Journal of Image Processing Theory and Applications (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/jipta.2023.060111
ISSN 2560-6239 Vol. 6 Num. 1

Remote Sensing Image Segmentation and Extraction Based on U-Net Convolutional Neural Network Model

Chen Xiong[1], Jiaqi Huang[2]

[1]Xi'an Jiaotong University City College, Xi'an, 710000, China
[2]Wuchang University of Technology, Wuhan, 430223, China

Keywords: U-Net; Data Enhancement; Remote sensing images; Image Segmentation

Abstract: Remote sensing images are essential for quickly acquiring large-scale ground information. Segmentation and extraction of high-resolution remote sensing images are widely used in many fields, such as agricultural monitoring, urban and rural planning, and map production and updating. In this paper, a U-Net convolutional neural network model is built on the Tensor Flow framework. A data enhancement strategy is specially designed for the training task of remote sensing image parcel segmentation to enhance the model's generalization ability. The experimental results choose accuracy as the evaluation index, and the final model accuracy can reach 0.9440. The remote sensing image parcel segmentation method proposed in this paper has high training efficiency and is suitable for high-accuracy remote sensing image segmentation and extraction.

1. Introduction

Analyzing remote sensing images has essential applications in many aspects, such as military, agriculture, urban planning, etc. Traditional techniques such as convolutional neural networks are one of the main methods to realize the processing of remote sensing images. New challenges are presented in the research topic of how to classify remote sensing images with high accuracy further, how to design a reasonable feature system, and choose a suitable classification model. Accurate and fast information on land parcels will provide the necessary support to national decision-making departments and is significant in reducing human and financial resource consumption.

There has been quite a lot of research in this area by domestic and foreign scholars. Yin Hao [1], for edge prediction and semantic segmentation, collaboratively modeled the construction of two types of two-branch convolutional neural networks to effectively solve the problem of high noise of edge ambiguity results and obtain high precision experimental results. Xiuli Fu et al. [2] proposed a convolutional neural network (CNN) model using SortMax classifier to effectively solve the problems of shallow structure classification models such as support vector machine (SVM) with difficult feature extraction and unsatisfactory classification accuracy in remote sensing images, and finally, the model classification accuracy reached 94.57% in the selected images. Danxin Zhao and Shengli Sun [3] proposed a new method of target detection by the residual network (ResNet) for the problem of low accuracy of aircraft target detection due to target orientation and size, shooting angle, and scene diversity, and the final result illustrates that the model has strong robustness to disturbances such as complex backgrounds, with an accuracy of 89.5%. Xiaoyuan Qu and Eternal

Zhang [4] used active learning combined with a U-net approach to effectively reduce the amount of data labeling, thus achieving the expected effect of the model. Yang JY et al.[5] extracted rural buildings in high spatial resolution remote sensing images of Bazhou City, Hebei Province, based on SegNet image semantic segmentation algorithm with deep convolutional neural network and compared and analyzed with traditional methods such as ISO clustering, maximum likelihood method, support vector machine (SVM) and random forest, and finally the experiments showed that SegNet could efficiently utilize high spatial resolution spectral information and spatial feature information of rural buildings in remote sensing images. Marco Castelluccio et al. [6] used CaffeNet and GoogLeNet for convolutional neural networks and three different learning modes for the semantic classification study of remote sensing scenes and finally proved the effectiveness and wide applicability of the method. Baoxuan Jin et al. [7] proposed an approach that combines object-oriented methods with deep convolutional neural networks (COCNN). For the COCNN method, the accuracy and kappa index coefficients are 96.2% and 0.96, respectively, on the basis of classification statistics, which are 8.98% and 0.1 higher than those of the CNN-only-based method, respectively. The final experimental results show that the COCNN method reasonably and effectively combines the object-oriented and deep learning methods, thus effectively solving the problem of inaccurate classification of typical features, which is better than that of the CNN-only using CNN has better classification accuracy.

In this paper, we investigate the performance of the U-Net convolutional neural network model in remote sensing image segmentation. We apply the U-Net convolutional neural network model to the Aerial Imagery Dataset from Wuhan University 2019 release. The original aerial data of this dataset is from the New Zealand Land Information Service website. A total of 2 types of objects are labeled within the dataset, golden building, green other. The experimental results on the test set have achieved significant results, and the accuracy in the results reaches 0.9440, which shows a finer segmentation and extraction ability compared with the traditional methods.

2. Experimental description and data enhancement

2.1 Data set description and statistics

Wuhan University 2019 released Aerial Imagery Dataset, the original aerial photography data from the New Zealand Land Information Service website. The dataset has 8,189 remote sensing images with 0.3m resolution and 512×512 pixels in size. The dataset contains a total of 18,7000 buildings. The dataset contains an image folder for remote sensing images and a labeled folder for segmented images, with the same number of label files as image files, as shown in the following example:

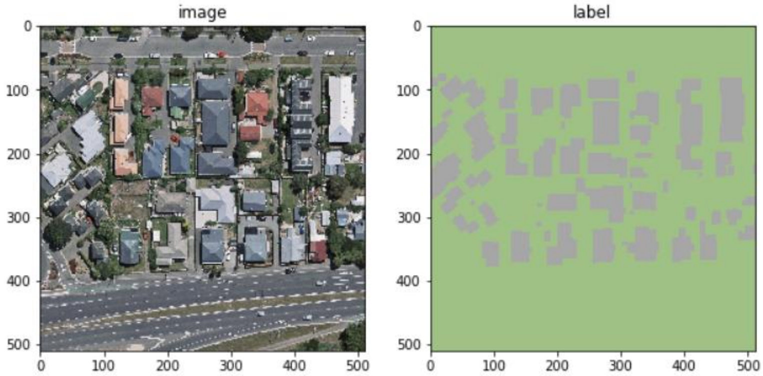

Figure 1: Data labels corresponding to original remote sensing images and remote sensing image plots

The types of parcels in the dataset are divided into two categories: building and other, and the distribution of the number of categories is shown in Figure 2.

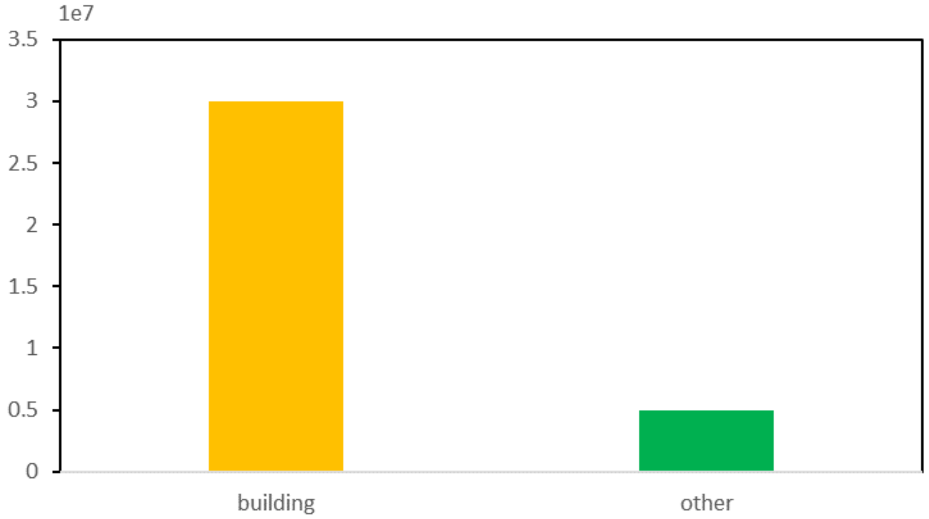

Figure 2: Data set label pixel statistics

2.2 Data Enhancement Strategy

In order to enhance the generalization performance and robustness of the model and to reduce the overfitting phenomenon of the network, some data enhancement methods are used to expand the data set when the sample size is small, such as: using up-and-down flip to make the model obtain rotation invariance; using cropping, stretching, etc. to make the model obtain size invariance; using adding noise to avoid the model learning useless high-frequency features; randomly changing saturation, etc. Since the amount of buildings in the dataset images is as high as 187,000, the training time will increase significantly if the dataset is further expanded.

In this study, a strategy of random data augmentation at the time of data loading was used in conducting the experiments, and the specific strategy was:

(1). Rotate randomly by an angle within [0,45 °].

(2). Translate horizontally and translate up and down by a distance within [0,0.1].

(3). A staggered tangent transformation keeps the point's x-coordinate (or y-coordinate) unchanged. In contrast, the corresponding y-coordinate (or x-coordinate) is translated by 0.5, and the magnitude of the translation is proportional to the perpendicular distance from the point to the x-axis (or y-axis).

(4). Zoom in the length or width direction. The parameter controls the image to be zoomed in and out to the same extent in both length and width directions.

(5). Randomly performs horizontal flip and up-down flip operations.

(6). Gaps in the image are filled in with the reflected mode, which conforms to the normal state. The corresponding label is transformed similarly when the image is flipped and scaled.

3. Introduction to Research Methodology and U-Net Network

3.1 U-Net Neural Network Model

The u-Net network was released in 2015, which has achieved good results in the semantic segmentation of medical images due to its fine volume and accurate extraction of linear image features, and it has been applied in various fields of semantic segmentation, one after another. The network structure is clear and presents a "tightening expansion" structure. It first uses convolution for downsampling to extract the features of each layer, then upsampling, and finally obtains an image of each pixel corresponding to its kind. The tightening phase gradually reduces the feature dimension and the number of parameters mainly through pooling, and the expansion phase recovers the details and dimensions of the image through the splicing of the number of features, forming a U-shaped structure, as shown in Figure 1. The most significant advantage of U-Net is its fast convergence speed and good segmentation effect, its symmetric structure is simple and easy to understand, and the model effect is excellent, so it has become one of the models for many network improvements. Therefore, this paper adopts the U-Net network structure to design a model for extracting image information in Figure 3.

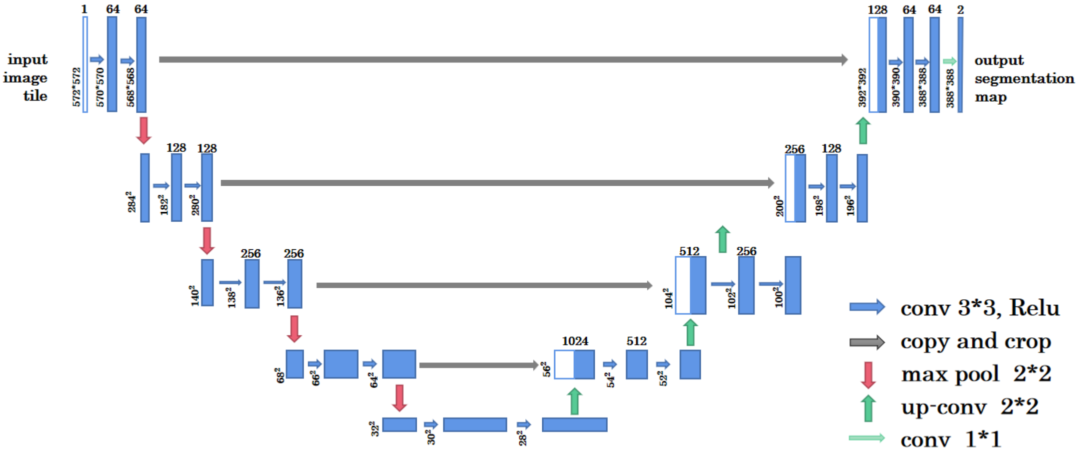

Figure 3: U-Net network structure diagram

The feature fusion method of U-Net is stitching, which splices the features in the channel dimension to form thicker features, while the fusion method of FCN is the sum of corresponding points and does not form thicker features.

3.2 Model Structure

The U-Net neural network used in this paper consists of 23 3x3 convolutional layers, 4 2x2 downsampling layers, 4 2x2 upsampling layers, and four skip connections to connect the feature maps generated by the deep and shallow networks, showing a "tightening-expansion" structure.

The U-Net single downsampling includes two 3x3 convolutional layers, a 2x2 pooling layer, an activation function of Relu (activation='relu'), compensation in the form of missing zeros (padding='same'), and a weight initialization parameter mode set to 'he_normal' (kernel_initializer='he_normal').

In the U-shaped structure, the feature information of the deep and shallow networks are stitched together using the Concatenate function, keeping the image dimension unchanged.

The reasons for choosing the U-Net network in this paper: (1) U-Net can use the corresponding labeled samples more effectively, based on a fully convolutional neural network, and still get more accurate segmentation by a small number of training images. (2) U-Net model has a good effect on binary classification. The soft-max and cross-entropy functions were used for the final energy function during the training process.

The Soft-max function is defined as:

$$p_k\left(X\right)=\frac{e^{\left(a_k\left(X\right)\right)}}{\sum_{k_1}^{k}e^{\left(a_{k_1}\left(X\right)\right)}} \tag{1}$$

The final soft-max of the image at each position of $p_{l(x)}^{(x)}$ combined with the cross-entropy function is:

$$E=\sum_{X\in\Omega}{}^{\text{`}} w(X)\log(p_{l(X)}(X)) \tag{2}$$

Highlight the importance of certain pixel points and introduce a weighting function:

$$w(X)=w_c\left(X\right)+w_0\cdot e^{\left(-\frac{(d_1\left(X\right)+d_2(X))^2}{2\sigma^2}\right)} \tag{3}$$

The weights of the network are initialized by a Gaussian distribution with:

$$f\left(x\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} \tag{4}$$

In this paper, Adam optimizer is used, and the learning rate is selected as '0.0001' according to the experience and running results.

Adam optimizer, based on the stochastic gradient descent algorithm (SGD), combines the advantages of both adaptive optimization (AdaGrad) and MSProp algorithms to calculate the update step by considering the gradient mean and the uncentered variance of the gradient synthetically.

Its main advantages are that the hyperparameters are well interpreted and usually require no adjustment or only a little fine-tuning; it is suitable for problems with sparse gradients or very noisy gradients; the integrated Adam counts in many cases as an optimizer with a better default working performance. For image boundary, the Adam optimizer can improve the boundary-blurring problem and improve the model accuracy. The learning rate control can make the model not easy to overfit; combined with the actual situation consideration, the Adam optimizer is more suitable for the problem of this paper.

4. Experimental study and analysis

4.1 Experimental environment and evaluation index

The experimental environment is a Win10 64-bit operating system server and an NVIDIA graphics card with 16GB of video memory. The hardware environment for the experiments is GPU GeForce GTX1080Ti, software Python3.8, and Tensorflow2.4.3.

The accuracy metric used in this paper is Accuracy. In remote sensing image segmentation, Accuracy is usually used to evaluate the correctness of an algorithm in classifying image pixels.

The calculation formula is:

$$\text{Accuracy}=\frac{\text{（TP+TN）}}{\text{（TP+TN+FP+FN）}} \tag{5}$$

In the above equation, TP denotes True Positive, i.e., the number of samples in which the model correctly determines positive cases as positive cases; TN denotes True Negative, i.e., the number of samples in which the model correctly determines negative cases as negative cases; FP denotes False Positive, i.e., the number of samples in which the model incorrectly determines negative cases as positive cases; and FN denotes False Negative, i.e., the number of samples in which the model incorrectly identifies positive cases as negative cases. In this paper, Accuracy is calculated to check the accuracy of the model. Since remote sensing images usually have high resolution and a large number of pixels, and the data set used in this paper has a large amount of data, data enhancement methods are used to process the data in order to not affect the experiment as much as possible.

4.2 Experimental protocol

Firstly, the data set is divided into a test set and a training set in the ratio of 1:9 for the data enhancement process. The data-enhanced images are not directly input to the network, but the original images and labels are first scaled and normalized, mapped to a uniform standard, and then the labels are one-hot encoded. And in this session, this paper defines the training data generator used to generate the training set and validation set in model training with a ratio of 4:1 and the test data generator used to pre-process the test images.

The key parameters are set to dropout=0.2, learning rate=1e-4, batch size=8, epoch=50, and after training, the validation set samples are predicted to obtain Accuracy accuracy metrics and evaluated for accuracy.

4.3 Experimental results

The Accuracy of the final model obtained from the training reached 94.40% for the predicted images obtained on the test set. The segmentation model trained by the method in this paper shows a strong segmentation ability. The segmentation boundaries between different types are quite accurate, overcoming the problem of an uneven sample size to some extent. However, there is still some room for improvement in the details at the edges. Comparison of prediction results are shown in Figure 4.

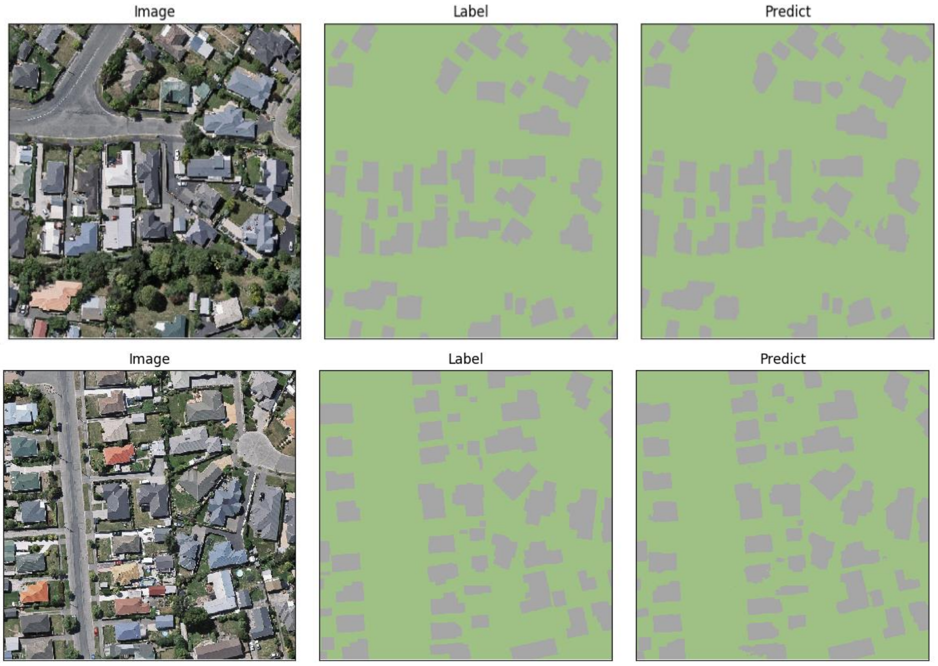

Figure 4: Comparison of prediction results

The model described in this paper was trained for 15 rounds, and the final model accuracy reached 94.40%, and the loss of each training round is shown in Figure 5, which indicates that the model is well-trained iteratively.

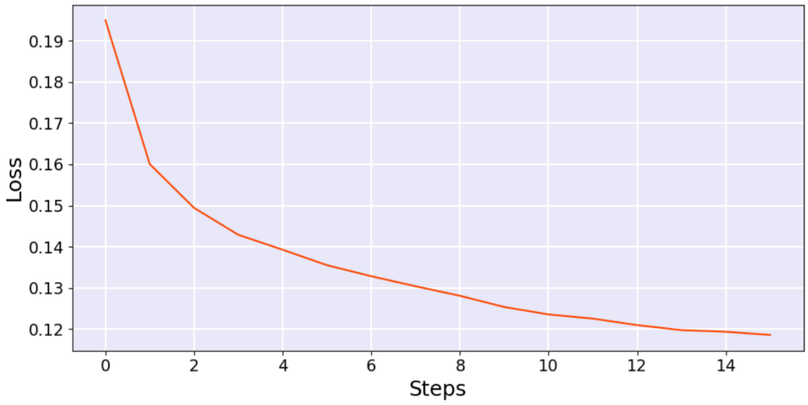

Figure 5: Training loss map

From the experimental results, it can be seen that the model obtained by using the method proposed in this paper has stronger segmentation ability for remote sensing images. This indicates that the problem of unbalanced sample distribution can be overcome to a certain extent when training with the method in this paper, and the obtained model is more robust and thus has finer segmentation ability.

5. Conclusion

The remote sensing image parcel segmentation method based on the U-Net convolutional neural network model studied in this paper was built by using a deep learning framework, and after the network was built, the model was trained on the Wuhan University Aerial Imagery Dataset dataset, and a random data enhancement strategy was added in the training process. The Accuracy of the result is 0.9440, which proves that the model trained by the method in this paper has a stronger segmentation ability.

The method in this paper also has some things that could be improved. It can be observed from the prediction images that the continuity of some regions is not good enough, and the segmentation results appear to be incomplete; some types of segmentation suffer from overfitting, and even very small parts scattered in other categories are segmented, and this over-sensitive segmentation ability should actually be avoided. Improving the continuity of segmentation results and avoiding overfitting of the model will be the focus of the next step and the key to improving the accuracy of prediction results.

References

[1] Yin Hao. Research on remote sensing image segmentation method based on convolutional neural network [D]. Shandong Agricultural University, 2021.

[2] Fu Xiuli, Li Lingping, Mao Kebiao et al. Remote sensing image classification based on convolutional neural network model [J]. High Technology Communication, 2017, 27(03): 203-212.

[3] Zhao Danxin, Sun Shengli. A new method of aircraft target detection based on ResNet for remote sensing images [J]. Electronic Design Engineering, 2018, 26(22): 164-168.

[4] Qu XY, Zhang E. Remote sensing image parcel target classification based on active learning [J]. Computers and Modernization, 2021(11): 50-55+60.

[5] Yang JY, Zhou ZX, Du ZJR, Xu QQ, Yin H, Liu R. Extraction of rural construction land based on SegNet semantic model for high-resolution remote sensing images [J]. Journal of Agricultural Engineering, 2019, 35(05): 251-258.

[6] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, Luisa Verdoliva. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. [J]. CoRR, 2015, abs/1508.00092.

[7] Baoxuan Jin, Peng Ye, Xueying Zhang, Weiwei Song, Shihua Li. Object-Oriented Method Combined with Deep Convolutional Neural Networks for Land-Use- Type Classification of Remote Sensing Images [J]. Journal of the Indian Society of Remote Sensing, 2019, 47(6).