Convolution Neural Networks in Image Processing

Sonal Borase1, Javed Shaikh2, Anupama Deshpande3
[1]PhD Scholar, JJT University, Jhunjhunu, India,
[2]Associate Professor, SKN Singhgad Institute of Technology & Science, Lonavala, India,
[3]Professor, JJT University, Jhunjhunu, Rajasthan, India.

Abstract: Convolutional Neural Network (CNN) is subset of artificial neural networks. It has become prominent in many computer vision applications. CNN is designed in such a way that it should learn automatically. Spatial hierarchies of features are used in algorithms through backpropagation by using multiple building blocks such as convolution layers, pooling layers, and fully connected layers. This article offers a perspective on the basic concepts of CNN and its applications in image processing domain. Convolution neural network have been successful in identifying faces, objects and traffic signs apart from powering vision in robots and self-driving cars. CNN provides a very exciting alternatives and other application which can play important role in today's computer science field. There are some Limitations also which are mentioned.

Index Terms - Artificial Neural Network, ANN, Convolutional Neural Network, CNN, LeNet-5 Structure, AlexNet Structure, VGG-16 Structure, Inception-v.

I. Introduction

When we look at the animal, we can identify the animal by their features. Similarly we can do that using computer, for the classification of image Convolution Neural Network (CNN) is useful [1]. A Convolution Neural Network is a Deep Learning algorithm which can take in an input image, assign importance to various objects in the image. Then it will be able to differentiate one from the other. CNN will classify the image and computer can identify the animal. This paper gives the review of Convolution Neural Network application in Image processing domain. We can say if a general neural network is inspired by a human brain, the convolutional neural network is inspired by the visual cortex system, in humans and other animals. A convolutional layer is one of the Convolution neural network. Convolutional neural network can also made up of pooling and dense layers along with convolution layer [2].

II. Comparison of ANN and CNN

A neural network is a linear combination of many layers. A combination between the previous layer's output and the current layer's weights and then it passes data to the next layer by passing through an activation function. Inputs of ANN are processed only in the forward direction so ANN is also known as a Feed-Forward Neural network. Input, hidden and Output, these are the layers of ANN Fig 1, the input is accepted by input layer, this input is processed by hidden layer and then the final production of result is by output layer. Essentially, each layer tries to learn certain weights [3]. ANN is a collection of connected units which can pass a signal from a unit to another. The number of units, their types, and the way they are connected to each other is called the network architecture.

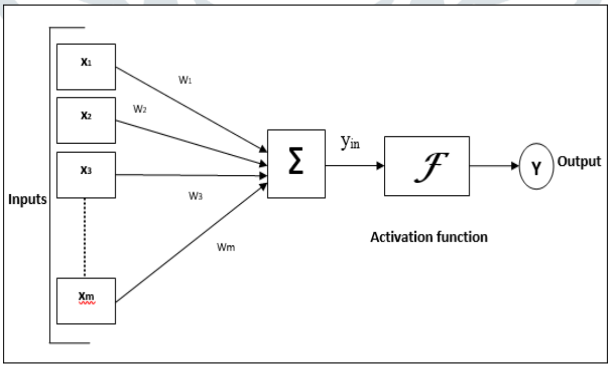

Fig.1. Artificial Neural Network

Convolutional Neural Network has convolution a mathematical operation between the previous layer's output and the current layer's kernel and then it passes data to the next layer by passing through an activation function. Fig2.shows the different layers of convolution neural network. Convolution neural network is a deep learning algorithm. A CNN has one or more layers of convolution units. From the previous layer, a convolution unit receives its input from multiple units which together create a proximity. Therefore, the input units share their weights.

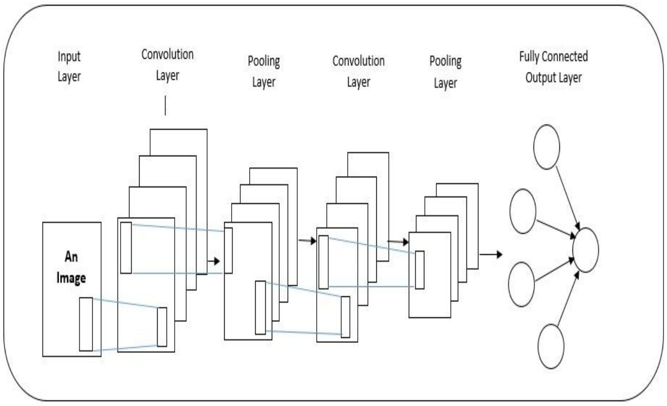

Fig.2. Convolution Neural Network Layers

III. RELATED WORK

This section discusses some works that helped us a lot during our research. Muhammad Abdullah et al [4], proposed method proves that using a well-trained CNN followed by RNN is equally effective for Video Facial Expression Recognition as for other similar tasks e.g. Action Recognition. Here CNN is trained to detect expressions from human faces. Error Correcting Output Code provides the CNN with the ability to reject or correct output errors to reduce character insertions and substitutions in the recognized text. Using these code words instead of letter images as the CNN target outputs makes it possible to construct an OCR for a new language without designing the letter images as the target outputs [5]. By analyzing independent visual modalities and their fusion with CNN and LSTM models, provides baseline recognition results of different pain levels [6]. This fusion of modalities helps to enhance recognition performance of pain levels in comparison to isolated ones. In paper [7], we present a multi-modal regression model that uses a convolutional neural network for recovering audio-visual synchronization of single-person speech videos. Convolutional Neural Networks in paper by Aiswarya S Kumar and Elizabeth Sherly [8] have been applied for recognizing the category of the principal entity in an image. That overall performance can be further improved by attaching the pre-trained CNN to an SVM.

IV. DIFFERENT MODELS OF CNN

There are different models of Convolution Neural Networks. There are two basic types of models of CNN. One is a mechanistic model and other is descriptive model. The mechanistic mode is in which internal parts of the model can be mapped to internal parts of the system of interest. Descriptive models of the visual system may be one that takes in an image and outputs an object label that aligns with human labels, as they are only matched in their overall input-output relationship so in a way that has no obvious relation to the brain.

- **LeNet-5**

This mechanistic model proposes subsampling layer. Here in Fig 3, a pattern of a convolutional layer followed by an average pooling layer. The original form of this model was proposed by Yann Lecun. This pattern is repeated two and a half times, then output feature maps. They are flattened and fed to a number of fully connected layers for interpretation and a final prediction [1].

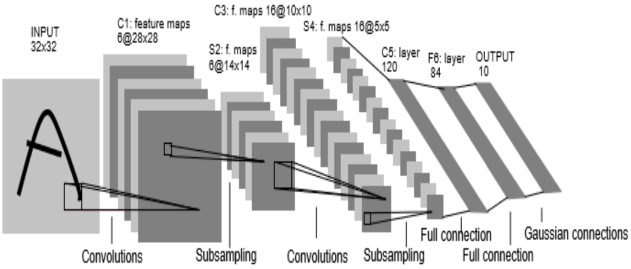

Fig.3. LeNet-5 Structure

- **AlexNet**

The architecture in Fig 4. is of AlexNet is deep. It is descriptive in nature. It was the first to show deep learning was effective in computer vision tasks. It extends upon some of the patterns established with LeNet-5. This model, Fig.4, has five convolutional layers in the feature extraction part of the model. It has three fully connected layers in the classifier part of the model [2].

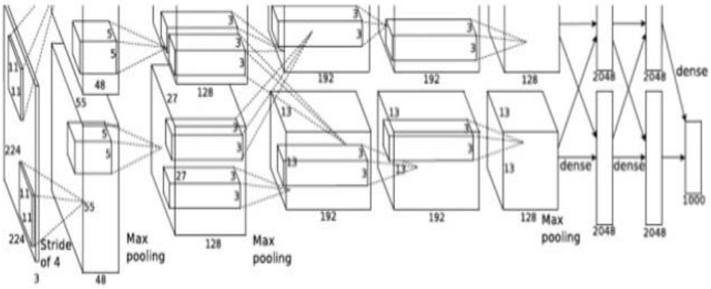

Fig.4. AlexNet Structure

- **VGG**

After AlexNet, the development of deep convolutional neural networks for computer vision tasks appeared to be a little bit of a dark art. From the name of the lab, the Visual Geometry Group at Oxford, this mechanistic model is referred to as VGG [4]. The most commonly a number of variants of the architecture were developed and evaluated from their performance and depth. There are 16 learned layers for VGG-16 and 19 learned layers for VGG-19. A schematic of the VGG-16 architecture trained on the ImageNet database is shown in Fig. 5. It can be used as a deep feature generator for producing semantic image vectors for our pavement images, when the fully-connected classifier is removed from the pre-trained VGG16 network, these semantic image vectors can then be trained and tested using another classifier like Neural Networks, Support Vector Machine for the prediction of label [3].

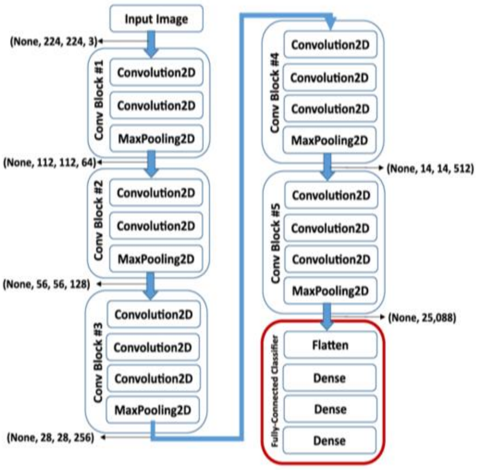

Fig.5. VGG-16 Structure

- **Inception-v1**

Inception model is a combination of different convolution layers. Fig.6 shows the Inception model where inception layer a combination of all layer. There 1×1 Convolutional layer, 3×3 Convolutional layer, 5×5 Convolutional layer. These layers with their output filter banks concatenated into a single output vector forming the input of the next stage.

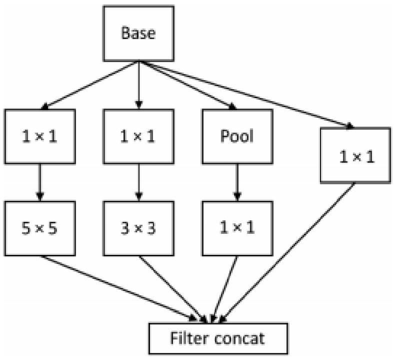

Fig.6. Inception-v1 Structure

- **Inception-v2**

Inception-ResNet-v2 is a convolutional neural network as shown in Fig.7. It has developed from Inception-v1. There is a factorization where factorize convolutions into smaller convolutions [9]. At the time of training, it reduces the computational cost but increases the memory consumption due to smaller convolutions.

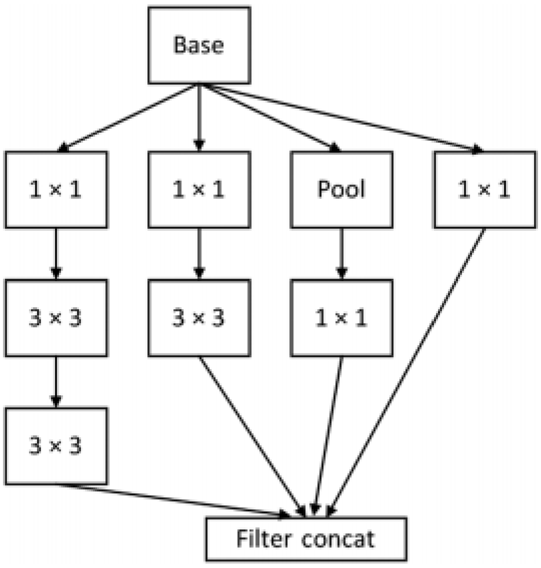

Fig.7. Inception-v2 Structure

- **Inception-v3**

Inception v3 is a widely-used image recognition model Fig.8. It is third version model in a series of Deep Learning Convolutional Architectures. The model itself is made up of symmetric and asymmetric building blocks, including convolutions, average pooling, max pooling, concatenates, dropouts, and fully connected layers [10].

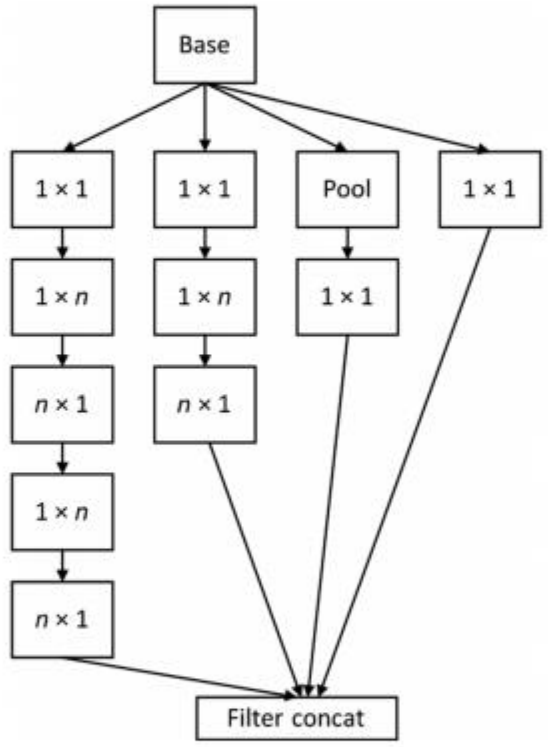

Fig.8. Inception-v3 Structure

V. Applications of CNN

The primary tasks of convolutional neural networks are to classify visual content, to recognize objects within is scenery, to gather recognized objects into clusters. On the basis of these functionality and uniqueness, it is useful in various fields [11]. The system with supervised machine learning classification algorithm, it uses CNN to deconstruct an image and identify its distinct feature. The system with unsupervised machine learning algorithm, it uses CNN to reduce the description of its essential credentials. Image classification makes the use of image tagging algorithm. This tagging includes recognition of objects and even sentiment analysis of the picture tone, this is possible by the use of CNN [12].The best example of image recognition CNN use is Medical Image Computing. The anomalies on the X-ray or MRI images with higher precision by using CNN medical image classification detection.

VI. Limitations of CNN

CNN is very dependent on human intercession. They are computationally expensive like any neural network model. It needs huge amount of data for analysis. The CNN would run into an over-fitting problem as if CNNs have millions of parameters and with small dataset because they needs massive amount of data to quench the thirst. The position and orientation of the object encoding into their predictions is not done by CNN. They completely lose all their internal data. The internal data such as the pose and the orientation of the object. They also route all the information to the same neurons. The predictions by looking at an image and then checking to see if certain components are present in that image or not, all these can be made by CNN. Then finally it classifies that image accordingly if they are present [13]. Max pooling in CNN loses valuable information. It does not encode relative spatial relationships between features. Because of this, CNN are not invariant to large transformations of the input data.

VII. Conclusion

By studying Convolution Neural Network we had concluded that as per as technology is developing day by day the need of Artificial Intelligence is increasing because of only parallel processing. In present time parallel processing is needed because with the help of this only we can save more time and money in any work related to computers and robots. In future, we have to develop more algorithms and problem solving techniques so that we can remove the limitations of the Artificial Neural Network. And if the Convolution Neural Network concepts combined with the Computational Automata we will definitely solve some limitations of this excellent technology.

References

[1] Loan N. N. Do Neda Taherifar Hai L. Vu, “Survey of neural network-based models for short-term traffic state prediction”, WILEY, 2018

[2] Gözde ÖZSERT YİĞİT , Buse Melis ÖZYILDIRIM “Comparison of Convolutional Neural Network Models for Food Image Classification”, IEEE 2017

[3] Kasthurirangan Gopalakrishnan, Siddhartha K., Khaitan b, Alok Choudhary a, Ankit Agrawal, “Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection”, ELSEVIER 2017

[4] Muhammad Abdullah, Mobeen Ahmad, Dongil Han, “Facial Expression Recognition in Videos”, IEEE

[5] Huiqun Deng, George Stathopoulos, and Ching Y. Suen, “Error-Correcting Output Coding for the Convolutional Neural Network for Optical Character Recognition”, 10th International Conference on Document Analysis and Recognition, IEEE, 2009

[6] Mohammad A. Haque, Ruben B. Bautista, Fatemeh Noroozi, Kaustubh Kulkarni, Christian B. Laursen, Ramin Irani, Marco Bellantonio, Sergio Escalera, Golamreza Anbarjafari, “Deep Multimodal Pain Recognition: A Database and Comparison of Spatio-Temporal Visual Modalities”, 13th IEEE International Conference on Automatic Face & Gesture Recognition, IEEE, 2018

[7] Toshiki Kikuchi and Yuko Ozasa, “Watch,Listenonce,Andsync:Audio-Visualsynchronizationwith Multi-Modalregressioncnn”, ICASSP, IEEE, 2018

[8] Aiswarya S Kumar and Elizabeth Sherly, “A convolutional neural network for visual object recognition in marine sector”, 2nd International Conference for Convergence in Technology (I2CT), IEEE, 2017

[9] Karen Simonyan, Andrew Zisserman,“Very Deep Convolutional Networks for Large-Scale Image Recognition”, Cornell University April 2015

[10] Wang et al,“Development of convolutional neural network and its application in image classification: a survey”, SPIE 2019

[11] J. U. Duncombe, “Infrared navigation—Part I: An assessment of feasibility (Periodical style),” *IJREAM* vol. ED-11, pp. 34–39, Jan. 1959.

[12] K. Teilo, “ An Introduction to Convolutional Neural Networks,” no. NOVEMBER 2015, pp. 0–11, 2016.

[13] I. Kokkinos, E. C. Paris, and G. Group, “Introduction to Deep Learning Convolutional Networks, Dropout, Maxout 1,” pp. 1–70.