Sensors and Materials, Vol. 35, No. 1 (2023) 317–332
MYU Tokyo

S & M 3170

Analysis of Urban Changes in High-resolution Remote Sensing Images Based on the Improved ResNet Model

Zongxia Xu,[1,2] Kui Zhang,[1,2*] Hanmei Liang,[1] Yanyan Zeng,[1] and Zhang Xuping[3]

[1]Beijing Institute of Surveying and Mapping,
No. 60 Nanlishi, Xicheng District, Beijing 100045, China
[2]Beijing Key Laboratory of Urban Spatial Information Engineering,
No. 60 Nanlishi, Xicheng District, Beijing 100045, China
[3]Beijing University of Civil Engineering and Architecture,
No. 15, Yongyuan Road, Daxing District, Beijing 102616, China

(Received October 31, 2022; accepted January 11, 2023)

Keywords: remote sensing image, ResNet, change detection, urban change discovery

"The overall urban planning of Beijing (2016–2035)" proposed "reduced development," which is highly concerned about the existing stock and highly sensitive to development variables. Facing the demand for the rapid discovery of changes in information regarding urban land cover elements, we make full use of the existing image and vector data resources accumulated over many years to carry out research on the discovery of urban change based on deep learning. To address the problems of low accuracy and poor anti-noise ability of the existing methods for the detection of changes in remote sensing images, a method for detecting change based on an improved Residual Network (ResNet) is proposed. By introducing a channel attention module, this method can make the network focus on information from the specific area of change in an image, thereby more efficiently completing the extraction and reconstruction of the features of a specific change. The effectiveness and reliability of this method are verified using a sample set based on the Beijing No. 2 image. By this method to achieve automatic all-element change polygon extraction, the accuracy, recall, and F1 are all above 85%, which is better than other models, enabling the rapid discovery and accurate location of urban spatial changes and providing strong technical support for innovative urban spatial monitoring and modes of supervision.

1. Introduction

Detection of changes in remote sensing images can assist in the analysis of urban changes and provide good technical support for monitoring feedback on the implementation of planning and data-driven fine governance. It can be widely used to update geographic data, monitor land cover/use, environmental changes, and urban development studies and is of great significance to national decision-making, serving people's livelihoods, and national construction. Traditional extraction algorithms for information on land cover often require a large amount of manual intervention and extensive prior knowledge and have the following problems: complex artificial

*Corresponding author: e-mail: 504719347@qq.com
https://doi.org/10.18494/SAM4222

ISSN 0914-4935
https://myukk.org/

feature design and poor versatility, insufficient utilization of spectral information, large computational complexity, weak generalization ability, and low robustness of the algorithm. With improvements in computer performance and the development of technologies such as big data processing and artificial intelligence, computer vision and deep learning methods have achieved unprecedented capabilities. Deep learning can identify advanced features from data and reduce the dependence on expert knowledge in the data production process. Improving the efficiency and accuracy of data processing is a significant task. At this stage, the application of deep learning technology to remote sensing image processing has become a research hotspot, and deep learning technology improves the intelligent detection of changes in remote sensing images.

The purpose of multiscenario classification of remote sensing datasets is to carry out semantic classification of the data according to the information in the images. The key to success lies in the extraction of image features. Owing to the increasingly rich information on features in data sources, research on scene classification in remote sensing images has advanced from low-level features based on color,(1) density,(2) and transform domain texture,(3) to intermediate features and now to deep learning. With the increase in the complexity of land cover changes and the diversity of remote sensing data, new methods to detect changes and new image processing algorithms continue to emerge. Semantic segmentation methods based on deep learning are mainly divided into two categories. One is a region-based convolutional network that is a candidate region-based learning method, and the other is a segmentation algorithm that can learn end-to-end, represented by a Fully Convolutional Network (FCN) and U-Net. In 2014, Girshick *et al*.(4) proposed the Region-based Convolutional Neural Network (R-CNN) method, which is a candidate box-based target detection and segmentation algorithm. The disadvantage of these methods is that the candidate regions overlap, and a large number of redundant feature maps are repeatedly calculated. He *et al*.(5) introduced a spatial pyramid pooling layer for R-CNN, which can overcome the drawbacks of R-CNN repetitive calculations and improve the efficiency of the algorithm. The basic idea of the end-to-end training method is to directly train the classifier to perform pixel-level classifications and to use the ground truth of the sample pixel level for supervised training. In 2014, the FCN model proposed by Long *et al*.(6) was the first fully convolutional image segmentation network that could accept any size input and end-to-end training. The feature map of the product layer is upsampled to restore it to the same size as the input image so that a prediction can be generated for each pixel while retaining the spatial information in the original input image and finally pixel-by-pixel classification is performed on the upsampled feature map. Chen *et al*.(7) added a fully connected conditional random field (CRF) to the end of the FCN and proposed the Deeplab model, which achieved an accuracy of 71.6% in the PASCALVOC2012 competition. In 2017, Liu *et al*.(8) proposed RefineNet (Refinement Network), which is a multipath refinement network. Each module of the model includes a residual convolution unit and a multiresolution fusion module to perform layer-by-layer decoding, and a chain pooling operation is used to expand the receptive field to enhance the context information. Pyramid scene parsing network (PSPNet)(9) is a pyramid parsing network that aggregates the contextual information of different regions through a pyramid pooling module, thereby improving the ability to obtain global information.

Remote sensing imagery is a special type of image data, and deep learning technology has been successfully applied to the task of detecting changes in remote sensing imagery. Su *et al.*[10] used an autoencoder and a stack mapping network to generate a feature difference map and combined it with the pixel information for detecting change, which has strong robustness to noise. Li *et al.*[11] proposed a stack-type constrained Boltzmann machine for the synthetic aperture radar (SAR) difference image change detection method and used a convolutional autoencoder neural network to determine the type of change in the multitemporal SAR images. Gong *et al.*[12] performed a binary classification of differential images based on a restricted Boltzmann machine (RBM) to achieve the detection of changes in SAR images. Wang[13] used semisupervised learning and sparse coding to improve the performance of unsupervised SAR image change detection. A large number of studies have shown that this is an effective method to monitor urban expansion by using satellite remote sensing images to obtain urban land use information and reveal the dynamic changes associated with urban expansion; this method also better reflects the situation in real time and is more reliable than the statistical data analysis method. Remote sensing technology is being gradually used in urban monitoring, for which land use and land cover (LULC) analysis is the most widely used application. For example, Hu *et al.*[14] conducted biophysical composition index (BCI) spatial analysis of multi-source remote sensing images to monitor the changes in impervious surfaces in urban areas.

Convolutional neural network (CNN) models such as Visual Geometry Group (VGG),[15] Inception Network (InceptionNet),[16] and Residual Network (ResNet)[17] have been proposed, and research based on CNN models has become the mainstream trend in the field of remote sensing image classification. The ResNet semantic segmentation network has also been gradually applied to the detection of changes in remote sensing images. By deepening the network structure, the effectiveness of feature extraction is enhanced, thereby determining the accuracy of dataset classification and change detection. However, this network has problems such as poor edge target segmentation accuracy and slow convergence. In view of this, a high-resolution remote sensing image change detection method based on an improved ResNet is proposed in this paper.

2. Related Work

ResNet is designed to better train deep neural networks. This model adds the idea of residual learning to the traditional CNN and satisfactorily solves the problems of gradient disappearance and accuracy reduction when the network is deep. If the number of layers is increased as much as possible, the capability of the model at detecting changes will also be greatly improved. It was proposed by He *et al.*[17] in the Microsoft Laboratory in 2015 and won first place in target detection of the Common Objects in Context (COCO) dataset and first place in image segmentation.

It is generally believed that the deeper the network, the stronger the performance. However, in fact, the capability of a network does not always increase with increasing depth, and network degradation will occur; that is, its capability is not as good as that of a shallow network (the accuracy of the test set and the training set is decreased.), and this result is not due to overfitting,

which will result in an increase in the accuracy of the training set and a decrease in the accuracy of the test set.

Even if only an identity mapping layer is added, the experimental result is that the network is degraded, and it becomes difficult to learn identity mapping as the depth increases. Therefore, ResNet was proposed to solve this problem. The stacked layers in the ResNet are called blocks. For a block, the function that can be fitted is $F(x)$. If the expected potential mapping is $H(x)$, instead of letting $F(x)$ directly learn the potential mapping, it is better to learn the residual $H(x)-x$, that is, $F(x):=H(x)-x$, so that the original forward path becomes $F(x)+x$, using $F(x)+x$ to fit $H(x)$. He *et al*. believed that this may be easier to optimize because it is easier to make $F(x)$ learn to be 0 than to let $F(x)$ learn to be an identity mapping, which can be easily achieved through L2 regularization. In this way, for redundant blocks, identity mapping can be achieved as long as $F(x)$ approaches 0, and the performance is not reduced.

2.1 Residual calculation method

The residual structure uses a shortcut connection method, which can also be understood as a shortcut. Let the eigenmatrices be added in every other layer, let $F(x)$ and x have the same shape, and then the addition is the addition of the numbers at the same position of the eigenmatrix. The block composed of $F(x)+x$ is called a ResidualBlock. As shown in Fig. 1, multiple similar ResidualBlocks are connected in series to form the ResNet. A residual block has two paths, $F(x)$ and x. The path fitting residual of $F(x)$ is the residual path, and path x is the identity mapping, which is called a "shortcut".

2.2 Two different residuals in ResNet

The residual path is roughly divided into two types. The residual structure in Fig. 2(a) is called the BasicBlock, and the convolution calculation plus the input directly calculates the output, which is composed of two 3 × 3 convolutional layers. The residual structure in Fig. 2(b)

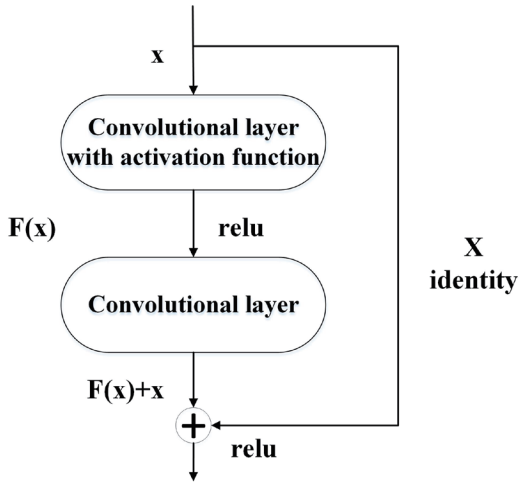


Fig. 1. Residual learning: a building block.

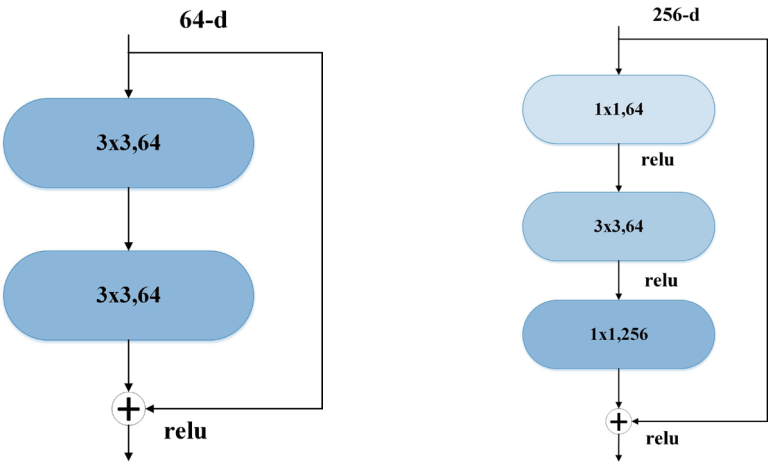


Fig. 2. (Color online) Two kinds of residual .

is called the bottleneck, and a 1 × 1 convolution is used to adjust the dimension, which is used to first reduce the dimension and then increase the dimension, mainly because of the practical consideration of reducing the computational complexity.

2.3 Shortcut for dimensionality reduction

The shortcut path can also be roughly divided into two types, depending on whether the residual path changes the number or size of the feature images. One type leaves the input x unchanged, while the other requires a 1 × 1 convolution for ascending and descending sampling. The main function of downsampling is to keep the shape of the output consistent with the output of the $F(x)$ path, which does not significantly improve the network performance.

3. Improved ResNet Method

In computer vision, a method that focuses attention on important areas of an image and discards irrelevant areas is called an attention mechanism. An attention-mechanism-based module helps improve considerably the accuracy of the detection of scene changes in remote sensing images. Because the attention mechanism is good at mining specific information from the data, the attention module and the ResNet network are integrated in the experiment, which then effectively extracts specific features from remote sensing images with complex backgrounds.

In image processing, the attention mechanism is a dynamic selection process for inputting important information to an image; this is achieved by the adaptive weighting of features. The attention mechanism has greatly improved the performance of many computer vision tasks, such as classification, object detection, semantic segmentation, and small sample detection. There are four general types of attention mechanisms: channel attention, spatial attention, temporal attention, and branch attention. In this study, two hybrid attention mechanisms were introduced:

channel and spatial attention mechanisms, i.e., the convolutional block attention model (CBAM). Given a feature map, the CBAM module can serially generate the map information of a feature of attention in the two dimensions of channel and space, and then the two types of map information about the feature are multiplied by the original input of the feature map for adaptive feature correction. The final feature map is generated. CBAM is a lightweight module that not only saves parameters and computational power but also ensures that it can be integrated into the existing network architecture as a plug-and-play module to improve performance.

3.1 Attention module

In CBAM, the attention module is divided into two parts: one is the channel attention module (CAM), and the other is the spatial attention module (SAM), as shown in Figs. 3 and 4, respectively.

CAM: The channel dimension is unchanged, and the spatial dimension is compressed. This module focuses on the meaningful information in the input image (The point of the classification task is the division of categories, and the change detection is similar to the binary classifications). The channel attention mechanism is realized through the relationship between the features. Each channel of the feature map is used as a feature detector, so the channel feature focuses on the useful information in the image.

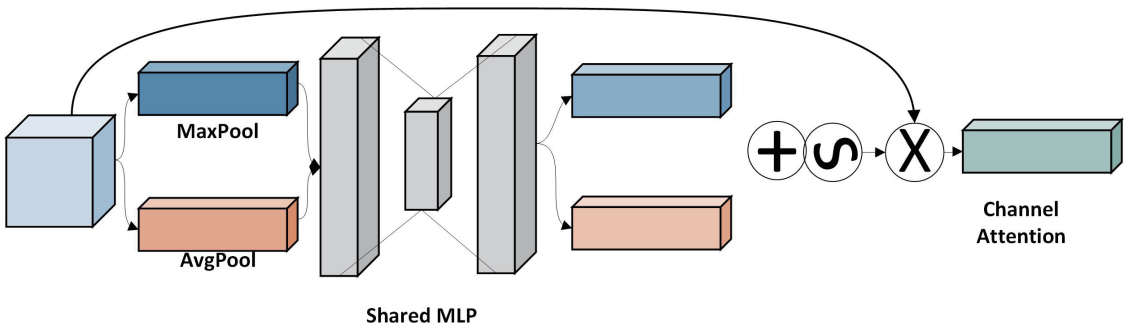


Fig. 3. (Color online) Structure of CAM.

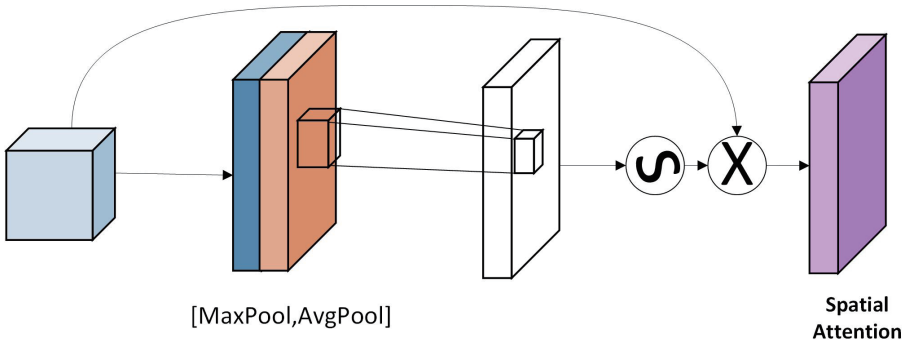


Fig. 4. (Color online) Structure of SAM.

To calculate the channel attention feature more efficiently, we need to compress the spatial dimension of the input feature map and perform the different pooling operations of average pooling and maximum pooling on the spatial content of the feature map to obtain two corresponding and different spatial context information sets that are then transferred to a shared multilayer perceptron (MLP) network to adaptively adjust the weights of different channels. Finally, the features so obtained are added and combined, and the results are processed through the normalization function to obtain the weight parameters, the feature vector is output, and then the product operation with the original feature can generate a new feature. The process can be expressed by

$$\begin{aligned} MC(x) &= \sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}(x)\right) + \mathrm{MLP}\left(\mathrm{MaxPool}(x)\right)\right) \otimes x \\ &= \sigma\left(W_1\left(W_0\left(\mathrm{Fcavg}\right)\right) + W_1\left(W_0\left(\mathrm{Fcmax}\right)\right)\right) \otimes x. \end{aligned} \tag{1}$$

Here, x refers to the input information, and W_1 and W_0 represent the weight coefficient of the shared network, which refers to the product operation of each pixel.

SAM: The spatial dimension is unchanged, and the channel dimension is compressed. This module focuses on the information regarding the location of the target. SAM is a supplement to the channel attention. Its purpose is to mine the most meaningful content information. To calculate the spatial attention, after inputting the feature map, average pooling and maximum pooling are first performed in the channel dimension, and then they are pooled. The feature maps thereby generated are then connected along the channel direction to generate an effective feature indicator. Finally, a convolution operation is used to mine the intrinsic relationship between information on the various locations, and then the normalization function is used to process the result to generate the final spatial attention feature map, which is multiplied by the input feature map to update the weight of the feature map.

3.2 Attention residual module

Through the attention module, the network can focus on the information from a specific region of the image, thereby more efficiently completing the extraction and reconstruction of specific features. Because a standard convolutional layer has a certain number of channels, each channel can adaptively learn a specific feature, and the attention module can focus on the extraction of specific features and ignore the existence of irrelevant features in all convolution operations. In this report, the attention module and the residual module are combined to form a deeper attention residual module, thereby making a new breakthrough in the ability of the network model to detect change. As shown in Fig. 5, the feature graph is first extracted by two convolution features, and the result is directly added to the input image through the dual attention mechanism module and output through the Rectifier Linear Unit (ReLU) activation function. The process can be expressed as

$$y = M_s(M_c(F(x,\{w_i\}))) + x. \tag{2}$$

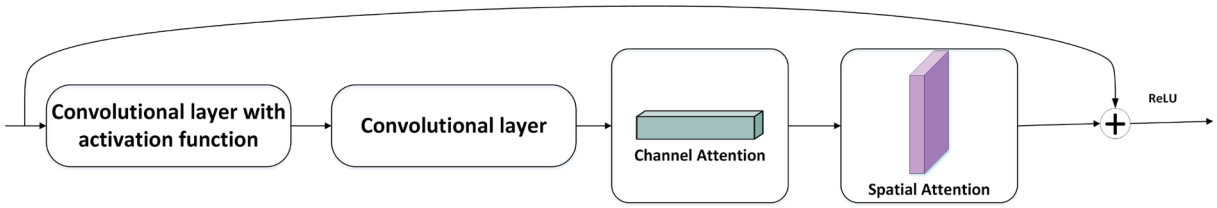


Fig. 5. (Color online) Structure of an attention residual module.

4. Experiments and Analysis

4.1 Dataset

The data used in the experiment in this paper are the remote sensing image of Beijing No. 2 and the historical labeling data (change labeling and category labeling data) within the scope of Beijing. In the experiment, two areas of the central urban area, A and B, were selected for prediction. There were more dense buildings and construction sites in A, whereas buildings in B were sparsely distributed. Region C was selected for training, and the dataset was configured according to 70% of the training set and 30% of the validation set. Figure 6 shows the distribution of the experimental area.

4.2 Experimental results and comparison

In the experiment, the size of the image input for model training is 512 × 512. The samples consisted of images of the same size before and after two periods and the corresponding change labels. Sample examples are shown in Fig. 7. The sample sets consist of 1000 pairs of images and 1000 labels. The initial learning rate is set to 0.0001, the batch size is 2, and a total of 200 trainings are performed. As the number of model training increases, the accuracy of the model increases, and the loss continues to decrease. When the model becomes stable, the network converges, and the training process ends.

After training, the data to be tested are predicted. In addition to the change detection experiment on the improved ResNet model, this study also uses three other semantic segmentation network models to perform two sets of change detection experiments on all elements in the image. The accuracy, recall, and F1 value were used as evaluation indicators. Accuracy refers to the proportion of pixels with correct predictions with respect to the total number of pixels; its formula is shown in Eq. (3). Recall refers to the ratio of the number of changed pixels in the detection results to the total number of actual changed pixels; its formula is shown in Eq. (4). The F1 value is an indicator used in statistics to measure the accuracy of a two-class model, which takes into account both the accuracy and recall of the model. The F1 value can be viewed as a weighted average of precision and recall of the model, and its formula is shown in Eq. (5).

$$precision = \frac{TP}{TP + FP} \tag{3}$$

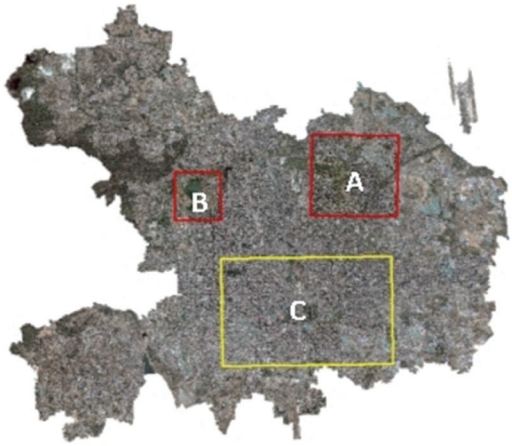


Fig. 6. (Color online) Distribution of experimental areas.

(a) (b) (c)

Fig. 7. (Color online) Sample example. (a) Early-stage image, (b) late-stage image, and (c) change label.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Here, *TP* is true positive, the prediction is positive, and actually, it is positive, which is correct. Where *FN* is false negative, the prediction is negative, and actually, it is positive, which is incorrect. Where *FP* is false positive, the prediction is positive, and actually, it is negative, which is incorrect. Where *TN* is true negative, the prediction is negative, and actually, it is negative, which is correct.

$$F1 = \frac{2PR}{P + R} \tag{5}$$

Here, *P* is the accuracy and *R* is the recall.

Two experimental areas (A and B) were selected for analysis. Two sets of experimental images were used as the test areas, and changes in the full-element detection of change for the four types of semantic segmentation networks, namely, Improved ResNet, ResNet, SegNet, and

UNet, were compared. At present, these algorithms have been well developed in the field of semantic segmentation and are widely used in the remote sensing of changes in image detection. We carried out comparative experiments.

Figures 8 and 9 show the results of the detection of change for the different semantic segmentation models: the predicted results for experiment area A are shown in Fig. 8, and the predicted results for experiment area B are shown in Fig. 9. The orange color represents the area of change. According to the results based on Unet, a large number of relatively fuzzy boundary changes are detected, and the segmentation results are relatively rough and not sensitive to details. This result arises because the method does not fully consider the relationship between pixels and lacks spatial consistency. SegNet, which uses the maximum pooling index to upsample the input feature map, has clear boundaries, but a large number of spurious changes are observed. Compared with the algorithm proposed in this study, the ResNet model misses more judgments on small areas and also indicates some pseudo-changes. The improved ResNet model combines the residual structure and the attention mechanism to make the network focus on the information in the specific region of the image, thereby making the network more

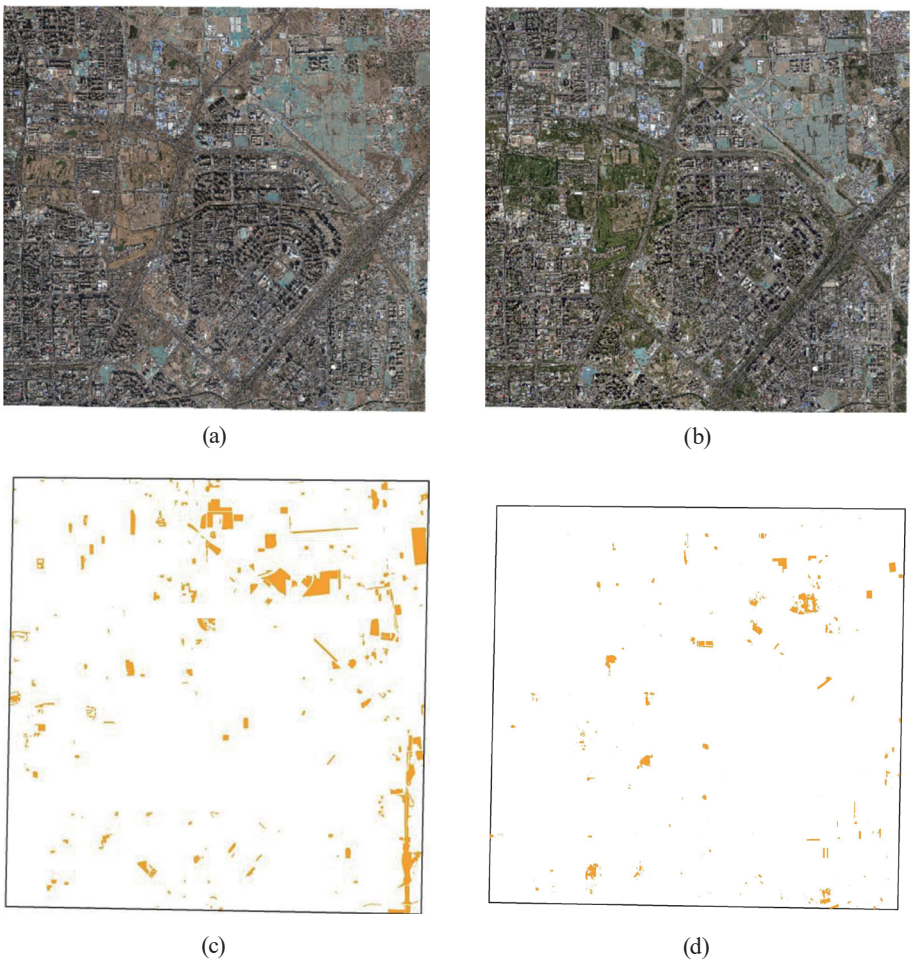
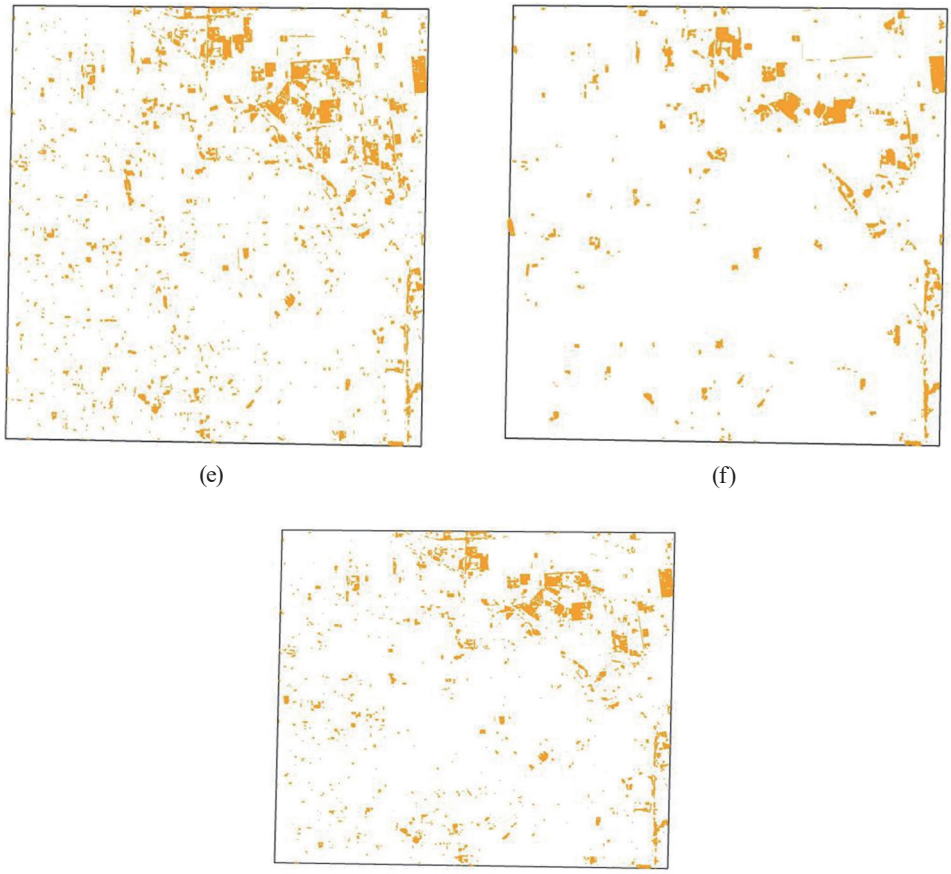


Fig. 8. (Color online) Results of detection of changes in experiment A. (a) T1 period image, (b) T2 period image, (c) true value, and (d) Unet.



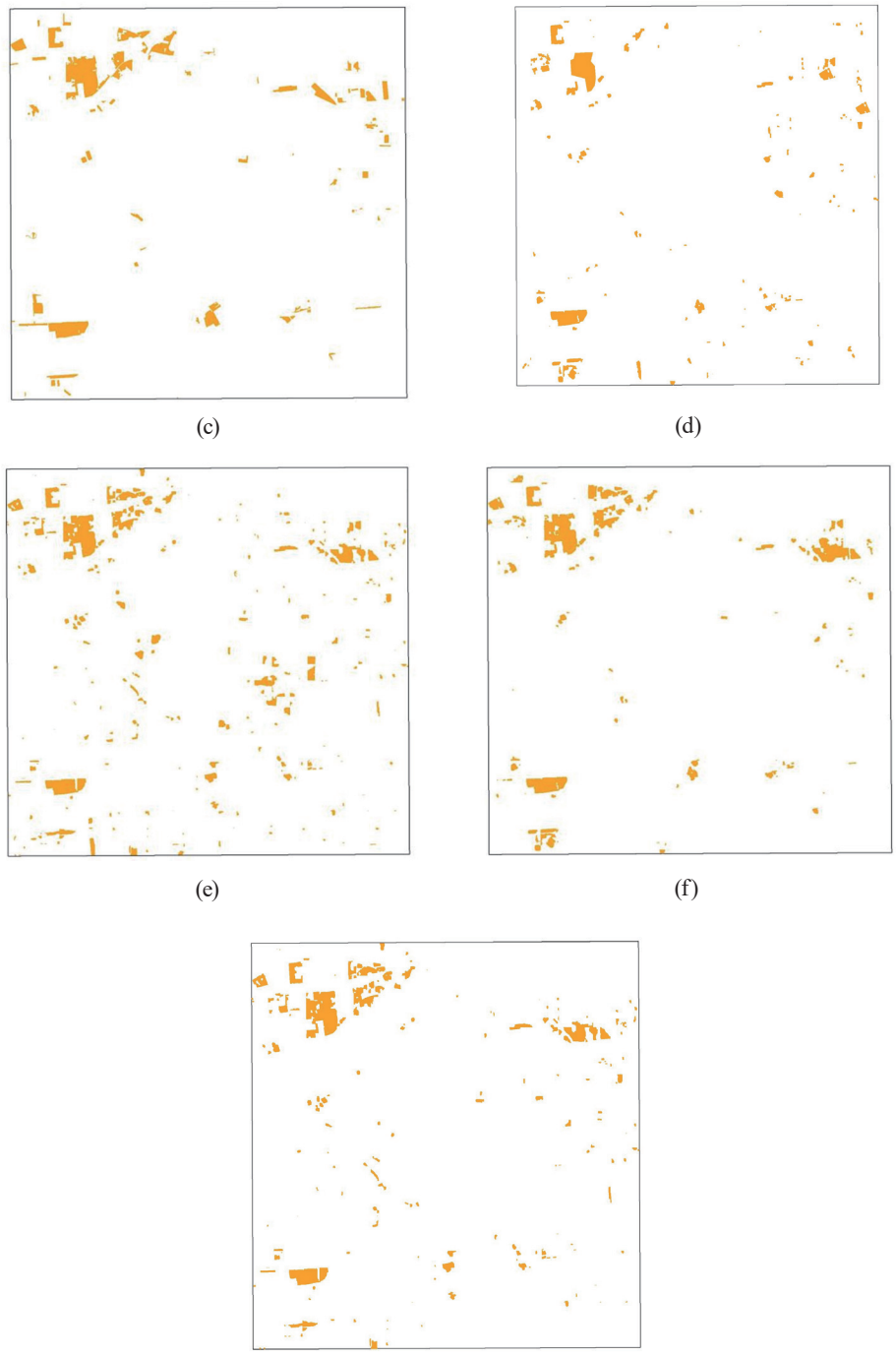
(e) (f)

(g)

Fig. 8. (Color online) (Continued) (e) SegNet, (f) Improved ResNet, and (g) ResNet.

(a) (b)

Fig. 9. (Color online) Results of detection of changes in experiment B. (a) T1 period image and (b) T2 period image.



(c) (d) (e) (f) (g)

Fig. 9. (Color online) (Continued) (c) True value, (d) Unet, (e) SegNet, (f) Improved ResNet, and (g) ResNet.

efficient. The extraction and reconstruction of specific features can be carried out to achieve better performance in the recognition of small changes, which to a large extent overcomes the effect of noise and has a better anti-interference ability with respect to "false changes," all of which have more practical effects on detection.

Tables 1 and 2 compare the accuracy of the experimental results, which can be seen from the data in the tables Because the improved ResNet model has a specific feature to enable extraction for a specific region, it can have a stronger recognition ability for small targets, and the accuracy, recall, and F1 of its detection of changes in all elements are all above 85%. Compared with other semantic segmentation models, the overall detection of change is better, and the detection is more robust. According to the test results, the data on the sparsely built-up area have a higher precision than that on the built-up area. In summary, the improved ResNet has certain advantages as well as stability in the detection of changes in all elements in remote sensing images and can be used as an effective method for the detection of changes in all elements in high-resolution remote sensing images. In addition, it can be applied to the detection of changes in areas of urban buildings, but the effect is better in sparsely built areas. The hardware environment used in this study is CPU Core (TM) 3.60 GHz, GPU NVIDIA Geforce RTX 2080 SUPER, memory 16 GB, and video memory 8 GB.

4.3 Analysis of urban changes

Through accuracy verification, the accuracy and recall of the results of automatic detection of changes based on remote sensing images can both surpass 85%, a level at which rapid and accurate extraction of information on urban spatial change to meet actual business needs can be achieved.

According to the continuous application of automatic change detection in remote sensing images, the role of changes in spots on maps is also constantly highlighted in the process of the dynamic monitoring of urban planning. According to technology for the detection of change, the update of information on urban topographic maps can be obtained in real time and presented by the changes in spots on maps. The production business has established that the changed areas on maps, when extracted, completely meet update requirements. Through the analysis of remote sensing images of the urban sub-center from 2017 to 2020, the detection rate of change exceeds 50%, which provides data in support of urban 3D update decisions. Through the analysis of the

Table 1
Comparison of the accuracy of Experiment A.

	Accuracy	Recall	F1
Unet	0.78	0.75	0.76
SegNet	0.80	0.86	0.83
ResNet	0.82	0.86	0.84
Improved ResNet	0.85	0.87	0.86

Table 2
Comparison of the accuracy of Experiment B.

	Accuracy	Recall	F1
Unet	0.81	0.80	0.80
SegNet	0.84	0.86	0.85
ResNet	0.85	0.87	0.86
Improved ResNet	0.87	0.88	0.87

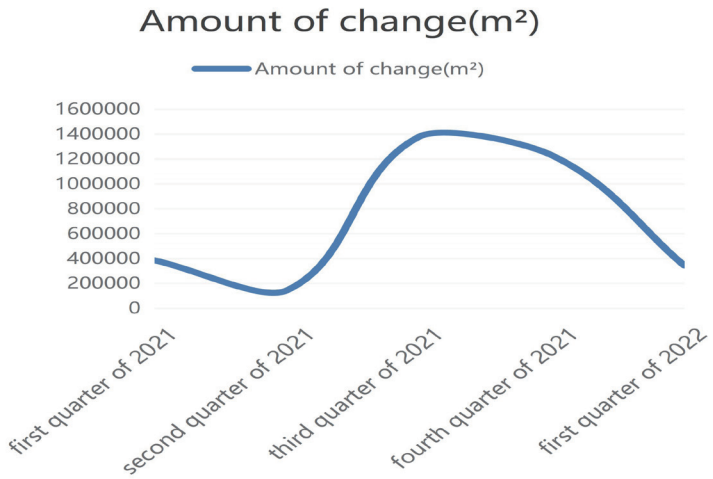


Fig. 10. (Color online) Results of quarterly change monitoring.

remote sensing images of the Chaoyang District from 2017 to 2020, the rate of change detected is approximately 10%, which can assist in the analysis of changes in the high-rise buildings in the region.

To meet the development requirements of Beijing, we dynamically monitor the changes in each quarter of the eastern and western districts. The multi-temporal remote sensing images of Beijing No. 2 are used to extract quarterly changes by automatic technology that detects such changes. The monitoring results of the past two years are shown in Fig. 10.

5. Conclusions

This report describes an algorithm for detecting change based on the deep neural network model. On the basis of the ResNet residual structure, an improved ResNet model is established. By introducing a CAM, the network can focus on the information of the specific areas in the image that change, thereby more efficiently completing the extraction and reconstruction of the specific features that change. Compared with the results from the SegNet and UNet network models, these results show that the accuracy, recall, and F1 value were better than those of other semantic segmentation models. The improved ResNet model can meet production needs in terms of the detection of changes in all elements, can obtain relatively accurate detection results, and has good accuracy and generalization in the detection of changes.

This network model can achieve large-scale, visualized dynamic detection of urban elements, providing a basis and theoretical method for urban renewal, analysis, task decision-making, and planning. The existing abundant image resources (including aerial photographs, satellite images, and others), and historical annotation data (topographic maps, land cover classification data, patches on maps, housing data, and relevant thematic data) are extracted automatically, and the module can assist operators in urban geographic data updating tasks to quickly locate the areas of change, improve the speed of operations, and reduce the workload. According to the application of the change detection results to the actual update service, a comparison between the actual update rate and the patterns detected in the changes can be obtained to assist in the

analysis of the actual update rate in more business; in the task of decision-making, the rate of change of some areas can be obtained to assist in the issuance of instructions at the decision-making level, such as providing data support for the 3D update decision of the city.

The remote sensing automatic change detection algorithm can be applied in practice. For real business scenarios in planning and natural resource management, automatic change detection algorithms can be customized to improve the utilization and utilization efficiency of high-resolution remote sensing images, reduce the threshold for using intelligent remote sensing interpretation technology, improve the dynamics and fineness of spatial monitoring, and provide strong technical support for innovative urban spatial monitoring and supervision modes.

At present, the analysis of urban change using remote sensing technology has the characteristics of rapid detection, but the actual production and application have high requirements for accuracy. The current means can only assist the rapid detection of changing areas, but the accuracy needs to be further improved. In the future, we hope to include multi-source data to improve the accuracy of the detection of change and better assist the analysis of urban changes.

Acknowledgments

This work was supported by Beijing Key Laboratory of Urban Spatial Information Engineering, No. 2020204 and task of deep-learning-based research on typical regional change detection technology and application of suspected illegal construction scenarios in Beijing.

References

1 C. Xin: Deep learning-based image classification and application research (University of Chinese Academy of Sciences Press, Beijing, 2017).
2 C. Liu, Z. X. Chen, and Y. Shao: J. Integrative Agric. **18** (2019) 506. https://doi.org/10.1016/S2095-3119(18)62016-7
3 W. X. Wang, M. F. Wang, and H. X. Li: J. Traffic Transport. Eng. (English Ed.) **6** (2019) 535. https://doi.org/10.1016/j.jtte.2019.10.001
4 R. Girshick, J. Donahue, T. Darrell, and J. Malik: 2014 IEEE Conf. Computer Vision and Pattern Recognition (2013). https://doi.org/10.1109/CVPR.2014.81
5 K. He, G. Gkioxari, P. Dollar, and R. Girshick: IEEE Trans. Pattern Anal. Mach. Intell. **42** (2020) 386. https://doi.org/10.1109/TPAMI.2018.2844175
6 J. Long, E. Shelhamer, and T. Darrell: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 3431. https://doi.org/10.1109/TPAMI.2016.2572683
7 LC. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and AL. Yuille: Proc. Int. Conf. Learning Representations (2014).
8 T. Liu, D. Yuan, H. Zhao, and J. Yin: IEEE Int. Conf. Robotics and Biomimetics (ROBIO) (2017). https://doi.org/10.1109/ROBIO.2017.8324475
9 H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia: 2017 IEEE Conf. Computer Vision and Pattern Recognition (2017). https://doi.org/10.1109/CVPR.2017.660
10 L. Z. Su: Remote sensing image change detection technology based on pixel information and deep learning (Xidian University Press, Xi'an, 2016).
11 Y. Li: Research on SAR image change detection method based on feature learning (Xidian University Press, Xi'an, 2016).
12 M. G. Gong, L. Z. Su, H. Li, and J. Liu: Comput. Res. Dev. **53** (2016) 123. https://kns.cnki.net/kcms/detail/11.1777.tp.20151116.1516.012.html
13 S. N. Wang: SAR image target recognition and change detection based on sparse feature learning (Xidian University Press, Xi'an, 2016).

14 Y. L. Hu, Y. B. Chen, Z. H. Zheng, Z. F. Wu, J. J. Li, and Z. W. Yang: Tropical Geogr. **38** (2018) (in Chinese). https://doi.org/10.13284/j.cnki.rddl.003059
15 X. L. Wang, Z. Z. Hu, and M. C. Mu: Traffic Inf. Safety **37** (2019) 95. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=JTJS201906013&DbName=CJFQ2019
16 B. Zhao, P. Li, and M. R. Dai: Comput. Syst. Appl. **28** (2019) 228. https://doi.org/10.15888/j.cnki.csa.006937
17 K. M. He, X. Y. Zhang, and S. Q. Ren: 2016 IEEE Conf. Computer Vision and Pattern Recognition (2016) 770. https://doi.org/10.1109/CVPR.2016.90

About the Authors

Zongxia Xu received her master's degree in Geographic Information Engineering from Capital Normal University, Beijing, China, in 2019. Since 2019, she has been working at the Beijing Institute of Surveying and Mapping, and since 2021, she has been an engineer. Her research interest is remote sensing image interpretation. (xuzongxia123@163.com).

Kui Zhang received his M.S. degree in the School of Information Engineering at China University of Geosciences (Beijing) in 2021. Currently he is an assistant engineer at the Beijing Institute of Surveying and Mapping in Beijing, China. He focuses on comprehending very high spatial resolution images. His specific research interests include Object-Based Image Analysis (OBIA) theory and image information extraction by deep learning methods. (504719347@qq.com)

Hanmei Liang received her master's degree in Human Geography from Capital Normal University, Beijing, China, in 2012. She served as a college-graduate village official in Daxing District of Beijing from 2012 to 2015, and has been working at the Beijing Institute of Surveying and Mapping since July 2015. Her research interests are mapping and geographic information. (hanmei710@163.com)

Yanyan Zeng received her B.S. degree from China University of Petroleum, Shandong, in 2010, and her PhD degree from the University of the Chinese Academy of Sciences, Beijing, in 2015. Since 2015, she has been a senior engineer at the Beijing Institute of Surveying and Mapping in Beijing, China. Her research interests are GNSS data processing, new fundamental surveying, and mapping. (zengyanyan1989@163. com)

Zhang Xuping received her B.S. degree from Capital Normal University, China, in 2020. Since 2020, she has been a master's degree candidate at Beijing University of Civil Engineering and Architecture. Her research interests are in remote sensing. (915028935@qq.com)