Summary of Research on Application of Deep Learning in Image Recognition

Don Yitong Ma

University of Wisconsin-Madison, Madison, 53703, United States

Abstract. With the rapid development of technology, artificial intelligence is gradually applied in the field of image processing. Traditional image processing and recognition takes human extraction of features as a means to describe and characterize low-level edge information in images, but a large number of features extraction causes certain troubles to image recognition, and it is difficult to extract deep-level information. In recent years, deep learning, as an important emerging technology in the field of image recognition, has broad application prospects, and the research on image recognition technology has important theoretical value and practical significance to promote the development of computer vision and artificial intelligence. It opens up a whole new direction in image processing research, thus meeting the demand for efficient image processing. This paper reviews the applications of deep learning in image recognition, analyzes the classical convolutional neural network model, and summarizes the applications of deep learning in the fields of license plate recognition, face recognition, and medical image recognition.

Key words: Deep learning, Mage recognition, Convolutional neural network.

1. Introduction

In recent years, deep learning technology has achieved excellent performance in many branch tasks of image processing. Especially since 2012 when the famous AlexNet network designed by Hinton et al.[1] achieved excellent results exceeding the accuracy of human recognition in a large-scale visual recognition challenge for the first time, the development of deep learning has ushered in an important turning point. Since then, deep learning has gradually been applied to many fields, and deep learning models and variants have achieved the most cutting-edge level in many fields.

2. Deep learning model in image recognition

2.1 Image recognition process

Image recognition by computer is usually divided into two major steps: image feature extraction and image classification prediction. Firstly, the input image is is preprocessed and processed into a form suitable for feature extraction, then the features of the image are extracted, after which the feature image is classified and predicted. After that, the feature images are classified and predicted, as shown in Figure 1.[2]

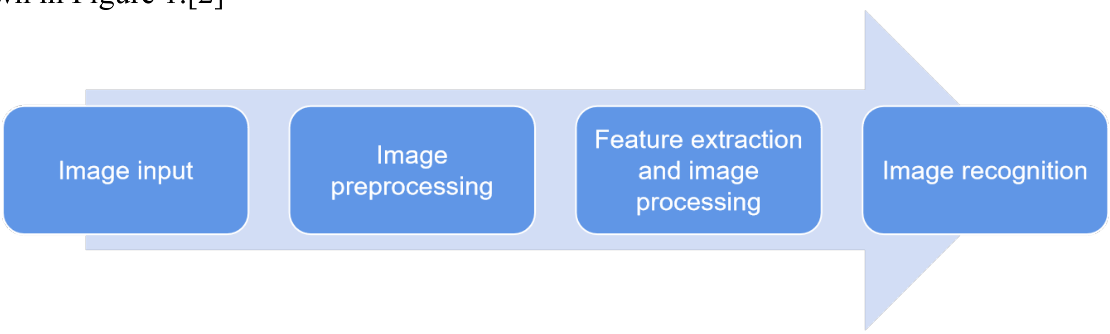

Figure 1. Image recognition process

Among them, image preprocessing is mainly for the purpose of eliminating interference, enhancing target image information, and better performing image feature extraction.

Table 1. Common preprocessing methods

1	Graying	Changing RGB image to gray image mainly using component method and weighted average method
2	Geometric transformation	Mainly using interpolation method to make spatial transformation on the image space to reduce image error information;
3	Image enhancement	Enhancing the target image information, including grayscale transformation, histogram correction and filtering methods. Traditional image recognition methods need to manually extract image features, and the recognition accuracy depends on the accuracy of feature extraction.

With the emergence of massive images in modern society, the resolution of images is getting higher and higher, and the traditional image recognition methods are not suitable for processing massive high-resolution images. The application of deep learning in the field of image recognition, by directly processing the input image, avoids the complex image feature extraction as well as data reconstruction process, and gets significant improvement in processing efficiency and accuracy.[3]

2.2 Analysis of Deep Learning Framework

With the increase of computer computing power and the emergence of big data, deep learning models have emerged. The main idea is to learn a set of models by using known data, so that the system can make corresponding predictions when encountering unknown data.

In the framework of deep learning, the algorithm structure of different models has more generality, such as convolutional neural network model and long-term short-term memory model, which can be divided into networking module, gradient descent optimization module and prediction module. This makes it possible to abstract a unified framework and greatly reduces the cost of writing modeling code. Some relatively general modules, such as the implementation of the basic network operator and various optimization algorithms, can be implemented by the framework. The modeler only needs to focus on data processing, configuring the networking approach, and stringing together the training and prediction processes with a small amount of code.

Deep learning not only enables end-to-end learning of models, but also drives AI into large industrial production, resulting in standardized, automated, and modularized common architectures. The deep learning models in different scenarios have certain generality, and the construction and training of deep learning models can be summarized in the following five steps, which are as follows. [4]

1). Data processing

Read data from local or URL, and complete pre-processing operations, such as data validation, format conversion, etc., to ensure that the model can be read. The data processing includes data import, data shape transformation, data set partitioning, data normalization processing and wrapper function, among which the data set partitioning and data normalization processing should be emphasized. [5] This is because the data set processing directly leads to whether the model can be applied properly in the next step. In general, the data set is divided into a training set and a test set, where the training set is used to determine the parameters of the model and the test set is used to judge the effect of the model. The data is normalized to make the model training more efficient and because each feature has the same range of values, the pre-feature weights represent the contribution of the variable to the prediction results. When the data is pre-processed, it can be called by the model. [6]

2). Model design

Model design is one of the key elements of deep learning models. It is equivalent to the hypothesis space of the model, i.e., the set of relationships that the model can express, and realizes the process of "forward computation" of the model.

3). Training configuration

Training configuration is one of the key elements of the deep learning model. The process is to set the solution-seeking algorithm, i.e., the optimizer, and to specify the computational resources. After the model is designed, it is necessary to find the optimal value of the model through the training configuration, i.e., the loss function to measure the goodness of the model.

4). Training process

The training process is one of the key elements of the deep learning model. Its model is able to call the training process cyclically, and each round includes three steps of forward computation, loss function and backward propagation.

5). Model saving

The trained model is saved and called when the model is predicted. It is because of the generality of deep learning in modeling and training process, only the model design, training configuration and training process are different when building different models, but the other steps are basically the same, so the deep learning framework is useful.

2.3 Convolutional Neural Networks (CNN)

In general, deep learning can be divided into supervised learning and unsupervised learning on the basis of whether the data contains .In the process of supervised learning, the mapping relationship between the marks and the features of the training data will be found, and the deviations in the learning process will be continuously corrected through the marks, and the prediction rate of learning will be continuously improved. Supervised learning mainly includes CNN, Recurrent Neural Networks (RNN) and Deep Stacked Networks (DSN).The training data of unsupervised learning does not have marks. Commonly used algorithms include Restricted Boltzmann Machine (RBM), and Deep Belief Network (DBN).[7]

Image data can be considered as a set of two-dimensional pixel value input. Convolutional Neural Networks (CNN) is composed of several convolution layers and fully connected layers and includes related weights and pooling layers. Such structure enables CNN to make full use of the two-dimensional structure of the data, so that the recognition image has a higher accuracy rate. CNN also uses standard backpropagation algorithm for training. With fewer parameter estimates, it is easier to train than other deep models. Figure 2 shows the basic model of image recognition based on CNN. First, the model divides the input image by region, then enters the image into CNN network for feature extraction and recognition, and finally classifies the image based on image features.[8]

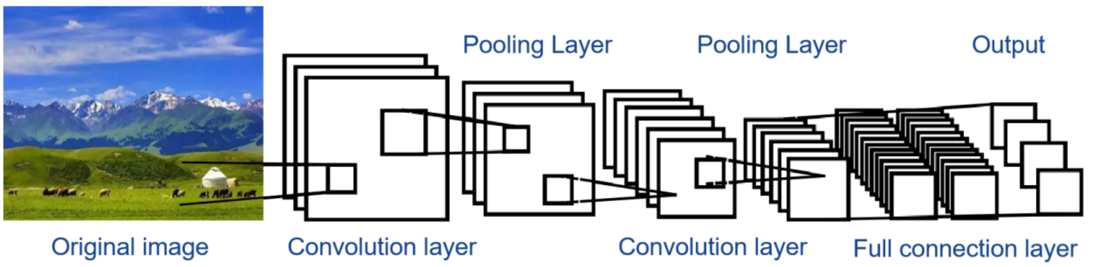

Figure 2. Graphical model of image recognition based on CNN

A typical convolutional neural network consists of basic modules such as Convolution, Pooling, activation functions such as Relu, Batch norm and Dropout. Batch normalization is an optimization method to increase the numerical stability, while Dropout prevents overfitting. The basic architecture of a convolutional neural network consists of a feature extractor and a classifier. The feature extractor is usually composed of several convolutional and pooling layers, and the convolutional and pooling processes continuously reduce the feature map and increase the number of feature maps. The feature extractor is usually followed by a classifier, which usually consists of a multilayer perceptron. In particular, after the last feature extractor, all the feature maps are expanded and arranged into a vector to obtain the feature vector, which is used as input to the classifier in the later layers.

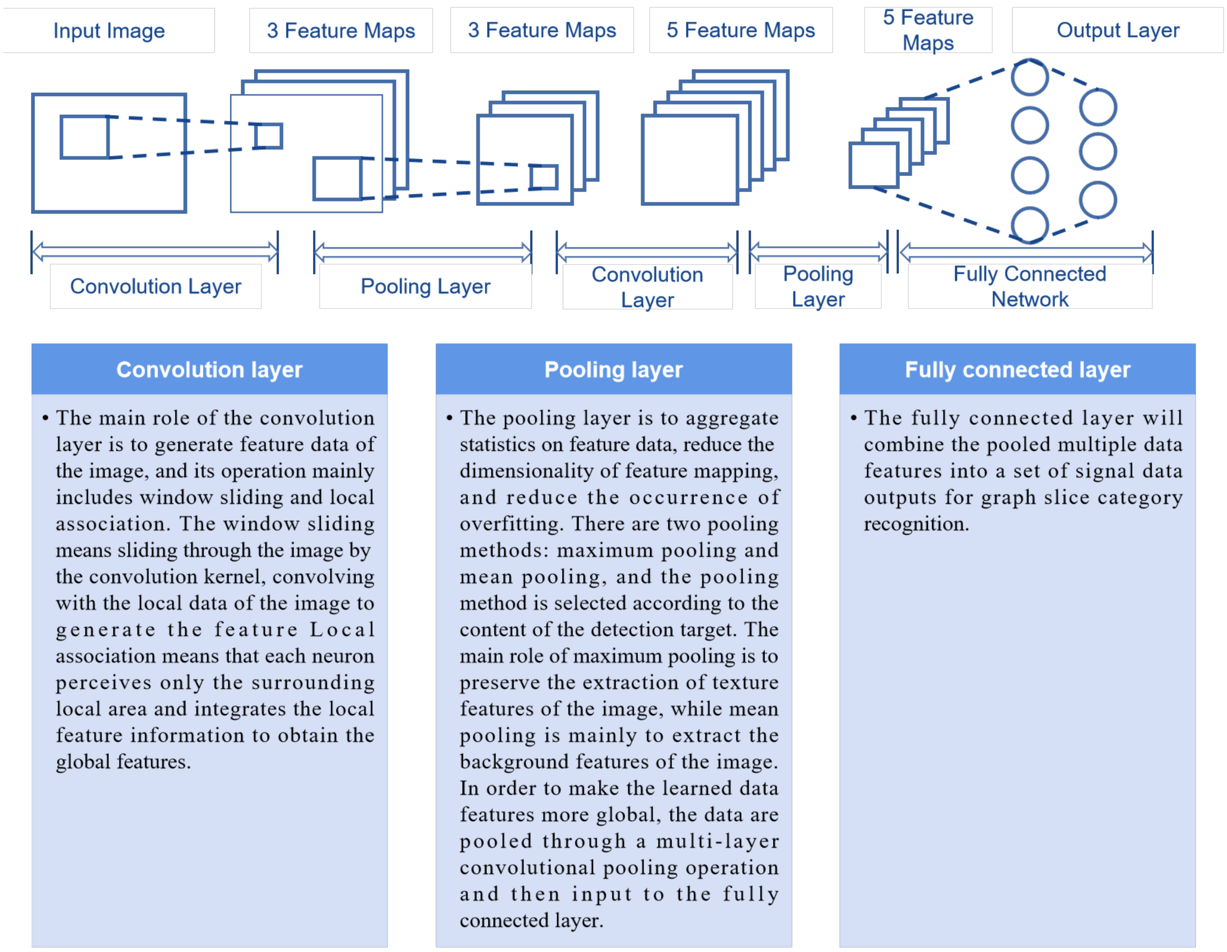

Figure 3. Schematic diagram of CNN network structure

3. Application of Deep Learning in Image Recognition

3.1 Deep learning-based license plate character recognition

With China's car ownership gradually increasing, license plate recognition plays an important role in the intelligent vehicle management system. Existing recognition algorithms are slow and inaccurate, and are easily misidentified by light and the angle of the license plate position relative to the fixed position of the camera. Compared with traditional recognition methods, deep learning-based recognition algorithms have more obvious advantages in terms of convenience, recognition accuracy and robustness. License plate recognition using deep neural networks can achieve automatic extraction of license plate image features with high accuracy rate.

At present, deep learning technology is the most widely used method in license plate recognition research. 2021, Xinwei Gao proposed Faster-RCNN based on deep learning for license plate localization, generating license plate extraction frame to extract license plates; using VGG16 network model to recognize characters, and finally completing the recognition of car license plates. The simulation results show that the recognition accuracy of license plate using Faster-RCNN combined with VGG16 is up to 99.2% in complex environment, and the recognition accuracy is better than other algorithms. [9]

3.2 Face recognition

Face recognition is one of the important applications of computer vision, and face recognition in a broad sense includes image acquisition, face detection, face alignment, feature representation, etc. Face recognition is widely used in many fields such as security, finance and e-government, etc.

Improving the performance of face recognition is very important for expanding the application field and practicality of face recognition. Traditional face recognition is limited in its accuracy due to factors such as illumination, pose and expression changes, occlusion, massive data and its own limitations. The emergence of deep learning has led to a breakthrough in face recognition technology, where learning algorithms learn discriminative face features directly from the original image, making it a great improvement in speed and accuracy. In 2021, Shuaifeng Huang gave an improved face recognition method based on LeNet-5, in which the structure of a typical convolutional neural network LeNet-5 is designed to consist of two convolutional sampling layers, a fully connected hidden layer and a classification output layer, which reduces the complexity of the network structure. In addition, the number of convolutional kernels is reduced, the pooling method and the classification output method are improved, and the computational complexity is reduced. It is demonstrated that the method improves the speed of single face recognition on the embedded platform while ensuring the training and testing accuracy. [10]

3.3 Medical image recognition

Medical images have the disadvantages of wide variety, low resolution, and heavy dependence on imaging equipment and imaging environment, which to a certain extent limit doctors to make effective diagnosis of patients' symptoms. The use of image processing technology combined with deep learning to diagnose the diseased parts of the human body is one of the most cutting-edge medical diagnosis methods today.

In recent years, the rapid development of deep learning models, especially the widespread use of CNN, has made it a mainstream trend to use neural network models to automatically extract and select features and perform classification. different variants of CNN models have been widely used in clinical disease diagnosis based on medical images, for example, based on Kaggle's public dataset of fundus images, Shanthi T et al. used improved AlexNet for the classification of diabetic retinopathy, and its accuracy can reach about 96.6%; based on VGG, the accuracy of benign and malignant classification of lung nodules using chest films can be as high as 99%. Currently, among the common CNN variants, ResNet and VGG have the best performance in medical image classification, so most of the tumor detection, brain neurological disease classification, and cardiovascular disease detection use these two models as baseline models for research. [11]

References

[1] Hinton G E. A practical guide to training restricted boltzmannmachines [J]. Momentum, 2012, 9(1):599-619.

[2] Zhangqi, Zhang Rongmei, Chen bin. A Review of Image Recognition Technology Research Based on Deep Learning[J].Journal of the Hebei Academy of Sciences,2019,36(03):28-36

[3] Bahrampour Soheil, Nasrabadi Nasser M, Ray Asok, et al. Multimodal Task- Driven Dictionary Learning for Image Classification. [J].IEEE Transactions on Image Processing, 2016, 25(1):24-38.

[4] Babri H A,Tong Y.Deep feedforward networks: application to pattern recognition[C]//Proceedings of International Conference on Neural Networks(ICNN’96), 1996:1422-1426

[5] Shao L, Wu D, Li X.Learning deep and wide:a spectral method for learning deep networks[J].IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(12):2303-2308.

[6] Zheng Yuanpan, Li Guangyang, Li Ye. Survey of application of deep learning in image recognition. Computer Engineering and Applications, 2019, 55(12) :20-36.

[7] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 7(18) :1527-1554.

[8] Zhang H, Yang B, Fang T, et al. Learning deep features for classification of typical ecological environmental elements in high-resolution remote sensing images[C]//Proceedings of 10th International Symposium on Computational Intelligence and Design (ISCID). Hangzhou: IEEE, 2017:223-227

[9] Gao Xiwei. Research on Vehicle License Plate Recognition Algorithm Based on Deep Learning[J]. Mechanical & Electrical Engineering Technology,2021,50(10):164-166.

[10] Huang Shuaifeng, Tang Lijuan, Liang Longbing. An improved embedded face recognition method based on LeNet-5[J]. Computing Technology and Automation,2021,40(04):91-96.

[11] Wang Lihui. Research progress and development trend of deep learning in medical imaging[J].Big Data Research.2020,6(06):83-104.