Hyperspectral image classification based on coupled dual-channel generative adversarial network

Jiaji Shi, Wen Yi, Yuan Liu, Xiaochen Lu*
College of Information Science and Technology, Donghua University, Shanghai 201620, China

ABSTRACT

The aim of hyperspectral image (HSI) classification is to define categories for the different labels that are assigned to each pixel vector. Generative adversarial network (GAN) can mitigate the limited training sample dilemma to some extent, but there are still two critical issues, namely balance collapse and insufficient sample diversity. In this article, we propose a Coupled Dual-Channel Generative Adversarial Network for HSI classification. It mainly consists of a Coupled Generative Network (CGN) and a Dual-Channel Discriminative Network (DDN). CGN achieves the reconstruction of HSI samples through cascaded convolutional layers, in which the label information of the samples is used to avoid the balance collapse, while DDN extracts spatial attention weights and spectral attention weights of the input true/false samples respectively, and performs feature mining in both spatial and spectral dimensions in a dual-channel style. In order to reinforce the detailed features of input samples in both spatial and spectral dimensions, we design a new Cascaded Spatial-Spectral Attention Block (CSSAB). Finally, feature maps at different scales are fused for final sample discrimination and classification, which can mitigate the effects of insufficient sample diversity. Experimental results on two HSI data sets demonstrate that the proposed CDGAN effectively improves the classification performance compared to some state-of-the-art GAN-based methods.

Keywords: Convolutional neural network, generative adversarial network, HSI classification, spatial-spectral, attention mechanism

1. INTRODUCTION

The fine spectral bands of hyperspectral image (HSI) provide rich semantic information on the ground surface in remote sensing field, so that the spectral bands and spatial distribution characteristics of HSI can be used for the application of feature classification. Deep learning models train classifiers while sufficiently extracting high-level features through nonlinear transformations, and continuously optimize themselves in a data-driven manner. Convolutional neural networks (CNNs) use local connections to extract spatial features and reduce the number of parameters by sharing weights. Owing to this superiority, HSI classification methods based on CNNs and their variants[1-7] have gained much popularity in recent years. However, although CNN models have greatly reduced the number of parameters, training these parameters still requires a large amount of labeled data which are known to be costly. Without enough sufficient labeled data, these CNN-based models are usually overtrained thus causing overfitting problems.

Generative Adversarial Networks (GANs)[8] have excellent generalization capabilities and can generate high-quality samples independent of the quality of unlabeled data. A GAN model usually contains a generator and a discriminator, where the generator generates fake data that are as real as possible, and the discriminator tries to discriminate real data from fake data. By adversarial competition between the two networks, both sides get the optimal results. This paper proposes a HSI classification method based on Coupled Dual-Channel Generative Adversarial Network (CDGAN). This method up samples the input data in a parameter-sharing coupled manner, and achieves the goal of fitting virtual samples through coupled generative networks. Moreover, a novel Cascaded Spatial-Spectral Attention Block (CSSAB) is introduced to refine the features of generated samples. In the discriminative network, the input samples are fused based on the spatial-spectral cascaded attention details, and then the key information is mined by a dual-channel spatial-spectral feature extraction structure. The classification of features is accomplished simultaneously during the confrontation between the generative network and the discriminative network. In this paper, we first introduce the related work pertaining to our topic. This is followed by a detailed explanation of our methodology used to develop the model. Subsequently, we present

*lxchen09@dhu.edu.cn; phone 183 4510 2382

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 1339506 · © 2024 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3049212

the results of our experiments and offer comparisons and detail analysis discussion for the presented models.

2. PROPOSED METHOD

2.1 Coupled dual-channel generative adversarial network model

The framework of the Coupled Dual-Channel Generative Adversarial Network (CDGAN) is shown in Figure 1, which consists of a Coupled Generative Network (CGN) and a Dual-Channel Discriminative Network (DDN) based on a coupled structure.

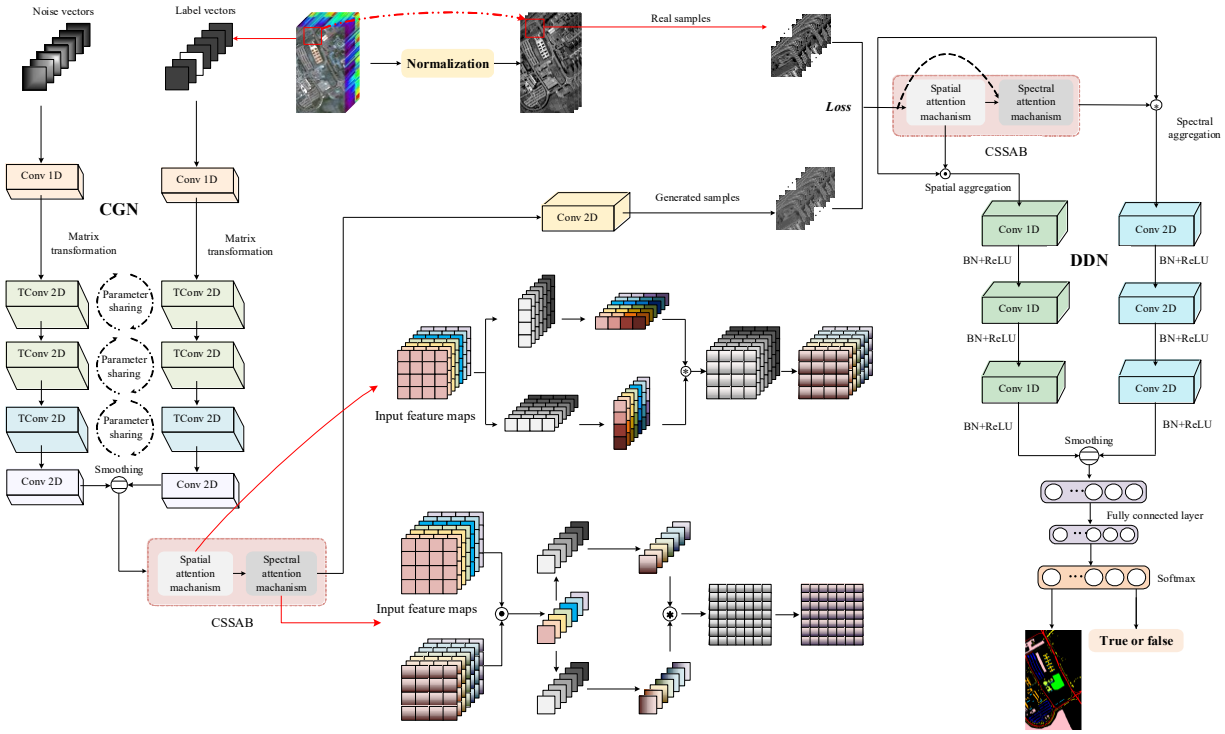

Figure 1. The framework of CDGAN for HSI classification.

2.1.1 Sample generation based on coupled generative networks. In CGN, the label vectors and the noise vector are fed into separated branches of the network. The feature map of the label vectors $\mathbf{y} \in \mathbb{R}^{1\times T}$ will first be fed into a one-dimensional convolutional layer for initial information mapping, and transformed into $\mathbf{y}_1 \in \mathbb{R}^{1\times 1\times (C/2)}$, where T is the class number. Each of the convolution layers is followed by a BN layer and a ReLU activation function, except for the last transpose convolution (TConv) layer. After BN and ReLU, the feature map $\mathbf{y}_1$ is then fed into three successive 2D TConv layers for spatial-scale upsampling to progressively reconstruct the spatial information.

In these three TConv layers, the first two convolutional kernels are both of size 3×3 containing $C/2$ channels, with an upsampling step of 1. The corresponding parameters of the last TConv layer are 5×5, $C/2$, 1, thus obtaining a new feature map $\mathbf{y}_2 \in \mathbb{R}^{9\times 9\times (C/2)}$. To ensure the spectral characteristics of the generated samples, a convolutional layer with a kernel of size 1×1 will accomplish the reconstruction of the spectral information of $\mathbf{y}_2$ with $\tilde{C}/2$ convolutional kernels and obtain $\mathbf{y}_3 \in \mathbb{R}^{9\times 9\times (C/2)}$. Similarly, the branch in which the noise vector $\mathbf{Z} \in \mathbb{R}^{1\times \ell}$ is located will accomplish the reconstruction of both spatial and spectral dimensions in the same way and obtain $\mathbf{Z}_3 \in \mathbb{R}^{9\times 9\times (C/2)}$. The two vectors are concatenated together in the spectral dimension to obtain $\mathbf{g} \in \mathbb{R}^{9\times 9\times C}$. After then, $\mathbf{g} \in \mathbb{R}^{9\times 9\times C}$ is fed into CSSAB to enhance the key information in the generated feature maps and suppress the representation of interference information. Ultimately, the feature map achieves the final detail reconstruction of the sample through a 2D convolutional layer of size 3×3, which completes the

generation of the sample $\tilde{\mathbf{g}} \in \mathbb{R}^{9\times9\times C}$ through a hyperbolic tangent activation function.

2.1.2 Discrimination and classification based on two-branch discriminative networks. In DDN, the input samples are first processed by CSSAB to obtain a spatial attention detail mask $\mathbf{A}_{spa} \in \mathbb{R}^{9\times9\times C}$ and a spectral attention detail mask $\mathbf{A}_{spe} \in \mathbb{R}^{C\times C}$. Subsequently, the spectral attention $\mathbf{A}_{spe}$ is matrix multiplied with the input sample $\mathbf{d}$, and the aggregation in the spectral dimension is performed guided by this weighting factor to obtain the spatial feature

. Similarly, the spatial attention $\mathbf{A}_{spa}$ is used to guide the aggregation of the spatial information on the input sample to obtain the spectral feature $\mathbf{d}_{spe} \in \mathbb{R}^{1\times C}$. Both feature maps are obtained and regularized through the BN layer, and at the end, the features will be concatenated together and sent to the FC layer for final fusion, where the discrimination and classification are achieved by the Softmax function.

In the adversarial process of CDGAN, DDN is always expected to identify the samples generated by CGN from the input samples and place them in class $T+1$, while the goal of CGN is to continuously fit the generated samples to approximate the true sample distribution, conforming to the former class b, in order to fight against DDN. In this process, the objective function of CDGAN is as follows:

$$\begin{aligned} L_G &= \sum_{i=1}^{\Lambda} l\ (D(G(z_i, y_i)), y_i) \\ L_D &= \sum_{i=1}^{\Lambda} l\ (D(h_i), y_i) + \sum_{i=1}^{\Lambda} l\ (D(G(z_i, y_i)), y_{T+1}) \end{aligned} \tag{1}$$

Among them, L_G is the objective function of the generator of CGN, L_D is the objective function of the discriminator of DDN, and $l(\cdot)$ denotes the cross-entropy loss function between different labels, and the process uses a total of Λ samples.

2.2 Cascaded spatial-spectral attention block

2.2.1 Spatial attention mechanism. For the input feature map $\mathbf{x} \in \mathbb{R}^{w\times h\times c}$, the spatial attention mechanism first feeds it into a convolutional layer $\mathbf{W}_h \in \mathbb{R}^{w\times 1\times c}$ with a kernel of size $w\times1$ to obtain the spatial embedding information $\mathbf{x}_h \in \mathbb{R}^{1\times h\times c}$ of the feature map on the *h*-th scale, and similarly, the spatial embedding information $\mathbf{x}_w \in \mathbb{R}^{w\times 1\times c}$ on the *w* scale will be obtained through a convolutional layer $\mathbf{W}_w \in \mathbb{R}^{w\times 1\times c}$ with a size of $1\times h$. All the convolutional layers described here aggregate spatial information on a channel-by-channel layer by *c* convolutional kernels in a depth-separated manner for HSI. Let $w(x_j) = \mathbf{W}_w \otimes x_j$ and $h(x_i) = \mathbf{W}_h \otimes x_i$ denote the results of the depthwise convolution (DWConv) operation, the parameters *i, j* denote the coordinate positions in the feature map **x**. After matrix transformation $\mathbf{T}$ of the two spatial embeddings, the embedded spatial information $\mathbf{x}_h^{\mathbf{T}} \in \mathbb{R}^{c\times h\times 1}$ and $\mathbf{x}_w^{\mathbf{T}} \in \mathbb{R}^{c\times 1\times w}$ perform matrix multiplication $*$ and then we can obtain the spatial correlation matrix $\mathbf{M}_{spa} \in \mathbb{R}^{h\times w\times c}$ of the input feature map on different channels, and after going through a channel-by-channel layer of Softmax function to obtain the final spatial attention distribution $\mathbf{A}_{\text{spa}} \in \mathbb{R}^{w\times h\times c}$:

$$a_{z(i,j)}^{spa} = \frac{\exp(m_{z(i,j)}^{spa})}{\sum_{i=1}^{w}\sum_{j=1}^{h}\exp(m_{z(i,j)}^{spa})} \tag{2}$$

where $a_{z(i,j)}^{spa}$ represents the spatial attention value of the feature map at *z*-th channel with spatial coordinates at (i, j), and $m_{z(i,j)}^{spa}$ is the corresponding value in $\mathbf{M}_{spa}$. Due to the Softmax function, the sum of each channel is 1.

2.2.2 Spectral attention mechanism. After the spatial attention distribution is obtained, it will be further used for the subsequent spectral attention generation. Analogous to GAP, a global pooling strategy based on the spatial attention distribution will be used to fuse the spatial information for the feature maps in this paper, namely:

$$e_z = \sum_{i=1}^{S}\sum_{j=1}^{S} a_{z(i,j)}^{spa} \times x_{z(i,j)}$$

By using the spatial attention distribution $\mathbf{M}_{spa}$ as a weighting factor, the input feature map is weighted to obtain the spectral feature vector $e \in \mathbb{R}^{1\times c}$. This summation process is maintained for each channel of the feature map. Subsequently, two point-by-point convolution layers $\mathbf{W}_v \in \mathbb{R}^{1\times c}$ and $\mathbf{W}_u \in \mathbb{R}^{1\times c}$ with a kernel of size 1, are used to perform spectral information mining respectively. Let $v(e) = \mathbf{W}_v \odot e$ and $u(e) = \mathbf{W}_u \odot e$ denote the results of the point-by-point convolutional operation, the symbol $\odot$ denotes the operation process, and in addition the residual connection Res. is applied on the generated feature vector. After matrix multiplication, the spectral correlation matrix spectral $\mathbf{M}_{spe} \in \mathbb{R}^{c\times c}$ is fed into the Softmax function to obtain the spectral attention mask $\mathbf{A}_{spe} \in \mathbb{R}^{c\times c}$ of the target, and the value of the coefficients in the mask can be calculated as follows:

$$a_{i,j}^{\text{spe}} = \frac{\exp(m_{i,j}^{spe})}{\sum_{i=1}^{c}\sum_{j=1}^{c}\exp(m_{i,j}^{spe})} \tag{3}$$

where $a_{i,j}^{spe}$ is the correlation attention coefficient between the *i*-th band and the *j*-th band, and $m_{i,j}^{spe}$ is the corresponding value in $\mathbf{M}_{spe}$.

3. EXPERIMENTS AND RESULTS

3.1 Experimental data and related settings

In order to verify the effectiveness of the proposed method, the UP dataset and the SS dataset are taken for experiments and comparisons, where 40 samples of each class in the UP dataset are obtained in a random way for the training of the experiments, and 20 samples of each class will be taken randomly for the SS dataset. In this process, overall accuracy (OA), average accuracy (AA), and Kappa measures are used to evaluate the performance of the proposed method, and all experimental results are obtained by counting the results of 10 independent runs.

In this experiment, all samples are normalized and a symmetric strategy is used as the padding pattern centered on the labeled image element to generate sample data with a spatial size of 9×9 for each dataset, and the full spectral band is retained, while the dimension of the input noise vector in CGN is set as $\ell = 100$. The batch size of the training samples is set to 16, both CGN and DDN are optimized by the Adaptive Moment Estimation (Adam) optimizer, with the initial learning rate set to 5×10^{-4} and the momentum is 0.9, and the optimization of the model is completed within 300 iterations cycles. To verify the effectiveness of the proposed method, Support Vector Machine (SVM)[9] based on Extended Morphological Profiles (EMP) method, Hybrid Spectral Convolutional Neural Network (HybridSN) method[10], 3D Convolutional GAN (3DGAN) method based on 3D convolution[11], Semi-Supervised GAN (SSGAN) method[12], and Capsule GAN (CapsGAN) method[13] based on capsule network are tested against each other. In these methods, the parameters of SVM are tuned manually. To compare with the presented method, the parameters related to HybridSN are set according to Reference[10], while the input of 3DGAN is set to 64×64, and the spacial size of both SSGAN and CapsGAN is set to 9×9, and their parameters are set according to References[12,13].

3.2 Results and analysis

The metrics of the two sets of experimental results are shown in Tables 1 and 2, and the corresponding classification visualization results are shown in Figure 2. Specifically, for the UP dataset, the CDGAN method proposed in this paper achieves the best values for OA and Kappa coefficients. The best result for the AA metric is achieved by the SSGAN method, while the AA value of CDGAN achieves the second-best result. For the GAN-based methods, the increase of fitted samples effectively alleviates the problem of overfitting and achieves good results in the categories of grass and metal slabs. Compared with other GAN-based methods, the method proposed in this paper can identify tarmac category better. For the SS dataset, CDGAN achieves the best results on all three metrics for the SS dataset, improving by 12.68%, 7.51%, 3.24%, 6.11%, 2.76%, and 0.66% respectively on the OA metric compared to the other methods and obtaining the best result of 98.50% on the AA metric, with the Kappa coefficient at least 0.65% ahead of the other methods. Furthermore, it

can be seen from the table that the method proposed in this article leads the other methods in terms of correct identification on most categories. In summary, the CDGAN method proposed in this paper achieves the best results on two publicly available datasets, validating the effectiveness of the method.

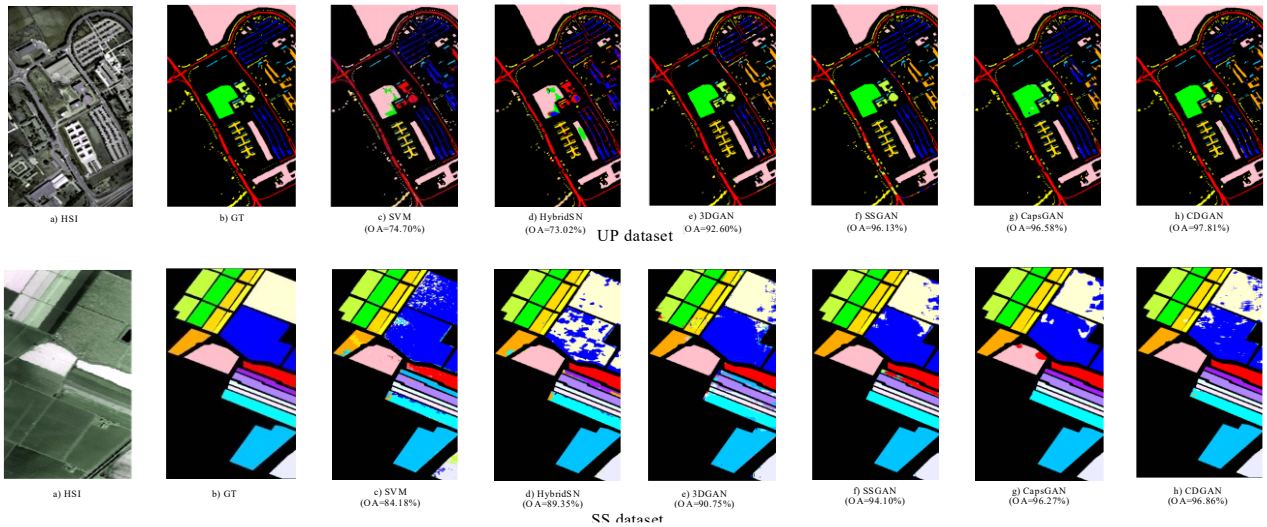

Figure 2. Visualization and classification maps of different approaches for the UP and SS datasets. Upper line is the UP dataset, lower line is the SS dataset. (a) HSI, (b) GT, (c) SVM, (d) HybridSN, (e) 3DGAN, (f) SSGAN, (g) CapsGAN, (h) CDGAN.

Table 1. Classification results of different approaches on the UP dataset.

Category	SVM	HybridSN	3DGAN	SSGAN	CapsGAN	CDGAN
Asphalt	77.22	56.41	88.26	**99.84**	96.29	97.23
Meadows	73.81	77.21	95.75	98.93	98.96	**99.53**
Gravel	24.42	16.58	83.33	**95.67**	92.98	94.35
Trees	99.87	96.86	86.72	97.95	**98.21**	98.05
Painted metal sheets	96.51	78.58	86.51	97.04	99.37	**99.54**
Bare soil	93.96	67.4	97.8	98.96	99.96	**99.32**
Bitumen	16.14	10.58	94.41	59.14	**99.48**	96.62
Self-blocking bricks	60.32	57.55	**90.64**	88.16	80.89	90.17
Shadows	99.89	99.88	79.97	96.74	98.92	97.32
OA (%)	74.7	73.02	92.6	96.13	96.58	**97.81**
AA (%)	68.4	69.61	87.46	**96.03**	93.05	95.92
Kappa (%)	72.62	71.39	90.11	94.9	95.45	**97.09**

Table 2. Classification results of different approaches on the SS dataset.

Category	SVM	HybridSN	3DGAN	SSGAN	CapsGAN	CDGAN
Brocoli_green_weeds_1	98.63	100	98.95	69.96	99.95	99.93
Brocoli_green_weeds_2	97.91	99.88	95.02	99.77	98.99	**99.86**
Fallow	88.45	93.22	98.58	98.94	99.13	**99.28**

Category	SVM	HybridSN	3DGAN	SSGAN	CapsGAN	CDGAN
Fallow_rough_plow	97.52	97.42	62.79	**99.89**	99.24	99.64
Fallow_smooth	92.55	93.5	65.1	87.07	98.24	**99.88**
Stubble	99.42	99.54	92.36	99.54	**99.89**	99.75
Celery	95.55	96.5	98.29	98.72	98.97	**99.83**
Grapes_untrained	64.47	91.22	91.27	94.02	91.54	**94.41**
Soil_vinyard_develop	91.11	99.93	97.02	99.64	99.22	99.72
Corn_senesced_green_weeds	87.3	95.32	88.6	94.01	**99.67**	98.66
Lettuce_romaine_4wk	78.28	91.85	97.84	99.65	**100**	**100**
Lettuce_romaine_5wk	92.56	94.94	98.02	99.44	99.67	**99.71**
Lettuce_romaine_6wk	90.44	98.36	78.94	99.81	98.29	**99.89**
Lettuce_romaine_7wk	94.99	96.86	93.88	76.86	98.21	97.41
Vinyard_untrained	77.15	61.65	85.81	**91.4**	87.5	87.34
Vinyard_vertical_trellis	98.69	100	95.26	98.87	**100**	**100**
OA (%)	84.18	89.35	90.75	94.1	96.27	**96.86**
AA (%)	86.65	89.73	88.19	91.2	98.15	**98.5**
Kappa (%)	82.21	88.21	89.7	93.42	95.85	**96.5**

4. CONCLUSION

In this paper, a GAN framework-based HSI classification approach is proposed for exploring HSI classification methods with small samples. Firstly, the noise vector is fed into the generator of CGN for generating the HSI samples. In order to increase the authenticity of the samples, the label vector is fed into CGN in parallel, and the spatial reconstruction of the samples is completed by coupled upsampling layers with shared parameters during sample generation, while the effective mapping of the noise is enhanced by learning of the label information through the deconvolution layers to improve the sample quality. In DDN, the true/false HSI samples learn spatial and spectral features separately in a dual-channel style, during which the spatial-spectral cascaded attention mechanism effectively facilitates the feature mining. During the competitive iteration in CDGAN, the task of classifying HSI under small samples is accomplished. The experimental results demonstrate the effectiveness of the proposed method, which has better classification results compared with the other five algorithms in this paper.

REFERENCES

[1] Cheng, G., Li, Z., Han, J., Yao, X. and Guo, L., "Exploring hierarchical convolutional features for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens. 56, 6712-6722 (2018).

[2] Gong, Z., Zhong, P., Yu, Y., Hu, W. and Li, S., "A CNN with multiscale convolution and diversified metric for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens. 57, 3599-3618 (2019).

[3] He, N., et al., "Feature extraction with multiscale covariance maps for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens. 57, 755-769 (2019).

[4] Haut, J. M., Paoletti, M. E., Plaza, J., Li, J. and Plaza, A., "Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach," IEEE Trans. Geosci. Remote Sens. 56, 6440-6461 (2018).

[5] Roy, S. K., Krishna, G., Dubey, S. R. and Chaudhuri, B. B., "HybridSN: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification," IEEE Geosci. Remote Sens. Lett. 17, 277-281 (2020).

[6] Cao, X., Yao, J., Xu, Z. and Meng, D., "Hyperspectral image classification with convolutional neural network and active learning," IEEE Trans. Geosci. Remote Sens. 58, 4604-4616 (2020).
[7] Zheng, Z., Zhong, Y., Ma, A. and Zhang, L., "FPGA: Fast patch-free global learning framework for fully end-to-end hyperspectral image classification," IEEE Trans. Geosci. Remote Sens. 58, 5612-5626 (2020).
[8] Goodfellow, I., et al., "Generative adversarial nets," Proc. NIPS, Montreal, QC, Canada, 2672-2680 (2014).
[9] Melgani, F. and Bruzzone, L., "Classification of hyperspectral remote sensing images with support vector machines," IEEE Trans. Geosci. Remote Sens. 42, 1778-1790 (2004).
[10] Roy, S. K., Krishna, G., Dubey, S. R. and Chaudhuri, B. B., "HybridSN: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification," IEEE Geosci. Remote Sens. Lett. 17, 277-281 (2019).
[11] Zhu, L., Chen, Y., Ghamisi, P. and Benediktsson, J. A., "Generative adversarial networks for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens. 56, 5046-5063 (2018).
[12] Zhong, Z., Li, J., Clausi, D. A. and Wong, A., "Generative adversarial networks and conditional random fields for hyperspectral image classification," IEEE Trans. Cybern. 50, 3318-3329 (2020).
[13] Wang, J., Guo, S., Huang, R., Li, L., Zhang X. and Jiao, L., "Dual-channel capsule generation adversarial network for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens. 60, 5501016 (2022).