Extraction of Suspected Illegal Buildings from Land Satellite Images Based on Fully Convolutional Networks

Yu PEI, Xi SHEN, Xianwu YANG, Kaiyu FU, Qinfang ZHOU*

Yunnan Map Institute, Kunming 650034, China

Abstract In the management of land resources and the protection of cultivated land, the law enforcement of land satellite images is often used as one of the main means. In recent years, the policies and regulations of the law enforcement of land satellite images have become more and more strict and been adjusted increasingly frequently, playing a decisive role in preventing excessive non-agricultural and non-food urbanization. In the process of the law enforcement, the extraction of suspected illegal buildings is the most important and time-consuming content. Compared with the traditional deep learning model, fully convolutional networks (FCN) has a great advantage in remote sensing image processing because its input images are not limited by size, and both convolution and deconvolution are independent of the overall size of images. In this paper, an intelligent extraction model of suspected illegal buildings from land satellite images based on deep learning FCN was built. Kaiyuan City, Yunnan Province was taken as an example. The verification results show that the global accuracy of this model was 86.6% in the process of building extraction, and mean intersection over union (mIoU) was 73.6%. This study can provide reference for the extraction of suspected illegal buildings in the law enforcement work of land satellite images, and reduce the tedious manual operation to a certain extent.

Key words Deep learning; Fully convolutional network; Semantic segmentation; Law enforcement of land satellite images; Extraction of suspected illegal buildings

DOI 10.19547/j.issn2152 - 3940.2025.01.016

In the law enforcement of land satellite images, newly-added illegal construction land accounts for most of the illegal types, and the results of the law enforcement of land satellite images, as basic data, have been widely used in natural resources, agriculture, environmental protection, emergency management, energy security, public security and other related industries, with broad application prospects and huge potential[1-2]. Buildings in remote sensing images tend to have more obvious features and higher recognition than other map spots, and the same is true for computer perspective. Studying the extraction of suspected illegal buildings in the law enforcement of land satellite images based on artificial intelligence technology can better support natural resource management and serve the construction of ecological civilization, support the needs of various industries, and serve social and economic development[3].

At present, the extraction of suspected illegal buildings mainly rely on manual extraction, with advantages of a long extraction cycle, wrong and miss extraction, and low overall efficiency. The main form is to determine the spatial position of land use change by manual visual interpretation of high-resolution satellite remote sensing images before and after different years. The extraction is mostly conducted based on the subjective experience of the operators to interpret the features of remote sensing images, and the quality cannot be guaranteed. Moreover, the speed of manual extraction is difficult to meet the current work. Therefore, in the context of big data and artificial intelligence, it is urgent to use advanced theories such as machine perspective and deep learning[4-6] to realize the automatic extraction of suspected illegal buildings.

In this paper, fully convolutional neural network, which has a high usage rate in deep learning, was applied to the extraction of suspected buildings in the law enforcement of land satellite images. Based on the cultivated land data, historical illegal records and remote sensing image information of Kaiyuan City, Honghe Hani and Yi Autonomous Prefecture, Yunnan Province over the years, the method for the extraction of suspected illegal buildings in the law enforcement of land satellite images based on fully convolutional neural network was mainly studied to improve the extraction speed and accuracy of suspected illegal buildings in the process of the law enforcement.

1 Fully convolutional network model

In recent years, the deep learning technology in the field of artificial intelligence is booming at an unprecedented speed, which profoundly reflects the simulation and transcendence of the complex information processing mechanism of the human brain. The core idea is to build a multi-layered network architecture that forms a deep neural network by stacking nonlinear data conversion modules (i.e. neurons or layers) on top of each other. Based on massive training data, this structure can gradually optimize its own parameters to approximate or represent extremely complex input-output relationships, and achieve highly flexible and powerful data representation capabilities[7-8]. With their high sensitivity to com-

Received: January 18, 2025 Accepted: February 25, 2025

* Corresponding author.

plex data patterns and learning ability, a deep learning algorithm has become a powerful tool for detecting changes in high resolution, and can accurately capture and analyze subtle changes in images using annotated datasets. It is widely used in satellite remote sensing, medical analysis, automatic program and other image recognition fields. This network model constructs a multi-layer convolutional neural network architecture, and makes use of a large number of high-resolution label data for in-depth analysis[9-11]. For example, Chen *et al.*[12] proposed a method based on temporal and spatial attention, and applied it to the detection of changes in remote sensing images, which greatly improved the detection accuracy. Using deep learning and change vector analysis (CVA), Wang Chang *et al.*[13] detected changes in three groups of remote sensing image datasets, and the results show that the change detection accuracy of deep learning method was higher than that of CVA method.

In the task of image classification, there are three kinds of convolutional neural network models with outstanding performance, namely AlexNet, GoogLeNet and Vgg16 Net. In 2015, Long *et al.* carried out a fully convolutional network transformation of these three models. Then, on the PASCAL VOC dataset organized by the PASCAL (Pattern Analysis, Statical Modeling and Computational Leaning) network, conducted experiments on image semantic segmentation, and proposed the concept of fully convolutional networks (FCN). The experimental results show that the model based on Vgg16 Net had the best effect in image semantic segmentation. Based on the above research results, the above FCN transformed based on Vgg16 Net was selected to extract building information. In this paper, the main task of extracting building information is to divide remote sensing images into two categories: buildings and background. Therefore, the number of categories to be classified in this network was set as 2. A deep learning model usually includes input layer, convolution layer, activation layer, pooling layer, fully connected layer, and output layer. Full convolution means that fully connected layer in the classification network are replaced by convolutional layers. FCN is derived from convolutional neural networks (CNN)[15], and CNN means connecting fully connected layer after the convolutional layer and finally outputting a result corresponding to the probability of various categories. FCN upsamples the convolutional output through the deconvolution layer to restore it to the size of the input image, so as to realize the classification prediction of a pixel. Compared with CNN, the fully connected layer of CNN needs a fixed size to determine the weight, while the input image of FCN is not limited by size, and neither convolution nor deconvolution depend on the overall size of an image. The network structure of FCN is shown in Fig. 1.

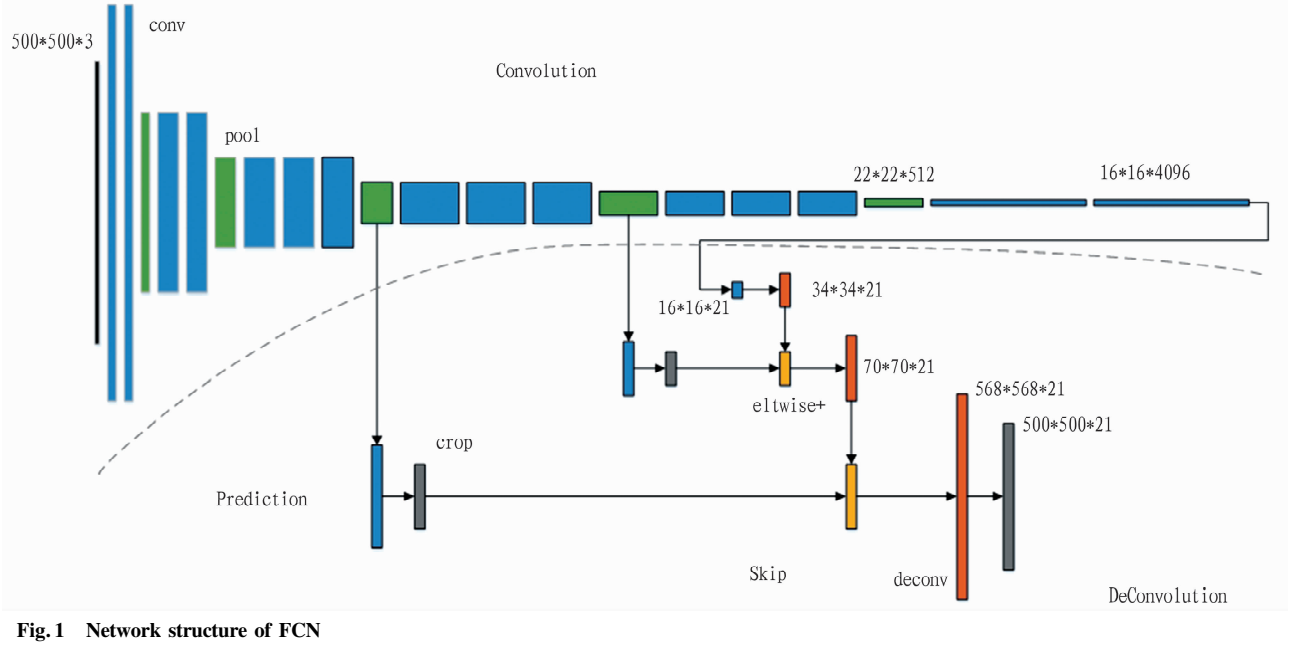


Fig. 1 Network structure of FCN

2 Construction of an intelligent extraction model of suspected illegal buildings from land satellite images

Aiming at the problems existing in the process of extracting suspected illegal buildings from land satellite images, an intelligent extraction model of suspected illegal buildings from land satellite images based on FCN was constructed. The function of the input layer is to read the samples. In this paper, the images of buildings with different sizes were used as the samples. Convolution layer, activation layer and pooling layer were used to extract sample features. The output layer predicts the classification results of the output pixel.

2.1 Convolution computation In mathematics, the convolution $h(x)$ of two functions $f(x)$ and $g(x)$ can be calculated as fol-

lows:

$$h(x) = \int_{-\infty}^{\infty} f(\tau) g(x-\tau) d\tau \quad (1)$$

In deep learning, discrete convolution can be calculated as follows:

$$y(n) = \sum_{-\infty}^{\infty} x \cdot h(n-i) \quad (2)$$

In the formula, x and h represents the weight of input and to be learned, respectively; i represents the current item; n is the total item. The core of convolution calculation is to use a convolution kernel of preset size to carry out the ergodic multiplication and summation operation on an image surface according to the given step size. In this process, the weight and bias values contained in the convolutional kernel are not artificially set, but are obtained relying on the automatic learning and optimization of the neural network in the training stage.

The pooling layer plays a key role in reducing the size of a feature map and model parameters. It achieves this function by moving a fixed-size window at a certain step on the feature map output by the convolutional layer. Within the area covered by each window, the pooling layer selects the maximum (namely performing the maximum pooling operation) or calculates the average (namely performing the average pooling operation), and uses the value as the output of that window. This mechanism not only effectively reduces the dimension of the feature map, but also further simplifies the number of parameters in the network, thus improving the training efficiency and generalization ability of the model.

2.2 Upsampling In FCN, the last several fully connected layers of the convolutional neural network are replaced by convolutional layers, and the output of the input image is a high-dimensional feature map after several convolutional pooling operations. In order to solve the problem of image size reduction caused by convolution and pooling, the feature map is upsampled to the same size as the input image, and then each pixel is classified to achieve the purpose of image segmentation.

The upsampling methods mainly include deconvolution, uppooling and bilinear interpolation. Bilinear interpolation was used in this paper, and the formula[14] is as follows:

$$y_{s,t} = \sum_{\alpha,\beta=0}^{1} \left| 1-\alpha-\left\{\frac{s}{n}\right\} \right| \left| 1-\beta-\{t/n\} \right| x_{\lfloor s/n \rfloor+\alpha,\lfloor t/n \rfloor+\beta} \quad (3)$$

In the formula, n is the sampling factor greater than 0, and is determined according to the magnification times of the feature map; s, t is the relative coordinate of the pixel after sampling, with the value range of $[1, n]$; α, β is 0 or 1, and is used to determine the position of the four pixels closest to the output pixel point; $\{\cdots\}$ denotes taking the decimal part; $\lfloor \cdots \rfloor$ means taking the integer part; $|\cdots|$ means taking absolute value; x is the pixel value of the input feature map; y is the pixel value of the output feature map by upsampling[16].

2.3 Update strategy of learning rate Learning rate varies dynamically according to a polynomial pattern. Its update strategy is formula (4):

$$new\_lr = \lambda \cdot initial\_lr \quad (4)$$

In the formula, *new _ lr* is the new learning rate obtained; *initial _ lr* is the initial learning rate; λ is obtained by parameters *initial _ lr* and *epoch*, and *epoch* is the number of iterations of the model.

2.4 Loss function In deep learning, the loss function is used to evaluate the performance of a model. The loss function quantifies the degree of classification errors of the model, and continuously adjusts the parameters during the training process to optimize its minimum, so as to make the data more fitting and increase its generalization ability. In this paper, cross entropy loss function was adopted, and its formula is as follows:

$$Loss = -\sum_{i=1}^{n} [\, y_i \ln \hat{y}_i + (1-y_i) \ln(1-\hat{y}_i) \,] \quad (5)$$

In the formula, n represents the total number of samples; i means the current sample; y_i stands for the true label; $\hat{y}_i$ is the prediction value of the model, namely a probability between 0 and 1.

The stochastic gradient descent (SGD) algorithm can be adopted for iterative training[17]. When the classification results of the model are the same as the true label, the minimum of the cross entropy loss function is 0. As the difference between them is greater, the value of the cross entropy loss function approaches infinity. By minimizing the cross entropy loss function, the prediction of the model is approximated to the true label to the greatest extent, and the accuracy and classification performance of the model are improved.

2.5 Method of accuracy evaluation In this paper, pixel accuracy (PA), global accuracy (GA), mean intersection over union (mIoU) and Kappa coefficient were used as evaluation indicators. Their formulas are as follows:

$$PA_i = \frac{P_{i,i}}{\sum_{m=1}^{n} P_{i,m}} \quad (6)$$

$$GA = \frac{\sum_{i=1}^{n} P_{i,i}}{\sum_{i=1}^{n}\sum_{m=1}^{n} P_{i,m}} \quad (7)$$

$$mIoU = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sum_{i=1}^{n} (x_i + y_i - x_i \cdot y_i)} \quad (8)$$

$$Kappa = \frac{N \cdot \sum_{i=1}^{n} P_{i,i} - \sum_{i=1}^{n} (\sum_{i=1}^{n} P_{i,m}) \cdot (\sum_{m=1}^{n} P_{m,i})}{N^2 - \sum_{i=1}^{n} (\sum_{i=1}^{n} P_{i,m}) \cdot (\sum_{m=1}^{n} P_{m,i})} \quad (9)$$

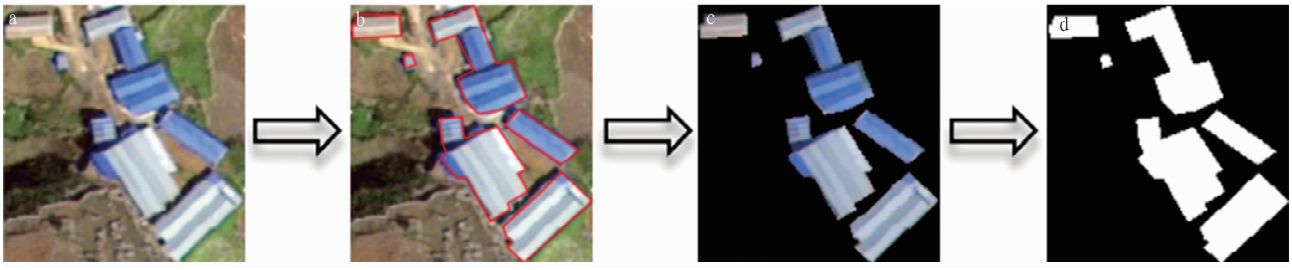
In the formulas, PA_i is the accuracy of pixel i; n represents the number of categories in semantic segmentation, namely the number of channels output by the last layer of the neural network; $P_{i,m}$ means the number of pixels that should belong to class i but are predicted to be class m; $P_{m,i}$ represents the number of pixels that should belong to class m but are predicted to be class i; N is the total number of pixels[16].

3 Case verification

3.1 Data sources Kaiyuan City, located in the southern end of Yunnan Plateau, is adjacent to Kangdian axis in the west and Qiangui Platform in the east, and is between two natural faults (the Red River and Nanpan River). The long-term geological tectonic activities have shaped a rich and diverse geomorphic landscape here, including fold belts, fault layers, subsidence basins and widely distributed karst mountains, as well as Zhongshan Val-

ley formed by the deep erosion of the Nanpan River and its tributaries, which together constitute a moderately shallow cut geomorphic map of Zhongshan Plateau. The topography of the region is distinctive, with mountains extending from north to south. The terrain is high in the southeast and low in the northwest, forming a significant fluctuation. The highest altitude reaches 2 775.6 m, while the lowest altitude is as low as 950 m, and the height difference between the two is as high as 1 825.6 m. This significant relative height difference further strengthens the three-dimensional terrain characteristics of the area. In this paper, the local area of Kaiyuan City is as the research area. The experimental data are based on the monitoring images of natural resources in the first quarter of 2024. The images include three bands of red, green and blue, and the spatial resolution is 1 m. The spatial reference coordinate system is CGCS2000_3_Degree_GK_Zone_34. The total image width is 65826 × 51988, and the pixel depth is 8 bits.

3.2 Building of a sample library On the images, buildings with independent spatial characteristics were marked by artificial vectorization, and the extraction locations were randomly and evenly distributed in urban areas and mountainous areas. The scale size of suspected illegal buildings is evenly distributed. A number field was added to the layer, and a separate number is filled. The steps of data preprocessing are shown in Fig. 2.



Note: a. Original image; b. Vector data; c. Clipping vector; d. Reclassification.

Fig. 2 Steps of data preprocessing

Using Python's GDAL library, the extracted vector layer was superimposed with an image, and the minimum external rectangle was clipped according to the pattern group. The clipped image was reclassified, and the core idea of the classification algorithm is to traverse pixel points. If they intersected with the range of vector elements, they were reclassified as pixel value 1. If they did not intersect with it, they were reclassified as pixel value 0. The coordinate information and size of the image before and after reclassification were kept consistent, and the number of band channels became 1 after clipping. The total number is 210.

The clipped image was used as the original image dataset, containing 3 bands of red (R), green (G) and blue (B). The reclassified image was used as the building label dataset, including 1 band. Cases are shown in Fig. 3.

Note: a, c. Original image; b, d. Building label.

Fig. 3 Cases of label dataset

3.3 Setting of related parameters The total number of images is 210, and 200 images participated in training, while 10 images are as the test set. The size of the training batch was set to be evenly divided by the number of samples, and it is 4 here. The number of iterations of the training data was set to 30. The initial learning rate was set to 0.000 1. The learning rate varies dynamically according to a polynomial pattern.

3.4 Monitoring of model training process The monitoring data of the training process are shown in Table 1 and Fig. 4. It can be seen that with the increase of the number of iterations, the global accuracy and mean intersection over union rose constantly. The learning rate gradually updated and decreased. Although the accuracy of buildings reached a peak of 84.3% at the 25th iteration, its global accuracy and mean intersection over union were not the highest, so it was not suitable for the final training model from the overall level.

3.5 Test results and accuracy verification The FCN model was obtained by training 200 manually labeled samples. After 30 iterations, the class accuracy was 78.6% for buildings and 90.3% for other classes. Global accuracy was 86.6%, and mean intersection over union was 73.6%. The training loss function value was 0.551 2.

In this paper, based on FCN, part of Kaiyuan City was taken as the research area, and 210 images of different sizes were reclassified into buildings and other categories. 200 images were randomly selected as training samples, and iteratively analyzed to

generate an intelligent extraction model of suspected illegal buildings. The remaining 10 images were as the prediction verification set. At the same time, 10 images from the Yunnan World Map website were selected to participate in the prediction and evaluate the accuracy of the model.

The comparison of prediction results of the test set is shown in Fig. 5. Color images are the original images, while white images are the label values of buildings; red images are the prediction values of buildings, and black images are the pixel values of other classes. As can be seen from the images, the FCN model can predict the location of buildings and achieve reclassification, which can be converted into vector elements by Python or other tools in the later stage. In order to ensure the universality of the inspection, the scale of the test set is different, and the number of buildings covered is different; the building style is also diversified.

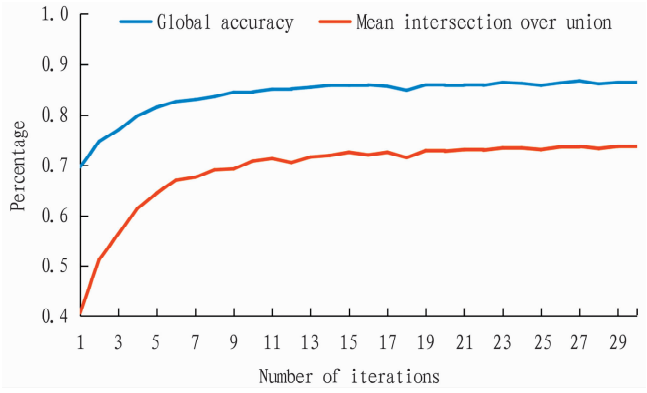


Fig. 4 Variation of global accuracy and mean intersection over union with the number of iterations during training

Table 1 Monitoring data of training process

Number of iterations	Function loss	Learning rate	Global accuracy	Building accuracy	Other accuracy	Mean intersection over union	Building overlap	Other overlap
5	0.813 4	0.000 087	81.5	63.8	89.7	64.5	52.2	76.8
10	0.642 7	0.000 072	84.6	81.5	86.0	70.9	62.6	79.2
15	0.599 8	0.000 055	85.9	79.4	88.9	72.6	64.1	81.2
20	0.557 7	0.000 038	85.9	81.3	88.1	72.8	64.6	81.1
25	0.583 3	0.000 021	85.9	84.3	86.7	73.1	65.4	80.8
30	0.551 2	0	86.6	78.6	90.3	73.6	65.0	82.2

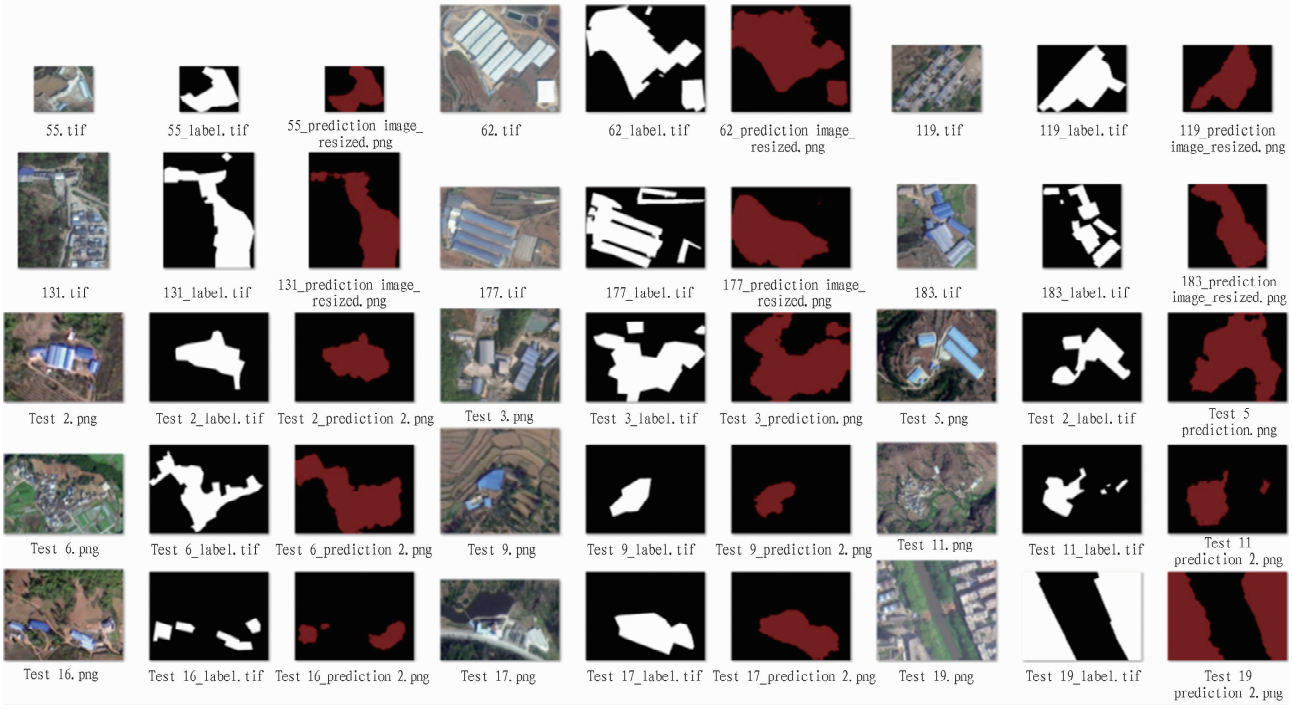


Fig. 5 Comparison of prediction results of the test set (part)

As shown in Table 2, the classification accuracy of buildings and other classes reached 84% and 95%, respectively. Kappa coefficient, an evaluation indicator, is used to measure the accuracy of classification and confirm whether remote sensing classification images is consistent with actual ground images[18]. Here, Kappa coefficient was 92.01%, indicating a high consistency between the prediction results and actual values. This is also true in terms of the generated results. Most of the predicted results show the characteristic of concentration. In the law enforcement work of land satellite images, suspected illegal buildings are mostly con-

centrated, and there is less detailed drawing, which is in line with expectations.

Table 2 Confusion matrix of extraction results of the test set

Number of actual pixels	Number of predicted pixels				
	All	Building	Others	Total	Correct rate
All	6 478 759	1 727 501	4 751 258	6 478 759	–
Building	1 465 726	1 371 387	261 775	1 633 162	0.84
Others	5 013 033	261 775	4 656 919	4 918 694	0.95
Total	6 478 759	1 633 162	4 918 694	6 551 856	–

Compared with the traditional manually extracted buildings, the accuracy of this experiment tended to be higher on the basis of saving time cost. Seen from the test results, the range of buildings extracted by manual experience was included in the prediction results of the test. Although the details of the local range cannot reach the detailed manual drawing, it can ensure that buildings were not missed, and only a small number of pixels were exceeded.

4 Conclusions

The research on methods for the intelligent extraction of suspected illegal buildings from land satellite images has great social and economic values for building a modern investigation, monitoring and supervision service system of natural resources, improving the service level of surveying and mapping geographic information, accelerating the development of social application of geographic information and remote sensing, and promoting the transformation of government departments' supervision methods[19]. In this paper, based on survey and monitoring images and the complex landform of Kaiyuan City, Yunnan Province, data pre-processing was optimized. Compared with the traditional method of building a sample library, this experiment clipped the smallest external rectangle in the form of "clustering" of multiple image spots, which greatly improved the authenticity and diversity of sample data and the accuracy of the model algorithm on the basis of data. The FCN model based on deep learning was used to achieve the intelligent extraction of suspected illegal buildings from land satellite images. The experimental results show that this separation model can be used to complete the repetitive work, and then the output results were secondarily analyzed and screened manually, which greatly sped up the extraction process and analysis accuracy. From the reflection spectrum of remote sensing ground objects, the spectral characteristics of buildings were more obvious than other ground objects, which increased the feasibility of the model to a certain extent.

There are the following problems in the experiment. Firstly, in the screening of samples, the number of samples was small, so that the features of model learning are relatively few. In general, the proportion of buildings in the samples was relatively small. Secondly, the frequency of data training was low, that is, 30 iterations of training were conducted due to the time cost. Thirdly, the accuracy of the test image was not high, resulting in blurred extraction boundary, which affected the classification accuracy of the model to a certain extent.

Based on the above problems, the future research directions are as follows. Firstly, 0.5 m high-resolution images or UAV orthographic remote sensing images are used for the test, and the number of samples should be increased as much as possible to improve the accuracy of the model. Secondly, a test on the number of iterations is conducted to constantly adjust parameters, optimize weight, and strive to achieve a better solution of the training model. Thirdly, based on the range of buildings extracted in this paper, it can be read by python or other tools and converted into vector elements for use and analysis in the law enforcement of land satellite images.

References

[1] DENG DZ. Deep learning-based cloud detection method for multi-source satellite remote sensing images[J]. Remote Sensing for Natural Resources, 2019, 35(4): 9 – 16.

[2] FAN H, WANG H, LIU WN, *et al.* Analysis of flood disaster impact based on domestic high-resolution satellite data[J]. Geospatial Information, 2019, 21(8): 65 – 67.

[3] China Surveying and Mapping. Top ten news of surveying and mapping geographic information in 2023[J]. China Surveying and Mapping, 2024(Z1): 33 – 36.

[4] ZHU XX, TUIA D, MOU L, *et al.* Deep learning in remote sensing: a comprehensive review and list of resources[J]. IEEE Geoscience and Remote Sensing Magazine, 2017, 5(4): 8 – 36.

[5] LI ST, SONG WW, FANG LY, *et al.* Deep learning for hyperspectral image classification: An overview[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(9): 6690 – 6709.

[6] DENG LT, ZHAO YR. Deep learning-based semantic feature extraction: A literature review and future directions[J]. ZTE Communications, 2023, 21(2): 11 – 17.

[7] ZHANG J, GAO Y. Application of deep learning remote sensing image interpretation technology in cultivated land protection[J]. Bulletin of Surveying and Mapping, 2023(8): 142 – 145.

[8] ZHANG LP, WU C. Advance and future development of change detection for multi-temporal remote sensing imagery[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(10): 1447 – 1459.

[9] GU JX, WANG ZH, KUEN J, *et al.* Recent advances in convolutional neural networks[J]. Pattern Recognit, 2018, 77: 354 – 377.

[10] MILLETARI F, NAVAB N, AHMADI SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3DVision (3DV). October 25 – 28, 2016, Stanford, CA, USA. IEEE, 2016: 565 – 571.

[11] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer – Assisted Intervention. Cham: Springer, 2015: 234 – 241.

[12] CHEN H , SHI Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection[J]. Remote Sensing, 2020, 12(10): 1662.

[13] WANG C, ZHANG YS, WANG X, *et al.* Remote sensing image change detection method based on deep neural networks[J]. Journal of Zhejiang University: Engineering Science, 2019, 54(11): 2138 – 2148.

[14] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640 – 651.

(To page 75)

atic structure chain between the historical and cultural space of Huai'an and the Grand Canal culture. The study hopes to provide a beneficial supplement for the research and practice of urban development of Huai'an and the protection of the human settlement environment of the Grand Canal.

References

[1] SHI AD. The co-temporal study of folk literature[J]. Folklore Studies, 2021(1): 106 – 117.

[2] LUAN FS. Luan Fengshi's archaeological collection[M]. Beijing: Cultural Relics Publishing House, 2017: 487 – 491.

[3] XING YM, TANG KX. Application of typology on the protection and renewal of canal city buildings[J]. Water Resources Planning and Design, 2019(1): 28 – 31.

[4] LI Z, LIU JR. Regional differences between northern and southern Chinese construction rulers[J]. Journal of Architectural History, 2023, 4(1): 18 – 30.

[5] HE Y, WANG CS. Evolution of spatial form of historic towns along the Beijing – Hangzhou Grand Canal: A case study of Hexia Ancient Town of Huai'an[J]. Urban Development Studies, 2023(2): 25 – 30.

[6] WU DL. Study on design characteristics and development context of traditional residential Houses in Lixiahe, Jiangsu from the early Ming Dynasty to the late Ming Dynasty[D]. Wuxi: Jiangnan University, 2024.

[7] XU CJ, LIAO L. Study on decorative patterns of traditional residential buildings in Hexia Ancient Town[J]. Popular Literature and Art, 2020: 66 – 67.

(From page 69)

[15] JONA E. GERLACH ANDRÉ, MARCUS M, *et al.* Convolutional neural network with data augmentation for object classification in automotive ultrasonic sensing[J]. The Journal of the Acoustical Society of America, 2023, 153(4): 2447 – 2459.

[16] CHEN M, WANG XQ. The study on extraction of seismic damage of buildings from remote sensing image based on fully convolutional neural network[J]. Technology for Earthquake Disaster Prevention, 2019, 14(4): 810 – 820.

[17] BOTTOU L. Stochastic gradient descent tricks[M]//Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 2012: 421 – 436.

[18] WU HB, DAI SY, WANG AL, *et al.* Collaborative classification of hyperspectral and LiDAR data based on CNN-transformer[J]. Optics and Precision Engineering, 2018, 32(7): 1087 – 1100.

[19] PANG M. Extraction of domestic satellite images patches based on deep learning[J]. Bulletin of Surveying and Mapping, 2024(4): 124 – 128, 134.