Transactions on Computational and Scientific Methods | Vo. 4, No. 11, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

The Evolution of Convolutional Neural Networks: Progress, Innovations, and Future Directions

Rylan Ashford
George Mason University, United States
rylan89@gmu.edu

Abstract: Convolutional Neural Networks (CNNs) have emerged as a cornerstone of modern deep learning, with their origins rooted in biophysical models inspired by the visual system's neural mechanisms. Since the inception of early models like LeNet-5, CNNs have achieved remarkable advancements in image classification, recognition, and target localization, demonstrating their versatility and efficacy across diverse fields. This paper traces the developmental trajectory of CNNs, from the foundational principles of deep learning to the latest structural innovations exemplified by models such as AlexNet, ResNet, and DenseNet. The study also examines lightweight architectures like MobileNet and ShuffleNet, designed for deployment in resource-constrained environments. Key conceptual advancements, such as network-in-network structures and new performance-enhancing elements, are highlighted. Despite these advancements, significant challenges remain, including the computational overhead of training and the optimization of hyperparameters like convolutional kernels, depth, and learning rates. This review underscores the ongoing need for research to refine CNN architectures and ensure their efficient application in both academic and industrial settings.

Keywords: Convolutional Neural Network; Structural Evolution; Deep Learning.

1. Introduction

Convolutional Neural Network (CNN), whose early model is called a neurocognitive machine, is a biophysical model inspired by the neural mechanisms of the visual system. A convolutional neural network can be regarded as a special kind of multilayer perceptron or feedforward neural network with local connectivity and shared weights, in which a large number of neurons are organized in a certain way to produce responses to overlapping regions in the visual field. Since its introduction, convolutional neural networks have made rapid progress, often topping and achieving remarkable results in large-scale competitions for image classification and recognition, target localization and monitoring, and playing an invaluable role in other fields of intelligence. The research results achieved by convolutional neural networks in various fields have made them one of the current research hotspots.

2. Research Background

In the present day, where deep learning is a big hit, many programs have been created that are powerful enough to overturn previous perceptions of machines, such as Alpha GO, which defeated Lee Seok, and the subsequent Alpha Go Zero, which was a blue-skinned program, and Te'o, a piano robot that made a splash at the Internet Spring Festival. These are the water that previous shallow networks could hardly reach. As a result, more and more researchers have devoted themselves to deep neural networks, which have gradually become the main form of deep learning, and the convolutional neural network introduced in this paper is one of the representative structures. The article firstly describes the development of convolutional neural networks; then understands the structure and the role of each part in it; then introduces the improvement methods of convolutional neural networks and several current improved networks; finally, the article introduces the specific application areas of convolutional neural networks and concludes with the problems that need to be solved in this field.

2.1 Deep Learning

Deep learning can actually be traced back to the time when perceptual neurons were proposed, and has subsequently undergone several waves of research and has been gaining popularity among researchers in recent years. Deep learning is based on multilayer neural networks and backward propagation algorithms, and is widely used in computer vision, natural language processing, speech recognition, and other fields. Neural network is a multi-layer model that gradually abstracts the original data through different modules, but retains the main features of the source data. After continuous abstraction of the neural network, it can theoretically fit any complex function. Figure 1 shows the structure of a neural network, divided into input units, hidden units and output units. Where, Hidden units H1 denotes the first hidden layer and Hidden units H2 denotes the second hidden layer.

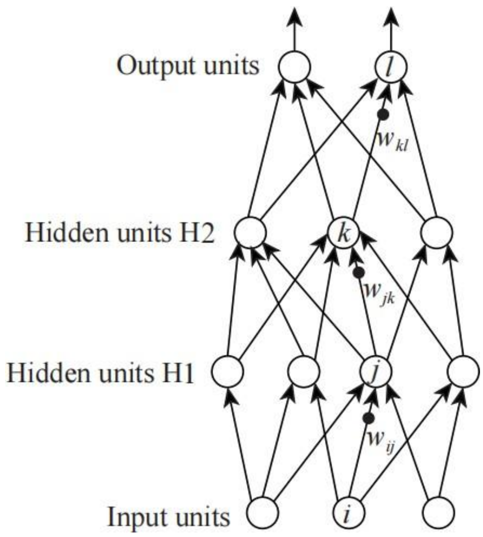

Figure 1. Structure of neural network

2.2 Back Propagation Algorithm

In 1986, Hinton et al. proposed the Backward Propagation Algorithm for error propagation, which led to the rediscovery of neural network research that had hit a low point. Essentially, the backward propagation of the gradient of the objective function in a multilayer neural network is a practical application of the chain rule in calculus. This algorithm is used to update the weights of the layers of the neural network. The input information is propagated forward from the input layer to the output layer, and the gradient of the computational error is then gradually backward propagated from the top output layer to the input layer, and once these gradients are computed the corresponding weights can be computed. The backward propagation algorithm also has its limitations. For example, it is prone to overfitting phenomena and phenomena such as gradient explosion or gradient disappearance, all of which may cause anomalies. To address these problems, many scholars have made investigations and proposed their own solutions in recent years.

3. Development of Convolutional Neural Network

Convolutional neural network, a model that was once forgotten for a while, has been able to grow to its present scale because of the efforts of countless researchers behind it. With the gradual improvement of current hardware technology and target requirements, the traditional convolutional neural network can no longer meet all the needs, so it still needs to make improvements. At present, these two aspects are mainly for structure and training method.

3.1 Improvement from the Structure

A more primitive and classical structure of convolutional neural network has been proposed above LenNet structural model, which is a more classical structure of convolutional neural network, it does not make any restrictions on the input form, convolutional kernel, etc., also this form is more convenient for later innovators. At present, the more mature improvements are divided into four main types: improvement from network input, improvement from feature fusion, restriction of convolutional kernel, and combination with other classifiers.

3.1.1 Improvement from Network Input

The input of a conventional convolutional neural network is hardly processed, so that although all the features we need are present, they are not particularly efficient for the algorithm. Therefore, some researchers propose to pre-process the objects that need to be processed in advance on the input side and discard the parts that are not needed, such as simple color layering, building scale pyramids, and other operations to improve the speed within an acceptable range when conditions allow.

3.1.2 Improvement from Feature Fusion

Convolutional neural network feature extraction is actually equivalent to layer-by-layer mapping, which extracts the features we need from the input content layer by layer and filters out the unwanted parts. But there is actually a loophole: those unwanted parts actually also contain the features of the input content, and the final content alone may not be the best, so we can consider fusing the results of this part of the mapping to make it closer to the results.

3.1.3 Restricted Convolution Kernel

The convolutional kernel is a fairly important part of the convolutional layer, so researchers have tried to make improvements from this aspect as well. They have even proposed to use other methods instead of convolution, such as using Gabor kernels, sparse silver snake matrices, etc. Of course, then, can it still be called a convolutional neural network or not? This is still open to debate, there is no valid method proposed yet.

3.1.4 Combination with other Classifiers

In the CNN overview, this paper introduces two important structures in the structure of convolutional neural networks: the pooling layer and the convolutional layer. In other words, the process of convolution is essentially a feature extraction process, and since it is a feature extraction, it is bound to involve classification, so we can also start from this aspect. The recall and accuracy of SVM as a classifier are 2.9% and 2.8% higher than those of traditional CNN methods, which definitely indicates that there is a lot of room for improvement in this method.

3.2 Improvements from the Training Algorithm

Compared with structural changes, changes in training algorithms have higher requirements on their own theoretical foundations. The two main aspects currently targeted are: improvements for nonlinear mapping functions and unsupervised training of networks.

3.2.1 Improvement of Nonlinear Mapping Function

CNN is used to adjust the range of period results by a nonlinear function after each mapping layer is completed. The Sigmoid function is generally used, but later researchers found that the human eye recognizes some sparse objects better than others, so a sparse representation is applied with the CNN to make it more closely match the human eye display. One of the best examples in this regard is the positive linear unit mentioned above, which has the effect of making the computed value of convolution 0 when it is less than 0, and constant if it is greater than 0, making the result obtain a good sparsity. Most of the current convolutional neural networks are used in the fully connected layer to improve their performance.

3.2.2 Unsupervised Training Algorithm

The original CNN required human labeling of data, and in the initial stage of artificial neural network, it was mainly for handwritten digit classification with less data volume, and this way could really save training time, but as the CNN continued to improve its function was not only limited to this, but also the license plate recognition, face recognition and even medical image recognition, which were difficult to get good training results just by the initial data volume. Therefore, it is obvious that artificial labeling of data is not feasible in the face of huge amount of data. Therefore, unsupervised training algorithms are an important and successful step in CNN improvement, and this feature was first adopted by J Ngiam et al. in 2011 for the sparse filtering algorithm. Specifically, they construct a feature distribution matrix and solve the sparsification problem once according to the direction of the features, and then normalize each sample feature by L2 parametrization to obtain a sample distribution matrix with features such as sample sparsity and activation time sparsity, etc. The unsupervised deep learning model can be formed by multi-layer cascade expansion of these sample distribution matrices.

4. Applications

Convolutional neural networks have unique advantages for image processing and computer vision, and play an important role in pattern recognition, target detection, target segmentation, and image classification. In addition, convolution, as a common mathematical operation, can also be used for natural language processing, signal processing, and image classification. In recent years, with the development of multimodal theory, convolutional neural networks and recurrent neural networks are often used in combination to build "video-image-text" models.

5. Conclusion

This paper composes the development line of research related to convolutional neural networks, from the proposal of deep learning to the development of convolutional neural networks today, the changes in the structure and methods of convolutional neural networks. It can be seen that from the earliest LeNet-5, AlexNet and later ResNet, DenseNet and other models with large scale and high accuracy, the research on convolutional neural networks has evolved, as well as some easy-to-deploy models with low overhead such as MobileNet and ShuffleNet. In this process, some significant ideas such as network-in-network have been born and some new elements have been introduced to improve the performance of the models. However, the training overhead of convolutional neural networks is still not small, and the selection of appropriate convolutional kernel, number of layers, and learning rate still needs to be studied more deeply in the industry.

References

[1] R X Wu, Q K Xiao. International Journal of Electronic Measurement Technology, 2019, 38(05): 86-90.

[2] W J Chen, P Zhou, H Y Yang, et al. A saliency detection model for the mechanism of channel-spatial joint attention [J]. Computer Engineering and Applications, 2021, 57(19): 214-9.

[3] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]//Proceedings of International Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.

[4] Yang Z, Wu B, LIAN W W, et al. Face recognition based on attention mechanism and depth identity mapping [J]. Sensors and Microsystems, 2020, 39(09): 150-3.

[5] N SRIVASTAVA, G E HINTON, A KRIZHEVSKY, et al. Dropout: a simple way to prevent neural networks from overfitting. [J]. Journal of Machine Learning Research, 2014, 15(1).

[6] M C FU. Simulation-Based Algorithms for Markov Decision Processes: Monte Carlo Tree Search from AlphaGo to AlphaZero [J]. Asia-Pacific Journal of Operational Research, 2019, 36(06).

[7] X Zhang, X Huang, W H Zhong, et al. FPGA implementation of Sigmoid function and its derivative function [J]. Journal of Fujian Normal University (Natural Science Edition), 2011, 27(02): 62-5.

[8] S P R BORRA, N V S S PRADEEP, N T S RAJU, et al. Face Recognition based on Convolutional Neural Network [J]. International Journal of Engineering and Advanced Technology (IJEAT), 2020, 9(4).

[9] Li Hui, Shi Bo. Face recognition algorithm based on convolutional Neural Network [J]. Software Guide, 2017, 16(03): 26-9.

[10] Forest in time. research on Deep Convolutional Neural Network Based on Sequence Pooling [J]; Journal of Anhui Agricultural University (Natural Science edition); 2006-06 Zhengzhou University, 2017.

[11] Y D Chen, L M Wang. face recognition Method based on convolutional Neural Network. journal of Northeast Normal University (Natural Science), 2016, 48(02): 70-6.

[12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv: 1409.1556, 2014.

[13] Y Ge. Research on Remote Sensing Image Retrieval based on CNN Migration Feature Fusion and Pooling [J]; Journal of Anhui Agricultural Sciences; 2006 -01 Nanchang University, 2019.

[14] Zhang Y N, QU L, Chen J W, et al. Multi-input Sigmoid excitation function neural network weight and structure determination method [J]. Application Research of Computers, 2012, 29(11): 4313-6 +51.