Research on water extraction technology of remote sensing image based on neural network

Qiang Zhang* 1, Xiao-nan Wang 1, Ya-nan You 2 and Han Li 1

1 Beijing Institute of Spacecraft System Engineering, postbox5142, Beijing, China
2 Beijing University of Posts and Telecommunications, No.10 Xitucheng Road, Haidian District, Beijing, China

bigtree000@163.com

Abstract. With the rapid development of space-based facilities and the increasing number of remote sensing satellites, more and more effective observation data can be obtained on the ground. The traditional manual translation method will not meet the requirements of remote sensing application. Aiming at the problem of water body extraction in remote sensing application, an automatic water body recognition method based on U-Net deep neural network is proposed. The results of water body NWDI index evaluation are considered comprehensively, and the interference elimination model is used to process the results of water body rough recognition, so as to further improve the extraction accuracy. Using the image data of a remote sensing satellite to verify the water body recognition method, the results show that the proposed method can automatically extract the water distribution area, and the extraction accuracy is better than 85%.

1. Introduction

Compared with the conventional image, the remote sensing image has the characteristics of wide coverage, rich targets and high spatial-temporal correlation. With the improvement of the spatial resolution of the remote sensing satellite, the target information in the image is more and more abundant, which makes the automatic translation and semantic analysis of the image more and more difficult. At the same time, with the rapid development of space-based facilities and the increasing number of remote sensing satellites, a large number of observation data will be obtained every day. If the image is interpreted and semantic analyzed manually, the cost is high and the efficiency is low, which seriously affects the operational application of remote sensing satellites. These developments make the automatic processing and accurate interpretation of remote sensing image become the current research hotspot.

In 2006, the upsurge of deep learning research in academic and application fields was started by Geoffrey Hinton[1]. Deep learning was found on the basis of early artificial neural network. By using the structure of multi-layer perceptron with multiple hidden layers, high-level feature attributes can be extracted, and rich internal information of data can be obtained. Due to the unique advantages of deep learning, especially the deep convolution neural network can automatically extract and analyze the target features, which has achieved remarkable results in the research. The related achievements are widely used in image processing, handwriting recognition, speech recognition, machine translation, content recommendation, automatic driving and other fields. In the research process, scholars have
proposed Google-Net, ResNet, VGG-Net, Alex-Net, R-CNN, Fast R-CNN, Faster R-CNN, FCN, U-Net and other deep learning networks\cite{2}\cite{3}.

Semantic segmentation is a very important basic part of remote sensing image interpretation. It is a pixel level image classification work to label each pixel in the image with a specific category. It is widely used in land resource survey, crop planting survey, water area distribution inspection, urban construction monitoring and other fields. Water area extraction is one of the applications of semantic segmentation, and it is also a very extensive basic processing work in remote sensing application. Aiming at the automation of water area extraction of remote sensing image, this paper proposes a processing method based on neural network. Using U-Net network as the basic network, the recognition model and interference elimination model are constructed, which can obtain high information extraction accuracy while processing automatically.

2. Convolutional neural network

Convolutional neural network is a kind of feedforward neural network. Its artificial neurons can respond to the sensory units in some converge units, and have excellent performance for wide image processing. Convolutional neural network includes input layer, convolutional layer, pooling layer, activation layer, fully connected layer and other important network units. Among them, the data input layer mainly processes the original image data such as de average, normalization and PCA/whitening. The convolution layer is composed of several convolution units whose parameters are optimized by back propagation algorithm. The purpose of convolution is to extract different features of input. The pool layer will continuously reduce the spatial size of the data, so the number of parameters and the amount of calculation will decrease, which also controls the over fitting to some extent\cite{4}. Generally speaking, the convolution layer of CNN is periodically inserted into the pooling layer. By introducing the activation layer into CNN and using the activation function to generate nonlinear factors, the nonlinear model can be better simulated. Each neuron in the whole connection layer will connect with all neurons in the previous layers, and the whole connection layer will integrate the local information with category differentiation in the convolution layer or pooling layer.

The input processing of convolution neural network is divided into three stages: convolution, nonlinear transformation (or activation) and down sampling (or pooling), as shown in Figure1.

![Figure 1. The process of convolution neural network.](image)

In the convolution neural network, the feature extraction mechanism mainly depends on the convolution kernel, which can be defined as:

\[
\begin{align*}
\hat{a}_{i,j} = f(\sum_{n} \sum_{m} w_{m,n}x_{i+m,j+n} + w_b)
\end{align*}
\]

(1)

Where, \((i,j)\) is a sample point in the characteristic map input by the current layer, \(w_{m,n}\) is the weight coefficient in the convolution kernel, \(w_b\) is the offset coefficient, and \(f(\cdot)\) is the activation function.

The nonlinear transformation in convolutional neural network refers to the nonlinear mapping function, which also becomes the activation function. In the stage of nonlinear transformation, convolution neural network maps the features extracted by convolution kernel nonlinearly. The nonlinear mapping functions commonly used in convolutional neural networks are a-equivalent
saturated nonlinear functions. In recent years, with the continuous improvement of convolution neural network technology, unsaturated nonlinear functions are often used in convolution neural network, such as ReLU (rectified linear units). Compared with the traditional saturated nonlinear function, ReLU has faster convergence speed, which is very positive for the improvement of neural network training efficiency[4].

3. Semantic segmentation based on convolutional neural network

Image semantic segmentation refers to the process of pixel level classification based on the semantic information of pixels in the image. Take the image as the input, output the corresponding category of each pixel. In the image semantic segmentation method based on deep learning, FCN, U-Net and DeepLab are widely used, among which the deconvolution mechanism in FCN is the key link to achieve semantic segmentation.

FCN (Fully Convolutional Networks) is the most basic network structure for image semantic segmentation, as shown in figure 2[5]. FCN focus image level classification to pixel level. For image level classification, convolutional neural network usually uses full connection layer to obtain the feature parameters for classification after convolutional layer, and FCN replaces the full connection layer with convolution layer. Therefore, the output of the network is no longer the category information but the “distribution map” related to the image semantic information. At the same time, in order to solve the impact of pooling on the feature matrix, FCN proposes to use the up sampling method to recover the feature matrix which is the same size as the input image.

Figure 2. Network structure diagram of FCN.

The accuracy of semantic segmentation results based on FCN is greatly affected by the sampling rate. When the degree of pooled down sampling in FCN is large, in the semantic segmentation results of the same large up sampling output, it is possible to accurately identify the boundary of the object, that is, the pooled down sampling destroys the boundary information of the object, while the up sampling process cannot recover the lost boundary information.

The pooling (down sampling) operation in FCN will result in the loss of object details, which restricts the further improvement of image semantic segmentation accuracy. In order to solve the problem of object detail information loss caused by over down sampling of shallow feature map, this paper uses the deep neural network architecture a combined with high and low feature map to carry out remote sensing semantic segmentation, and the processing flow is shown in Figure3. U-Net has a shrinking path to capture semantics and an expanding path to accurately locate, which constitutes a U-shaped network structure combing high-level and low-level feature maps. U-Net can drive end-to-end training with fewer samples, and has higher image processing speed[6].
4. Water Area Extraction Method

In the water area extraction, it is often affected by buildings, their shadows and hills, resulting in extraction information errors and affecting the extraction accuracy. Therefore, in the water area extraction, three models are constructed based on U-NET network. They are water body recognition model, building and shadow recognition model, hill recognition model and so on. Among them, building and shadow recognition model and hill recognition model are used for interference elimination. Three kinds of models are extracted and eliminated step by step to improve the recognition accuracy.

In the process of processing, firstly, a few typical images are randomly selected to mark, forming three kinds data sets for model training. The first is the RGB channel of fusion image, marking water area and background area. The second is the RGB channel of the fusion image, which can label the building and its shadow, and the background area. The third is the RGB channel of the fusion image, which can mark the hilly area and the background area.

After the data set is obtained, the data is expanded, rotated, mirrored and flipped, the contrast and brightness of the image are changed, and the number of samples is expanded.

According to the three data sets, adjust the model parameters to complete the training, and get three models, which are used to segment water, buildings and shadows, and hilly areas.

In the process of image water extraction, for each pixel, the NWDI assisted neural network water recognition model is used to get the preliminary results of water extraction. The results of water segmentation are input into building and shadow segmentation model, and the influence of building, shadow and hill is eliminated to further improve the accuracy of segmentation. The whole processing flow is shown in Figure 4.
5. Extraction Results
Using 10 remote sensing satellite images as input data, the water area in Shantou area is extracted, and the extracted depth neural network and water area extraction method are verified.

The final treatment results are shown in Figure 5, where blue represents the water area. Table 1 lists the analysis and processing results of 10 satellite images. The final extraction accuracy is measured by map index, and the recognition accuracy is 85.8%.
Table 1. Statistical table of Shantou regional water area extraction results based on a remote sensing satellite image.

| Remote sensing image | Water pixel number | Total number of pixels | Water body ratio(%) |
|----------------------|--------------------|------------------------|---------------------|
| image 1              | 339,100,668        | 826,838,475            | 41.01               |
| image 2              | 224,775,529        | 825,482,100            | 27.23               |
| image 3              | 118,239,804        | 826,642,318            | 14.30               |
| image 4              | 60,128,699         | 815,381,238            | 7.37                |
| image 5              | 14,060,273         | 754,356,864            | 1.86                |
| image 6              | 12,023,252         | 754,844,440            | 1.59                |
| image 7              | 46,971,102         | 754,137,120            | 6.23                |
| image 8              | 412,073,780        | 799,172,855            | 51.56               |
| image 9              | 229,746,809        | 798,975,780            | 28.76               |
| image 10             | 78,577,924         | 798,607,378            | 9.84                |

6. Conclusion

In this paper, aiming at the problem of semantic segmentation in remote sensing data processing, an automatic water area extraction method based on deep neural network is proposed. Using U-NET as the basic framework, a water body recognition model is constructed, and the interference extraction models such as building and shadow extraction and hill extraction model, which have great influence on water body recognition, are integrated with water body recognition model and water body index on this basis, the interference extraction model is used to eliminate the interference, and further improve the accuracy of information extraction. The result of segmentation is evaluated by using map index, and the accuracy is better than 85%. The proposed method can be used for automatic extraction of water area from remote sensing satellite image data.

References

[1] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2015, 313(5786):504-507.
[2] Krizhevsky A, Sutskever I, Hinton G E, et al. ImageNet Classification with Deep Convolutional Neural Networks[C]. Neural Information Processing Systems, 2012: 1097-1105.
[3] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]. International Conference on Learning Representations, 2015, 56-58.
[4] Chen X Y, Xiang S M, Liu C L, et al. Vehicle detection in satellite images by hybrid deep convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(10): 1797—1801.
[5] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, 2015:3431-3440
[6] SU Jianmin, YANG Lanxin, JING Weipeng. U-Net based semantic segmentation method for high resolution remote sensing image [J]. Computer Engineering and Applications, 2019, 55(7): 207-213. (in Chinese)