Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2021.020205
ISSN 2790-2692 Vol. 2, Issue 2: 33-39

Convolutional Neural Network in Image Recognition

Mayank Abdullah*

Case Western Reserve University, USA

**corresponding author*

Keywords: Convolutional Neural Network, Image Recognition, Genetic Algorithm, Simulation Experiment

Abstract: Image recognition is an important direction in computer vision. After years of research, image recognition technology has made great progress. This paper mainly studies the application of convolutional neural network (CNN) in image recognition. This paper first analyzes the basic mechanism of CNN, and on this basis, uses genetic algorithm to optimize CNN for image recognition application. Experimental results show that the optimal connection structure searched by this method can significantly improve the effect of CNN image recognition and speed up the training speed of the model.

1. Introduction

Today, with the rapid development of artificial intelligence and big data, people's living standards have been significantly improved, and their lifestyles have also been greatly changed. The rapid progress of Internet technology and the popularity of electronic devices make people's life increasingly convenient and fast, while the exponential growth of image data also brings new challenges [1]. Compared with text information, image contains richer and more vivid information. How to efficiently use image data to solve various problems related to production and life has become one of the key research topics in the field of image processing. Deep learning works by simulating the working mechanism of human brain, and its development comes from the influence of artificial neural network model, while the traditional artificial neural network adopts the mode of fully connecting neurons to transmit data layer by layer, resulting in a relatively complex network structure of the model and a huge amount of calculation [2]. In deep learning application in the field of image recognition, the convolution weights of neural network because of its local connection and sharing characteristics, effectively reduce the model of the process of training and the problem of large number, at the same time can be active learning, the characteristics of the image without professional manual design different characteristics for different tasks, a series of achievements in the field of image recognition. However, with the development of The Times and the data blowout, a large number of image data are without labels. At present, manual labeling is mainly adopted, which is time-consuming and laborious, and seriously lacks efficiency. Meanwhile, the accuracy rate is also affected by the relevant practitioners themselves, and the label data cannot be obtained at a low cost and quickly [3, 4].

Traditional image classification algorithms, mainly through manual or machine learning algorithms to find the differences between images, combined with image processing algorithms, to obtain qualitative or quantitative features, so as to achieve the purpose of classification. In terms of feature extraction, mainly using local binary pattern, such as gradient direction histogram feature extraction algorithm to extract image background, color, size, etc at the bottom of the visual characteristics, but these feature extraction methods are not good generalization ability, such as the same kind of image in different background, using local binary pattern for texture feature has the very big difference, In addition, the designer's prior knowledge and cognitive understanding of classification task should also be used in feature extraction [5, 6]. In terms of classifier, mainly includes the KNN classifier, decision tree method and support vector machine (SVM), although these classifiers are in a large extent, improve the ability of image classification model, but there are different processing large and background of image data, the classification accuracy of classifiers is a far cry from a real demand, so the traditional classifier is not applicable in complex background image. Artificial Neural Network (ANN), which uses the connections of neurons to construct Neural Network, adjusts parameters through Network learning. Since then, the neural network theory has been continuously enriched, and finally ushered in the explosion of deep learning, which provides a new theoretical guidance for the further development of image recognition technology and creates a new situation in the field of image recognition [7]. Among them, CNN (CNN) shines brilliantly and is widely used by many researchers [8].

In recent years, deep learning has developed rapidly due to its high practicability and universality. Depth study of convolution neural network technology not only has good capability of image feature extraction, also has excellent generalization performance, making it become the focus in the current, and in the current study, gradually appeared the CNN based target detection network, to achieve the goal in the image of the detection and location, This provides new ideas and techniques for image classification and object detection.

2. Optimal CNN Based on Image Recognition

2.1. CNN Structure

CNN is the most commonly used network in deep learning networks, which is well applied to every domain [9] in our daily life. Its purpose is to solve the problem of multiple parameters and large scale caused by multiple layers of deep network, which can reduce its complexity in the field of image processing, so as to make the final model relatively efficient. To some extent, the layout of CNN is closer to that of biological neural network, which fully benefits from the principle of local connection and parameter sharing. It can not only maintain a deep network structure, but also reduce weight parameters, so the model has a good generalization ability [10, 11].

(1) Input layer

Current CNNs can input the following data: one-dimensional data, usually time or spectrum sampling; Two-dimensional data, usually image data, can be grayscale images or color images containing three channels; Three-dimensional data, usually video data. Since image processing is the most frequent in CNN at present, two-dimensional CNN is the most commonly used [12, 13].

(2) Convolutional layer

The convolutional computing layer is the most important layer in the CNN structure, and the convolutional layer contains a different number of convolutional filters (also known as convolution kernels), which can extract different features of the input image data [14]. The calculation process of the convolution operation is shown in Equation (1), where L represents the number of layers of the neural network, K represents the convolution kernel, Mj represents the JTH feature map of channel M, I represents the ith feature in Mj, and B represents the bias.

$$X_j^L = f\left(\sum_{i \in M_j} X_i^{L-1} * K_{i,j}^L + b_j^L\right) \tag{1}$$

The convolution kernel slides in the input image in sequence from left to right and from top to bottom, and is multiplied and summed by bit with the corresponding pixels in the same size area in the input image. The distance of each slide is called the convolution step [15]. In the process of convolution operation, for the input image and convolution kernel with a certain size, the number of feature extraction is determined by the step size, and the smaller the step size, the more times the feature is extracted [16].

(3) Pooling layer

Pooling layer pooling is generally followed by convolution operation. Traditional CNNs are usually constructed by conV-pooling. Pooling shall have two functions: on the one hand, retain the main features, reduce the calculation of feature parameters and network layer, reduce feature size and prevent overfitting; on the other hand, keep the translation invariance of features [17].

(4) Activation function layer

Sigmoid function is the earliest and most widely used activation function in neural network models. The input value of this function is real number, and the output range is 0-1[18]. In the network, the relationship between neurons in each layer is continuous multiplication, so the use of Sigmoid function in the deep neural network layer will lead to the compression of the output value of neurons in each layer within the interval of 0-1, and the final result of continuous multiplication will tend to 0, which will cause the phenomenon of vanishing gradient, and thus the network cannot be optimized. In addition, the output of this function is not centered on 0. In the process of back propagation, the update direction will be carried out in one direction, which slows down the convergence speed. The mathematical expression of this function is shown in Equation (2).

$$f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

(5) Fully connected layer

The fully connected layer is a tiling structure composed of multiple neurons. The features are stretched by convolution, the previously extracted features are reintegrated together, the high-level features are represented by a one-dimensional vector, and the output is transmitted to the classifier for final classification. Feature integration through fully connected layer reduces the influence of feature location on classification.

2.2. CNN Optimization

Genetic algorithm has a high degree of parallelism, randomness and adaptability, which is a relatively easy to understand and simple to operate evolutionary algorithm. In this paper, genetic algorithm is used to optimize CNN.

Pre-train CNN. A CNN is pre-trained by the backpropagation algorithm until the model converges and the model parameters are saved. The purpose of pre-training is to fix the parameters of the model.

Initialize the population. M binary sequences are randomly generated as the initial population P0, and the iteration counter t=0. The number of elements of each binary sequence in P0 is determined by the number of input and output feature map channels of the convolutional layer to be optimized. For example, in a typical CNN, the number of feature channels in S2 layer is 6, the number of channels in C3 layer is 16, and the number of elements in each binary sequence is 96. Each element represents the connection state of the corresponding position.

Decode to the corresponding connection structure. Each binary sequence in the current population Pt corresponds to each channel connection mode and is separately applied to the pre-trained CNN.

Calculate population fitness. The CNNs with different connection structures are propagated forward to obtain the cross-entropy corresponding to each connection structure in the current population Pt, which is the fitness of the current population.

Determine whether the iteration of adaptive genetic algorithm stops. That is, judge whether the counter t is equal to the total number of iterations N. If so, proceed to the selection optimal decoding; otherwise, perform the selection, crossover and mutation operations.

Selection, crossover and mutation operations. For the selection operation, the elite selection mechanism is used here to directly inherit the first C individuals with the best fitness to the next generation.

The optimal individual is selected and decoded into the corresponding connection structure. The individual with the optimal fitness in the current population is selected as the coding sequence of the optimal connection structure, which is decoded into the corresponding feature channel connection structure and applied in the pre-trained CNN.

Of characteristics through the above based on adaptive genetic algorithm in optimizing the structure of the channel connection method, can delete all the characteristics of the connection diagram redundant connections between structure, automatic search out the best connection structure in the face of the current study tasks, and this method has universality, to some extent can promote to optimize the convolution model of feature extraction ability and generalization ability, Thus, the accuracy of image recognition is improved.

3. Image Recognition Simulation Experiment

3.1. Experimental Data set

The ImageNet10 dataset is composed of 10 classes randomly selected from ImageNet, and the image size is uniformly set to 96×96.

The CIFAR-10 dataset contains 60,000 images with a total of 10 class datasets.

3.2. Experimental Environment and Configuration

Table 1. Experimental environment configuration

Configuration	Parameter
CPU	Intel Core i7-11700
GPU	RTX 3080
Memory	32GB
Deep learning framework	Tensorflow

As shown in Table 1, in the experiment, Tensorflow deep learning framework is used, and all the experiments are in the same NVIDIA

RTX3080 runs on a GPU to achieve the purpose of using GPU acceleration. In addition, in the process of CNN training, the experiment uses stochastic gradient descent (SGD) to train the

GA-CNN model.

4. Analysis of Experimental Results

4.1. Comparison of Different Models

The experiment verifies the correctness of the scheme by comparing the classification accuracy of Alexnet model, GoogLenet model and GA-CNN model, which are the mainstream CNNs currently used for image recognition. The results are shown in Table 2.

Table 2. Comparison of results between Alexnet, GoogLenet and GA-CNN

Data set	Alexnet	GoogLenet	GA-CNN
ImageNet10	80.14%	84.53%	88.72%
Cifar-10	81.35%	86.95%	91.46%

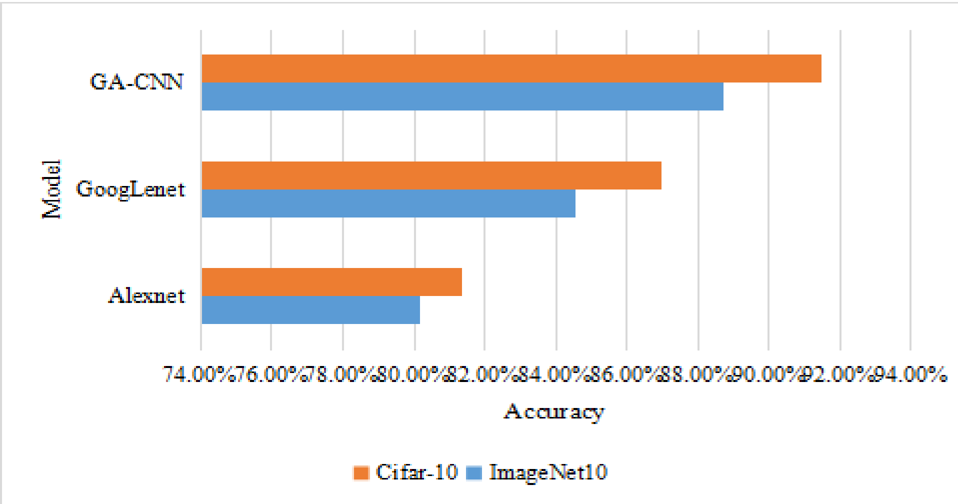

Figure 1. Comparison of accuracy of three models

As shown in FIG. 1, it can be seen that the results of GA-CNN have been significantly improved. This is mainly because the image processed by Alexnet model is only a simple channel data, which does not have sufficient feature expression ability for a complex image data. As a result, the classification accuracy is lower than that of GA-CNN model.

4.2. Comparison before and after Optimization

Table 3. Comparison of CNN algorithm before and after optimization

Data set	Before optimization	After optimization
ImageNet10	86.31%	88.72%
Cifar-10	89.06%	91.46%

As shown in Table 3, the recognition results after optimization are improved to a certain extent compared with those before optimization.

4.3. Comparison of Different Learning Rates

As shown in Figure 2, among the three groups of learning rates of 0.1, 0.01 and 0.001, on the two different groups of data sets, the classification effect with learning rate of 0.001 is the best. In the training process, if the learning rate is too large, the optimal solution may be skipped. In the early stage of the experiment, if the learning rate is too small to train the model, the learning process of the model will be very slow. Further, if the learning rate is set small enough, it may converge to the local optimal solution.

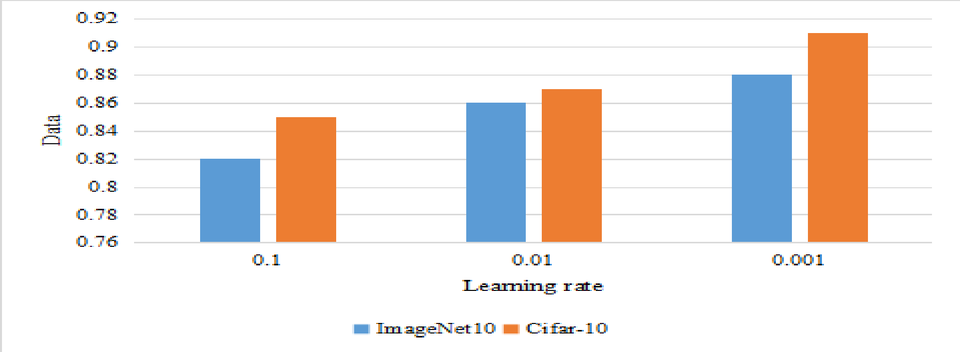

Figure 2. Classification effect of different learning rates on different data sets

5. Conclusion

Convolution weights of neural network is a Shared, the characteristics of the local connection model, made great achievements in the field of image recognition, because it can experiment end-to-end feature extraction, without professional design manual features of different according to different tasks, so not only caused great attention in academia, at the same time also has been widely used in industry, liberate the workforce effectively, Make the production activities become more efficient and fast. In this paper, CNN is applied to the problem of image recognition. Although part of the problem is solved, there are still some problems to be solved. With the deepening of CNN, the model parameters will be larger, and the training needs more data, but the deeper the model is, the higher the classification accuracy will be. How to balance the depth of the model with the size of the training dataset should be studied.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Campilho A, Karray F, Bart T. [Lecture Notes in Computer Science] Image Analysis and Recognition Volume 10882 || Fully CNN for Lungs Segmentation from Chest X-Rays. (2018) 10.1007/978-3-319-93000-8(Chapter 9):71-80. https://doi.org/10.1007/978-3-319-93000-8_9

[2] Vidushi, Agarwal M, Rajak A, et al. Assessment of Optimizers Impact on Image Recognition with CNN to Adversarial Datasets. Journal of Physics: Conference Series. (2021) 1998(1):012008. https://doi.org/10.1088/1742-6596/1998/1/012008

[3] Sharan R V, Moir T J. Acoustic Event Recognition Using Cochleagram Image and Cnns. Applied Acoustics. (2019) 148(MAY):62-66. https://doi.org/10.1016/j.apacoust.2018.12.006

[4] Dat N, Ki K, Hong H, et al. Gender Recognition from Human-Body Images Using Visible-Light and Thermal Camera Videos Based on a CNN for Image Feature Extraction. Sensors. (2017) 17(3):637. https://doi.org/10.3390/s17030637

[5] Uddin M Z, Khaksar W, Torresen J. Facial Expression Recognition Using Salient Features and CNN. IEEE Access. (2017) (99):1-1. https://doi.org/10.1109/ACCESS.2017.2777003

[6] Siregar B, Wirtjes J S, Nababan E B, et al. Development of a Person's Facial Expression Recognition System Using a CNN. Journal of Physics: Conference Series. (2021) 1882(1):012127 (10pp). https://doi.org/10.1088/1742-6596/1882/1/012127

[7] Joshi S, Kumar R, Dwivedi A. Hybrid DSSCS and CNN for Peripheral Blood Cell Recognition System. IET Image Processing. (2020) 14(17):4450-4460. https://doi.org/10.1049/iet-ipr.2020.0370

[8] Natarajan, Sudha, Baradkar, et al. Traffic Sign Recognition Using Weighted Multi-CNN. IET Intelligent Transport Systems. (2018) 12(10):1396-1405. https://doi.org/10.1049/iet-its.2018.5171

[9] Tomimori H, Chen K T, Baba T. A CNN with Incremental Learning. Journal of Signal Processing. (2017) 21(4):155-158. https://doi.org/10.2299/jsp.21.155

[10] Paoletti M E, Haut J M, Plaza J, et al. A New Deep CNN for Fast Hyperspectral Image Classification. Isprs Journal of Photogrammetry & Remote Sensing. (2017) 145PA (NOV.):120-147. https://doi.org/10.1016/j.isprsjprs.2017.11.021

[11] Abtahi T, Shea C, Kulkarni A, et al. Accelerating CNN with FFT on Embedded Hardware. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. (2018) (99):1-24.

[12] Clark T, Nyberg E. Creating the Black Box: A Primer on CNN Use in Image Interpretation - ScienceDirect. Current Problems in Diagnostic Radiology. (2020) 49(6):365-367. https://doi.org/10.1067/j.cpradiol.2019.07.004

[13] Chandio A A, Asikuzzaman M, Pickering M R. Cursive Character Recognition in Natural Scene Images using a Multilevel CNN Fusion. IEEE Access. (2020) (99):1-1. https://doi.org/10.1109/ACCESS.2020.3001605

[14] Lvarez-Canchila O I, DE Arroyo-Pérez, A Patiňo-Saucedo, et al. Colombian Fruit and Vegetables Recognition Using Cnns and Transfer Learning. Journal of Physics: Conference Series. (2020) 1547(1):01(2020) (6pp). https://doi.org/10.1088/1742-6596/1547/1/012020

[15] Kwak Y, Song W J, Kim S E. Speckle-Noise-Invariant CNN for SAR Target Recognition. IEEE Geoscience and Remote Sensing Letters. (2019) 16(4):549-553. https://doi.org/10.1109/LGRS.2018.2877599

[16] Rivera P, Anazco E V, Choi M T, et al. Trilateral CNN for 3D Shape Reconstruction of Objects From A Single Depth View. Image Processing, IET. (2019) 13(13):2457-2466. https://doi.org/10.1049/iet-ipr.2019.0532

[17] Jaffe L, Zelinski M, Sakla W. Remote Sensor Design for Visual Recognition with CNNs. IEEE Transactions on Geoscience and Remote Sensing. (2019) 57(11):9090-9108. https://doi.org/10.1109/TGRS.2019.2925813

[18] Hirasawa T, Aoyama K, Tanimoto T, et al. Application of Artificial Intelligence Using a CNN for Detecting Gastric Cancer in Endoscopic Images. Gastric Cancer Official Journal of the International Gastric Cancer Association & the Japanese Gastric Cancer Association. (2018) 87(Suppl 1):1-8.