

Research Article

Land Use Classification Using Improved U-Net in Remote Sensing Images of Urban and Rural Planning Monitoring

Xiaoling Xie , Lin Ye , Xueqin Kang , Lei Yan , and LiQin Zeng 

School of Architecture and Urban Planning, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China

Correspondence should be addressed to Xiaoling Xie; xiexl@mail.lzjtu.cn

Received 27 January 2022; Revised 11 March 2022; Accepted 21 March 2022; Published 12 April 2022

Academic Editor: Ahmed Farouk

Copyright © 2022 Xiaoling Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problem of low accuracy and efficiency of existing land use classification methods for high-resolution remote sensing image segmentation, a land use classification method using improved U-Net in remote sensing images of urban and rural planning monitoring is proposed. First, taking the high-resolution remote sensing images of different remote sensing satellites as the data source, the remote sensing images in the data source are registered and cropped so that the pixels at the corresponding positions represent the same geographical location. Then, the encoder of the U-Net model is combined with the residual module to share the network parameters and avoid the degradation of the deep network. The dense connection module is integrated into the decoder to connect the shallow features with the deep features, so as to obtain new features and improve the feature utilization rate. Finally, the depthwise separable convolution is used to process the spatial and channel information of the convolution process separately to reduce the model parameters. Experiments show that the pixel accuracy, recall rate, precision rate, and average intersection-over-union ratio of the proposed land use classification method based on improved U-Net are 92.35%, 80.56%, 83.45%, and 86.75%, respectively, which are better than the compared methods. Therefore, the proposed method is proved to have good land use classification ability.

1. Introduction

Human beings' excessive demand and consumption of land resources have brought various problems, such as land degradation, land desertification, sharp reduction of forest land and grassland area, soil erosion, serious land pollution, and so on. These problems directly lead to the paralysis of some systems and the extinction of rare species [1]. If human beings create value without paying attention to the protection of nature and blindly destroy it, the living conditions will be greatly reduced and the future development will be limited. Therefore, it is necessary to reasonably classify and use land [2–4].

With the rapid development of high-speed imaging sensors and remote sensing technology, the resolution of remote sensing image can be as high as 0.41 m. On the one hand, the maturity of high-resolution remote sensing image technology provides accurate information for understanding the world; on the other hand, it brings great challenges to the

automatic and intelligent interpretation of remote sensing image. How to make full use of these high-quality data is an important research direction [5–8]. High-resolution remote sensing images contain rich and complex surface information and record more ground object details. They are widely used in agriculture, industry, military, and other fields. The study of ground information through high-resolution remote sensing images can not only facilitate resource survey, disaster monitoring, urban planning, and military defense but also contribute to the improvement of intelligent technologies such as unmanned driving, crop planting statistics, remote sensing mapping, and so on [9–12].

Deep learning has strong feature fitting ability, so it is widely used in remote sensing image scene classification. This method extracts image features layer by layer in an end-to-end way, fuses them to form high-level features, and finally generates the semantic description of the image, that is, category label [13–17]. Due to the use of real semantic

labels, deep learning can extract feature representations highly related to image categories and achieve high classification accuracy [18].

The widely used land use classification methods mainly include visual interpretation, supervised classification, unsupervised classification, and so on [19]. In recent years, with the development of artificial intelligence, the deep learning method has been gradually introduced into land use classification [20]. Zhang et al. [21] combined the advantages of convolutional neural network and Markov random field and proposed a variable precision rough set (VPRS) model to quantify the uncertainty in the classification of VFSR image by convolutional neural network (CNN). Maggiori et al. [22] proposed a CNN model for remote sensing image classification. CNN was used to learn the contextual features of image labels on a large scale. This network was composed of four stacked convolution layers. The image was down-sampled, and the relevant features were extracted. In order to overcome the shortcomings of large-scale labeled remote sensing image datasets, Scott et al. [23] proposed a method of combining transfer learning with deep convolutional neural network (DCNN), which used TL to guide DCNN. This method retained the depth visual features learned from image corpora in different image domains, so as to improve the robustness of DCNN to remote sensing image data. Mou et al. [24] proposed a full Conv-Deconv network for unsupervised spectral spatial feature learning of hyperspectral images, which could be trained end-to-end. Mou et al. [25] proposed a hyperspectral image classification method based on recurrent neural network, which used the newly proposed activation function Pre-tanh to replace the traditional tanh function for hyperspectral sequence data analysis. Naushad et al. [26] proposed a land classification method combining VGG16 and wide residual network and fine-tuned the model with transfer learning. Considering pixel-level classification and boundary mapping, Dong et al. [27] proposed a new feature integration network (FE-Net), which included two stages: multi-scale feature encapsulation and enhancement. However, in the face of high-resolution remote sensing images, the above methods are difficult to effectively mine data features, and there are many model parameters, which have the problems of low accuracy and efficiency.

Aiming at the problem of low accuracy and efficiency of existing land use classification methods for high-resolution remote sensing image segmentation, a land use classification method using improved U-Net in remote sensing images of urban and rural planning monitoring is proposed. The innovations of the proposed method are as follows:

- (1) The encoder is combined with the residual module to share the network parameters and avoid the degradation of the deep network. The dense connection module is used to cascade the upper features with the deep features.
- (2) The convolution modules in the model are replaced by depthwise separable convolution. The spatial and channel information of the convolution process are processed separately. The direct correlation between

spatial and channel is removed, so the amount of parameters of the model is effectively reduced.

2. Construction of Dataset

2.1. Overview of Satellite Information Used in Datasets.

The multi-temporal remote sensing images in the dataset used in this paper come from different remote sensing satellites. QuickBird series satellites were launched by American company DigitalGlobe in 2001. It is one of the first commercial satellites in the world to provide sub-meter resolution. Its orbital altitude is about 450 km, its mass is about 1018 kg, the regression cycle is 1 to 6 days, and the corresponding actual area of a single image is about 272.25 km². The provided resolutions include panchromatic 0.61 to 0.71 meters and multi-spectral 2.44 to 2.88 meters. The sensor can detect four different bands: 450–520 nm blue band, 520–600 nm green band, 630–690 nm red band, and 760–900 nm near-infrared band.

Landsat series satellites are led by NASA in the United States to observe and study global change. At present, only Landsat7 and Landsat8 are in service. Landsat7 satellite carries enhanced thematic plotter sensors. Compared with the original sensor device, the device significantly improves the image resolution and positioning quality by setting absolute calibration on the satellite. A single image covers an area of 32375 km². Landsat8 has an orbital height of about 705 km, an orbital period of 99 minutes, and a revisit period of 16 days. The satellite is mainly equipped with two different sensors, namely, land imager and thermal infrared sensor. The two sensors provide 11 different bands, which can provide richer synthesis schemes through different combinations. The effects of different band combinations are shown in Figure 1, from which it can be clearly seen that the natural color map is closer to the daily images, while the vegetation analysis map focuses on the plant color, which can clearly distinguish vegetation from the surrounding environment. The atmospheric penetration map can shield some highlights to reduce the impact of light.

2.2. Detailed Introduction of Dataset.

Most remote sensing images in the dataset are central and eastern China and contain a wide variety of changes, such as bare land changing into roads, farmland changing into high-speed rail tracks, wasteland changing into factories, urban building demolition, new buildings, mountain reclamation, river diversion, ore mining, and so on. The dataset contains 13680 pairs of multi-temporal remote sensing images with a resolution of 256 × 256, and the corresponding area of each image is about 2.5 km². Some images do not come from the same satellite, so first, each pair of images need to be registered and cropped so that the pixels at the corresponding positions represent the same geographical location. However, even so, there is still the problem of angle deviation caused by different satellite shooting angles of some buildings. For example, Figure 2 shows images of the same building obtained by different remote sensing satellites at the same location. It can be seen from the figure that these buildings

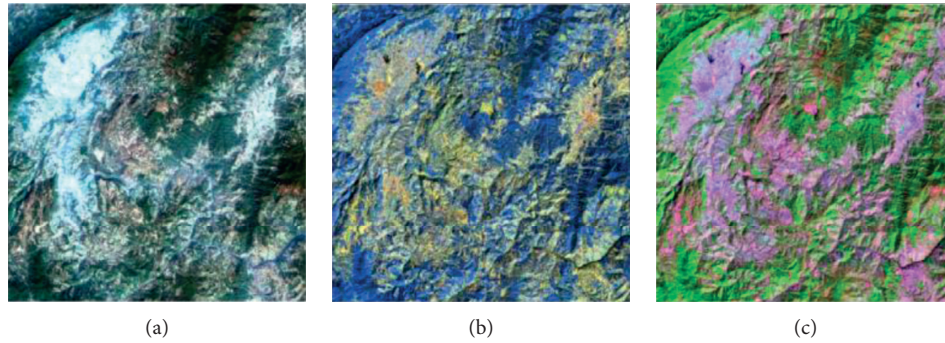


FIGURE 1: Composite maps of different band combinations of Landsat8 satellite. (a) Natural color. (b) Extraatmospheric view. (c) Vegetation analysis.



FIGURE 2: Comparison of angle deviation of buildings at different times.

only have some angular deviation, but in fact, there is no change in ground features. However, these deviations cannot be adjusted by simple image registration and can only be solved by using the recognition ability of the algorithm itself.

3. Land Use Classification of Remote Sensing Images Based on Improved U-Net

3.1. Technical Roadmap. Typical CNN models include AlexNet, VGG, GoogleNet, etc., all of which contain the most basic hierarchy. The development of these networks is characterized by complex networks, increasing parameters and deepening layers. However, there is no evidence that the number of layers of the network is directly proportional to the accuracy of the model, nor it is directly related to the image, region, and classification requirements. Therefore, blindly deepening the network is not desirable. Moreover, the deep layers of the network have high requirements for hardware, which directly affects the efficiency. It is the best situation to achieve a certain balance.

However, these networks are suitable for image recognition. During the test, a picture is input into the network and then a category label value is output, which cannot classify each pixel. The goal of this paper is to classify remote sensing images at pixel level.

U-Net is the optimization of FCN, and the extraction effect is significantly improved compared with FCN. Both U-Net and FCN have encoder-decoder structures, which are simple but effective. The encoder is combined with the

residual module to share the network parameters and avoid the degradation of the deep network. The dense connection module is used to cascade the upper features with the deep features, which is conducive to extracting new features and improving the reuse rate of feature information. Therefore, the improved U-Net is used to classify land use. This paper first preprocesses the remote sensing image data and then classifies the study area by deep neural network, including sample making, model training, prediction, and classification, as shown in Figure 3.

3.2. Improved U-Net Model

3.2.1. Basic U-Net Model. The most noticeable feature of U-Net is the integration of low-dimensional features and high-dimensional features, making full use of the semantic features of images. Figure 4 shows the schematic diagram of U-Net structure. The U-Net structure is symmetrical left and right, and the images are input into the U-Net structure. First, the high-dimensional feature map with low resolution is obtained after several convolution down-sampling operations through the encoder, i.e., compression channel on the left side of the network. The network structure on the left side is a Gaussian feature pyramid structure from low dimension to high dimension; then, enter the decoder on the right side of the network, i.e., the expansion channel, and the input feature maps are subjected to a series of deconvolution upsampling operations to generate the feature maps of the corresponding size with the original pyramid step by step. Finally, a prediction result map at the same pixel level as the input image is output. The biggest difference between U-Net and FCN structure is that when decoding, the high-dimensional features of this layer are fused with the low-dimensional features in corresponding pyramid layer for upsampling, which considers both the high-dimensional and low-dimensional features in the image.

3.2.2. Improved U-Net Model. The proposed model consists of an encoder, decoder, and connection block. The encoder is composed of 8 residual modules, 2 residual modules constitute a network layer, and each residual module is composed of two 3×3 convolution layer and quick

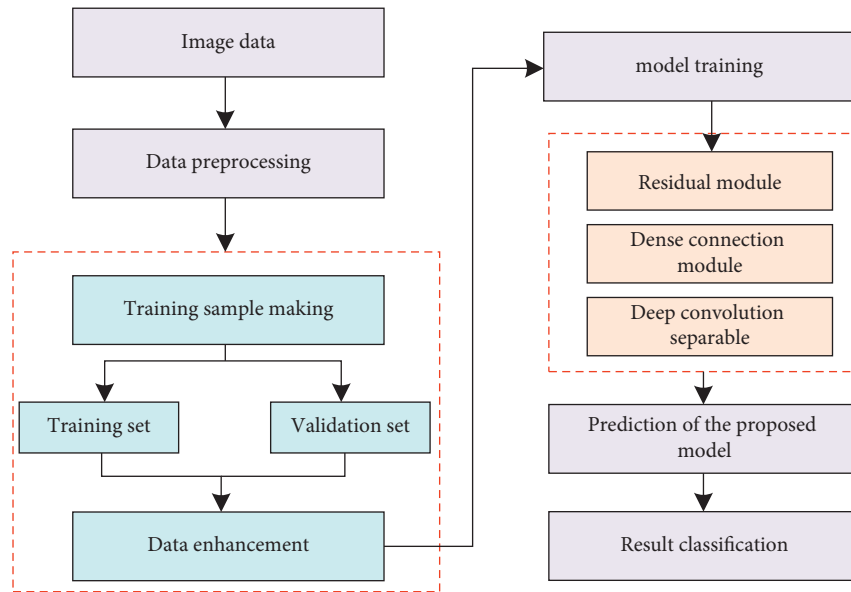


FIGURE 3: Proposed technology roadmap.

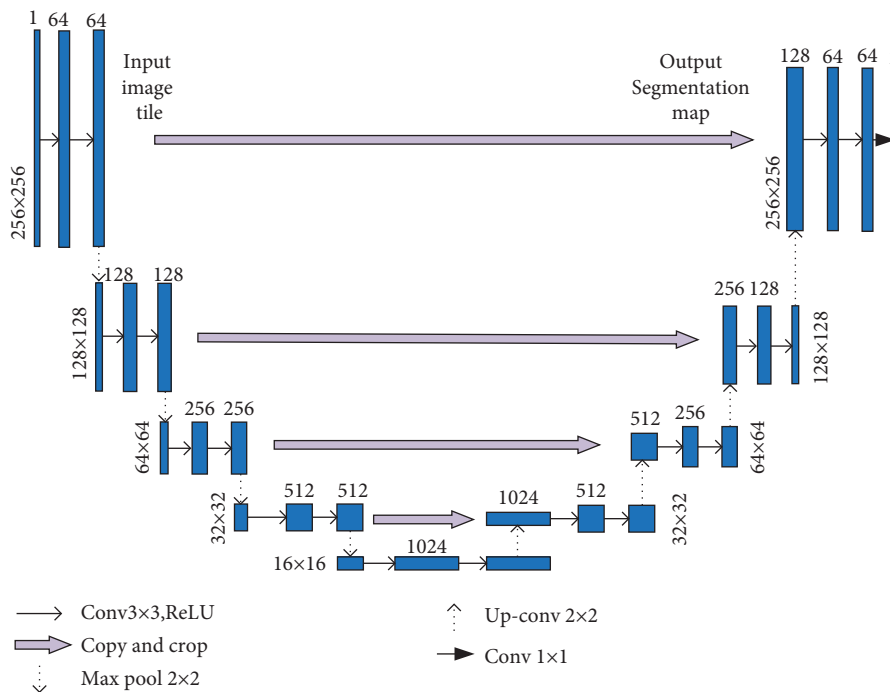


FIGURE 4: Schematic diagram of U-Net structure.

connection. Each network layer of the encoder is connected through the convolutional and max-pooling layers with step of 2 and convolutional kernel size of 2×2 . The image information is convoluted to extract local features, and the quick connection operation of the residual module fuses the input global information with local features, so that the model can capture more abundant feature information, and the network is not easy to degrade. Each layer of the decoder consists of a dense connection module, which uses layer to layer connection by transpose convolution with step of 2 and convolutional kernel size of 3×3 . The dense connection

module consists of four 3×3 convolution layers. In addition, the output of the residual module is combined with the relatively symmetrical dense connection module by cascading, so that the model integrates more shallow features. The connecting block consists of four 3×3 convolution layers. The encoder output is connected to the decoder. In order to obtain a lighter network model, less convolution kernels than U-Net are used in each layer, and batch normalization and ReLU activation functions are added after each convolution layer to prevent overfitting. The model structure is shown in Figure 5.

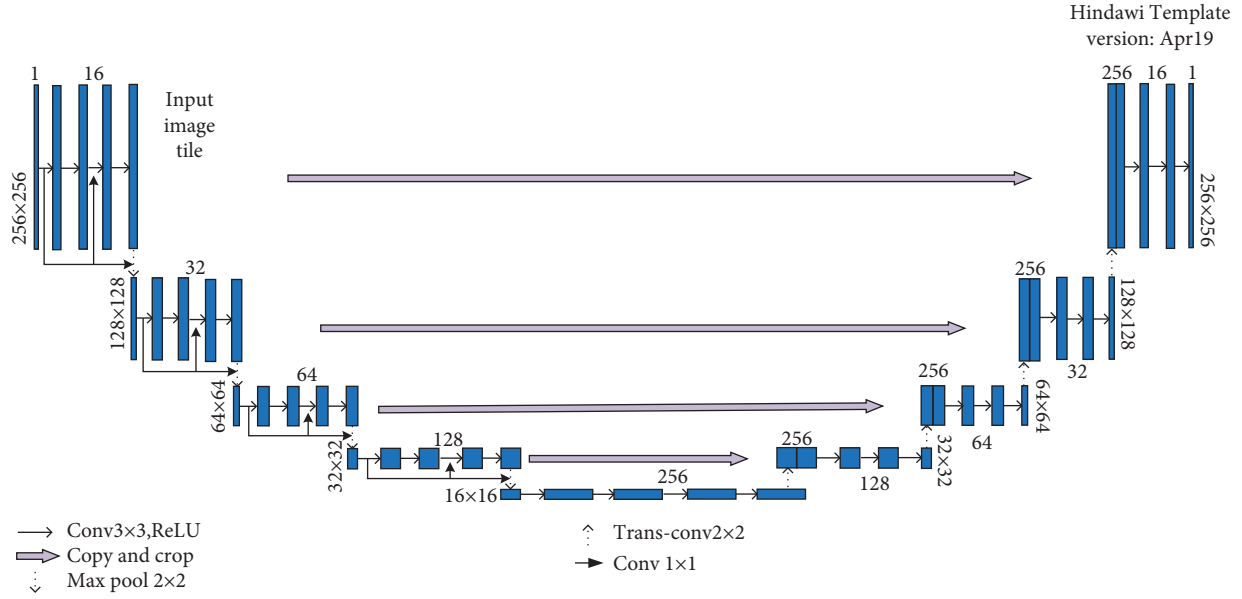


FIGURE 5: Improved U-Net model.

3.2.3. Depthwise Separable Convolution. The general convolution is replaced by depthwise separable convolution to reduce the amount of U-Net parameters. In Xception structure, the convolution kernel is three-dimensional, including width, height, and channel dimensions. The traditional convolution process is the unified processing of spatial and channel information, and the depthwise separable convolution processes the spatial and channel separately to remove the direct correlation between them. In this way, the redundant processing process is removed, the model parameters are reduced accordingly, the network becomes simpler, and the universality is strengthened. The operation of depth separable convolution includes first performing spatial convolution in the depth direction (convolution on each input channel, respectively), followed by a pointwise convolution mixing of the output channels together. The pointwise convolution is with convolutional kernel size of 1×1 , which is applied to the second step of depth separable convolution to expand the depth of the image.

Supposing that the size of convolution kernel is $D_k \times D_k$, the size of input data is $D_f \times D_f \times M$ and the size of output data is $D_f \times D_f \times N$, and the parameter quantity of general convolution is

$$D_k \times D_k \times M \times N \times D_f \times D_f. \quad (1)$$

The depth separable convolution parameters are obtained by adding the parameter quantities of pointwise convolution and depth separable convolution:

$$D_k \times D_k \times M \times D_f \times D_f + M \times N \times D_f \times D_f. \quad (2)$$

Using the property that depthwise separable convolution can reduce the parameter quantity, some traditional convolutions are replaced by depthwise separable convolution on the basis of U-Net. The parameter quantity is reduced to about 1/3 of the original parameter quantity, and the model inference time is about 5/6 of the original time.

3.3. Loss Function Design. The fundamental principle of the design criterion of loss function is to directly reflect the network model according to the characteristics of function. In the field of deep learning, the loss functions commonly used include Euclidean loss function, hinge loss function, softmax cross-entropy, and contrastive loss function. In order to be applicable to the proposed improved U-Net structure as much as possible, log loss function is used as the loss function in the experiment. The log loss function is the loss function corresponding to sigmoid, and its formula is

$$L(Y, P(Y|X)) = -\log P(Y|X), \quad (3)$$

where L is the value of cross-entropy, X is the value of sample data, and Y is the value of predicted data. In order to facilitate network calculation, the log loss function is mainly used for maximum likelihood estimation. Because the derivation of maximum likelihood is very cumbersome, the logarithm is first calculated, and then the derivative and extreme points are calculated. As the name suggests, the loss function is the sum of the losses of each type of sample. If the result is negative, it is the minimum loss of maximum likelihood estimation.

4. Experiment and Analysis

4.1. Experimental Setup and Evaluation Index. The computer system of the experimental environment is Windows 10, and the programming language used is Python. The deep learning framework used in the experiment is TensorFlow, and the high-level neural network API-Keras is used in the construction of deep learning network (graphics card: NVIDIA GeForce GTX 1060 6 GB; memory: 16GB).

In this paper, pixel accuracy (PA), recall (RC), precision (PR), and mean intersection over union (MIoU) are used as evaluation indexes. PA is the most commonly used evaluation index, which indicates the proportion of correctly predicted pixels in all pixels. MIoU calculates the proportion between the intersection and union of two sets. In the remote sensing image change detection task, these two sets represent the changed region and the unchanged region. The formulas are shown in (4) and (5):

$$PA = \frac{\sum_{i=0}^k P_{ij}}{\sum_{i=0}^k \sum_{j=0}^k P_{ij}}, \quad (4)$$

$$MIoU = \frac{1}{K+1} \sum_{i=0}^k \frac{P_{ij}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}}, \quad (5)$$

where k represents the number of object categories; p_{ij} represents the number of pixels that belong to the category i but are recognized as the category j ; and p_{ii} indicates the number of pixels correctly identified.

Recall represents the proportion of the change area correctly recognized by the algorithm in the change area of the original image, and precision represents the proportion of the number of pixels in the correct change area predicted in the prediction map to the number of pixels in all the real reference change areas. The calculation methods of these two indicators are shown in formulas (6) and (7).

$$RC = \frac{TP}{TP + FN}, \quad (6)$$

$$PR = \frac{TP}{TP + FP}, \quad (7)$$

where TP represents the pixel marked as the change region in the reference image, and the recognition result of the algorithm for the region is also the pixel of the change region; FP represents a pixel marked as an unchanged region in the reference image, and the recognition result of the region by the algorithm is a changed region; FN represents a region marked as a change region in the reference image, and the recognition result of the region by the algorithm is an unchanged region; and TN represents the unchanged region marked in the reference image, and the recognition result of the region by the algorithm is also the unchanged region.

4.2. Loss Curve and Accuracy of Network Training Process.

The loss function is used to evaluate the difference between the predicted value and the real value of the model. The better the loss function, the better the performance of the model. Figure 6 shows the loss curve and accuracy diagram of the experimental training process of the proposed method, including the loss curve and accuracy diagram of the training set and the verification set. When the epoch is 20, the loss value and accuracy are stable, Train_loss and Val_loss have converged, and the difference between them is very small. Finally, Train_loss is stable at about 85, Val_loss is stable at about 82, and accuracy is stable at about 0.74.

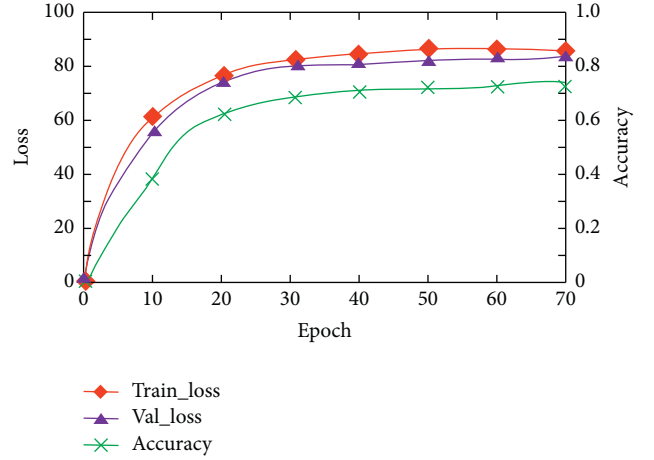


FIGURE 6: Loss curve and accuracy during training.

4.3. *Comparative Experiments of Different Models.* In order to prove the performance of the proposed algorithm, the methods in [26, 27] are compared with the proposed method under the same experimental conditions. The experimental results are shown in Figures 7 and 8.

By observing the experimental results, the PA of the method in [26] is only 91.24%, the RC is 65.33%, and the PR is 72.76%. The PA of the method in [27] is 91.87%, the RC is 71.24%, and the PR is 92.35%. The PA, RC, and PR of the proposed method are 92.35%, 80.56%, and 83.45%, respectively. Compared with other methods, it can be seen that the proposed method is the only model with more than 80% in PA, RC, PR, and MIoU indexes. MIoU is 86.75%, 10.16% higher than that in reference [26] and 4.30% higher than that in reference [27]. This is because the proposed method combines the residual module with the encoder to share the network parameters and avoid the deep network degradation. In addition, by combining the output of the residual module with the relatively symmetrical dense connection module in a cascade way, the model integrates more shallow features and improves the ability of the model to mine data features. The core of the comparison method is to optimize the model parameters, which cannot extract the deeper information from the remote sensing image data. Therefore, when dealing with the remote sensing image classification task, the values of each index are lower than those of the proposed method.

Figure 9 shows the comparison diagram of different methods for land change detection and plant change detection. Because the two models in [26, 27] cannot fully and effectively integrate the information of different depths, the edge information noise shown on the prediction map is serious. Although the approximate change region can be extracted, there is an obvious gap compared with the proposed method. The proposed improved U-Net model is superior to the other two models in terms of edge information and recognition integrity. This is mainly due to the residual module and dense connection module in the model, which improves the ability of the model to mine data features. In addition, the depth separable convolution is used to process the spatial and channel information of the

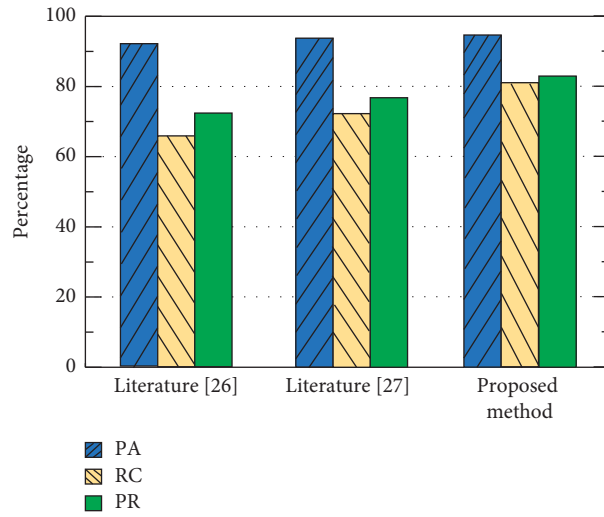


FIGURE 7: Comparison of PA, RC, and PR indexes of different methods.

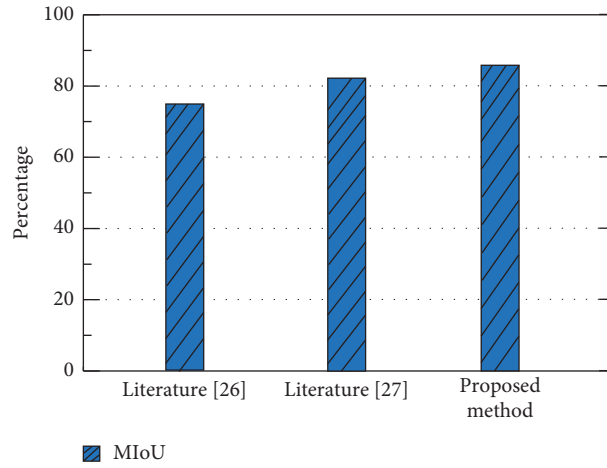


FIGURE 8: Comparison of MIoU indexes of different methods.

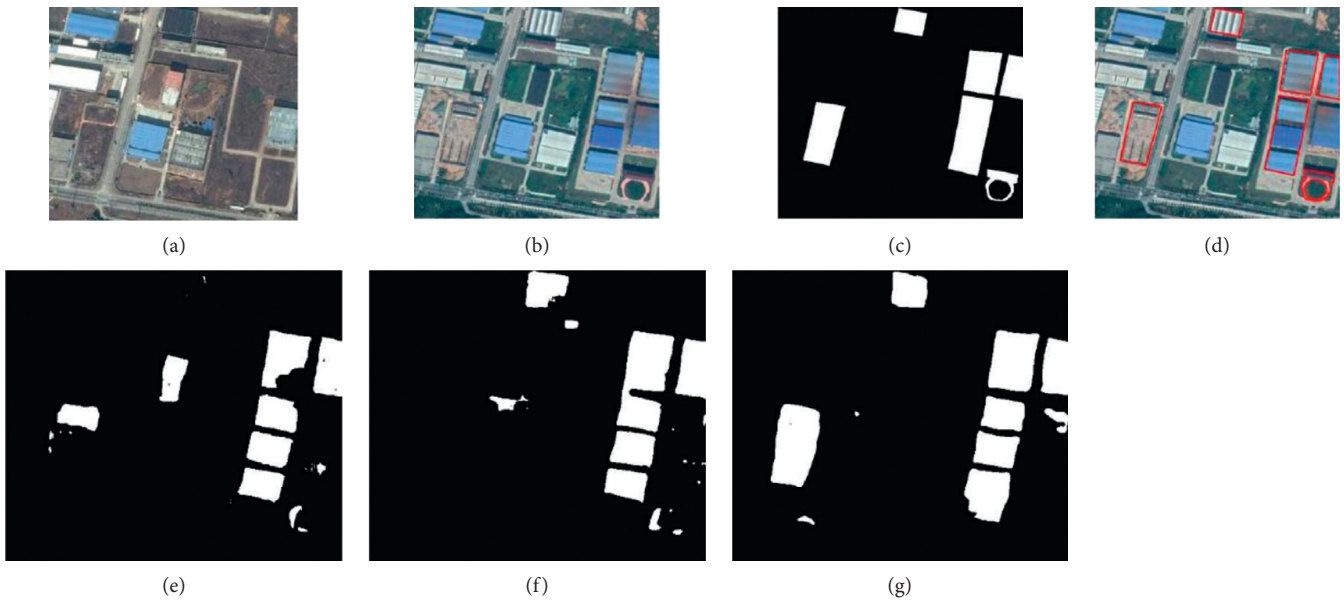


FIGURE 9: Comparison of different methods for change detection. (a) T1 time image. (b) T2 time image. (c) Reference change area. (d) Change area range. (e) Literature [26]. (f) Literature [27]. (g) Proposed method.

convolution process separately to remove the direct correlation between them, which can provide relatively independent shallow information for fusion with the main module while completing their respective tasks. Therefore, experiments show that the proposed method is feasible and efficient for land use classification in remote sensing images.

5. Conclusion

In view of the low accuracy and efficiency of existing land use classification methods for high-resolution remote sensing image segmentation, a land use classification method using improved U-Net in remote sensing images of urban and rural planning monitoring is proposed. By combining the encoder and residual module of U-Net model and integrating the dense connection module into the decoder, the data mining ability of the model is improved, and the depthwise separable convolution is used to process the spatial and channel information of the convolution process, respectively, so as to reduce the model parameters. Compared with other methods, it can be seen that the proposed method is the only model with more than 80% in PA, RC, PR, and MIoU indexes. The classification system in this paper is not detailed enough and does not involve secondary classes. In the future, a variety of data types can be integrated for more detailed classification. In addition, how to use multiple GPUs to train the network model at the same time is also one of the focuses of future research.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Research on rural planning in Northwest China from the perspective of smart contraction: a case study of Gansu Province) (no. 51968037).

References

- [1] K. Jia, X. Wei, X. Gu, Y. Yao, X. Xie, and B. Li, "Land cover classification using Landsat 8 operational land imager data in Beijing, China," *Geocarto International*, vol. 29, no. 8, pp. 941–951, 2014.
- [2] Y. Jia, H. T. Li, H. Y. Gu, and Y. S. Han, "Study on the technology and method of land cover classification for geographic national conditions surveying," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, no. 2, pp. 61–66, 2013.
- [3] S. Paul, J. Li, and R. Wheate, "Application of object oriented image classification and Markov chain modeling for land use and land cover change analysis," *Journal of Environmental Informatics*, vol. 31, no. 1, pp. 30–40, 2018.
- [4] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, no. 5, pp. 277–293, 2017.
- [5] K. M. Bergen, D. G. Brown, J. F. Rutherford, and E. J. Gustafson, "Change detection with heterogeneous data using ecoregional stratification, statistical summaries and a land allocation algorithm," *Remote Sensing of Environment*, vol. 97, no. 4, pp. 434–446, 2005.
- [6] T.-A. Teo and C.-H. Huang, "Object-based land cover classification using airborne lidar and different spectral images-Terrestrial," *Atmospheric and Oceanic Sciences*, vol. 27, no. 4, pp. 491–498, 2016.
- [7] D. Phiri and J. Morgenroth, "Developments in Landsat land cover classification methods: a review," *Remote Sensing*, vol. 9, no. 9, pp. 967–976, 2017.
- [8] D. Yemshanov, D. W. McKenney, and J. H. Pedlar, "Mapping forest composition from the Canadian National Forest Inventory and land cover classification maps," *Environmental Monitoring and Assessment*, vol. 184, no. 8, pp. 4655–4669, 2012.
- [9] H. Yang, S. Li, J. Chen, X. Zhang, and S. Xu, "The standardization and harmonization of land cover classification systems towards harmonized datasets: a review," *ISPRS International Journal of Geo-Information*, vol. 6, no. 5, pp. 154–160, 2017.
- [10] R. Khatami, G. Mountrakis, and S. V. Stehman, "Predicting individual pixel error in remote sensing soft classification," *Remote Sensing of Environment*, vol. 199, no. 7, pp. 401–414, 2017.
- [11] Y. Wang, M. Y. He, J. Xiang, and Z. M. Zhou, "Review of land cover classification based on remote sensing data," in *Proceedings of the 3rd International Conference. Wireless Community. Sensor Network*, pp. 69–78, Wuhan, China, December 2016.
- [12] M. S. Regasa, M. Nones, and D. Adeba, "A review on land use and land cover change in Ethiopian basins," *Land*, vol. 10, no. 6, pp. 585–593, 2021.
- [13] R. Khatami, J. Southworth, C. Muir et al., "Operational large-area land-cover mapping: an Ethiopia case study," *Remote Sensing*, vol. 12, no. 6, pp. 954–964, 2020.
- [14] R. Lucas, N. Mueller, A. Siggins et al., "Land cover mapping using digital earth Australia," *Data*, vol. 4, no. 4, pp. 143–156, 2019.
- [15] L. Peng, X. Wang, and T. Chen, "Multifunctional land use value mapping and space type classification: a case study of Puge County, China," *Natural Resource Modeling*, vol. 32, no. 4, pp. 2012–2020, 2019.
- [16] B. Stelmach-Fita, "European land use spatial data sources and their role in integrated planning: opportunities and challenges for Poland," *Land*, vol. 10, no. 11, pp. 1138–1148, 2021.
- [17] K. Tatsumi, Y. Yamashiki, M. A. Canales Torres, and C. L. R. Taibe, "Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data," *Computers and Electronics in Agriculture*, vol. 115, no. 14, pp. 171–179, 2015.
- [18] W. Chen, X. Li, H. He, and L. Wang, "A review of fine-scale land use and land cover classification in open-pit mining areas by remote sensing techniques," *Remote Sensing*, vol. 10, no. 1, pp. 15–24, 2018.
- [19] W. Y. Yan, A. Shaker, and N. El-Ashmawy, "Urban land cover classification using airborne LiDAR data: a review," *Remote Sensing of Environment*, vol. 158, no. 3, pp. 295–310, 2015.

- [20] M. M. Mishaa, A. D. Andrushia, and T. M. Neebha, "Image based land cover classification for remote sensing applications-A review," in *Proceedings of the 2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, pp. 152–155, Coimbatore, India, May 2021.
- [21] C. Zhang, I. Sargent, X. Pan, A. Gardiner, J. Hare, and P. M. Atkinson, "VPRS-based regional decision fusion of CNN and MRF classifications for very fine resolution remotely sensed images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4507–4521, 2018.
- [22] E. Maggiori, Y. Tarabalka, and G. Charpiat, "Fully convolutional neural networks for remote sensing image classification," in *Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5071–5074, Beijing, China, July 2016.
- [23] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land-cover classification of high-resolution imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 4, pp. 549–553, 2017.
- [24] L. Mou, P. Ghamisi, and X. X. Zhu, "Unsupervised spectral-spatial feature learning via deep residual Conv-Deconv network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 391–406, 2017.
- [25] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639–3655, 2017.
- [26] R. Naushad, T. Kaur, and E. Ghaderpour, "Deep transfer learning for land use and land cover classification: a comparative study," *Sensors*, vol. 21, no. 23, pp. 8083–8089, 2021.
- [27] S. Dong, Y. Zhuang, Z. Yang, P. Long, C. He, and L. Teng, "Land cover classification from VHR optical remote sensing images by feature ensemble deep learning network," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1396–1400, 2019.