Remote Sensing Image Recognition Method Based on Faster R-CNN

Chao Ma[1]
Shanghai Academy of Agricultural Sciences ,Shanghai, Shanghai, China
machao@saas.sh.cn

Xianyong Yi[4]
King Abdullah University of Science & Technology , Thuwal, Jeddah, KSA
xianyong.yi@kaust.edu.sa

Jinzhao Li [2]
Northwestern Polytechnical University , Xi'an, ShanXi, China
lijinzhao@mail.nwpu.edu.cn

Linyi Li[5*]
Shanghai Academy of Agricultural Sciences ,Shanghai, Shanghai, China
*Corresponding author's e-mail: lly@saas.sh.cn

Zecong Wang [3]
Hainan University , Haikou, Hainan, China
20171619310067@hainu.edu.cn

***Abstract*—This paper proposes a method for remote sensing image recognition based on Faster R-CNN. Using Faster R-CNN model and ZFNet as the basic network, experiments show that the accuracy rate of Architecture, Greenhouses and Paddy field recognition is 90.67%, 93.85%, 83.33%, and the average recognition accuracy reached 89.28%. At the same time, compared with the recognition results of recognition detection methods such as CNN and TT-RICNN, it was found that the proposed Faster R-CNN model has better recognition performance well, with good recognition detection accuracy.**

Keywords- recognition; R-CNN; CNN; TT-RICNN

I. INTRODUCTION

Remote sensing image is an important strategic resource, which is widely used in military and civil fields such as military reconnaissance, land and resources monitoring, traffic monitoring, etc. enhancing the processing ability of remote sensing image and improving the accuracy and resolution of target recognition are the most important issues in remote sensing image research, which are also the focus of many researchers. With the advent of information age, eep convolution neural network has made great achievements in the field of computer vision. Compared with the traditional neural network, deep convolution neural network has broken through the limitation of layers and greatly improved the accuracy. At present, the application of convolution neural network in remote sensing impact target recognition has achieved fruitful results. There is a high degree of recognition in the accuracy of target recognition. However, throughout the current research status, it is mainly focused on single target recognition, and the research on multi-target recognition needs to be further explored. In this regard, a Faster R-CNN recognition algorithm based on deep convolution neural network is proposed to identify multi-target, such as Paddy field, Architecture and Greenhouses, The recognition results show that the recognition performance is good and the average recognition rate is over 89%, which shows that the recognition method has good universality [1-3].

II. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural network(CNN) was originally designed from the concept of "receptive field" put forward by Hubel and Wiesel in the study of cat's visual cortex. After that, Fukushima proposed a neural cognitive machine on this basis, which became the first implementation model of convolutional neural network. Later, Yann Lecun and others proposed the structure of Le Net-5 classic convolutional neural network, Convolution neural network is a milestone in the development of convolution neural network. In essence, it is a deep learning model and a deep learning method based on artificial neural network.[4-5]

A. Convolution neural network structure

The structure of convolutional neural network is generally composed of four parts: convolution layer, pooling layer, full connection layer and output layer. Among them, Le Net-5 and AlexNet are the most classical convolutional neural network structures. With the deepening of research, the deep-seated network structures continue to emerge, including ZF Net and RESNET. The structure of convolutional neural network model is shown in Figure 1 [6-8].

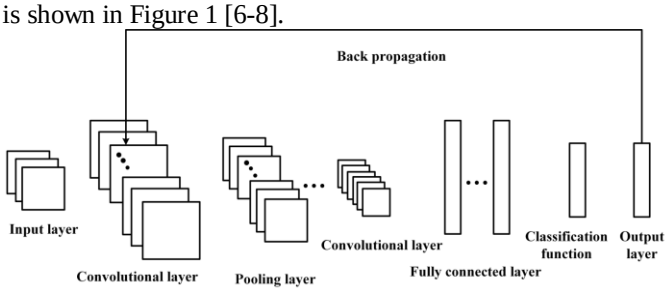

Figure 1. Convolutional neural network structure

The main function of the convolution layer is to extract features from the input image by convolution operation. The formula of the convolution layer is as (1), where X_j^l represents the j-th output feature map of the fiest layer, X_i^{l-1} represents the i-th feature map of the l-1 layer, $f(x)$ is the activation function, M_j represents the set of input features, $k_{j,i}^l$ and b_i^l represent the additive bias of convolution kernel and X_j^l respectively. Convolution kernel is obtained by back propagation algorithm. At the beginning, convolution kernel is initialized randomly, and its number represents the number of features.

$$X_j^l = f(\sum_{j \in M_j}(X_i^{l-1} * k_{j,i}^l) + b_i^l) \quad (1)$$

The pooling layer is located at the back of the convolution layer. Its main function is to reduce the mining of input image features, so as to reduce the computational complexity, improve the receptive field, and ultimately enhance the generalization ability. Therefore, the pooling layer is also called the downsampling layer. The operation of the pooling layer also introduces the activation function. The specific expression formula of the operation is as (2), Among them, β_j^l and b_i^l represent the multiplicative bias and additive bias of X_j^l respectively, and $s(x)$ is the sampling function. The operation of pool layer only reduces the dimension of each channel's feature map, and does not reduce the number of features.

$$X_j^l = f(\beta_j^l s(X_i^{l-1}) + b_i^l) \quad (2)$$

The full connection layer is usually located at the end of the network structure, which can be more than one. The main function of the full connection layer is classification. After the full connection layer, the two-dimensional feature map obtained from the previous processing will be transformed into one-dimensional feature vector. The application of activation function in the convolutional neural network structure is helpful to improve the nonlinear expression ability. In recent years, the most commonly used activation function in convolutional neural network is the ReLu function. Compared with the original sigmoid and tanh functions, the ReLu function has faster convergence speed and faster calculation, and plays a good role in alleviating the gradient dispersion problem. The expression of the ReLu function is shown in formula (3).

$$f(x) = \max(0, x) \quad (3)$$

B. Characteristics of convolutional neural network

Convolution neural network has three characteristics: sparse connection, downsampling and weight sharing. Through the above characteristics, image feature translation, scaling and distortion invariance are realized.

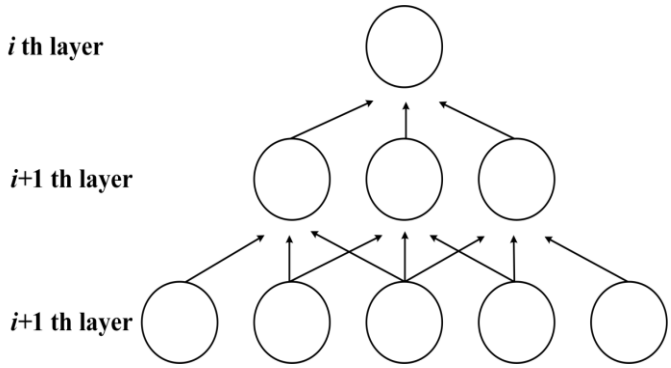

Figure 2. Sparse connection

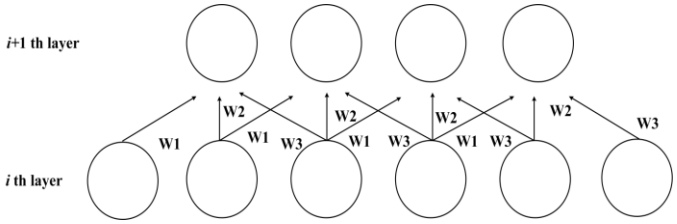

Figure 3. Weight sharing

Sparse connection is mainly manifested in convolution structure, and the form of sparse connection is shown in Figure2, which adopts local connection, which greatly reduces the parameters and calculation amount, The task processing efficiency of the convolution neural network structure model is greatly improved. The weight sharing is that each convolution kernel performs convolution operation with the feature image to generate a new feature map. In fact, all pixels in the image share the same weight value. The weight sharing map is shown in Figure3, which makes the convolution neural network structure have translation invariance through the weight sharing, At the same time, it can greatly reduce the parameters and further improve the calculation efficiency; downs-ampling is reflected in the pooling layer. Through the downsampling processing of the pooling layer, the network structure has the memory function, thus realizing the invariance of distortion and scaling, and also greatly reducing the operation complexity.

III. REMOTE SENSING IMAGE RECOGNITION

This study mainly uses the Faster R-CNN model and simultaneously uses data expansion and other measures to realize the identification and detection of multiple types of remote sensing image targets. The specific identification and detection process and results are described below:

A. Faster R-CNN algorithm and implementation

Faster R-CNN is a commonly used area-based target recognition and detection algorithm, which is mostly used for the recognition and detection of a single category. Faster R-CNN is composed of RPN and Fast R-CNN, the former is a regional generation network, which mainly identifies targets And determining the location, the latter is a target recognition network, which mainly recognizes the size and classifies the

C.

target. Faster R-CNN recognizes and detects the target as shown in Figure 4, which is divided into four steps. First, the image features are extracted and generated. Feature map, secondly use RPN to generate candidate region boxes; then the feature map of the pooling layer and the proposal information are used to obtain the relative position of the target, and the final result is input to the fully connected layer; the final connected layer is used for classification and use The rectangular box indicates the precise target position.

In order to make the RPN and Fast R-CNN network share the convolutional layer, a four-stage algorithm is used to implement the shared feature calculation. In the first step, the RPN network is trained through the ImNetNet pre-trained network model; in the second step, the previous step is used to train The obtained RPN output area is recommended to train the Fast R-CNN network; in the third step, the trained Fast R-CNN network is obtained in the second step to initialize the RPN network. During the training, care must be taken to keep the convolution layer unchanged. Only the parameters unique to the RPN network are adjusted; in the fourth step, the shared convolutional layer remains unchanged, and the region of interest generated in the previous step is used to fine-tune the fully connected layer of the Fast R-CNN network. The RPN and Fast R-CNN networks were trained, and convolution layer parameter sharing was also implemented.

Considering that there are not enough training samples to train the Fast R-CNN model, this research uses ZFNet as the basic network structure. According to the first three convolutional layers of the ZFNet network structure, there are no class-related features, so the Fast R-CNN model is being trained In the first and second stages of the method, the parameters of the first three convolution layers can be directly fixed, and only the fourth convolution layer can be fine-tuned.This operation greatly reduces the data size required for training, making use of Fast R-CNN Model recognition and detection of multiple types of remote sensing image targets become possible.

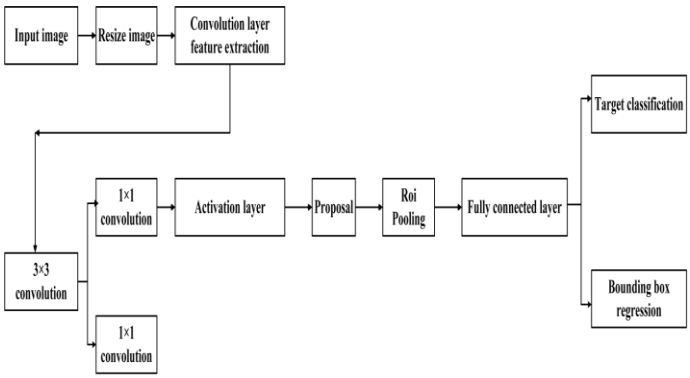

Figure 4. Model framework

B. Experiment

This experiment uses the positive sample set in the NWPU VHR-1 0 data set, and randomly divides the sample set into a training set, a test set, and a validation set with a ratio of 2:2:1. Write Faster R-CNN algorithm on Python3.5 Tensor Flow was used to build the model, and GE-Force GTX1060 graphics card was used. In order to test the proposed model recognition and detection performance, the average accuracy rate (mAP) was used for evaluation. The final detection results are shown in Table 1. The detection effect diagram is shown in Figure 5.

TABLE I. STATISTICS OF REMOTE SENSING IMAGE TARGET DETECTION

Category	Verification sample number	Correct sample number	Accuracy rate (AP)
Aircraft	150	136	90.67%
Greenhouses	62	61	93.85%
Paddy field	168	140	83.33%
Total	383	337	89.28%

RESULTS

++++

Figure 5. Remote sensing image target detection results

C. Analysis of experimental results

According to the detection results in Table 1, the average accuracy of the proposed model for multi class target recognition and detection of remote sensing image is 89.28%, which is significantly improved compared with the recognition accuracy of traditional remote sensing algorithm CNN and T-T-R, as shown in Table 2. It can be seen that fast r-cnn has good performance in multi class target recognition, It is a fast and effective method of remote sensing image recognition.

TABLE II. RECOGNITION ACCURACY OF DIFFERENT MODELS

Category	SSCBow	CNN	T-RICNN
Architecture	50.6%	70.1%	88.4%
Greenhouses	50.8%	56.9%	77.3%
Paddy field	33.4%	84.3%	85.3%
mAP	44.93%	70.43%	83.67%

IV. CONCLUSION

In this study, a Faster R-CNN model based on deep convolution neural network is proposed, which takes zfnet as the basic network structure and reduces the operational parameters by fine tuning the parameters of high-level convolution layer. The experimental results show that the recognition accuracy of Architecture, Greenhouses, Paddy field and other remote sensing image targets is more than 89%, and the recognition performance is good .

ACKNOWLEDGMENT

This work was supported by Hu Nong Qing Zi (2018) No. 1-30 Development of Rice Disease Diagnosis System Based on Image Recognition.

REFERENCES

[1] C. Samson, L. Blanc-Féraud, G. Aubert, et al. “A Level Set Model for Image Classification”. International Journal of Computer Vision, vol. 40, pp. 187–197, December 2000.

[2] N.K. Alham, M. Li, Y. Liu,,et al. “A MapReduce-based distributed SVM ensemble for scalable image classification and annotation”. Computers & Mathematics with Applications, vol. 66, pp. 1920–1934, December 2013.

[3] Y. Liu, J. Guo, J. Lee. “Halftone Image Classification Using LMS Algorithm and Naive Bayes”. IEEE Trans Image Process, vol. 20, pp. 2837–2847, October 2013.

[4] P. Moeskops, M.A. Viergever, A.M. Mendrik, et al. “Automatic Segmentation of MR Brain Images With a Convolutional Neural Network”. IEEE Transactions on Medical Imaging, vol. 35, pp. 1252-1261, May 2016.

[5] L. Zhang, J. Liu, B. Zhang, et al. “Deep Cascade Model-Based Face Recognition: When Deep-Layered Learning Meets Small Data”. IEEE Transactions on Image Processing, vol. 29, pp. 1016-1029, May 2016.

[6] J.W. Lu, V.E. Liong, G. Wang, et al. “Joint Feature Learning for Face Recognition”. IEEE Transactions on Information Forensics and Security, vol. 10, pp. 1317–1383,July 2015.

[7] S.H. Gao, Y.T Zhang, K. Jia, et al. “Single Sample Face Recognition via Learning Deep Supervised Autoencoders”. IEEE Transactions on Information Forensics and Security, vol. 10, pp. 2108–2118, October 2015.

[8] J.W. Lu, G. Wang, J. Zhou. “Simultaneous Feature and Dictionary Learning for Image Set Based Face Recognition”. IEEE Transactions on Image Processing. vol. 26, pp. 4042–4054, August 2017.