International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 11 | Nov 2020 www.irjet.net p-ISSN: 2395-0072

Convolutional Neural Network (CNN) for Image Detection

Prof. Sujata Bhairnallykar[1], Aniket Prajapati[2], Anurag Rajbhar[3], Sahil Mujawar[4]

[1-4]*TE. Computer Engineering, Saraswati College of Engineering, Navi Mumbai, Maharashtra 410210, INDIA*

Abstract - *The most impressive forms of artificial neural network(ANN) architecture is the Convolutional Neural Network (CNN). The use of CNNs is to solve difficult image pattern recognition tasks with its accurate and simple design, offers a simple method of getting started with ANNs. Convolution neural network has been used in the field of digital image processing from such a long time and also in speech recognition and natural language processing(NLP), and gained great success. Similar to deep learning algorithms the convolutional neural networks also have algorithms that can train large datasets with millions of parameters, in form of 2D images as input and convolve or frame it with filters to produce the desired outputs. In this paper, CNN models are primarily built to evaluate its performance on image recognition and detection datasets. How the CNN's recognizes the handwritten alphabets are depicted with suitable example in this article. The deep learning algorithms are inspired from the biological function of our neurons in our visual cortex. These algorithms mimic the functionality of our cerebral cortex. After reading this paper we assume that readers will have some knowledge of machine learning and artificial neural networks.*

Key Words*: *Deep Learning, Handwritten digit Recognition, Object Detection, Convolutional Neural Networks, computer vision, Image recognition.

1. INTRODUCTION

Convolutional neural network (CNN) as a successful model of deep learning, has the ability of hierarchical learning features, and the research shows that the feature extracted by CNN has a stronger ability of discrimination and generalization than hand -crafted feature[5].

Convolutional Neural Network is the widely used deep learning framework which was inspired by the visual cortex of animals[7].

Deep Learning is an implementation of the artificial neural networks with multiple hidden layers to mimic the functions of the human cerebral cortex[1].The MNIST dataset is a dataset containing handwritten digits and tests the performance of a classification algorithm. Handwritten digit recognition is an image classification and recognition problem and there have been recent advancements in this field[1].

Deep learning has emerged as an approach for achieving promising results in various applications like image recognition, speech recognition, natural language understanding, signal processing, face recognition, prediction of bioactivity of small molecules etc[2].

Another important aspect of CNN, is to obtain abstract features when input propagates toward the deeper layers[11].

2. NEURAL NETWORK

Neural networks are device that computes parallely, which is a replication of the brain. The main purpose is to develop a system that can perform various computational tasks efficient and faster than the traditional systems[4].

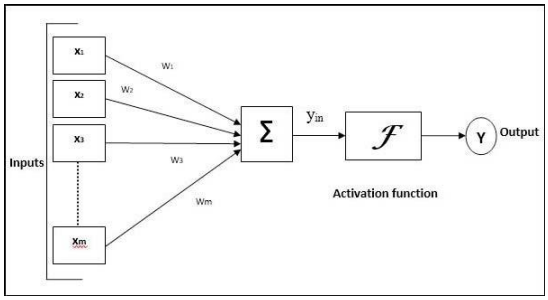


Fig. Neural network

Problem:

But here the problem was if there is a larger input it was difficult to get data and may cause over fitting.

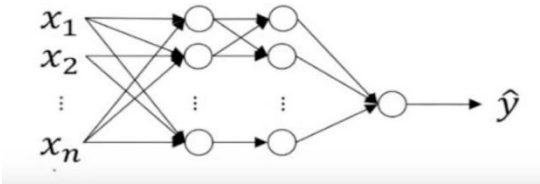


Fig: NN with hidden layers.

Consider an example we have 5 million inputs and 500 hidden layers and we require 5*500 million dimensional matrix, which is much more to deal with. Hence we came up with the concept on the top of neural network as CNN.

3. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are also known as convent. Convents are very similar to normal neural networks which can be visualized as a collection of neurons arranged as an acyclic graph[9]. The main difference from a neural network is that a hidden layer neuron is only connected to a subset of

neurons in the previous layer. Because of this sparse connectivity it is capable of learning features implicitly. Generally CNN consists of four components namely:

(1) Convolution layer

(2) ReLU layer(Activation Function)

(3) Pooling layer

(4) Fully connected layer.

Functionality of each component has been illustrated below. CNN is a model that is gaining attention because of its classification capability based on contextual information[2].

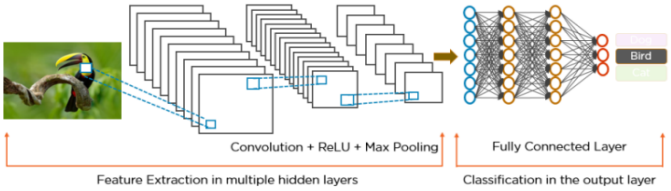


Fig: Typical CNN structure

Imagine there's an image of a bird, and you want to identify whether it's really a bird or some other object. The first thing you do is feed the pixels of the image in the form of arrays to the input layer of the neural network (multi-layer networks used to classify things). The hidden layers carry out feature extraction by performing different calculations and manipulations. There are multiple hidden layers like the convolution layer, the ReLU layer, and pooling layer, that perform feature extraction from the image. At the end fully connected layer identifies the object in the image.

4. METHODOLOGY

Now let's see an example of the image which has the character 'X' in it and how it is going through the steps of CNN layers for its identification.

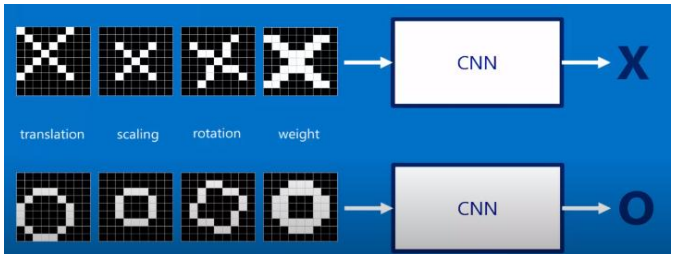


Fig: Identification

4.1 CONVOLUTION LAYER

Basis of CNN is convolutional layers. In CNN, every image is represented in the form of an array of pixel values. More precisely convolutional layers are able to detect patterns in images.

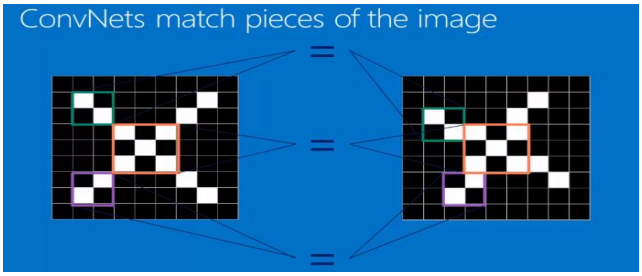


Fig Pattern Matching

With each convolutional layer we need to specify the number of filters the layer should have. These filters are actually what detect the patterns. Filter detect could be edges, corners, circles, squares, etc (objects). These kind of geometric filters are what we'd see at the start of our network the deeper our network goes the more sophisticated these filters become.

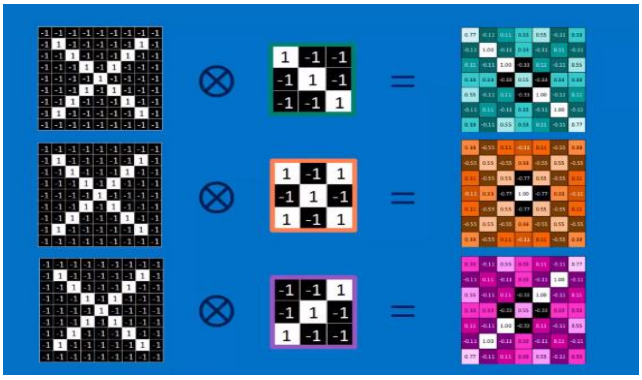
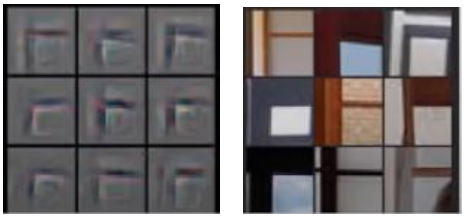


Fig: Filters for character.

So in later layers rather than edges in simple shapes our filter may be able to detect specific objects like eyes, ears, hair or fur feathers, scales and beaks. Even deeper layers filters are able to detect more sophisticated objects like full dogs, cats, birds, etc. When adding a convolutional layer to a model we also have to specify how many filters we want the layer to have, a filter can technically just be thought of as a relatively small matrix for which we decide number of rows and columns that this matrix has and the values within the matrix are initialized with random numbers. We will slide this matrix over the every m*n block of pixels from the entire image, this sliding is actually referred to as convolving. These filters are called as pattern detectors.

Consider an example of more complex filter for pattern detection in dogs, edges, corners.



Zero Padding:

During convolving the convolved image is smaller than original image and there are chances we may lose some data for that we use padding. Convolution reduces the image dimension. Zero padding is a technique that allows us to preserve the original input size. Zero padding occurs when we add a border of pixels all with value zero around the edges of our input this kind of adds a padding of zeroes around the outside of the image.

4.2 RECTIFIED LINEAR UNIT(ReLU)

There are many types of activation functions like sigmoid, ReLU for activation of pixels in the convolved layers of the image. Once the feature maps are extracted, the next step is to move them to a ReLU layer. ReLU performs an element-wise operation and sets all the negative pixels to 0. It introduces non-linearity to the network, and the generated output is a rectified feature map. Below is the graph of a ReLU function. The original image is scanned with multiple convolutions and ReLU layers for locating the features.

$$f(x) = \max(0, x)$$

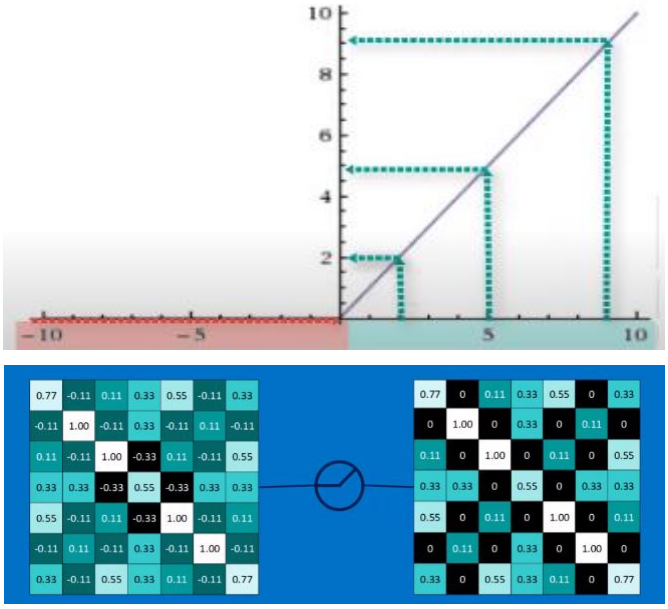


Fig: Rectified image map and rectified image

A stack of images become a stack of images with no negative values. This rectification is done with all the convolved images of the layer.

4.3 POOLING LAYER

Pooling is performed after non-linear activation, where the pooling layer helps reduce the number of parameters and avoids overfitting, and it also serves as a smoothing measure to eliminate unwanted noise. Pooling is also of many types like min pooling, max pooling, average pooling, etc. Pooling mainly used to reduce the image size using the filter and stride. Max pooling is reducing the resolution of the given output of a convolutional layer, the network will be looking at larger areas of the image at a time going forward which reduces the amount of parameters in the network and consequently reduces computational load additionally, max pooling may also help to reduce overfitting. It mainly focuses on the higher valued pixels which have the values for edges, curves, etc and ignoring the lower value pixels.

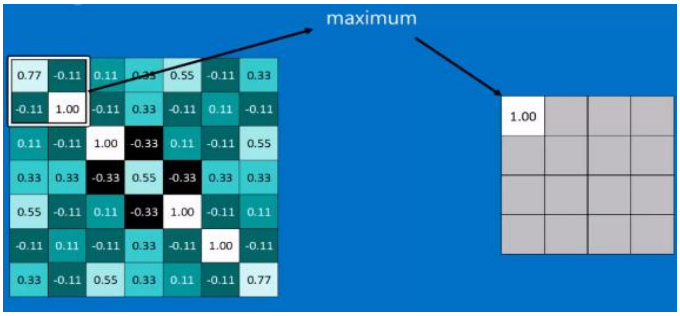


Fig: Pooling with 2*2 matrix_size and stride of size 2

Just like rectification pooling is also done with all the convolved images present in the layer[Fig: Pooling every image].

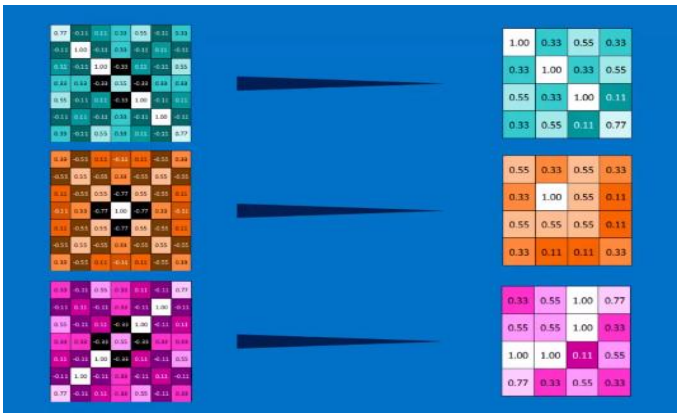


Fig: Pooling every image

Deep stacking:

These three layers can be repeated several(or many) times. This process of layering is known as deep stacking. This process in turn helps in more precise detection of details in images [Fig: Deep stacking].

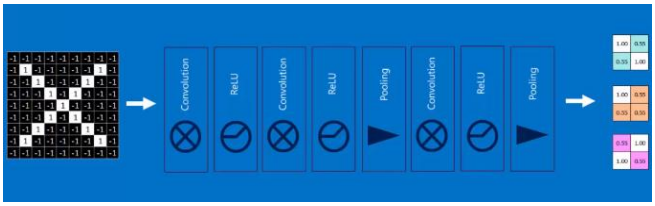


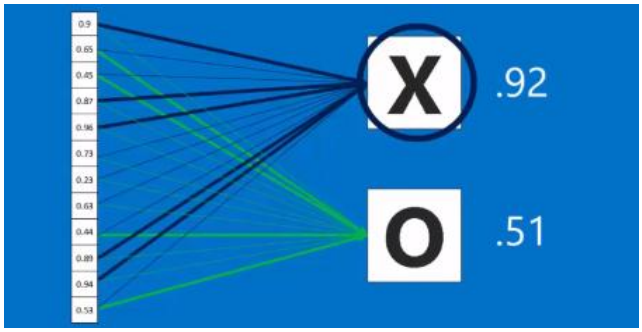
Fig: Deep stacking.

4.4 FULLY CONNECTED LAYER

Fully Connected Layer is simply, feed forward neural networks[9]. Fully Connected Layers form the last few **layers** in the network[9]. The fully connected layer (FC) layers are usually found towards the end of CNN architectures and can be used to optimize objectives such as class store. Neurons in a fully connected layer have full connections to all activations in the previous layer, as seen in regular Neural Networks[9].

In fully connected layer every value(neuron) gets a vote.

Vote depends on how strong a value predicts image.



4.5 OVERALL EXPLANATION

Here's how exactly CNN recognizes any image:

- The pixels from the image are fed to the convolutional layer that performs the convolution operation.
- It results in a convolved map.
- The convolved map is applied to a ReLU function to generate a rectified feature map.
- The image is processed with multiple convolutions and ReLU layers for locating the features.
- Different pooling layers with various filters are used to identify specific parts of the image.
- The pooled feature map is flattened and fed to a fully connected layer to get the final output.
- The output will have the value ranging from (0 to 1) showing probability of match.

A set of pixels becomes a set of votes.

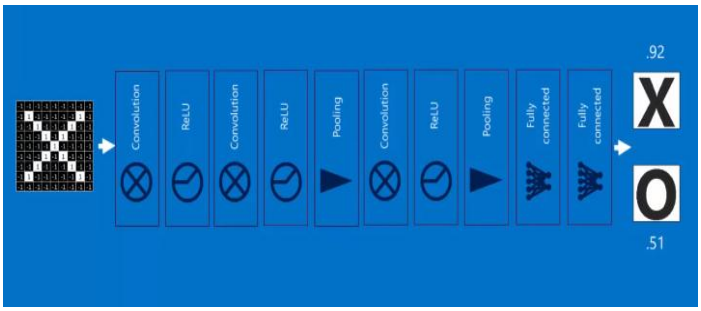


Fig: Working of CNN.

5. CONCLUSIONS

In this work, an experiment is carried out on the image dataset with training of the model and then classifying the image as of character "X" or "O", based on the features extracted. Similarly we can classify different types of objects, images etc[4]. The accuracy on training set may also be improved further by adding more hidden layers. And this system can be implemented as a assistance system for machine vision for detecting nature language symbols[1]. Another future scope using this objects in image detection techniques we can design a robot which can learn thing by watching videos. For example the robot can cook, paint, etc. Also there can be optimization in recognition technique so that computer can be able to detect images in the playing videos.

REFERENCES

[1] Rahul Chauhan, Kamal Kumar Ghanshala, R.C Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition", First International Conference on Secure Cyber Computing and Communication (ICSCCC), Volume: 8, Issue: August 2018, pp: 278-282.

[2] Sakshi Indolia, Anil Kumar Goswami, S.P. Mishra, Pooja Asopa, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach", International Conference on Computational Intelligence and Data Science(ICCIDS), Volume: 5, Issue: May 2018, pp 679-688.

[3] Nadia Jmour, Sehla Zayen, Afef Abdelkrim, "Convolutional Neural Network for image classification", Institute of Electrical and Electronics Engineers (IEEE), Volume: 2, Issue: February 2018, pp: 397-402.

[4] Sanjay Kumar, Manish Kumar, "A Study on the Image Detection Using Convolution Neural Networks and Tenser-Flow", International Conference on Inventive Research in Computing Applications (ICIRCA), Volume:2, Issue: February 2018, pp 1080-1083.

[5] Wang Zhiqiang, Liu Jun, "A Review of Object Detection Based on Convolutional Neural Network", 36th Chinese Control Conference, Volume: 7, Issue: 26-28 July 2017, pp 11104- 11109.

[6] Juidong Yang, Jianping Li, "Application of Deep Convolutional Neural Network", Institute of Electrical and Electronics Engineers (IEEE), Volume: 7, Issue: July 2017, pp 229-232.

[7] Neena Aloysius, Geetha M, "A Review on Deep Convolutional Neural Networks", International Conference on Communication and Signal Processing, Volume: 4, Issue: 6-8 April 2017, pp 0588-0592.

[8] Saad Albawi Tareq Abed Mohammed, Saad Al-Zawi, "Understanding of a Convolutional Neural Network", Institute of Electrical and Electronics Engineers (IEEE), Volume: 1, Issue: January 2017.

[9] http://www.google.com