Semisupervised Variational Generative Adversarial Networks for Hyperspectral Image Classification

Chao Tao, Hao Wang, Ji Qi, and Haifeng Li

***Abstract*—Although the hyperspectral image (HSI) classification is extensively investigated, this task remains challenging when the number of labeled samples is extremely limited. In this article, we overcome this challenge by using synthetic samples and proposing semisupervised variational generative adversarial networks (GANs). In contrast to conditional GAN (previously used for the generation of HSI samples), the proposed approach has two novel aspects. First, an encoder-decoder network is extended to the semisupervised context using an ensemble prediction technique. Through this technique, our deep generative model can be trained using limited labeled samples (only five samples per class) with a large number of unlabeled samples. Second, we build a collaborative relationship between the generation network and the classification network. This property enables that our model can produce meaningful samples that can contribute to the final classification. The experiments on four benchmark HSI datasets demonstrate that the proposed model can achieve an increase of $>10\%$ in overall classification accuracy compared with the baseline model without using the generated sample. We also show that the proposed model can achieve better and more robust performance for HSI classification than other generative methods as well as semisupervised methods, especially when the labeled data are limited.**

***Index Terms*—Deep learning, generative adversarial networks (GAN), hyperspectral image classification, semisupervised learning, variational auto-encoder (VAE).**

I. Introduction

HYPERSPECTRAL imaging sensors can capture images with pixels represented as high-dimensional spectral vectors that range from visible spectral to short-wave infrared bands [1]. These sensors enable hyperspectral image (HSI) to be a powerful tool for various applications, such as land cover classification [2], [3], environmental protection [4], [5] and precision agriculture [6], [7].

Methods used for HSI classification can be roughly divided into three types according to the ratio of label samples in all training samples: unsupervised, supervised learning, and semisupervised. Unsupervised learning algorithms can be easily applied in the HSI processing area because no labeled samples are required. Many methods, such as fuzzy clustering [8] and fuzzy C-Means [9], have demonstrated impressive results in HSI classification. However, verifying the relationship between clusters and classes based on limited priori knowledge is difficult. Supervised learning classification mainly includes support vector machine (SVM) [10], [11] and random deep forest [12], [13]. However, obtaining a large number of labeled samples for these supervised learning methods is time-consuming and costly in HSI [14]. Therefore, the semisupervised learning method has extensively attracted the attention of researchers. Its basic idea is based on the assumption of data distribution, which uses a small number of labeled samples and a large number of unlabeled samples to build a high-performance learner for image classification. Commonly used semisupervised learning methods include the generated model [15], the transductive SVM method [16], the self-learning method [17], [18] [19], and the graph-based method [20], [21], [22]. However, such algorithms are less tolerant by unlabeled samples. Moreover, if the unlabeled samples are not properly selected, they may induce the classifier to make mistakes.

In this context, we focus on two main challenges for HSI classification. First, due to the high dimensionality of HSIs, general optical image classification methods in the computer vision domain fail to obtain refine classification maps of HSIs. Second, it is time-consuming and costly to obtain a large number of labeled samples for HSIs. With the development of imaging technologies, hyperspectral sensors have provided fine spatial resolution as well as detailed spectral information. Convolutional neural networks (CNNs), especially deep CNNs has proven to be promising in extracting hierarchical features. A series of end-to-end CNN models have been proposed to extract features of HSIs and achieved promising results [23]–[26]. These methods are based on the assumption that deep models require a large number of training data. Nonetheless, these papers overlook the scarcity of labeled training samples in HSIs. On the other hand, a large number of unlabeled samples contain efficient information to exploit. Several semisupervised learning methods [20], [22] suggest to exploit information contained in real data samples, included unlabeled samples and limited labeled samples. Although they achieved good performance, these results may heavily depend on unlabeled data selection rather than models. Recently, a generative adversarial network (GAN) [27] was proposed to capture the data distribution by playing a Nash equilibrium game between a generative network G and a discriminative network D. With the help of adversarial learning, GAN can create an efficient generator to synthesize

Manuscript received January 4, 2020; revised February 7, 2020; accepted February 9, 2020. Date of publication February 24, 2020; date of current version March 17, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 41771458, Grant 41301453, and Grant and 41871364, in part by the Natural Science Foundation of Hunan Province under Grant 2017JJ3378 and in part by the Fundamental Research Funds for the Central Universities of Central South University under Grant 1053320182684. *(Corresponding author: Haifeng Li.)*

The authors are with the School of Geosciences and InfoPhysics, Central South University, Changsha 410083, China (e-mail: kingtaochao@126.com; haowang@csu.edu.cn; erenturing@csu.edu.cn; lihaifeng@csu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.2974577

higher quality samples in an unsupervised manner. Such virtual samples may enrich the dataset and introduce variability. Therefore, more efficient semisupervised training strategies based on GANs for HSI classification are valuable to explore.

Due to the advantages of GAN, some approaches tried to solve the above problem based on GANs. The basic idea is to apply such deep generative models to generate realistic and diverse samples and then work for a supervised classification with a small number of labeled samples and a large number of newly generated samples. In [28], a 3D-GAN is proposed to process image classification tasks as well as synthetic data. However, it requires a certain number of labeled samples (e.g., 200 samples per class) for adversarial learning, which ignores to exploit the information contained in unlabeled samples. The aforementioned methods in [29] and [30] proposed semisupervised GANs for HSI classification based on conditional GANs. However, conditional GANs model [31], [32] has the problem of unstable training [33], gradient vanishing [34], and mode collapse [35]. Although [30] proposes to use conditional random field to smooth the semisupervised process, the instability of the training process of conditional GANs may influence the quality of the generated samples.

In addition to GANs, variational auto-encoder (VAE) [36] can also generate samples with the smooth training process, while its imperfect reconstruction loss usually causes the generated image to be blurry [37]. To improve the quality of the generated samples, and to make the training of GAN more stable, conditional variational autoencoder generative adversarial network (CVAEGAN) [38] takes advantage of the combination of VAE and GAN, and applies feature matching objective for conditional adversarial learning to synthesize images for a specific identity. However, CVAEGAN does not clearly build the relationship between the generative network G and the classification network C in model training stage, which may lead to generate just realistic but not enough meaningful samples for learning classifiers. Besides, it is a supervised generative model. When there are not enough labeled samples for training, the variety of synthesized samples may be degraded, and thus this property may greatly restrict the model in limited labeled data condition such as hyperspectral image classification application.

In this article, we further explore the potential of using GANs for HSI classification under the extremely limited labeled condition (e.g., five samples per class), and proposed a semisupervised variational GAN to solve the problems mentioned above. In contrast with CVAEGAN, the proposed framework is extended to the semisupervised context by using an ensemble prediction technique [39] to ensure that the deep generative model can be trained with limited labeled samples. Besides, we build a collaborative optimization mechanism between the generative network G and the classification network C in the process of sample generation. This property enables our model to produce meaningful samples which can contribute to the final classification. From the experiments on three widely used hyperspectral data sets and one Landsat remote sensing image, we observed that the proposed model can generate diverse but category-keeping samples using only five labeled samples per class and a large number of unlabeled samples. Such samples enable the introduction of more variability and can possibly enrich the dataset. As a result, the classification performance is significantly improved compared to the baseline without using the generated samples.

The rest of this article is organized as follows. Section II describes the background of CVAEGAN. Then, Section III presents the proposed semisupervised variational generative method and the details of using adversarial samples for HSI classification. The experimental results are shown in Section IV. Finally, Section V concludes this article.

II. Background

CVAEGAN is a generative framework that takes advantage of the combination of VAE and GAN, and applies feature matching objective for conditional adversarial learning to synthesize images for a specific identity. This framework [Fig. 1(a)] contains four parts: 1) an encoder network E to learn the relationship between the latent space and the real image space; 2) a generative network G to synthesize samples; 3) a discriminative network D to distinguish between real and synthesized samples; 4) a classification network C to measure class probabilities for real images.

Mathematically, suppose we have a dataset x, which is parametrized by latent factors z. The encoder network E maps inputs to a latent variable space with a specific distribution, unit Gaussian $\mathcal{N}(0,1)$, and outputs the mean and covariance of the latent vector in practice, i.e., μ and ϵ, respectively. The gap between prior $p(z)=\mathcal{N}(0,1)$ and posterior $q_\phi(z|x)$ can be reduced by KL loss

$$\mathcal{L}_{KL}=\frac{1}{2}(\mu^T\mu+\text{sum}(\exp(\epsilon)-\epsilon-1)). \tag{1}$$

Then, a mapping from x to latent variables z is obtained by sampling $z=\mu+\boldsymbol{r}*\exp(\epsilon)$, where $r\sim\mathcal{N}(0,I)$. After that, CVAEGAN reconstructs the latent variables to the original image space by conditional adversarial learning. The discriminative network D tries to distinguish real data from synthesized data. The generative network G tries to fool the network D, and the classification network C tries to introduce specific labeled information for generating process.

Suppose we have auxiliary information c, here, c could be any kind of available extra information, e.g., class labels. Classification network C takes in x as input and outputs a corresponding predicted vector. The output can represent the posterior probability $P(c|x)$ by using softmax function. And the network C tries to minimize the following loss function

$$\mathcal{L}_{C_0}=-\mathbb{E}_{x\sim P_{\text{data}}(x)}[\log P(c|x)]. \tag{2}$$

Generative network G accepts latent variables z and auxiliary information c as input to generate new samples $G(z,c)$. Discriminative network D is a binary classifier, which is performed to maximize $\log(D(x))$ and minimize $\log(D(G(z,c)))$

$$\begin{aligned}\mathcal{L}_D=&-\mathbb{E}_{x\sim P_{\text{data}}(x)}[\log D(x)]\\&-\mathbb{E}_{z\sim P_z(z)}[\log(1-D(G(z,c)))].\end{aligned} \tag{3}$$

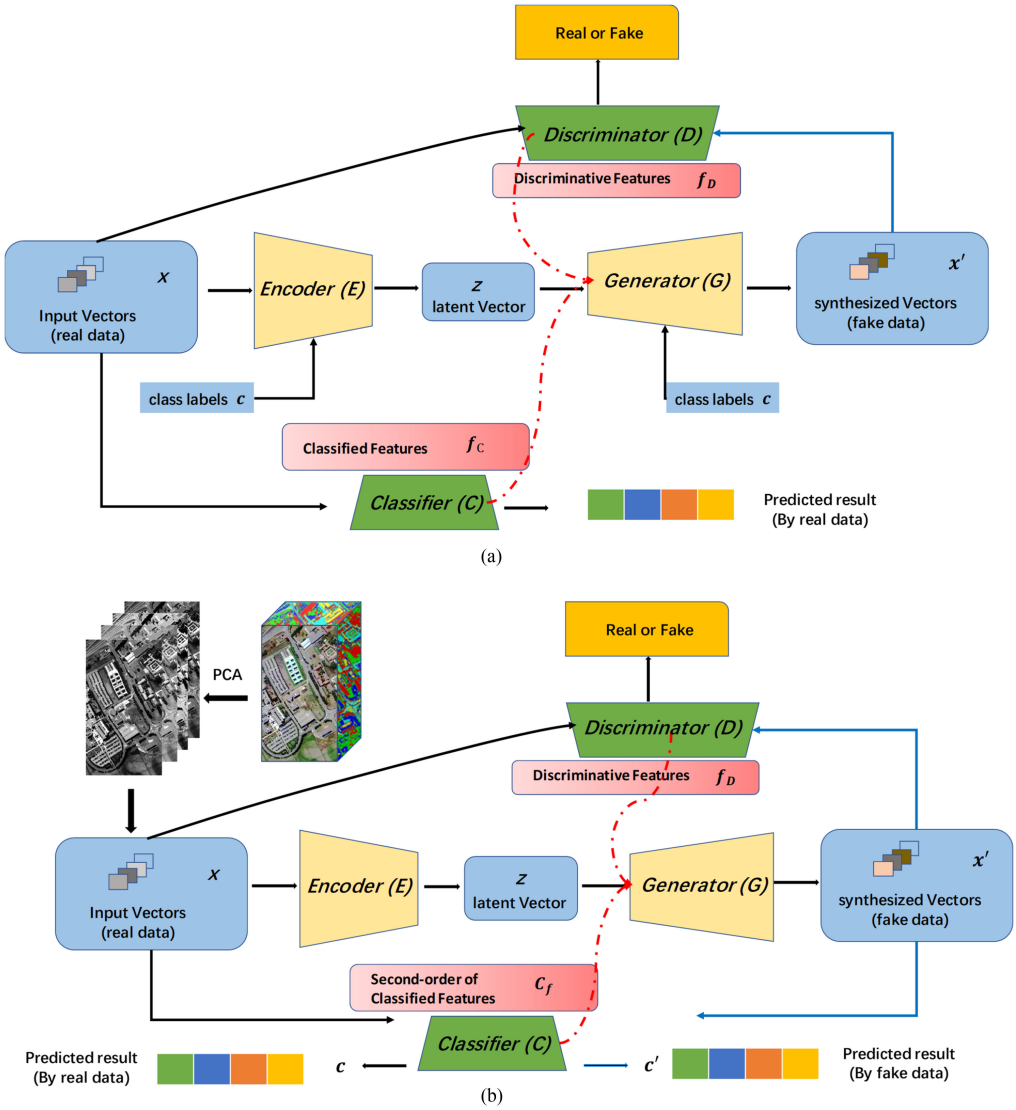


Fig. 1. Illustration of the structure of CVAEGAN and the proposed SSVGAN. The encoder network E maps the input x to a latent vector z. The generator G generates fake data x by sampling from a distribution $P(x|z)$, and tries to learn the real data distribution by the gradients given by the discriminator D and the classifier C. The encoder-decoder framework sets up the relationship between the latent space and real data space. The variability in the synthesized samples can be obtained by imposing Gaussian noise with the latent vector.

To avoid training instability problem and synthesize sharper samples [40], CVAEGAN applies feature matching in the implementation of conditional adversarial learning. Concretely, the network G tries to minimize the following loss function:

$$\mathcal{L}_{G_0} = \frac{1}{2}\Big(\| x - G(z,c) \|_2^2 + \| f_D(x) - f_D(G(z,c)) \|_2^2 + \| f_C(x) - f_C(G(z,c)) \|_2^2 \Big) \tag{4}$$

where f_D and f_C are the features of an intermediate layer of discriminative network D and classification network C, respectively.

III. Methodology

CVAEGAN is capable of generating realistic and diverse samples with fine-grained category labels, such as faces or flowers. However, it has two shortcomings. First, it needs a lot of training samples to train the network while the labeled data is usually limited in HSI application. Second, it just considers how to generate realistic samples but does not clearly build the relationship between the generative network G and the classification network C. Thus, in this article, we propose a semisupervised variational GAN called semi-supervised variational generative adversarial network (SSVGAN) [Fig. 1(b)] to solve these problems. The architecture of the proposed network SSVGAN (shown in Table I) is as similar as CVAEGAN which also contains four parts, but we use different training strategy and optimization mechanism. In the following sections, we begin by describing the semisupervised training framework of SSVGAN (Section III-A). Then we proposed a collaborative optimization mechanism between the generative network G and the classification network C in model training stage (Section III-B). We finally show how to utilize

TABLE I
NETWORK OF THE PROPOSED SSVGAN

Nets	No.	Layer	BN	Stride	Activation
E	1	Conv:$5*5*32$	NO	2	ReLU
	2	Conv:$5*5*128$	YES	2	ReLU
	3	Conv:$5*5*256$	YES	2	ReLU
	4	FC:2048	YES	-	ReLU
	5	FC:20	NO	-	NO
G	1	Conv transpose:$5*5*128$	YES	2	ReLU
	2	Conv transpose:$5*5*32$	YES	2	ReLU
	3	Conv transpose:$5*5*10$	NO	2	Tanh
D	1	Conv:$5*5*32$	NO	2	LeakyReLU
	2	Conv:$5*5*128$	YES	2	LeakyReLU
	3	Conv:$5*5*256$	YES	2	LeakyReLU
	4	FC:512	YES	-	LeakyReLU
	5	FC:1	NO	-	Sigmoid
C	1	Conv:$5*5*32$	NO	2	ReLU
	2	Conv:$5*5*128$	YES	2	ReLU
	3	Conv:$5*5*256$	YES	2	ReLU
	4	FC:512	YES	-	ReLU
	5	FC:class num	NO	-	Softmax

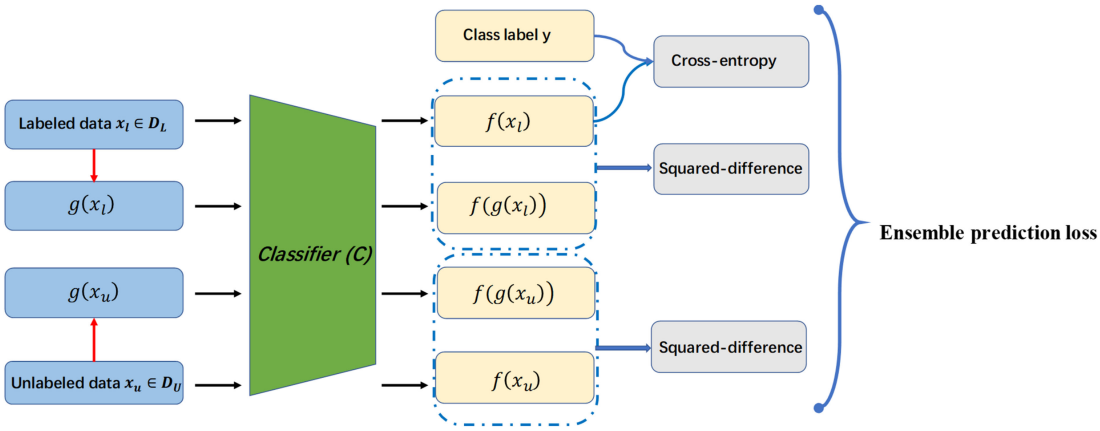


Fig. 2. Illustration of ensemble predictions.

the proposed generative model for generating HSI samples (Section III-C) and used them in a traditional classification workflow (Section III-D).

A. Semisupervised Training Framework of SSVGAN

Labeled data are usually limited in HSI classification application. When the number of labeled samples are insufficient for training, the generalization performance of the deep neural network may be degraded, and thus this property may greatly restrict the model in limited labeled data condition. To solve this problem, we first explore an ensemble prediction technique to train the classification network C by simultaneously optimizing a standard supervised classification loss on labeled samples along with an additional unsupervised consistence loss term imposed on labeled and unlabeled data (Fig. 2).

Mathematically, suppose we have a limited amount of labeled samples $D_L = \{(x_{i,l}, y_i)\}_i^{N_l}$ with an abundant amount of unlabeled samples $D_U = \{(x_{i,u})\}_i^{N_u}$, where $N_u \gg N_l$. In the training stage, the network C tries to minimize the following loss function:

$$\mathcal{L}_C = -\frac{1}{N_l}\sum_{x_i \in D_L} \log P(c|x_i) + \frac{1}{N_u}\sum_{x_i \in D_L \cup D_U} \| f(x_i) - f(g(x_i)) \| \quad (5)$$

where the first term is the standard supervised loss function evaluated for labeled inputs only. The second term is the added unsupervised consistence loss term evaluated for labeled and unlabeled inputs. Here, $f(x_i)$ and $f(g(x_i))$ represent the class probability of x_i and $g(x_i)$, respectively. $g(x_i)$ is a stochastic input augmentation function used to impose Gaussian noise or random transformation on the original data x_i. The second term is used to penalize the inconsistent prediction between two realizations of the same input. By using unsupervised consistence loss term, we can use a large number of unlabeled samples to better learn the distribution of training samples. Moreover, it can encourage the decision boundary lie in the low-density regions of the data distribution which is helpful for classification, because

small prediction error between two realizations of the same input can be satisfied only if the decision boundary traverses a low-density path.

B. Collaborative Optimization Mechanism Between the Generative Network G and the Classification Network C

Another limitation of CVAEGAN is that it does not clearly build the relationship between the generative network G and the classification network C in model training stage, which may lead to generate just realistic but not enough meaningful samples for learning classifiers. To solve this problem, we further build a semisupervised collaborative optimization mechanism between network G and network C in an end-to-end learning manner.

Specifically, the semisupervised collaborative optimization mechanism contains two training processes.

In classification network C training stage, we choose both generated samples and real data as the input of network C, with the aim of providing diverse information to assist the classification network training. Therefore, the (5) can be rewritten as

$$\mathcal{L}_C = -\frac{1}{N_l}\sum_{x_i\in D_L}\log P(c|x_i) + \frac{\lambda}{N_u}\sum_{x_i\in D_L\cup D_U}\parallel f(x_i)-f(G(x_i))\parallel . \tag{6}$$

Here, $G(x_i)$ is the generative function instead of stochastic input augmentation function on the original data x_i. λ is a ramp-up weight that trades off the semisupervised with classification accuracy on the labeled data. In the initial training state, since the generated samples are far away from the real samples in the data space, λ is set to zero. As the training epoch increases, λ is ramping up to the max weight (empirically, multiply the labeled percentage in a dataset by 10) during the first 80 training epochs. In the generative network G training stage, similar to CVAEGAN, we also use the feature matching strategy [defined in (4)] to link the generative network G and the classification network C. However, different from CVAEGAN using pairwise feature matching based on pixel-wise loss, we use the global second-order statistics (covariances) loss in this stage. This is because that the classification network C can rapidly achieve convergence if we use semisupervised training mechanism and pixel-wise loss function simultaneously. In other words, the whole network is easily overfitting, which makes the generated samples realistic but lose diversity. Thus, in order to alleviate the overfitting problem, we choose the global second-order statistics to measure the global deviations of classified features between the real x and the generated $G(z,c)$. Therefore, the (4) will be rewritten as

$$\mathcal{L}_G = \frac{1}{2}\Big(\parallel x-G(z,c)\parallel_2^2 + \parallel f_D(x)-f_D(G(z,c))\parallel_2^2 + \parallel C_f(x)-C_f(G(z,c))\parallel_2^2\Big) \tag{7}$$

where C_f is covariance matrices of the features of an intermediate layer of classification network C.

Algorithm 1: Optimization Procedure of SSVGAN.

Input: The labeled samples D_L, and unlabeled samples D_U.
Output: The parameters of whole network, $\theta = \{\theta^C, \theta^E, \theta^G, \theta^D\}$

```
while θ^G not converge do
  for one epoch do
    Require: x ∈ D_L ∪ D_U
    Update θ^D by minimizing L_D in (3)
        θ^D ←+ −∇_{θ^D}(L_D)
    Update θ^E by minimizing (L_KL + L_G) in (1)
      and (7)
        θ^E ←+ −∇_{θ^E}(L_KL + L_G)
    Update θ^G by minimizing L_G in (7)
        θ^G ←+ −∇_{θ^G}(L_G)
    Update θ^C by minimizing L_C in (6)
        θ^C ←+ −∇_{θ^C}(L_C)
  end for
end while
```

The training of SSVGAN is performed in a semisupervised way, and the overall parameters $\theta = \{\theta^C, \theta^E, \theta^G, \theta^D\}$ for network C, E, G, and D are optimized by minimizing the following objective function:

$$\mathcal{L} = \mathcal{L}_C + \mathcal{L}_{KL} + \mathcal{L}_G + \mathcal{L}_D \tag{8}$$

where $\mathcal{L}_C$ is the semisupervised loss defined in (6), $\mathcal{L}_{KL}$ is the KL loss defined in (1), $\mathcal{L}_D$ and $\mathcal{L}_G$ are the conditional loss defined in (3) and (7). The detailed optimization process of SSVGAN is shown in Algorithm 1.

C. Generation of HSI Samples Using SSVGAN for HSI Classification

This section describes the use of the proposed generative network for generating HSI samples. For HSI analysis, researchers demonstrated that the redundancy from interband correlation is very high, and the data structure in the spectral dimension can be reduced without significant loss of the useful information for subsequent utilizations [1]. Moreover, training our networks using such high-dimensional samples is difficult. Thus, we apply principal components analysis (PCA) for dimensionality reduction, which can reduce the spectral dimensions to a suitable scale and speed up the convergence of model network training.

Then, spatial sampling is performed for each nonedge pixel, and the sampling matrix is a size of 32×32 centered on the pixel. The samples will be used as the inputs of the proposed model, and the label of the center pixel in the sampling matrix will be used as the sample label. Subsequently, we train the proposed model using limited labeled samples and large quantities of unlabeled samples.

The proposed network can explicitly learn the relationship between the real image space and the latent space. When the network is trained, we can use it to generate samples by simply modifying the latent vector that corresponds to the samples. The sample generation procedure consists three steps:

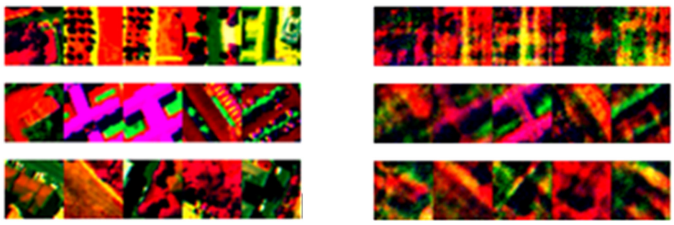


Fig. 3. False color map for real and generated data on four datasets (have been processed by PCA), where three PCs are regarded as the red, green, and blue bands. (a) Real training data. (b) Generated data by SSVGAN.

Step 1: Randomly select one labeled samples x from a category c, and then extract the mean and covariance of the latent vector, i.e., μ and ϵ using the encoder network E.

Step 2: Obtain a latent vector z by impose partial Gaussian noise in sampling: $z^{'} = z + \boldsymbol{r} * \nu$, where $z = \mu + \boldsymbol{r} * exp(\epsilon)$, $r \sim \mathcal{N}(0, I)$ and ν is a sparse matrix to slightly modify the value of one or two dimensions.

Step 3: Generate a fake sample using $z^{'}$ and the generative network G.

We can generate samples from category c as much as you want by repeating these three steps. (Fig. 3) shows some selected real and generated samples by SSVGAN on four datasets. Intuitively, a gap between the fake samples and the real samples is still observed, whereas generated samples have captured some detailed attributes from the real samples.

D. HSI Classification With Generated Samples

Though the classifier C in the proposed SSVGAN model can be used for classification, it is a shallow network designed specifically to enhance the training stability of the proposed GAN rather than extracting discriminant features. Meanwhile, achieving satisfactory classification performance under limited labeled sample conditions is difficult even if the ensemble prediction technique is used for training. Due to this concern, in our work, we follow the traditional classification workflow, which first extract hand-craft spectral-spatial feature [41] from both real and generated samples, and then embedding it into a linear SVM for final classification. Specifically an input hyperspectral image $\mathbf{I}$ with d principal components (PC) is provided, and spectral-spatial feature for an arbitrary pixel x_0 in $\mathbf{I}$ is extracted as follows:

$$\mathbf{U} = \{f_0, \text{sort}([f_1, f_2 \dots f_{N*N-1}])\} \in R^{d \times (N*N+1)} \quad (9)$$

where $f_0 \in R^{d*1}$ is the reduced spectral vector associated to the pixels x_0, and $f_1, f_2 \dots f_{N*N-1}$ are reduced spectral vectors associated to the pixels in the $N * N$ square neighborhood of x_0. The function **Sort ()** is used to reorder $N * N - 1$ reduced spectral vectors based on the corresponding value of the first PCs in ascending order. Here, sort function can make spectral-spatial features invariant to local image rotation [41].

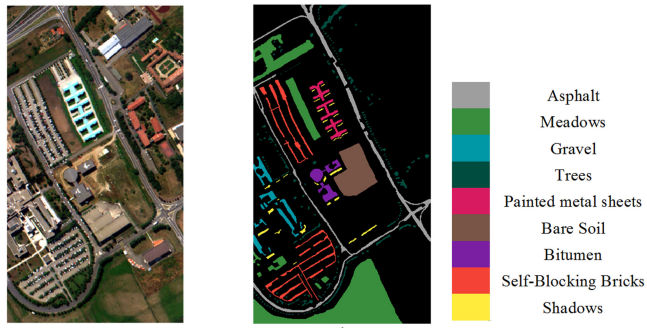


Fig. 4. University of Pavia. (a) False color image. (b) Ground reference map.

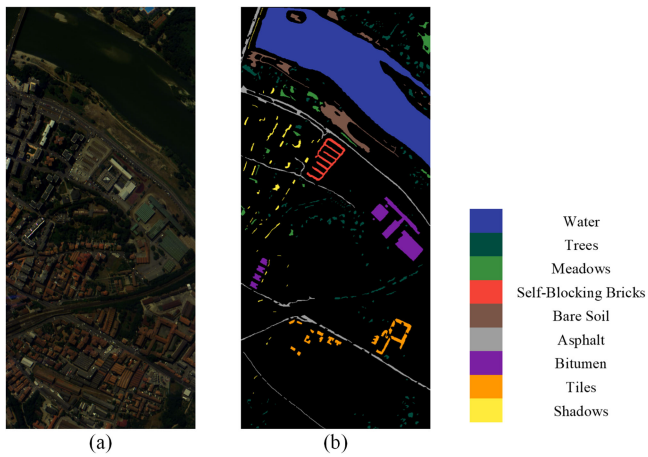


Fig. 5. Pavia Centre. (a) False color image. (b) Ground reference map.

IV. Experiment Results

A. Data Description

In our experiments, three widely used hyperspectral datasets and one Landsat remote sensing image are adopted to validate the proposed method. The following is a detailed description of these datasets:

University of Pavia: This dataset was acquired by the ROSIS sensor during a flight campaign over Pavia, Northern Italy. The number of spectral bands for Pavia University is 103 with a wavelength range of 0.43−0.86 μm. Pavia University contains 610 × 340 pixels with a spatial resolution of 1.3 m. Image ground truths differentiate nine classes each. Fig. 4 shows the false color images and its ground truth, where bands 57, 34, and 3 are regarded as the red, green, and blue bands, respectively.

Pavia Centre: This dataset was acquired by the ROSIS sensor during a flight campaign over Pavia, Northern Italy, with a size of 1096 × 492 and 102 spectral bands in the wavelength range of 0.43−0.86 μm. The spatial resolution is 1.3 m. The available ground reference map covers nine classes of interest. Fig. 5 shows the false color images and its ground truth, where bands 57, 34, and 3 are regarded as the red, green, and blue bands, respectively.

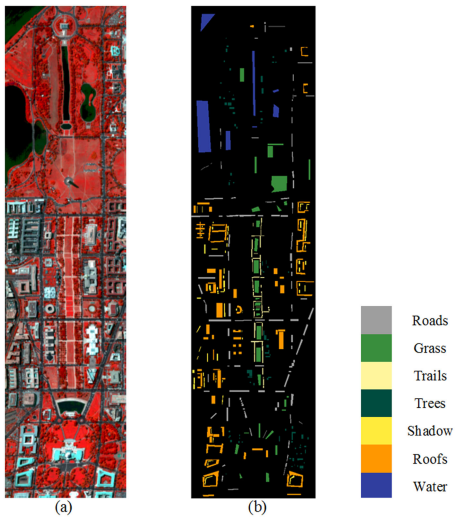


Fig. 6. DC Mall. (a) False color image. (b) Ground reference map.

DC Mall: This dataset was an airborne hyperspectral data flightline over the Washington DC Mall provided with the permission of Spectral Information Technology Application Center of Virginia. The sensor system response in 210 bands in the 0.4–2.4 μm region of the visible and infrared spectrum. DC Mall has 1208 $\times$ 307 pixels with a spatial resolution of 2 m. Bands in the 0.9 and 1.4 μm region, where the atmosphere is opaque, have been omitted from the dataset, thereby leaving 191 bands. The available ground reference map covers seven classes of interest. Fig. 6 shows the false color images and its ground truth, where bands 76, 42, and 3 are regarded as the red, green, and blue bands, respectively.

Jiamusi: This dataset was acquired by the Landsat8/OLI (Operational Land Imager) sensor over Jiamusi, Western China with a size of 256 $\times$ 256 and 7 spectral bands. The spatial resolution is 30 m. We manually label the data with four covered classes of interest. Fig. 7 shows the false color images and its ground truth, where bands 4, 3, and 2 are regarded as the red, green, and blue bands, respectively.

In the experiments, five labeled samples per class are used from the ground truth together with the unlabeled samples from the input image to train the proposed SSVGAN. Afterward, several samples are generated per class using the trained network, and combine them with the limited truth samples per class for SVM classifier training.

To evaluate the quality of our approach, we assess the performance of classification by computing three common measures: the overall accuracy (OA), the average accuracy (AA), and the Kappa coefficient on the available reference data. Five experiments are performed on a PC with 3.7 Hz Core E5-1620 CPU, GeForce GTX 1080Ti and 24-GB RAM:

Experiment I aims at statistically examine the effect of different parameter values on final image classification performance.

Experiment II aims to evaluate the quality of synthesized samples on HSI classification and explores whether the samples synthesized from our model SSVGAN can be used for data augmentation to improve HSI classification model.

Experiment III compares different strategic generative methods to generate samples for HSI classification and demonstrates the advantage of the proposed SSVGAN.

Experiment IV compares different semisupervised methods for HSI classification to validate the quality of classification results, especially when the number of labeled samples is extremely limited.

Experiment V tests the effect with a semisupervised, supervised, or without classification network C in the proposed SSVGAN on the final image classification.

B. Classification Results

1) Parametric Analysis: In our method, there are two main parameters directly affecting the quality of generated samples. The first one is the batch size used in training the proposed generative network. The second one is the number of retained PCs after PCA.

To assess the impact of batch size, we have analyzed a minibatch size of 8 to a larger minibatch size of 128 on four datasets in Fig 8. As can be seen, we get the highest OA when the batch size is set to 16 or 32. The accuracy gradually decreases as the batch size rises. When the batch size is set to 128, the accuracy has slipped almost 1%. It shows that very large batch size may harm the generalization ability of the proposed generative model. In addition, batch size can affect computation time as well as performance. According to the experiment, we set the batch size as 32, considering both performance and computation time.

In addition, the PCA-3, PCA-10, PCA-20 and PCA-50 in Table II means that we preserve 3, 10, 20, and 50 PCs, respectively. Since the Jiamusi contains seven bands, we only compute PCA-3 on this dataset. As can be seen, when we retained ten PCs, the dataset PaviaCe and DC Mall can obtain the best classification accuracy, and PaviaU dataset can obtain the second best classification accuracy. Moreover, if the number of retained PCs is not sufficient (e.g., three PCs) or too excessive (e.g., fifty PCs), the classification results prone to decrease. Thus, we set the number of retained PCs as ten in the following experiments.

2) Result Analysis: In this experiment, the quality of the synthesized samples on HSI classification is evaluated and explores whether the samples synthesized from our model SSVGAN can be used for data augmentation to remarkably train the HSI classification model. We extract the hand-craft spectral-spatial feature [41] from limited labeled samples and embedding it into a linear SVM as the baseline classification. And we use the same limited labeled samples and amounts of unlabeled samples to train our model SSVGAN to synthesize samples. Following, detailed experiments are conducted by adding synthetic samples for each class. We use both real and synthetic samples to train another linear SVM for the comparison. All the experiments are repeated ten times, and Tables III–VI show the mean and standard deviation classification results for these four test images.

Two observations were made from the result:

In contrast with the baseline, which only used five labeled samples for each class, the proposed method improves its performance by adding synthetic samples for these four datasets. The

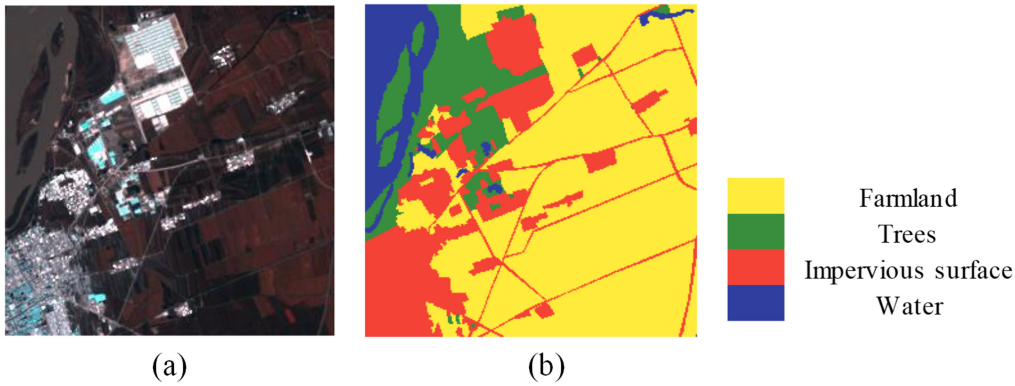


Fig. 7. Jiamusi. (a) False color image. (b) Ground reference map.

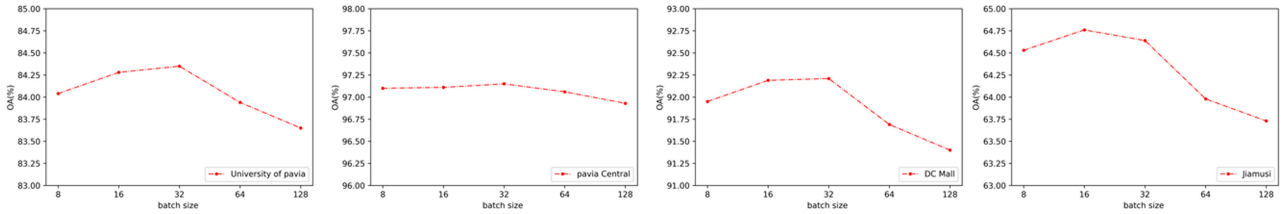


Fig. 8. Parameters analysis of the batch size (From left to right: University of Pavia, Pavia Centre, DC Mall, and Jiamusi, respectively).

TABLE II
PARAMETRIC ANALYSIS OF THE NUMBER OF RETAINED PCS

Dataset		PCA-3	PCA-10	PCA-20	PCA-50
PaviaU	OA(%)	79.00±1.57	84.01±0.66	**84.35±0.19**	82.77±0.22
	AA(%)	79.15±2.24	81.79±1.26	**85.35±0.47**	80.64±0.75
	Kappa*100	73.62±1.86	79.70±0.79	**80.27±0.24**	78.09±0.30
PaviaCe	OA(%)	95.71±0.14	**97.15±0.05**	94.95±0.12	94.91±0.13
	AA(%)	91.65±0.32	**94.73±0.18**	94.62±0.24	94.58±0.21
	Kappa*100	92.58±0.21	**95.01±0.09**	93.76±0.15	93.73±0.11
DC Mall	OA(%)	82.33±2.87	**92.21±0.21**	87.67±0.65	87.56±0.39
	AA(%)	82.61±3.00	**90.63±0.32**	90.93±0.44	90.15±0.39
	Kappa*100	78.50±3.40	**90.37±0.26**	85.03±0.76	84.72±0.47
Jiamusi	OA(%)	64.76±0.82	-	-	-
	AA(%)	66.97±1.04	-	-	-
	Kappa*100	45.87±0.88	-	-	-

Bold signifies the best accuracy obtained in different conditions of retained principal components.

TABLE III
CLASSIFICATION RESULTS OBTAINED BY PAVIAU WITH FIVE REAL SAMPLES AND SYNTHETIC SAMPLES BY SSVGAN FOR EACH CLASS

University of Pavia	Baseline	SSVGAN
Asphalt	74.30	83.13±0.54
Meadows	70.00	91.34±0.19
Gravel	76.08	80.76±0.62
Trees	76.50	88.82±0.65
Painted metal sheets	99.09	99.45±0.12
Bare Soil	80.60	76.45±0.67
Bitumen	97.54	90.30±0.41
Self-Blocking Bricks	59.73	57.97±0.86
Shadows	97.22	99.95±0.13
OA(%)	74.65	84.35±0.19
AA(%)	81.23	85.35±0.47
Kappa*100	68.89	80.27±0.24

TABLE IV
CLASSIFICATION RESULTS OBTAINED BY PAVIA CENTRE WITH FIVE REAL SAMPLES AND SYNTHETIC SAMPLES BY SSVGAN FOR EACH CLASS

Pavia Centre	Baseline	SSVGAN
Water	99.89	99.98±0.01
Trees	88.05	95.30±0.19
Meadows	95.23	92.38±0.00
Self-Blocking Bricks	100.00	99.92±0.23
Bare Soil	73.18	82.33±0.23
Asphalt	99.02	99.53±0.13
Bitumen	84.59	83.88±0.58
Tiles	99.47	99.31±0.24
Shadows	100.00	100.00±0.00
OA(%)	96.13	97.15±0.05
AA(%)	93.27	94.73±0.18
Kappa*100	93.23	95.01±0.09

TABLE V
CLASSIFICATION RESULTS OBTAINED BY DC MALL WITH FIVE REAL SAMPLES AND SYNTHETIC SAMPLES BY SSVGAN FOR EACH CLASS

DC Mall	Baseline	SSVGAN
Roads	97.01	96.96±0.20
Grass	92.54	96.35±0.35
Trails	83.28	90.86±0.50
Trees	84.76	96.85±0.12
Shadow	83.83	69.19±0.31
Roofs	54.17	84.74±0.78
Water	87.19	99.46±0.00
OA(%)	79.79	92.21±0.21
AA(%)	83.25	90.63±0.32
Kappa*100	75.67	90.37±0.26

TABLE VI
CLASSIFICATION RESULTS OBTAINED BY JIAMUSI WITH FIVE REAL SAMPLES AND SYNTHETIC SAMPLES BY SSVGAN FOR EACH CLASS

Jiamusi	Baseline	SSVGAN
Farmland	35.26	64.59±1.57
Trees	30.51	42.17±1.30
Impervious surface	44.60	73.47±0.45
Water	87.81	87.68±0.82
OA(%)	39.32	64.76±0.82
AA(%)	49.54	66.97±1.04
Kappa*100	15.49	45.87±0.88

TABLE VII
COMPARED STRATEGIC CLASSIFICATION RESULTS OBTAINED BY PAVIAU WITH FIVE REAL SAMPLES AND SYNTHETIC SAMPLES FOR EACH CLASS

University of Pavia	SSVGAN	ACGAN	CVAEGAN
Asphalt	83.13±0.54	74.56±0.06	79.80±0.81
Meadows	91.34±0.19	80.50±0.19	90.79±0.49
Gravel	80.76±0.62	57.01±0.23	81.08±1.14
Trees	88.82±0.65	87.24±0.20	84.44±0.39
Painted metal sheets	99.45±0.12	99.69±0.00	96.33±0.25
Bare Soil	76.45±0.67	55.01±0.44	69.84±0.93
Bitumen	90.30±0.41	84.57±0.23	93.40±0.36
Self-Blocking Bricks	57.97±0.86	57.10±0.11	55.64±1.24
Shadows	99.95±0.13	99.07±0.11	99.12±0.13
OA(%)	84.35±0.19	74.43±0.10	82.08±0.18
AA(%)	85.35±0.47	77.19±0.16	83.38±0.64
Kappa*100	80.27±0.24	68.06±0.12	77.35±0.22

TABLE VIII
COMPARED STRATEGIC CLASSIFICATION RESULTS OBTAINED BY PAVIA CENTRE WITH FIVE REAL SAMPLES AND SYNTHETIC SAMPLES FOR EACH CLASS

Pavia Centre	SSVGAN	ACGAN	CVAEGAN
Water	99.98±0.01	100.00±0.00	100.00±0.00
Trees	95.30±0.19	92.27±0.44	94.22±0.36
Meadows	92.38±0.00	94.38±0.79	93.33±0.00
Self-Blocking Bricks	99.92±0.23	99.44±0.36	99.68±0.38
Bare Soil	82.33±0.23	79.02±0.94	80.35±0.20
Asphalt	99.53±0.13	99.75±0.00	99.75±0.00
Bitumen	83.88±0.58	83.57±0.50	84.01±0.50
Tiles	99.31±0.24	99.52±0.43	99.63±0.24
Shadows	100.00±0.00	100.00±0.00	100.00±0.00
OA(%)	97.15±0.05	96.78±0.06	97.01±0.04
AA(%)	94.73±0.18	94.22±0.38	94.55±0.18
Kappa*100	95.01±0.09	94.37±0.11	94.76±0.07

TABLE IX
COMPARED STRATEGIC CLASSIFICATION RESULTS OBTAINED BY DC MALL WITH FIVE REAL SAMPLES AND SYNTHETIC SAMPLES FOR EACH CLASS

DC Mall	SSVGAN	ACGAN	CVAEGAN
Roads	96.96±0.20	97.38±0.16	95.64±0.31
Grass	96.35±0.35	97.08±0.23	97.77±0.40
Trails	90.86±0.50	85.71±0.65	85.04±1.58
Trees	96.85±0.12	98.72±0.14	97.46±0.43
Shadow	69.19±0.31	67.62±0.69	68.38±0.71
Roofs	84.74±0.78	73.99±1.07	81.71±1.31
Water	99.46±0.00	99.58±0.01	99.29±0.01
OA(%)	92.21±0.21	89.13±0.27	91.07±0.35
AA(%)	90.63±0.32	88.58±0.42	89.33±0.68
Kappa*100	90.37±0.26	86.69±0.32	88.99±0.41

TABLE X
COMPARED STRATEGIC CLASSIFICATION RESULTS OBTAINED BY JIAMUSI WITH FIVE REAL SAMPLES AND SYNTHETIC SAMPLES FOR EACH CLASS

Jiamusi	SSVGAN	ACGAN	CVAEGAN
Farmland	64.59±1.57	60.14±5.64	66.78±4.57
Trees	42.17±1.30	28.21±7.69	17.35±3.02
Impervious surface	73.47±0.45	49.91±10.09	66.36±7.50
Water	87.68±0.82	87.82±0.75	89.41±1.66
OA(%)	64.76±0.82	54.79±5.15	61.14±2.41
AA(%)	66.97±1.04	56.52±6.04	59.97±4.19
Kappa*100	45.87±0.88	30.77±7.07	38.22±3.24

proposed method can achieve approximately 11.38%, 1.78%, 14.7%, and 30.38% improvement in Kappa compared with baseline. The samples generated by the proposed model introduces variability, thereby improving the classifier training process.

Another interesting observation is that although the classification accuracy of most categories has improved, some categories still decline. In particular, the classification performance of the same category significantly differ under various datasets. For example, the category "Bare Soil" has reduced accuracy after adding generated samples in the University of Pavia dataset. However, its classification accuracy is improved to 82.33% in Pavia Centre. From the perspective of data collection, although they are under the same label category, significant differences in the surface features are expressed under different datasets. The Bare Soil in Pavia Centre is adjacent to the river, but it is surrounded by buildings in the University of Pavia. Affected by the surrounding environment, the spectral information of a marginal Bare Soil pixel in the University of Pavia dataset is more mixed. Thus, the model is easy to generate mixed Bare Soil samples which are useless for classification. Another example of performance difference is under the "shadow" category in University of Pavia and DC Mall datasets. The resolution of University of Pavia and DC Mall dataset is 1.3 and 2 m, respectively. This resolution complicates the spectral information of Shadow category pixels in DC Mall.

3) Comparison Analysis in Generative Methods for HSI Classification: In this experiment, we compared the proposed SSVGAN with two state-of-the-art *generative methods* in limited labeled condition (e.g., five labeled samples per class). The first one is auxiliary classifier generative adversarial network (ACGAN) [32], which is a supervised framework with

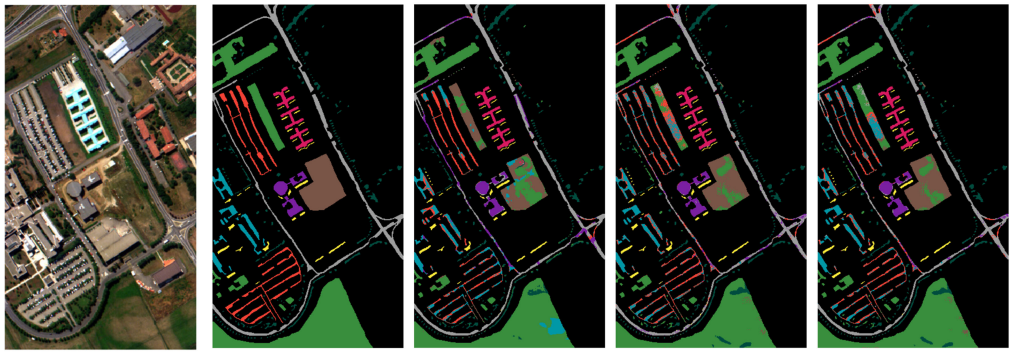


Fig. 9. University of Pavia. (a) False color image. (b) Ground truth. (c) ACGAN. (d) CVAEGAN. (e) SSVGAN.

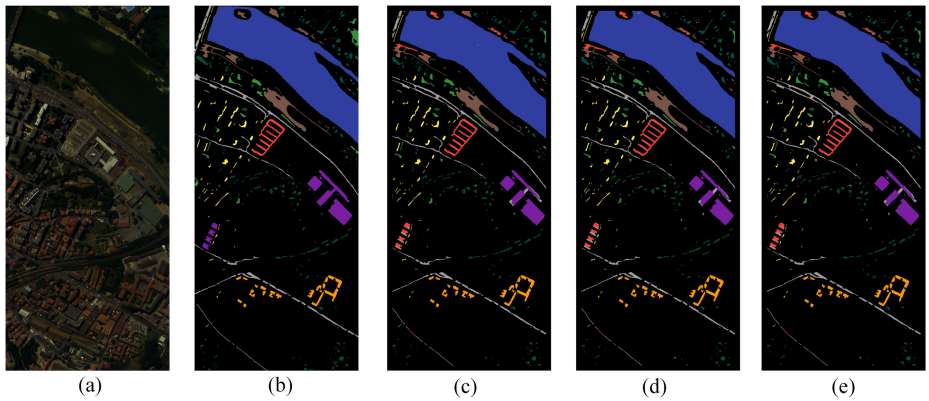


Fig. 10. Pavia Centre. (a) False color image. (b) Ground truth. (c) ACGAN. (d) CVAEGAN. (e) SSVGAN.

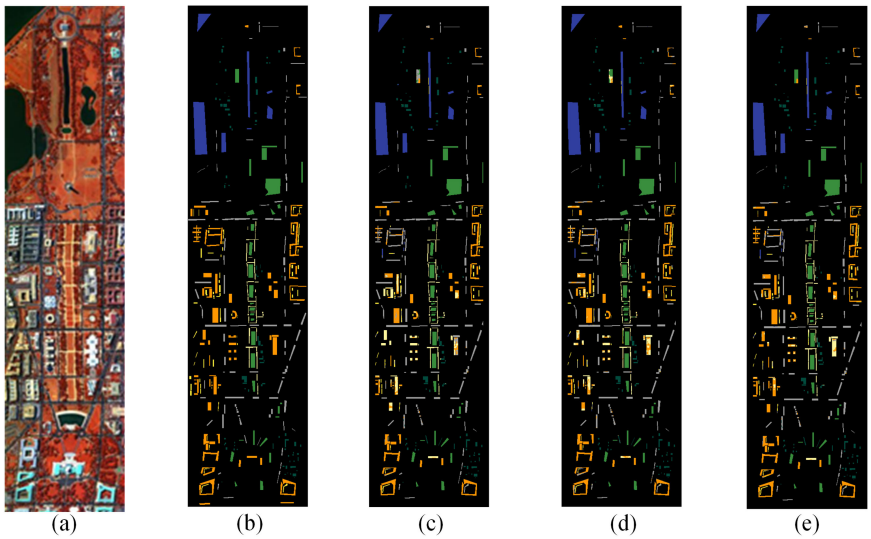


Fig. 11. DC Mall. (a) False color image; (b) Ground truth. (c) ACGAN. (d) CVAEGAN. (e) SSVGAN.

a generative network G, a discriminative network D, and a supervised classification network C. The G network consists 4 convolution layers with 256, 128, 64, and 10 channels. For the network D network and network C, we use the same network architecture settings shown in Table I. For training the ACGAN, we use the Adam solver with the best-fixed learning rate of $2.0 \times 10–4$ in the training stage. The second one is CVAEGAN [38], we use the same network architecture settings shown in Table I. For training the CVAEGAN, we use the Adam solver with the best-fixed learning rate of $5.0 \times 10–5$. For the proposed SSVGAN, we use the Adam solver with a fixed learning rate of $1.0 \times 10–4$. We trained each model for 10 000

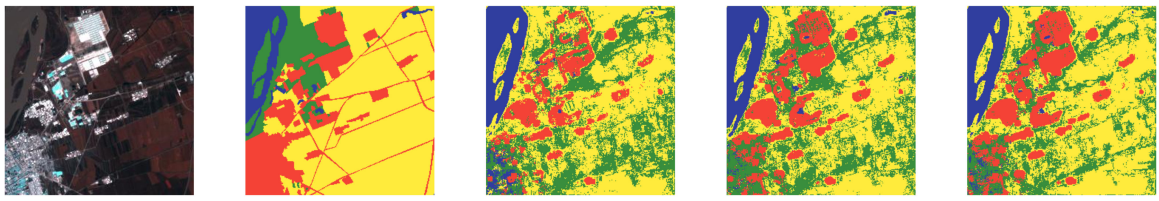


Fig. 12. Jiamusi. (a) False color image. (b) Ground truth. (c) ACGAN. (d) CVAEGAN. (e) SSVGAN.

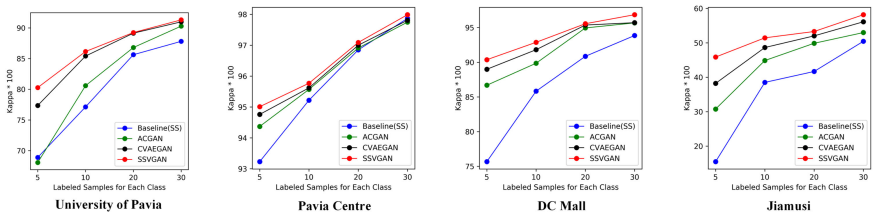


Fig. 13. Classification results obtained by four datasets with 5-10-20-30 labeled samples for each class (From left to right: University of Pavia, Pavia Centre, DC Mall and Jiamusi, respectively). Each marker in the figure represents the AA.

TABLE XI
COMPARED SEMISUPERVISED CLASSIFICATION RESULTS OBTAINED BY PAVIAU WITH FIVE REAL SAMPLES FOR EACH CLASS

University of Pavia	SSVGAN	SS-LPSVM	BLGL	RLDE
Asphalt	83.13±0.54	82.04±7.24	78.29±4.17	78.86±4.14
Meadows	91.34±0.19	85.59±2.79	90.23±4.28	81.94±1.62
Gravel	80.76±0.62	68.11±3.15	57.82±8.67	78.99±2.17
Trees	88.82±0.65	75.74±3.26	71.53±1.79	95.77±1.08
Painted metal sheets	99.45±0.12	98.98±0.13	99.19±0.13	99.96±0.09
Bare Soil	76.45±0.67	53.52±5.39	54.61±7.56	93.00±1.93
Bitumen	90.3±0.41	88.75±1.86	84.35±4.28	94.90±0.53
Self-Blocking Bricks	57.97±0.86	50.82±5.87	79.17±2.87	55.19±2.26
Shadows	99.95±0.13	95.37±1.38	96.45±1.64	99.53±0.41
OA(%)	84.35±0.19	77.45±1.18	79.04±1.34	84.01±1.24
AA(%)	85.35±0.47	78.54±0.88	79.76±0.61	84.38±0.92
Kappa*100	80.27±0.24	71.60±2.78	73.57±2.16	78.86±1.51

TABLE XII
COMPARED SEMISUPERVISED CLASSIFICATION RESULTS OBTAINED BY PAVIA CENTRE WITH FIVE REAL SAMPLES FOR EACH CLASS

Pavia Centre	SSVGAN	SS-LPSVM	BLGL	RLDE
Water	99.98±0.01	99.78±0.07	100±0.00	100.00±0.00
Trees	95.30±0.19	87.77±0.85	94.38±1.28	95.55±0.41
Meadows	92.38±0.00	95.23±1.36	93.91±0.63	80.95±0.42
Self-Blocking Bricks	99.92±0.23	98.8±0.21	99.24±0.38	99.94±0.12
Bare Soil	82.33±0.23	83.6±3.66	73.43±3.35	83.20±1.65
Asphalt	99.53±0.13	98.92±0.36	99.51±0.03	99.02±0.34
Bitumen	83.88±0.58	84.81±1.55	86.61±1.73	83.79±0.64
Tiles	99.31±0.24	96.76±0.65	97.15±0.87	99.43±0.60
Shadows	100.00±0.00	99.39±0.06	100±0.00	100±0.00
OA(%)	97.15±0.05	96.23±0.52	96.30±0.34	97.04±0.19
AA(%)	94.73±0.18	93.33±0.31	93.32±0.36	93.66±0.33
Kappa*100	95.01±0.09	93.40±0.69	93.51±0.66	94.82±0.17

TABLE XIII
COMPARED SEMISUPERVISED CLASSIFICATION RESULTS OBTAINED BY DC MALL WITH FIVE REAL SAMPLES FOR EACH CLASS

DC Mall	SSVGAN	SS-LPSVM	BLGL	RLDE
Roads	96.96±0.20	97.07±1.40	94.01±2.08	96.87±0.66
Grass	96.35±0.35	92.73±2.38	98.26±0.65	96.11±0.74
Trails	90.86±0.50	85.71±1.01	84.74±0.97	86.79±1.85
Trees	96.85±0.12	85.90±0.21	95.98±0.63	95.80±0.47
Shadow	69.19±0.31	73.23±3.14	73.51±1.86	69.94±2.72
Roofs	84.74±0.78	55.63±5.81	67.29±2.20	81.46±2.91
Water	99.46±0.00	98.94±0.35	98.66±0.37	99.46±0.00
OA(%)	92.21±0.21	84.57±1.22	86.96±0.79	90.98±0.82
AA(%)	90.63±0.32	85.87±0.73	87.60±0.84	89.49±0.79
Kappa*100	90.37±0.26	81.20±1.35	84.08±0.94	88.85±1.14

TABLE XIV
COMPARED SEMISUPERVISED CLASSIFICATION RESULTS OBTAINED BY JIAMUSI WITH FIVE REAL SAMPLES FOR EACH CLASS

Jiamusi	SSVGAN	SS-LPSVM	BLGL	RLDE
Farmland	64.59±1.57	70.36±4.43	71.81±3.47	73.85±4.24
Trees	42.17±1.30	19.63±3.96	18.92±1.99	18.95±6.37
Impervious surface	73.47±0.45	44.66±12.11	45.28±12.79	58.33±4.76
Water	87.68±0.82	87.72±0.08	87.73±0.25	87.81±0.16
OA(%)	64.76±0.82	58.36±2.10	59.28±2.57	63.52±1.21
AA(%)	66.97±1.04	55.58±3.91	55.93±4.26	59.73±2.77
Kappa*100	45.87±0.88	30.62±5.15	31.36±5.89	38.99±3.53

iterations with a batch size of 32 until they converge. And the best retained PCs shown in Table II on four datasets have been applied.

In the experiments, we use both real and synthetic samples from different generative methods to train another linear SVM for the comparison. From compared classification results of generative methods in Tables VII–X, the SSVGAN shows a better performance in terms of accuracies. For the DC Mall dataset, the SSVGAN exhibits the highest OA, AA, and Kappa*100, with an improvement of 3.08%, 2.05%, and 3.68% over ACGAN, respectively. On the other hand, our SSVGAN approach outperforms CVAEGAN by 1.14%, 1.30%, and 1.38% in terms of OA, AA, and Kappa*100, respectively. For the University of Pavia, Pavia Centre, and Jiamusi datasets, we can obtain the similar results. These results indicate that the semisupervised collaborative optimization mechanism is conduce to generate meaningful samples compared with other generative methods. Classification accuracies are also evaluated in a visual perspective. Figs. 9–12(c)–(e) show the classification maps for ACGAN, CVAEGAN, and SSVGAN in four datasets.

TABLE XV
EFFECT WITH A SEMISUPERVISED, SUPERVISED OR WITHOUT CLASSIFICATION NETWORK C IN THE PROPOSED SSVGAN ON FINAL CLASSIFICATION

Dataset		PaviaU	PaviaCe	DC Mall	Jiamusi
Without the network C in SSVGAN	OA(%)	79.53±0.23	96.94±0.06	87.99±0.34	55.84±3.41
	AA(%)	79.70±0.27	94.48±0.21	87.97±0.38	56.36±4.66
	Kappa*100	74.00±0.13	94.65±0.10	85.34±0.40	31.90±4.84
With a supervised network C in SSVGAN	OA(%)	81.53±0.35	96.95±0.06	90.62±0.38	60.83±2.19
	AA(%)	81.32±0.42	94.52±0.24	89.31±0.47	61.05±3.57
	Kappa*100	76.54±0.23	94.76±0.12	88.67±0.37	38.56±3.26
With a semi-supervised network C in SSVGAN	OA(%)	84.35±0.19	97.15±0.05	92.21±0.21	64.76±0.82
	AA(%)	85.35±0.47	94.73±0.18	90.63±0.32	66.97±1.04
	Kappa*100	80.27±0.24	95.01±0.09	90.37±0.26	45.87±0.88

Moreover, we investigate the behavior of the different generative methods regarding to the different number of labeled samples in Fig. 13. We can gain the following observations from the results: Compared to the other generative methods, SSVGAN exhibits the highest classification improvement on four datasets. As labeled samples increases, the classification gap from different generative methods has been closer. However, the proposed SSVGAN provides a significant improvement in the small number of labeled samples conditions. This difference is due to the fact that unlabeled samples have been exploited in SSVGAN. In other words, with the help of a semisupervised classification network C, SSVGAN can learn well to synthesize meaningful samples for enhancing hyperspectral classification performance.

4) Comparison Analysis (Semisupervised Methods for HSI Classification): In the previous section, several generative methods for HSI classification have been included to give a comprehensive comparison. It demonstrates that the proposed model has excellent abilities in image classification. In this section, we further compared the proposed method with three well-known semisupervised methods for HSI classification including SS-LPSVM [42], BLGL [43], and RLDE [44], where SS-LPSVM, BLGL are graph-based methods and RLDE is a cotraining method. For these compared methods, we use the optimal parameter settings based on experiments. It is noted these compared semisupervised methods directly use limited labeled samples and some unlabeled samples for classification. While our generative model aims to generate meaningful samples using labeled and unlabeled real samples, then we use both real labeled and synthetic samples for classification. For a fair comparison, we adopt the same ratio of unlabeled samples, approximately 70%, for this comparison experiment. The compared classification results are shown in Tables XI−XIV.

For the DC Mall dataset, SSVGAN exhibits the best OA, AA, and Kappa*100, with improvements of 7.64%, 4.76%, and 9.17% over SS-LPSVM, respectively. Our approach outperforms BLGL by 5.25%, 3.03%, and 6.29% in terms of OA, AA, and Kappa*100, respectively. It also increases 1.23%, 1.14%, and 1.52% in terms of OA, AA, and kappa*100 compared with that of RLDE. For the University of Pavia, Pavia Centre and Jiamusi datasets, we can obtain the similar results. Additional insights into the effect of SSVGAN on HSI classification challenge are provided. To expand the labeled training set, SS-LPSVM reference [42] uses a spatial-spectral graph to propagate labels from labeled samples to unlabeled samples. By contrast, SSVGAN can improve the classification performance with the help of generated samples in the semisupervised context. This demonstrates that the generated samples from SSVGAN capture some meaningful details which can contribute to classification.

5) Test on the Effect With a Semisupervised, Supervised or Without Classification Network C in the Proposed SSVGAN: As mentioned in previous sections, we use a semisupervised classification network C in the proposed SSVGAN and build a collaborative optimization mechanism between the generative network G and the classification network C in the process of sample generation, hoping that the generated samples can contribute to the final classification. Thus, in this section, we test the effect with a semisupervised, supervised, or without classification network C in the proposed SSVGAN on the final image classification. As can be seen in Table XV, if we remove the classification network C in SSVGAN, the final classification performance will decrease for all four test images. The classification OA has been descended by 4.82%, 0.21%, 4.22%, and 8.92% on four datasets, respectively. If we use the supervised classification network C in SSVGAN, the final classification performance has decreased by 2.82%, 0.2%, 1.59%, and 3.93% on four datasets, respectively. This result shows the importance of the use of the classification network C with semisupervised training strategy in SSVGAN.

V. CONCLUSION

In this article, a semisupervised variational GAN is proposed to generate auxiliary data to overcome the challenge task of HSI classification when the number of labeled samples is extremely limited. We propose to build a collaboration relationship between the generative network and the classification network to generate samples which can contribute to classification. Experiments on four benchmark datasets show that these generated samples significantly improve classification performance and can achieve better and more robust performance than other generative methods, which is recently used to generate HSI samples. However, the proposed method still has two limitations. First, as the generation network G and classification network C are hard to simultaneously achieve convergence, the first convergent network is easily overfitting, which makes the generated samples lose diversity. Second, it is known that pixels of different entities may share a similar spectral signature in HSI, which would lead to poor discriminant using original spectral feature space for feature matching. To promote our approach for HSI

classification, we will balance the generation and classification process and consider apply the metric learning method [45], [46] in our future work.

REFERENCES

[1] J. A. Richards and J. Richards, *Remote Sensing Digital Image Analysis*, vol. 3. Berlin, Germany: Springer, 1999.

[2] A. Santara *et al.*, "Bass Net: Band-adaptive spectral-spatial feature learning neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5293–5301, Sep. 2017.

[3] B. Song, P. Li, J. Li, and A. Plaza, "One-class classification of remote sensing images using kernel sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 4, pp. 1613–1623, Apr. 2016.

[4] C. Tao, Y.-j. Wang, B. Zou, Y.-L. Tu, and X.-l. Jiang, "Assessment and analysis of migrations of heavy metal lead and zinc in soil with hyperspectral inversion model," *Spectrosc. Spectral Anal.*, vol. 38, no. 6, pp. 1850–1855, 2018.

[5] C. Tao, Y. Wang, W. Cui, B. Zou, Z. Zou, and Y. Tu, "A transferable spectroscopic diagnosis model for predicting arsenic contamination in soil," *Sci. Total Environ.*, vol. 669, pp. 964–972, 2019.

[6] C. M. Gevaert, J. Suomalainen, J. Tang, and L. Kooistra, "Generation of spectral–temporal response surfaces by combining multispectral satellite and hyperspectral UAV imagery for precision agriculture applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3140–3146, Jun. 2015.

[7] E. R. Hunt, C. S. Daughtry, S. B. Mirsky, and W. D. Hively, "Remote sensing with simulated unmanned aircraft imagery for precision agriculture applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 11, pp. 4566–4571, Nov. 2014.

[8] Y. Zhong, A. Ma, and L. Zhang, "An adaptive memetic fuzzy clustering algorithm with spatial information for remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1235–1248, Apr. 2014.

[9] S. Niazmardi, S. Homayouni, and A. Safari, "An improved FCM algorithm based on the svdd for unsupervised hyperspectral data classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 831–839, Apr. 2013.

[10] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," in *Proc. 27th AIPR Workshop: Advances Comput. Assisted Recognit.*, 1999, vol. 3584, pp. 221–232.

[11] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[12] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[13] T. K. Ho, "The random subspace method for constructing decisionforests," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[14] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[15] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 318–322, Mar. 2013.

[16] U. Maulik and D. Chakraborty, "Learning with transductive SVM for semisupervised pixel classification of remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 77, pp. 66–78, 2013.

[17] I. Dópido, J. Li, P. R. Marpu, A. Plaza, J. M. B. Dias, and J. A. Benediktsson, "Semisupervised self-learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4032–4044, Jul. 2013.

[18] J. Wang, N. Jiang, G. Zhang, B. Hu, and Y. Li, "Automatic framework for semi-supervised hyperspectral image classification using self-training with data editing," in *Proc. 7th Workshop Hyperspectral Image Signal Process.: Evolution Remote Sens.*, 2015, pp. 1–4.

[19] B. Cui, X. Xie, S. Hao, J. Cui, and Y. Lu, "Semi-supervised classification of hyperspectral images based on extended label propagation and rolling guidance filtering," *Remote Sens.*, vol. 10, no. 4, pp. 515–533, 2018.

[20] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.

[21] N. Jamshidpour, S. Homayouni, and A. Safari, "Graph-based semi-supervised hyperspectral image classification using spatial information," in *Proc. 8th Workshop Hyperspectral Image Signal Processing: Evolution Remote Sens.*, 2016, pp. 1–4.

[22] Y. Shao, N. Sang, C. Gao, and L. Ma, "Spatial and class structure regularized sparse representation graph for semi-supervised hyperspectral image classification," *Pattern Recognit.*, vol. 81, pp. 81–94, 2018.

[23] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, 2015.

[24] W. Li, G. Wu, F. Zhang, 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.

[25] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018.

[26] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[27] I. Goodfellow *et al.* "Generative adversarial nets," in *Proc. Advances Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[28] 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.

[29] 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.

[30] Z. Zhong, J. Li, D. A. Clausi, and A. Wong, "Generative adversarial networks and conditional random fields for hyperspectral image classification," *IEEE Trans. Cybern.*, early access, May 30, 2019, doi: 10.1109/TCYB.2019.2915094.

[31] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[32] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 2642–2651.

[33] L. Mescheder, S. Nowozin, and A. Geiger, "The numerics of GANs," in *Proc. Advances Neural Inf. Process. Syst.*, 2017, pp. 1825–1835.

[34] I. J. Goodfellow, "On distinguishability criteria for estimating generative models," 2014, *arXiv:1412.6515*.

[35] S. Arora, R. Ge, Y. Liang, T. Ma, and Y. Zhang, "Generalization and equilibrium in generative adversarial nets (GANs)," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 224–232.

[36] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013, *arXiv:1312.6114*.

[37] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," in *Proc. Advances Neural Inf. Process. Syst.*, 2016, pp. 658–666.

[38] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "CVAE-GAN: fine-grained image generation through asymmetric training," in *Proc. IEEE Inf. Conf. Comput. Vision*, 2017, pp. 2745–2754.

[39] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," 2016, *arXiv:1610.02242*.

[40] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Advances Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.

[41] C. Tao, Y. Tang, C. Fan, and Z. Zou, "Hyperspectral imagery classification based on rotation-invariant spectral–spatial feature," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 980–984, May 2014.

[42] L. Wang, S. Hao, Q. Wang, and Y. Wang, "Semi-supervised classification for hyperspectral imagery based on spatial-spectral label propagation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 97, pp. 123–137, 2014.

[43] Y. Gao, R. Ji, P. Cui, Q. Dai, and G. Hua, "Hyperspectral image classification through bilayer graph-based learning," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2769–2778, Jul. 2014.

[44] D. Ou, K. Tan, Q. Du, J. Zhu, X. Wang, and Y. Chen, "A novel tri-training technique for the semi-supervised classification of hyperspectral images based on regularized local discriminant embedding feature extraction," *Remote Sens.*, vol. 11, no. 6, p. 654, 2019.

[45] Y. Dong, B. Du, L. Zhang, and L. Zhang, "Dimensionality reduction and classification of hyperspectral images using ensemble discriminative local metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2509–2524, May 2017.

[46] Y. Gu and K. Feng, "Optimized Laplacian SVM with distance metric learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1109–1117, Jun. 2013.

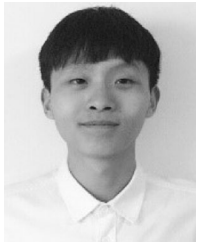
Chao Tao received the B.S. degree from the School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China, in 2007, and the Ph.D. degree from the Institution of Patter Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, China, in 2012.

He is currently an Associate Professor with the School of Geosciences and Info-Physics, Central South University, Changsha, China. He has authored more than 30 peer-reviewed articles in international journals from multiple domains such as remote sensing and computer vision. His research interests include computer vision, machine learning, deep learning, and their applications in remote sensing.

Dr. Tao has been frequently serving as a Reviewer for more than four international journals including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL), IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTAR), Photogrammetric Engineering & Remote Sensing (PE&RS), and International Society for Photogrammetry and Remote Sensing (ISPRS) *Journal of Photogrammetry and Remote Sensing* (JP&RS). He is also a Communication Evaluation Expert for the National Natural Science Foundation of China.

Hao Wang received the B.S. degree, in 2018, from the School of Geosciences and Info-Physics, Central South University, Changsha, China, where he is currently working toward the M.S. degree.

His research interests include remote sensing and computer vision, machine learning, deep learning, and their applications in remote sensing.



Ji Qi received the B.S. degree, in 2018, from the School of Geosciences and Info-Physics, Central South University, Changsha, China, where he is currently working toward the M.S. degree in the School of Geosciences and Info-Physics, Central South University, Changsha, China.

His research interests include remote sensing and computer vision, machine learning, deep learning, and their applications in remote sensing.

Haifeng Li received the master's degree in intelligence transportation system from the South China University of Technology, Guangzhou, China, in 2005, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009.

He is currently a Professor with the School of Geosciences and Info-Physics, Central South University, Changsha, China. He was a Research Associate with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, in 2011, and a Visiting Scholar with the University of Illinois at Urbana-Champaign, Urbana, USA, from 2013 to 2014. He has authored over 30 journal papers. His current research interests include geo/remote sensing big data, machine/deep learning, and artificial/brain-inspired intelligence.

Dr. Li is a Reviewer for many journals.