Semisupervised Hyperspectral Image Classification With Cluster-Based Conditional Generative Adversarial Net

Wenzhi Zhao, Xuehong Chen, Yanchen Bo, and Jiage Chen

Abstract— **Hyperspectral image classification is a challenging task when a limited number of training samples are available. It is also known that the classification performance highly depends on the quality of the labeled samples. In this work, a cluster-based conditional generative adversarial net (CCGAN) is proposed as an effective solution to increase the size and quality of the training data set. The proposed method is able to automatically select the most representative initial samples with a subtractive clustering-based strategy, which keeps the diversity for sample generation. Moreover, compared to the traditional semisupervised classification frameworks, the CCGAN is able to generate realistic spectral profiles by considering the class-specific labels. Experiments on well-known Pavia University data set demonstrate that the proposed CCGAN can significantly boost the classification accuracy, even using a small number of initial labeled samples.**

Index Terms— **Generative adversarial nets (GANs), hyperspectral images, image classification, semisupervised learning (SSL).**

I. Introduction

With the fast development of remote sensing technology, both the spectral and spatial resolution of remote sensing image has been significantly improved. In this regard, hyperspectral imagery may contain hundreds of spectral bands whilst the spatial resolution can be as high as submeter for each pixel. Therefore, fine-scale hyperspectral images play an important role in urban mapping, environmental management, crop analysis, and mineral detection [1], [2]. The accurate identification of the semantic label of each pixel is a prerequisite to realizing these applications. However, with the increased spectral resolution of hyperspectral images, the difficulty of image interpretation also has significantly elevated.

In general, there are two obstacles that prevent efficient hyperspectral imagery interpretation. On the one hand, the ground truth data are quite scarce due to the expensive and time-consuming process of human-based data labeling. Given the limited number of training samples, the classification performance decreases as the dimensionality of hyperspectral image increases, which is called the "Hughes phenomenon" [3]. On the other hand, when the spatial resolution gets finer, the intraclass variation introduced by the complex textures and illumination changes also have dramatically increased. Different targets may share similar spectral properties that have greatly challenged the classification. In order to decrease variability from the same class and increase the separability between similar classes, it is important to select the most representative samples [4]–[6]. Thus, the quality of training samples also plays an important role in image classification, as it provides the separable boundary information in the feature space [7], [8].

Hyperspectral images contain hundreds of continuous narrow spectral bands in the electromagnetic spectrum. Due to the similarity between adjacent pixels, the randomly selected samples may highly redundant in the spectral domain. Under such circumstance, classifiers usually face the challenges that come from the overfitting phenomenon. Moreover, redundant samples provide highly correlated information that impacts the performance of sample augmentation. In order to handle these problems, there are two ways to find the most representative samples, namely, unsupervised and semisupervised methods [9]. The unsupervised methods can reveal data structure of samples without any label information. It aims to identify potential cluster centers to find the most representative samples, such as k-means and subtractive clustering. Therefore, unsupervised methods focus on finding homogeneous clusters exploration on limited training samples which can keep the diversity of samples. Complementary, semisupervised learning (SSL) techniques offer a way to generate new training samples from unlabeled samples. The SSL methods try to find the most representative class-specific samples from both labeled and unlabeled samples, such as active learning (AL) [10]–[13] and generative adversarial net (GAN) [14]. However, existing methods mainly focus on exploring surface-type samples from the existing data pool which locks the diversity of selected samples. Therefore, how to find representative samples whilst keep them in diversity are the most challenging tasks in the field of hyperspectral image classification. Recently, few research works have focused on GAN and its abilities in diversified samples generation [15], [16]. Still, the diversity of input samples is limited and it lacks class-specific structures that

Manuscript received March 13, 2019; revised April 28, 2019; accepted June 18, 2019. Date of publication July 3, 2019; date of current version February 26, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC1508903, in part by the China Postdoctoral Science Foundation under Grant 2018M640087 and Grant 2019T120063, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018NTST01. *(Corresponding author: Yanchen Bo.)*

The authors are with the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China, and also with the Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China (e-mail: boyc@bnu.edu.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2019.2924059

could further boost the hyperspectral image classification accuracy.

In order to accurately classify hyperspectral images, a cluster-based sample generation method is proposed to increase the number of training samples in Fig. 1. Different from the commonly used data argumentation strategies, an unsupervised-based subtractive clustering is applied to reduce the redundancy inside labeled training samples. Based on unsupervised measurements, the most representative samples can be automatically selected. Then, we propose a cluster-based conditional GAN (CCGAN) framework to combine the idea of conditional GAN and SSL strategy for class-specific sample generation. The objective of CCGAN is to generate labeled samples based on the limited number of training samples. The contributions of this study are as follows.

1) The subtractive clustering strategy is applied to reduce sample redundancy and find the most representative samples for data structure exploration.
2) The CCGAN is proposed to automatically generate the class-specific samples in the spectral domain.
3) An SSL framework is proposed to force CCGAN iteratively updates generated samples for efficient hyperspectral image classification.

The rest of this letter is organized as follows. Section II presents the proposed cluster-based spectral-spatial sample generation method. Section III presents the data sets, while the experimental results are discussed in Section IV. Final conclusions are given in Section V.

II. Methodology

A. Subtractive Clustering for Sample Selection

Due to the similarity of adjacent image pixels, subtractive clustering strategy is applied to reduce data redundancy and improve the representativeness of the selected samples. In subtractive clustering, each individual pixel is regarded as a potential cluster center. Suppose the sample is $x_i, i \in (1, \ldots, n)$ and each contains h bands $x \in R^h$ that represents the spectral data. The potential measurements of each sample are calculated as

$$F(x_i) = \sum_{j=1}^{n} e^{(-\alpha||x_i - x_j||^2)} \tag{1}$$

where x_i represents a hyperspectral pixel and α is a positive distance factor that determines the size of the neighborhood for a cluster. By using this potential function, the maximum value for a sample x^{c-1} in the training data set can be represented as

$$F_1^{\max} = \max F(x^{c-1}). \tag{2}$$

The $F_1^{\max}$ indicates the potential function of the first cluster center x^{c-1}. The following cluster centers can be calculated using the following equation:

$$F_c(x_i) = F_{c-1}(x_i) - F_{c-1}^{\max} e^{(\alpha||x^{c-1} - x_i||^2)} \tag{3}$$

where $F_{c-1}(x_i)$ is the previous potential function, and $F_c(x_i)$ is the new potential function and x^{c-1} is the last found cluster center. This iteration continues to a predefined number of clusters. For a specific class, there are c clusters are obtained by subtractive clustering. Each cluster shares similar spectral properties according to the potential measurements. Therefore, in order to find the most representative samples, N samples are randomly selected from c clusters.

B. Semisupervised Learning With CCGAN Framework

Although traditional GAN is widely used in spectral-spatial sample generation, still it only exploits data structure under the real/fake condition without any class-specific information. To remedy this, we introduced a CCGAN to generate class-specific samples to achieve the purpose of improved classification. Different from traditional GAN, the CCGAN requires an additional term $y, (y \in Y)$ as a condition for sample generation. Here, Y is the embedding space used to condition the generator that drawn from the training data set. To be specific, the generator G defines a conditional density model $p_g(x|y)$ to replicate the empirical distribution of training samples $p_d(x, y)$. The objective of generator G can be formulated as

$$p_g(x, y) = p_g(x|y)p_y(y) \approx p_d(x, y). \tag{4}$$

For the spectral information generation, the 1-D deconvolution neural network is applied to generate realistic samples. In the hyperspectral image, suppose there are b spectral bands. For an image pixel i, the spectral profile of selected samples can be represented as vector x_i^b. Conventionally, GAN framework contains two contradictory players, namely, a generator and a discriminator, and they are positioned in adversarial game. The generator G is designed to generate realistic samples to fool discriminator; at the same time, the discriminator D is tasked to distinguish samples whether from the training data or the generator. The objective of GAN can be formulated as a minimax value function:

$$\min_G \max_D (E_{x \sim p_d(x)}[\log \mathrm{D}(x)]) + (E_{z \sim p_z(z)}[\log \mathrm{D}(G(z))]) \tag{5}$$

where z is a noise space used to initialize the generator, values $z \in Z$ are sampled from the noise distribution. $x, (x \in X)$ represents the samples selected by subtractive clustering or outputs from the generator.

For a well-trained generator G with L layers, the input noise z is transformed into generated spectral samples x_{gen} for a specific label y

$$x_{\text{gen}} = h_L(z|y). \tag{6}$$

Suppose we have a batch of real samples $(x_i, y_i)_{i=1}^n$ that x_i paired with condition label y_i. Let $z_i \sim p_z(z)$ be the noise data generated from the noise distribution. The loss function for the generator can be formulated by the discriminator

$$l_G = -E_{z,y} \log D(G(z_i, y_i)). \tag{7}$$

For CCGAN, the discriminator can be regarded as a combination of unsupervised classifier D_{unsup} and supervised classifier D_{sup}. For the unsupervised classifier D_{unsup} from CCGAN, it is trained to discriminate fake samples from the real ones. If the semisupervised problem with k classes, the discriminator has $k + 1$ outputs where the class $k + 1$ represents the fake

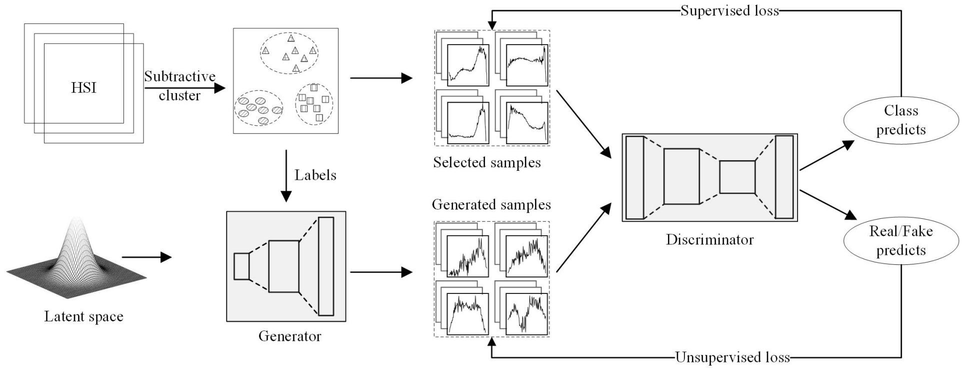

Fig. 1. Workflow of CCGAN and semisupervised hyperspectral image classification.

samples generated from the generator. Thus, the loss function of the unsupervised classifier can be formulated as

$$l_{D_{\text{unsup}}} = -E(x \sim p_g(x)\log(D_{\text{unsup}}(y = k+1|x)) - E(x \sim p_d(x))\log(1 - D_{\text{unsup}}(y = k+1|x)). \quad (8)$$

The term $D_{\text{unsup}}(y = k+1|x)$ is the probability of x being a fake example, and $1 - D_{\text{unsup}}(y = k+1|x)$ is the probability of x being a real example. The unsupervised loss term l_{unsup} is the same as the regular discriminator loss except that it measures the discrepancies between the real and fake samples. In addition, for the supervised term of the discriminator, the loss function can be represented as

$$l_{D_{\text{sup}}} = -E(x, y)\log(D_{\text{sup}}(y|x, y \leq k)). \quad (9)$$

The supervised loss term is a log conditional probability for labeled samples which is a standard cost as in supervised learning setting. In general, the final objective function for SSL of CCGAN is

$$L = \min_G \max_D (l_G + l_{D_{\text{unsup}}} + l_{D_{\text{sup}}}). \quad (10)$$

For the SSL of CCGAN, the objective function consists of three terms. The first term is to minimize the log conditional probability for the generated samples. The second term is to maximize the log probability of the generated samples. The third term is to maximize the log conditional probability for labeled data, which is the standard cost as in supervised learning formulation. When the training process finished, the discriminator from CCGAN can be used to classify samples for SSL.

III. Experiment

To illustrate the sample generation ability of the proposed CCGAN, the experiments are conducted on the well-known hyperspectral image data sets for algorithm evaluation. The Pavia University data set was acquired by the ROSIS sensor during a flight campaign over Pavia, Northern Italy. The ROSIS sensor provides 115 spectral bands within spectral ranges from 0.43 to 0.86 μm. After removing noise-effected bands, the remaining 102 spectral bands are available for our experiment. The sizes of this data set are 610 × 340 pixels with the spatial resolution of 1.3 m per pixel. There are nine classes of land cover types inside of this data set. A 10% of the labeled samples are randomly selected for CCGAN training and another 10% samples are randomly selected for evaluation.

TABLE I
Detailed Information About Configuration of the CCGAN

Name	Layer	Kernel	Stride	Features	Activation
G	1	1×7	1	128	ReLu
	2	1×5	2	256	ReLu
	3	1×10	2	256	ReLu
	4	1×10	3	128	tanh
	5	1×20	1	1	ReLu
D	1	1×10	1	50	ReLu
	2	1×10	2	100	ReLu
	3	1×10	2	200	tanh
	4	1×3	1	50	tanh
	5	1×3	1	*c*/1	tanh

A. Experimental Setting

For the purpose of spectral profile generation, the CCGAN framework is applied to mimic spectral distribution pattern from the real spectral profiles. To capture the spectral distribution pattern, we developed a 1-D generator to formulate the real distribution of spectral profiles. Specifically, the deconvolutional neural network can automatically convert 1-D noises into realistic spectral profiles given the condition y. The detailed information about CCGAN configuration is listed in Table I. In order to highlight the effectiveness of the CCGAN model, we included one-dimensional CNN (1D-CNN), two-dimensional CNN (2D-CNN), and the traditional GAN models for comparison. For 1D-CNN, it contains an input layer, a convolutional layer with the sizes of 11 × 1, a max-pooling layer and a fully connected layer for classification. The 2D-CNN consists of three convolutional layers and two-pooling layers, and the traditional GAN is similar to CCGAN, except the conditional term in the generator.

B. CCGAN Performances and Analysis

The CCGAN is able to generate realist samples when the optimal status is achieved during the training process. In the final state, the generator can produce realist fake samples to deceive the discriminator who tries to pick up the fake ones. In this section, we take an insight look at the generator and discriminator, to better understand the mechanism that working behind the CCGAN. To serve this purpose, the quality of generated samples, representative of feature projections and loss function optimizations are included.

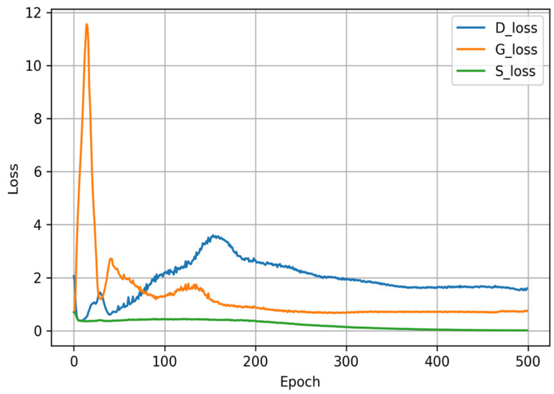

Fig. 2. Loss function optimization during training process. D_loss and G_loss represent the error rates for unsupervised discrimination. S_loss means supervised loss function values.

1) Loss Function Optimization: To train the CCGAN, there are three loss functions to be minimized. During the training process, the iteration is set to 500 times, and the learning rates are 0.001 and 0.0001 for discriminator and generator, respectively. The curves of loss function optimization are shown in Fig. 2. In general, the discriminator and generator demonstrate the contradictory pattern to each other. The generator is harder to generate realist samples when discriminator error rates are low. Similarly, a well-trained generator is capable to fool the discriminator with realist fake samples. In the end, the discriminator and generator are stabilized with enough training iterations. Meanwhile, for the supervised loss function, the error rates keep dropping during the entire training process. Supervised loss function and discriminator loss function share the same parameter when optimizing CCGAN. In this way, the CCGAN is able to learn both real data distributions and class-specific patterns for semisupervised classification.

2) Sample Visualization and Analysis: For better understanding the sample generation power of the CCGAN, we visualized the spectral profiles for three specific classes, namely, Asphalt, Meadows, and Gravel, as shown in Fig. 3. Different from the previous studies on hyperspectral sample generation, the CCGAN is designed to mimic real data distribution under class-specific conditions. From the figure, we conclude that the CCGAN is able to capture class-specific spectral profiles. For example, the generated meadow profiles have a low reflectance band around 60 and a high reflectance band around band 80, which is similar to the original data. However, different from the original smooth spectral profiles, the generated samples have random variations across the entire spectral bands. The reason for fuzzy prediction is that the CCGAN can only capture the spectral pattern in general sense, but overlooked the spectral continuity between adjacent bands. In general, the well-trained CCGAN has great ability to mimic the data distribution patterns on 103 spectral bands. Thus, it is possible to utilize CCGAN to enrich training samples and improve the classification accuracy.

To quantitatively measure the quality of generated samples, we projected training samples into 2-D feature space, as shown in Fig. 4. For the original spectral profiles, 10% of

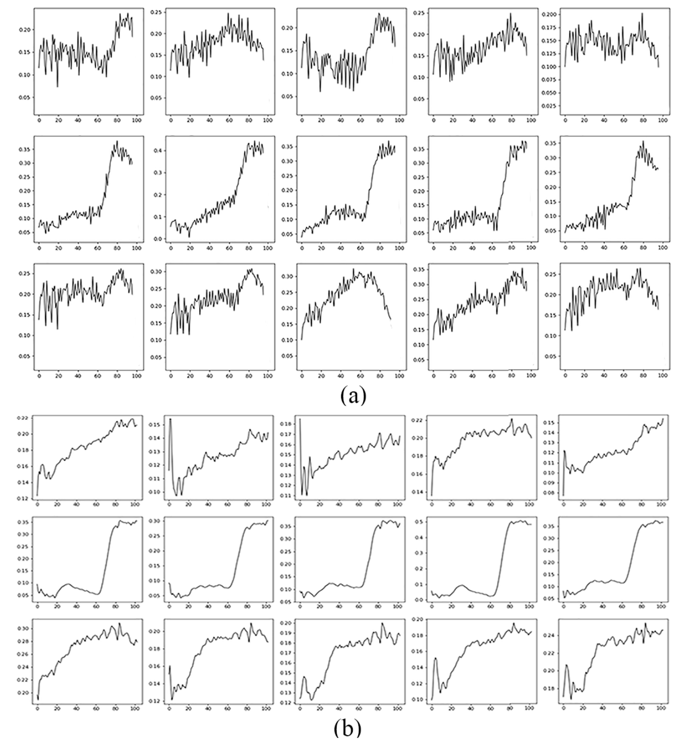

Fig. 3. (a) Generated spectral profiles against (b) original spectral data for class Asphalt, Meadows, and Gravel, respectively.

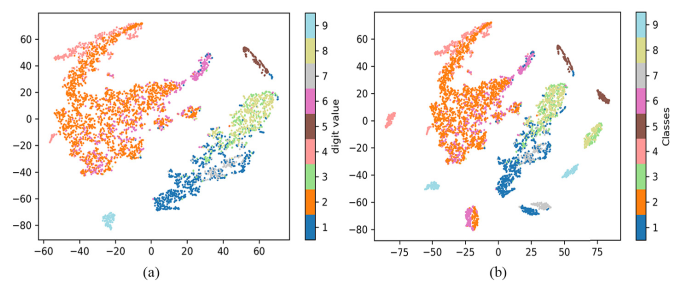

Fig. 4. Projection of the (a) cluster-based samples and (b) enriched samples on Pavia University data set. Different colors represent different classes, respectively. For clarity, we map the 10% of training samples and 200 generated samples on 2-D space using t-SNE embedding.

training samples were randomly selected and projected into lower dimensional space, as demonstrated in Fig. 4(a). The separation ability of the original data set is limited since the selected samples share similar spectral profiles, such as class 8 (bricks) and class 3 (gravel). Moreover, the imbalanced selected samples further decrease the separability of the training data set. To solve this problem, the CCGAN generates additional samples to make the training data set more balanced. In this experiment, we generated two hundred additional samples for each class to increase the separability of the training data set. For a better understanding of the generated samples, we projected the enriched training samples into two dimension space, as demonstrated in Fig. 4(b). From this figure, the generated samples are capable to increase the separability of training data set by filling up the feature space effectively. In addition, the enriched data set can boost

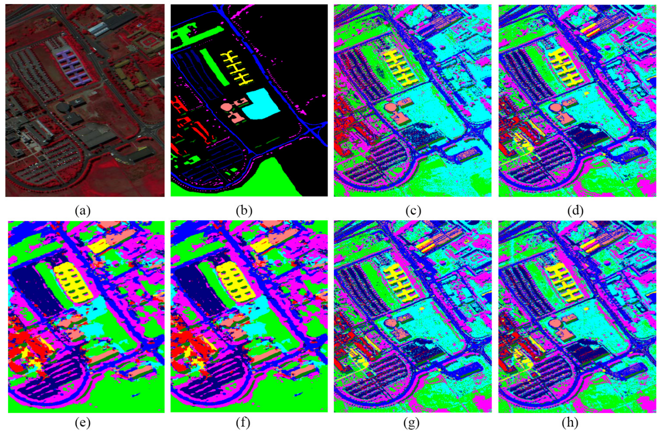

Fig. 5. Classification maps of the Pavia University scene with different strategies. (a) Original data. (b) Reference map. (c) 1D-CNN. (d) AL. (e) 2D-CNN. (f) 3D-CNN. (g) GAN. (h) CCGAN.

TABLE II
CLASSIFICATION RESULTS OF DIFFERENT MODELS ON THE PAVIA UNIVERSITY DATA

Class	1D-CNN	AL	2D-CNN	3D-CNN	GAN	CCGAN
1	80.94	80.85	77.39	83.37	78.06	90.32
2	70.37	80.06	98.89	96.88	90.68	96.14
3	77.32	85.64	56.74	78.29	81.95	82.03
4	85.93	86.43	92.75	93.71	84.79	95.76
5	99.70	89.97	99.78	99.53	97.39	99.01
6	93.26	92.78	47.27	75.36	70.34	83.64
7	95.41	92.59	80.08	88.64	80.91	90.51
8	84.47	86.76	96.69	97.05	87.25	97.28
9	92.08	90.74	96.30	92.86	92.43	94.37
OA	79.55	77.43	86.18	87.94	85.81	92.48
AA	86.61	85.89	82.88	82.37	80.67	89.74
Kappa	74.28	70.39	81.22	80.62	79.35	88.35

the classification performances by increasing the number of training samples.

C. Classification Results and Comparison

To demonstrate the performance of CCGAN, we compared the proposed method with other state-of-the-art methods. The classification maps are shown in Fig. 5, and the detailed information about classification accuracies are reported in Table II. For 1D-CNN, 2D-CNN, and 3D-CNN, the spectral bands were directly fed into the deep learning framework. However, due to the limited number of training samples, the overall accuracy of such frameworks only reach 79.55%, 86.18%, and 87.94%, respectively. In order to increase the number of training samples, the traditional GAN utilizes a generator to convert noises into unlabeled samples. Although the traditional GAN significantly increased the number of training samples, it is hard to explore class-specific deep features. Thus, the CCGAN achieves the best classification results by generating realistic class-specific training samples.

IV. CONCLUSION

In this letter, we proposed a CCGAN for hyperspectral image classification. Compared to the traditional GAN model, our proposed method can generate realistic samples for each specific class. In general, there are two merits for this CCGAN method: 1) it can automatically explore the most representative features for sample generation and 2) it utilizes the conditional GAN to generate realistic samples for hyperspectral image classification. Based on these merits, we tested the CCGAN on a well-known hyperspectral data set, and the classification accuracies are significantly higher than the traditional ones. In the future, how to reduce the noises for generated spectral profiles is a question worth exploring.

REFERENCES

[1] L. Zhang, L. Zhang, D. Tao, and X. Huang, “On combining multiple features for hyperspectral remote sensing image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.

[2] Z. Wang, B. Du, L. Zhang, L. Zhang, and X. Jia, “A novel semisupervised active-learning algorithm for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3071–3083, Jun. 2017.

[3] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, “Hyperspectral image unsupervised classification by robust manifold matrix factorization,” *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019.

[4] X. Guo, X. Huang, L. Zhang, L. Zhang, A. Plaza, and J. A. Benediktsson, “Support tensor machines for classification of hyperspectral remote sensing imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3248–3264, Jun. 2016.

[5] W. Zhao, S. Du, and W. J. Emery, “Object-based convolutional neural network for high-resolution imagery classification,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3386–3396, Jul. 2017.

[6] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, “Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.

[7] B. Ayerdi, I. Marqués, and M. Graña, “Spatially regularized semi-supervised ensembles of extreme learning machines for hyperspectral image segmentation,” *Neurocomputing*, vol. 149, pp. 373–386, Feb. 2015.

[8] W. Zhao, W. J. Emery, Y. Bo, and J. Chen, “Land cover mapping with higher order graph-based co-occurrence model,” *Remote Sens.*, vol. 10, no. 11, p. 1713, 2018. [Online]. Available: http://www.mdpi.com/2072-4292/10/11/1713

[9] H. Wu and S. Prasad, “Semi-supervised deep learning using pseudo labels for hyperspectral image classification,” *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.

[10] S. Samiappan and R. J. Moorhead, “Semi-supervised co-training and active learning framework for hyperspectral image classification,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 401–404.

[11] Z. Zhang, E. Pasolli, M. M. Crawford, and J. C. Tilton, “An active learning framework for hyperspectral image classification using hierarchical segmentation,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 640–654, Feb. 2016.

[12] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, “Manifold regularized sparse NMF for hyperspectral unmixing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[13] Z. Wang, B. Du, L. Zhang, and L. Zhang, “A batch-mode active learning framework by querying discriminative and representative samples for hyperspectral image classification,” *Neurocomputing*, vol. 179, pp. 88–100, Feb. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231215018676

[14] Z. He, H. Liu, Y. Wang, and J. Hu, “Generative adversarial networks-based semi-supervised learning for hyperspectral image classification,” *Remote Sens.*, vol. 9, no. 10, p. 1042, Oct. 2017.

[15] Y. Zhan, D. Hu, Y. Wang, and X. Yu, “Semisupervised hyperspectral image classification based on generative adversarial networks,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 212–216, Feb. 2018.

[16] 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.