JAGAN: A Framework for Complex Land Cover Classification Using Gaofen-5 AHSI Images

Weitao Chen, *Member, IEEE*, Shubing Ouyang, Jiawei Yang, Xianju Li, Gaodian Zhou, and Lizhe Wang, *Fellow, IEEE*

***Abstract*—Owing to their powerful feature extraction capabilities, deep learning-based methods have achieved significant progress in hyperspectral remote sensing classification. However, several issues still exist in these methods, including a lack of hyperspectral datasets for specific complicated scenarios and the need to improve the classification accuracy of land cover with limited samples. Thus, to highlight and distinguish effective features, we propose a hyperspectral classification framework based on a joint channel-space attention mechanism and generative adversarial network (JAGAN). To relearn feature-based weights, a higher priority was assigned to important features, which was developed by integrating a two-joint channel-space attention model to obtain the most valuable feature via the attention weight map. Additionally, two classifiers were designed in JAGAN: sigmoid was used to determine whether the input data were real or fake samples produced by the generator, while Softmax was adopted as a land cover classifier to yield the prediction type labels of the input samples. To test the classification performance of the JAGAN model, we used a self-constructed complex land cover dataset based on GaoFen-5 AHSI images, which consists of mixed landscapes of mining and agricultural areas from the urban-rural fringe. Compared with other methods, the proposed model achieved the highest overall classification accuracy of 86.09%, the highest kappa amount of 79.41%, the highest F1 score of 85.86%, and the highest average accuracy of 82.30%, indicating the JAGAN can effectively improve the classification accuracy for limited samples in complex regional environments using GF-5 AHSI images.**

***Index Terms*—Attention mechanism, generative adversarial network, Gaofen-5 (GF-5), hyperspectral remote sensing, land cover classification.**

I. Introduction

LAND cover information is essential for a variety of geospatial applications, such as urban planning, regional administration, and environmental management [1]. Furthermore, it serves as the basis for understanding changes on earth's surface and related socioecological interactions [2].

More and more people utilize remote sensing images to extract land cover information [67], [68], among which hyperspectral remote sensing images are characterized by "image-spectral integration" and have been widely used to extract quantitative information in agriculture, rock and mineral identification, and environmental science. Subsequently, this method has been widely used to obtain surface quantitative information [3]–[9]. Particularly, conducting land cover classification in complex geographical scenarios is advantageous owing to its rich spectral information [10], [11]. However, in complicated environments with substantial amounts of data and spatial structures resulting from multiple bands, the automatic classification of land cover using hyperspectral remote sensing images remains a challenging task owing to the number of details on surface elements, complex spectral characteristics of surface objects, high dimensionality of the spectral bands, and limited training samples [12]–[16].In the early stages of hyperspectral image classification research, most methods aim to utilize its spectral features during classification [17], including the K-nearest neighbor (KNN) [18], spectral angle [19], extreme learning machine [20], and support vector machine (SVM) [21], [22]. However, these methods ignore interpixel spatial information [23], which limits any improvements to the classification accuracy. Spatial features are effective at improving the hyperspectral data representation and classification accuracy [10], [14], [24]–[28]. Although spatial features achieve optimal results for improving the classification accuracy, their performance is poor under conditions with limited samples. From another perspective, the dataset quality also affects the classification accuracy. Some studies have established large-scale remote sensing image datasets, which contribute to promoting the development of classification research [29]–[31]. As deep learning technology and computing power continue to advance, deep learning-based methods have been used in hyperspectral image classification owing to their strong deep-level feature extraction capability [9], [32]–[38]. These methods include deep convolutional neural networks (CNNs), autoencoders (AEs), deep belief networks (DBNs), and generative adversarial networks (GANs) [32]; [36], [39]–[41]. CNN-based methods are most widely used in the remote sensing community and can improve accuracy. Previous studies have reduced the run time in these algorithms [42]. However, CNNs exhibit poor performance with insufficient training samples. AEs have been used in hyperspectral image classification owing to their unsupervised feature learning capabilities [43]–[49]. Overall, AEs deliver limited effects for improving the accuracy of hyperspectral image classification because they yield a compressed feature representation of high-dimensional

Manuscript received November 18, 2021; revised January 7, 2022; accepted January 13, 2022. Date of publication January 25, 2022; date of current version February 10, 2022. This work was supported in part by the Fundamental Research Funds for the Natural Science Foundation of China under Grant U1803117, Grant U21A2013, and in part by the Aerospace Research Fund under Grant D040104. *(Corresponding author: Lizhe Wang.)*

The authors are with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: wtchen@cug.edu.cn; oysb@cug.edu.cn; yangjw@cug.edu.cn; ddwhlxj@cug.edu.cn; zhougaodian@cug.edu.cn; lizhe.wang@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2022.3144339

hyperspectral image data, especially with limited samples. DBNs have also been successfully applied to hyperspectral image classification [45], [50], [51]. However, under complex surface conditions, the DBN-based model requires a higher computational load for enhanced classification performance owing to miscellaneous surface objects, broken image patches, and varying spatial and geometric characteristics of surface objects. In this case, reverse propagation may induce gradient disappearance.

Deep models usually have overfitting problems in hyperspectral image classification owing to limited training samples [49], [52]–[55]. Therefore, developing effective deep model training strategies to limit overfitting is essential. GANs represent one of the available strategies for limiting overfitting [56]. During hyperspectral image classification, the discriminator training process can proceed in an effective manner to prevent immediate overfitting with insufficient training samples. Additionally, samples generated by GAN can be used as virtual samples. He *et al.* [57] proposed an early semisupervised learning method for hyperspectral and GAN, which enables the full use of limited labeled classification based on a three-dimensional (3-D) bilateral filter samples and sufficient unlabeled samples. Zhu *et al.* [58] used generated fake samples based on GAN to serve as training samples for hyperspectral image classification, which significantly improved the classification performance and alleviated the overfitting phenomenon during training. Zhan *et al.* [59] proposed a semisupervised framework for hyperspectral images based on 1D-GAN with limited labeled samples. Wang *et al.* [60] proposed a Caps-Triple GAN framework for hyperspectral image classification of 1D-CNN. The spatial features can be learned by the generator, thus, further improving the classification performance. Feng *et al.* [61] proposed a new multiclass spatial-spectral GAN, which serves as a solution to the lack of discriminative information in the generated samples and the inability to consider both spatial and spectral information. Feng *et al.* [62] proposed a collaborative learning and attention mechanism GAN, which yields a distribution of generated samples in the spectral and spatial dimensions similar to that of authentic hyperspectral images, thereby eliminating errors and confusing information. Wang *et al.* [54] proposed a dual-channel fusion capsule GAN for hyperspectral image classification by integrating the capsule network with GAN to eliminate the mode collapse and gradient disappearance problem inherent in the traditional GAN. Wang [69] developed GAN-based HSI classification methods, in which a regularization method of adaptive DropBlock to alleviate the mode collapse issue and a single classifier designed for the discriminator to deal with an imbalanced training data problem. To solve the problem of selecting a fixed number of bands and an adaptive number of bands for HSI classification, respectively, Feng *et al.* [70] proposed a method of reinforcement learning for semisupervised band selection.

Although the aforementioned GAN-based methods have yielded significant progress in hyperspectral classification, some issues still require solutions. First, potential gradient disappearance results in slow or even failed network convergence; the more completely the discriminator is trained, the more severe the disappearance of the generator gradient. The second problem is pattern crash, where samples generated by GAN feature a single data model that tends to have excessive data of a certain type and minimal data of other types. Third, we must further address overfitting problems caused by a limited training set size with high-dimensional features and the efficiency of spectral–spatial exploitation [49], [53]–[55].

The attention mechanism has been widely used to obtain significant features during hyperspectral image classification [13], [26]–[28]. Zhu *et al.* [27] proposed an end-to-end residual spectral–spatial attention network to improve classification accuracy. Zhang *et al.* [28] added a spatial attention mechanism to optimize the classification of hyperspectral images using a spectral partitioning residual network. However, the majority of hyperspectral classification models based on GAN use a single attention mechanism, such that the extraction of key features and reduction of disturbance from neighboring surface objects remains difficult, particularly in complicated surface environments.

To effectively obtain more beneficial spatial and spectral features from hyperspectral images, we propose a framework based on the GAN and channel-space joint attention mechanism (JAGAN). Gaofen-5 (GF-5) (advanced hyperspectral imager, AHSI) data were used for land cover classification of mixed landscapes along the urban–rural fringe and surface mining areas in this article. The main contributions of this article are as follows.

1) To improve the land cover classification accuracy using hyperspectral images with limited samples, this article proposes the JAGAN framework. Compared with common CNN networks, the GAN-based JAGAN can make full use of limited training samples, while a channel-space joint attention module was added to the framework to relearn low and high-level feature-based weights, assign higher priority to important features, highlight and distinguish effective features, and weaken information not conducive to classification.
2) In the channel and spatial attention modules, the results of the maximum and average pooling were integrated, which not only remains the most significant part of the features, but also retains the overall expression effect among the features. Compared with the simple attention method, this method can extract more discriminative channel space features to obtain better classification results.
3) A GF-5 AHSI semantic segmentation dataset for a mixe-d landscape from the urban–rural fringe and mining areas was constructed (Download:[1]). The dataset contains 120 bands and includes six land cover types: surface-mined land, construction land, bare land, road, cropland, and water. This dataset supplements currently available hyperspectral remote sensing datasets.

[1]https://drive.google.com/drive/folders/1-43T06aWQVj9eEwKB_edlWYPuWhgAv_L

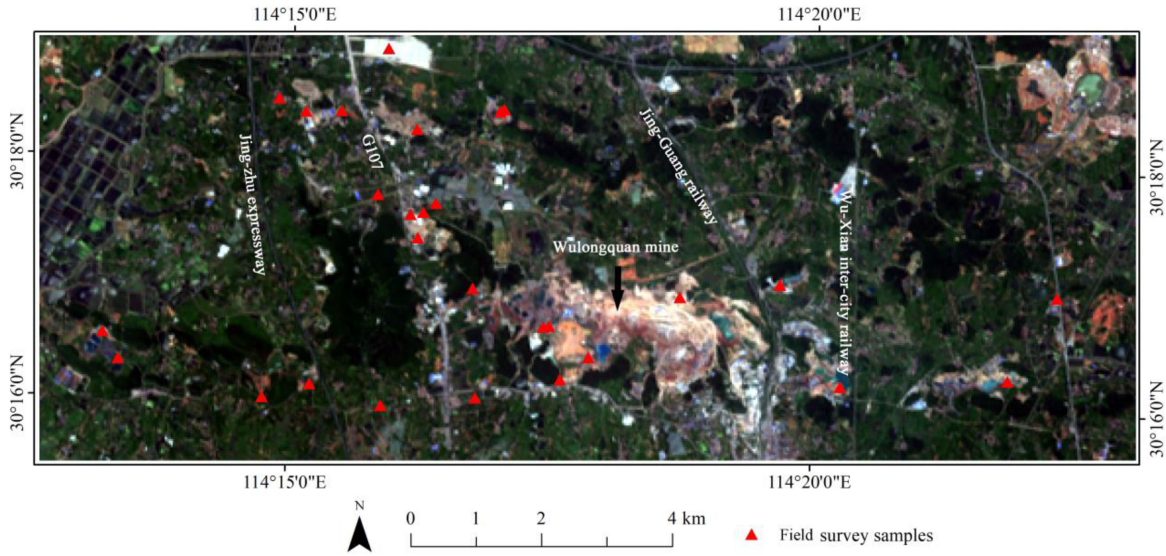


Fig. 1. GF-5 Advanced HyperSpectral Imager's true color image of the study area. RGB = band 59, 38, and 20.

TABLE I
MAIN CHARACTERISTICS OF THE ADVANCED HYPERSPECTRAL IMAGER SENSOR ON BOARD GF-5

Track Height (km)	705
Spectral Range (μm)	0.39–2.51
Total Number of Bands	330
Spectral Resolution (nm)	5 (VNIR),10(SWIR)
Ground Sampling Distance (m)	30
Band Width (km)	60

II. METHODS

A. GF-5 AHSI Dataset Construction

1) GF-5 Data Description and Preprocessing: The GF-5 satellite was launched from China on May 9, 2018, equipped with a visible short-wave infrared (SWIR) (AHSI) (see Table I). Its spectral range extends from 400 to 2500 nm, in which the visible near-infrared (VNIR) and SWIR are 5 and 10 nm, respectively. The swath width was 60 km, and the spatial resolution was 30 m [63]. Since correction for atmospheric and topographic effects is an important processing step to improve data quality, the GF-5 AHSI was radiometrically calibrated and orthorectified using ENVI 5.5 software (The Environment for Visualizing Images, ENVI), where orthorectification correction was using digital elevation model of the ASTER GDEM Version 2 (v2), with 30-m postings and 1×1 degree tiles.

The GF-5 AHSI had a total of 330 bands. To remove redundant information, principal component analysis (PCA) [64] was used to reduce data dimensionality in this article. As the variance and cumulative contribution rate of the first 10 components of the PCA reached 0.9999, the first 20 components of the PCA were selected for subsequent analysis to use the spectral information in this article completely.

2) GF-5 AHSI Dataset Construction: The study area, located in Jiangxia District, Wuhan City, Hubei Province, China, covers an area of 109.4 km^2. As it is a mixed landscape with mining and agriculture areas, the types of surface objects are complex. Particularly, Wulongquan mining areas are characterized by several types of open-pit mining land, including stopes, dumps, solid waste, and mine transfer sites, which feature significant 3-D topographic characteristics, interclass similarity, and intraclass heterogeneity. The study area possesses 218×561 pixels from GF-5 images. Fig. 1 shows a true-color image of the study area.

Employing ArcGIS 10.4 software, manual labeling methods were used to train and test sample points (see Fig. 2) containing real labels. Based on the features of the surface objects and the interpretability of the spatial resolution of GF-5, land cover was divided into six types in this article: surface-mined land, construction land, forest land, road, crop land, and water. Table II lists detailed information on this classification scheme. The GF-5 dataset was divided into two components: the first was an image containing the original image data, while the second was the land cover type labels with heights and widths identical to the image. Each label value represented the type of image pixel at its corresponding location.

The production process was as follows. First, the "tif" file from GF-5 was read through Python's GDAL library to obtain the geographic coordinates of the upper left corner of the image; the relative coordinate of each pixel was then obtained based on the geographic coordinate and pixel size. Thus, the final matrix file of $218 \times 561 \times 20$ was generated by writing the pixel value in the relative position of each pixel. Furthermore, we exported the category and geographic coordinate information of the generated sample points to a "txt" file via ArcGIS 10.4. The "txt" file was read through the Python script to obtain the relative coordinates of the sample points according to the geographic coordinate and pixel size. Finally, the relative coordinate location of each sample point was written as a value representing the category

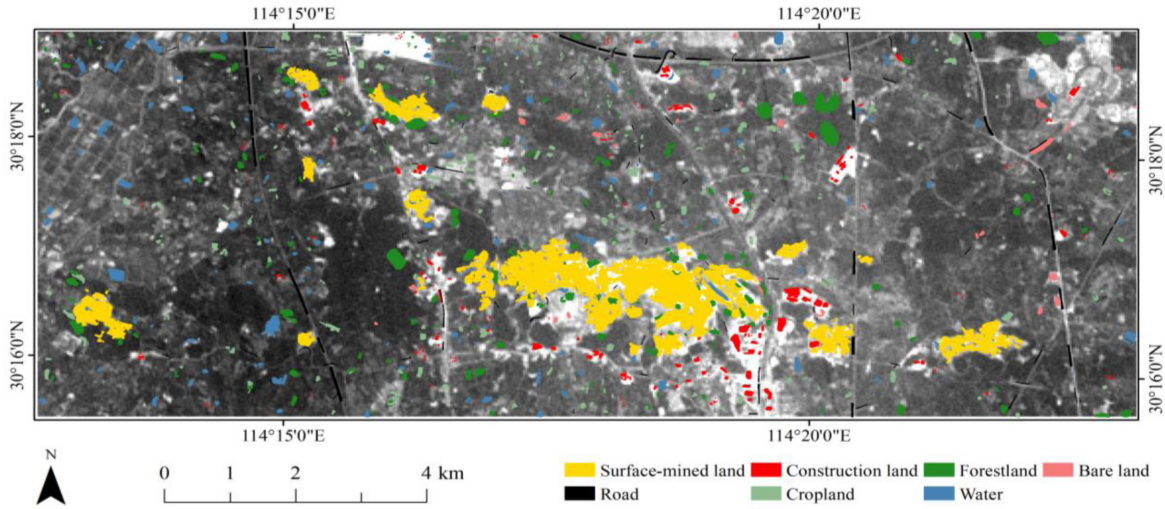


Fig. 2. Spatial distribution of the training and test samples proposed in this study.

TABLE II
LAND COVER CLASSIFICATION SCHEME FOR THE GF-5 DATASET USED IN THIS ARTICLE

Category	Description
Surface-mined areas	Mining operation area in the study area, including mining and abandoned stopes, beneficiation yard, and waste dump, among others.
Crop land	Area where crops are cultivated and cultivable in the study area, including paddy fields, greenhouses, dry land, and fallow land.
Forest land	Area where trees are planted in the study area, including nurseries, orchard, woodland, stress vegetation, and shrubs.
Water	Enclosed area containing water in the study area, including ponds and mining sumps.
Roads	Roads in the study area, including asphalt and cement roads.
Construction land	Urban, rural, civil, and commercial land in the study area, including urban land, rural residential areas, and other construction land.

TABLE III
FURTHER DETAILED DESCRIPTIONS OF EACH SAMPLE POINT

Category	Number of Sample Points	Polygon Area (km2)
Surface-mined area	4,838	4.3279
Road	484	0.4364
Water	1,018	0.9304
Crop land	919	0.8309
Forest land	1,512	1.3480
Construction Land	549	0.4891
Total	9,320	8.3627

of the sample point. Among them, 1 to 6 were designated for surface-mined areas, roads, water, crop land, forest land, and construction land, respectively, and other unmarked pixel tags were labeled as 0, ultimately generating the matrix file of land cover type labels. Table III lists the sample point numbers for each type.

B. JAGAN Framework Construction

The adopted loss function and framework of JAGAN are based on the Auxiliary Classifier GANs(ACGAN)[71], compared with original GAN, which only judges whether the input sample is true or false. The discriminator D generates the probability distribution P(S|X) = D(X) from real training data and fake data supplied by generator G(X). The purpose of the D network is to maximize the log-likelihood of the right source

$$L_D = E\left[\log P(S = \text{real}|X\text{real})\right] + E\left[\log P(S = \text{fake}|X\text{fake})\right]. \quad (1)$$

The generator G is trained to minimize the following likelihood:

$$L_G = E\left[\log P(S = \text{fake}|X\text{fake})\right] \quad (2)$$

ACGAN's network design, unlike classic GAN, can be used for multiclass image classification. The input of discriminator D is the real training data with corresponding class labels c and the fake data generated by G. The discriminator D output branch is used to distinguish real and fake data, but it also produces the classification label distribution. The loss function of ACGAN consists of two parts: the log-likelihood of the right source of input data Ls and the log-likelihood of the right class labels Lc

$$Ls = E\left[\log P(S = \text{real}|X\text{real})\right] + E[\log P(S = \text{fake}|X\text{fake})] \quad (3)$$

$$Lc = E\left[\log P(C = c|X\text{real})\right] + E\left[\log P(C = c|X\text{fake})\right]. \quad (4)$$

So D is optimized to maximize the Ls + Lc, while G is optimized to maximize Lc – Ls.

Fig. 3 shows the specific framework based on JAGAN. In JAGAN, generator G receives 100-D noise vectors and real

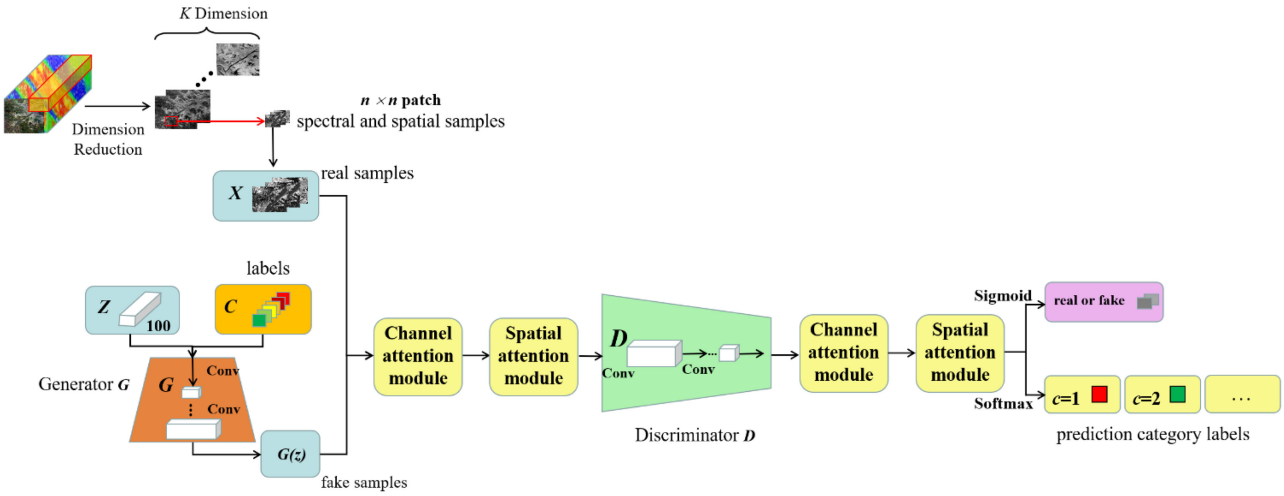


Fig. 3. JAGAN network architecture proposed in this study.

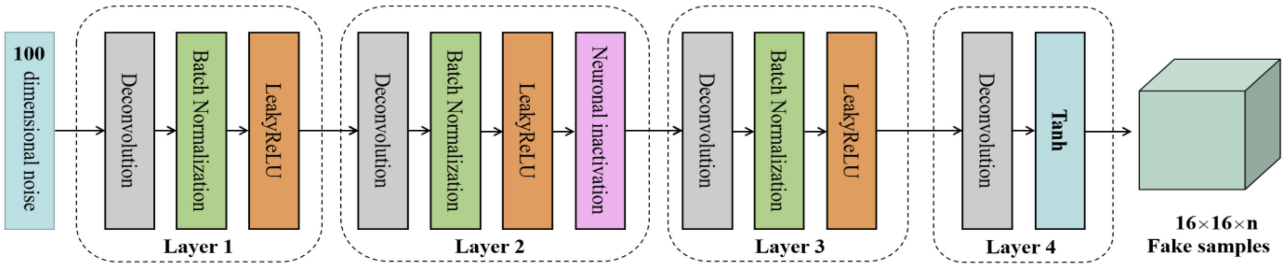


Fig. 4. Generator network structure developed in this article.

category labels, sampling them via deconvolution to generate fake samples, G(z), with shapes and sizes identical to those of the real data. The false sample dimension was the same as that in the real data after PCA dimensionality reduction, which was 20 bands. Discriminator D receives real and fake samples as inputs and uses a step-size convolution to perform down-sampling for feature extraction. Two classifiers were designed. First, the sigmoid classifier was used to determine whether the input data were from real or fake samples produced by the generator. Second, Softmax was adopted as a surface object classifier to output the prediction category labels of the input samples. Before the real and fake samples were input in front of the discriminator, as well as the SoftMax and sigmoid classifiers, the channel-space joint attention module was added to obtain the most valuable information via the attention weight map to improve the classification accuracy. The input of both generator G and discriminator D increased the sample label information; the parameters were optimized according to multiclassification loss in the network training. Therefore, compared with the traditional GAN, the proposed JAGAN was capable of optimizing the loss function more reasonably while simultaneously effectively utilizing the spectral and spatial features of the GF-5 AHSI images.

1) Generator Network Structure Construction: To improve the quality of the self-generated samples, the generator produced fake samples with sizes identical to those of the real samples for adversarial training with the discriminator. This article adopted a 16 × 16 neighborhood size as the input. Therefore, to generate fake samples with identical sizes to the real samples, the generator network had a total of four layers, as shown in Fig. 4. Table IV lists the generator network parameters.

TABLE IV
GENERATOR NETWORK PARAMETERS

Convolutional Layer Dimension	Batch Standardization	Step Size	Pixel Fill	Activation Function
4 × 4 × 128	YES	2	1	LeakyReLU
4 × 4 × 64	YES	2	1	LeakyReLU
4 × 4 × 32	YES	2	1	LeakyReLU
4 × 4 × 20	NO	2	1	Tanh

The generator first received 100-D noise vectors, which were a randomly generated set of sample values conforming to the standard normal distribution. The noise vector was first reshaped into a 3-D tensor with a size of 2 × 2 × 128 via the first deconvolution network, which featured a convolution and step size of 4 × 4 × 128 and 2, respectively. The noise vector was then reshaped into a 3-D tensor with a size of 4 × 4 × 64 via the second deconvolution network, which featured a convolution and step size of 4 × 4 × 64 and 2, respectively, while using neuronal inactivation to prevent overfitting. Finally, a 3-D tensor with a size of 8 × 8 × 32 was produced via the third deconvolution

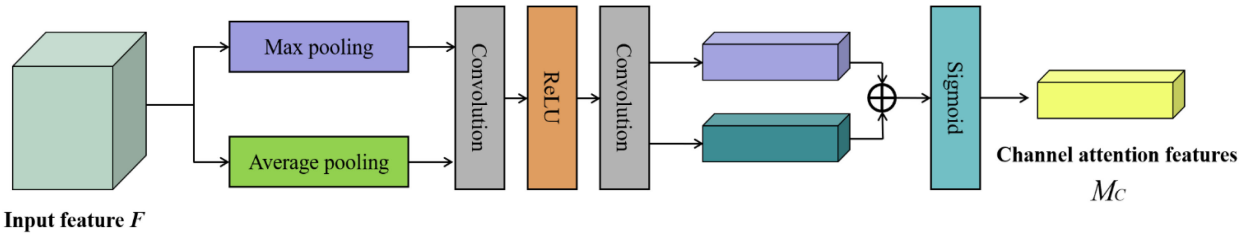


Fig. 5. Channel attention module proposed in this article.

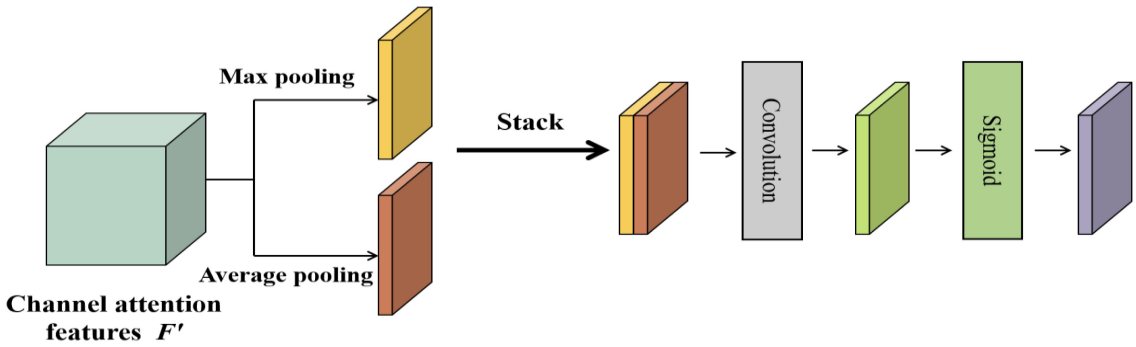


Fig. 6. Space attention module developed in this article.

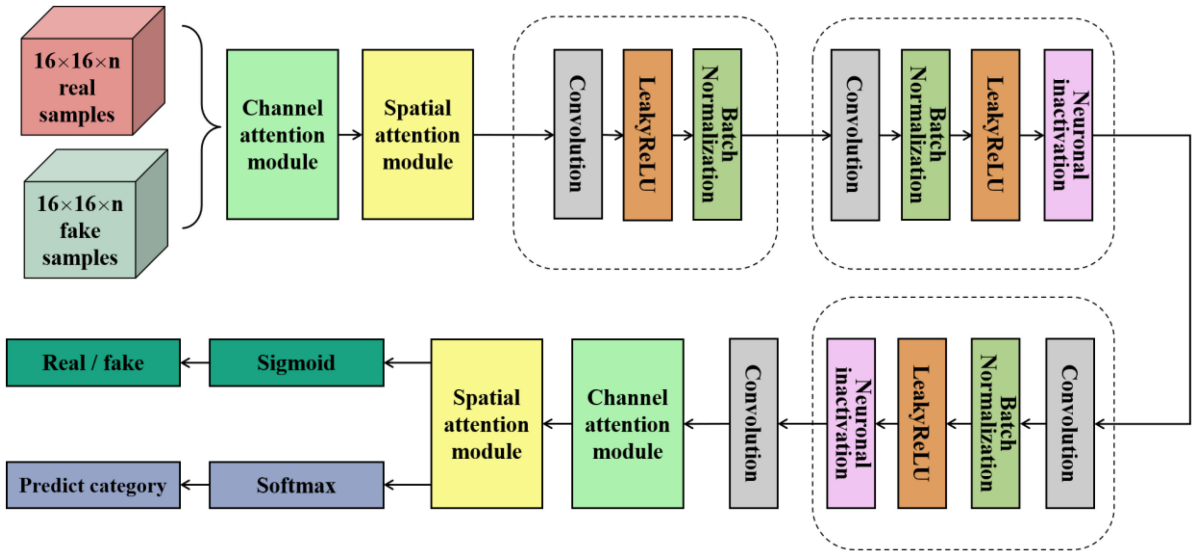


Fig. 7 Discriminator network structure developed in this article.

network, which featured a convolution and step size of 4 × 4 × 32, and 2, respectively, and a 3-D tensor with a size of 16 × 16 × 20 via the fourth deconvolution network, which featured a convolution and step size of 4 × 4 × 20, and 2, respectively. As the original GF-5 image was reduced to 20 bands through PCA and a 16 × 16 spatial neighborhood was simultaneously selected, the 100-D noise vector was mapped to a tensor of 16 × 16 × 20 with the same size as the real sample through a four-layer deconvolution network, i.e., false samples.

2) Joint Attention Module Construction: In this article, the channel-space joint attention module was utilized before discriminator input and output into the sigmoid classifier, as well as the SoftMax classifier. The feature diagram was run through two modules. First, a feature graph featuring high H, wide W, and dimension C was obtained as the input feature F, i.e., C×H×W. The feature graph was then passed through the channel attention module to obtain a channel attention diagram M_c. By multiplying the corresponding elements with the original feature,

a feature diagram was obtained, as follows:

$$F' = Mc(F) \otimes F \tag{5}$$

where $\otimes$ represents the multiplication operation of the corresponding elements.

The resulting feature diagram F' was then input into the spatial attention module and a spatial attention diagram M_s by multiplying M_s and the corresponding elements of F'. The feature diagram F'' was obtained as follows:

$$F'' = Ms(F') \otimes F'. \tag{6}$$

In the channel and spatial attention modules, the results of the maximum and average pooling were integrated. Maximum pooling preserves the most prominent part of the feature and ignores the overall expression effect of the feature. Average pooling considers the overall expression effect of features and weakens the differences between features. Therefore, the characteristics of the two pooling methods were comprehensively considered in the joint attention module; their results were fused to achieve optimal feature expression. The following are the realization processes for the attention modules in the channel and spatial domains.

2) a) Channel Attention Module: Fig. 5 shows the channel attention module used in this article. First, we input a *C*-dimensional feature graph with average and maximum pooling to aggregate spatial information, followed by obtaining two *C*-dimensional pooling feature diagrams, F_avg and F_max. Subsequently, the two pooled feature graphs were input into the multilayer perception with a hidden layer to obtain two 1 $\times$ 1 $\times$ C channel attention graphs. To reduce the parameters, the number of hidden layer neurons was C/r, where r is the compression ratio. Finally, we added the corresponding elements of the two-channel attention graphs obtained through multilayer perception. By activating the feature graph upon the addition of the sigmoid activation function, the final channel attention graph M_c was obtained. Equation (3) shows the entire process

$$\begin{aligned} &M_c(F) \\ &= \sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}\left(F\right)\right) + \mathrm{MLP}\left(\mathrm{MaxPool}\left(F\right)\right)\right) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{\mathrm{avg}}\right)\right) + W_1\left(W_0\left(F^c_{\mathrm{max}}\right)\right)\right). \end{aligned} \tag{7}$$

Using the interchannel relation of features, we can obtain a $1 \times 1 \times C$ channel attention graph. The weight of each dimension on the abovementioned map represents the importance and relevance of the key information in the feature layer corresponding to that dimension.

2) b) Space Attention Module: After the channel attention map and original feature map of the input channel attention module were multiplied by the corresponding elements, they were input into the subsequent spatial attention module, as shown in Fig. 6. First, the refined feature diagram of the channel attention was input into the spatial attention module. We performed maximum and average pooling along the channel direction to obtain 2-D feature diagrams, F_avg and F_max, respectively, both of which had a size of 1 $\times$ H $\times$ W. We then dimensionally concatenated the two obtained feature diagrams to obtain a spliced feature diagram with a size of 2 $\times$ H $\times$ W. Finally, the splicing feature diagram was passed through the convolutional layer with a convolution kernel size of 7 $\times$ 7, as well as the sigmoid activation function to generate the spatial attention diagram, $\mathrm{M_s}$. Equation (4) shows this process

$$\begin{aligned} M_s\left(F\right) &= \sigma\left(f^{7\times7}\left(\left[\mathrm{AvgPool}\left(F\right);\mathrm{MaxPool}\left(F\right)\right]\right)\right) \\ &= \sigma(f^{7\times7}\left(\left[F^s_{\mathrm{avg}};F^s_{\mathrm{max}}\right]\right). \end{aligned} \tag{8}$$

The resulting spatial attention diagram M_s and feature diagram M_c refined by the channel attention were multiplied by the corresponding elements to obtain the final output.

3) Discriminator Network Structure Construction: The JAGAN model extracted the spatial and spectral features of the GF-5 image through discriminator D. Simultaneously, the discriminator and generator conducted adversarial training to determine whether the generated samples were fake. Fig. 7 illustrates the discriminator network structure. Table V lists the discriminator network parameters.

The discriminator simultaneously received both real and fake samples generated by a generator with an identical size of 16 $\times$ 16 $\times$ 20. The spatial features were extracted through 16 $\times$ 16 spatial neighborhoods, while the spectral features were extracted in the 20-D spectral domain. Among them, 16 $\times$ 16 represents the spatial neighborhood of the pixel, while the 20th dimension is the band of the PCA after dimensional reduction.

True and fake sample data first passed through the joint attention module, whose output size was identical to that of the

TABLE V
DISCRIMINATOR NETWORK PARAMETERS USED IN THIS ARTICLE

NO.	Convolutional Layer Dimension	Batch Standardization	Step Size	Pixel Fill	Activation Function
1	4 × 4 × 32	YES	2	1	LeakyReLU
2	4 × 4 × 64	YES	2	1	LeakyReLU
3	4 × 4 × 128	YES	2	1	LeakyReLU
4	4 × 4 × 64	NO	2	1	NO
5	64 × n_classes	NO	-	-	Softmax
6	64 × 2	NO	-	-	Sigmoid

TABLE VI
NUMBER OF TRAINING AND TESTING SAMPLES USED IN THIS ARTICLE

Category	Training Samples	Test Samples
Surface-mined areas	200	4638
Roads	200	284
Water	200	818
Crop land	200	719
Forest land	200	1,312
Construction land	200	349
Total	1,200	8,120

TABLE VII
RESULTS OF THE CLASSIFICATION ACCURACY FOR THE METHODS USED IN THIS ARTICLE. HERE, 1 TO 6 REPRESENT SURFACE-MINED AREAS, ROADS, WATER, CROP LAND, FOREST LAND, AND CONSTRUCTION LAND

Category	KNN	SVM	2D-CNN	3D-CNN	3D-GAN	JAGAN
1	68.41 ± 0.64	72.70 ± 0.56	70.70 ± 0.69	75.25 ± 0.42	**94.58 ± 0.08**	93.68 ± 0.11
2	70.63 ± 0.86	71.33 ± 0.34	85.31 ± 0.26	86.36 ± 0.23	83.33 ± 0.56	**86.63 ± 0.38**
3	64.04 ± 0.36	72.52 ± 0.65	88.86 ± 0.19	86.92 ± 0.16	84.41 ± 0.49	**88.91 ± 0.26**
4	45.44 ± 1.38	55.80 ± 1.88	59.39 ± 1.11	**72.79 ± 0.53**	58.44 ± 2.53	60.61 ± 1.85
5	54.64 ± 1.55	61.55 ± 0.78	**79.26 ± 0.73**	75.84 ± 0.49	73.68 ± 0.78	74.27 ± 0.80
6	63.90 ± 0.34	72.78 ± 0.77	89.40 ± 0.14	**91.40 ± 0.08**	85.43 ± 0.26	90.53 ± 0.09
OA	63.58 ± 0.64	69.33 ± 0.72	74.24 ± 0.59	77.39 ± 0.53	85.37 ± 0.59	**86.09 ± 0.62**
AA	61.18 ± 0.65	67.78 ± 0.33	78.82 ± 0.47	81.43 ± 0.33	79.98 ± 0.56	**82.30 ± 0.55**
F1-score	66.91 ± 0.40	71.65 ± 0.55	75.42 ± 0.24	78.44 ± 0.22	85.02 ± 0.22	**85.86 ± 0.29**
Kappa	48.35 ± 0.72	55.74 ± 0.28	63.16 ± 0.44	67.13 ± 0.70	78.02 ± 0.45	**79.41 ± 0.23**

TABLE VIII
RUNNING TIME OF DIFFERENT CLASSIFICATION METHODS

Method	Training time (s)	Test time (s)	epoch
KNN	2.27	19.22	/
SVM	2.40	1.61	/
2D-CNN	1485.58	1.06	30000
3D-CNN	1122.20	1.32	30000
3D-GAN	1049.22	7.18	100
JAGAN	1106.12	7.19	100

input module. The feature diagram of 16 × 16 × 20 output by the attention module was converted into a feature diagram with a size of 8 × 8 × 32 via the first-layer convolutional network with a convolution and step size of 4 × 4 × 32, and 2, respectively. A feature diagram of 4 × 4 × 64 was output via the second layer of the convolutional network with a convolution and step size of 4 × 4 × 64, and 2, respectively. A feature diagram of 2 × 2 × 128 was output via the third layer of the convolutional network with a convolution and step size of 4 × 4 × 128, and 2, respectively. A feature diagram of 1 × 1 × 64 was output via the fourth layer of the convolutional network with a convolution and step size of 4 × 4 × 64, and 2, respectively.

The output feature diagram extracted the important features through another joint attention module, and the output remained a feature diagram of 1 × 1 × 64. The feature map was converted into a 1-D vector and then input into a fully connected layer with the Sigmoid activation function, which outputs the probability of whether the sample was real. The feature diagram of 1 × 1 × 64 was input into a fully connected layer with Softmax as the activation function to determine the type of input sample and output the corresponding label.

The discriminator network enabled classification while conducting adversarial training with the generator using the Sigmoid and SoftMax activation functions. Using a channel-space joint attention module can simultaneously achieve better classification effects in terms of complicated series and interconnecting surface object types covering construction land for urban and rural residencies, as well as mining areas.

C. Precision Evaluation Methods

JAGAN and other popular deep learning models were evaluated using the overall accuracy (OA), recall (class accuracy), F1-Score, Kappa , and average accuracy (AA, the average recall of all classes). The evaluation metrics are defined as follows:

$$\text{Overall Accuracy} = \frac{\sum_a P_{aa}}{\sum_a t_a} \tag{9}$$

$$\text{Recall} = \frac{P_{aa}}{t_a} \tag{10}$$

$$\text{Precision} = \frac{P_{bb}}{t_b} \tag{11}$$

$$F1 - \text{Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{12}$$

$$\text{Kappa} = \frac{P_o - P_c}{1 - P_c} \tag{13}$$

$$P_o = \frac{\sum_a P_{aa}}{\sum_a t_a} \tag{14}$$

$$P_c = \frac{\sum_k \left(\sum_b P_{kb} * \sum_a P_{ab}\right)}{\left(\sum_a t_a\right)^2} \quad k \in [1, K] \tag{15}$$

where P_{ab} denotes the number of samples of class a predicted to belong to class b, $t_a = \sum_b P_{ab}$, which is the total number of samples belonging to class a, $t_b = \sum_a P_{ab}$, which is the total number of samples belonging to class b, and K is the number of classes.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Comparison of Selected Methods

To verify the superiority of JAGAN in this article, comparative experiments were conducted between JAGAN and five other methods, namely KNN, SVM, 2D-CNN [65], 3D-CNN [66], and

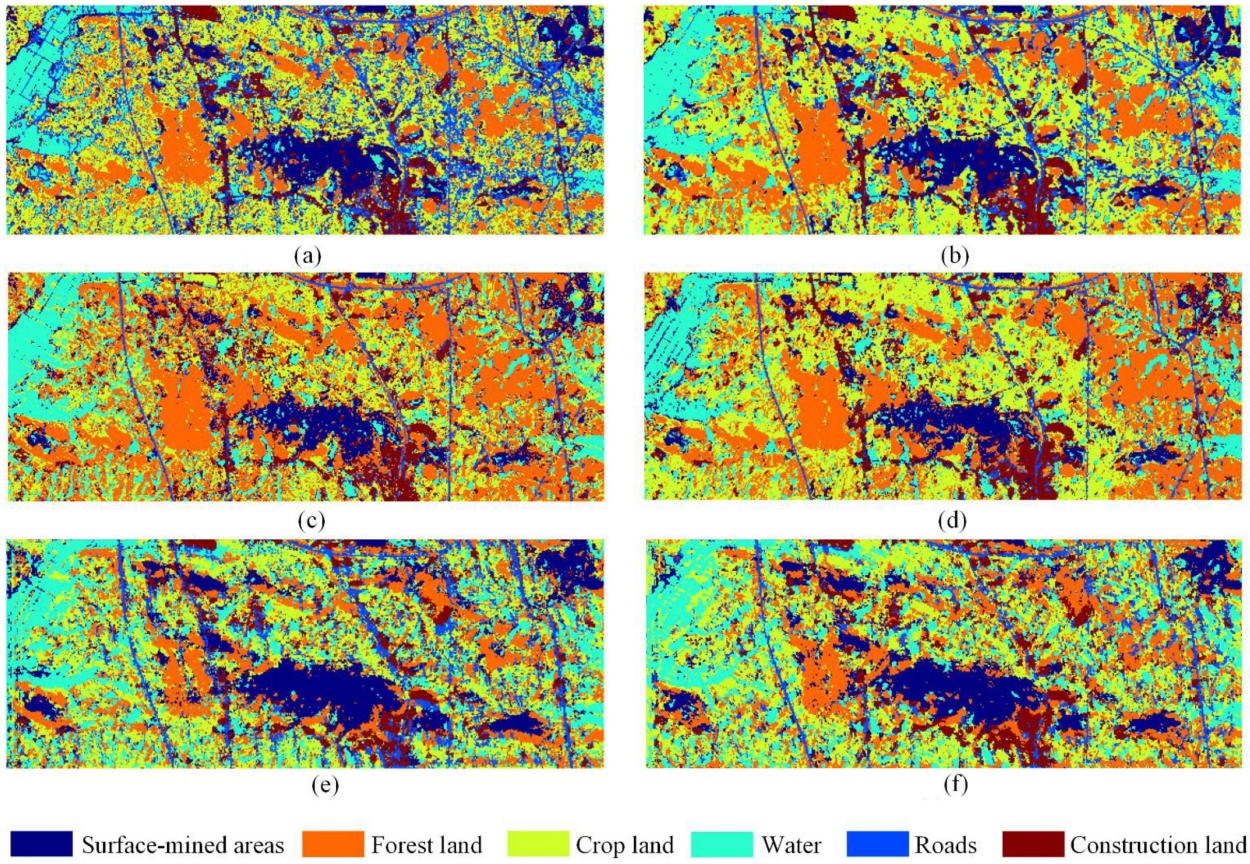


Fig. 8. Classification results for each method in the study area.

3D-GAN [58]. Among them, KNN and SVM are classic machine learning classifiers that serve as contrast benchmarks for other classifiers based on deep learning. As popular hyperspectral classifiers based on deep learning, 2D- and 3D-CNN enable the comprehensive exploitation of spatial and spectral features. 3D-GAN serves as a benchmark for comparisons with adversarial generation networks. Meanwhile, comparative experiments on various networks were conducted using limited samples to verify that JAGAN has superior capabilities with respect to insufficient labeled sample data via the generator.

B. Experimental Environment and Parameter Setting

The environment and framework used were Python v3.6 and Pytorch 1.1.0. The hardware configuration was 64 GB of memory, the CPU was an Intel (R) Xeon (R) Silver 4210, and the GPU was RTX2080ti GPU with 11 GB of memory. A total of 200 labeled sample points for each class were randomly selected, and all the remaining points were used for testing. Table VI lists detailed information on the experiments.

C. Experimental Results

Table VII lists the results of the experiments. The KNN and SVM algorithms performed the worst, but the methods based on deep learning had significantly better performance. Methods based on GAN performed better than those based on CNNs. Compared with 3D-GAN, the classification accuracy of JAGAN was better, especially for small targets, such as roads and water. This indicates that using the channel-space joint attention module can easily obtain useful features relevant to the current output, as well as more recognizable feature representations of different land cover categories, thus improving the classification accuracy. Based on the GF-5 datasets used in this article, as the proportion of the mining area in the entire test set was excessively large, it had a specific impact on the OA model. Therefore, we could not evaluate the quality of the model with only the OA; we must also compare the average performance effect on each class and the classification accuracy of each class in combination with the AA index. Although JAGAN was superior to the 3D-CNN and other methods in terms of the OA, its accuracy in some categories was slightly lower than that of 3D-CNN. A partial improvement occurred in terms of the AA index. As we focused more on the classification effect of mining areas, JAGAN still had certain advantages. In addition, Table VIII compares the running time of different classification methods. Due to the deep network architecture, deep learning-based methods cost more training time than SVM and KNN. Among deep learning-based methods. The GAN consumes longer time than 2DCNN, 3DCNN in terms of a single training epoch and test time. However, the number of training epochs is larger for 2dcnn and 3dcnn to achieve optimal accuracy, so overall training time is longer than GAN. Among

GAN-based methods, the addition of the attention mechanism makes JAGAN slightly longer than 3DGAN in terms of training and testing time. In summary, the time spent by JAGAN is not much higher than other deep learning-based methods, but most accuracy metrics are over other methods.

Using the abovementioned well-trained models, a classification map was created for the entire study area, whose results are shown in Fig. 8. Fig. 8 shows that the KNN and SVM methods had poor results; there was even the phenomenon of salt and pepper in the KNN results. Among the deep learning-based methods, the 2D-CNN yielded predictions that were more biased toward forest, while 3D-GAN yielded predictions biased toward misjudgments in forest and mining areas. We can conclude that JAGAN and 3D-CNN were superior for entire region predictions, 3D-CNN for road predictions, and JAGAN for mining area predictions. More detailed features were extracted from small target objects via 3D-CNN, while the generator in JAGAN was more often generated in mining areas, thus achieving a better classification effect.

IV. Discussion

A. Performance of JAGAN With Limited Samples

To explore whether JAGAN can improve the classification accuracy with limited samples via data expansion by generating fake samples using generators, 20, 30, and 50 sample points for each category were selected for training; the remaining labeled samples were used for tests and methods for small sample tests.

Table IX lists the results.

The CNN-based method had a relatively large demand for training data. For limited samples (taking 50 sample points for training as an example), the OA of 2D-CNN was 64.30% and Kappa was 52.42%, while 3D-CNN had an OA of 68.01% and Kappa of 55.96%. However, for a similar case, GAN-based methods yielded better classification results, where GAN had an OA of 75.23% and Kappa of 62.87%, while JAGAN had an OA of 76.69% and Kappa of 63.66%. The experimental results demonstrated that the generator in GAN can serve as a data augmentation strategy to supplement the data volume during training, thus improving the classification accuracy. Compared with 3D-GAN, JAGAN yielded improvements in several accuracy evaluation indicators. Taking 50 samples as an example, the OA, AA, F1-score, and Kappa increased by 1.4, 2.3, 0.8, and 0.8%, respectively. The adopted joint attention module was effective for key feature extraction under limited samples and classification accuracy improvement. However, the experimental results showed that CNN-based methods are superior to GAN-based methods for the AA. At the same time, CNN-based methods were more balanced between classes, whereas GAN-based methods were more inclined to produce distinctions between certain categories.

TABLE IX
Classification Results for Different Methods With Limited Samples

Method	Metrics	20 for each category	30 for each category	50 for each category
KNN	OA	52.17 ± 0.68	52.39 ± 0.51	57.11 ± 0.71
	AA	50.90 ± 0.56	53.92 ± 0.46	55.68 ± 0.67
	F1-score	55.14 ± 0.55	56.15 ± 0.53	60.25 ± 0.60
	Kappa	37.0 ± 1.26	38.21 ± 1.10	42.67 ± 1.53
SVM	OA	58.2 ± 0.76	599 ± 0.86.6	65.08 ± 0.76
	AA	57.03 ± 0.66	60.25 ± 0.77	63.88 ± 0.62
	F1-score	60.73 ± 0.53	62.81 ± 0.43	61.37 ± 0.51
	Kappa	43.99 ± 1.53	46.21 ± 0.99	52.06 ± 1.42
2D-CNN	OA	57.35 ± 1.59	61.17 ± 1.23	64.30 ± 1.12
	AA	55.38 ± 2.03	**65.26 ± 1.36**	68.78 ± 1.43
	F1-score	59.19 ± 1.02	63.12 ± 0.95	65.72 ± 0.88
	Kappa	42.85 ± 2.22	48.77 ± 1.66	52.42 ± 2.04
3D-CNN	OA	59.6 ± 0.89	62.14 ± 0.96	68.01 ± 0.84
	AA	**62.8 ± 0.97**	63.55 ± 0.88	**69.27 ± 0.78**
	F1-score	62.8 ± 0.85	1649 ± 0.68	69.66 ± 0.83
	Kappa	47.13 ± 1.32	49.41 ± 1.43	55.96 ± 1.22
3D-GAN	OA	66.80 ± 2.26	70.79 ± 1.59	75.23 ± 1.03
	AA	56.70 ± 1.99	56.96 ± 1.84	66.14 ± 1.12
	F1-score	66.51 ± 1.55	69.36 ± 1.33	74.99 ± 0.86
	Kappa	51.1 ± 1.72	55.55 ± 2.06	62.87 ± 1.53
JAGAN	OA	**68.74 ± 2.02**	**72.47 ± 1.21**	**76.69 ± 1.14**
	AA	60.25 ± 1.72	62.90 ± 1.64	68.46 ± 1.57
	F1-score	**68.52 ± 1.40**	**72.30 ± 0.75**	**75.71 ±0.75**
	Kappa	**53.7 ± 1.95**	**58.79 ± 1.62**	**63.66 ± 1.49**

B. Ablation Study on Attention Mechanisms

In order to further investigate the role of the joint channel-space attention mechanism for limited label samples, the results of the ablation study on attention mechanism with 200 labeled samples for training are presented in Table X.

In comparison with JAGAN (No.6), on the one hand, JAGAN without attention mechanism of average pooling operation (No.8) decreases in accuracy by 0.3%, 1.45%, 8.37%, and 1.05% for OA, AA, F1-score, and kappa, respectively. On the other hand, the accuracy drops by OA 0.09%, AA 2.77%, F1-score 8.32%, and Kappa 0.93% for JAGAN without attention mechanism of max-pooling operation (No.7). These indicate that both operations are used to learn not only important feature information, but also to retain helpful background information.

The following discussion explores the effect of different attention modules and module positions on classification accuracy when both the maximum pooling and average pooling operations (No.1–6) are used in the attention module. The JAGAN without the attention mechanism (NO.1) was not the best performer in OA. However, it was the worst performer in the rest of the metrics, indicating that the attention mechanism is vital for land cover classification in GF-5 AHSI dataset. In terms

TABLE X
CLASSIFICATION RESULTS FOR ABLATION EXPERIMENTS OF ATTENTION MECHANISMS

No.	Average or Max-pooling	Before first convolution block of discriminator		After last convolution block of discriminator		OA (%)	AA (%)	F1-score (%)	Kappa (%)
		Channel attention	Spatial attention	Channel attention	Spatial attention				
1	both	×	×	×	×	85.81 ± 0.59	79.85 ± 0.79	78.43 ± 1.02	77.40 ± 0.94
2	both	√	√	×	×	85.58 ± 0.75	81.69 ± 0.78	84.88 ± 0.35	78.51 ± 0.55
3	both	×	×	√	√	85.78 ± 0.64	81.47 ± 0.77	85.13 ± 0.33	78.62 ± 0.44
4	both	√	×	√	×	**86.57 ± 0.61**	80.68 ± 1.29	79.17 ± 0.93	78.51 ± 0.87
5	both	×	√	×	√	86.39 ± 0.71	80.54 ± 0.69	79.09 ± 0.98	78.24 ±1.02
6	both	√	√	√	√	86.09 ± 0.62	**82.30 ± 0.55**	**85.86 ± 0.29**	**79.41 ± 0.23**
7	average	√	√	√	√	86.00 ± 0.84	79.53 ± 0.81	77.54 ± 1.28	78.48 ± 1.30
8	max	√	√	√	√	85.79 ± 0.61	80.85 ± 1.17	77.49 ± 0.85	78.36 ± 1.02

× represents the module is not used, √ represents the module is used.

of attention mechanism placement analysis, JAGAN (NO.6) maximum improves OA, AA, F1-score, and kappa by 0.51%, 0.83%, 0.98%, and 0.9%, respectively, when compared to the placement before (NO.2) or after (NO.3) the discriminator alone, indicating that the extraction of important channel and spatial information from both low- and high- level semantic features is superior to only from low- or high- level semantic features. Compared with the use of only one attention mechanism (NO.4, 5), JAGAN (NO.6) resulted in a maximum decrease of 0.48% in OA, but maximum increases of 1.76%, 6.77%, and 1.17% in AA, F1 scores, and kappa, respectively, revealing that utilizing joint attention mechanism was more favorable for the limited imbalanced dataset. Meanwhile, the variances of the JAGAN (NO.6) in each of these cases were 0.62%, 0.55%, 0.29%, and 0.23%, respectively, which were the lowest among all compared networks. These suggest that the JAGAN is more stable in terms of classification.

V. CONCLUSION

To effectively obtain more beneficial spatial and spectral features from hyperspectral images, JAGAN model based on the channel-spatial joint attention mechanism and GAN is proposed. Tests of the JAGAN model were carried out on mixed landscapes in GF-5 AHSI data. The results indicated that the proposed JAGAN model can first focus on the key features and then provide higher weights to key features via the joint attention module, thereby increasing the classification accuracy. Second, the network focused more on obtaining useful features associated with the current output and yielding more recognizable feature representations for different categories, thereby improving the classification accuracy of small targets. Finally, the generator produced fake samples similar to real samples to attain a data expansion effect and improve the classification accuracy for small samples.

Future studies will focus on automatically determine appropriate initialization parameters according to the characteristics of the GF-5 data to optimize the classification model further. In addition, to validate the performance of the JAGAN framework further, more experiments can be conducted on publicly available state-of-the-art hyperspectral remote sensing image classification datasets, and we are also going to try to apply the JAGAN framework on semisupervised multispectral classification.

REFERENCES

[1] X. Liu *et al.*, "Classifying urban land use by integrating remote sensing and social media data," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 8, pp. 1675–1696, 2017, doi: 10.1080/13658816.2017.1324976.

[2] L. Cassidy *et al.*, "Social and ecological factors and land-use land-cover diversity in two provinces in Southeast Asia," *J. Land Use Sci.*, vol. 5, no. 4, pp. 277–306, 2010, doi: 10.1080/1747423X.2010.500688.

[3] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 12, pp. 2325–2329, 2000, doi: 10.1126/science.290.5500.2323.

[4] M. Govender, K. Chetty, and H. Bulcock, "A review of hyperspectral remote sensing and its application in vegetation and water resource studies," *Water SA*, vol. 33, no. 2, pp. 145–151, 2007, doi: 10.4314/wsa.v33i2.49049.

[5] H. Grahn and P. Geladi, "*Techniques and Applications of Hyperspectral Image Analysis*. Hoboken, NJ, USA: Wiley, 2007.

[6] F. D. Van der Meer *et al.*, "Multi-and hyperspectral geologic remote sensing: A review," *Int. J. Appl. Earth Observ. Geoinformat.*, vol. 14, no. 1, pp. 112–128, 2012, doi: 10.1016/j.jag.2011.08.002.

[7] T. Adão *et al.*, "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1110, doi: 10.3390/rs9111110.

[8] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017, doi: 10.1109/MGRS.2017.2762087.

[9] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas., "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018, doi: 10.1109/ACCESS.2018.2812999.

[10] Z. Zheng, Y. Zhong, A. Ma, and L. P. Zhang, "FPGA: Fast patch-free global learning framework for fully end-to-end hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5612–5626, Aug. 2020, doi: 10.1109/TGRS.2020.2967821.

[11] Z. Zheng, A. Ma, L. Zhang, and Y. Zhong, "Deep multisensor learning for missing-modality all-weather mapping," *ISPRS J. Photogramm. Remote Sens.*, vol. 174, pp. 254–264, 2021, doi: 10.1016/j.isprsjprs.2020.12.009.

[12] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013, doi: 10.1109/MGRS.2013.2244672.

[13] X. Tong *et al.*, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111322, doi: 10.1016/j.rse.2019.111322.

[14] A. Vail, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," *Remote Sens.*, vol. 12, no. 15, 2020, Art. no. 2495, doi: 10.3390/rs12152495.

[15] H. Sun, X. Zheng, and X. Lu, "A supervised segmentation network for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 30, pp. 2810–2825, Feb. 2021, doi: 10.1109/TIP.2021.3055613.

[16] C. Cruz-Ramos, B. P. Garcia-Salgado, R. Reyes-Reyes, V. Ponomaryov, and S. Sadovnychiy, "Gabor features extraction and land-cover classification of urban hyperspectral images for remote sensing applications," *Remote Sens.*, vol. 13, no. 15, 2021, Art. no. 2914, doi: 10.3390/rs13152914.

[17] P. Ghamisi *et al.*, "Advanced spectral classifiers for hyperspectral images: A review," *Geosci. Remote Sens.*, vol. 5, no. 1, pp. 8–32, 2017, doi: 10.1109/MGRS.2016.2616418.

[18] H. Huang and X. L. Zheng, "Hyperspectral image classification with combination of weighted spatial-spectral and KNN," *Opt. Precis. Eng.*, vol. 24, no. 4, pp. 873–881, 2016, doi: 10.3788/OPE.20162404.0873.

[19] Y. Sohn and N. S. Rebello, "Supervised and unsupervised spectral angle classifiers," *Photogrammetric Eng. Remote Sens.*, vol. 68, no. 12, pp. 1271–1282, 2002.

[20] J. Echanobe, I. del Campo, K. Basterretxea, and V. Martínez, "Genetic algorithm-based optimization of ELM for on-line hyperspectral image classification," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2017, pp. 4202–4207.

[21] D. K. Jain *et al.*, "An approach for hyperspectral image classification by optimizing SVM using self organizing map," *J. Comput. Sci.*, vol. 25, no. 1, pp. 252–259, 2017, doi: 10.1016/j.jocs.2017.07.016.

[22] Y. N. Chen *et al.*, "Feature line embedding based on support vector machine for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 1, Jan. 2021, Art. no. 130, doi: 10.3390/rs13010130.

[23] N. Audebert, B. L. Saux, and S. Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019, doi: 10.1109/MGRS.2019.2912563.

[24] X. Kang, X. Xiang, S. Li, and J. Atli Benediktsson, "PCA-Based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017, doi: 10.1109/TGRS.2017.2743102.

[25] H. C. Li *et al.*, "Gabor feature-based composite Kernel method for hyperspectral image classification," *Electron. Lett.*, vol. 54, no. 10, pp. 628–630, 2018, doi: 10.1049/el.2018.0272.

[26] H. Sun, X. Zheng, X. Lu, and S. Wu, "Spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, May 2020, doi: 10.1109/TGRS.2019.2951160.

[27] M. Zhu, L. Jiao, F. Liu, S. Yang, and J. Wang, "Residual spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 449–462, Jan. 2021, doi: 10.1109/TGRS.2020.2994057.

[28] X. Zhang, S. Shang, X. Tang, J. Feng, and L. Jiao, "Spectral partitioning residual network with spatial attention mechanism for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jun. 2021, Art. no. 5507714, doi: 10.1109/TGRS.2021.3074196.

[29] G. S. Xia *et al.*, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017, doi: 10.1109/tgrs.2017.2685945.

[30] G. S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.

[31] Y. Long *et al.*, "On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and Million-AID," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4205–4230, Apr. 2021, doi: 10.1109/JSTARS.2021.3070368.

[32] L. P. Zhang *et al.*, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *Geosci. Remote Sens.*, vol. 4, no. 2, pp. 22–40, 2016, doi: 10.1109/MGRS.2016.2540798.

[33] W. Li, G. Wu, F. hang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017, doi: 10.1109/TGRS.2016.2616355.

[34] J. Yang, Y. Q. Zhao, and C. W. Chan, "Learning and transferring deep joint spectral-spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 67, no. 99, pp. 1–14, Aug. 2017, doi: 10.1109/TGRS.2017.2698503.

[35] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018, doi: 10.1109/TGRS.2018.2841823.

[36] L. Ma *et al.*, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, 2019, doi: 10.1016/j.isprsjprs.2019.04.015.

[37] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-d-2-d CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 271–281, Feb. 2020, doi: 10.1109/LGRS.2019.2918719.

[38] Q. Lv, W. Feng, Y. Quan, G. Dauphin, L. Gao, and M. Xing, "Enhanced-random-feature-subspace-based ensemble CNN for the imbalanced hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3988–3999, Mar. 2021, doi: 10.1109/JSTARS.2021.3069013.

[39] F. Wang *et al.*, "Residual attention network for image classification," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3156–3164.

[40] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.

[41] Q. Yuan *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, 2020, Art. no. 111716, doi: 10.1016/j.rse.2020.111716.

[42] S. Mei, X. Chen, Y. Zhang, J. Li, and A. Plaza, "Accelerating convolutional neural network-based hyperspectral image classification by step activation quantization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2021, Art. no. 5502012, doi: 10.1109/TGRS.2021.3058321.

[43] X. Ma, H. Wang, and J. Geng, "Spectral–spatial classification of hyperspectral image based on deep auto-encoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4073–4085, Sep. 2016, doi: 10.1109/JSTARS.2016.2517204.

[44] Y. Pu *et al.*, "Variational autoencoder for deep learning of images, labels and captions," *Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 2353–2360, 2016.

[45] P. Zhong, Z. Gong, S. Li, and C.-B. Schönlieb, "Learning to diversify deep belief networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3516–3530, Jun. 2017, doi: 10.1109/TGRS.2017.2675902.

[46] S. Paul and D. N. Kumar, "Spectral-spatial classification of hyperspectral data with mutual information based segmented stacked autoencoder approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 138, pp. 265–280, 2018, doi: 10.1016/j.isprsjprs.2018.02.001.

[47] R. Lan, Z. Li, Z. Liu, T. Gu, and X. Luo, "Hyperspectral image classification using k-sparse denoising autoencoder and spectral–restricted spatial characteristics," *Appl. Soft Comput.*, vol. 74, pp. 693–708, 2019, doi: 10.1016/j.asoc.2018.08.049.

[48] Y. Cai, Z. Zhang, Z. Cai, X. Liu, and X. Jiang, "Hypergraph-structured autoencoder for unsupervised and semisupervised classification of hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Feb. 2021, doi: 10.1109/LGRS.2021.3054868.

[49] Z. Chen *et al.*, "Self-Attention-Based conditional variational auto-encoder generative adversarial networks for hyperspectral classification," *Remote Sens.*, vol. 13, no. 16, 2021b, Art. no. 3316, doi: 10.3390/rs13163316.

[50] K. Tan, F. Wu, Q. Du, P. Du, and Y. Chen, "A parallel gaussian–bernoulli restricted boltzmann machine for mining area classification with hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 627–636, Feb. 2019, doi: 10.1109/JSTARS.2019.2892975.

[51] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019, doi: 10.1109/TGRS.2019.2907932.

[52] H. Li, J. Li, X. Guan, B. Liang, Y. Lai, and X. Luo, "Research on overfitting of deep learning," in *Proc. IEEE 15th Int. Conf. Comput. Intell. Secur.*, 2019, pp. 78–81.

[53] W. Yu, M. Zhang, Z. He, and Y. Shen, "Convolutional two-stream generative adversarial network-based hyperspectral feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–10, May 2021, doi: 10.1109/TGRS.2021.3073924.

[54] J. Wang, S. Guo, R. Huang, L. Li, X. Zhang, and L. Jiao, "Dual-channel capsule generation adversarial network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Jan. 2021, doi: 10.1109/TGRS.2020.3044312.

[55] H. Liang, W. Bao, and X. Shen, "Adaptive weighting feature fusion approach based on generative adversarial network for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 2, Jan. 2021, Art. no. 198, doi: 10.3390/rs13020198.

[56] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020, doi: 10.1145/3422622.
[57] Z. He *et al.*, "Generative adversarial networks-based semi-supervised learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 10, 2017, Art. no. 1042, doi: 10.3390/rs9101042.
[58] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018, doi: 10.1109/TGRS.2018.2805286.
[59] Y. Zhan *et al.*, "Semi-supervised classification of hyperspectral data based on generative adversarial networks and neighborhood majority voting," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 5756–5759, doi: 10.1109/IGARSS.2018.8518846.
[60] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, "Caps-TripleGAN: GAN-assisted capsnet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7232–7245, Sep. 2019, doi: 10.1109/TGRS.2019.2912468.
[61] J. Feng, H. Yu, L. Wang, X. Cao, X. Zhang, and L. Jiao, "Classification of hyperspectral images based on multiclass spatial–spectral generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5329–5343, Aug. 2019, doi: 10.1109/TGRS.2019.2899057.
[62] J. Feng *et al.*, "Generative adversarial networks based on collaborative learning and attention mechanism for hyperspectral image classification," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1149, doi: 10.3390/rs12071149.
[63] Y. Liu *et al.*, "The advanced hyperspectral imager: Aboard China's gaoFen-5 satellite," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 4, pp. 23–32, Dec. 2019, doi: 10.1109/MGRS.2019.2927687.
[64] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, 1987, doi: 10.1007/1-84628-124-5_3.
[65] V. Sharma *et al.*, "Hyperspectral CNN for image classification & band selection, with application to face recognition," Technical Report KUL/ESAT/PSI/1604, ESAT, Leuven, Belgium, 2016.
[66] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, Jan. 2017, Art. no. 67, doi: 10.3390/rs9010067.
[67] Y. Chen, L. Tang, Z. Kan, M. Bilal, and Q. Li, "A novel water body extraction neural network (WBE-NN) for optical high-resolution multispectral imagery," *J. Hydrol.*, vol. 588, 2020, Art. no. 125092, doi: 10.1016/j.jhydrol.2020.125092.
[68] Y. Chen, Q. Weng, L. Tang, Q. Liu, X. Zhang, and M. Bilal, "Automatic mapping of urban green spaces using a geospatial neural network," *GIScience Remote Sens.*, vol. 58, pp. 1–19, 2021, doi: 10.1080/15481603.2021.1933367.
[69] J. Wang, F. Gao, J. Dong, and Q. Du, "Adaptive dropblock-enhanced generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5040–5053, Jun. 2021, doi: 10.1109/TGRS.2020.3015843.
[70] J. Feng *et al.*, "Deep reinforcement learning for semisupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, Feb. 2022, doi: 10.1109/TGRS.2021.3049372.
[71] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANS," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2642–2651.

Weitao Chen (Member, IEEE) was born in Wugang, China. He received the B.E. degree in land resource management from Jiaozuo Institute of Technology, Jiaozuo, China, in 2003, and the M.E degree in quaternary geology and the doctor's degree in environmental science and engineering from China University of Geosciences (CUG), Wuhan, China, in 2012 and 2006, respectively.

He is a Professor with the School of Computer Science, CUG. He has authored and coauthored more than 30 papers. His main research interests include machine learning and remote sensing of environment.

Prof. Chen is a Member of IEEE.

Shubing Ouyang was born in Fuzhou City, China, in 1990. She received the B.S. degree in geology from the Wuhan University of Engineering Science, Wuhan, China, in 2012 and the M.S. degree in mineral resource prospecting and exploration in 2015 from the China University of Geosciences, Wuhan, China, where she is currently working toward the Ph.D. degree in geoscience information engineering with the School of Computer Science.

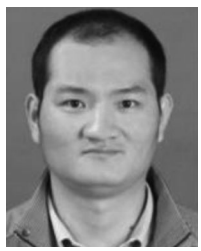
Her research interests include geoscience information processing, remote sensing image processing, and deep learning.

Jiawei Yang received the B.S. and M.S degrees from the China University of Geoscience, Wuhan, China, in 2018 and 2021, respectively.

His research interests include remote sensing image processing, computer vision, and deep learning.

Gaodian Zhou received the B.Eng. and M.E. degrees in resource exploration engineering, in 2014 and 2017, respectively, from the China University of Geosciences, Wuhan, China, where he is currently working toward the Ph.D. degree in geoscience information engineering with the School of Computer Science.

His research interests include semantic segmentation, remote sensing image process, and big data.



Xianju Li received the B.S. degree in geomatics engineering, the M.S. degree in geodesy and survey engineering, and the Ph.D. degree in surveying and mapping from China University of Geoscience, Wuhan, China, in 2009, 2012, and 2016, respectively.

Since 2016, he has been an Associate Professor with the School of Computer Science, China University of Geosciences. He has authored and coauthored more than ten papers. His main research interests include remote sensing image processing and analysis, computer vision, and machine learning.

Lizhe Wang (Fellow, IEEE) received the B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the Doctor of Engineering degree from the University Karlsruhe (*Magna Cum Laude*), Karlsruhe, Germany, in 2008.

He is a ChuTian Chair Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include HPC, e-Science, and remote sensing image processing.

Prof. Wang is a Fellow of IET and SPIE.