WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n1.28
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Application of Convolutional Neural Network in Image Recognition

Kaijie Lai

Xi'an Jiaotong-Liverpool University, China
Kaijie.Lai21@student.xjtlu.edu.cn

ABSTRACT

Image recognition is of great significance for both human social development and daily life. This paper mainly discusses the application research of convolutional neural network (CNN) in the field of image recognition. Firstly, the basic structure and working principle of CNN, including the core components of convolution layer, pooling layer and full connection layer, are introduced, and then the structure and characteristics of the middle layer are explained. Then, this paper compares the advantages and disadvantages of different CNN models such as LeNet-5, AlexNet, GoogLeNet, ResNet, etc., and analyzes their application in face recognition, medical image recognition, traffic recognition, character recognition and other fields. Finally, this paper summarizes the application of CNN in image recognition, and points out the problems faced in its practical application, such as model complexity, computational resource requirements, etc., aiming to provide reference for further research of image recognition technology.

KEYWORDS

Image recognition; Convolutional neural network; CNN model; Deep learning; Practical application

1. INTRODUCTION

With the rapid development of industrialization, produced a lot of image information, the traditional image recognition technology is difficult to handle such a large image data and meet the requirements of speed and precision, big data and deep learning technology arises at the historic moment, based on the convolution neural network image recognition method has become the mainstream of image recognition algorithm. The development of convolutional neural network can be traced back to the 1980s, its structure is simple neural network structure, with the development of deep learning, due to the convolutional neural network with feature extraction ability, its superior performance in image recognition task, it gradually become the mainstream method in the field of image recognition. In particular, in 2012, Krizhevsky proposed a convolutional neural network model AlexNet, which is used to identify ImageNet data sets, including 8 layers of convolution layers and 3 layers of fully connected layers, using non-linear activation function and Dropout, etc. At the same time, using deep network design and parallel computing means, the error of ImageNet has been reduced by about 12.4%. In 2014, the VGG model was proposed by Simonyan and Zisserman, which improves the accuracy of image classification by increasing the depth of the neural network. Google also proposed the Inception model in the same year, which uses a multi-scale convolution kernel and a parallel network structure to improve the performance of neural networks. In 2015, He et al. proposed a structure ResNet with residual connectivity, so that deeper networks do not cause performance degradation, further driving the development of convolutional neural networks. In 2017, Huang et al. proposed the densely connected DenseNet, connecting the feature graph of the previous layer with

the feature graph of the latter layer, enabling the network to better share features during the training process. In 2020, the Vision Transformer model, proposed by Dosovitskiy et al., performs feature extraction by introducing a self-attention mechanism and performs well in many image classification and image segmentation tasks. In order to better improve the performance of image recognition, many studies began to enhance the performance of convolutional neural networks by combining other machine learning techniques.

This paper summarizes and discusses the research of image recognition based on convolutional neural network, including introduces the basic structure and working principle of the model, highlights the structure and characteristics of convolution network, compare the advantages and disadvantages of commonly used convolution neural network model, also discusses the convolution neural network in face recognition, medical image recognition, traffic recognition, character recognition, application, finally, reviews the application of convolution neural network in image recognition, in order to provide corresponding reference for further research of image recognition technology.

2. THE BASIC PRINCIPLES OF THE CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural network (CNN) is a forward neural network structure. The core component is a neural network model composed of convolutional layer and pooling layer, including input layer, convolutional layer, pooling layer, full connection layer and output layer. The structure of CNN is shown in Figure 1. First, after the image is input into the CNN network, the features in the image are extracted after multiple convolution and pooling operations, and finally the image features are transported to the fully connected layer for image recognition.

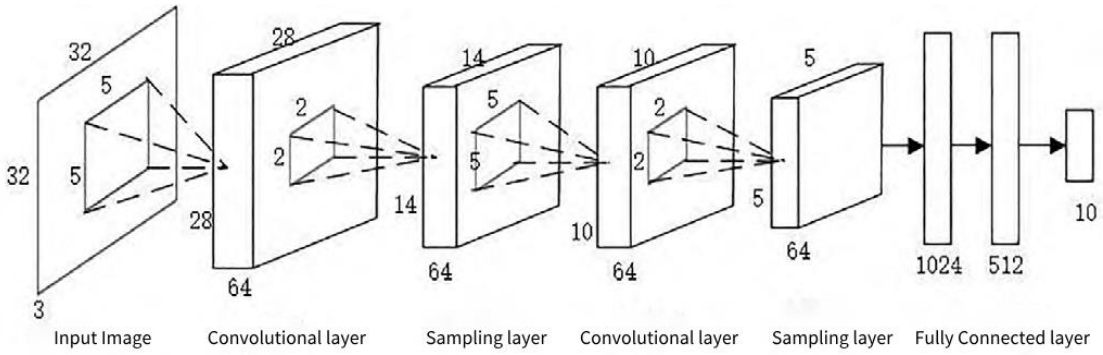

Figure 1. Convolutional neural network structure diagram

2.1. The Convolution Layer

In the convolution layer, the convolution operation by one or more convolution kernel (also known as filter) on the front image area corresponding convolution operation, then according to a certain step for sliding operation, extract image pixel features, image feature synthesis after the activation function activation, complete an input to the output feature extraction process, the convolution feature diagram reflects the fusion features of the front image. In the actual image processing task, the convolutional neural network can design multiple layers to extract the image information of the front layer by layer, and depict the features of the original image until it is transferred to the fully connected layer. In the feature extraction process, the main work of the convolution layer is to train fewer parameters to extract the feature information of the data picture. The convolutional layer uses convolution kernel to perform each local feature of the input image to convolution and extract feature information, which is essentially a process of neural network to extract external features of the image by learning the local features. Specifically, the convolutional layer is composed of multiple convolution cores. Each convolution kernel is regarded as a template, and each pixel of the input

image is calculated with the convolution kernel to generate an output value. Then then the convolution kernel is convoluted on the input image according to a certain step size, and finally the output feature graph is obtained. This process is equivalent to the feature extraction of the input image, generating a useful feature image. The convolution is calculated as follows:

$$h_{ij}=\sigma(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}w_{mn}x_{i+m,j+n}+b) \tag{1}$$

Where: h, is the output after convolution; $X_{i+m, j+n}$ is the element of the row m and column n in the input; W_{mn} is the weight in the convolution kernel; b is the bias term; σ is the activation function; M and N are the size of the convolution kernel. In practical application.

In all cases, the filter parameters of the convolutional layer need to be constantly adjusted by the back propagation algorithm to obtain a better filtering effect and output feature diagram.

2.2. Convolutional Neural Network Training

For the supervised learning process, in the training process of convolutional neural network, the network will automatically learn the parameter values in each layer network. A training process includes information extraction from the input layer to the output layer, learning process and error back propagation from output to input to realize the parameter update process. The forward transfer process mainly introduces the convolution operation of the convolution layer, and the error back propagation shows the BP process of the traditional neural network. The operation of convolution operation is shown in equation (2):

$$X_j^l = f(\sum_i X_{iM_j}^{l-1} * \boldsymbol{K}_{ij}^l + b_j^l) \tag{2}$$

X_j^l is the j-th feature graph of the l-layer in the network; f is the activation function; M_j is the subgraph participating in the convolution operation in the j-th feature graph of the l-1 layer in the network; $\boldsymbol{K}_{ij}^l$ is the convolution kernel matrix, and the convolution kernel matrix of each input feature graph in the convolution operation; b_j^l is the bias of the j-th feature graph of the l-layer in the network, and each output feature graph has a bias amount.

2.3. Pooling Layer

The pooling layer is also known as the lower sampling layer. For the convolved feature map, the selected step size uses a certain size image area to take out the typical characteristic value of the area. There are usually three kinds of eigenvalues: maximum value, average value and random value. The corresponding pooling methods are maximum pooling, average pooling and random pooling respectively. The random pooling methods have randomness and are randomly selected according to the probability matrix according to the size. The pooling operation changes multiple values in the pooling region into one value, and the width and height of the feature graph are reduced to several times of the original, which reduces the feature vector output of the convolutional layer, greatly reduces the number of parameters and the calculation amount in the convolutional network, and at the same time ensures the translation invariance of the data. The number of channels in the feature map remains unchanged after pooling operation.

2.4. Fully Connected Layer

Fully connected layer is a traditional multilayer perceptron, appeared in the end of the network structure, each layer of neurons and the previous layer of neurons are connection, the input layer image features by multilayer feature extraction after highly fusion, get the image of high

representation, finally by the regression classification model to get the corresponding response value to the output layer.

3. A CONVOLUTIONAL NEURAL NETWORK MODEL COMMONLY USED IN IMAGE RECOGNITION

In the field of image recognition, there are many kinds of CNN network structure models commonly used for image recognition, such as LeNet-5, AlexNet, GoogLeNet, ResNet and other network models. The commonly used network models and their advantages and disadvantages are analyzed and compared below.

(1) LeNet model

The LeNet model was first proposed by YAN L C in 1998. Lay the foundation of the convolutional neural networks, The LeNet network structure is shown in Figure Figure 2, Composed of two convolutional layers, two fully connected layers, and one output layer, Also known as the LeNet-5, The basic model structure is input-> conv 1-> pool1-> conv 2-> pool2-> fc 1-> fc 2-> output, Middle with two convolution-pooling stacks, The size of the convolution kernel issue 1 Zhang Songlan: Image recognition based on the convolutional neural network review 77 are 55. The convolution layer is used to extract the spatial features of the image, and the activation function uses softmax; the pooling layer adopts average pooling to reduce the dimension of the image data; each node in the fully connected layer and the input node, which gathers the comprehensive feature information extracted by the previous convolutional graph. This model was mainly used for the recognition and classification of handwritten characters in its early stage.

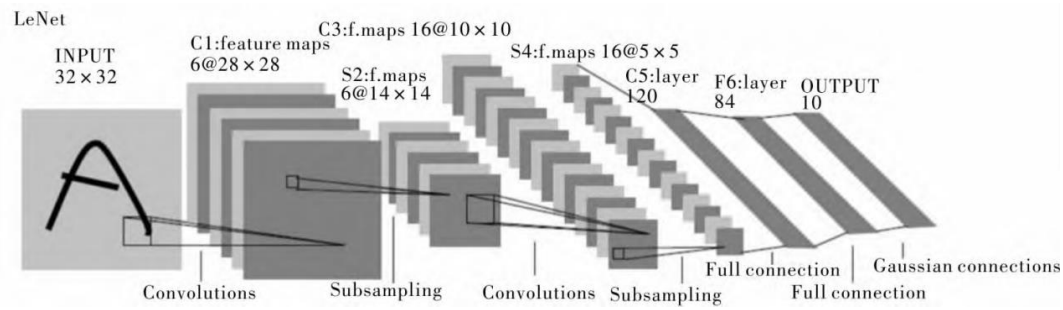

Figure 2. LeNet Network Structure

(2) AlexNet is a kind of convolutional neural network, which contains 5 convolutional layers and pooling layers, which can effectively extract various features of images and improve the accuracy of model recognition. In addition, ReLU activation function and Dropout layer are used to further improve the performance of the model; GPU parallel computing technology greatly improves the speed and efficiency of model training for large-scale image recognition tasks. However, the AlexNet model is relatively complex, which requires large computing resources and storage space during running time, and the amount of data required to be processed during training is very large, so it requires a long training time and a large amount of training data.

(3) VGGNet is a relatively deep convolutional neural network, containing 16 or 19 convolution layers and pooling layers, which can extract more complex image features, and has achieved very good results in some challenging image recognition tasks. Because VGGNet adopts small convolution kernel, the model has small complexity, can effectively extract the features of the image, and has good recognition performance. However, the extremely deep network structure of VGGNet leads to its great computational complexity, which requires more computing resources and even GPU, and is prone to model overfitting. It needs to adopt appropriate methods and lack of attention module, which leads to insufficient information processing in regions with high correlation in the input data.

(4) GoogLeNet is a relatively novel convolutional neural network, with 22 layers, the network model introduced the Inception module, global average pooling layer innovative structure, can greatly reduce the number of parameters of the model, improve the accuracy, the model can also learn the characteristics of different sizes and different direction, this to solve sharpening, fuzzy, rotating deformation has a strong robustness, at the same time at the same time has strong real-time. However, the model of GoogLeNet is very deep, especially in the middle of training, there is the risk of gradient disappearance and explosion, so it requires fine parameter initialization and regularization, and even uses deep residual networks such as ResNet to solve. The model structure is also complex, and the calculation time and cost are large.

(5) ResNet is a very deep convolutional neural network, with up to 152 layers. The main innovation of the network is the use of residual learning method, which effectively solves the problem of gradient disappearance in the training of a very deep network. ResNet Deeper neural networks can be constructed, thus improving the classification accuracy, such as ResNet-101 and ResNet-152 have shown good results. However, the network structure of ResNet is relatively complex, and its training time is longer than that of VGG and other models, so it needs a larger training set and computing resources to support it.

4. APPLICATION OF CONVOLUTIONAL NEURAL NETWORK IN IMAGE RECOGNITION

4.1. Handwritten Digit Recognition

Handwritten digit recognition is one of the earliest applications of CNN in image recognition. LeNet network can achieve 98% accuracy in handwriting character recognition, and put into use in practical system. Design of convolutional network model with Inception of GoogLeNet model and verified on the MNIST dataset can achieve 99.66% accuracy. Combined with SIP technology and FPGA technology, convolutional neural network was used to realize microsystem design, HLS to design convolutional layer and pooling layer in CNN, generate IP core time-sharing multiplexing microsystem, realize MicroVGGNet lightweight model, and the recognition accuracy of MNIST handwritten digital data set reached 98.1%.

4.2. Face Recognition

Face recognition technology is a common technical means in daily life, which is widely used in video surveillance, access control system, e-commerce and other aspects. Methods based on convolutional neural networks have achieved very good performance on publicly available datasets such as LFW and YTF. For example, the models of the DeepID series achieved near-perfect results on these datasets. CNN has achieved remarkable results in the field of face recognition. For example, FaceNet system, which can achieve as high as 99.63% accuracy by training deep neural network. This system adopts a novel triplet loss function to make the neural network compress the space of face features when extracting face features, thus improving the accuracy. For CNN training parameters, computational problems, experts put forward on the face recognition algorithm lightweight MobileNet convolutional neural network model, will improve MobileNet network and regional generation network RPN fusion, in the LFW face database and self-built small database training test, recognition accuracy reached 97.54%, increased by 0.2%, running speed increased by 21.3%, model parameters reduced by 88%.

4.3. Medical Image Recognition

Convolutional neural network has made great achievements in medical image processing. Many research teams at home and abroad are committed to using convolutional neural network to solve

various medical image processing problems, such as skin cancer classification, medical image analysis, detection of brain and liver organs, etc. For example, researchers from Stanford University developed a deep learning based skin cancer classification algorithm that can accurately identify various types of skin cancer with an accuracy of 91%; researchers from Leibniz Information Center and German Cancer Research Center used convolutional neural network to effectively detect muscle tumor and lung cancer in radiological imaging, and researchers at Oxford University developed an automatic segmentation algorithm based on convolutional neural network, which can accurately detect organs such as brain and liver.

4.4. Traffic Identification

In the field of traffic, image recognition technology is widely used in many aspects, including license plate recognition, traffic sign recognition, traffic flow prediction, Lane offset warning, these technologies play an important role in the construction of smart cities. For example, a traffic sign detection algorithm based on deep convolutional neural network is proposed in the literature, which is 99% accuracy and fast in real time. In short, traffic identification technology has become an important research direction in the field of artificial intelligence, and has been widely used in traffic control, intelligent traffic and other fields. The development of neural convolutional network in traffic identification will continue to promote, which will further promote the rapid development of smart city and intelligent transportation industry.

4.5. Human Action Recognition

The traditional human action recognition methods are divided into feature extraction methods based on human motion information and feature extraction methods based on spatial and temporal points of interest. Action recognition in the human body is divided into two separate processes: feature extraction and action recognition. In deep learning, end-to-end training method can be unified training and learning of feature extraction and action recognition. Most of the action recognition task is used for video detection. The spatiotemporal collaborative convolution model proposed by temporal convolution and spatial convolution effectively reduces the complexity of the model. The multimodal system based on action, scene, object and sound features proposed by Roig et al. effectively improves the action recognition ability in scenes lacking the ability to detect and understand behavior-related contextual information.

4.6. Crop Testing

There are also a lot of applications in convolution neural networks. The LeNet network model was used for image recognition in multiple clusters of kiwi fruit recognition. Zhou Shengan et al. Used MobileNetV3's lightweight convolutional neural network model to detect small and medium fruit defects, and the experimental results showed a good balance in terms of training time, detection speed and accuracy. Some experts constructed three neural network models of VGG 16, ResNet50 and InceptionV3 on apple tree leaves, and found that using transfer learning can improve the convergence rate and accuracy of the model.

5. SUMMARY

As a new deep learning method, convolutional neural network has been developed in various fields of computer vision with the rapid development of hardware computing power. Convolutional neural networks are used in all aspects of life because of their own efficient —— local connectivity, parameter sharing, as well as good scalability and robustness. Image recognition is one of them. Although the convolutional neural network is powerful, there are still many problems in its development process. This paper summarizes and discusses the research of image recognition based

on convolutional neural network, including introduces the basic structure and working principle of the model, highlights the structure and characteristics of the middle layer of convolutional network, compares the advantages and disadvantages of common convolutional neural network model, and discusses the application of face recognition, medical image recognition, traffic recognition, the application of convolutional neural network in image recognition, in order to provide corresponding reference for further research of image recognition technology.

REFERENCES

[1] Luo Fugui, Song Qian, Qin Yunchu, and so on. Research on the Application of Convolutional Neural Network in Image Recognition [J]. Computers and Information Technology, 2024,32 (3): 51-54. DOI:10.3969/j.issn.1005-1228.2024.03.014.

[2] Nie Jianzheng. Research on transfer learning algorithm based on Convolutional Neural Network and its application in Fine-grained Image Recognition [D]. Shaanxi: Xidian University, 2018. DOI:10.7666/d.D01663960.

[3] Wang Long. Research on semantic segmentation algorithm based on convolutional neural network and its application in image recognition of sweet pepper [D]. Jiangsu: Jiangsu University, 2021.

[4] Chen Beibei. Application of Convolutional Neural Network in Remote Remote Image Recognition [D]. Hebei University of Geosciences, 2018.

[5] Gai Rongli, CAI Jianrong, Wang Shiyu, et al. Review of the application of convolutional neural networks in image recognition [J]. Small Microcomputer System, 2021, 42 (9): 1980-1984. DOI:10.3969/j.issn.1000-1220.2021.09.030.

[6] Time rise. Application of Convolutional neural network in traffic image recognition [D]. Gansu Province: Northwest Normal University, 2021.

[7] Conshoman. Research and Application of Image Recognition Method Based on Convolutional Neural Network [D]. Hebei: Yanshan University, 2021.

[8] Li Xiaoyang, Xie Hengyi, Han Zhenhui, et al. Research on the Application of Convolutional Neural Network in Image Recognition Technology [J]. Energy and Environmental Protection, 2020, 42 (6): 73-76.

[9] Wang Xinyu. Application of convolutional neural network in finger vein recognition [D]. Sichuan: Southwest University of Science and Technology, 2022.

[10] Liu Bingzai. Research on Chinese Verification code identification technology based on Convolutional Neural Network [D]. Wuhan University of Light Industry, 2023.