ISSN 2616-5775 Vol. 6, Issue 4: 9-13, DOI: 10.25236/AJCIS.2023.060402

Research on Improved Algorithm of Image Classification Based on Convolutional Neural Network

Hongyan Liu

Chongqing Jiaotong University, Chongqing, 40074, China
3295703284@qq.com

Abstract: *With the continuous exploration of artificial intelligence, the convolutional neural network, as one of the representative algorithms, has also developed rapidly. The convolutional neural network extracts more high-dimensional and abstract features from the data, summarizes the distributed feature representation of the data, and discovers complex nonlinear relationships. Due to the rapid increase in the amount of calculations in the era of big data, the structure of convolutional neural networks is also more complex, so the difficulty of computing tasks continues to increase. Aiming at these difficulties, this paper optimizes the convolutional neural network model AlexNet. This paper first introduces the basic principles of artificial neural networks and related technologies of convolutional neural networks, and analyzes the development prospects and research directions of convolutional neural network algorithms. Then introduce the convolutional neural network model AlexNet, and analyze and summarize its shortcomings.*

Keywords: *CNN, AlexNet, Picture detection*

1. Introduction

With the continuous exploration of artificial intelligence, the convolutional neural network has been one of the most representative algorithms that has seen rapid development in recent years. One of the key strengths of CNNs is their ability to extract high-dimensional and abstract features from the data, enabling them to identify complex nonlinear relationships between data. However, with the exponential growth in the amount of data being generated, the structure of CNNs has become increasingly complex, resulting in more computational challenges[1].

In this essay, we will explore the applications of Convolutional Neural Networks in image detection, and discuss their strengths and limitations. Additionally, we will discuss the development prospects and research directions of convolutional neural network algorithms. Furthermore, we will analyze the convolutional neural network model AlexNet, one of the seminal models in the field of computer vision, and evaluate its strengths and shortcomings. Finally, we will propose an improved network model that addresses some of the limitations of AlexNet, with the aim of further enhancing the accuracy and efficiency of CNNs for image detection tasks.

2. Background

The development of Convolutional Neural Networks (CNNs) can be traced back to the early 1980s when the idea of using neural networks to process visual information was first proposed. However, due to limited computing resources and the lack of large labeled datasets, progress in the The field was slow until the early 2010s[2]. In 2012, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) marked a turning point for the field of computer vision. In this competition, a CNN model called AlexNet achieved a top-5 error rate of 15.3%, significantly outperforming other models that had been developed up to that point. This breakthrough in the field of computer vision helped spur a wave of research focused on improving the performance of CNNs in various image detection tasks[3].

Since then, researchers have been exploring new ways to improve the performance of CNNs in various image detection tasks. One of the key areas of research has been the development of more efficient architectures, such as MobileNets and SqueezeNets, which are designed to run on low -power devices such as mobile phones and embedded systems. Another area of research has been the

ISSN 2616-5775 Vol. 6, Issue 4: 9-13, DOI: 10.25236/AJCIS.2023.060402

development of CNN models that can handle more complex tasks, such as semantic segmentation, where each pixel in an image is assigned a label based on the object it belongs to[4].

Overall, the development of Convolutional Neural Networks has enabled significant progress in the field of computer vision, particularly in the domain of image detection. Ongoing research is focused on addressing the limitations of CNNs, developing more efficient and accurate models, and abending their caps handle more complex tasks. These advancements are expected to have a significant impact on fields such as healthcare, transportation, and security, where accurate image detection is crucial for making decisions and ensuring safety.

3. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a type of artificial neural network that is specifically designed for image processing tasks. The basic building blocks of a CNN are convolutional layers, which use a set of filters to extract features from an input image[5]. These features are Then passed through a series of other layers, including pooling layers, activation layers, and fully connected layers, to generate an output that represents the class probabilities of the input image[6].

One of the key features of CNNs is their ability to learn features directly from the data, without the need for manual feature engineering. This is achieved through the use of convolutional filters, which are small matrices that are slid over the input image to extract features such as edges, corners, and other patterns[7]. These filters are learned through backpropagation, a process in which the network adjusts the weights of its neurons based on the error between its predictions and the actual labels of the training data[8].

In addition to their ability to learn features directly from the data, CNNs also incorporate several techniques to improve their performance, such as dropout regularization, batch normalization, and data augmentation. Dropout regularization helps prevent overfitting by randomly dropping urleon train out some, batch normalization normalizes the activations of each layer to improve the stability of the network. Data augmentation involves creating new training examples by applying random transformations such as rotation, scaling, and flipping to the original images, which can help prevent overproveizability and improvement[9].

Overall, the fundamentals of Convolutional Neural Networks involve the use of convolutional layers to extract features from images, along with other techniques to improve their performance and prevent overfitting. These techniques have enabled CNNs to achieve state-of-the-artide performance on a wide range of image processing tasks, including image classification, object detection, and semantic segmentation.

3.1. Convolution layer and convolution kernel

In a convolutional neural network (CNN), convolutional layers are the key building blocks for performing feature extraction on input images. Convolutional layers work by applying a set of learnable filters (also called convolutional kernels) to an input image.

A convolution kernel is a small matrix of numbers that defines the weights for the convolution operation. During training, the values of the convolution kernels are learned through backpropagation, which adjusts the weights of the kernels according to the error between the predicted output of the network and the actual output[10].

The convolution operation itself involves sliding a kernel over the input image in a systematic manner, computing the dot product at each location between the kernel and the corresponding input pixel. This produces a feature map that highlights the presence of specific features (such as edges or textures) in the input image[11]. By using multiple convolutional kernels in one layer, the network can learn to detect multiple features at different scales and orientations. The output of a convolutional layer usually passes through other layers, such as pooling and activation layers, before being fed into subsequent convolutional or fully-connected layers for further processing. Collectively, convolutional layers and kernels are key components of CNNs, allowing the network to learn and extract features directly from input images. By using multiple convolutional kernels, CNNs can learn increasingly complex and abstract input representations, thereby achieving state-of-the-art performance on various image processing tasks[12].

3.2. Pooling layer

The main function of the pooling layer is to downsample the data, reduce the spatial dimension of the feature map, remove unimportant data information with low reference value, retain most important information, and further reduce the number of parameters. Image is a data type with particularly important spatial information. Its pixel values have regional and gradual changes. According to the distribution of pixels in a certain area, the content of its adjacent area can be predicted with a high probability. This property is also called image. The "static nature". Therefore, the statistics of the pixel features of the current area can express the overall feature information of the area[13]. For example, the mean value, maximum value, sum, etc. of a certain feature information in the area covered by the statistical convolution kernel are used to represent the overall features in the range. Common pooling methods include Max Pooling, Average Pooling, and Sums Pooling. For maximum pooling, the pixel with the largest pixel value in each region of the image is selected and output as the convolution result. In addition to selecting the maximum value, the mean value, weighted sum, etc. of the area covered by the convolution kernel can also be calculated and output. A large number of practical results have proved that the maximum pooling is more suitable as the first choice for the pooling layer[14].

3.3. Fully connected layer

A fully connected layer (FCL) is a type of layer commonly used in neural networks, including convolutional neural networks (CNN), responsible for generating the final output of the network. An FCL is sometimes called a dense layer because every neuron in the layer is connected to every neuron in the previous layer[15]. In CNNs, the FCL is usually located at the end of the network, after several convolutional and pooling layers, and is responsible for classifying the input image. The output of the preceding convolutional layers is typically flattened into a one-dimensional vector and passed through one or more FCLs that perform a series of matrix multiplications and activation functions to produce the final output of the network. The number of neurons in the FCL typically decreases as the network progresses towards the output layer, with the number of neurons in the final layer equal to the number of classes being classified. During training, the weights of the FCL are adjusted via backpropagation, using a loss function that measures the difference between the network's predicted output and the actual labels of the training data[16].

A potential disadvantage of FCLs is that they can be prone to overfitting, especially if the network has a large number of parameters. To mitigate this, techniques such as dropout regularization can be used to randomly drop some neurons during training, reducing the risk of overfitting[17].

4. Basic Architecture Features

The classic architecture of the convolutional neural network was originally designed and applied by Yann LeCun, and it has been proved in principle that such a combination can extract and process image features very well. Finally, the results of a large number of experimental comparisons also illustrate the structure. The model below has the best effect under the same experimental environment. In the following years of network model optimization process, most convolutional neural network models also adopt similar structural models[18].

The first convolutional neural network model LeNet-5 was designed by Yann LeCun, which created the basic structure of the convolutional neural network, and various subsequent improved models basically followed this structure. After LeNet-5 was proposed, it was applied to the classification and recognition of handwritten characters in the project. At that time, many American banks used the network to extract and recognize the characters on the check. However, the convolutional neural network is widely used by AlexNet, which continues the basic structure of LeNet-5, but has been improved in many details, and can complete larger-scale recognition tasks. Its excellent performance makes convolution into Networks are widely used in the field of computer vision. AlexNet adopted some relatively new technical means at that time, the most innovative one is the successful application of ReLU activation function, Dropout, LRN and other methods in the network architecture, and achieved very good results[19]. At the same time, AlexNet also uses GPU to accelerate the training of deep convolutional neural network and complete a huge amount of calculation. The AlexNet network structure model has a total of more than 600 million connections, more than 60 million weights and 650,000 network nodes. Its structure has 5 convolutional layers. After the first two convolutional layers, the data is LRN normalized and pooled. Dimensionality reduction operation, the subsequent three consecutive convolutional layers are used to

ISSN 2616-5775 Vol. 6, Issue 4: 9-13, DOI: 10.25236/AJCIS.2023.060402

perform multi-angle and multi-directional feature extraction on the feature map, and then the result is presented in the form of approximate probability through the SoftMax network layer, and the category represented by the maximum value is selected as Prediction results for the current input data[20].

Conv 11×11, s4, 96/ReLU
LRN
Max Pool 3×3, s2
Conv 5×5, s1, 256/ReLU
LRN
Max Pool 3×3, s2
Conv 3×3, s1, 384/ReLU
Conv 3×3, s1, 384/ReLU
Conv 3×3, s1, 256/ReLU
Max Pool 3×3, s2
FC4096/ReLU
FC4096/ReLU
FC1000

Figure 1: AlexNet

The main new technologies used in the AlexNet model are as follows: (1) ReLU has been successfully used as the activation function of the convolutional neural network. Through the analysis and verification of the experimental results, the performance of this function in the deep network is better than that of the Sigmoid activation function. It solves the gradient problem that is prone to occur in deep neural network training. No one had previously discovered the superiority of the ReLU function in artificial neural networks. (2) Propose the Dropout method, which is used to randomly "delete" the hidden layer neuron nodes during the training process. The principle is to pre-set the probability. Before each iteration of the network, the hidden layer nodes are randomly deactivated according to this probability value, and then the data is forward-propagated in the modified network model, and the back-propagation of the error is also maintained in this way. This step is repeated throughout the network until the next batch of data passes through the network model. In each iteration, the structure of the network model passed by the data is different, which reduces the dependence of the training results on the model structure to a certain extent, makes the screening of feature information more flexible, and the total weight of the distribution will be more accurate. During the training of the model, some features will be affected by the structure, causing their weights to be too large or too small, and when trained in another model, their weights will have a large difference, which is the internally designed network Model. In practical applications, such interference factors should be avoided as much as possible. The design of the Dropout method realizes the function of regularization, and plays the role of "average" and "combination" on the model. "Average" refers to training and optimization under different network structures, and finally averages the performance results of multiple network models. "Combination" is the processing method of Dropout, which is equivalent to generating multiple unique "micro-models" in the training phase is aimed at different batches of data, and finally these models are combined into the final "integrated network model". At the same time, in the iterative training of the network, the relationship between neurons is different in each iteration, further reducing the It reduces the complex dependencies between neurons, reduces the binding phenomenon between feature information, and avoids the problem of model overfitting. The concept and principle of dropout have been proposed very early, but it has not been applied to deep learning. AlexNet It is the first artificial neural network model using this method, and the test results also prove the effectiveness of this method in artificial neural networks[21].

References

[1] Li Q., et al. Medical image classification with convolutional neural network. in 2014 13th international conference on control automation robotics & vision (ICARCV). 2014. IEEE.

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

[3] Li Z., et al., A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 2021.

ISSN 2616-5775 Vol. 6, Issue 4: 9-13, DOI: 10.25236/AJCIS.2023.060402

[4] O' Shea K. and R. Nash, An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458, 2015.
[5] Lee H. and H. Kwon, Going deeper with contextual CNN for hyperspectral image classification. IEEE Transactions on Image Processing, 2017. 26(10): p. 4843-4855.
[6] Traore B.B., B. Kamsu-Foguem, and F. Tangara, Deep convolution neural network for image recognition. Ecological informatics, 2018. 48: p. 257-268.
[7] Kim P. and P. Kim, Convolutional neural network. MATLAB deep learning: with machine learning, neural networks and artificial intelligence, 2017: p. 121-147.
[8] Wu J., Introduction to convolutional neural networks. National Key Lab for Novel Software Technology. Nanjing University. China, 2017. 5(23): p. 495.
[9] Al-Saffar A. A. M., H. Tao, and M. A. Talab. Review of deep convolution neural network in image classification. in 2017 International conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET). 2017. IEEE.
[10] Sultana F., A. Sufian, P. Dutta. Advancements in image classification using convolutional neural network. in 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). 2018. IEEE.
[11] Sun Y., et al., Automatically designing CNN architectures using the genetic algorithm for image classification. IEEE transactions on cybernetics, 2020. 50(9): p. 3840-3854.
[12] Han D., Q. Liu, and W. Fan, A new image classification method using CNN transfer learning and web data augmentation. Expert Systems with Applications, 2018. 95: p. 43-56.
[13] Sun M., et al., Learning pooling for convolutional neural network. Neurocomputing, 2017. 224: p. 96-104.
[14] Gholamalinezhad H. and H. Khosravi, Pooling methods in deep neural networks, a review. arXiv preprint arXiv:2009.07485, 2020.
[15] Basha S.S., et al., Impact of fully connected layers on performance of convolutional neural networks for image classification. Neurocomputing, 2020. 378: p. 112-119.
[16] Zhang C.-L., et al. In defense of fully connected layers in visual representation transfer. in Advances in Multimedia Information Processing–PCM 2017: 18th Pacific-Rim Conference on Multimedia, Harbin, China, September 28-29, 2017, Revised Selected Papers, Part II 18. 2018. Springer.
[17] Sainath T.N., et al. Convolutional, long short-term memory, fully connected deep neural networks. in 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). 2015. Ieee.
[18] Macukow B. Neural networks–state of art, brief history, basic models and architecture. in Computer Information Systems and Industrial Management: 15th IFIP TC8 International Conference, CISIM 2016, Vilnius, Lithuania, September 14-16, 2016, Proceedings 15. 2016. Springer.
[19] Iandola F.N., et al., SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360, 2016.
[20] Yuan Z.-W. and J. Zhang. Feature extraction and image retrieval based on AlexNet. in Eighth International Conference on Digital Image Processing (ICDIP 2016). 2016. SPIE.
[21] Wang S.-H., et al., Alcoholism identification based on an AlexNet transfer learning model. Frontiers in psychiatry, 2019. 10: p. 205.