Image Recognition based on Convolution Neural Network

Xin Hu ,Hua Ouyang ,Yang Yin
College of electrical engineering, Naval University of engineering, Wuhan, China
1810347286@qq.com, 78ouyang@163.com, reeyan@163.com
Corresponding Author: Yang Yin Email: reeyan@163.com

2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC) | 978-1-7281-5244-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ITAIC49862.2020.9339197

Abstract—**This paper introduces the development of image recognition technology, summarizes several classic traditional algorithms, and then introduces the use of convolution neural network for image recognition, and summarizes the application and development of CNN in various fields in recent years, and proposes to recognize images by deep learning in complex or high-risk environment similar to substation Don't take the place of manpower to detect the working state of instruments and equipment, which has a develop-ment prospect. In the end, it puts forward some suggestions on the direction of image recognition using deep learning. In the future, we need to improve and optimize the model, inte-lligent algorithm and data set.**

Keywords—image recognition, deep learning, convolution al neural network, data set, target detection algorithm

I. INTRODUCTION

Image recognition develops step by step from the earliest word recognition to the later digital image processing and recognition as well as the very popular object recognition. Its mathematical essence belongs to the mapping problem from pattern space to category space. The earliest image recognition technology can be traced back to the 1940s. The main steps of traditional image recognition include image acquisition, image preprocessing, feature extraction and image recognition. In the 1990s, deep learning gradually emerged. Because of its strong learning ability and efficient feature expression ability, image recognition technology has a great development. In recent years, it has been widely used in transportation, communication, medical and other fields. This paper mainly discusses the development of image recognition based on deep learning in recent years and how to use it The convolution neural network of gate is used to recognize the target.

II. THE MODEL OF CONVOLUTION NEURAL NETWORK

A. Tradition algorithm

traditional image recognition is generally divided into two parts: image feature extraction and classification. The traditional method has two prominent shortcomings: first, it needs to rely on manual design, and often relies on the knowledge accumulated by the prior, which is relatively blind. Secondly, the classifier structure is simple, and it may not be able to map the complex features to the final recognition results, which has certain limitations.

Sift, a feature extraction algorithm, was proposed by David Lowe in 1999 [1] and improved in 2004. This transform has scale invariance and is essentially a local feature descriptor. It can determine the location and scale of key points by using the position of interest points in image detection. Based on the local gradient direction of the image, the direction is assigned to each key point, and the local gradient of the image is measured in the field of each key point, and finally expressed by a feature vector. After that, bay, H., tuytelaars, T. and van gool, L jointly proposed surf [2] in 2006. Similar to sift, surf must also consider how to determine the location of points of interest. The difference is that sift uses dog instead of LOG operator to find its local extremum in scale and image as feature points, while surf method is based on Hessian matrix, which greatly reduces the operation time by integrating images. There are decision trees and support vector machines (SVM) which are well-known. The original support vector machine algorithm [3] was invented by Vla-dimirn. Vapnik and Alexey ya. Chervonenkis in 1963. Support vector machine is a classic binary classification algori-thm, and its hyperplane is more robust. In recent years, SVM has been continuously improved, making it widely used in deep learning. However, with the rapid development of comp-uter technology, the complexity and diversity of images are improving. The traditional image recognition technology is not suitable for processing highresolution data images in various situations. The accuracy of traditional algorithm for extracting local feature points in massive image recogni-tion is relatively low, and the calculation is difficult. The image recognition technology based on deep learning can dire-ctly process the input image, avoid the process of extracting image features one by one, and greatly reduce the amo-unt of calculation, so it has been recognized and applied in the professional field. The convolution neural network has become a research hotspot in the field of image recognition.

B. The proposal of CNN

Convolutional neural network (CNN) was initially inspired by the perception of biological visual cell

differentiation. Biological observation objects are processed by hierarchical visual cortex, and the color, texture and contour are judged by layer by layer scanning, and then the higher-level features of objects are abstracted to recognize objects. In 1959, Hubel [4] found that there are simple cells and complex cells in the primary visual cortex of cats, which undertake different levels of visual perception. These cells are very sensitive to the sub regions of the visual input space, which we call the receptive field. The convolution neural network was formally proposed by Yann Lecun [5] in 1998. The backpropagation algorithm was introduced into the convolutional neural network, which has achieved great success in hand-written digit recognition. Furthermore, convolutional neural network has really started a new chapter of deep learn-ing. Various improved models based on convolutional neural network have been proposed and applied in various fields.

A typical convolutional neural network is composed of convolution layer, convergence layer and full connection layer like Fig1. The structure diagram is as follows:

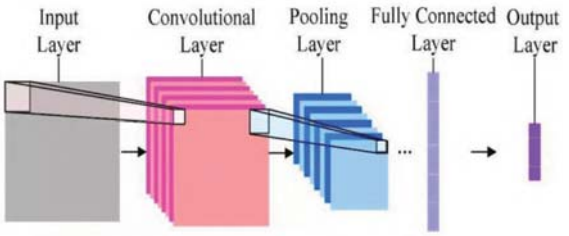

Fig.1 Convolutional Neural Network Structure

During the development of convolutional neural network step by step, many classic network structures have been born. Here, five representative CNN classical networks are listed according to the time line

1. Lenet was proposed by Yann Lecun [6] in 1998, and the MNIST training set was successfully trained. Moreover, the handwritten numeral recognition system based on lenet-5 was used by many American banks in the 1990s to recognize handwritten digits on cheques.

2. In 2012, alexnet proposed by krizhevsky [7] and others was rated as the first modern deep convolution network model, because it first used the technology of modern deep convolution network to improve the accuracy of the model, and it won the championship of Imagenet image classification competition in 2012.

3. In 2014, researchers from Oxford University's visual geometry group and Google deepmind company jointly developed a new deep convolution neural network: vggnet [8], and won the second place in ilsvrc2014 competition classification project.

4. The concept [9] network is composed of multiple concept modules and a small number of aggregation layers. In 2014, as the earliest concept V1: Google concept net, it appeared in the ilsvrc 2014 competition for the first time, and won the first place in the competition at that time.

5. In 2015, Kaiming he [10] and others proposed RESNET to improve the transmission efficiency of information by adding directly connected edges to the nonlinear convolution layer. Because of its "simple and practical" coexistence, many methods are based on resnet50 or resnet101, and are widely used in detection, segmentation, recognition and other fields, and it won the first place in the classification task of Imagenet competition. In Fig2, we can see it completely.

Fig.2 Convolution Network Development

III. APPLICATION OF CNN IN VARIOUS FIELDS

In recent years, convolutional neural networks have been applied in various fields:

1. Medical image is very dependent on equipment imaging, and it has many kinds and poor image quality, which is not conducive to the diagnosis of the patient's condition. The medical image recognition based on convolution neural network is very helpful to determine the location of lesions.Liu Ji [11] and others classified the cervical images and used the convolutional neural network (CNN) model for discrimination, which greatly improved the efficiency; Coudray [12] and others trained the CNN model to judge the pathological condition of lung cell tissue and distinguish whether the detected cells were pathological changes. The prediction result of this model is as high as 97%, which can be applied to common cancer diagnosis.

2. Nowadays, animal husbandry has gradually shifted from traditional stocking to intensive and precise farming. Ai-ming at animal head portraits of animal husbandry in Qinghai Tibet Plateau, lamaojie [13] designed a darket-53 fram-ework based on yolov3 algorithm by using convolutional neural network to identify and count animal heads in Qinghai Tibet Plateau, and its map can reach 87.89%.

3. In terms of agriculture, Li Linfang [14] and others put forward improvements on the current manual and mecha-nical methods for fruit quality detection. Compared with the traditional classification method, it has better recognition speed and accuracy, and the model reliability reaches 94.34%.

4. With the rapid development of electronic intelligent devices in recent years, face recognition has become a hot topic. Carrying face recognition technology on electronic intelligent devices, how to make the recognition rate higher is the research direction. The limitations of traditional recognition, including low

contrast accuracy, make the requirements can not be met. On the basis of convolutional neural network, Fu [15] and others proposed a new CNN architecture guided convolutional neural network (guided CNN), which solved the problem of low matching degree of face images at different resolutions. The deepid model proposed by Tang Xiaoou's team [16] has carried out a detailed and essential analysis on the internal structure of CNN. By using deep learning, the development of face recognition technology has gone to a higher level.

Not only these, in some special environments, in the face of the impact of complex and changeable environmental factors, when manual operation is not convenient, the use of deep learning for image recognition is more secure and convenient. At present, for example, the inspection work of substation in society is almost manual operation, and the degree of trouble and danger is self-evident. In particular, the area of the substation is generally large, which is very hard for manual work, and a demand to replace manual inspection work is generated. In my opinion, the intelligent inspection technology based on deep learning has a very broad application prospect. Its intellectualization is embodied in the construction of the model which can conduct deep learning according to the convolutional neural network (CNN). For the condition monitoring and detection of a large number of instruments, if the real-time films collected by the camera are put into the convolution neural network model for automatic identification of targets, and then analyze whether the working status of each machine and equipment in the substation is normal It is very meaningful for the safety monitoring of substation.

IV. PROSPECT OF CNN DEVELOPMENT IN THE FUTURE

There are three key points that need to be improved in depth learning for image recognition

One is the reliability of the model, which requires us to build a network architecture wit:h excellent performance. This also includes several aspects: 1. Stack cross design of convolution layer, convergence layer and full connection layer; 2. Convolution kernel size, weight selection, loss function application and so on. This point needs to be debugged and compared step by step through experiments, so as to determine the relatively optimal scheme.

The second is the continuous improvement of intelligent algorithm to improve the overall performance of the model. A good algorithm can reduce the computational difficulty, shorten the running time, and improve the accuracy of image recognition. For target detection and recognition, the r-cnn algorithm proposed by girshick [17] and others in 2014 is to improve the CNN convolution neural network like Fig3. On this basis, the candidate box for the region generated from the input image is added, and the region is fused with similar feature regions by selective search method; in 2015, girshick R [18] proposed fast In r-cnn algorithm like Fig4, Rio (region of interest) is proposed, and softmax is introduced for image classification. Compared with r-cnn, the improvement is to map the proposed region to the last feature map of CNN, so that each image only needs to extract features once, and classification is also solved at the end of convolutional neural network; in 2015, Ren s [19] proposed faster R-cnn algorithm mainly solves the problem of large number of candidate regions in fast r-cnn algorithm like Fig5, and many regions are repeatedly extracting the same feature, which is a waste of time. It proposes a concept of RPN (region recommendation network), which uses the method of extracting candidate regions from the convoluted feature map, which greatly reduces the time spent on extracting candidate regions. Therefore, the improvement of the algorithm has always been an important part of the study of deep learning.

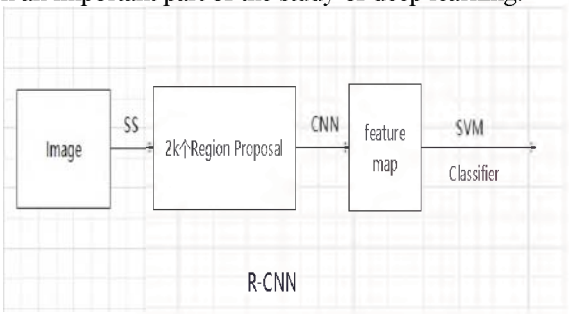

Fig.3 R-CNN

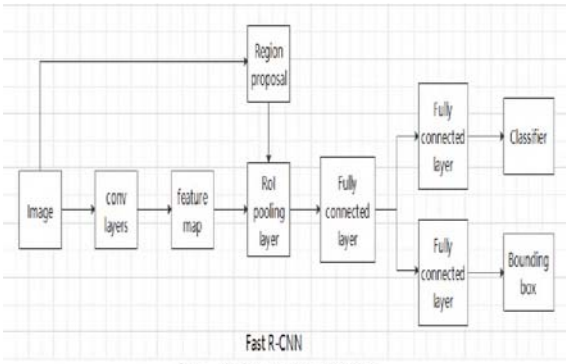

Fig.4 Fast R-CNN

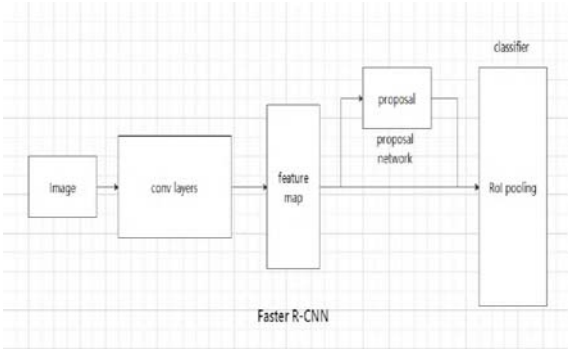

Fig.5 Faster R-CNN

The third is the construction of data sets. Why does the image recognition technology in some fields of our country still stay on the traditional method and can't rely on deep learning for application and development. One important reason is the shortage of data sets. The limited data set is not enough to train a network model well, which will lead to the low precision when the actual detection is needed, so the application efficiency can not meet the requirements, and the construction of deep learning model is not of great significance. So the construction of data set is very important for deep learning. Liu Yanlin [20] aimed at the problem of the shortage of marine fish school images, FishBase and wildfish were used as the basic databases to expand the data in the way of web crawler, and then data cleaning was carried out to construct a large-scale image data set of marine fish schools, which provided great convenience for the research on the classification of marine fish schools.

However, the domestic data sets in various fields are r elatively scarce, and need to be supplemented. In view of this weak point, we know that the current methods of coll ecting data sets are usually manually photographed and st ored, and then post processed to select high-quality image s for construction. The result of this is that it is very time-consuming and energy-consuming, which may not meet t he required target economic effect value. In view of the la ck of image data, we can fully consider the use of video d ata acquisition in the future. The method of collecting pic tures should not only be a small amount of interception, b ecause the target may only be collected. The continuity of video data should be used to fully collect the target featu res in various environments to construct a comprehensive data set as far as possible.

The application potential of image recognition based on convolution neural network is very huge. In the future, continuous improvement and enhancement are carried out for the above problems, and deep learning will play a strong role in promoting the image field.

REFERENCES

[1] Lowe D G.Distinctive image features from scale-invariant keypoints[J].International Journal of Computer Vision,2004, 60:91-110.

[2] Bay H,Ess A,Tuytelaars T,et al.Speeded-up robust features(SURF) [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[3] Xu Yuan, Xu Xiaoliang, Li cainian, et al. Pedestrian detection based on SVM classifier and hog feature extraction [J]. Computer Engineering, 2016,42 (1): 56-60,65.

[4] Hubel D H,Wiesel T N,1959. Receptive fields of single neurones in the cat's striate cortex[J]. The Journal of physiology, 148(3):574-591.

[5] Lecun Y, Boser B, Denker J S, et al., 1989. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1(4):541-551.

[6] Yann Lecun, Leon Bottou, Y. Bengio, Patrick Haffner. Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 1998.

[7] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In NIPS, 2012.

[8] VGG16: K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition.In ICLR, 2015.

[9] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions. In CVPR, 2015.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren. Deep Residual Learning for Image Recognition. In CPVR, 2016.

[11] Liu Ji, sun Rencheng, Qiao Songlin. Application of deep learning in medical image recognition [J]. Journal of Qingdao University (NATURAL SCIENCE EDITION), 2018 (1): 69-74.

[12] Coudray N，Moreira A L，Sakellaropoulos T，et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning[J]. bioRxiv，2017：197574.

[13] La Maojie, an Jiancai rang. Animal image recognition based on convolutional neural network [J]. Software, 2020,41 (08): 43-45.

[14] Li Linfang, Wang Jianjun, Wei Zheng, Feng Xiangrong, an Jinliang. Study on orange classification and recognition based on convolutional neural network [J]. Journal of Henan University of science and Technology (NATURAL SCIENCE EDITION), 2020,48 (03): 68-73.

[15] Fu T，Chiu W，Wang F.Learning guided convolutional neural networks for cross-resolution face recognition[C]// Proceedings of 27th International Workshop on Machine Learning for Signal Processing（MLSP）.Tokyo：IEEE，2017：1-5.

[16] Taigman Y, Yang M, Ranzato M，et al.DeepFace：closing the gap to human-level performance in face verification[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.Columbus：IEEE，2014：1701-1708.

[17] Girshick R,Donahue J,Darrell T,et al.Rich feature hierarchies for accurate object detection and semantic segmentation[C].Computer Vision and Pattern Recognition,2014.

[18] Girshick R. Fast RCNN[C]. 2015 IEEE International Conference on Computer Vison(ICCV). IEEE, 2015.

[19] Gao Lu. Design and Implementation of Switch State Indentification System for Substation Based on Deep Learning [D]. Chinese Academy of Sciences (Shenyang Institute of Computing Technology),2019.

[20] Liu Yanlin. Construction and application of large scale marine fish image database [D].University of Chinese Academy of Sciences (Shenzhen Institute of advanced technology, Chinese Academy of Sciences),2020.(20).