Research on Change Detection Method of High-Resolution Remote Sensing Images Based on Subpixel Convolution

Xin Luo, Xiaoxi Li, Yuxuan Wu, Weimin Hou, Meng Wang, Yuwei Jin, and Wenbo Xu

***Abstract*—Remote sensing image change detection method plays a great role in land cover research, disaster assessment, medical diagnosis, video surveillance, and other fields, so it has attracted wide attention. Based on a small sample dataset from SZTAKI Air-Change Benchmark Set, in order to solve the problem that the deep learning network needs a large number of samples, this work first uses nongenerative sample augmentation method and generative sample augmentation method based on deep convolutional generative adversarial networks, and then, constructs a remote sensing image change detection model based on an improved DeepLabv3+ network. This model can realize end-to-end training and prediction of remote sensing image change detection with subpixel convolution. Finally, Landsat 8, Google Earth, and Onera satellite change detection datasets are used to verify the generalization performance of this network. The experimental results show that the improved network accuracy is 95.1% and the generalization performance is acceptable.**

***Index Terms*—Change detection, DeepLabv3+, deep convolutional generative adversarial networks (DCGAN), deep learning, subpixel convolution.**

I. Introduction

REMOTE sensing image change detection is a process of extracting natural or artificial change areas from two or more images in the same scene at different times through a series of methods. It has important applications in land use or cover, disaster assessment, medical diagnosis, video monitoring, and other fields. Especially when a natural disaster occurs, the technology of change detection can quickly and effectively identify the disaster area. The use of remote sensing image change detection technology after some natural disasters, such as volcanoes, earthquakes, tsunamis, and debris flows, can effectively assess disasters, rationally allocate disaster relief personnel, and quickly and effectively reduce losses caused by natural disasters. When applied to urban buildings, it can timely monitor the demolition, reconstruction, and expansion of buildings, and reduce safety risks and eliminate the phenomenon of illegal occupation of land [1]. Moreover, remote sensing change detection technology can also monitor the growth of plants, so as to reasonably allocate forestry resources. When applied to crops, it is also helpful to adjust the planting plan of crops to increase yields [2].

In recent years, satellite remote sensing technology has made a continuous progress. A lot of information contained in remote sensing images, which can clearly reflect the detailed information and spatial structure characteristics of objects, provides good conditions for processing, interpretation, and analysis. In other words, satellite remote sensing technology has stepped into the submeter-level era, and the fine change detection of ground objects has become a problem that should be studied in depth. However, more detailed information means not only more processing but also a smaller gap between changed and unchanged areas. The difficulty of obtaining a graph of change detection results increases. The reason is that the increase of spatial resolution affects the spectral resolution of images, which makes the variance among targets of the same class increase, and it is difficult to identify the changed region. Besides, obtained data may be interfered by light, humans, and sensors. The traditional change detection method needs a fine preprocessing, or it will lead to large errors [3]–[7]. Even so, the final change detection map may still contain a large number of pepper and salt noiselike false alarm points, which seriously limits its application in practice [3], [8]–[11]. Therefore, it is an urgent need to update the existing remote sensing image change detection methods.

The most widely used method in the field of remote sensing image change detection is multitemporal classification. This method first obtains a classification result map, and then, compares the classification result map in order to obtain change detection results. Traditional classification methods include minimum distance classifier (MDC) [12], maximum likelihood classification (MLC) [13], support vector machine (SVM) [14],

Manuscript received August 31, 2020; revised October 18, 2020, November 6, 2020, and December 1, 2020; accepted December 7, 2020. Date of publication December 11, 2020; date of current version January 8, 2021. This work was supported in part by the Science and Technology Program of Sichuan under Grant 2017GZ0327, in part by the Science and Technology Program of Hebei under Grant 20355901D, in part by the Science and Technology Program of Hebei under Grant 19255901D, in part by the National Defense Science and Technology Key Laboratory of Remote Sensing Information and Image Analysis Technology of China under Grant 6142A010301, and in part by the Chinese Air-Force Equipment Pre-Research Project under Grant 10305∗∗∗02 *(Corresponding authors: Xin Luo; Weimin Hou.)*

Xin Luo, Xiaoxi Li, Yuxuan Wu, Meng Wang, Yuwei Jin, and Wenbo Xu are with the School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: luoxin@uestc.edu.cn; wulala@std.uestc.edu.cn; wuyuxuan@std.uestc.edu.cn; wangmengkkk@std.uestc.edu.cn; yuwei_jin@163.com; xuwenbo@uestc.edu.cn).

Weimin Hou is with the School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China (e-mail: hwm@hebust.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3044060

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

artificial neural network (ANN) [15], decision tree (DT) [16], etc. These methods can only extract the shallow features of images. The extraction performance of deep and abstract features is not very ideal, and the adaptability to complex samples is very poor, resulting in inaccurate estimation of classifier parameters, inferior classification results, and low accuracy of change detection results after spending a lot of time and resources.

With strong learning and feature extraction capacity, deep learning can exhibit excellent performance even during the processing of satellite data [17]–[20]. In 2015, Ronneberger *et al.* [21] proposed U-Net network, which greatly improved the performance of semantic segmentation. During the period of 2015–2018, from Deeplabv1 to DeepLabv3+ model [22]–[25], the segmentation efficiency and accuracy were effectively improved by adding inception module, ResNet structure, and feature pyramids. Due to the outstanding performance of deep neural network, scholars have applied it to remote sensing image change detection, and have developed many theories and methods. The change detection methods based on results can be divide into two categories: 1) binary change map and 2) multiple change map called postclassification or direct classification (obtained by comparison classes) [26], [27]. In 2015, Zhong *et al.* [28] proposed a change detection algorithm based on the pulse coupled neural network (PCNN) and normalized moment of inertia (NMI). In 2016, Gong *et al.* [18] combined deep learning with remote sensing image change detection for the first time, and proposed a new SAR image change detection method. In 2017, Wu proposed an unsupervised change detection method based on stable feature acceleration algorithm (SURF) and SVM [19]. These methods basically use the neural network to classify the images before and after the change, and then, use difference methods to generate change detection maps. In the experiments, the images before and after the change are overlapped to train neural networks. The change detection problem is transformed into a dichotomous problem so that the change region can be obtained directly. In 2019, Mou *et al.* [29] proposed a novel recurrent convolutional neural network (ReCNN) architecture, which is trained to learn a joint spectral–spatial–temporal feature representation in a unified framework for change detection in multispectral images. Liu *et al.* [30] proposed the use of unsupervised band expansion techniques to generate artificial spectral and spatial bands to enhance the change representation and discrimination for change detection from multispectral images. As to CD in multitemporal VHR images, Saha *et al.* [31] proposed a novel unsupervised context-sensitive framework, called deep change vector analysis (DCVA), which exploits convolutional neural network (CNN) features. Su *et al.* [32] presented a land cover classification training pipeline with DeepLabv3+ to realize a change detection method that can identify land cover changes from aerial imagery. Compared with traditional methods, change detection methods based on deep learning have obvious advantages. First, its classification precision is higher, and the rate of omission and error in the obtained detection results graph is lower. Second, these kinds of methods reduce tedious steps in traditional change detection processes and make detection speed faster.

Data augmentation is often used in machine learning, especially when it comes to deep learning methods. This is owing to the fact that deep learning requires a certain amount of training and testing samples. Large-scale datasets with labels are essential for CNN-based models to achieve high recognition rates or good classification results. Training CNN models using smaller scale datasets usually results in overfitting. In particular, when using deep learning methods for change detection, a large number of images of the same region in different phases are required. The acquisition of datasets can be labor intensive, severely limiting the application of this technique in the field of change detection.

The key to solving this problem is to increase the amount of sample data, and the difficulty of obtaining change detection samples has led to a variety of data augmentation methods. The existing data augmentation methods are divided into two categories—nongenerative data augmentation and generative data augmentation.

Commonly used nongenerative data augmentation methods include the following—geometric transformations (such as rotation, translation, and scaling), Gaussian noise addition, lighting simulation, and tone stretching. Nongenerative methods can only rely on the real image to bring simple changes to the image, and the representation of image contour and detail changes is far from satisfactory. Generative data augmentation methods are a technique for generating images based on known information. In recent years, with the development of deep learning, more generative data augmentation methods have emerged. In 2014, Goodfellow *et al.* [33] proposed a generative model based on deep learning, called generative adversarial network (GAN). Later, in 2015, Radford *et al.* proposed a deep CNN (DCNN) with more layers. Deep convolution GAN (DCGAN) solves the problems of training instability, pattern collapse, and internal covariate transformation that exist in GAN [34]. Krizhevsky *et al.* [35] trained a large DCNN that can classify the 1.3 million high-resolution images of LSVRC-2010 ImageNet training set into 1000 different categories.

In this article, we use both nongenerative and DCGAN-generative methods for sample augmentation to handle the problem that deep learning networks require a large number of samples and that change detection samples are difficult to obtain. After that, a deep convolutional network-based framework for remote sensing image change detection was constructed using the DeepLabv3+ network, which is superior in current classification algorithms. At the same time, the subpixel convolution [36] is used to replace the inverse convolution layer in DeepLabv3+ to enhance the up-sampling effect so as to accomplish change detection (no other researchers have conducted research). To some extent, our work resolved the problem of insufficient samples when deep learning methods are used for change detection, and improved the accuracy of change detection.

II. Data and Research Methods

First, we carried out the sample augmentation experiment. DCGAN is used for generative data augmentation. The generated data are the sample data, and the change areas are manually marked. Then, we study a change detection algorithm based on deep convolution network. In this article, we

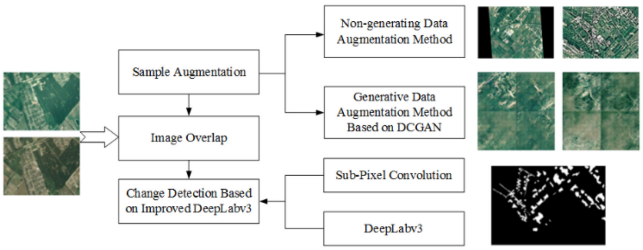

Fig. 1. Proposed change detection scheme for high-resolution remote sensing images.

Fig. 2. Examples of partial training samples from SZTAKI AirChange Benchmark Set.

exploit DeepLabv3+, which is more outstanding than classification algorithms, and improve the algorithm according to the characteristics of the generated sample dataset. Subpixel convolution is used to replace the deconvolution layer in the original DeepLabv3+ network to realize end-to-end training, and prediction of change detection. Fig. 1 is our proposed change detection scheme for high-resolution remote sensing images.

Because deep learning needs a large number of samples, we must deal with the lack of sufficient sample data before studying the change detection network. On the other hand, the final detection results may be severely affected by the differences in the images from different sensors. Therefore, our research work still uses the images from the same sensor to build the dataset and perform sample augmentation in order to meet the requirement of the deep learning change detection network for enough samples.

A. Sample Augmentation

In fact, change detection usually needs the images of the same area in different time phases. Acquiring this kind of data often consumes a lot of time and resources. If the time interval between the images of two phrases is too small, it is very difficult to detect changing areas, and the possibility of change is vague. Due to factors such as sensors, illumination, angle, etc., there are many differences between different datasets. Meanwhile, deep learning requires a large amount of sample data during the training process, and the preparation of the dataset often takes a long time [37]. Therefore, very small training samples for change detection have become an urgent problem to be solved. In summary, this work combines generative data augmentation with a deep neural network-based change detection method. The generative data augmentation method is applied to generate more change sample data, which can reduce the consumption of temporal and material resources to a certain extent.

The dataset used in this study is the SZTAKI AirChange Benchmark Set produced by DEVA Lab [38], [39]. This dataset contains 13 pairs of 3-band images, 952 $\times$ 640 in size, 1.5 m/pixel in resolution, and contains binary change masks. When marking the change area, the following differences were considered as relevant changes:

1) new built-up regions;
2) building operations;
3) planting of large group of trees;
4) fresh plough-land; and
5) groundwork before building over.

The ground truth does not contain change classification, only binary change–no change decision for each pixel. In the mask maps, the white area is the changed part and the black area is the unchanged part. Fig. 2 shows the pictures of some samples of this dataset.

B. Sample Augmentation Method

In reality, a large amount of sample data is often needed in the training process of deep neural network, otherwise it is easy to bring about network overfitting. Therefore, this article uses both the nongenerated data augmentation method and the generated data augmentation method to enlarge the quantity of change samples.

1) Nongenerative Data Augmentation Method: First, a variety of nongenerative data augmentation methods are used to expand the data scale. Each group of original images was randomly converted into different forms by means of inversion, rotation, sharpness adjustment, and Gaussian noise addition [40]. In the process of data augmentation, the images of time phase 1 and 33wq2 are processed simultaneously, and only one of them is shown here in Fig. 3.

The sample augmentation conducted by the nongenerative method can increase the volume of training data for deep learning networks and relieve the risk of model underfitting.

2) Generative Data Augmentation Method: In this study, we also adopt the method of generative augmentation based on DCGAN [34]. The basic structure of GAN network is composed of a generative model G and a discriminative model D. The input of discriminator D is the output of generator G, and it is designed to judge the probability that the samples generated by the generator G are true or false. So, the output discriminator D is binary. If the output is 0, the input image is false; if the output is 1, the input image is real. If the input to generator G is a random noise, a false sample image is generated from this noise in order to "deceive" the discriminator D.

The network structure of generator G is a deconvolutionlike neural network. Its output layer is a 100-dimensional random vector with a uniform distribution. Four fractional strided convolutions are adopted to carry out convolution operation, similar to replacing pooling layers with deconvolution, to learn

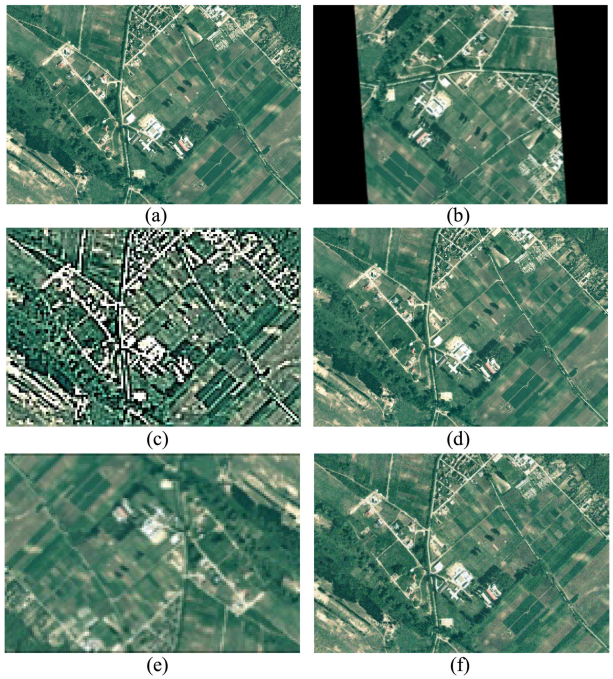

Fig. 3. Example of nongenerative data augmentation. (a) Original picture. (b) Random rotation. (c) Sharpness adjustment. (d) Left and right flip. (e) Up and down flip. (f) Random Gaussian noise addition.

the model's own spatial sampling. The network structure of discriminator D is similar to a CNN, using convolutions with steps instead of spatial pooling to perform spatial up-sampling. Like a "two-player game," generator G and discriminator D learn from each other and fight against each other. With the increase of network training times, the samples generated by generator G become more and more realistic. After thousands of iterations, the generator network converts a random noise vector from a potential space to a real sample for a dataset, and the discriminator network finds it difficult to identify whether it is true. At this time, the output of discriminator D is 0.5. Training GAN is a very intuitive process. Both generator G and discriminator D networks are trained at the same time. As the network training time increases, their performances will both improve, and the samples generated by generator G are more and more realistic. After thousands of iterations, if everything goes well, the generator network can "perfectly" generate realistic fake images. At that time, a "perfect" generator model G is built, which can be used to generate sample images for deep learning. The basic structure of DCGAN is shown in Fig. 4.

It is assumed that the distribution of training samples from real pictures is p_g and the input noise random variable of generator G is $p_z(z)$. The purpose of the generator is to turn this series of noises into an image whose distribution is within p_g. But at the beginning of training, the generator cannot directly generate the images with a distribution falling into p_g. Therefore, the distribution of the images generated by the generator model is assumed to be $p_z(z)$, and the fitting function between $p_z(z)$ and p_g is denoted as $G(z; \theta_g)$. The goal of training is to get each $p_z(z)$ close to p_{g} by learning θ_g, and for this sake, we need to introduce

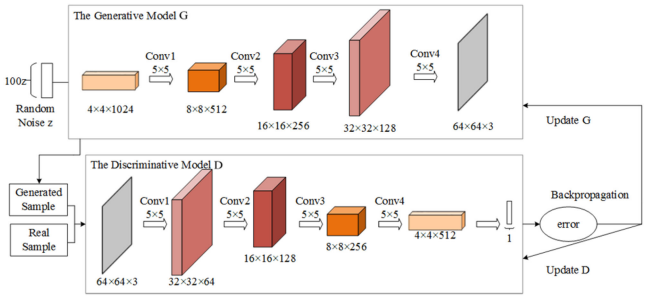

Fig. 4. DCGAN structure [23].

discriminator D. Suppose $D(x; \theta_d)$ is the probability that the output x comes from the training sample. The optimal solution of $G(z; \theta_g)$ and $D(x; \theta_d)$ is obtained by the alternating training of two networks. In this study, the images before and after the change are regarded as real samples, and DCGAN is adopted to generate new samples. The alternate training process of DCGAN is essentially the maximum and minimum optimization problem. It can be divided into two parts.

a) Training of generator network G*:* Fix discriminator network D and optimize parameters of network G. Since $D(G(\mathrm{z}))$ represents the probability that D network judges whether the image generated by G is real, maximizing the objective function $D(G(z))$ is equivalent to minimizing $1-D(G(z))$, namely, minimizing the objective function $V(D,G)$, as shown in (1). After "judging," the discriminator transfers its gradient back to G to update the parameters of the network G

$$\min_{\mathrm{G}} \mathrm{V}(\mathrm{D},\mathrm{G}) = \mathrm{E}_{\mathrm{z}-\mathrm{P}_{\mathrm{z}}(\mathrm{z})}\left[\log(1-\mathrm{D}(\mathrm{G}(\mathrm{z})))\right]. \tag{1}$$

b) Training of discriminator network D*:* Fix generator network G and optimize parameters of network D. To enable the discriminator network D to discriminate samples more effectively, we need to maximize the discriminant result of a true sample and minimize the discriminant result of a false sample, which is equivalent to minimizing $D(G(\mathrm{z}))$ and maximizing $1-D(G(z))$. Hence, during training network D, the objective function $V(D, G)$ should be maximized, as shown in (2). The parameters of network D are updated through the back propagation of errors. Fig. 5 shows a part of the training samples generated by DCGAN

$$\begin{aligned}\max_{\mathrm{D}} \mathrm{V}(\mathrm{D},\mathrm{G}) &= \mathrm{E}_{\mathrm{x}-\mathrm{P}_{\mathrm{data}}(\mathrm{x})}\left[\log \mathrm{D}(\mathrm{x})\right] \\ &+ \mathrm{E}_{\mathrm{z}-\mathrm{P}_{\mathrm{z}}(\mathrm{z})}\left[\log(1-\mathrm{D}(\mathrm{G}(\mathrm{z})))\right].\end{aligned} \tag{2}$$

C. DeepLabv3+ Network and Optimization

1) DeepLabv3+ Network Model: DeepLabv3+ model combines deep CNN with dense conditional random fields (CRFs) [25]. First, it uses atrous convolution, which has a unique advantage over ordinary convolutions in obtaining dense features of the input images. Usually, ordinary convolution operations will make the feature map smaller and smaller. When mapping the convolution results to the corresponding position in the original image, the feature response of the whole image is sparse. However, atrous convolution will not reduce the resolution of the

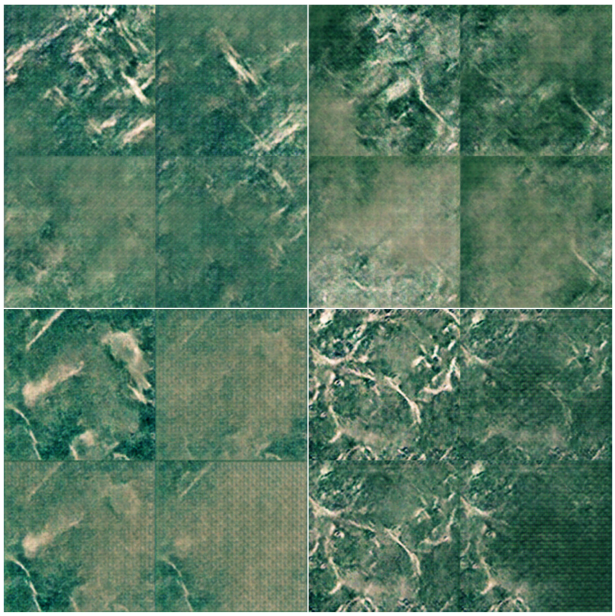

Fig. 5. Some samples generated by DCGAN.

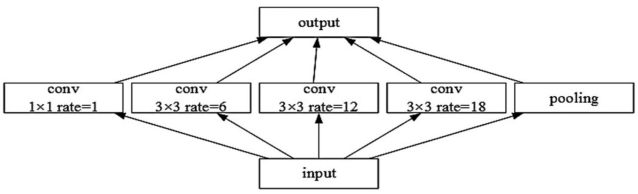

Fig. 6. ASPP structure.

feature map and not change the amount of computation, so that the generated feature map will appear denser when being located on the original image [35], [41].

Furthermore, DeepLabv3+ still uses a multiple-scale strategy to get a better classification effect, further improving atrous spatial pyramid pooling (ASPP) and suggesting a more generic framework [42], [43]. Compared to the previous DeepLab series, DeepLabv3+ adds batch normalization (BN) in its ASPP structure. By this means, even if a larger learning rate is chosen, the network can still maintain fast training and convergence speed. In addition, BN can take the place of the local response normalization layer. In order to cope with the effect of normalized input data on the learning features of the next layer in the network, two learnable parameters, scaling parameter γ and offset parameter β, are added. These two parameters enable the network to adaptively adjust the distribution of layer features, thereby making the distribution of hidden output features of the entire network more stable and enhancing the learning capability of the model. The brief diagram of ASPP in DeepLabv3+ is presented in Fig. 6. Here, the input is the output of ResNet, and the output result is finally sent to the decoder module.

The last but not the least, DeepLabv3+ network utilizes an encoder–decoder structrue [25]. In the encoder stage, the ResNet network is first used to extract the original image features. Then, the ASPP module is applieded to extract the input features of

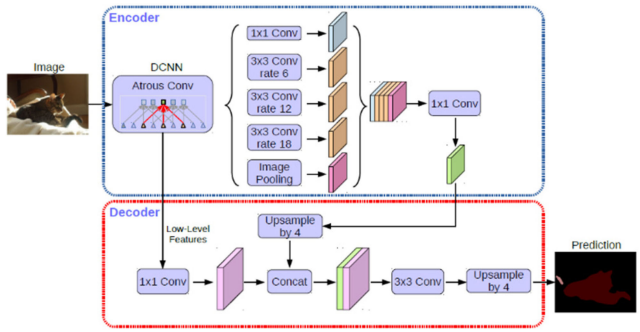

Fig. 7. DeepLabv3+ structure [28].

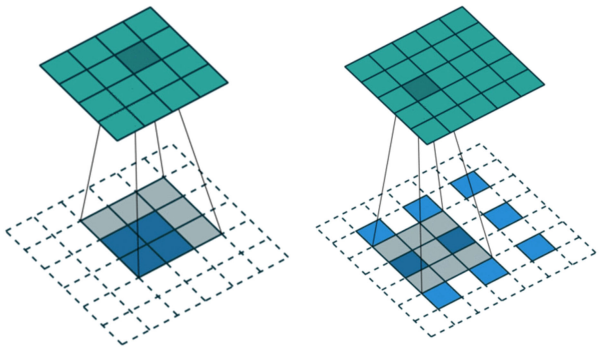

Fig. 8. Examples of deconvolution operations. (a) Filling 0 in the surrounding and gap. (b) Filling 0 at intervals.

images through filtering in multiple scales and multiple visible fields of view and perform the pooling operation. Thus, the encoded multiscale context information is obtained. In the decoder stage, the low-level features and the features processed by the ASPP module are concatenated to perform convolution again. Finally, up-sampling is executed to gradually restore spatial information to capture the finer target boundaries of images. The specific steps are as follows—the features extracted by the encoder are first up-sampled by four times, which is called the feature map F_1. The feature map F_2' of the same scale as F_1 extracted from the encoder is convolved by 1×1, and the number of channels is reduced to obtain the feature map F_2. Then, connect F_1 and F_2 to obtain F_3. Perform common 3×3 convolution on F_3 to adjust F_3 slightly, and directly up-sample by four times to get the segmentation result. Fig. 7 shows the DeepLabv3+ structure.

2) SP-DeepLabv3+ Network Model: In the DeepLabv3+ network, the image needs to be up-sampled for the sake that the details of images will be lost due to continuous convolution and pooling, and the up-sampled method of DeepLabv3+ is deconvolution. Deconvolution is also called transpose convolution. It can be seen from Fig. 8 that deconvolution is equivalent to the up-sampling operation. While the size of its output image has increased, it cannot restore the original image, since this process is mainly implemented through filling 0 in the input image. Generally, there are two filling patterns, filling 0 in the surrounding and filling 0 at intervals, as illustrated in Fig. 8.

Although deconvolution can enlarge the size of the output image, it cannot restore the information of the original image

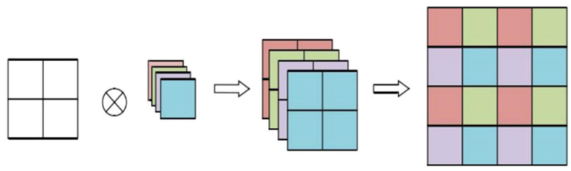

Fig. 9. Structure of subpixel convolutional layer.

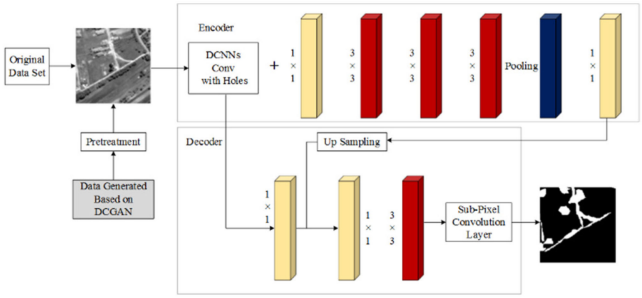

Fig. 10. Diagram of SP-DeepLabv3+ network structure.

properly. Therefore, subpixel convolution is exploited in this study to replace deconvolution [36]. It is generally believed that the smallest unit of an image is "pixel." Although at the macrolevel, there is no gap between each pixel, at the microlevel, there are still smaller units between pixels. Due to the limitations of sensor technology and photosensitive elements, these smaller units cannot be revealed directly. They are called "subpixel." The structure of the subpixel convolutional layer is shown in Fig. 9. This is a two-fold up-sampling scheme based on subpixel convolution. Set k as the magnification of up-sampling, and the input image of a single channel is convolved to obtain feature maps of the same size, but the number of feature maps changes to k^2. If the input image is of multichannel, assuming the number of channels is c, the number of channels after convolution becomes k^2c. The pixels are then rearranged, with each pixel having k^2 channels, into $k \times k$ image blocks. Finally, by connecting all the $k \times k$ image blocks of each pixel together, a higher resolution image as the output is constructed. In the DeepLabv3+ network, constant convolution and pooling operations lose some of the detailed information about the image. Although deconvolution can recover some useful information, compared to deconvolution with filling 0 for up-sampling, subpixel convolution is equivalent to expanding a pixel by a factor of k^2, since each extension point is associated with that pixel, which can maximize the image details and avoid the loss of critical information. Therefore, our research uses a subpixel convolutional layer instead of deconvolution layer in the DeepLabv3+ network, i.e., the final upsampling method is modified. Subpixel convolution is used to replace the final up-sampling layer of the network, which can introduce more nonlinear transformations and ensure the continuity of the enlarged image, thus offering better segmentation performance. Moreover, the operation of pixel rearrangement is simpler, which renders the computational complexity reduced.

The overall structure of our improved DeepLabv3+ network is demonstrated in Fig. 10, which is represented below by SP-DeepLabv3+ (subpixel DeepLabv3+).

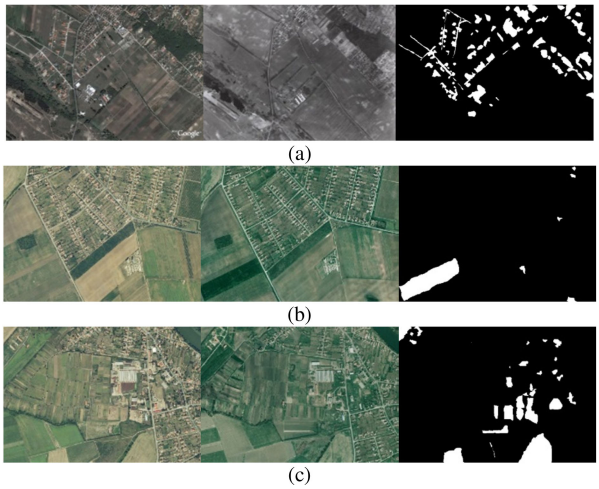

Fig. 11. Three examples of the experimental samples with manually marked areas. (a) Area 1. (b) Area 2. (c) Area 3.

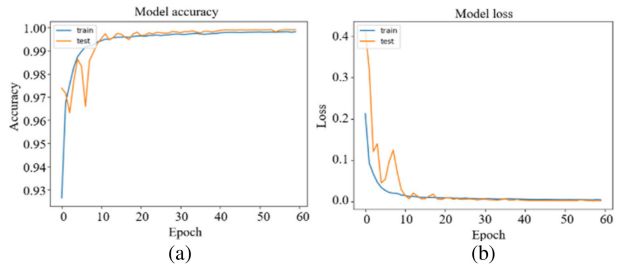

Fig. 12. Training and test result curves of the original DeepLabv3+ network. (a) Accuracy curves of network training. (b) Loss curves of network training.

III. Analysis of Experimental Results

A. Remote Sensing Image Change Detection Based on the DeepLabv3+ Network Before and After Improvement

This study uses the TensorFlow open source framework. Operating system is Win10, the computer is configured with an Intel Core i9-9900K CPU, an RTX 2080Ti graphics card, and a 64 GB RAM, and the Python version is 3.7.1. The dataset used in the experiment consists of two parts. One part is SZTAKI Airchange Benchmark Set; the total 13 groups of original images were expanded into 78 pairs by nongenerative data augmentation. Then, the images were cut into 1206 pairs of 256 × 256 size each. The other part consists of 146 pairs generated by DCGAN network, of which 114 pairs were useful. Of the sample data, 70% was randomly chosen for training, and the remaining images served as the test set. The proportion of training and testing data is 7:3. Before inputting into the network for training, the images of time phase 1 and 2 should be registered. Since the images of the two phases are all of three bands, the input image is six bands. "batch_size" represents the number of learning images in a batch, and its value is eight. The random gradient descent method was used to train the network. The training images of three groups of different regions are as given in Fig. 11.

The loss curve and accuracy curve before improvement are exhibited in Fig. 12.

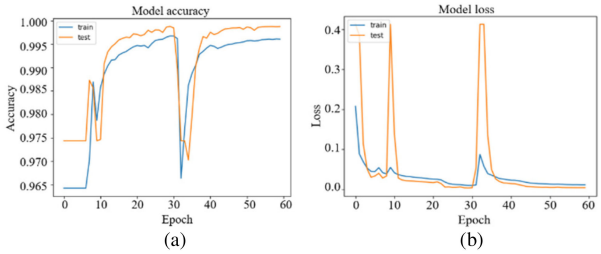

Fig. 13. Training and test result curves of the SP-DeepLabv3+ network. (a) Accuracy curves of network training. (b) Loss curves of network training.

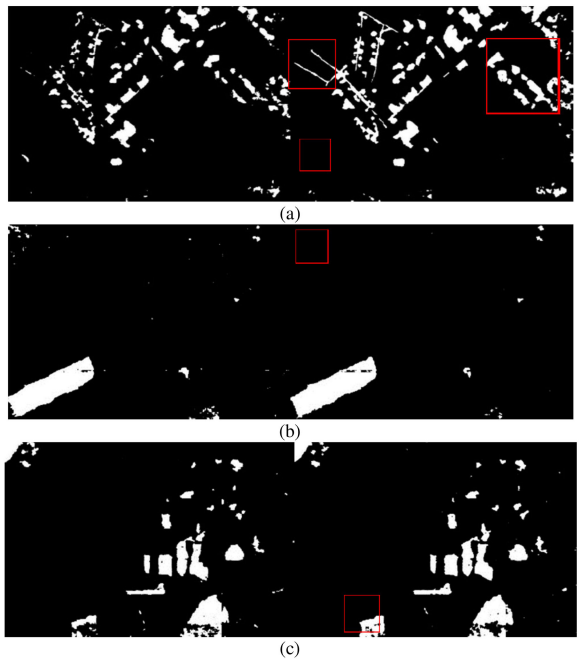

Fig. 14. Some change detection results of DeepLabv3+ (left) and SP-DeepLabv3+ (Right). (a) Area 1. (b) Area 2. (c) Area 3.

The accuracy curve and loss curves of the improved network SP-DeepLabv3+ are shown in Fig. 13. It can be found that when the epoch reached 50 times, the network began to converge, and its convergence was mostly completed after 60 times of training.

Some change detection results of DeepLabv3+ and SP-DeepLabv3+ were displayed in Fig. 14. It can be drawn from Fig. 14 that the performance of the improved network is better than that of the original one. The areas marked in red boxes are obvious improvements. In order to evaluate the improved network further, the obtained change detection maps are compared with the manually marked reference area. The result comparison of different change detection methods is listed in Table I.

Additionally, we also compare the DeepLabv3+ networks before and after improvement with two popular detection networks, U-Net [21] and STANet [44]. Some test images are given in Fig. 11, and some results are shown in Fig. 15. It can be found that the networks based on DeepLabv3+ are all superior to the U-Net network. The reason is that the performance of the original DeepLabv3+ network is excellent, and our improvement further enhances the performance of the network.

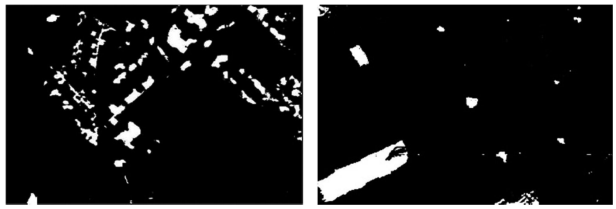

Fig. 15. Some change detection results of the U-Net network.

Moreover, we also compare the proposed method with OTSU and CVA, two traditional change detection methods [3], [8]. The comparison results are also presented in Table I. As revealed in Table I, the deep learning-based method outperforms the traditional method in all indicators, and the improved DeepLabv3+ has the highest accuracy. Although the deep learning approaches involved in this work have resulted in low omission detection rates, omission detections are still unavoidable. As indicated in Table I, the processing speed of traditional methods is obviously the fastest (all < 1 s), but their accuracy is all poor. The reason is that the traditional algorithms are relatively simple, and they are implemented through highly efficient codes differing from deep learning methods. STANet and U-Net have higher processing speed because they do not divide large-scale images into blocks of standard size like the networks based on DeepLabv3+. However, their change detection performances for our dataset are not very ideal. Specially, since the public STANet network pretrained model was primarily based on building change data, the omission rate of STANet is high. It discriminated most detection areas into the unchanged, which results in high OA and SP, and low ER. The time complexity of methods based on DeepLabv3+ is similar, and the average processing time of a 1048×724 resolution image is about 10 s. Although the time cost of SP-DeepLabv3+ network is slightly higher than that of DeepLabv3+ network, the change detection accuracy of the former is significantly enhanced.

B. Verification of Network Generalization

The verification experiment in this work was carried out on the SZTAKI AirChange Benchmark Set, whose augmented version served as training and test samples for the network. However, in order to verify whether the change detection network proposed in this work has generalization performance to some degree, we also conducted detection experiments on three other public datasets. But even so, since the deep learning network is relatively sensitive to differences among images from different sensors, the internal parameters of the trained network will vary significantly, and the comparability of experimental results is not very strong. For images from different sensors, the superior strategy is to regenerate the samples and train and test the change detection network.

1) Results for Landsat 8 Data: In order to testify the proposed DeepLabv3+ network generalization performance, another public change detection data utilized in network testing are derived from the Landsat 8 satellite images and provided on the United States National Geological Survey (USGS) website.

TABLE I
CHANGE DETECTION RESULT COMPARISON OF DIFFERENT METHODS

Evaluation Index (EI)	SP-DeepLabv3+	DeepLabv3+	U-Net[21]	STANet[44]	OSTU[3]	CVA[8]
Kappa Coefficient (Kappa)	**0.75**	0.64	0.58	0.19	0.35	0.19
Overall Accuracy (OA)	**95.1%**	93.6%	92.8%	95.1%	87.7%	84.4%
Omission Rate (OR)	**4.6%**	5.2%	3.4%	78.1%	5.0%	78.7%
Detection Error Rate (ER)	**5.6%**	18.0%	21.2%	2.5%	37.2%	5.0%
Sensitivity (SS)	**94.4%**	81.2%	68.6%	21.8%	65.9%	21.3%
Specificity (SP)	**95.2%**	94.9%	94.0%	97.4%	94.2%	95.0%
Balanced Accuracy (BA)	**94.8%**	88.1%	81.3%	59.6%	80.1%	58.1%
F1-Score (F1)	**77.8%**	69.7%	70.1%	21.5%	64.3%	24.7%
Time Cost	12.71 s	12.65 s	6.06 s	1.10 s	0.26 s	0.28 s

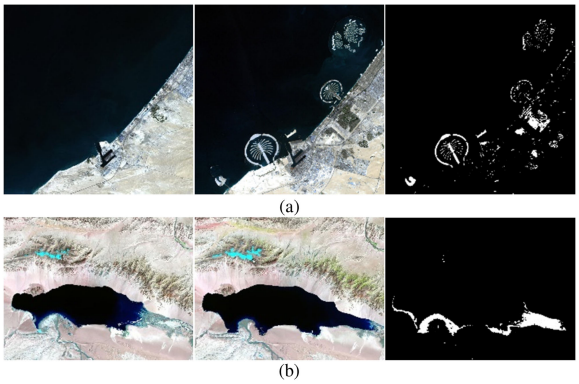

Fig. 16. Testing images of Landsat 8. (a) Area 1. (b) Area 2.

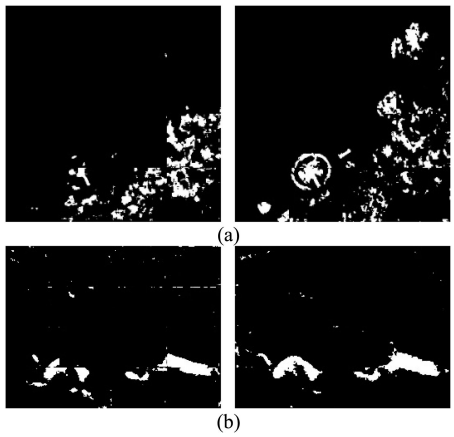

Fig. 17. Change detection results of DeepLabv3+ (left) and SP-DeepLabv3+ (right) for Landsat 8 data. (a) Area 1. (b) Area 2.

TABLE II
CHANGE DETECTION RESULT COMPARISON OF THE DEEPLABV3+-BASED METHODS ON LANDSAT 8 DATA

EI	Kappa	OA	OR	ER	SS	SP	BA	F1
DeepLabv3+	0.41	95.7%	48.6%	3.1%	51.3%	96.9%	74.1%	43.1%
SP-DeepLabv3+	0.56	96.4%	20.0%	3.1%	80.0%	96.8%	88.4%	57.7%

The data contain remote sensing images of eight areas in 30 m resolution, but the sizes of each image group are inconsistent. These images were acquired over Dubai separately in 2000 and 2012, and over Ayakkum Lake separately in 2003 and 2012. The two chosen image groups are, respectively, 1600 × 1600 and 1600 × 1206 in size. The typical regions of these two images chosen for generalization testing are shown in Fig. 16.

The original and improved DeepLabv3+ networks were, respectively, tested with two image groups, and the change detection results are shown in Fig. 17. By comparing the change detection result of the two DeepLabv3+-based networks with the manual marks, the detection indicators can be calculated, which are presented in Table II.

It can be seen that the performance of the two DeepLabv3+-based methods all decline. This phenomenon may be owing to the training samples of the two DeepLabv3+-based networks being not derived from Landsat 8 data and the difference in resolution between Landsat 8 satellite images and SZTAKI dataset. Although the detection performances of the two networks have all deteriorated, the DeepLabv3+-based algorithm still has great advantages, especially the improved DeepLabv3+ network. In the future, we can further enhance the performance of this network by enriching the samples of the training dataset.

2) Results for Google Earth Data: The second public testing data are the change images of the 2011 earthquake of the Pacific coast of Tōhoku, which are downloaded from Google Earth. The images of phase 1 were acquired on August 14, 2009, and the images of phase 2 were acquired on March 14, 2011. The spatial resolution of these images is one meter, and their size is 11008 × 6400. Also, two images of areas, an airport and a building region, are applied in generalization testing, as shown in Fig. 18. The change areas are marked by ourselves.

The original and improved DeepLabv3+ networks were, respectively, tested with two image groups, and the change detection results are shown in Fig. 19. By comparing the change detection result of the two DeepLabv3+-based networks with the manual marks, the detection accuracy indicators can be calculated, which are presented in Table III.

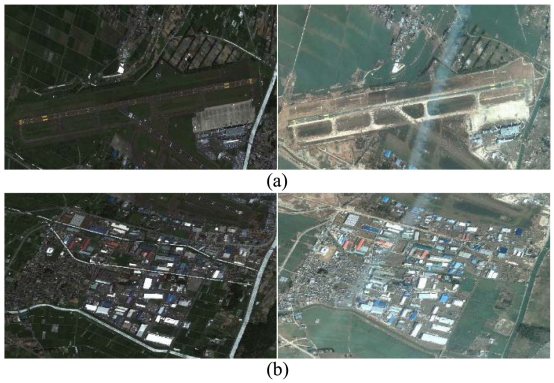

Fig. 18. Testing images of Google Earth. (a) Area 1: Airport. (b) Area 2: Buildings.

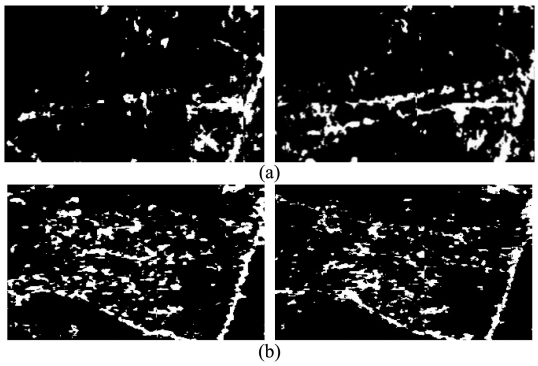

Fig. 19. Change detection results of DeepLabv3+ (left) and SP-DeepLabv3+ (right) for Google Earth data. (a) Area 1. (b) Area 2.

TABLE III
CHANGE DETECTION RESULT COMPARISON OF THE DEEPLABV3+-BASED METHODS ON GOOGLE EARTH DATA

EI	Kappa	OA	OR	ER	SS	SP	BA	F1
DeepLabv3+	0.53	88.9%	23.5%	8.3%	76.5%	91.7%	84.1%	59.0%
SP-DeepLabv3+	0.66	91.2%	18.7%	5.1%	81.3%	94.9%	88.1%	70.0%

It can be found that compared with the detection results for Lantsat 8 data, the detection accuracy for Google Earth data is better. The reason may be that there are slight resolution differences between the Google Earth data and the SZTAKI training samples. Hence, it can be inferred that the generalization performance of the network may be affected by image resolution.

3) Results for Onera Satellite Change Detection (OSCD) Data: The third public testing data selected in this article are OSCD, which were taken by the Sentinel-2A satellite. Phase 1 is of 2015, and Phase 2 is 2018. It contains 24 image pairs. The shooting area is on a global scale, including Brazil, the United States, Europe, the Middle East, and Asia. The spatial resolution of the images includes 10, 20, and 60 m. Two typical areas were chosen for generalization testing, as shown in Fig. 20.

The original and improved DeepLabv3+ networks were, respectively, tested with two image pairs, and the change detection results are shown in Fig. 21. By comparing the change detection result of the two DeepLabv3+-based networks with the manual marks, the detection accuracy indicators can be calculated, which are presented in Table IV.

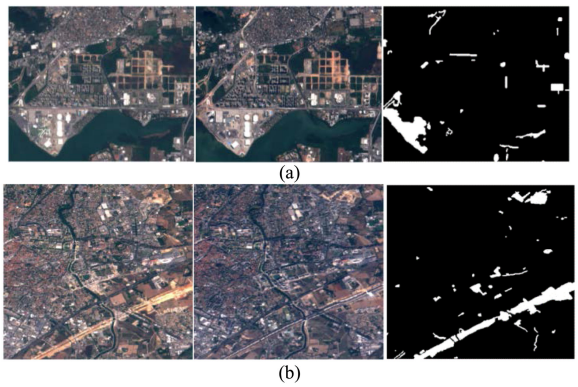

Fig. 20. Some examples of OSCD dataset. (a) Area 1. (b) Area 2.

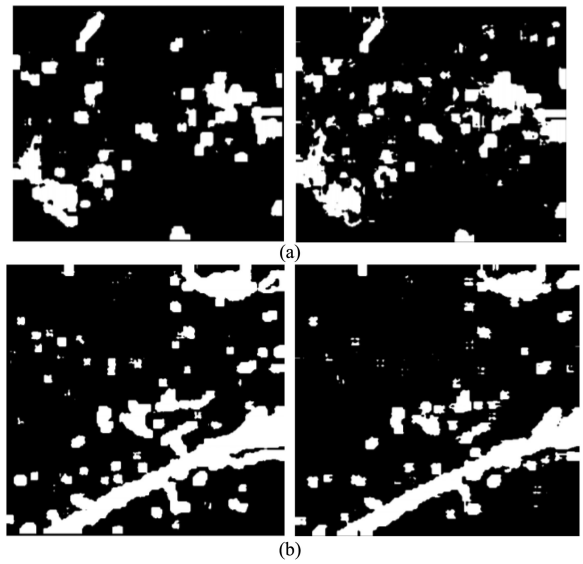

Fig. 21. Change detection results of DeepLabv3+ (left) and SP-DeepLabv3+ (right) for OSCD data. (a) Area 1. (b) Area 2.

TABLE IV
CHANGE DETECTION RESULT COMPARISON OF THE DEEPLABV3+-BASED METHODS ON OSCD DATA

EI	Kappa	OA	OR	ER	SS	SP	BA	F1
DeepLabv3+	0.39	75.2%	8.7%	35.5%	69.9%	83.7%	76.8%	39.0%
SP-DeepLabv3+	0.44	83.6%	5.3%	27.2%	71.4%	89.1%	80.3%	44.8%

It can be seen that the accuracy of the results on the OSCD dataset is the lowest, which may also be owing to the resolution of the images. Therefore, the generalization performance of the improved DeepLabv3+ network can be further ameliorated by increasing the diversity of the training dataset in future research.

IV. Conclusion

In this work, the change detection problem of high-resolution remote sensing images is converted into a biclassification problem. For the first time, the deep neural network DeepLabv3+, which is a forefront method of image classification, is applied to implement image change detection, and a more accurate detection scheme is proposed. In order to solve the problem of insufficient change detection samples in deep learning network training, this work uses data augmentation methods, including the traditional nongenerative data augmentation methods, and the generative data augmentation method DCGAN. The data generated by DCGAN network effectively supplement the sample dataset. Then, an improved DeepLabv3+ network (SP-DeepLabv3+) was adopted to realize remote sensing image change detection. The network was improved by replacing the deconvolution layer with subpixel convolutions, which enhanced the overall accuracy of the network to 95.1%. Finally, in order to verify the generalization capability of our proposed detection network, the data from Landsat 8, Google Earth, and OSCD dataset are used in network testing, and the results show that the proposed network has acceptable generalization performance. In the future, we will further study the generalization problem of deep learning networks for change detection and the problem of change detection between images from different sensors.

References

[1] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.

[2] P. Zhang *et al.*, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, Jun. 2016.

[3] J. Gao *et al.*, "Infrared image change detection of substation equipment in power system using Markov random field," in *Proc. Int. Conf. Comput. Intell. Inf. Syst.*, Nanjing, China, Apr. 2017, pp. 332–337, doi: 10.1109/ciis.2017.54.

[4] L. Jia, M. Li, P. Zhang, Y. Wu, L. An, and W. Song, "Remote-sensing image change detection with fusion of multiple wavelet kernels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3405–3418, Jan. 2016.

[5] M. N. Sumaiya and R. S. S. Kumari, "Logarithmic mean-based thresholding for SAR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1726–1728, Sep. 2016.

[6] W. Li, M. Lu, and X. Chen, "Automatic change detection of urban land-cover based on SVM classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Milan, Italy, Jul. 2015, pp. 1686–1689, doi: 10.1109/IGARSS.2015.7326111.

[7] M. H. Kesikoglu *et al.*, "The determination of coastline changes using artificial neural networks in Yamula dam lake, Turkey," in *Proc. 8th Int. Conf. Inf. Technol.*, Amman, Jordan, Aug. 2017, pp. 737–740, doi: 10.1109/ICITECH.2017.8079936.

[8] H. Zhuang, K. Deng, H. Fan, and M. Yu, "Strategies combining spectral angle mapper and change vector analysis to unsupervised change detection in multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 681–685, May 2017.

[9] J. Li and R. M. Narayanan, "A shape-based approach to change detection of lakes using time series remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2466–2477, Dec. 2003.

[10] H. Q. Qing *et al.*, "Landslide change detection based on spatio-temporal context," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Fort Worth, TX, USA, Jul. 2017, pp. 1095–1098, doi: 10.1109/IGARSS.2017.8127147.

[11] P. Lv, Y. Zhong, J. Zhao, and L. Zhang, "Unsupervised change detection based on hybrid conditional random field model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4002–4015, Jul. 2018.

[12] V. Alberga, "A study of land cover classification using polarimetric SAR parameters," *Int. J. Remote Sens.*, vol. 28, no. 17, pp. 3851–3870, Aug. 2007.

[13] O. Hagner and H. Reese, "A method for calibrated maximum likelihood classification of forest types," *Remote Sens. Environ.*, vol. 110, no. 4, pp. 438–444, Oct. 2007.

[14] X. Niu and Y. Ban, "Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 1–26, Sep. 2012.

[15] P. D. Heermann and N. Khazenie, "Classification of multispectral remote sensing data using a back-propagation neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 1, pp. 81–88, Feb. 1992.

[16] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sens. Environ.*, vol. 86, no. 4, pp. 554–565, Aug. 2003.

[17] P. J. Du *et al.*, "Review of hyperspectral remote sensing image classification," *J. Remote Sens.*, vol. 20, no. 2, pp. 236–256, Feb. 2016.

[18] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.

[19] L. Wu, B. Liu, and B. Zhao, "Unsupervised change detection of remote sensing images based on SURF and SVM," in *Proc. Int. Conf. Comput. Intell. Inf. Syst.*, Nanjing, China, Apr. 2017, pp. 214–218, doi: 10.1109/CIIS.2017.39.

[20] C. Robert, "Machine learning, a probabilistic perspective," *Chance*, vol. 27, no. 2, pp. 62–63, 2014.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, Nov. 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[22] L. C. Chen *et al.*, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *Comput. Sci.*, vol. 4, pp. 357–361, 2014, doi: 10.1080/17476938708814211.

[23] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[24] L. C. Chen *et al.*, "Rethinking atrous convolution for semantic image segmentation," *Compt. Vis. Pattern Recog. (CVPR)*, 2017, *arXiv: 1706.05587v3*.

[25] C. L. Chieh *et al.*, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, Sep. 2018, pp. 833–851. [Online]. Available: https://link.springer.com/conference/eccv

[26] H. Mahdi and S. S. Teymoor, "Hyperspectral change detection: An experimental comparative study," *Int. J. Remote Sens.*, vol. 39, no. 20, pp. 7029–7083, Apr. 2018, doi: 10.1080/01431161.2018.1466079.

[27] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.

[28] Y. Zhong, W. Liu, J. Zhao, and L. Zhang, "Change detection based on pulse-coupled neural networks and the NMI feature for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 537–541, Mar. 2015.

[29] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019.

[30] S. Liu, Q. Du, X. Tong, A. Samat, and L. Bruzzone, "Unsupervised change detection in multispectral remote sensing images via spectral-spatial band expansion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3578–3587, Aug. 2019.

[31] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.

[32] R. Su and R. Chen, "Land cover change detection via semantic segmentation," *Comput. Vis. Pattern Recog. (CVPR)*, 2019, *arXiv: 1911.12903*.

[33] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Jun. 2014, pp. 2672–2680.

[34] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *Comput. Sci.*, 2015, *arXiv: 1511.06434v2*.

[35] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Conf. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 2012, pp. 1097–1105, doi: 10.1145/3065386.
[36] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 1874–1883, doi: 10.1109/CVPR.2016.207.
[37] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, Beijing, China, Mar. 2006, pp. 1458–1465, doi: 10.1109/ICCV.2005.239.
[38] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multi-layer conditional mixed Markov model," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3416–3430, Oct. 2009.
[39] C. Benedek and T. Szirányi., "A mixed Markov model for change detection in aerial photos with large time differences," in *Proc. Int. Conf. Pattern Recognit.*, Tampa, FL, USA, Dec. 2008, pp. 1–4, doi: 10.1109/ICPR.2008.4761658.
[40] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jun. 2016, pp. 2169–2178, doi: 10.1109/CVPR.2006.68.
[41] C. Peng *et al.*, "Large kernel matters-improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Mar. 2017, pp. 4353–4361.
[42] Y. Shao *et al.*, "An automatic change detection technology for remote sensing data using Gaussian mixture model," in *Proc. IEEE 6th Int. Conf. Agro-Geoinform.*, Fairfax, VA, USA, Jun. 2017, pp. 1–4, doi: 10.1109/Agro-Geoinformatics.2017.8047063.
[43] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[44] H. Chen and Z. Shi, "A spatial-temporal attention-based method and new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1662.

Xin Luo was born in Chengdu, Sichuan, China, in 1977. She received the B.S. degree in electrical technology, the M.S. degree in power electronics and electric drive, and the Ph.D. degree in pattern recognition and intelligent systems from the Northwestern Polytechnical University, Xi'an, China, in 1999, 2004, and 2008, respectively.

From 2008 to 2012, she was a Lecturer with the School of Communication and Information Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China. Since 2012, she has been an Associate Professor with the School of Resources and Environment, UESTC. From 2014 to 2015, she was a Visiting Scholar with the Department of Mathematics, University of California, Santa Barbara, Santa Barbara, CA, USA. She is an expert of Sichuan Branch of Chinese National Remote Sensing Center. She is the author of more than 20 articles, more than ten inventions, and holds a patent. Her research interests include remote sensing image processing and application, machine learning, pattern recognition, computer vision, and 3-D modeling.

Dr. Luo is a member of the Chinese Society for Geodesy Photogrammetry and Cartography. She was a recipient of the First prize of Science and Technology Progress Award of Surveying and Mapping Geographic Information of Sichuan province in 2017, the IEEE International Conference on Knowledge Innovation and Invention Best Conference Paper Award in 2018, and the Second prize of Science and Technology Progress Award of Surveying and Mapping Geographic Information of China Province in 2020.

Xiaoxi Li received the B.S. degrees in spatial information and digital technology, in 2018 from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, where she is currently working toward the M.S. degree in surveying and mapping science and technology.

Her research interests include remote sensing image stereo matching, 3-D reconstruction, and point cloud segmentation.

Yuxuan Wu received the B.S. degree in spatial information and digital technology, in 2019 from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, where he is currently working toward the M.S. degree in surveying and mapping.

His research interests include intelligent acquisition and processing of remote sensing images.

Weimin Hou received the B.S. degree in electrical engineering from Chongqing University, Chongqing, China, in 1994, and the Ph.D. degree in signal and information processing from the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, in 2007.

Since 2007, he has been an Associate Professor with the School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, China. He is the author of a book and more than 20 articles. His research interests include environment remote sensing, image processing and application, array signal processing, wireless communications, and artificial intelligence.

Dr. Hou is a Senior Member of the Chinese Institute of Electronics.

Meng Wang received the B.S. degree in applied physics from Jilin University, Changchun, China, in 2017, and the M.S. degree in electronic and communication engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2020.

She is currently a Product Manager with Shenzhen Yizhangtong Intelligent Technology Company, Ltd., Shenzhen, China. Her research interests include image recognition and semantic segmentation.

Yuwei Jin received the B.S. degree in surveying and mapping engineering from the Southwest University of Science and Technology, Mianyang, China, in 2015. He is currently working toward the Ph.D. degree with the University of Electronic Science and Technology, Chengdu, China.

His research interests include remote sensing image processing, semantic segmentation, and deep learning.

Wenbo Xu received the B.S. degree in biomedical engineering from Chongqing University, Chongqing, China, in 1994, the M.S. degree in communication and information systems from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 1998, and the Ph.D. degree in cartography and geographic information systems from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2004.

From 2004 to 2012, he was a Lecturer and Associate Professor with the School of Automation Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China. From 2013 to 2014, he was a Visiting Scholar with the Department of Geography, University of Iowa, Iowa City, IA, USA. Since 2014, he has been a Professor with the School of Resources and Environment, UESTC. He is the author of a book, more than 50 articles, more than ten inventions, and holds three patents. His research interests include environment remote sensing, image processing and application, and geographic information system.

Dr. Xu was a recipient of the First prize of Chinese Science and Technology Progress Award of Surveying and Mapping, in 2019 and the First prize of Science and Technology Progress Award of Surveying and Mapping Geographic Information of Sichuan province in 2017.