Research Advanced in Technology and Applications of Image Recognition Based on Deep Learning

Zihan Lu [1], Guanyu Qiao [2,*], Shaozhe Zhang [3]

[1] Nanjing Forestry University, Nanjing, Jiangsu Province, 210037, China

[2] University of Liverpool, Liverpool, Merseyside, L69 3BX, United Kingdom

[3] Adcote School Shanghai, Shanghai, 200000, China

* Corresponding Author Email: Guanyu.Qiao22@student.xjtlu.edu.cn

Abstract. Image recognition is a fundamental research problem in the field of computer vision, which aims to build models to recognize the category of a given image. Early image recognition methods mainly relied on artificially designed features, and their performance could not meet actual application needs. Thanks to the rapid development of deep learning technology, convolutional neural networks can extract deep information from images to better represent image semantics. Therefore, image recognition based on convolutional neural networks has attracted a lot of research attention in recent years and has been widely used in many fields. Focusing on image recognition and its applications, this paper first introduces common image recognition frameworks, and focuses on the latest research progress in image recognition based on deep learning. Then, this article summarizes the scientific research results of deep learning in agriculture, medicine, transportation and other application fields in recent years. Finally, this paper discusses the development trend of deep learning in the field of image recognition, and expounds the future development trend of image recognition.

Keywords: Image recognition; deep learning; applications.

1. Introduction

Image recognition has always been a basic research issue in the field of computer vision, which is the technical basis for many downstream visual understanding tasks and has attracted a lot of research attention. As the most accurate and largest source of information, the value of image recognition is further highlighted in the Internet era led by digitalization and intelligence. Massive image data contains rich knowledge, which can help humans relieve work intensity and improve work efficiency.

Image recognition technology can be traced back to the 1940s. However, limited by insufficient technology and imperfect hardware facilities, image recognition technology did not begin to flourish until the 1990s. Traditional image recognition methods are mainly based on the idea of machine learning, which extracts low-level features at the pixel level to describe image content and relies on classifiers to realize recognition. Representative traditional image recognition algorithms include backpropagation algorithm, Bayesian classification method, etc. These methods mainly use shallow-level structural models and require manual preprocessing of images during feature extraction, which inhibits the accuracy of image recognition. Thanks to the powerful feature representation capabilities of convolutional neural networks, image recognition methods based on deep learning transform the understanding of image content into the semantic understanding of low-level features and map them to high-level domains layer by layer, which further promotes recognition accuracy and speed. With the continuous optimization of the convolutional neural network structure, the recognition accuracy of general images has exceeded 99%, and the recognition speed has reached 5 times that of manual recognition. In addition to its excellent performance in terms of data, the intelligent identification system also has the advantages of a longer working cycle, lower cost, and the ability to work in dangerous environments.

At present, image recognition has been widely used in all aspects of human social life such as military affairs, transportation, biomedicine, agriculture, and automation. Taking agriculture as an example, broccoli is a plant that is extremely prone to over-ripening and rotting, so the identification

of its maturity is crucial. The traditional manual identification method not only consumes a lot of manpower and material resources, but also easily leads to missing the best picking period because it cannot be monitored 24 hours a day. However, using image recognition can not only solve these problems, but also make individual picking suggestions for each region of broccoli, or even each plant. In terms of transportation, using car keys to open doors also faces many problems, such as the loss of keys and battery limitations. However, using image recognition to analyze car owner characteristics can not only solve the problem of lost keys, but also ensure vehicle safety. In medicine, there are often many symptoms that are difficult to diagnose with the human eye, and parts of the body that are difficult to observe with the human eye. Or issues that are not suitable for diagnosis by a real person due to gender and privacy considerations. In these cases, image recognition systems can make recommendations to doctors.

Based on the knowledge of existing literature, this paper will review the development, application, advantages, disadvantages, and prospects of image recognition technology. Specifically, this paper first introduces common image recognition frameworks and focuses on the latest research progress in deep learning-based image recognition. Then, it summarizes the scientific research achievements of deep learning in agriculture, medicine, transportation, and other application fields in recent years, and analyzes the shortcomings in the research process. Finally, the development trend of deep learning in the field of image recognition is discussed, and the future development trend of image recognition is elaborated.

2. Representative image recognition methods

2.1. Development of Image Recognition

Image recognition is an important technology in the fields of artificial intelligence and computer science. It is often applied for identifying various patterns and objects in images. The development of image recognition technology has gone through three stages: text recognition, digital image processing and recognition, and object recognition. The earliest stage of text recognition research dates back to 1950. Digital image processing and recognition have a history of nearly 50 years, and their characteristics such as compressibility, ease of transmission, and minimal distortion have provided significant momentum for the rapid development of image recognition technology. Object recognition is a technology based on the perception and cognition of the environment and objects in three-dimensional space within the field of computer vision. Today, the flourishing field of artificial intelligence combines object recognition, digital image processing, and recognition techniques, leading to functional applications in various industries and domains, making it a focal point in modern society.

2.2. Main Methods of Image Recognition

2.2.1 Fractal Features

The concept of "fractals" originated from a paper titled "How Long Is the Coast of Britain?" published by a French-American mathematician in 1967. From a mathematician's perspective, irregular and non-smooth mountains and rivers are difficult to describe using traditional Euclidean geometry. When idealized tools are used for measurement, the result tends to increase infinitely. These theories inspired mathematicians to appreciate the irregularity of curve lengths and eventually led to the creation of a new field in geometry – fractals. However, fractal features are generally applied to images with natural textures that exhibit fractal characteristics for texture segmentation and recognition. Some artificial objects do not possess fractal characteristics, limiting the application of image recognition based on fractal features [1].

2.2.2 Wavelet Moments

Wavelet moments, formed by combining wavelet analysis and invariant moments, possess both the good time-frequency localization characteristics of wavelet analysis and the geometric

transformation invariance of invariant moments. They were widely used in early computer vision and image reconstruction fields. However, the effectiveness of image recognition using wavelet moments fluctuates significantly within different threshold ranges, making it challenging to establish a standardized threshold. Additionally, the computational complexity of wavelet moments limits their continued use in rapidly evolving image recognition technology [2].

2.2.3 Neural Networks

Neural networks are mathematical models of a large number of perception and learning-capable neurons found in biological brains, aiming to mimic the characteristics of neural networks in humans and animals. Convolutional Neural Networks (CNNs) are a type of feedforward neural network with convolutional operations and deep structures. They consist of input layers, intermediate layers, and output layers. CNNs have outstanding image processing capabilities, making them one of the most commonly used technologies in modern computer vision and image recognition. When using CNN for image recognition, the image is directly input into the model, and the structure of the image itself can be preserved without the preprocessing and feature extraction process in the traditional algorithm, thereby reducing the complexity of model processing [3]. The difference from other neural networks is that the matrix multiplication operation in one or more layers of CNN is replaced by a convolution operation, which reduces a large number of parameters by using the advantages of multi-layer neural network and image locality, and improves model training speed [3].

The basic structure of CNN is a feature extraction layer and a feature mapping layer. The local features are extracted by connecting the receptive fields between the layers. The feature map structure mainly uses the Sigmoid function to operate the convolutional neural network to ensure its displacement invariance. Popular CNN models include LeNet, AlexNet, and GoogleNet, among others. Taking LeNet as an example, its structure is shown in Figure 1. CNN usually consists of input layer, intermediate layer and output layer. As the most basic and important layer in CNN, the convolutional layer convolves or multiplies the pixel matrix of a given image to generate an activation map for the given image. Another important network layer is the pooling layer, which calculates the local average and extracts features twice, which can reduce the feature resolution. Initially, CNN was widely used in target recognition tasks, and now it also shows excellent performance in target tracking, pose estimation, text detection and recognition, visual saliency detection, action recognition, scene labeling and other tasks.

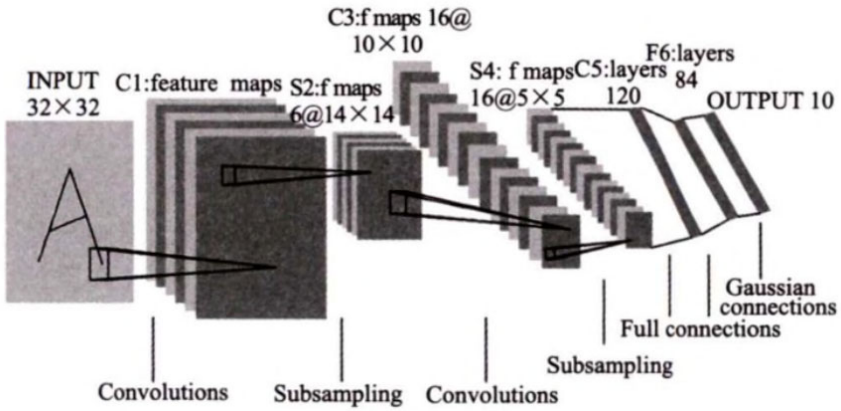

Figure 1. Structure diagram of LeNet-5

3. Application of image recognition

3.1. Application of image recognition in agriculture

Agriculture, as the most basic and largest industry among the three major industries, occupies a very important position. Therefore, the application of image recognition can improve agricultural production efficiency and reduce the number of hungry people. In agriculture, image recognition is

commonly used to identify the maturity of crops. Compared with traditional manual recognition, intelligent recognition increases the recognition speed by 400% while sacrificing some accuracy. There are three main neural network models when it comes to identifying maturity according to the differences in the recognition tasks.

(1) Classification-based. Classification neural networks detect plant maturity by assigning category labels to different stages of flower morphology, fruit color, etc. Features are generally extracted through a deep convolutional neural system (DCNN), converted into a matrix, and compared with labels. Khosravi et al. successfully identified four growth stages of two olive varieties through RGB images and DCNN [4].

(2) Detection-based. The object detection neural network identifies the growth stages of target individuals in images by assigning different category labels to the growth stages of the same plant. Compared with the classification neural network, although the target detection neural network is not universal between different types of plants, it has higher accuracy when identifying a single plant and can distinguish the target crop from a group of crops. Psilukis et al. effectively identified and classified the three degrees of doneness of broccoli by building Faster R-CNN and CentreNet [5].

(3) Segmentation-based. Segmentation neural networks are also used to evaluate the status of flowers or fruits. Tian et al. used the Mask R-CNN algorithm and the MASU R-CNN algorithm to segment apple blossoms of three maturity levels with a success rate of 96.43% [6].

Among all neural network models, RetinaNet has the highest accuracy and fastest speed. In a test, Faster R-CNN completed the maturity analysis of broccoli in 9 seconds with an accuracy of 83%. RetinaNet completed the classification in 6.5 seconds and won in accuracy by 14 percentage points, with an accuracy rate of 97% [7]. The reason why RetinaNet is powerful lies in the function Focal Loss it uses. Traditional R-CNN is generally two-stage. The first step is to delete useless information, and the second step is to compare the information. RetinaNet abandoned this structure and instead used Focal Loss to process useless information, and then used CNN to analyze it. Due to the powerful exclusion capability of Focal Loss, RetinaNet is more comfortable in subsequent analysis.

Although the intelligent identification system can identify very efficiently, it still faces many problems. One of the difficulties is the identification and analysis of rare crops. In the case of ackee fruit, due to its high rarity, there is no complete dataset for training. For this problem, rare plants can be planted artificially and then the data set can be completed through a large number of cases, or the intelligent identification system can be continuously updated and iterated during the identification process. Then again, there isn't an exact measure of plant maturity. This limits the intelligent identification system to the task of assisting picking. Neither scientists nor administrators can conduct high-precision data analysis through the identification system. For this problem, you can create a standardized index to express maturity, such as dividing the cycle from the first picking to the second ripening of the plant into 1000 equal parts, using 0 and 1000 to represent freshly picked and fully matured plants respectively. plant. This is not only convenient for scientific experiments, but also convenient for consumers to choose their favorite maturity.

3.2. Facial Recognition

Face recognition is the most representative application of image recognition, which aims to use computer learning to extract personal information through facial features. Due to its good stability, face recognition has gradually been applied to monitoring systems, smart payment, public security systems and other applications. In addition, facial expression recognition, as a derivative task of face recognition, has also received extensive attention in recent years.

Method 1: Convolutional Neural Network (CNN) models are widely used in image classification, object detection, feature extraction, and other areas. They are a type of supervised learning network model with autonomous learning capabilities. Traditional CNNs cannot extract local facial features effectively. To enhance the accuracy of facial recognition, Wang Jiaxin and Lei Zhichun proposed an algorithm that combines facial feature fusion with convolutional neural networks for facial recognition [8].

Method 2: Facial emotion recognition, built upon facial recognition, involves two processes: feature extraction and classification. Significant breakthroughs have been made in deep learning-based research on facial emotion recognition. However, most images in databases are ideal with good lighting, frontal views, and clarity, making them limited in natural environments due to factors like lighting conditions and shooting angles. Addressing these issues to improve recognition accuracy, Wang Yuqi combined an encoder-decoder framework-based algorithm with an attention mechanism, reducing the loss of facial details [9].

3.2.1 Facial Recognition

Using convolutional neural networks based on image recognition technology, facial features can be extracted. To enhance the efficiency of facial feature recognition, LBP (Local Binary Pattern) operators and DCT (Discrete Cosine Transform) techniques can be incorporated [8]. LBP operators describe local feature textures, allowing effective classification of image detail features. By compressing facial image data with DCT, it can be observed that the low-frequency portion contains most of the energy. Image reconstruction through the extraction of DCT low-frequency coefficients can preserve most facial feature information with some error compared to the original image [8]. Additionally, to improve facial recognition rates, an efficient facial recognition algorithm can be formed by combining convolutional neural networks with feature fusion. After training and feature classification, it can effectively extract facial feature information, achieving the recognition and extraction of facial information characteristics of drivers.

3.2.2 Facial Emotion Recognition

Facial emotion recognition, built upon facial recognition, involves two processes: feature extraction and classification. Significant breakthroughs have been made in deep learning-based research on facial emotion recognition. However, most images in databases are ideal with good lighting, frontal views, and clarity, making them limited in natural environments due to factors like lighting conditions and shooting angles. To address these issues, Wang et al. proposed an algorithm based on the encoder-decoder framework, combined with attention mechanisms and Long Short-Term Memory (LSTM) [9]. This approach accurately identifies facial features. The encoder in this framework uses an RNN (Recurrent Neural Network). RNN+LSTM is used for dynamic recognition of facial expressions, followed by feature extraction using Convolutional Neural Networks (CNN). Through this method, we can apply image recognition technology based on CNNs and recurrent neural networks, particularly LSTM, to drive facial emotion recognition. This can help determine their emotional state and provide reminders related to emotions and behavioral safety [9].

3.3. Application of image recognition in medicine

3.3.1 Image recognition on Lung tumor

The research report in recent years states that cancer still poses a great threat to people all over the world while they do expect to live a better life now and in the future. As far as the data shows, the rate of men who suffer from the lung tumor is 31.5% [10]. At the same time, the death rate has reached over 27%, which can be reduced if doctors can diagnose the diseases quickly and correctly. However, due to the lack of experience, degree of fatigue, no clear early symptoms of lung tumor, some diagnoses can be inaccurate. As a result, image recognition is used as the auxiliary means to extract the key information rather than miss them and provide diagnosticians with comprehensive and objective information, offering convenience to all medical workers.

It is possible to quantitatively assess the mitochondrial structure of lung tumor cell lines to achieve high classification accuracy between healthy mitochondrial tissue and other tumor cells if CNN-based image recognition technology is being used, which simulates the function of the human brain. This method can improve the higher recognition rate and allow for earlier diagnosis of the patient's state of illness, making early treatment reduce the mortality rate caused by lung tumors. Meanwhile, multiple medical data such as text, image, 3D data, intermedia data, and video data are converted into a unified slide format. Although, it may lead to noise and information value issues. Therefore,

preprocessing of the original images is not unnecessary, and the image interpolation technique is used to make the images readable and ensure that the data enters the model training with decreasing difficulty and less cost. What's more CT images, PET images, and CT/PET fusion images are used in the experiments to prevent the overfitting caused by insufficient data sets and all of three methods can achieve accuracy rates, which can reach over 90% [11].

In addition, based on the idea of standard back-propagation (BP) learning algorithm and PCA used to identify the characteristic gases of lung cancer, the presence of lung tumors can be easily identified. At the same time, with multiple algorithms and diverse approaches, doctors can use increasingly comprehensive and analyzable data processing methods to diagnose cancer quickly and accurately, which is good for the patients who are suffering from the pain of cancer.

3.3.2 Image recognition on COVID-19

COVID-19 features rapid transmission and strong infectivity in the early stage of the infection. Therefore, whether doctors can recognize abnormal pulmonary nodule states in the lung and distinguish COVID-19 infection accurately and correctly can improve the accuracy and reduce the probability of false positive reactions.

In the era of big data and big information, the acquisition of medical images is easier than ever. However, due to the wide variety of medical images, there may be missing data in the specific areas of medical images. Meanwhile, even if a lack of data exists, it is possible to synthesize valid medical image data through GAN and synthesize more valid data effectively and efficiently [12]. At the same time, with the use of the Keras structure, selective modifications can be made to the structure of the convolutional neural network to create the system with self-learning training samples, which extract images from the training set, split their content and names, convert them into data, and train the model. Then, input the random image to test the trained model, optimize and classify the training results. Based on the procedure that had been done, use the InceptionV3 function for the transfer learning. At last, output the training results and prediction, which can reduce the possibility of the manual error. The overall data includes two types of datasets (training data and validation data), both of which have a correct recognition rate of up to 96% for infected and uninfected individuals.

There are also detections based on the two convolutional neural network models (GoogleNet and ResNet) [13], which is widely applied in the field of medical image recognition in the two rounds of experiments. The first round of experiments was conducted to build an experimental model that meets the experimental conditions. In the second experiment, the setting parameters achieved a high accuracy rate, the results of which showed that GoogleNet achieved over 90 % accuracy in most situations. With sufficient data support, it is estimated that the stability and accuracy of COVID-19 image recognition can be guaranteed. Compared to the application of ResNet, it not only has a high accuracy rate, but also reduces the training time by almost 55 %. Both experiments above meet certain clinical requirements and can reduce the failure rate of human judgment in clinical diagnosis, which has significant clinical implications for COVID-19.

4. Challenges and Countermeasure

4.1. Challenges

Nowadays, the public is paying increasing attention to their privacy rights, which may result in the lack of guarantee of quality and quantity, especially in the practical application of image recognition. Secondly, professionals in some areas may not be proficient in computer-related knowledge. Meanwhile, it is difficult for them to explain the professional perspective with their aspects of knowledge, which may lead to resistance towards image recognition, Therefore, it has a long way for the widespread application of image recognition. Third, new technology lacks the relevant laws and regulations. Due to the lack of laws and regulations, it is hard to determine individual responsibility and obligation in the accidents and it may cause conflicts between individuals [14].

4.2. Countermeasure

First, in the measure of collecting data, despite the limited amount of data in the single database, the sharing and collaboration of image recognition data from multiple local or multi-domain databases can solve the problems of obtaining data effectively and efficiently and avoid some kind of problems like overfitting, which is caused by 'quality and quantity' problems. What's more, with the use of GAN technology, it is easy to generate higher-quality image recognition data. In most cases, the same effects as those mentioned above can be achieved. At the same time, collaboration in diverse fields is needed. The collaboration can not only help the experts in the various fields understand the basic principles of image recognition, but also improve the machine learning models in order to apply the image recognition in the multiple fields. Then, a pilot run of the image recognition before implementing it world-while can attract more citizens to use and consolidate the trust of users and those being served in the application of image recognition. Lastly, experts in various fields should collaborate with each other to address the gaps in the relevant laws and regulations. This ensures the basic rights and interests of users in the application of image recognition and establishes a detailed division of responsibilities when unexpected accidents occur, which provides the legal basis for the application of the technology in multiple fields.

5. Conclusion

Image recognition has always been a fundamental task in the computer vision community and a foundation for many high-level visual tasks, such as object detection, tracking, segmentation, etc. It is not rare to see that image recognition can play a positive role in various fields such as agriculture, transportation, and medicine, not only relieving the work pressure of practitioners but also improving working efficiency. Compared with the relevant research, this paper provides a description and summary of the future and challenges encountered in the application of image recognition systems based on convolutional recognition networks in the diverse fields. Specifically, this paper systematically introduces representative methods of image recognition in the fields of agricultural pest and disease recognition, facial recognition, and medical image recognition. This paper also summarizes the challenges of image recognition in different application scenarios and proposes prospects, which we believe can bring new insights to the development of the field of image recognition.

Author's Contribution

All the authors contributed equally, and their names were listed in alphabetical order.

References

[1] Li Honggui, Li Xingguo, Li Guozhen, et al. Infrared image recognition based on fractal feature. Infrared and Laser Engineering, 1999(01):22-26.

[2] Zhou Yi. Research on the image content ecognition techonology based on wavelet moment. South Jiaotong University, 2010.

[3] Gai Rongli, Cai Jianrong, Wang shiyu, et al. Research review on image recognition based on deep learning. Journal of Chinese Computer Systems, 2021,42(09):1980-1984.

[4] Khosravi, H.; Saedi, S.I.; Rezaei, M. Real-time recognition of on-branch olive ripening stages by a deep convolutional neural network. Sci. Hortic. 2021.

[5] Psiroukis, V.; Espejo-Garcia, B.; Chitos, A.; Dedousis, A.; Karantzalos, K.; Fountas, S. Assessment of Different Object Detectors for the Maturity Level Classification of Broccoli Crops Using UAV Imagery. Remote Sens. 2022, 14, 731.

[6] Tian, Y.N.; Yang, G.D.; Wang, Z.; Li, E.; Liang, Z.Z. Instance segmentation of apple flowers using the improved mask R-CNN model. Biosyst. Eng. 2020, 193, 264–278.

[7] Li Jingbo, Li Changchun, Fei Shuaipeng, Ma Chunyan, Chen Weinan, Ding Fan, Wang Yilin , Li Yacong, Shi Jinjin, Xiao Zhen. Wheat Ear Recognition Based on RetinaNet and Transfer Learning

[8] Wang Jiaxin, Lei Zhichun. A convolutional neural network based on feature fusion for face recognition. Laser ＆ Optoelectronics Progress, 2020,57(10):339-345.

[9] Wang Yuqi. Research on Facial Emotion Recognition Based on Deep Learning.China Academic Journal Electronic Publishing House, 2020(14):36-37+63.

[10] Lu Hui, LI Feng, Hu Qing, PANG Zhonghao, Chen Shengjie, Jiang Zhihua, Gu Tianyi, Ji Hongyun. Ability of artificial intelligence in differentiating benign and malignant nodules of early lung cancer. Chian Academic Journal Electronic Publishing House, 2021, 26(12):1870-1875.

[11] Liang Mengmeng, Zhou Tao, Xia Yong, Zhang Feifei, Yang Jian. Multimodal lung tumor image recognition based on randomized fusion and CNN. Journal Of Nanjing University(Nature Science), 2018,54(04):775-785.

[12] Zhong Congyue, Yang Xiaoling. COVID-19CT Image recognition system based on convolutional neural network. Comupter and Information Technology, 2022,30(03):12-14+40.

[13] Wang Qiyao, Wang Jiangqing. Deep learning based novel coronavirus pneumonia CT image recognition. China Computer & Communication, 2020, 32(17):62-64.

[14] Wang Tianren, Li Yining,Wang Hongyi, et al. Current status and research progress on mediacal image data augmentation technology. China Academic Journal Electronic Publishing House, 2021,28(03):34-37+44.