lines, the general neural system isn't adaptable for picture grouping as handling such an expansive information is computationally extremely costly and not doable. The other test is that countless can prompt over-fitting. Be that as it may, with regards to pictures, there is by all accounts little connection between's two firmly arranged individual pixels. This prompts the possibility of convolution.

**Convolution :**

**Convolution is a scientific activity on two capacities to create a third capacity that communicates how the state of one is adjusted by the other. The term convolution alludes to both the outcome work and to the way toward registering it [1]. In a neural system, we will play out the convolution activity on the info picture grid to decrease its shape. In beneath model, we are convolving a 6 x 6 grayscale picture with a 3 x 3 framework called channel or part to create a 4 x 4 grid. In the first place, we will take the speck item between the channel and the initial 9 components of the picture framework and fill the yield network. At that point we will slide the channel by one square over the picture from left to ideal, through and through and play out a similar computation. At last, we will create a two-dimensional initiation delineate gives the reactions of that channel at each spatial position of info picture lattice.**

Challenges with Convolution

**1- Shrinking output**

One of the enormous difficulties with convolving is that our picture will ceaselessly contract in the event that we perform convolutional tasks in different layers. Suppose in the event that we have 100 concealed layers in our profound neural system and we perform convolution task in each layer than our picture size will recoil a tad after each convolutional layer.

**2- Data lost from the image corners**

### The second drawback is that the pixels from the side of the picture will be utilized in few yields just though the center district pixels contribute all the more so we lose information from the sides of our unique picture. For instance, the upper left corner pixel is associated with just a single of the yield yet center pixel contributed in somewhere around 9 yields.

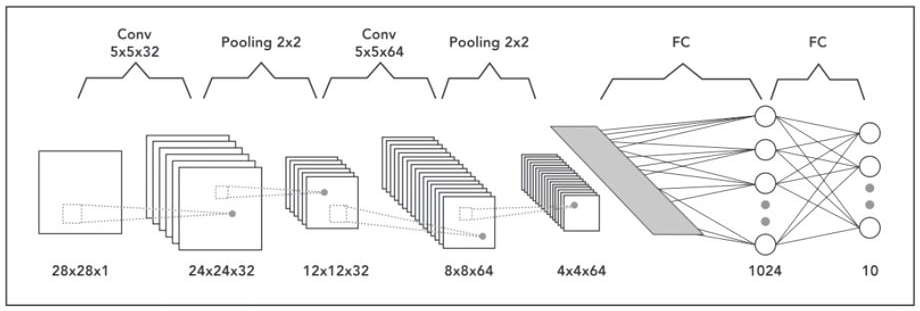
### **Padding**

In order to solve the problems of shrinking output and data lost from the image corners, we pad the image with additional borders of zeros called zero padding. The size of the zero padding is a hyperparameter. This allows us to control the spacial size of the output image. So if we define F as the size of our filter, S as the stride, N as the size of the image, and P as the amount of padding that we require, then the image output size is given by the following

We can see that by using zero padding as 1, we have preserved the size of the original image. There are two common choices of padding. ‘Valid’ where we use P = 0 means no padding at all and ‘Same’ where the value of P is selected such that the size of the output image is equal to input image size. As far as filter size ‘F’ is concerned it is a recommended practice to select the odd number. Common choices are 1, 3, 5, 7…etc.

### **The Architecture of Convolutional Neural Network**

A neural network that has one or multiple convolutional layers is called Convolutional Neural Network (CNN). Let’s consider an example of a deep convolutional neural network for image classification where the input image size is 28 x 28 x 1 (grayscale). In the first layer, we apply the convolution operation with 32 filters of 5 x 5 so our output will become 24 x 24 x 32. Then we will apply pooling with 2 x 2 filter to reduce the size to 12 x 12 x 32. In the second layer, we will apply the convolution operation with 64 filters of size 5 x 5. The output dimensions will become 8 x 8 x 64 on which we will apply pooling layer with 2 x 2 filter and the size will reduce to 4 x 4 x 64. Finally, we will pass it through two fully connected layers to convert our image matrix into a classification matrix.

Architecture of Convolutional Neural Network