CONVOLUTIONAL NEURAL NETWORKS

Prepared for:

Dr. Navneet Goyal

Instructor In-charge, Machine Learning (BITS F464)

Department of Computer Science & Information Systems, BITS Pilani

Prepared by:

Namit Shrivastava	**2020A2PS1767P**
Aryan Raj	**2020A2PS1777P**
Yashvardhan Prasad	**2020A2PS1775P**

Convolutional Neural Network

Neural Network

Inspired by the human brain, a neural network is a branch of machine learning designed such that it simulates the way biological neurons signal to each other.
They consist of node layers—an input layer, multiple hidden layers, and an output layer. Each node is linked to another node, and each connection has a weight and a threshold associated with it. If the output of a node is above the threshold value, that node is activated and data is sent to the next layer of the network. Else, no data is forwarded from that node.
There are several types of neural networks, each with their own uses and data types.

- Convolutional Neural Network (CNN): often used for computer vision, image processing and image classification tasks.
- Feed Forward Neural Network: used for pattern recognition and computer vision tasks. A Multi-Layer Perceptron model: used for complex classifications and machine translation tasks.
- Recurrent neural networks (RNNs): used for speech recognition, text to speech conversion and other natural language processing (NLP) tasks. Recently, RNNs have been developed to learn longer-term dependencies and are called Long Short-Term Memory Networks (LSTM)
- Modular Neural Networks: used for stock market predictions and compression of input data.

Structure of a CNN

Convolutional neural networks do better than other neural networks when they are fed images, speech, or audio signals. They comprise of three layers:

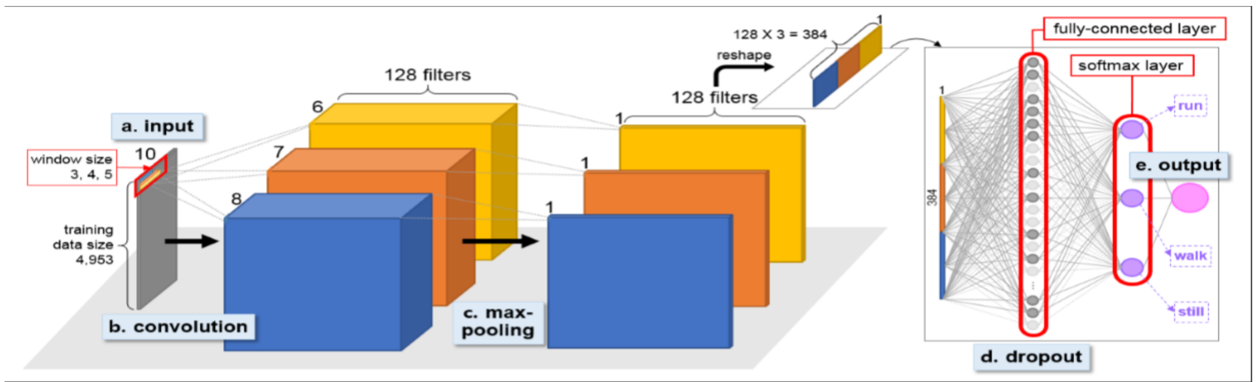
1. Convolutional layer
   Foundation of CNN. Major part of the processing takes place in this layer. A convolutional layer requires input data, a filter, and a feature map.

2. Pooling layer
   also called "downsampling," reduces the number of parameters in the input (dimensionality reduction).

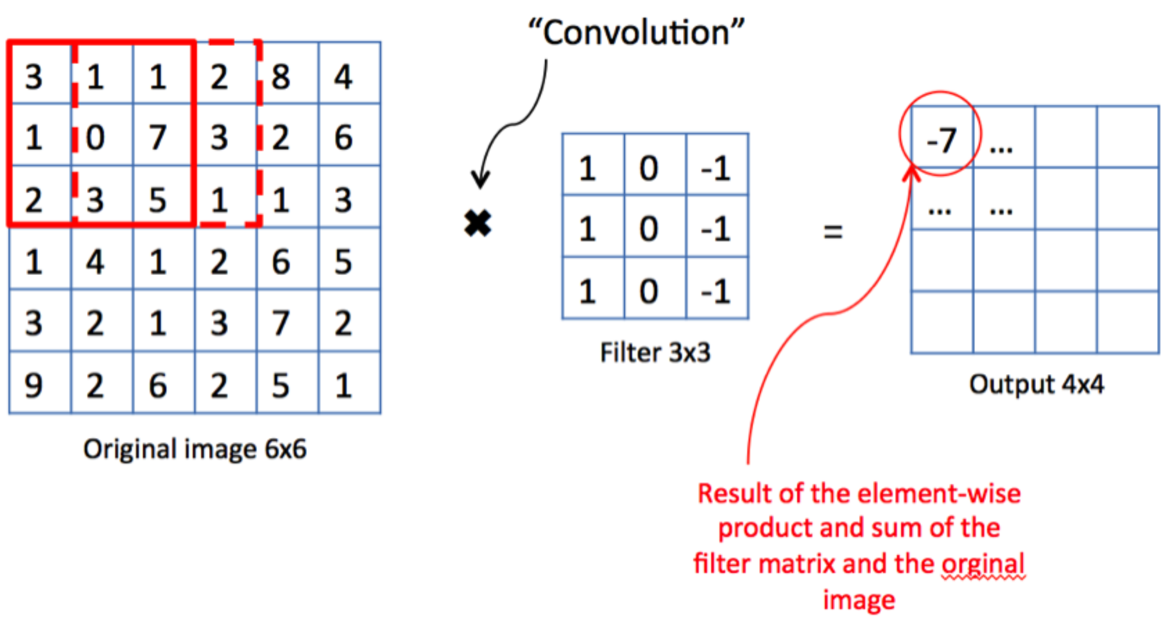
3. Fully-connected layer
   Also referred to as the FC layer. This layer is in charge of classifying the data based on the features that were taken from the previous layers and the filters that were used in them.

A neural network can have multiple convolutional layers, but the fully connected layer is the final layer. With each new layer, the CNN gets more complicated and can recognize more parts of the image. As the image data moves forward through the CNN's layers, it starts to identify larger parts of the objects in the image until it identifies the right one.

Working of CNN



An image can be considered to be a matrix of pixel values. A CNN is able to successfully capture the spatial and temporal characteristics of an image through the application of relevant filters. The architecture provides a better fit to the image dataset due to the reduction in the number of parameters involved and the reusability of weights.



A) CONVOLUTIONAL LAYER

Let the input be an image. Images are composed of pixels. Consequently, a 3D image can be viewed as a matrix of pixels. It will possess three dimensions—height, width, and depth. CNN is equipped with a feature detector, also known as a kernel or a filter, which moves across the image's fields to determine whether the feature is present (the convolution process).

The feature detector is made up of a 2D array of weights, each of which represents a different part of the image. Typically, the filter size is a 3x3 matrix, but the size can vary depending on the circumstance. In addition, this determines the size of the output field. The filter is then applied to a portion of the matrix, and a dot product between the input pixels and the filter is calculated. The resultant dot product is then fed to an output array. The filter is then moved by one stride, and this procedure is repeated until the filter has traversed the entire matrix or image. The final output of this process is known as a feature map. Each output value in the feature map is not required to correspond to each pixel in the input image. It must only connect to the filter being applied (also known as the "receptive field"). In convolutional and pooling layers the output array does not have to correspond directly to each input value. Due to this property, these layers are often called "partially connected" layers.

The feature detector does not change its weights as it traverses the image; this is known as parameter sharing. During training, certain parameters, such as the weight values, are modified by backpropagation and gradient descent. But the size of the output volume is affected by three hyperparameters that must be set before training the neural network.

1. The number of filters influences the output's depth. For instance, three distinct filters would generate three unique feature maps, creating a three-dimensional depth.

2. The number of pixels that the kernel passes over in the input matrix in each step is called a stride. Although stride values of two or greater are uncommon, a larger stride reduces output.

3. There are three types of padding:
    - Valid padding: Also known as no padding, in this case the last convolution is discarded if the dimensions do not align.
    - Same padding: This padding ensures that the output layer has the same dimensions as the input layer.
    - Full padding: This type of padding increases the size of the output by appending zeros to the input's border.

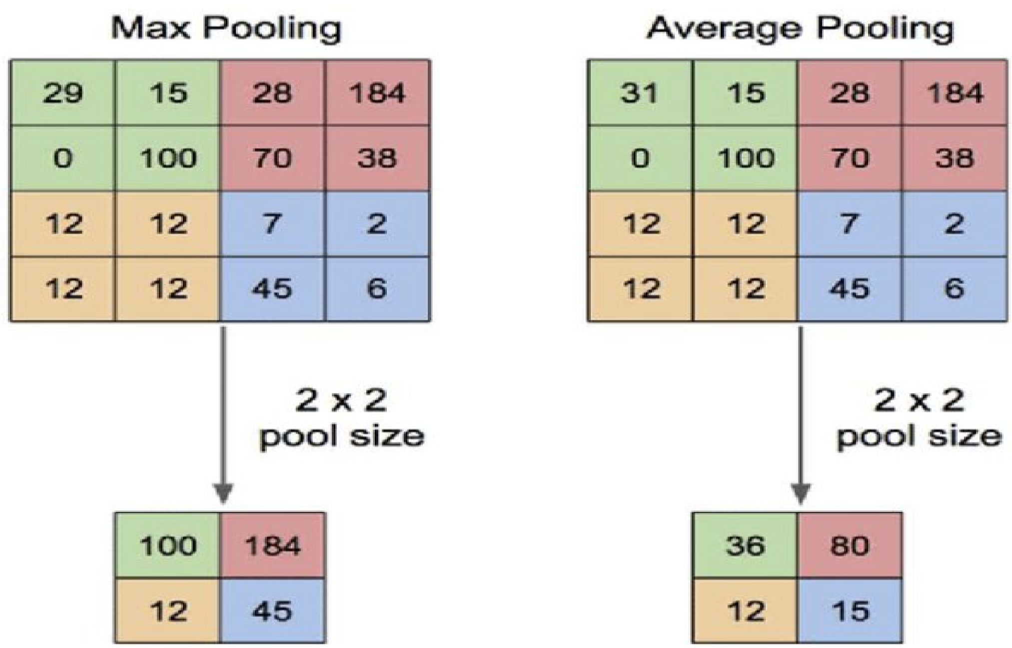
    When the input image exceeds the filter size, zero-padding is used . All of the elements that don't fit in the input matrix are set to zero, so the output is either bigger or the same as the input.

CNN introduces non-linearity into the model by applying a Rectified Linear Unit (ReLU) transformation to the feature map. The convolutional layer transforms the image into numbers, enabling the neural network to interpret and extract relevant patterns.

B) POOLING LAYER

Also referred to as "downsampling," it conducts dimensionality reduction by decreasing the number of input parameters. In the pooling operation a weightless filter is passed across the entire input, similar to convolution. The kernel takes the values in the receptive field and uses an aggregation function on them to fill the output array. There are two main types of pooling:

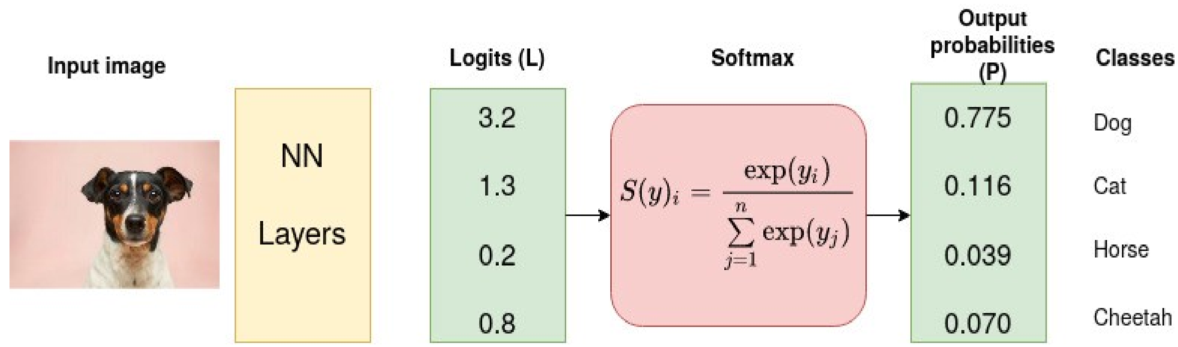
- **Max Pooling**: Pixel with the greatest value is selected from the receptive field and sent to the output array. This method is utilized more frequently than average pooling.
- **Average Pooling**: average value within the receptive field is calculated and sent to the output array.



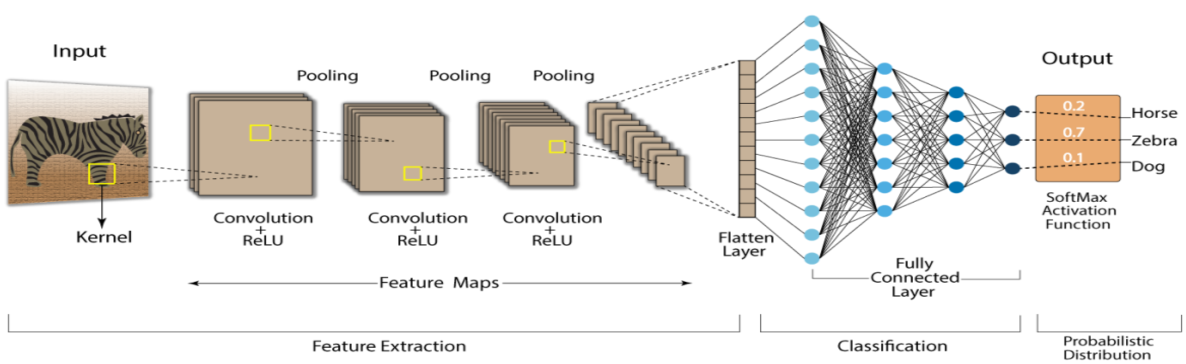
CNN receives a number of benefits from the pooling layer, despite the fact that a significant amount of information is lost. They help in reducing complexity, enhancing efficiency, and reducing the risk of overfitting.

C) FULLY-CONNECTED LAYER

Each node in the output layer is connected directly to a node in the previous layer. This layer is in charge of classifying the data based on the features that were taken from the previous layers and the filters that were used in them. The FC layer uses a softmax activation function for classification. A Softmax activation function scales the logits into probabilities. Logits are the unnormalized final scores of the model, representing the score for each possible output class. A Softmax activation function takes a vector of logits as input, applies the softmax function to each value, and returns the normalized scores (calculating a probability ranging from 0 to 1) for each possible output class. The class with the greatest probability is our final result.



Convolution Neural Network (CNN)



References

1) Education, I. C. (2020, October 20). What are Convolutional Neural Networks? Retrieved November 28, 2022, from

   https://www.ibm.com/cloud/learn/convolutional-neural-networks

2) G. (2021, July 11). Basic Overviews on Convolutional Neural Networks. Retrieved November 28, 2022, from

   https://guymonahan.medium.com/basic-overviews-on-convolutional-neural-networks-f205180116be