International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-7, Issue-8, August 2021
ISSN: 2395-3470
www.ijseas.com

Detailed Investigation on Convolutional Neural Network in Deep Learning

Subhana T B[1], Prof. Shamna A R[2]

[1] Musaliar College of Engineering, Trivandrum, India

[2] Musaliar College of Engineering, Trivandrum, India

Abstract

Deep learning has become an interesting field for researchers in recent times. Convolutional Neural Network (CNN) is a deep learning approach that is popularly used for solving complex problems. The limitations of traditional machine learning approaches are overcome by the development of CNN. It has become popular in image recognition, object detection, and speech recognition. The purpose of this study is to provide knowledge and understanding of the various aspects of Convolutional Neural Network. CNN is designed to automatically and adaptively study spatial hierarchies of features through backpropagation by using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers. This study explains the different layers of Convolutional Neural Network and how these layers work.

Keywords: *Deep Learning, Convolutional Neural Network, Convolution, Pooling*

1. Introduction

Deep learning algorithms have made great strides in the field of computer vision. Deep learning is the implementation of artificial neural networks (ANNs) with multiple hidden layers to mimic the functions of the human cerebral cortex. Layers of deep neural network extract multiple features, thus providing multiple levels of abstraction. The Conventional Neural Network (CNN) is a well-known in-depth learning architecture inspired by the natural visual perception mechanism of living things. Instead of handcrafted features, convolutional neural networks are used to automatically learn a hierarchy of features that be used for classification purposes. This is achieved by successively convolving the input image with learned filters to build up a hierarchy of feature maps.

In 1959, Hubel & Wiesel discovered that cells in the animal visual cortex are responsible for detecting light in receptive fields. Inspired by this discovery, Kunihiko Fukushima proposed the neocognitron in 1980, which could be considered as the predecessor of CNN. The neocognitron model consists of "S-cells" and "C-cells". The "S-cells" sit on the first layer of the model and are connected to the "C-cells" on the second layer of the model and its overall idea is to capture the “simple-to-complex” concept and turn it into a computational model for visual pattern recognition. In 1980, LeNet was the first work in modern Convolutional Neural Networks by LeCun et al. It was specifically designed to classify handwritten digits and was successful in recognizing visual patterns directly from the input image without preprocessing. But due to lack of adequate training data and computing power, this architecture failed to work well on complex problems.

Since 2006, several methods have been developed to overcome the difficulties encountered in training deep CNNs. Most notably, Krizhevsky et al. proposed a classic CNN architecture named AlexNet and showed significant improvements over previous methods in the image classification task. The overall architecture of this method is similar to LeNet but with a deeper structure. With the success of AlexNet, many works like ZFNet, VGGNet, GoogleNet and ResNet have been proposed to improve its performance. The trend observed from the evolution of architecture is that networks are getting deeper e.g., ResNet is about 20 times deeper than AlexNet and 8 times deeper than VGGNet. By increasing depth, the network can reach the target function easily with increased nonlinearity and get better feature representations. However, this increases the complexity of the network, which makes the network be more difficult to optimize and easier to get overfitting.

2. Deep Learning

Deep learning is a computer software that mimics the network of neurons in a brain. It is a subset of machine learning and is called deep learning because it makes use of deep neural networks. Deep Neural Networks (DNNs) are such types of networks where each layer can perform complex operations such as representation and abstraction that make sense of images, sound, and text. Deep learning has become the fastest-growing field in machine learning, and it is being used by increasingly more companies to create new business models.

Deep learning algorithms are constructed with connected layers.

- The first layer is called the Input Layer
- The last layer is called the Output Layer
- All layers in between are called Hidden Layers.

Each Hidden layer is composed of neurons that are connected to each other. The neuron will process and then propagate the input signal received from the layer above it. The strength of the signal given by the neuron in the next layer depends on the weight, bias and activation function. The network consumes large amounts of input data and runs them through multiple layers thus, the network can learn increasingly complex features of the data at each layer.

2.1 Types of Deep Learning Network

Artificial Neural Network (ANN): Artificial Neural Network, is a group of multiple perceptron/ neurons at each layer. ANN is also known as a Feed-Forward Neural network because inputs are processed only in the forward direction. In this, all the perceptrons are organized within layers, such that the input layer takes the input, and the output layer generates the output. The middle layers do not link with the outside world and therefore it is named as hidden layers. Each of the perceptrons contained in one single layer is associated with each node in the subsequent layer. It can be concluded that all of the nodes are fully connected. It does not contain any visible or invisible connection between the nodes in the same layer. There are no back-loops in the feed-forward network.

Recurrent neural networks (RNNs): Recurrent Neural Network are yet another variation of feed-forward networks. Here each of the neurons present in the hidden layers receives an input with a specific delay in time. The Recurrent neural network mainly accesses the preceding info of existing iterations. It not only processes the inputs but also shares the length as well as weights crossways time. It does not let the size of the model to increase with the increase in the input size. However, the only problem with this recurrent neural network is that it has slow computational speed as well as it does not contemplate any future input for the current state. It has a problem with reminiscing prior information.

Convolutional neural networks (CNN): CNN is a multi-layered neural network with a unique architecture designed to extract increasingly complex features of the data at each layer to determine the output. CNN's are well suited for perceptual tasks. To achieve the best accuracy, deep convolutional neural networks are preferred more than any other neural network.

3. Convolutional Neural Network

A Convolutional Neural Network is a Deep Learning algorithm which can take in an input image, assign importance to various aspects in the image and can differentiate one from the other. The pre-processing required in a CNN is much lower as compared to other classification algorithms. The name “convolutional neural network” indicates that the network undergoes a mathematical operation called convolution in place of general matrix multiplication in at least one of their layers. CNN is inspired by the organization of animal visual cortex and designed to automatically and adaptively learn spatial hierarchies of features, from low- to high-level patterns. CNN architecture is composed of three types of layers: convolution, pooling, and fully connected layers. The convolution and pooling layers perform feature extraction whereas the fully connected layer maps the extracted features into final output such as classification. A convolution layer plays a key role in CNN, which is composed of a stack of mathematical operations, such as convolution and a specialized type of linear operation.

3.1 CNN Architecture

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of a series of convolutional layers that convolve with a multiplication or other dot product. The activation function is commonly a ReLU layer, and is subsequently followed by additional convolutions such as pooling layers, fully connected layers and normalization layers, referred to as hidden layers because their inputs and outputs are masked by the activation function and final convolution.

Steps in Convolutional Neural Network

- Step 1: Convolution
- Step 1b: ReLU Layer
- Step 2: Pooling
- Step 3: Flattening
- Step 4: Full Connection

Convolutional Layer: The convolutional layer is the core building block of a CNN. The layer's parameters consist of a set of learnable filters or kernels, which have a small receptive field, but extend through the full depth of the input volume. During the forward pass, each filter is convolved across the width and height of the input volume, computing the dot product between the entries of the filter and the input and producing a 2-dimensional activation map of that filter. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input. The three elements that enter into the convolution operation are Input image, Feature detector or Filters or Kernels Feature map or Activation Map

Convolution is a linear operation that involves the multiplication of an array of input data and a two-dimensional array of weights, called a filter or a kernel. The filter is smaller than the input data and the dot product is applied between a filter- sized patch of the input and the filter. The filter is applied systematically to each overlapping part or filter-sized patch of the input data, left to right, top to bottom. The output obtained by multiplying the filter with the input array once is a single value. As the filter is applied multiple times to the input array, the result is a two-dimensional array of output and is called feature map. Convolutional neural networks develop multiple feature detectors and use them to develop several feature maps which are referred to as convolutional layers. Through training, the network determines important features in order to scan images and categorize them more accurately.

Rectified Linear Unit (ReLU): It removes negative values from an activation map by setting them to zero. ReLU increases the nonlinear properties of the decision function and of the overall network without affecting the receptive fields of the convolution layer. In the rectification process all the black elements are removed from it and only those carrying a positive value i.e., the grey and white colors are retained.

Pooling Layer: The pooling also known as down sampling layer is responsible for reducing the spacial size of the activation maps. This helps to decrease the computational power required to process the data through dimensionality reduction. Also it is useful for extracting dominant features, thus maintaining the process of effectively training of the model. There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel. On the other hand, Average Pooling returns the average of all the values from the portion of the image covered by the Kernel. The Convolutional Layer and the Pooling Layer, together form the i-th layer of a Convolutional Neural Network. Depending on the complexities in the images, the number of such layers may be increased for capturing low-levels details.

Flattening: Flattening is the process of converting the data into a 1-dimensional array for inputting it to the next layer. The output of the convolutional layers is flattened to create a single long feature vector and is connected to the final classification model, called a fully-connected layer.

Fully Connected Layer: Fully connected layers connect every neuron in one layer to every neuron in another layer. The flattened matrix goes through a fully connected layer to classify the images. The objective of a fully connected layer is to take the results of the convolution and pooling process and use them to classify the image into a label.

3.2 Different types of CNN Architecture

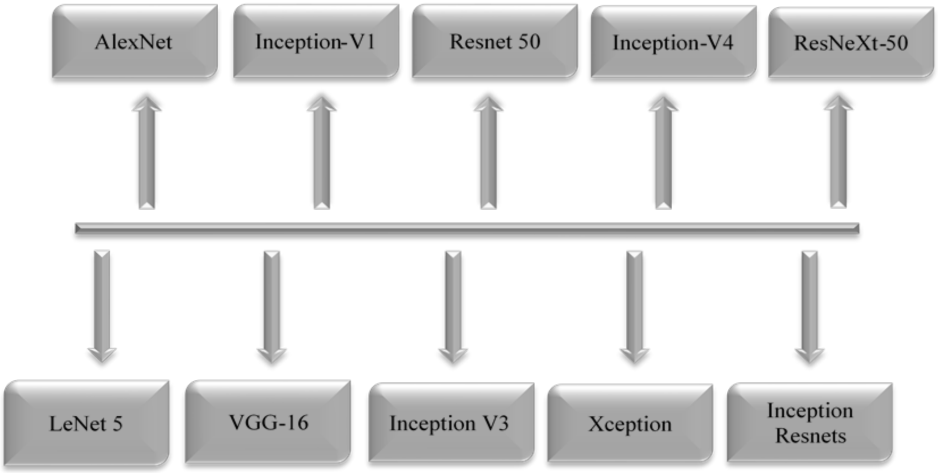

Fig 1. Different types of CNN Architecture

LeNeT 5: LeNet-5 has 2 convolutional and 3 fully connected layers. It has trainable weights and a sub-sampling layer. LeNet5 has about 60,000 parameters. It is developed by Yann LeCunn as he applied a backdrop style to Fukushima's convolutional neural network architecture.

AlexNeT: AlexNet has 8 layers, 3 fully-connected and 5 convolutional. AlexNet had 60 million parameters.. AlexNet developers successfully used overlapping pooling and Rectified Linear Units.

VGG-16 : VGG-16 has 13 convolutional and 3 fully-connected layers. It used ReLUs as activation functions, just like in AlexNet. VGG-16 had 138 million parameters.

Inception-V1: Inception-v1 had 22 layers along with 5 million parameters. The strongest feature of this network was the improved usage of computer resources inside the neural network. Instead of stacking convolutional layers atop each other, this network stacked dense modules which had convolutional layers within them.

Inception-v3: A successor to Inception-v1, Inception v-3 had 24 million parameters and ran 48 layers deep. Inception v3 could classify images into a total of 1000 categories, including keyboard, pencil, mouse, and many other animals. This model was trained on more than one million images from the ImageNet database.

ResNet-50: It consists of 50 layers of ResNet blocks, each block having 2 or 3 convolutional layers. ResNet 50 had 26 million parameters. The basic building blocks for ResNet-50 are convolutional and identity blocks.

Xception: Xception was 71 layers deep and had 23 million parameters. It was based on Inception-v3.. Xception practically is a CNN based solely on depth-wise separable convolutional layers.

Inception-v4: With 43 million parameters and an upgraded Stem module, Inception-v4 is touted to have a dramatically improved training speed due to residual connections. It is developed by Google researcher. Inception v4 had undergone uniform choices for each grid size. It has deeper network, Stem improvements, and the same number of filters in every convolution block.

Inception-ResNets: The Inception-ResNet had 25 million parameters and 32 towers. It was a combination of Inception v4 and ResNet-50.

ResNeXt-50: At 50 layers deep and sporting 25.5 million parameters, ResNeXt-50 was trained on more than a million images from the ImageNet dataset. An improvement over ResNet, ResNeXt-50 displayed a 3.03% error rate with a considerable relative improvement of 15%.

4. Application Of CNN

- Decoding Facial Recognition: Facial recognition is broken down by a convolutional neural network into major components such as identifying every face in the picture, focusing on each face despite external factors, such as light, angle, pose, etc. and identifying unique features. All the collected data are compared with already existing data in the database to match a face with a name.
- Historic and Environmental Collections: CNNs are also used for more complex purposes such as natural history collections. These collections act as key players in documenting major parts of history such as biodiversity, evolution, habitat loss, biological invasion, and climate change.
- Analysing Documents: Convolutional neural networks can also be used for document analysis. This is not just useful for handwriting analysis, but also has a major stake in recognizers. For a machine to be able to scan an individual's writing, and then compare that to the wide database it has, it must execute almost a million commands a minute. Using CNN and newer algorithms, the error rate has been brought down to a minimum of 0.4% at a character level.
- Grey Areas: Introduction of the grey area into CNNs is posed to provide a much more realistic picture of the real world. Humans can understand that the real world plays out in a thousand shades of grey. Allowing the machine to understand and process fuzzier logic will help it understand the grey area us humans live in and strive to work against. This will help CNNs get a more holistic view of what human sees.
- Understanding Climate: CNNs can be used to play a major role in the fight against climate change, especially in understanding the reasons why we see such drastic changes and how we could experiment in curbing the effect.

- Advertising: CNNs have already introduced a different world of advertising by introducing programmatic purchases and personalized advertising based on data.
- Other Interesting Field: CNN is preparing for the future with driverless cars, robots that can mimic human behaviour, assistants to human genome mapping projects, earthquake and natural disaster predictions, and perhaps self-diagnosis of medical problems.

5. Advantages and Limitations of CNN

Advantages

- The main advantage of CNN is that it automatically detects the important features without any human supervision.
- CNN is also computationally efficient. It uses special convolution and pooling operations and performs parameter sharing. This enables CNN models to run on any device, making them universally attractive.
- Another reason why CNN is so popular is because of their architecture, the best thing is that it does not require feature extraction.
- Deep convolutional networks are flexible and work well on image data.

Limitations

- A Convolutional neural network is significantly slower due to an operation such as maxpool.
- 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.
- CNN requires a large Dataset to process and train the neural network.

6. Conclusion

As an outstanding representative of deep learning, CNN is now widely used in various fields. We can see that since AlexNet was proposed in 2012, research in deep learning has been extremely rapid and new technologies emerges almost every year or even every few months. New technologies are often accompanied by new network structures and deeper network training methods, and constantly creating new accuracy records in areas such as image processing, speech recognition, and natural language processing. So far, the research on convolutional neural networks is still in a period of rapid development, and the technology of CNN is changing rapidly. Of course, one of the driving forces that cannot be ignored is that we have faster GPU computing resources to experiment with, and very convenient open-source tools such as TensorFlow, Theano, etc. that allow researchers to quickly explore and try.

References

[1] Yuhang Dong, Zhuocheng Jiang, Hongda Shen, W. David Pan Lance A. Williams, Vishnu V. B. Reddy, William H. Benjamin, Jr., Allen W. Bryan, Jr., “Evaluations of Deep Convolutional Neural Networks for automatic identification of Malaria infected cells”, IEEE Conference (2017)

[2] Kewen Yan, Shaohui Huang, Yaoxian Song, Wei Liu, Neng Fan., “Face Recognition based on Convolutional Neural Network”, IEEE Conference (2017)

[3] Umme Aiman, Virendra P. Vishwakarma., “Face Recognition Using Modified Deep Learning Neural Network”, IEEE Conference (2017)

[4] Goutham Reddy Kotapalle, Sachin Kotni., “Security using image processing and Deep Convolutional Neural Networks”, IEEE Conference (2018)

[5] Jinesh Mehta, Eshaan Ramnani, and Sanjay Singh., “Face Detection and Tagging Using Deep Learning”, IEEE Conference (2018)

[6] Dongya Jia, Shengfeng Yu, Cong Yan, Wei Zhao, Jing Hu, Hongmei Wang, Tianyuan You., “Deep Learning with Convolutional Neural Networks for Sleep Arousal Detection”, IEEE Conference (2018)

[7] Bol´ıvar Chacua, Iv´an Garc´ıa, Paul Rosero, Luis Su´arez, Iv´an Ram´ırezy, Zhima Simba˜na, Marco Pusda., "People Identification Through Facial Recognition Using Deep Learning", IEEE Conference (2019)

[8] Saibal Manna, Sushil Ghildiya, Kishankumar Bhimani., "Face Recognition from video using Deep Learning", IEEE Conference (2020)

[9] "Face Recognition Based on Convolutional Neural Network", Kewen Yan, Shaohui Huang, Yaoxian Song, Wei Liu, Neng Fan. IEEE Conference (2017)

[10] "Research on Face Recognition Based on Deep Learning", Xiao Han, Qingdong Du. IEEE Conference (2018)

[11] D. H. Hubel, T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex", The Journal of physiology (1968).

[12] K. Fukushima, S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition", Competition and cooperation in neural nets', (1982).

[13] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of IEEE (1998).

[14] Jiuxiang Gua, Zhenhua Wangb, Jason Kuenb, Lianyang Mab, Amir Shahroudyb, Bing Shuaib, Ting Liub, Xingxing Wangb, Li Wangb, Gang Wangb, Jianfei Caic, Tsuhan Chenc, "Recent Advances in Convolutional Neural Networks", Science Direct (2017)

[15] Y. Sun, X. Wang, and X. Tang., "Deeply learned face representations are sparse, selective, and robust," Proc. IEEE Conf. Comput. Vis.Pattern Recog. pp. 2892-2900 Jun. 2015.

[16] M. Paulin, J. Revaud, Z. Harchaoui, F. Perronnin, and C. Schmid., "Transformation pursuit for image classification," In Computer Vision and Pattern Recognition (CVPR), 2014.

First Author Subhana T B, Degree: B tech in Electronics and Communication Engineering (Graduated in 2021)

Second Author Shamna AR, Degree: BE, ME, Pursuing PhD, Associate Professor, Department of Electronics and communication, Musaliar college of engineering chirayinkeezhu.