ISSN 2278-3091
Volume 11, No.2, March - April 2022
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse011122022.pdf
https://doi.org/10.30534/ijatcse/2022/011122022

Application of Deep Convolution Neural Network for Image Classification: A Review

Mehrunnissa
Department of Computer System Engineering, University of Engineering and Technology, Peshawar, Pakistan,
mehro48@gmail.com

Received Date : February 2, 2022 Accepted Date :March 5, 2022 Published Date : April 06, 2022

ABSTRACT

With the development of big data information and success in computer vision problems, more hidden layers in CNNs give it a greater and complicated structure and more powerful characteristic. Convolutional Neural Networks (CNN) provide an opportunity for automatically gaining knowledge of the domain specific features. The convolutional neural network is model and skilled by means of the deep leaning of neural networks and the set of rules has made great achievements in computer vision considering the fact that it's a creation. This paper first explains the upward push and structure of deep learning and convolution neural network (CNN), and summarizes the structure or shape of CNN, and its different operations like convolution, feature extraction and pooling operation of convolution neural network. Development of convolution neural network model primarily based on deep learning in image classification are reviewed, an intensive literature survey of Convolution Neural Networks which is the broadly used framework of deep learning. With Alex Net or ImageNet because the base model of image classification in CNN model, we've got reviewed all the versions emerged over the years to fit various programs and a small discussion on structure and working of CNN.

Key words: Artificial Intelligence, Convolution neural network, Deep learning, Image Classification, Machine learning.

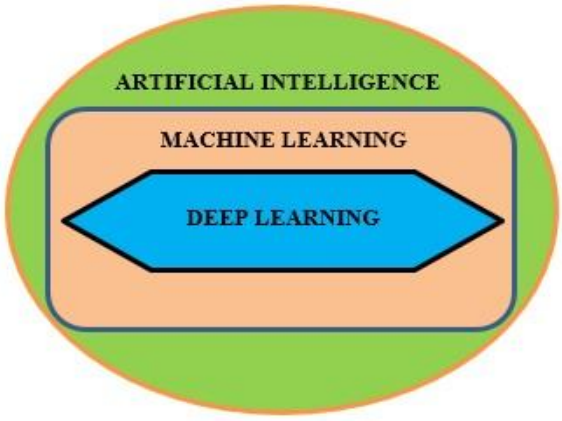
1. INTRODUCTION

Image recognition technology is a powerful type of application in artificial intelligence in which computer systems detect and recognize objects for example faces in photographs. Nowadays image recognition has already evolved into something very advanced and sophisticated, with many mobile applications like Google Photos offering easy access to search for images through face recognition or label creation. There are also industries focusing on image detail analysis for facial expressions or even brand logos. Something interesting about neural networks is how there are a lot of different types to investigate. A particularly interesting type of artificial neural network is the convolutional one. Within these convolution neural networks, the first primarily based neuron which operates in such a way that it uses the nearby connectivity between neurons and the hierarchical structure of many different nerve networks with specific outputs in such way it gets more complicated with time. When looking into such type of neural network means there are options where there will be multiple and individual output nodes, but we focus on a few main features Neural models are a type of computer system modeled after the neural systems of the human brain. They are designed to recognize complex patterns for large datasets and can be applied for image recognition on the internet or facial recognition used by security systems or police departments. The neural cognition machine decomposes a visible sample into many sub samples, and those sub sample features are processed by hierarchical cascaded characteristic planes so that the version is very good even within the case of small targets of the recognition ability of the target object. After that, the researchers started performing experiments to explain the use of an artificial neural network (a low version with simplest one hidden layer node) known as a multi-layer sensor instead of manually extracting functions and using a simple stochastic gradient descent approach to train the model, and similarly some researchers proposed a back propagation algorithm for calculating the errors of gradients, which finally turned into a very effective approach.

In 1990, LeCun et al. [1] studied the identification of handwritten digits in images and proposed training of convolution neural network models with gradient back propagation algorithm. The model that is used was the MNIST dataset, in which a convolution neural network is a multi-layer artificial neural network that is intended to handle two-dimensional input data. This means that it has multiple layers (planes) made up of multiple smaller (self-determining) neurons, which are linked to each other via their layered counterparts. This kind of revolutionary model replicates the way our brain works and how information travels through the axonal pathways to different regions enabling them to recognize what they see or hear. The proposed approach enables us to decrease the complications that might arise due to computation of training process in the network by using the weights in the time dimension and it applies a similar process to that of the training process for speech recognition which has been used widely and is still used till date.

2. Artificial Intelligence Vs Machine learning Vs Deep learning

Machine learning is a subset of artificial intelligence that provides systems the ability to automatically learn from data and improve systems. Deep Learning is a subfield of machine learning, which focuses on algorithms based on patterns in the human brain which makes it possible for deep learning models to deliver better results than traditional models. The difference between deep learning and regular machine learning is that deep learning algorithms require a lot of data to operate as they depend upon stacked layers that process in a hierarchical fashion which is why they are perfect when data is huge Machine learning algorithms can be used with less data, but this can result in decreased accuracy, while the parallel processing unit can perform similar tasks to that of a GPU. Thirdly, in machine learning and deep learning algorithms, features must be identified and put into the system for better training of the system. With more features, the performance gets better. On the other hand, because of deep learning algorithms having many parameters which take longer to adjust (than machine learning), training these algorithms can take up a lot of time compared to training a typical machine learning algorithm [2]. Figure 2 illustrates the relationship

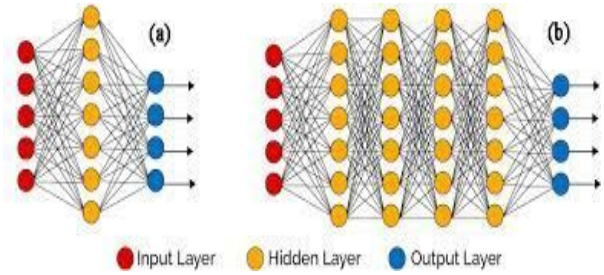


between artificial intelligence, deep learning, and machine learning.

Figure 1: Artificial Intelligence Vs Machine Learning Vs Deep Learning

3. Deep learning in image classification

Many machine learning algorithms like linear regression, logistic regression is used to train machines to make predictions from given data, but these algorithms deal with limited amount of data. The performance of these algorithms is less for huge amount of data. So, in terms of data, deep learning is a best choice. In deep learning, systems think and learn like humans by using the concept of neural network, consisting of a hierarchy of layers, whereby each layer in networks transform s the input data into more abstract representations. Between input and output layers, there are series of hidden layers which are used to identify the input features and create new features based on the data. The more layers a network has, the higher the level of features it will learn. Each next layer of DNN uses the output from the previous layer as input. As compared to DNN, Artificial Neural Network is much different which is only good at learning weights with one hidden layer but does not contain many layers and hence it cannot learn complex features [3]. Deep learning can be expensive and need huge datasets to train itself. The data and features are exponential related. For example: If you



have 10 features then you are required to provide at least 100 data values [4]. Figure 3 demonstrates the simple neural and deep neural network structure.

Figure 2: (a) Simple Neural Network, and (b) Deep Neural Networks [5]

4. Convolution neural networks

Convolution Neural network (CNN) is the most widely used deep learning framework which was inspired by means of the visible cortex of animals. It had been extensively used for recognition of objects duties but now it has been also examined in other fields like tracking of objects, pose estimation, detection of text in an image, visual saliency detection, motion recognition, scene labeling and much more [5]. Convolution Neural Networks by way of LeCun et al. in 1990 and later progressed on it and it turned into specially designed to classify handwritten digits and turned into a success in recognizing visual styles without delay from the input image and without any preprocessing. However, because of lack of sufficient training facts and computing power, this structure didn't perform nicely in complex troubles [6]. Later in 2012, Krizhevsky et al had come up with a CNN model that succeeded in bringing down the error fee on ILSVRC competition. A CNN is a special case of the neural network which contains one or more convolution layers, which are followed by one or more fully connected layers. The design of a CNN is inspired by visual mechanism, the visual cortex, in the brain. Deep learning recognizes the objects in an image by using convolution neural network. This paper specifically points out factors: synthetic neural network with multiple hidden layers has a very effective characteristic studying capacity. The features extracted through the training model have extra summary and greater basic expression of the unique records of original input data. CNN models are a class of neural networks suitable for processing grid-like topology data, which vary from 1D time-series data to 2D images [7]. CNN models rely on affine transformation, which involves a vector of inputs being multiplied by a matrix

(also called kernels, or filters) to produce an output. The multiplication by a matrix is referred to as convolution operation. Typically, a bias vector is added to the result of the matrix multiplication. Next, a non-linear function, called an activation function, is applied to the output of a forementioned operations. After the non-linear activation function, a pooling operation is typically applied [8].

5. CNN WORKING PRINCIPAL

We want from computer to be able in differentiating between all the given images and figure out the unique features that makes a car or that makes a bottle a bottle. When we look at an image of car, we can classify it as such if the picture has features that are identifiable such as type or structure. In a similar way, the computer is able perform image classification by examining the low-level features such as edges and curves, and then building up to more abstract concepts through a series of convolution layers. When a computer takes an input image, it will process its array of pixel values which depends on the resolution and size the image, it will examine a 32 x 32 x 3 array of numbers (The 3 refer s to RGB values). For example, if there is a color image of JPG format and the size is 480 × 480. Its array will be 480 × 480 × 3in which each of these numbers is given a value from 0 to 255 that describes the pixel intensity at that point [9]. When we perform image classification, these are the only inputs available to the computer. The idea is that you give the computer array of numbers, and it will output numbers that describe the image probability that has a certain class (80 for cat, .15 for dog, .05 for bird, etc.) [10]. Convolutional Neural Networks (CNN) are better models for classification and detection of objects in visual tasks, specifically traffic sign recognition (TSR). Changing the order of connections in the convolutional and pooling layers can improve the accuracy, efficiency, and cost-effectiveness of the systems [11].

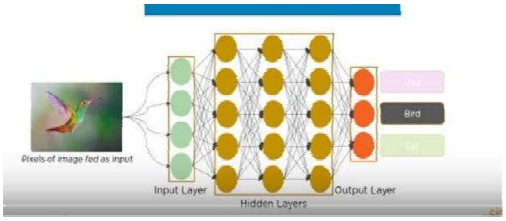


Figure 3: layers of CNN [12]

6. LAYERS OF CNN

Image classification in deep learning works on three layers

1) **Input layer**.

2) **Hidden layer**: hidden layer consists of the following layers

- Convolution layer
- RELU layer
- Pooling layer
- Fully connected layer

3) Output layer.

1) Input layer: It accepts the pixels of input image in the form of array.

2) Hidden layer: This layer carries out feature extraction by some calculations.

Hidden layer has three more sub layers:

- **Convolution layer**: This layer uses a matrix filter and performs convolution operation to detect patterns in an image. Several convolution filters are used that performs convolution operation. For example, an image 5*5 whose pixels values are 0 or 1. Sliding the filter matrix over the image and computing the dot product to detect patterns [13].

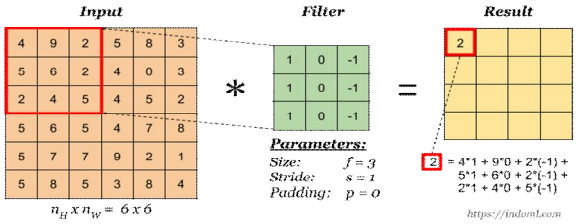


Figure 4: Convolution operation in CNN [14]

- **RELU layer**: activation function applied to convolution layer to get a rectified feature map of the image.
- **Pooling layer**: it is a down sampling operation that reduces the dimensionality of the feature map it uses multiple layers to detect edges and corner etc.

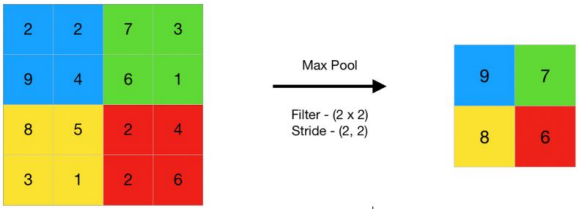


Figure 5: Pooling layer [15]

The last process after pooling is flattening. Flattening is the process of converting all the resultant to 2-dimensional array from pooled feature map into a single linear vector.

Figure 6: Array to single linear vector conversion [16]

3) **Output layer**: The final layer is the output layer that identifies the image. e.g., image of a bird was fed as an input layer. Passing through hidden layers finally the output layer recognizes the bird [17].

7. SOME CNN ARCHITECTURES

1. ImageNet:

ImageNet is a dataset of over 15 million labeled images of high-resolution that belongs to approximately 22,000 categories. The images were collected from the web and labeled by humans. The ImageNet dataset is trained through convolution neural networks which consists of many convolutions and fully connected layers. First, we must import image data set in MATLAB and then trained dataset using CNN as described above [18].

2. NASNet*:*

Zoph et al. from Google Brain proposed NASNet (Neural Architecture Search with Energy-Based Route). They followed an approach initially tracing a building block with a small dataset, CIFAR-10, a set of 60,000 32x32 color images. That was then applied to a larger dataset: ImageNet, which contained over 1 million high-resolution pictures arranged in 1350 categories. The idea we learned was the reduction of neural network layers via the Scheduled Drop Path technique which significantly improved generalization. Their architecture consisted of blocks and cells organized into "normal" and "reduction" cell layers where the former returned a larger feature map for each layer and the latter returned smaller feature maps in four instances by two. Consequently, NASNet obtained classification accuracies up to 82.7% on the ImageNet dataset while using fewer parameters.

3. AlexNet:

This structural design of Alex Krizhevsky, Ilya Sutskever & Geoff Hinton, which is in a convolutional neural network with the field of computer vision and was the very first popular work of this kind. The network is related to LeNet but instead by alternating convolution layers and pooling layers AlexNet had all the convolution layers stacked collectively. Although this network is much bigger and deeper than LeNet, though it were able to win the ILSVRC-2012 competitions with high level of accuracy [19].

4. ZFNet*:*

Zeiler and Fergus proposed the ZFNet (2013) framework which improved some of the basic design choices made in AlexNet, thereby bringing down the error rate significantly. While there were several improvements to partial layer features already in play (e.g., ResNets), it was the authors of ZFNet who glued a few pieces together to reach a breakthrough by looking at deeper features of the pixel domain. Specifically, they de-convolved their convolutions with noise addition in order to visualize layers and unpooled to "see" deeper into their networks. Based on those insights, using their novel approach for learning filters (called hard expansion), they reduced overall classification error rate by quite some margin [20].

5. GoogleNet:

The Inception architecture from Szegedy et al from Google was awarded with the top score at the fourth annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) for classifying objects in images and has become a model for deep learning researchers to follow. First introduced in Google's GoogLeNet, it has 20 layers of modules known as 'Inceptions' that have much fewer parameters than earlier models. The success of the GoogLeNet Inception model paved the way for more refined versions that are currently being developed: Inception-v1, Inception-v2 and eventually Inception-v3[21].

6. LeNet

By LeCun et al. In 1990 this was one of the new works in convolution Neural Networks and later it was advanced in 1998. In their work, the tasks of Handwritten Digit accomplished the use of convolution neural networks. It is an application for analyzing zip codes, digits etc but the lack of excessive computing machines at that point caused a break in the use of CNN [21][22].

8. CONCLUSION

This paper provides a comprehensive review of the literature on the contemporary state of the art of Convolutional Neural Networks for image classification and detection. Layer based details of CNNs are outlined. Even though CNNs are successful in many applications, there is no hypothetical evidence to explain the reasons why it performs so well. But in this work, a proper review and use of convolution neural networks for computer vision has been done. CNN models that feature low model sizes and achieve satisfactory accuracy have a promising future. It is our strong belief that with the necessary real-time hardware capabilities and successful resolution of design constraints, CNNs would enable a safer and more efficient self-driving cars and a rise in their popularity. The entire structure and functioning of CNN are not captured in this paper but satisfactory efforts have been made in the direction of it. Hopefully this article will be useful for vision researchers beginning to work on convolution neural networks.

REFERENCES

1. Y. LeCun, C. Cortes. MNIST handwritten digit database [EB / OL]. Http://yann.Lecun.Com / exdb /mnist.2010.
2. Ramachandran R, Rajeev DC, Krishnan SG, P Subathra, Deep learning an overview,

IJAER,Volume 10, Issue 10, 2015, Pages 25433-25448.

3. J. Feng, J. Liu, and C. Pan, "Complex Behavior Recognition Based on Convolutional Neural Network: A Survey," 2018 14th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), 2018, pp. 103-108, doi:10.1109/MSN.2018.00024.
4. D. H. Hubel and T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, The Journal of physiology, 1968.
5. Y. Dong, Q. Liu, B. Du and L. Zhang, "Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification," in IEEE Transactions on Image Processing, vol. 31, pp. 1559-1572, 2022, doi: 10.1109/TIP.2022.3144017.
6. M. Jaderberg, K. Simonyan, A. Zisserman. Spatial transformer networks. Advances in Neural Information Processing Systems. Montréal, Canada: [s. N] 2015: 2008-2016.
7. F. M. Bianchi, J. Grahn, M. Eckerstorfer, E. Malnes and H. Vickers, "Snow Avalanche Segmentation in SAR Images with Fully Convolutional Neural Networks," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 75-82, 2021, doi: 10.1109/JSTARS.2020.3036914.
8. Alex Krizhevsky, Sutskever I, and Hinton G.E, Imagenet classification with deep convolution neural networks. In NIPS, 2012.
9. M.Mandal,"Introduction to convolution neural networks(CNN)" by free encyclopedia towards data,May 1, 2021, available at-https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/ (accessed on 23,March,2022).
10. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, Going deeper with convolutions, CoRR, 2014.
11. T. Turay, T. Vladimirova: Toward Performing Image Classification and Object Detection with Convolutional Neural Networks, School of Engineering, University of Leicester, Leicester LE1 7RH, U.K.
12. A.Choulwar "The Art of convolution neural network" by free encyclopedia towards data,April 12,2019 available at-https://medium.com/@achoulwar901/the-art-of-convolutional-neural-network-abda56dba55c(Accessed on 23,March,2022).
13. "CNN| Introduction to pooling layer" by free encyclopedia towards data,29,July 2021, available at-https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/ (Accessed on 23,March ,2022).
14. P.Kongslip "CNN:Step3-Flattening" by free encyclopedia towards 22,July,2019, available at-https://medium.com/@PK_KwanG/cnn-step-2-flattening-50ee0af42e3e. (Accessed on 23, March,2022).
15. Deepika Jaswal, Sowmya.V, K.P.Soman ''Image Classification using CNN'' International Journal of Scientific & Engineering Research, Volume 5, Issue 6, June-2014 ISSN 2229-5518.
16. D. W. Ruck, S. K. Rogers, M. Kabrisky. Feature selection using a multilayer perception Journal of neural Network Computing, 1990, 2 (2): 40-48.
17. Toshev and C. Szegedy, Deep pose: Human pose estimation via deep neural networks, in CVPR, 2014.
18. J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, Decaf: A deep convolution activation feature for generic, 2014**.**
19. J. Fan, W. Xu, Y. Wu, and Y. Gong, Human tracking using convolution neural networks, Neural Networks, IEEE Transactions, 2010.
20. M. Jaderberg, A. Vedaldi, and A. Zisserman, Deep features for text spotting, in ECCV, 2014.
21. Syed Hussain,Munther Abulkibash,Samir Tout A Survey of traffic Sign recognition Systems Based on convolution Neural Networks, Michigan University, 2018.
22. Ali Yazdizadeh, Zachary Patterson, and Bilal Farooq, Ensemble Convolutional Neural Networks for Mode Inference in Smartphone Travel Survey, IEEE transactions on intellegent transportation systems, VOL. 21, NO. 6, June 2020.