

1. Convolution Neural Network

In neural networks, Convolutional neural network (CNN) is one of the main categories to do images recognition, images classifications. CNN model design for working with two-dimensional image data but we can also use with one-dimensional and three-dimensional data.

CNN takes input image, process it and classify under certain categories. Image is array of pixel and size of this array depends on image resolution. A convolution is a linear operation that involves multiplication of set of weight with input image data and this multiplication performed between an array of input data and array of weight called filter or a kernel. In CNN, neurons are arranged in three-dimension, width, height and depth.

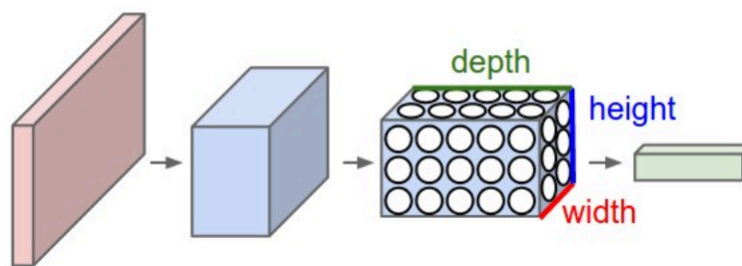


Fig. 1 Convolution Network

As Fig. 1 shows, CNN arranges its neurons in three dimensions (width, height and depth), as visualized in one of the layers. Every layer of a Convolution Network transforms the 3D input volume to a 3D output volume of neuron activations. Here, red input layer holds the image, width and height would be the dimensions of the image, and the depth would be 3 (Red, Green, Blue channels).

In Deep Learning CNN models, input image passes through a series of convolution layers with filters or kernels, pooling layer, fully connected layers and apply appropriate function to classify an object.

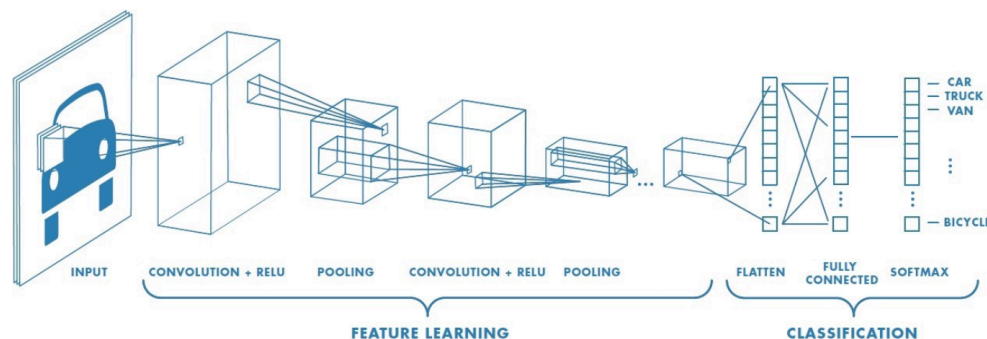


Fig. 2 Neural Network with multiple convolution layers

Fig. 2 shows a complete flow of CNN to process an input image and classifies the objects based on values.

1.1 Convolution Layer

Convolution is the first layer to extract feature from an input image. Convolution keep the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel.



Fig. 3 Image matrix multiplication with kernel matrix

As shown in Fig. 3, convolution image matrix multiplies with filter matrix and gives output which called feature map. The filter matrix shifts over the input image data matrix and its calls stride. For instance, if stride is 1 then filter. Move to 1 pixel at a time and so on. Image with different filters can perform operation such as sharpen the edges, edge detection, and blur etc.

1.2 Pooling Layer

Pooling layers reduce the number of parameters when input image is too large. Pooling reduce the dimensionality of map but retains as much as important information. Pooling layer is created on top of filter in convolution layer. There is different type of pooling such as max pooling, average pooling and sum pooling. Max pooling takes largest element from feature map, average pooling takes average of feature map and sum pooling takes sum of feature map.

1.3 Fully Connected Layer

In fully connected layer works same as neural network in which all the feature combined together and feed it to the model. Fully connected layer uses the activation function like sigmoid and SoftMax to classify the output. Fig. 4 shows CNN architecture with convolution layers and relu activation function, pooling layers and fully connected layers.

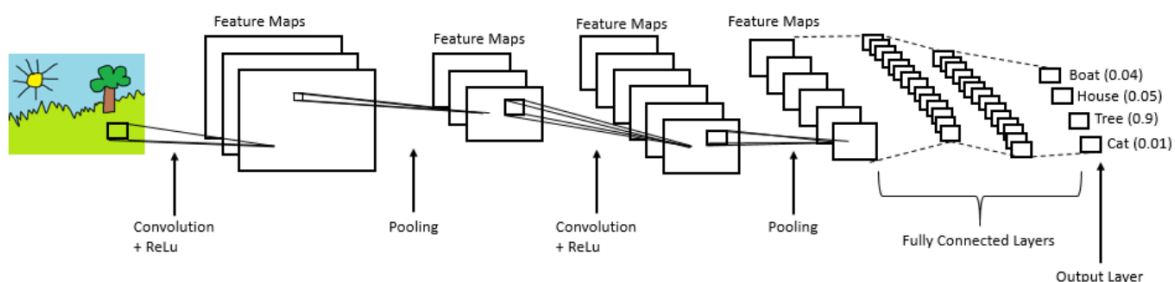


Fig. 4 CNN architecture

1.4 Type of Convolution Neural Network

1.4.1 One Dimensional CNN

One dimensional CNN also known as Convo1D, kernel slides along one dimension. Data like time series uses Convo1D neural network. Convo1D is also used on audio data as it can be represented sound and texts as time series data.

1.4.2 Two Dimensional CNN

Two dimensional CNN also known as Convo2D, is generally used on image data. It called two dimensional CNN because kernel slides along two dimensions on the input data. This can detect edges, distribution of color in image which makes model robust in image classification.

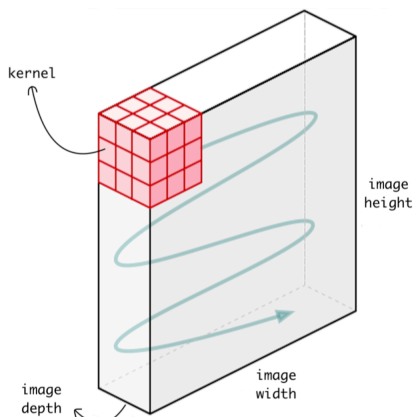


Fig. 5 Kernel Slides along Image

As shown in Fig. 5, Kernel slides along in two dimensions, input shape represent height, weight and depth of the image and kernel size represent the height and width of the kernel and kernel depth will be the same as depth of the image. 2D image has three dimensions where third dimension represent the color channels.

1.4.3 Three Dimensional CNN

Three dimensional CNN also known as Convo3D, the kernel slides in three dimensions.

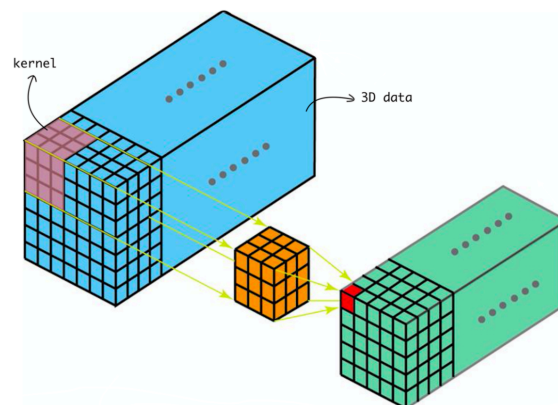


Fig 6. Kernel Slides on 3D data

As shown in Fig. 6, kernel slides in three dimensions, Convo3D is mostly used with 3D image data such as MRI and CT scan data. We can use Convo3D to classify this medical data and extract feature from it. In Convo3D, input shape has 4 dimensions, where fourth dimension represents the number of color channels. Kernel size represents height, width and depth of the kernel and fourth dimension of the kernel will be the same as the color channels.