A Review On Convolutional Neural Networks

Aastha Valecha

Undergraduate Student
Bachelor of Technology
Indian Institute of Technology(IIT), Delhi, India

Abstract : Over the past few decades, deep neural networks and artificial neural networks(ANN) have been one of the most powerful tools for solving problems in real life. It has become prevalent these days because of its ability to handle a massive amount of data. It is interesting to see its implications in the field of data handling, especially pattern recognition. One of the most valuable tools in working with visual data is Convolutional Neural Networks. CNN has shown excellent performance in machine learning problems. In this paper, we will look at the crucial elements of CNN and how they work. We will also look at the practical application of CNN in object identification. This paper also lists the shortcomings and the prospects of the CNN model.

Index Terms – *CNN, Object Classification, Convolution, image processing.*

I. INTRODUCTION

A convolutional neural network (CNN) is a kind of artificial neural network utilized in picture recognition and processing. This is particularly designed to process pixel information. Convolutional neural networks, additionally referred to as ConvNets, had been first introduced in the 1980s by Yann LeCun, a French computer science researcher. In deep learning, it is a class of deep neural networks, most typically implemented to investigate visual imagery.

II. LIMITATIONS OF TRADITIONAL NEURAL NETWORKS

A multilayer perceptron (MLP) is a class of feedforward artificial neural network. Traditionally, MLP has been used for image processing. A MLP includes three layers of nodes namely, an input layer, a hidden layer, and an output layer. Besides for the input nodes, every node is a neuron that makes use of a nonlinear activation feature. Supervised learning approach referred to as backpropagation for training is used in MLP. The multiple layers and non-linear activation of MLP distinguish it from a linear perceptron. It is able to distinguish information that isn't always linearly separable.

There are several drawbacks of MLP's, specifically with regards to image processing. MLPs use one perceptron for every input (for example, pixel in a photograph, multiplied by three in RGB case). The quantity of weights hastily turns into unmanageable for huge pictures. MLP is now deemed inadequate for cutting-edge superior computer vision responsibilities. It has the feature of completely connected layers, wherein every perceptron is attached with each other perceptron. Drawback is that the quantity of overall parameters can develop to a very excessive number of perceptrons. This can lead to computational inefficiency and problems in fitting the model can arise due to this.

Another drawback of MLP's is that they react in a different way to an input (image) and it's shifted version. One of the primary issues is that spatial data is misplaced when the image is flattened into an MLP. It disregards the spatial information.

III. ADVANTAGES OF CNN IN IMAGE PROCESSING

The Convolutional Neural network (CNN) has given outstanding performance in lots of computer vision and machine learning problems. The key advantages of CNN are listed as follows-

3.1 Feature Learning

The most remarkable difference among CNNs and conventional ANNs is that CNNs are basically used within the discipline of pattern recognition of images. This permits us to encode image-precise features into the structure, making the network efficient. They can capture relevant features from an image/video at different levels similar to a human brain. This is called feature learning.

3.2 Feature Extraction

Useful Attributes can be detected from a CNN that has been trained already. CNN's training weights should be fed on the data and then modified for performing a particular task.

3.3 Parameter Sharing

Through parameter sharing, we are able to reduce the quantity of weights inside a conv layer. Parameter sharing is utilized in all convolutional layers in the network. Training time is reduced due to parameter sharing; this is the advantage of the reduction of the number of weight updates that need to take place throughout the backpropagation.

The main strengths of CNNs are to provide an efficient dense network which performs the prediction or identification efficiently.

IV. CNN ALGORITHM

The neurons inside a CNN are structures from neurons of the input image. Input image has the spatial dimensionality of three due to height, width and depth. The layers are arranged in such a manner that the simpler patterns like curves, plane lines, horizontal and vertical edges are detected easily and the more complicated patterns like the face recognition and object classification can be done after that.

CNN consists of five layers, namely input layers, convolutional layers, pooling layers, fully-connected layers,and output layers. When these layers are stacked together, a CNN structure can be shaped.

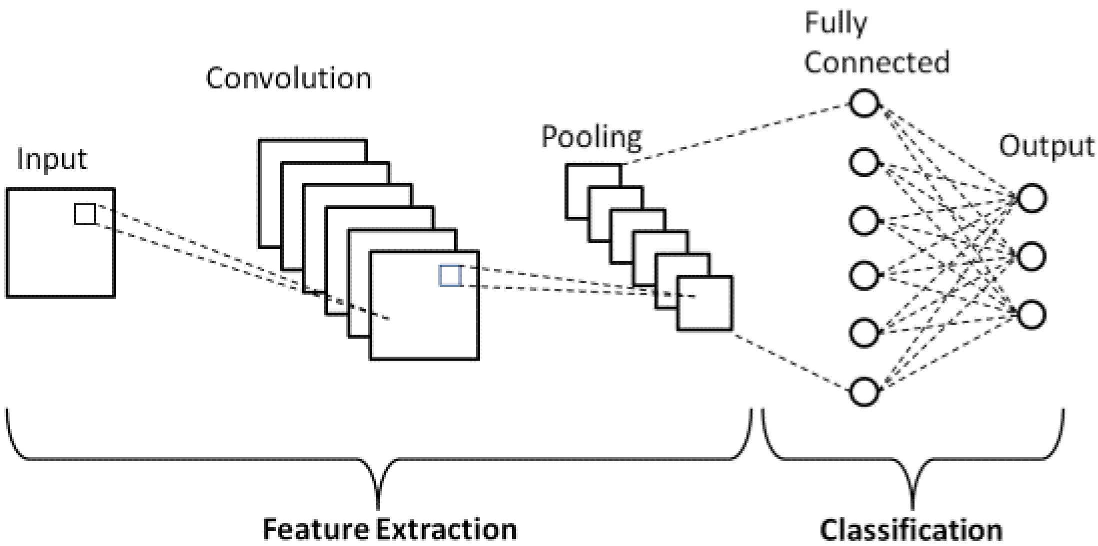

Figure 1 - Architecture of the Convolutional Neural Networks showing three layers and input/output structure

4.1 Input Layer

The input layer in CNN contains image information. Image data is represented via a three-dimensional matrix. We need to reshape the image into a single column. As observed in other kinds of ANN, the input layer will maintain the pixel values of the image.

4.2 Convolution Layer

The convolution layer is the primary building block of the CNN. It contains the crucial portion of the network's computational load. Convolution preserves the relationship among pixels by learning image features and the usage of small squares of input data. It is a mathematical operation that takes inputs consisting of an image matrix and a filter or kernel. The convolutional layer then determines the output of neurons which are connected to local regions of the input via the calculation of the scalar product among their weights and the region linked to the input volume.

4.3 Pooling Layer

The Pooling layer decreases the spatial size of the Convolved feature. This layer decreases the computational strength required to process the information because it reduces the dimensionality of the matrix. The pooling layer will then simply carry out downsampling along the spatial dimensionality of the given input, further lowering the number of parameters inside that activation. Pooling layers segment would reduce the number of parameters when the images are too massive. Spatial pooling, also known as subsampling or downsampling, reduces the dimensionality of every map, however it keeps crucial information.

4.4 Fully-connected Layers

Neurons in the fully-connected layer have full connectivity with all neurons within the previous and succeeding layer. This is the reason why it can be computed through the standard matrix multiplication accompanied via a bias impact. The fully-connected layers carry out the same responsibilities as that of a general ANNs and then produce the class scores from the activations that are to be used for classification. It is also recommended that ReLu can be used among these layers, to enhance overall performance.

Through this simple method of transformation, CNNs are capable of remodelling the original input layer through the use of convolutional and downsampling techniques to provide class scores for classification and regression purposes. Convolutional Neural Networks vary to other kinds of Artificial Neural networks in which rather than specializing in the whole lot of the problem domain, information about the specific type of input is exploited. This in turn permits for a far less complicated network architecture to be set up.

Understanding the architecture of CNN isn't enough to solve the problems. The creation and optimisation of these models can take quite some time, and can be pretty complicated.

V. Application of CNN in Object Detection

Studies have been taking place at a really rapid pace in the domain of object detection and the outcomes are simply impressive. As the term indicates, object Detection identifies and localizes an object within an image. numerous techniques and strategies are proposed to resolve the object detection challenges. CNN is a type of feed-forward neural network and works on the principle of weight sharing. Convolution is an integration displaying how one function overlaps with another function and is a blend of two functions being multiplied. The object Detection Deep learning Algorithms can be defined as follows-

1. An ‘encoder’ takes the source image and runs a series of processes in layers to learn and extract features. The features assist to locate and label each object.
2. A ‘decoder’ or a regressor takes the output of the encoder as the input. Using this predicts the location and size of each object, highlighting it with a bounding box. The location of image in the 2-D plain is the output of the decoder in terms of the coordinates x and y.

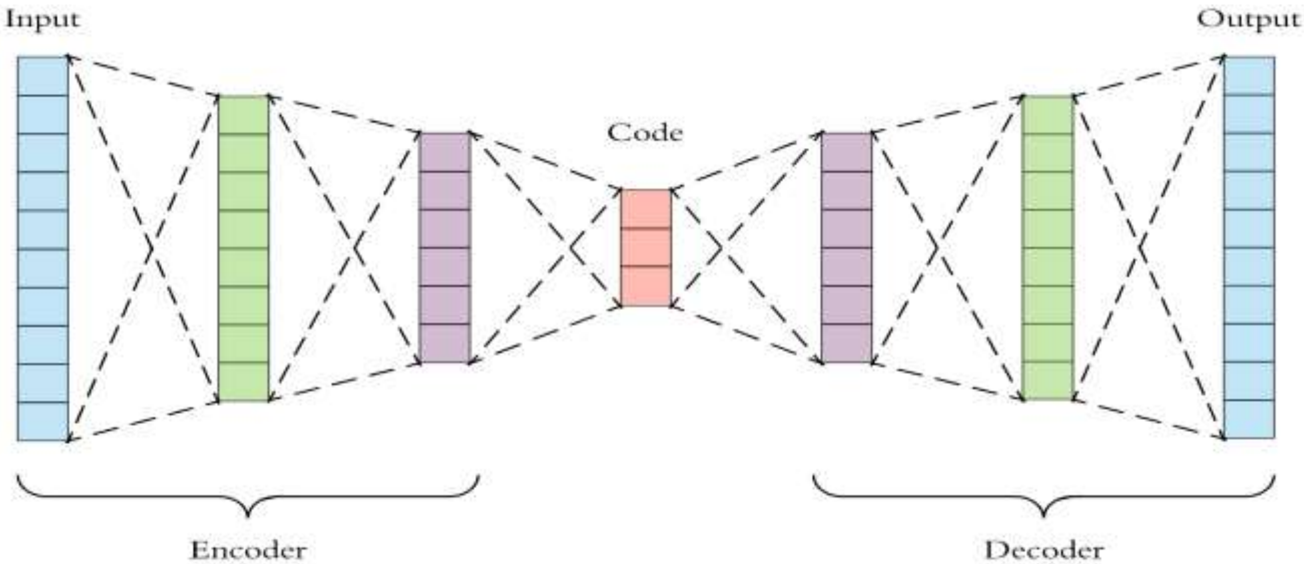

Figure 2 - Encoder-Decoder structure in CNN

Object detection holds it's applications in various domains like surveillance, human computer interaction, robotics, transportation, and retrieval.

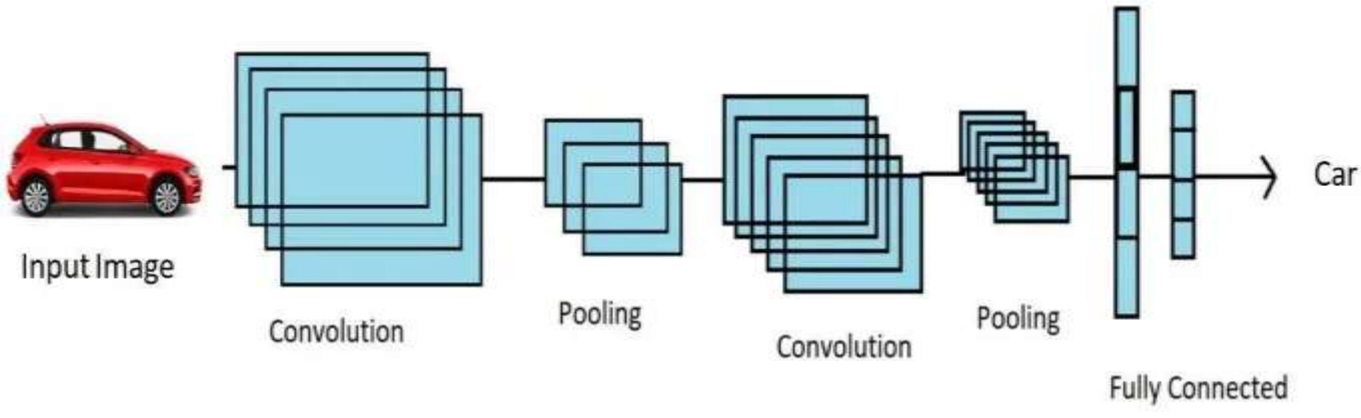

Figure 3- CNN Architecture for Object Detection

The limitation of CNN working with real life images is that they have multiple objects that belong to various image categories hence the computation of such an image will be very costly. To resolve this issue we have built on the CNN algorithm to produce more efficient algorithms like R-CNN, Fast R-CNN, and Faster R-CNN. These models are computationally more feasible for real life problems like facial expression detection, and autonomous vehicles.

- R-CNN (Region-Based Convolutional Neural Networks)

To bypass the problem of choosing a huge number of regions, Ross Girshick et al. proposed a method wherein we use selective search to extract just a few selective regions from the image and he referred to them as region proposals. Consequently, now, rather than classifying a huge number of regions, selected regions can be worked upon.

- Fast RCNN

In fast RCNN, we feed the input image to the CNN, which in turn generates the convolutional feature maps. With the usage of these maps, the regions of proposals are extracted. RoI pooling layer is basically used for reshaping all of the proposed regions into a set size. These regions are then fed into a fully-connected network.

- Faster RCNN

Faster RCNN is an enhancement of Fast RCNN. Fast RCNN makes use of the selective search for generating regions of interest, while faster RCNN makes use of the "region proposal network". RPN takes image feature maps as an input and generates a fixed number of object proposals, each with an objectness score as output. The expected region proposals are then reshaped by the use of a RoI pooling layer which is in turn used to categorise the image in the proposed region and predict the offset values for the bounding image boxes.

VI. Limitations of CNN

Although CNN's have proved their metal in a variety of problems, it does has some limitations.

6.1 Data Requirement

CNN deals with data obtained from images as there pixel values. We are usually handling high-dimensional data and this creates the need for a large amount of data to be fed in a convolutional neural network to work effectively.

6.2 Time efficiency

CNN network deals with heavy data, hence the time required for computation increases. If the CNN has several layers then the training process takes a lot of time if the computer doesn't consist of a good GPU.

VII. Conclusion

Convolutional neural networks have had groundbreaking results over the past decade in numerous fields related to pattern recognition. From image classification to voice recognition, convolutional neural networks have provided great results in problem solving. The key aspect of CNN that makes it so different is reducing the number of parameters in neural networks. This achievement of CNN has contributed a lot in improving the computational efficiency and hence made it more adaptable by researchers and developers all over the world. It allows us to solve large models and complicated tasks, which are not quite possible to be solved via the traditional artificial neural networks.

References

[1]http://scholar.google.co.in/scholar_url?url=https://cs.nju.edu.cn/wujx/paper/CNN.pdf&hl=en&sa=X&ei=vrj6YKKdCZLoyQT8xLGoBw&scisig=AAGBfm148Ejn5IEBu6A0Ua41U763WkOMXQ&nossl=1&oi=scholarr

[2]http://scholar.google.co.in/scholar_url?url=https://arxiv.org/pdf/1511.08458&hl=en&sa=X&ei=vrj6YKKdCZLoyQT8xLGoBw&scisig=AAGBfm0foAzkh1jVEBJ_ML8Jn2DFR_n3lA&nossl=1&oi=scholarr

[3]https://towardsdatascience.com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac

[4]https://medium.com/data-science-bootcamp/multilayer-perceptron-mlp-vs-convolutional-neural-network-in-deep-learning-c890f487a8f1

[5]Understanding Parameter Sharing (or weights replication) Within Convolutional Neural Networks | by Richmond Alake | Towards Data Science

[6]https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

[7]https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[8]https://www.researchgate.net/publication/285164623_An_Introduction_to_Convolutional_Neural_Networks

[9]https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/

[10]https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

[11]https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.upgrad.com%2Fblog%2Fbasic-cnn-architecture%2F&psig=AOvVaw0mYrLe-8CuBN4OXKhfTHHW&ust=1627203140807000&source=images&cd=vfe&ved=0CAwQjhxqFwoTCMi4yKOq-_ECFQAAAAAdAAAAABAD

[12] https://www.upgrad.com/blog/basic-cnn-architecture/

[13]https://www.analyticsvidhya.com/blog/2021/05/20-questions-to-test-your-skills-on-cnn-convolutional-neural-networks/#:~:text=Input%20Layer%3A%20The%20input%20layer%20in%20CNN%20should,x%201%20before%20feeding%20it%20into%20the%20input.

[14]https://medium.com/@iamvarman/how-to-calculate-the-number-of-parameters-in-the-cnn-5bd55364d7ca#:~:text=Output%20Layer%3A%20This%20layer%20is%20the%20fully%20connected,Fully-connected%20layer%2C%20as%20it%20as%20a%20convolutional%20layer.

[15] Object Detection using Regions with CNN features | by Varun Pusarla | Analytics Vidhya | Medium

[16] Object Detection Using CNN - 360DigiTMG

[17] https://www.sciencedirect.com/science/article/pii/S1877050918308767

[18]https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FConvolutional-encoder-decoder-network-based-detection-of-crop-lines-using-semantic_fig1_340616012&psig=AOvVaw0E7GKT5LRk--JEsFXA0z6y&ust=1627256931004000&source=images&cd=vfe&ved=0CAwQjhxqFwoTCMDr6NHy_PECFQAAAAAdAAAAABAD

[19]Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks (neurips.cc)

[20]S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031

[21]https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

[22]https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/

[23]https://iq.opengenus.org/disadvantages-of-cnn/#:~:text=Minor%20Drawbacks%20of%20CNN%3A%201%20A%20Convolutional%20neural,Dataset%20to%20process%20and%20train%20the%20neural%20network

[24]machine learning - What are the drawbacks of fully-convolutional neural networks? - Computer Science Stack Exchange

[25]https://www.google.com/url?sa=i&url=https%3A%2F%2Ftowardsdatascience.com%2Fapplied-deep-learning-part-3-autoencoders1c083af4d798&psig=AOvVaw1EwceWZlhH0cCZyYBmJwsh&ust=1627392486054000&source=images&cd=vfe&ved=0CAwQjhxqFwoTCLj9t8zrgPICFQAAAAAdAAAAABAQ