Research on Application of Convolutional Neural Network in Remote Sensing Image Recognition

Chenlong Bao, Ying Chen

College of International Studies, National University of Defense Technology, Nanjing, Jiangsu, 210039

Keywords: remote sensing image; convolutional neural network; recognition; application

Abstract: With the advancement of science and technology in recent years, the amount of remote sensing image data has become more and more abundant and has important research value. As a current research hotspot, remote sensing image recognition has attracted more and more people to study it. Based on this, the application of convolutional neural network in remote sensing image recognition is studied in this article. First, the basic principles of convolutional neural network are introduced, including image preprocessing, image local connection, zero-padded, multi-convolution kernel application, pooling processing and pre-processing. It then analyzes the current research status of remote sensing image recognition, briefly analyzes the problems of deep learning in remote sensing image processing, and predicts its future applications. Finally, it elaborates the application of convolutional neural network in remote sensing image recognition. Only by intensifying research can we better utilize the application value of convolutional neural networks in remote sensing image recognition, and then promote the development of remote sensing image processing technology.

1. Introduction

Image recognition technology is an important field of artificial intelligence. It has important research and application value in many fields such as navigation, map and terrain registration, natural resource analysis, weather forecast, environmental monitoring, physiological lesion research, etc. [1]. With the increasing of data throughput and computing power, researchers continue to deepen their research on human visual models. Relying on efficient algorithm design and training with large amounts of data, it is possible to combine machine learning algorithms with images. After the introduction of deep learning algorithms, it brought new vitality to image recognition [2].

2. Convolutional neural network principle

A complete image recognition system mainly includes: image sensors, image acquisition and digitization processing, image pre-processing, image feature extraction, and image recognition algorithms. Image recognition algorithms are embodied in other parts besides acquisition and digitization, which are the most important part of the image recognition system. In the process of image recognition, the extracted image features are different according to different needs, and for the same image, different people may recognize different results, such as the same picture that often appears on the Internet, people of different ages or in different states see completely different content. That is to say, there may be deviations in the recognition process for different people, so try to let the machine understand is easy to be biased, that is, the universality and generalization ability of traditional artificially designed recognition algorithms are weak. This is mainly because traditional image recognition methods usually use linear mapping to extract features during the feature extraction process. The linear features of the picture, but the distribution structure of most of the local images is very complicated, using only linear mapping is difficult to extract all the features of the image, so there is no way to accurately identify it. Specifically, the traditional image recognition method solves a small range of problems and weak generalization ability. When the task changes, you need to find a new solution. This cannot meet the needs for high efficiency and universality in artificial intelligence. Therefore, finding a universal and efficient image recognition

Published by CSP © 2020 the Authors

method has become the focus of academic research. The convolutional neural network proposed by Yann Le Cun in 1998 is the source of many excellent image classification methods [3]. The basic principle of convolutional neural networks is as follows.

(1) Image preprocessing

For a black and white image, the computer stores the pixel values in the form of an array. The value ranges from 0 to 255. 0 means the darkest is black, 255 means the brightest is white. Through the above processing, an image can be processed into a matrix of size windth × heigth. For a color picture, each pixel is composed of the three primary colors of red (R), green (G), and blue (B). The color image is processed as a three-dimensional tensor of windth × heigth × depth, where depth is an array containing three RGB color values.

(2) Local connection of images

In the traditional image recognition process using artificial neural networks, each pixel of the picture is usually used as the input of the feedforward neural network, and the image label is used as the output for training. The classifier is formed by continuously adjusting the weights between the networks. In order to complete the recognition, however, using this method often requires a lot of computing resources, and when the sample size is too large or the picture is too large, it is often difficult to obtain the optimal result due to the limitation of computing resources. The network uses local connection to solve the above problems well. Convolutional neural networks introduce the concept of a convolution kernel. Convolution kernels are generally a 3 × 3 or 5 × 5 matrix. The convolution kernel is based on a certain step size which is convolved with pixels of the same size in the original image. By scanning the full image, the obtained result is used as the hidden layer of the network. In this way, the information of the plane structure of the picture can be retained. After applying the convolution kernel in each dimension, it also serves as the hidden layer of the network.

(3) Zero-padding

In the process of using the convolution kernel to process the original image, the original image will be reduced after being processed by the convolution kernel, which may cause the problem of edge information being discarded. Therefore, zeros can be used to complete the image around, so that the size of the original image does not change after convolution.

(4) Multi-convolution kernel application

When performing image processing, multiple convolution kernels can be applied to process the original image at the same time. Likewise, the output is used as a node in the hidden layer.

(5) Pooling

After the convolution, a pooling operation can be added. Commonly used are average pooling and maximum pooling. After the pooling process, the redundant information after the convolution operation can be effectively removed, thereby reducing the impact of invalid information to improve recognition accuracy.

(6) Feedforward neural network

After using the convolution kernel to process the original graph information, the completed hidden layer can be constructed. In this process, the hidden layer can be convolved to form multiple hidden layers. Finally, all hidden layer nodes are used as feedforward neural networks. Input, training with labels as output, and finally forming a multi-layer neural network, which is a convolutional neural network, which can achieve accurate recognition of images.

3. Research status of remote sensing image recognition

CNN has unique advantages in processing high-dimensional image data. According to the rich spatial and spectral information contained in hyperspectral images, some literatures have proposed methods to transform spectral information into images. One is to convert to grayscale images using CNN to extract the texture features for classification; the other is to convert to waveform graphs and use CNN to train the fluctuation features for classification. The experiments show that the Gaussian kernel SVM method with PCA dimension reduction is superior when there are many types of samples, and it is also partially superior when the number of samples is small. At the stage, previous researchers carried out research on different hyperspectral image features as input information for

CNN. Based on the CNN classification algorithm developed in the spectral domain, by constructing a 5-layer network structure, and then analyzing the spectral information of the pixels one by one, the full spectrum is input at the input end segment set, calculate the cost function value through neural network, realize the extraction and classification of spectral features, and the test accuracy rate is 90.16%. In the study of spatial neighborhood information as input, the spatial neighborhood information of each pixel is used as CNN input of the framework, at the same time, in order to alleviate gradient dispersion, improve network execution efficiency and classification accuracy, the activation function ReLU was designed, and the table was studied. The mini-batch stochastic gradient descent method can improve the execution efficiency of the CNN framework, and the test accuracy reaches 97.57% [4]. CNN is also widely used in the detection of target detectors and the extraction of buildings. Use DCNN to build a water body recognition model, first use maximum stable extreme region algorithm segmented the high-resolution remote sensing image of the drone, input the target sub-area to be identified, and introduced the DCNN water body recognition model to identify the water body. Experiments have shown that the recognition accuracy is up to 95.36%. The detection research has a large difference. Using the CNN algorithm to identify SAR image targets under different activation function applications, the test accuracy is more than 95%, and at the same time, it is concluded that the ReLu function is the most suitable activation function. In the extraction recognition and classification, CNN training and testing were performed on images of rural buildings and non-buildings under the Caffe Net learning framework, and the building recognition rate reached 95.00% [5]. Of course, there are also some CNN models in the research. The classification test and accuracy evaluation were performed, but due to the network properties and parameter selection of CNN, the line feature extraction was blurred, the classification boundary was rough, and the classification effect was affected. From the specific application analysis above, different activation functions and different convolution kernels will affect the test accuracy, the choice of network properties and parameters will also have some impact.

Deep learning has developed rapidly in remote sensing image processing, but there are also many problems, summarized as follows:

(1) DBN can better extract image features when taking advantage of the characteristics of unsupervised learning, and based on the complex and limited remote sensing image data, the structure parameters of the DBN model gradually explored are of great benefit to future research, but the selection of network parameters requires the intervention of artificial and prior knowledge, and it is difficult to determine appropriate test parameters [6].

(2) CNN has obvious advantages when processing high-dimensional images like hyperspectral data, but the limited samples of remote sensing images limit the generalization ability of the CNN algorithm. In addition, the choice of different activation functions, convolution kernels, and network parameters will affect the test accuracy, and proper selection will greatly enhance CNN execution efficiency.

(3) SAE uses the characteristics of unsupervised classification in practical applications. At the same time, in terms of dimensionality reduction and feature extraction, it generates fewer reconstruction errors than the PCA method, but it still needs to be combined with other classifiers to obtain high accuracy of classification and recognition, and the need for parameter optimization and manual labeling of samples, etc., these make SAE not completely unsupervised learning.

Combining the above problems, the predictive analysis of the application of future deep learning in remote sensing image classification and recognition is as follows:

(1) Further improve and tune the deep learning network parameters and functions. Different network parameters and function choices will have different effects on the accuracy of the research project. A reasonable choice will greatly help the experiment.

(2) Research on deeper model levels and complex model structures. Deep-level models will bring more accurate results, and complex structures will reduce the degree of overfitting, thereby improving model learning capabilities and also benefiting large-scale training data [7].

(3) In-depth discussion of sample expansion. Existing remote sensing image data does not meet the needs of deep learning training. Researchers need to combine image data with translation,

rotation, and zoom to generate more effective data to improve model accuracy.

(4) Build a deep learning model structure combined with multiple algorithms. At present, a single deep learning model is difficult to achieve the accuracy required for remote sensing image processing, so researchers often use improved models to improve the accuracy, such as the combination of SAE and SVM algorithms.

4. Application of convolutional neural network in remote sensing image recognition

Convolutional neural networks modify and improve the functions and structures of the layers of traditional neural networks. Through local connection, weight sharing, and space sharing, the original information of the image can be fully retained during the feature extraction process. The network has a supervised learning method to achieve accurate recognition of images while greatly improving the training efficiency of the network, which can realize a large number of image recognition tasks [8]. At present, convolutional neural networks are widely used in the field of face recognition specific implementation process, the feature extraction method based on the convolution kernel is used to extract features from small samples under changing conditions within complex classes. At the same time, a sparse matrix is introduced to carry out network structure and training complexity [9]. Related research results show that convolutional neural networks can efficiently and accurately deal with face recognition problems.

In addition, related studies have also applied convolutional neural networks to the prediction of stock prices. The historical stock price data is analogized to the pixel features of the image. The convolution kernel is used to process the historical stock price data to extract the characteristics of stock changes, so as to realize the future identification of stock prices, that is, prediction of future prices [10].

4.1 Network Structure

In a specific implementation, first input an image, and after a series of convolution and pooling operations, extract the features of the image to generate a feature map; then use the candidate region to generate a network and process on the feature map to generate different scales and aspect candidate target area; finally, a classification regression network is used to output the category of the generated target candidate area based on the features in the candidate area.

4.1.1 Feature extraction network

The feature extraction network in the CNN algorithm is a convolutional neural network. The network structure can select a suitable network as the feature extraction network of the CNN algorithm. Common feature extraction networks are ZF network, VGG16 network, and Alex Net network. The amount of training data often determines the performance of the trained network model in deep learning. The larger the amount of training data, the better the performance of the trained network model, and the smaller the amount of training data, even if a better network structure is used, the training performance of the network model will also be poor.

Therefore, in deep learning, the network trained on a large data set is often used in its own network structure, and its own data set is used to fine-tune the network to improve the performance of the network. The CNN algorithm uses the large data set image pre-trained model as a feature extraction network.

4.1.2 Candidate area generation network

The candidate region generation network uses the RPN network, which is a fully convolutional network. The input of the RPN network is a feature map output by the feature extraction network, and it outputs rectangular candidate regions with multiple scales and aspect ratios. The network first uses a 3 × 3 sliding window slides on the feature map, then maps each position that the sliding window passes into a 256-dimensional feature vector, and finally inputs these 256 feature vectors into two fully connected layers, one full output of the connected layer is 2 × 9 = 18 scores, and the output of the other fully connected layer is 4 × 9 = 36 correction parameters. In target detection and recognition, because the shape and size of the target to be detected are different, the CNN algorithm is nine reference rectangular frames which are designed at each sliding window position to match various targets. These nine reference rectangular frames correspond to areas of 1282, 2562, and 5122, and aspect ratio lengths of 1: 1, 1: 2, and 2: 1. Each rectangular reference frame corresponds to 4 correction parameters, which represent the position of the rectangular reference frame. You can modify the correction parameters to get 9 candidate areas for each sliding window position. There are two points,represents whether the candidate area contains the target to be tested.

The network first performs a convolution operation on the feature map through a 3 × 3 convolution kernel to form a feature vector; then uses two 1 × 1 convolution kernels of two sizes to fully connect the layers to obtain the scores of the candidate regions and correct the parameters; finally, normalize the scores through the softmax layer to obtain the confidence level of whether the candidate region contains the target to be measured. The loss function of the candidate region generation network is a multi-tasking loss function. The training tasks for category confidence and correction parameters are unified.

4.1.3 Classification regression network

After the candidate regions are extracted, the classification and regression operations need to be performed on the candidate regions. The input of the classification regression network is the feature map and candidate regions output by the feature extraction network, such as the candidate regions extracted from the network output. The output is the confidence that the candidate region corresponds to each category degree and candidate region correction parameters. Due to the different sizes and shapes of candidate regions and the number of features included, the classification regression network first passes a Roi Pooling Layer, turning the features in the candidate area into feature maps of the same size and shape, and then going through two fully connected layers. The initialization of these two fully connected layers uses the full connection of the pre-trained model in the large data set Image Net. The parameters corresponding to layers fc6 and fc7. Finally, the category score and correction parameters of the candidate area are output through fc/cls and fc/bbox_reg, and the score corresponding to the candidate area category is normalized through a softmax layer, and the confidence corresponding to the candidate area category is output degree.

In the CNN algorithm, the Roi Pooling layer uses only one scaled pooling layer, which converts feature maps contained in candidate regions of different sizes into feature vectors of the same dimension. During the training process of convolutional neural networks, its performance is often depending on the number of training sets, the smaller the training set, the more prone to overfitting. The dropout method is an effective method for preventing overfitting in CNN algorithms in classification and regression networks. This method refers to the neural network training, some hidden layer nodes are randomly selected to stop their work. The hidden layer nodes will not update the parameters during this training, the parameter values will remain unchanged, and the normal operation of the network will not be affected.

4.2 Activation function selection

According to the structure of the neural network, it can be seen that the output of the neuron structure is the weighted sum of all input data, which will cause the entire neural network to be a linear model that can only be used for linearly separable problems. In order for the neural network to solve linearly inseparable problems, the output of each neuron passes a non-linear function, so the entire neural network model is non-linear, and this non-linear function is the activation function. Since the neural network model is used in the training process, the gradient descent method adjusts the parameters of the neural network, so the non-linear activation function used needs to be differentiable everywhere in the function definition domain. The commonly used activation functions include ReLU function, sigmoid function, and tanh function. Models for different activation functions design the experiment, because it only considers the effect of different activation functions on the loss value, so the number of iterations run is set to a fixed value of 10,000 times, and the learning rate is set to 1e-3. Although the activation function is a sigmoid function and a tanh function, the model value of the loss function continues to decline and stabilize. When the activation function is the ReLU function, the model's loss of function values initially decline faster loss function value within a period of time intermediate shock in the model, but the model can be eventually converged and stabilized, and converges in the case of loss of function values than the loss function of the sigmoid function and tanh function models is small, so the activation function used is the Re LU function.

4.3 Model Parameter Tuning

4.3.1 Weight parameters

In CNN algorithm, the loss function used in training the candidate region generation network is a multi-tasking loss function, including two sub-items of classification loss function and regression loss function. The relative importance of these two sub-items depends on the weight parameter to determine. In the default state, when the weight parameter is 1, it means that the two sub-items are equally important. When the weight parameter is too large, it means that the regression loss function has a large contribution to the loss function; when the weight parameter is too small, indicates that the sub-item of the classification loss function has a large contribution to the loss function. After analysis of experimental data, when the weight parameter is 0.1, the obtained mAP value is the largest, which is 0.863. Therefore, the the parameter weights adopted by the CNN-based remote sensing image recognition model is 0.1.

4.3.2 Training times

The performance of the convolutional neural network is not only related to the learning rate, but also affected by the number of times the model is trained. If the number of times the model is trained is too small, the model may not reach convergence and stop training. If the number of times of training is too large, then the subsequent training process after the model converges is a waste of computing resources. Because the alternating training method is used, the number of training times for the candidate region generation network and the number of training for the classification regression network need to be determined. In the experiment, the maximum training times are set to 15000 when the number of training times reaches 10,000, the loss function of the candidate region generation network converges, and when the number of training times reaches 10,000, the loss function of the classification regression network converges. Therefore, the training number of the candidate region generation network is set to 10,000, and the number of trainings for the classification regression network is set to 10,000.

The experimental verification shows that the average accuracy of the CNN algorithm can meet the needs of remote sensing image recognition.

5. Conclusion

At present, deep learning-based remote sensing image recognition is still in its infancy and has a lot of room for development. In the past, due to resource constraints, in the face of a small number of remote sensing images, we often required the data on the opponent to do a lot of special processing, and then train a simple classifier model to achieve our target recognition work. It includes special processing such as image preprocessing, complex feature extraction, feature transformation, etc., to obtain an ideal shallow level representation of remote sensing images. Today, with the rapid development of remote sensing technology, we can easily obtain a large number of high-resolution optical remote sensing images. With the development of machine learning theory, image expressions have evolved from shallow features to high-level semantic features, and people's understanding of images has become more and more deep. High-resolution optical remote sensing image target recognition is an important part of image recognition. The application of convolutional neural networks in the field of image classification has achieved great results. Many scholars at home and abroad have also borrowed the principle of convolutional neural networks in the research of remote sensing image recognition. Some scientific research results have been obtained, but there are still places that need to be improved and perfected in practical applications. How to better apply deep neural networks in the algorithm model of remote sensing image classification has become an urgent problem in the research of neural networks.

References

[1] Zhang Hongqun, Liu Xueying, Yang Sen, Li Yu. Supervised remote sensing image retrieval based on deep learning[J]. Journal of Remote Sensing. 2017 (03): 115-118.

[2] Luo Jianhao, Wu Jianxin. Review of fine-grained image classification based on deep convolution features [J]. Journal of Automation. 2017 (08): 98-101.

[3] Xue Dixiu. Research on medical image cancer recognition based on convolutional neural network [D]. University of Science and Technology of China. 2017.

[4] Li Yandong, Hao Zongbo, Lei Hang. Review of convolutional neural networks[J]. Computer Applications. 2016 (09): 201-203.

[5] Xu Yihui, Mu Xiaodong, Zhao Peng, Ma Yi. Scene classification of remote sensing images using multi-scale features and deep networks[J]. Journal of Surveying and Mapping. 2016 (07): 56-58.

[6] Liu Yang, Fu Zhengye, Zheng Fengbin. Scene classification of high-resolution remote sensing image based on neural cognitive computing model[J]. Systems Engineering and Electronics. 2015 (11): 236-238.

[7] Yangzhou, Mu Xiaodong, Wang Shuyang, Ma Chenhui. Scene classification of remote sensing image based on multi-scale feature fusion[J]. Optics and Precision Engineering. 2018 (12): 189-191.

[8] Zha Zhihua, Deng Hongtao, Tian Min. Research on image segmentation algorithm based on convolutional neural network[J]. Wireless Internet Technology. 2019 (13): 34-36.

[9] Wang Zhen, Gao Maoting. Design and implementation of image recognition algorithm based on convolutional neural network[J]. Modern Computer (Professional Edition). 2015 (20): 273-275.

[10] Jason, Wu Kuilin, Zhu Jiasong, Li Qingquan. Research on superpixel-level Gabor feature fusion method for hyperspectral image classification[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition). 2018 (01): 102-104.