Research Article

Land Use Classification Method of Remote Sensing Images for Urban and Rural Planning Monitoring Using Deep Learning

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

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

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

Received 24 February 2022; Revised 18 March 2022; Accepted 21 March 2022; Published 11 April 2022

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

Aiming at the problems that most existing segmentation methods are difficult to deal with the imbalance of remote sensing image distribution and the overlap of segmentation target edges, a land use classification method of remote sensing image for urban and rural planning monitoring based on deep learning is proposed. Firstly, the U-Net is improved by pooling index upsampling and dimension superposition. The improvement can not only extract high-level abstract features but also extract low-level detail features, so as to reduce the loss of image edge information in the process of deconvolution. Then, the batch normalization and scaling exponential linear unit (SeLU) are used to improve the U-Net model. Finally, the improved U-Net model is applied to the classification of remote sensing images of land use types to realize dynamic monitoring. The experimental analysis of the proposed method based on TensorFlow deep learning framework shows that its total accuracy exceeds 94%. The segmentation effect of land use types in remote sensing images is good.

1. Introduction

Land cover is a complex of various elements of the surface covered by natural environmental species and artificial buildings, including surface vegetation, soil, water area, and various buildings. It has multiple forms or states in time and space [1]. In recent years, with the continuous development of China’s economy and the accelerated pace of urbanization, a large amount of land has been continuously developed and requisitioned. Especially in the eastern coastal areas, buildings and road networks extend to rural areas and primitive forest areas, and a large proportion of vegetation has been destroyed, resulting in Earth shaking changes in the original land cover [2, 3]. Therefore, using remote sensing technology to obtain high-resolution images of land cover to classify and study the surface is not only conducive to environmental protection but also conducive to the steady improvement of the national economy and the healthy development of urban and rural areas [4].

Remote sensing is a technology that detects electromagnetic radiation information on the Earth's surface from different working platforms far away from the ground, such as balloons, aircraft, various spacecraft, and artificial Earth satellites, and then draws professional images and data and sends them back to the ground receiving station [5]. With the continuous increase of observation means and the continuous iterative upgrading of automatic observation equipment, the massive data collected by satellite remote sensing observation and aerial remote sensing have also far exceeded the limit of manual processing [6]. Therefore, how to process these data more intelligently and quickly is also a difficult problem for researchers. Without effective processing means, massive observation data cannot play an important role, which greatly wastes public resources. At the same time, high-resolution images will inevitably bring more abundant ground information. While providing better source data for land cover classification technology, it is also accompanied by more diversified and complex surface features (geometry, scattered distribution, and target image coverage), which puts forward higher accuracy requirements for classification technology [7].

At present, there has been some research on remote sensing image classification at home and abroad. The traditional methods can be divided into parametric methods...
and nonparametric methods, including maximum likelihood method and support vector machine [8, 9]. In addition, random forest algorithm is also widely used in image classification. For example, Boell et al. [10] extracted and combined spectral and texture features for remote sensing image classification. Boell et al. [10] also used two dimensionality reduction feature selection methods as classifiers based on high-order feature statistics, which effectively improve the classification accuracy, but the classification efficiency is low. Shao et al. [11] used multiscale segmentation to construct the word bag representation of high-resolution images. The segmented patch is used as “image document,” and the spatial information is imported from the two-level image, so as to more accurately infer the potential theme features in the image and realize the classification of land use. However, the processing efficiency of high-precision and complex remote sensing images needs to be improved. Wang et al. [12] proposed a feature-level fusion method based on discriminant correlation analysis. In the classification stage, the decision-level fusion form is adopted to further improve the performance of scene classification and effectively improve the accuracy of land use classification, but the classification calculation cost is high. Wang et al. [13] proposed a random simulation classification algorithm of feature space indicators, which extended the statistical method from two-dimensional geographical space to m-dimensional image feature space to deduce the variation function of feature space indicators, and realized the accurate classification of land cover and use types in Duolun County, Inner Mongolia, and Huangfengqiao forest farm, Hunan, China. Zhang et al. [14] proposed a discriminant sketch subject model with structural constraints for SAR image classification. The proposed model represents the local structure mode of the image based on the sketch and then retains the local image manifold information in terms of structure and texture, which realizes high classification accuracy and time efficiency but also requires high resources and energy consumption cost.

In recent years, the further development of artificial intelligence and machine learning technology has brought new opportunities to massive data processing. Deep learning has certain advantages in land use classification of remote sensing images. It not only can extract complex features from a large number of data but also does not need artificial feature extraction in specific fields. Deep learning is a very efficient learning algorithm, which surpasses the accuracy of traditional algorithms and is the main direction of future research in the field of remote sensing [15]. Zhang et al. [16] proposed a convolutional neural network for fine segmentation of remote sensing images and extracting useful land feature information. Lei et al. [17] proposed a novel capsule network based on spectral-spatial features to improve the performance of capsule network in hyperspectral remote sensing image classification and significantly reduce the calculation cost of the model, but the calculation efficiency needs to be further optimized. Dong et al. [18] proposed a feature integration network based on deep learning, including two stages: multiscale feature fusion and enhancement. Combined with the loss function, this method realizes accurate and efficient remote sensing image classification, which is better than the current methods based on deep learning. However, the accuracy and classification efficiency of high-precision remote sensing image processing should be further improved.

Although deep learning has solved some problems in remote sensing image processing, it still faces some challenges and new opportunities. For example, in the problem of remote sensing image recognition, although the existing algorithms have made some work, they still face challenges in the aspects of different satellite data migration, classification accuracy, classification diversity, and so on. Therefore, this paper proposes a land use classification method of urban and rural planning monitoring remote sensing images using deep learning. The innovations of this paper are summarized as follows:

1. Due to the diversity of land use types, the individual segmentation targets of remote sensing images are small. Therefore, the proposed method uses multi-scale feature extraction methods such as pooling index upsampling to improve the segmentation accuracy of images.

2. For better image segmentation, the proposed method optimizes the U-Net model by batch normalization and using scaling exponential linear unit (SeLU) as the activation function to improve its classification performance.

2. Classification Model Based on Improved U-Net

2.1. Problem Description. Nowadays, the rapid development of deep learning benefits from a large number of training data in the big data environment. For example, there are 15 million pieces of data in ImageNet dataset and 9 million pieces of data in Open Images Dataset. In many immature fields, such as remote sensing image semantic segmentation, it is difficult to obtain such a level of data with good semantic annotation [19]. Therefore, how to train a good segmentation model in a small sample of remote sensing image dataset is the focus of researchers [20].

In order to solve the above two problems, an improved neural network model is proposed in this paper. The algorithm has an encoder-decoder structure, including a contraction path to capture semantics and an extension path to obtain segmentation map. The improved U-Net neural network model realizes end-to-end training, requires less remote sensing training images, and can obtain faster segmentation efficiency.

2.2. Improved U-Net Model

2.2.1. Basic U-Net Model. The basic U-Net model has a typical encoder-decoder structure, as shown in Figure 1. The encoding part adopts convolution and maximum pooling layer to obtain low-level spatial features. The decoding part on the right uses upsampling and convolution to recover the size of the feature map and uses the concatenate method to
fuse the different feature information of the encoder-decoder. The U-Net structure is symmetrical, which can make the model converge quickly in the case of few samples.

As shown in Figure 1, the basic U-Net architecture has five layers, and the overall network is in a symmetrical structure. The left is called the contraction path and the right is called the expansion path. On the left is a typical convolution neural network structure, with 4 downsampling and 5 groups of convolution, and each group of convolution is composed of two convolution operations. The size of convolution core is 3×3, the sampling core is 2×2, and the convolution step is 2. The input data size is 572×572×1. The characteristic graph is continuously reduced through convolution operation, and the number of channels is continuously increased. The final output data size of the left is 28×28×1024. On the right side, upsampling is performed first, and the size of convolution kernel is 2×2. Then, it is superimposed with the characteristic map of the corresponding layer of the left contraction path and then convoluted. The convolution and upsampling in this process need to use the activation function. The activation function of rectified liner unit (ReLU) is generally used in the U-Net layer, which can improve the gradient disappearance and speed up the training speed. The last layer of the network is 1×1 convolution, which is used to output the classification results. The activation function is generally converted into a probability map by sigmoid or softmax function.

2.2.2. Improved U-Net Model. U-Net was first used in medical image segmentation. Similarly, according to the specific task of land use classification of remote sensing image, fine-tuning design is carried out. The improved U-Net model structure is shown in Figure 2. The overall symmetrical structure of the network is mainly composed of two parts: the encoder on the left half of the diagram and the decoder on the right half of the diagram.

As shown in Figure 2, in the encoder part of the U-Net remote sensing image extraction model, two repeated convolution layers and one pooling layer are continuously used for feature image extraction and feature image dimensionality reduction. The convolution layer of each contraction path (encoder layer) adopts multiple 3×3 and uses the ReLU function as the activation function. Using multiple 3×3 convolution kernels can not only retain more detailed information of the image but also reduce the amount of network parameters and improve the calculation speed [21]. Each pooling layer adopts maximum pooling to reduce the size of the feature map to half of the original size. In the decoder part of the network, the upsampling method...
is continuously used to gradually restore the size of the feature map, and the skip connection is used to fuse the features with the convolution feature map extracted from the corresponding contraction path.

U-Net adopts the method of channel superposition for feature stitching and fusion, that is, in order to further improve the segmentation accuracy, U-Net stacks the features extracted in the encoder process with the corresponding decoder process, so as to obtain the feature maps of more channels [22, 23]. This feature fusion method can not only retain the feature information in the downsampling process to the greatest extent but also make the network contain both high-level semantic features and low-level detail features in the upsampling process. In addition, U-Net has a unique "U-shaped symmetrical structure," which is conducive to fine image segmentation. In the whole network, the input image is gradually downsampled for 4 times in the encoder, which reduces the image by 2, 4, 8, and 16 times, respectively, and continuously extracts the advanced features of the image. However, in the decoder part, the feature map is upsampled for four times, and the extracted advanced features are gradually restored to the size of the input image, so as to ensure that the edge of the output segmentation result will not be too rough. Thanks to the structural characteristics of U-Net, the network has good segmentation effect even in the case of small data sample set.

2.3. Network Model Optimization Strategy

2.3.1. Batch Normalization. The concept of batch normalization was proposed by Google in 2015. It is a skill to accelerate network training by reducing internal covariate offset. In a statistical sense, training is the process of learning data distribution, training the network model on the training set, and evaluating the performance of the network model on the test set, which is based on a strong assumption that the training set and the test set have the same data distribution. Deep neural network contains many layers, and the input of each layer is an independent data distribution [24]. The stochastic gradient descent (SGD) algorithm is usually used to train the network. During the training process, the model parameters change continuously, and the input data distribution of each layer also changes (internal covariate offset), resulting in many training problems: the learning speed is too slow, the learning effect depends heavily on the initial data distribution, and the gradient explosion and gradient disappearance in the backpropagation process [25, 26]. The
proposal of batch standardization solves the above problems. During training, use a large initial learning rate to improve the training speed and use less or no dropout and regularization to control overfitting, so as to reduce the sensitivity of the model to the initial weight. Specifically, batch processing is divided into two steps: normalization and transformation.

Normalization:
\[ \mu = \frac{1}{m} \sum_{i=1}^{m} x_i, \]
\[ \sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu)^2, \]
\[ \bar{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \vartheta}}, \]
where \( m \) means batch_size, \( \mu \) represents the mean value, \( \sigma^2 \) represents the partial variance, \( \vartheta \) is a constant set to keep the value stable, \( \bar{x}_i \) is the normalized value, and \( x_i \) indicates input data. In fact, the above normalization destroys the feature distribution learned in the network layer to a certain extent. Therefore, the transformation operation is needed.

Transformation:
\[ y_i = \alpha \bar{x}_i + \beta, \]
where \( \alpha \) and \( \beta \) are the parameters to be learned and \( y_i \) is the output after batch normalization.

The specific process of batch normalization is shown in Algorithm 1.

Input
Mini-batch input: \( x = \{x_1, x_2, \ldots, x_m\} \)
(1) Begin
(2) Calculate the mean of batch data:
\[ \mu = \frac{1}{m} \sum_{i=1}^{m} x_i \]
(3) Calculate variance of batch data:
\[ \sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu)^2 \]
(4) Standardization:
\[ \bar{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \vartheta}} \]
(5) Scale transformation and migration:
\[ y_i = \alpha \bar{x}_i + \beta \]
(5) Return learning parameters \( \alpha \) and \( \beta \).
End
Output normalized network response:
\[ y_i = BN_{\alpha,\beta}(x_i). \]

3.2.1. SeLU Activation Function. When the input of ReLU activation function is in the negative half axis, it will lead to the "death" of neurons, while SeLU can well avoid the "death" of neurons. SeLU activation function was proposed by Günter Klambauer and others. SeLU can automatically normalize the sample distribution to 0 mean and unit variance after passing through the activation function [27]. Self-normalization ensures that the gradient will not explode or disappear during training, and the effect is better than batch normalization.

The expression of SeLU is as follows:
\[ SeLU(z) = \begin{cases} 
  z, & z > 0, \\
  ye^z - y, & z \leq 0,
\end{cases} \]
where \( y \) is the scaling factor.

The activation functions such as ReLU are used in the case that the negative half axis slope is gentle, so that when the variance of activation is too large, it can be reduced to prevent gradient explosion, but the positive half axis slope is simply set to 1. The positive half axis of SeLU is greater than 1. When the variance is too small, it can increase and prevent the gradient from disappearing. In this way, the activation function has a fixed point, so that the output of each layer after network deepening is 0 and the variance is 1.

3. Experiment and Analysis
The experiment is based on Ubuntu 16.04 LTS operating system and uses the deep learning framework TensorFlow to build an improved U-Net model. Other software and hardware environments are shown in Table 1.

3.1. Dataset and Evaluation Index. Five kinds of public datasets of GID remote sensing images are used as experimental data. Five GID datasets contain 150 GaoFen-2 (GF-2) satellite images labeled at the pixel level. They are dense buildings, farmland, forest, grassland, and water body, and the annotation scales of different semantic categories vary greatly. In the experiment, the data composition of five GID datasets is shown in Table 2.

In addition, pixel accuracy (PA) and mean intersection over union (mIoU) were used to evaluate the performance of the proposed method, and precision, recall, F1 value, and accuracy were used to evaluate the effect of improving the U-Net model. The confusion matrix is shown in Table 3.

Among them, TP (true positive) means the number predicted to be positive when the real value is positive; FN (false negative) refers to the quantity predicted to be negative when the real value is positive; FP (false positive) refers to the quantity whose true value is negative and predicted to be positive; and TN (true negative) refers to the number predicted to be negative when the true value is negative.

The proportion of correct pixels in the total pixels, which is calculated as follows:
\[ PA = \frac{\sum_{i=0}^{k} P_{ij}}{\sum_{i=0}^{k} \sum_{j=0}^{k} P_{ij}}. \]
where \( P_{ij} \) is the correct number of pixels and \( P_{ij} \) is the total number of pixels.

mIoU is the ratio of intersection and union of two sets, which is calculated as follows:
\[ MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{\sum_{j=0}^{k} P_{ij}}{\sum_{j=0}^{k} P_{ij} + \sum_{j=0}^{k} P_{ji}}. \]

Precision is defined as
\[ precision = \frac{TP}{TP + FP}. \]
Recall is defined as
\[ recall = \frac{TP}{TP + FN}. \]
The F1 value is defined as
\[ F1 = 2 \cdot \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \]  
(8)

Accuracy is defined as
\[ \text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \]  
(9)

3.2. Loss Curve and Accuracy of Network Training Process.

In the training process of the improved U-Net model, three training samples are input into the training set each time, and the parameters are updated iteratively through the cross-entropy loss function and Adam optimizer. A total of 110 epochs of all samples were trained with Keras framework to complete the training of the improved U-Net model. The learning accuracy and loss value changes of training data and test data during the training process are shown in Figure 3.

As can be seen from Figure 3, with the increasing number of training epochs, the learning accuracy and loss value gradually remain stable, and the improved U-Net model gradually tends to converge. Taking the test data as an example, the accuracy and loss value of the improved U-Net model are about 93% and 0.1, respectively, which ensures better classification performance.

3.3. Comparison of Different Methods. For two groups of images, the segmentation effects of different methods are shown in Figure 4. From left to right are the original image, the label image, the segmentation result of Shao et al. [11], the segmentation result of Lei et al. [17], and the segmentation result of the proposed method, in which the box area is the selected detail area.

It can be seen from the segmentation map of lakes in the red box area on the left of Figure 4 that the lighter lakes are similar to the surrounding environment color in visual performance. Although the method in [11] uses multiscale information, it lacks efficient learning network, so the segmentation result is the worst, dividing most of the water area into farmland. Compared with the methods in [11, 17], the proposed method has the best segmentation effect. The method can better integrate large-scale information, so the segmentation effect is the best. Similarly, the segmentation map on the right of Figure 4 shows that the proposed method is ideal for the segmentation of building areas and farmland. The proposed method adopts the improved U-Net model, obtains better feature extraction ability by stacking features, and optimizes the network model by using batch normalization and SeLU activation function, so as to obtain more ideal segmentation results.

3.4. Performance Comparison of Different Models. In order to quantitatively analyze the classification performance of the proposed model, it is compared with the methods in [11, 17]. The average precision, recall, F1 value, and total accuracy in the task of land remote sensing image segmentation are shown in Figure 5.

As can be seen from Figure 5, compared with other methods, the segmentation effect of the proposed improved U-Net model is the best, and the mean values of precision, recall, F1, and the total accuracy are about 94.3%, 91.5%, 93.6%, and 94.1%, respectively. The method proposed in this paper adopts the improved U-Net to obtain multichannel features through feature stacking, which is helpful for image classification and can effectively identify most complex environments. The method in [11] constructs the word bag
representation of high-resolution images through multiscale segmentation and realizes the classification of land use combined with spatial information. However, it lacks a high-performance processing model, which has poor segmentation effect for the complex urban-rural fringe environment, with a total accuracy of about 85%. The method in [17] uses a new capsule network based on spectral-spatial features to classify remote sensing images, which can effectively improve the segmentation accuracy. The mean precision value of the method in [17] is about 4.6% higher than that of the method in [11]. However, the method in [17] does not consider multiscale features, which affects the segmentation effect, and its F1 value is reduced by about 9% compared with the proposed method.
4. Conclusion

At present, the research of remote sensing image segmentation generally tends to be combined with deep learning technology, but most of the research is not mature. Its segmentation accuracy and running time cannot lead to large-scale deployment and application in real-time environment. Therefore, this paper proposes a land use classification method of urban and rural planning monitoring remote sensing image using deep learning. Among them, the U-Net is improved by batch normalization, SeLU activation function, and feature fusion, and it is used for the classification of land use in remote sensing images to realize the reliable monitoring of urban and rural planning. The experimental results based on TensorFlow show the following:

1. The improved U-Net model adopts pooling index upsampling and skip connection to obtain better classification effect, and its accuracy and loss value are about 93% and 0.1, respectively.

2. The proposed method uses improved U-Net for remote sensing image segmentation, which can better extract edge features. Also, the proposed method uses multiscale image classification to improve the segmentation effect, and its total accuracy exceeds 94%.

The proposed method does not do the dynamic change of land use for several years. In the next research, we can try to use the proposed method to do the dynamic change of all ground objects in a certain area so that we can quickly and intuitively see the land use situation in a few years and provide guidance for the macro-control of the government or relevant departments. In addition, the test accuracy is greatly affected by the samples, and the sample needs to achieve high accuracy, which requires field verification, and the workload is large. If there are few samples in a certain place, the accuracy of the test results of this kind of ground features will be low. Moreover, if the image imaging time is different, it will also affect the accuracy. The generalization of samples is poor, and images in different years need different sample sets. Moreover, the production of samples requires a certain professional ability and can be more proficient in identifying various features in the image, which is not universal.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

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

References

[1] E. Lichtblau and C. J. Oswald, “Classification of impervious land-use features using object-based image analysis and data fusion,” Computers, Environment and Urban Systems, vol. 75, no. 5, pp. 103–116, 2019.

[2] J. Guo, L. Wang, D. Zhu, C.-Y. Hu, and C.-Y. Xue, “Subspace learning network: an efficient ConvNet for PolSAR image
classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1849–1853, 2019.

[3] M. O. Aderele, T. S. Bola, and D. O. Oke, “Land use/land cover changes of ago-owu forest reserve, osun state, Nigeria using remote sensing techniques,” *Open Journal of Forestry*, vol. 10, no. 4, pp. 401–411, 2020.

[4] O. Charfi Marrakchi, C. Masmoudi Charfi, M. Hamzaoui, and H. Habae, “Improvement of sentinel-1 remote sensing data classification by DWT and PCA,” *Journal of Sensors*, vol. 2021, no. 12, pp. 1–12, 2021.

[5] A. Shaker, W. Y. Yan, and P. E. Larocque, “Automatic land-water classification using multispectral airborne LiDAR data for near-shore and river environments,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, no. 6, pp. 94–108, 2019.

[6] L. Matikainen, K. Karila, P. Litkey, E. Ahokas, and J. Hyppä, “Combining single photon and multispectral airborne laser scanning for land cover classification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, no. 6, pp. 200–216, 2020.

[7] D. O. Gemeda, M. Demeke, and S. Legesse, “Assessment of land use, land cover change dynamics using gis and remote sensing of soro district, hadya zone, southern Ethiopia,” *Nigerian Journal of Technological Research*, vol. 14, no. 1, pp. 62–74, 2019.

[8] B. Sergio, C. Martiño, D. Nefati, G. Manuel, N. Horacio, and Jorge Novo, “Fully automatic multi-temporal land cover classification using Sentinel-2 image data - ScienceDirect,” *Procedia Computer Science*, vol. 159, no. 5, pp. 650–657, 2019.

[9] Z. Chen, Y. Hu, and Y. Zhang, “Effects of compression on remote sensing image classification based on fractal analysis,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4577–4590, 2019.

[10] M. Boell, H. R. Alves, M. Volpato, D. Ferreira, and W. Lacerda, “Exploiting feature extraction techniques for remote sensing image classification,” *IEEE Latin America Transactions*, vol. 16, no. 10, pp. 2657–2664, 2019.

[11] H. Shao, Y. Li, Y. Ding, Q. Zhuang, and Y. Chen, “Land use classification using high-resolution remote sensing images based on structural topic model,” *IEEE Access*, vol. 8, no. 8, pp. 215943–215955, 2020.

[12] X. Wang, M. Xu, X. Xiong, and C. Ning, “Remote sensing scene classification using heterogeneous feature extraction and multi-level fusion,” *IEEE Access*, vol. 8, no. 3, pp. 217628–217641, 2020.

[13] Q. Wang, H. Sun, R. Li, and G. Wang, “A new stochastic simulation algorithm for image-based classification: feature-space indicator simulation,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, no. 6, pp. 145–165, 2019.

[14] Y. Zhang, F. Liu, L. Jiao, S. Yang, L. Li, and M. Yang, “Discriminative sketch topic model with structural constraint for SAR image classification,” *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, no. 4, pp. 5730–5745, 2020.

[15] O. Youme, T. Bayet, J. M. Dembele, and C. Cambier, “Deep learning and remote sensing: detection of dumping waste using UAV,” *Procedia Computer Science*, vol. 185, no. 2, pp. 361–369, 2021.

[16] C. Zhang, Y. Chen, X. Yang et al., “Improved remote sensing image classification based on multi-scale feature fusion,” *Remote Sensing*, vol. 12, no. 2, pp. 213–223, 2020.

[17] R. Lei, C. Zhang, W. Liu et al., “Hyperspectral remote sensing image classification using deep convolutional capsule network,” *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 2, pp. 8297–8315, 2021.

[18] S. Dong, Y. Zhuang, Z. Yang, L. Pang, H. Chen, and T. Long, “Land cover classification from VHR optical remote sensing images by feature ensemble deep learning network,” *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1396–1400, 2020.

[19] M. M. Hasan, R. Islam, and M. S. Rahman, “Analysis of land use and land cover changing patterns of Bangladesh using remote sensing technology,” *American Journal of Environmental Sciences*, vol. 17, no. 3, pp. 71–81, 2021.

[20] S. Noppitak and O. Surinta, “Ensemble convolutional neural network architectures for land use classification in economic crops aerial images,” *ICIC Express Letters*, vol. 15, no. 6, pp. 531–543, 2021.

[21] R. Interdonato, D. Ienco, R. Gaetano, and K. Ose, “DuPLO: a DUal view Point deep Learning architecture for time series classification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 149, no. 3, pp. 91–104, 2019.

[22] H. Wei, L. Z. Wang, R. Y. Feng et al., “Sample generation based on a supervised Wasserstein Generative Adversarial Network for high-resolution remote-sensing scene classification,” *Information Sciences*, vol. 539, no. 6, pp. 177–194, 2020.

[23] P. Zhang and S. Hu, “Fine crop classification by remote sensing in complex planting areas based on field parcel,” *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, vol. 35, no. 21, pp. 125–134, 2019.

[24] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, “Deep learning in remote sensing applications: a meta-analysis and review,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, no. 6, pp. 166–177, 2019.

[25] J. Hu, D. Hong, and X. X. Zhu, “MIMA: MAPPER-induced manifold alignment for semi-supervised fusion of optical image and polarimetric SAR data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9025–9040, 2019.

[26] S. Shen, P. Yue, and C. Fan, “Quantitative assessment of land use dynamic variation using remote sensing data and landscape pattern in the Yangtze River Delta, China,” *Sustainable Computing: Informatics and Systems*, vol. 23, no. 9, pp. 111–119, 2019.

[27] W. Xiao, H. Ren, and X. Lu, “Vegetation classification by using UAV remote sensing in coal mining subsidence wetland with high ground-water level,” *Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery*, vol. 50, no. 2, pp. 177–186, 2019.