Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics (CONF-CIAP 2023)
DOI: 10.54254/2753-8818/5/20230387

The Study for Convolutional Neural Network and Corresponding Applications

Xingchen Chen*

School of Engineering, University of Edinburgh, City of Edinburgh, the United Kingdom

s2269664@ed.ac.uk

Abstract. Deep learning is a key technological tool in the field of image identification with wide application prospects because of its significant benefits in feature extraction and model fitting. Deep learning entails numerous stages of non-linear transformations. The primary implementation of the current deep learning technique is the deep neural network, the connection pattern of which takes its reference from the way that the visual cortex of animals is organized. Among all the deep learning methods, Convolutional Neural Network (CNN) is one of the most renowned means to process image. It has excellent performance in terms of large-scale image processing. A convolutional neural network consists of convolutional layers, a fully connected layer, a pooling layer, and associated weights. Convolutional neural networks have fewer parameters to consider than other deep, feed forward neural networks, making them an attractive deep learning architecture. There are several traits in CNN, including pooling, shared weights, and local connections. With the help of these features, the network's complexity and the quantity of training parameters can be decreased, and the model's level of invariance to scale, shift, and distortion can be increased, along with its robustness and fault tolerance. This paper firstly summarized the history of convolutional neural network, then briefly discussed the components like the neuron and multilayer perception. The main structure of the CNN is showed afterwards. The paper also mentioned the features and the applications of CNN, mainly in the field of Computer Vision (CV).

Keywords: Convolutional neural network, Deep learning, Machine learning, Image processing.

1. Introduction

Artificial Neural Networks (ANN) are algorithmic mathematical models that simulate the composition and operation of biological nervous systems while processing data in distributed parallel. By internally altering the weight relationship between neurons and neurons, ANN accomplishes its goal of processing information. A feedforward neural network made up of numerous convolutional layers and pooling layers is known as a convolutional neural network (CNN). CNN performs very well in multiple tasks, especially in image processing.

After researching the visual cortex of the cat brain in 1962, two biologists by the names of Hubel and Wiesel put out the idea of the "receptive field" [1], which would later have a significant impact on the creation of artificial neural networks. The Receptive Field is the size of the area on the input image where each layer of the convolutional neural network's feature map's pixels are mapped. The explanation that a location on the feature map correlates to an area on the input map is the more often accepted one. Fukushima, the creator of the convolutional neural network, presented a weight-sharing convolutional neural layer and a neurocognitive machine in 1980 based on the receptive field theory of biological neurology [2]. LeCun created the convolutional neural network in 1989 by fusing the weight-sharing convolutional neural layer and the back-propagation algorithm [3], and he first applied it successfully to enhance the USPS's handwritten character recognition system. LeCun again increased the accuracy of handwritten character recognition in 1998 when he developed the traditional network model LeNet-5 of the convolutional neural network [4].

The main components of a CNN are an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. Convolutional layers and pooling layers are frequently utilized, and they are alternatively set. To put it another way, a pooling layer connects to a convolutional layer, which is connected to a pooling layer, and so on. Local connections exist between each neuron and its input in the output feature map of the convolutional layer. The input value of the neuron is the sum of the local input, bias value, and matching connection weight. This method is comparable to the convolution technique, which is where CNN gets its name.

2. CNN overview

2.1. Neurons

The fundamental processing nodes of artificial neural networks is neurons. It often has several inputs and a single output, and Figure 1 shows the structural model.

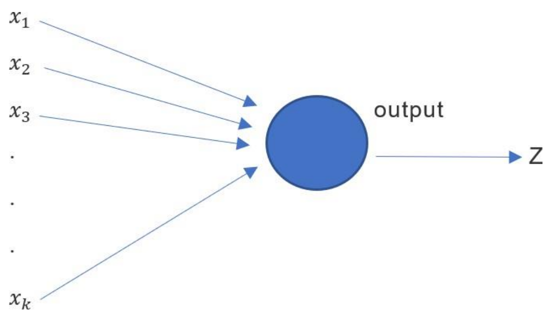

Figure 1. The structure of neurons.

x_1, x_2, $x_3$and x_k represent the input signal; k input signals are simultaneously input to the neuron. The output Z is given by formula:

$$z = \delta\,(w1x1 + w2x2 + \cdots + wkxk + b) \tag{1}$$

The connection between the input signal x and the neuron is represented by the weight values w1, w2, and wk; the internal state of the neuron is represented by b. The function of activation is. ReLU function, sigmoid function, or tanh(x) function are some examples [5].

2.2. Multilayer Perceptron

The input layer, the hidden layer (which may have one or more layers), and the output layer make up the Multilayer Perceptron (MLP) neural network model. It can resolve linear inseparable issues that a single-layer perceptron cannot resolve. A topology diagram of a multi-layer perceptron network is shown in Figure 2.

DOI: 10.54254/2753-8818/5/20230387

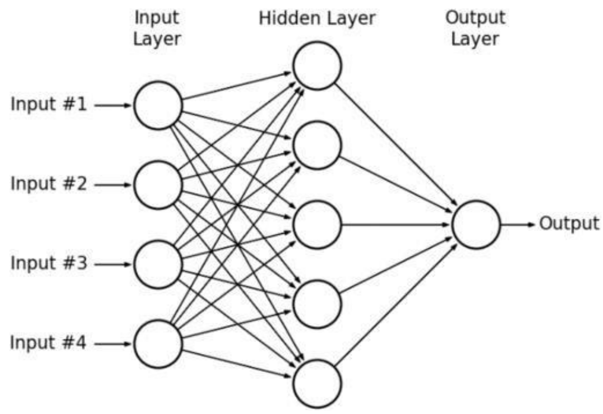

Figure 2. The structure of MLP.

2.3. Convolutional Neural Networks
A neural network with a convolutional structure is the convolutional neural network (CNN). CNN can use less memory because of its convolutional structure. Local receptive field, weight sharing, and the pooling layer are CNN's three key operations. These successfully cut down on the network's parameter count and solve the overfitting issue.

2.3.1. Network structure. The convolutional neural network is a multi-layer supervised learning neural network. The feature extraction function of the convolutional neural network is implemented by the two hidden modules, the convolutional layer and pooling layer. Iterative training is widely used to improve the network's accuracy. By using the gradient descent approach to inversely adjust the weight parameters in the network layer by layer, the network model optimizes the loss function. Alternate convolutional layers and max pooling layers make up the convolutional neural network's lower hidden layer. The fully connected layer, also known as the hidden layer and logistic regression classifier, is the top layer of a traditional multi-layer perceptron. The feature image created by feature extraction by the convolutional layer and the pooling layer is the input of the first fully connected layer. The final output layer is a classifier, which can classify the input image using logistic regression or SoftMax regression [6].

Convolutional, pooling, and full connection layers make up the convolutional neural network structure. Each layer features several feature maps and each of the layers uses a convolution filter to extract an aspect of the input, and each of which has a number of neurons.

2.3.2. Local receptive field and weight sharing. The core elements of a convolutional neural network are the pooling layer, the local receptive field, and weight sharing. These techniques enable us to reduce the complexity of the network parameters and provide a specific level of displacement, scaling, and nonlinear deformation stability for the network [6].

Each neuron only needs to sense the local aspects of the image because the spatial connection of the image is local. As a result, the receptive field is used. Additionally, by combining several local neurons that were gained by these sensations at a higher level, the global information while utilizing fewer connections can be obtained.

Because sharing among various neurons can cut down on the number of parameters that need to be solved, the weight-sharing method is employed. Using various filters to deconvolve the image will produce various feature maps. All neurons in the first hidden layer are able to recognize the same characteristics in different regions of the image thanks to weight sharing, which involves applying convolution to the image using the same convolution kernel. The convolutional network can adapt well to the small-scale movement of the image because its main capability is to detect the same sort of characteristics in multiple positions.

2.3.3. Convolutional layer, pooling layer, and fully connected layer. The convolutional layer's functionality is extracting the information and features in the input image. These features are reflected in the pixels of the image, such as the texture and color.

The convolution operation is like the convolution in mathematics but is simpler. The image seen by the computer is a matrix of numbers. The image has several channels and consists of several matrices. The convolution operation is to perform the cross-correlation operation first from left to right and then from top to bottom through each channel's convolution kernel (the convolution kernel is usually 3×3). So, the convolution operation will also retain the position information. It is like a small window, sliding from the upper left corner to the lower right corner step by step. The cross-correlation operation means that the corresponding positions are multiplied and added together.

Max and average pooling are two common pooling operations. The pooling layer is calculated by sliding an $n \times n$ matrix window and finding the maximum or average value in the matrix.

The pooling layer's goals are to choose the features that were recovered from the convolutional layer, lower their dimension, and enhance their receptive field. It let a pixel after pooling corresponds to an area in the previous picture. Because the pooling layer is not backpropagated, the pooling method can lower the number of variables in the feature map and decrease the amount of computation.

A fully connected layer generally follows the pooling layer. The fully connected layer converts all feature matrices of the pooling layer into a one-dimensional feature vector. For picture categorization and generating output results, the fully connected layer is often placed at the end of the convolutional neural network structure [7].

From the pooling layer to the fully connected layer, the data will be trimmed and transformed into a lower-dimensional vector. The number of categories that must be output is the dimension of this vector. Then, using the SoftMax function, express each value of this vector as a probability.

3. Features of CNN

The CNN classification model differs from conventional models in that it may accept a two-dimensional picture as a direct input and output the classification outcome.

It has various benefits that conventional technology lacks, such as strong fault tolerance, parallel processing, and self-learning capabilities. As a result, samples may have excellent resolution, quick running speed, good adaptive performance, and more significant errors. Additionally, it can deal with ambiguous inference rules, complex environmental data, and imperfect prior knowledge. Through structural restructuring and weight reduction, it incorporates the feature extraction function into the multi-layer perceptron, skipping the labor-intensive image feature extraction procedure before recognition. Due to their significantly better generalization capability than earlier methods, convolutional neural networks have been employed for pattern categorization, object detection, and object recognition [7].

Using the backpropagation algorithm, CNN, a feedforward neural network, can solve the network's unsolved parameters and recover its topology from a two-dimensional image.

The feature extraction and pattern classification are entirely put into a black box to eliminate manual feature extraction when CNN is used to identify displacement, scaling, and other types of distortion invariance in 2-D or 3-D images. The parameters are then achieved by continuous optimization.

Sharing local weights has a special structure that is similar to the natural biological neural network, giving CNN distinct advantages in voice and picture processing.

4. Applications of CNN

4.1. Image classification

Before the convolutional neural network, the method for the image classification task is to extract the feature information of the picture manually, using methods like contour detection, edge detection,

LBP, HOG, and HAAR. Such an explicit method of extracting image features not only consumes much time but there are still other serious problems waiting for engineers to solve. For example, the image's features change when the light's direction changes; the object's rotation can also make a difference, leading to the failure of previous pattern recognition methods.

In the ImageNet image classification competition in 2012, the AlexNet network model won because Alex used and improved the CNN model [8]. Since then, convolutional neural network has thrived in image classification. Many express companies have begun using convolutional neural network models to recognize handwritten fonts on express orders, saving business costs as much as possible and improving their system operation efficiency. Aside from image classification, a more challenging task is to perform object recognition on the overall image. An image usually has more than one category; for example, a picture may include multiple categories like a dog, a house, and a tree. In recent years, autonomous and assisted driving have become very popular, and convolutional neural networks effectively make it possible to process images collected by in-vehicle sensors.

4.2. Facial recognition

Face recognition technology is a frequently employed technological tool in daily life and is utilized extensively in e-commerce, access control systems, video surveillance, and other areas. Face image capture, face positioning, face recognition pre-processing, identity verification, and other elements make up a face recognition system. The functional modules of the face recognition system are primarily comprised of face capture and tracking, face recognition comparison, face modelling and retrieval, person identification, and image quality detection during the actual development process. One of the key technologies in the face recognition system, which has a direct impact on the system's accuracy, is face recognition comparison. Prior to CNN, face comparison and recognition algorithms were primarily dependent on PCA, LDA, and LBP and their upgraded algorithms, which were slow and did not achieve the system's anticipated recognition efficiency. Researchers have put forth a number of facial recognition algorithms since the CNN model that are more accurate than conventional algorithms.

With the help of CNN, facial recognition has become one of the most significance research fields of deep learning. DeepFace, proposed by Taigman, and DeepID, proposed by Tang Xiaoou, were relatively successful models in face recognition [9]. Google also proposed their model, FaceNet. In the LFW and YouTube database tests, the accuracy obtained by FaceNet is 99.63% and 95.12%, respectively [10].

4.3. Medical image recognition

With the continuous development of medical imaging technology, pathological identification plays an increasingly important role in medical diagnosis. Machine learning in artificial intelligence can help complete the automatic recognition of medical image diagnosis, digitally assisting the process of medical diagnosis and reducing the workload of medical workers. The convolutional neural network (CNN) is a very efficient machine learning technique that has produced remarkable results by simulating the human image identification process. Automatic recognition through CNN can reduce pathologists' workload and provide clinicians with more objective analysis results [11, 12].

5. Conclusion

In this paper, the history and structure of CNN and its applications were introduced. Although CNN has gained success in a lot of experimental measurements, much work is still worthy of further research. First, because modern CNNs are becoming more sophisticated, training them also necessitates massive databases and powerful computers. The tag database collection process involves a lot of manual labor. People are therefore eager to create an unsupervised CNN learning technique. Second, selecting the right hyperparameters poses a substantial challenge when using CNNs for novel applications. The intrinsic dependencies of these hyperparameters can make tweaking more challenging. Such as learning rate, convolution filtering kernel size, number of layers, etc., call for a great deal of expertise and knowledge. Finally, there is still no one, comprehensive hypothesis explaining CNNs. The CNN model's current operating mode is still a mystery. Right now, it would be wise to devote more time to learning CNN's fundamental operating principles. Deep CNN and computer neuroscience require more in-depth study, much as biological visual perception mechanisms motivated early CNN creation.

DOI: 10.54254/2753-8818/5/20230387

References

[1] Hubel D et al. 1968 Receptive fields and functional architecture of monkey striate cortex. The Journal of Physiology 215-243.

[2] Fukushima K et al. 1982 Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition In Competition and cooperation in neural nets (pp. 267-285) Springer Berlin Heidelberg.

[3] LeCun Y et al. 1989 Handwritten digit recognition with a back-propagation network Advances in neural information processing systems 2.

[4] LeCun Y et al. 1998 Gradient-based learning applied to document recognition Proceedings of the IEEE 86(11) 2278-2324.

[5] Nair V et al. 2010 Rectified linear units improve restricted boltzmann machines. In Icml.

[6] Gu J et al. 2018 Recent advances in convolutional neural networks Pattern recognition 77 354-377

[7] Daidaidemao 2022 A detailed introduction to CNN [Online] CSDN.

[8] Krizhevsky A et al. 2017 Imagenet classification with deep convolutional neural networks Communications of the ACM 60(6) 84-90.

[9] Taigman Y et al. 2014 Deepface: Closing the gap to human-level performance in face verification In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 1701-1708.

[10] Schroff F et al. 2015 Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 815-823.

[11] Yu Q 2019 Semantic segmentation of intracranial hemorrhages in head CT scans In 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) pp. 112-115.

[12] Sarvamangala D R 2021 Convolutional neural networks in medical image understanding: a survey Evolutionary intelligence 1-22.