Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/8/20230287

Studies advanced in image classification based on deep learning

Hongwei Chen

Department of Materials Science and Engineering, Shenzhen MSU-BIT University, Shenzhen, Guangdong Province, 518172, China

1811031228@mail.sit.edu.cn

Abstract. Accurate and efficient image classification is one of the important research topics in image analysis, and it has also been a research hotspot in the computer vision community.Deep neural networks are increasingly being used for picture categorization and processing in recent years as a result of advancements in machine learning technology.In this article, we introduce the research progress of image recognition technology based on depth learning, including the design ideas, principles, structures, advantages and disadvantages of several depth neural networks. In addition, we quantitatively compared the recognition results of different methods on classic image classification datasets. Lastly, we review some current issues with picture classification and talk about the direction of deep neural network development in the use of image classification technology.

Keyword: image classification, feature expression, convolutional neural network, deep learning.

1. Introduction

The classification and recognition of target objects has always been the core content of image analysis, and is also one of the research hotspots in the computer vision community. Image recognition aims to build models to predict the category of specific objects in a given image. It is widely employed in a variety of industries, including driverless vehicles, intelligent agriculture, medical image recognition, and more. Massive volumes of picture data from many areas have been steadily accumulating in recent years due to the Internet's rapid expansion. This data includes a wealth of helpful information and has the potential to be valuable commercially. How to process and recognize images has become a major issue, attracting more and more researchers' attention.

For image recognition tasks, feature representation is the key to achieving accurate classification. Early image recognition methods mainly relied on manual features such as texture, color, contour, and so on. In traditional image recognition, only one type or one label, can be recognized, resulting in low recognition efficiency. Moreover, the probability of error in the process of image recognition is extremely high. In today's image more and more, a small change may appear the recognition error will be bigger. So the requirements for picture recognition technology are getting higher and higher. Only by improving the accuracy and efficiency value of recognition can we meet the needs of social development. These methods are time-consuming and laborious, and requires high technical requirements for designers. In addition, a major problem with such manual feature based recognition methods is their weak generalization ability, which means that the results may fluctuate significantly when the trained model is migrated to similar application scenarios. Therefore, although artificial design features have been widely used, they are not an extensible approach.

Image classification approaches, such artificial neural networks, active learning, support vector machines, and so forth, have evolved into artificial intelligence in recent years as a result of the progressively manifesting benefits of artificial intelligence in processing information. Compared to ordinary artificial neural networks, deep neural networks have more computational levels. They may significantly increase the accuracy of picture classification with a lot of unknown information by using statistical learning methods on vast data and extracting image information from the standpoint of computer vision. Therefore, deep neural networks are becoming a hot topic in image classification research.

Finding recognition methods suitable for different fields is extremely important. We should compare and study the characteristics of various image recognition technologies based on different situations to determine which one is better suited to the needs of experiments. Improving accuracy and increasing time efficiency are important indicators that we should comprehensively consider. We review recent research on deep learning-based image classification algorithms in this article to introduce the state of the art in image classification, with a focus on deep learning-based image recognition approaches. We also report and compare the performance of representative classification algorithms on some common datasets. Also, the primary issues in image classification research are highlighted. This is followed by a thorough examination of how image classification algorithms will be used and developed in the future.

2. Method

2.1. Overview of deep neural networks

The first attempt of artificial neural network can be traced back to 1958, where Rosenblat proposed a neural network model named perceptron consisting of two layers of neurons. However, experiments have found that the perceptron can only perform simple linear classification tasks. After then, Rumelhar et al. devised the renowned Back Propagation (BP) technique to address the issue of excessive computation brought on by raising the number of processing layers, which significantly broadened the range of solutions for non-linear classification jobs [1]. However, neural networks also face the problem of too long training time and network optimization difficulties caused by local optimization problems.

The deep neural network has had a resurgence with Hinton et al2006 .'s proposal of a deep neural network machine learning approach based on the notion of human brain learning. A multi-level neural network called a Deep Neural Network (DNN), sometimes referred to as Deep The output features of one layer are used as the input features for the subsequent layer in learning.After multi-level training of nonlinear transformations of original features, a deep neural network can transform the features of initial samples into another feature space based on demand, thereby learning feature representations that have better effects on existing inputs [2]. For data with very rich information dimensions, deep neural networks can more fully and accurately mine the potential information of massive data through the establishment of multi-level mathematical models. As a result of continuous research on deep neural networks, models including deep belief networks (DBNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) have been presented in recent years; these models will be discussed in more detail in the following subsections.

2.2. Deep belief networks

The Deep Confidence Network (DBN), an unsupervised learning model that can be taught layer by layer from bottom to top and optimized using a global back propagation strategy, is composed of many Restricted Boltzmann machines that are stacked [3]. The Restricted Boltzmann machine (RBM) is a network model that employs random states. The RBM only has two states, active and inactive, like a switch.The structure of RBM is different from other neural networks in that there is no connection between the cells within a single layer, while the cells between layers are fully connected. RBM can well analyze complex data, but for data with higher complexity or requiring deeper analysis, a multi-level network model DBN needs to be introduced.As seen in Figure 1, a series of RBMs will be acquired by continuous learning. The output of the trained RBM will be utilized as the input of the following RBM. Due to the difficulty of performing the global optimization of a DBN with several hidden layers, a layer-by-layer training technique is used, which may be broken up into two parts. Initially, just one layer of BRM neurons is trained at a time, and the current layer's output is utilized as the following layer's input.; Then, after all layers have been trained, use the wake-sleep algorithm for fine tuning.

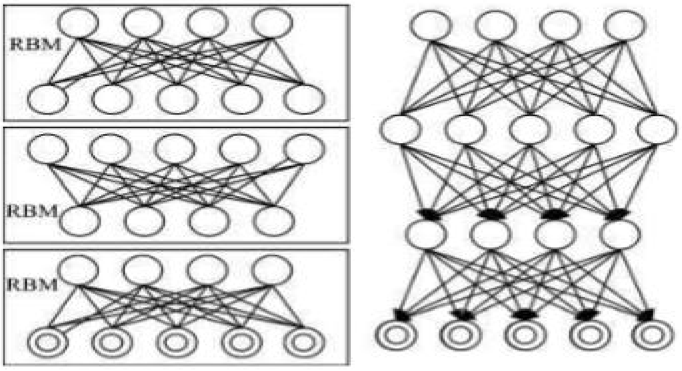

Figure 1. The structure of RBM and DBN.

2.3. Convolutional neural networks

Convolutional neural network (CNN) is a feedforward neural network, whose basic structure includes two layers [4-5]: (1) layer that extracts features. The portion of the upper layer is used as input to the current neuron, and the current neuron extracts the characteristics of the region, using the characteristics of the region to determine the corresponding relationship between it and other features; (2) layer of feature mapping. Several planar feature maps with identical weights for each neuron on the plane make up each computational layer of a CNN.

The foundation of convolutional neural networks is the input-output mapping connection.There was no distinct mathematical model linking input and output prior to learning.In order to build a model, CNN trains the convolutional network by discovering several mappings between inputs and outputs..In contrast to early manual features, CNN learns from training data and avoids explicit feature extraction. CNN networks may also learn concurrently since neurons on the same CNN plane have identical weights[6]. The four primary elements of CNN's basic network topology are the input layer, convolution layer, pooling layer, fully connected layer, and output layer, as illustrated in Figure 2.

(1) Input layer.An image processing technique's input layer typically reflects the picture's pixel matrix. The input layer is used to accept the given image data, which is often required with a three-dimensional tensors and combined with various data enhancement strategies, such as folding, flipping, etc.

(2) Convolution layer.The convolution layer, whose fundamental component is convolution operation, is the brain of the convolution neural network.The operation is defined as a inner product (multiplying and summing elements one by one) on images (different data window data) and filter matrices (a set of fixed weights2) A filter in CNN computes convolution on the local input data. The data window keeps panning and sliding after each local data computation until all data has been calculated.

(3) Pooling layer. The pooling procedure uses a location's output in the input matrix as the general statistical features of nearby areas, mostly using Average Pooling, Max Pooling, and other

methods.Pooling simply means specifying a value on the area to represent the entire area. Super parameters of pooling layer: pooling window and pooling step length. Pooling can also be seen as a convolution operation.

(4) Fully connection layer. The final classification results are often provided by 1 to 2 fully connected layers at the conclusion of CNN after numerous rounds of convolutional layer and pooling layer processing.After several iterations of convolution and pooling processing, it is possible to say that the information in the image has been abstracted into features with higher information richness.It is possible to think of the convolutional layer and pooling layer as an automated picture feature extraction procedure.The whole connectivity layer must still be used once the extraction is finished in order to complete the classification process.

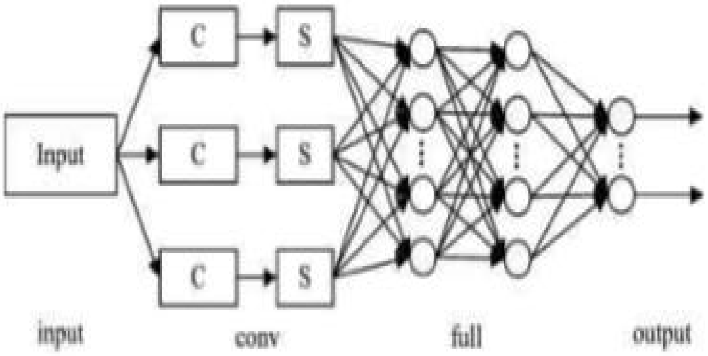

Figure 2. The basic structure of CNN.

2.4. Recursive neural network

Recursive neural network (RNN) is a special series of neural networks, with various feedback relationships between artificial neurons. The output of a particular layer of neurons is not only connected to the output of the layer before it, but also acts as that layer's input at the subsequent instant.For instance, the i-layer neuron's input at time t includes both the output of that layer at that moment and its own output at time t-1.

Recursive neural networks (RNNs) are built to handle sequential data, such a string of text or changes in the stock market. Text, sound, and video are just a few examples of the various sequential data formats that exist.A network's output at any given time is linked to its input as well as its output from a previous time or several occasions.These sequential sorts of data frequently exhibit temporal correlations. Recursive neural networks may be used to tackle a wide range of issues, including voice recognition, language modelling, machine translation, and other issues. These networks have specific memory functions that are strongly connected to sequences and lists. Recurrent neural networks can't manage this association as effectively as feedforward neural networks because they lack memory, hence the output of earlier moments cannot be communicated to later times.

3. Experiment and performance analysis

3.1. Common dataset

Several large-scale data sets, primarily MNIST, CIFAR, ImageNet, etc., have been suggested for the construction of supervised classification models [7-8]. The MNIST dataset, which includes 60k training pictures and 10k test images with a size of 28×28, is a collection of handwritten numbers from 0 to 9. The CIFAR dataset includes CIFAR-10 and CIFAR-100. The first has 50,000 training photographs and 10,000 test photos from 10 categories, while the second contains 50,000 training photos and 10,000 test photos from 100 categories.The ImageNet dataset contains over 14 million full-size tagged images with approximately 22000 categories.

3.2. Performance

To compare the results of different methods on common datasets, we quantitatively report the results of representative methods on ImageNet datasets, as shown in Figure 3. It can be observed that the classical ResNet can achieve a accuracy of 87.6% on the ImageNet dataset, whose performance can be further improved to 91.2%. Compared with the ResNet, the NFNet replaces the Batch Normalization in ResNet with the adaptive gradient clipping technique, which can boost both the recognition speed and accuracy. All the results show that the existing image recognition networks have progressively tended to be mature and can achieve results of remote human classification errors.

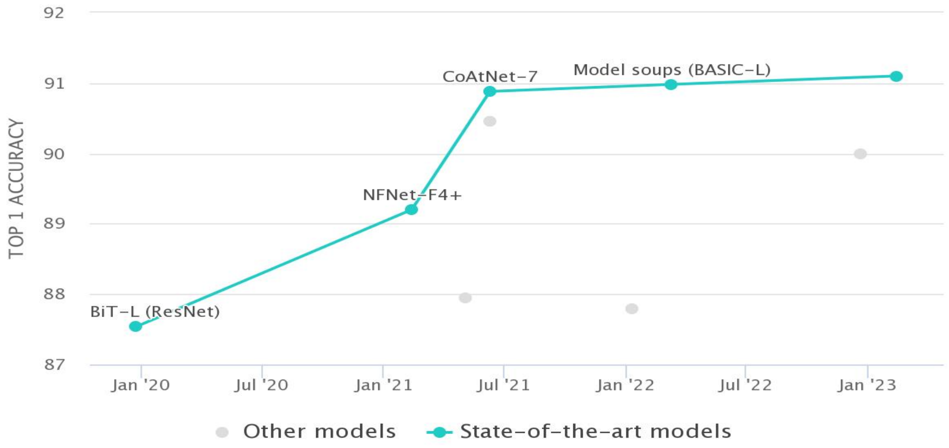

Figure 3. The recognition accuracy of representative models on ImageNet data set.

4. Discussion

Although the rapid development of computers has provided many methods for image recognition, there are still many problems to be solved. The problem of light weight images can only be used in places with better graphics cards or equipment. Therefore, we should improve the related technologies to help the group to use the portable devices so that they can be better used in life. When the data set is processed, there may be a problem that when there is a large difference in the distribution of the number of types of pictures, the recognition will cause a large error. The program cannot identify exactly what data we want. Moreover, when the number of data sets is large, the time required for training and testing is too long, resulting in low efficiency. In the process of image processing, there are some methods that lose useful information, which leads to problems in the final recognition. Therefore, we cannot set up a program to achieve the best results for image recognition in a certain category or field.

Regarding the field of image recognition application: First, it can be applied to the recognition of license plates. With more and more vehicles these days, the annoyance and non-compliance with traffic rules also increases. Therefore, a complete license plate recognition system is needed to better supervise the behavior of vehicles and maintain the normal operation of society. People's desire for a better life has become stronger with the growth of the economy, yet occasionally we glimpse a picture without knowing where it is.Therefore, it is necessary to establish a landscape image recognition system to help people better understand the world. A nation's strength is determined by its citizens' fundamental standards of living and whether those in need receive assistance, not by the standard of living of its wealthiest citizens.Now, the number of hearing-impaired people is increasing. According to the data of the World Federation of Hearing Impaired [10], there are currently more than 7×10^7 people in the world using more than 300 sign languages to communicate. According to relevant media data, the number of language barriers and hearing disabilities in my country exceeds 3×10^7 [9-10]. Hence, in order to better integrate them into society and increase their happiness, it is vital to build a sign language picture recognition system to aid them in communicating with everyday people.As the quality of living in today's society improving, individuals are becoming more self-conscious.For everyday wear seems to have become an essential issue every day. In order to help individuals with the problems of not understanding how to dress in life, we can automatically match the matching of garments and pants based on the gathering of numerous matching data sets. Or you can match the most appropriate outfit according to each person's body shape and face shape.

In the field of image recognition, there is still a lot of room for us to study and continue to improve. Finding the most suitable image recognition system that meets the development needs of each field has become a direction that needs continuous improvement and progress in the future. Future research will need to focus on ways to lower recognition error and increase recognition efficiency. On the original basis to go on improving and innovation, to help people live a better life.

5. Conclusion

Nowadays, great progress has been made in image recognition technology, which is also used in various industries. The accuracy of recognition has been greatly improved. Processing time is also very efficient. In this article, we highlight recent developments in the field of deep learning-based image identification, including representative algorithms, design principles, and benefits and drawbacks.However, each recognition method has different effects in different fields, so sometimes it is impossible to quickly find an optimal solution. In a new field, which method we choose may have the highest efficiency. In the study of several identification methods introduced in this paper, basically every experiment adopts a comparison method to select the optimal solution process, which maximizes the feasibility of the experiment. It can be seen that a relatively mature system has been formed in the field of image recognition, which is more closely related to real life. I believe that the development of computer recognition technology can enable people to live a better and convenient life and better enjoy life.

References

[1] Jiang Li. 2013. Research on BP Network Learning Method Based on Particle Swarm Optimization and Simulated Annealing Algorithm [D]. Anhui University. Doctoral thesis.

[2] Taylor G W, Hinton G E, Roweis S T. 2006. Proc. Int. Conf. Modeling Human Motion Using Binary Latent Variables//Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4-7.

[3] Sang K K, McMahon, Olukotun. 2009. Proc. Int. Conf. A highly scalable Restricted Boltzmann Machine FPGA implementation[C]// International Conference on Field Programmable Logic & Applications. IEEE.

[4] Goodfellow, I., Bengio, Y., Courville, A.． J. 2016. Deep learning (Vol. 1)． Cambridge MIT press, 2016: 326-366

[5] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang L, Wang G and Cai J. 2015. J. Recent advances in convolutional neural networks. arXiv preprint arXiv:1512.07108.

[6] Cao Tiancheng. 2022. J. Multi label natural scene image recognition. Digital Technology and Applications, 2022,40 (11): 28-30.

[7] http://yann.lecun.com/exdb/mnist/

[8] https://www.cs.toronto.edu/~kriz/cifar.html

[9] MURRAY J. World federation of the deaf[EB/OL]. http://wfdeaf.org/our-work/, 2020.

[10] Gu Chong. Language barrier recovery needs more than 30 million people in China [EB/OL].

[11] https://m.btime.com/item/router?gid=40ea0atodav8fk8lcqahek43ilu, 2020.