Article

Generative Adversarial Networks-Based Semi-Supervised Learning for Hyperspectral Image Classification

Zhi He * [ID], Han Liu [ID], Yiwen Wang and Jie Hu

Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, China; liuh88@mail2.sysu.edu.cn (H.L.); yizhigiser@126.com (Y.W.); verohu@163.com (J.H.)
* Correspondence: hezh8@mail.sysu.edu.cn; Tel.: +86-020-8411-3057

Abstract: Classification of hyperspectral image (HSI) is an important research topic in the remote sensing community. Significant efforts (e.g., deep learning) have been concentrated on this task. However, it is still an open issue to classify the high-dimensional HSI with a limited number of training samples. In this paper, we propose a semi-supervised HSI classification method inspired by the generative adversarial networks (GANs). Unlike the supervised methods, the proposed HSI classification method is semi-supervised, which can make full use of the limited labeled samples as well as the sufficient unlabeled samples. Core ideas of the proposed method are twofold. First, the three-dimensional bilateral filter (3DBF) is adopted to extract the spectral-spatial features by naturally treating the HSI as a volumetric dataset. The spatial information is integrated into the extracted features by 3DBF, which is propitious to the subsequent classification step. Second, GANs are trained on the spectral-spatial features for semi-supervised learning. A GAN contains two neural networks (i.e., generator and discriminator) trained in opposition to one another. The semi-supervised learning is achieved by adding samples from the generator to the features and increasing the dimension of the classifier output. Experimental results obtained on three benchmark HSI datasets have confirmed the effectiveness of the proposed method, especially with a limited number of labeled samples.

Keywords: hyperspectral image (HSI); semi-supervised classification; generative adversarial networks (GANs); three-dimensional bilateral filter (3DBF)

1. Introduction

A hyperspectral image [1–4] contains hundreds of continuous narrow spectral bands, spanning the visible to infrared spectrum. Hyperspectral sensors have attracted much interest in remote sensing for providing abundant and valuable information over the last few decades. With the useful information, HSI has played a vital role in many applications, among which classification [5–7] is one of the crucial processing steps that has received enormous attention. The foremost task in hyperspectral classification is to train an effective classifier with the given training set from each class. Therefore, sufficient training samples are crucial to train a reliable classifier. However, in reality, it is time-consuming and expensive to obtain a large number of samples with class labels. This difficulty will result in the curse of dimensionality (i.e., Hughes phenomenon) and will induce the risk of overfitting.

Much work has been carried out to design suitable classifiers to deal with the above-mentioned problems in the last decades. In general, those methods can be categorized into three types, i.e., unsupervised, supervised and semi-supervised methods. Unsupervised methods focus on training models from large unlabeled samples. Since no labeled samples are required, the unsupervised

methods can be easily applied in the hyperspectral processing area. Many unsupervised methods, such as fuzzy clustering [8], fuzzy C-Means method [9], artificial immune algorithm [10], graph-based method [11], have demonstrated impressive results in hyperspectral classification. However, one cannot ensure the relationship between clusters and classes with too little priori knowledge.

Supervised classifiers, which are widely used in hyperspectral classification, can yield improved performance by utilizing the priori information of the class labels. Typical supervised classifiers include the support vector machine (SVM) [12,13], artificial neural networks (ANN) [14] and sparse representation-based classification (SRC) [15,16], etc. SVM is a kind of kernel-based method that aims at exploring the optimal separating hyperplane between different classes, ANN is motivated by the biological learning process of human brain, while SRC stems from the rapid development of compressed sensing in recent years. Versatile as the supervised classifiers are, their performance heavily depends on the number of labeled samples. In contrast to the urgent needs of labeled samples, they ignore the large number of unlabeled samples to assist classification.

Semi-supervised learning is designed to alleviate the "small-sample problem" by utilizing both the limited labeled samples and the wealth of unlabeled samples that can be easily obtained without significant cost. The semi-supervised methods can be roughly divided into four types: (1) generative models [17,18], which estimate the conditional density to obtain the labels of unlabeled samples. (2) Low density separation, which aims to place boundaries in regions where few samples (labeled or unlabeled) existed. One of the state-of-the-art algorithms is the transductive support vector machine (TSVM) [19–21]. (3) Graph-based methods [22–26] that utilize labeled and unlabeled samples to construct graphs and minimize the energy function, and thus, assigning labels to unlabeled samples. (4) Wrapper-based methods, which apply a supervised learning method iteratively and a certain amount of unlabeled samples are labeled in each iteration. The self-training [27,28] and co-training [29,30] algorithms are commonly-used wrapper-based methods.

Notably that the samples within a small neighborhood are likely to belong to the same class and thus, the spatial correlation between neighboring samples can be incorporated into the classification to further improve the performance of the classifiers. For instance, the spatial contextual information [31–39] can be extracted by various spatial filters. Segmentation methods (e.g., watershed segmentation [40] and superpixel segmentation [41,42]) can also be adopted to exploit the spatial homogeneity of the HSI. One can also use the spatial similarity of neighboring samples [43–45] in the classification stages. Regularizations [15,46–53] can be added in the classifiers to refine the classification performance. Different from the above-mentioned vector/matrix-based methods, there are some three-dimension (3D)/tensor-based methods [34,54–57] that respect the 3D nature of the HSI and process the 3D cube as a whole entity. The 3D/tensor-based methods have demonstrated considerable improvement since the joint spectral-spatial structure information is effectively exploited.

However, most of the aforementioned methods can only extract features of the original HSI dataset in a shallow manner. Deep learning [58], which can hierarchically obtain the high-level abstract representation, has recently become a hotspot in the image processing area, especially in hyperspectral classification. Typical deep architectures involve the stacked autoencoder (SAE) [59], deep brief network (DBN) [60] and convolutional neural networks (CNN) [61]. The above-mentioned classification frameworks are supervised, which require a large number of labeled samples for training. Recently, a semi-supervised classifier based on multi-decision labeling and contextual deep learning (i.e., CDL-MD-L) is proposed by [62], which has demonstrated promising results in hyperspectral classification.

In this paper, a generative adversarial networks (GANs)-based semi-supervised method is proposed for hyperspectral classification. To extract the spectral-spatial features, we extend the existing two-dimensional bilateral filter (2DBF) [36,63,64] into its three-dimensional version (i.e., 3DBF), which is a non-iterative method for nonlinear and edge-preserving smoothing. The 3DBF is suitable for spectral-spatial feature extraction since it respects the 3D nature of the HSI cube. Subsequently, the outputs of the previous step can be utilized to train GANs [65,66], which are promising neural

networks that have been the focus of attention in recent years. In this paper, the GANs are trained for semi-supervised classification of HSI to use the limited labeled samples and vast of unlabeled samples. The semi-supervised learning is performed by adding samples from the generators to the extracted features and increasing the dimension of the classifier output.

Compared to the existing literature, the contribution of this paper lies in two aspects:

1. We extract the spectral-spatial features by the 3DBF. Compared to the vector/matrix-based methods, the structural features extracted by the 3DBF can effectively preserve the spectral-spatial information by naturally obeying the 3D form of the HSI and treating the 3D cube as a whole entity.
2. We classify the HSI in a semi-supervised manner by the GANs. Compared to the supervised methods, the GANs can utilize both limited training samples and abundant of unlabeled samples. Compared to the non-adversarial networks, the GANs take advantage of the discriminative models to train the generative network based on game theory.

The remaining part of this paper is organized as follows. Section 2 describes the proposed semi-supervised classification method in detail. Section 3 reports the experimental results and analyses on three benchmark HSI datasets. Finally, discussions and conclusions are drawn in Sections 4 and 5.

2. Proposed Semi-Supervised Method

The conceptual framework of the proposed method is shown in Figure 1, which is composed of two parts: (1) feature extraction; (2) semi-supervised learning. The spectral-spatial features of the original HSI cube I can be extracted by the 3DBF, which is a 3D filter that can obey the 3D nature of the HSI and extract the spectral-spatial features simultaneously. Subsequently, GANs are utilized in the feature space for semi-supervised classification by taking full advantage of both the limited labeled samples and the sufficient unlabeled samples. The classification map can be achieved by visualizing the classification results of different samples.
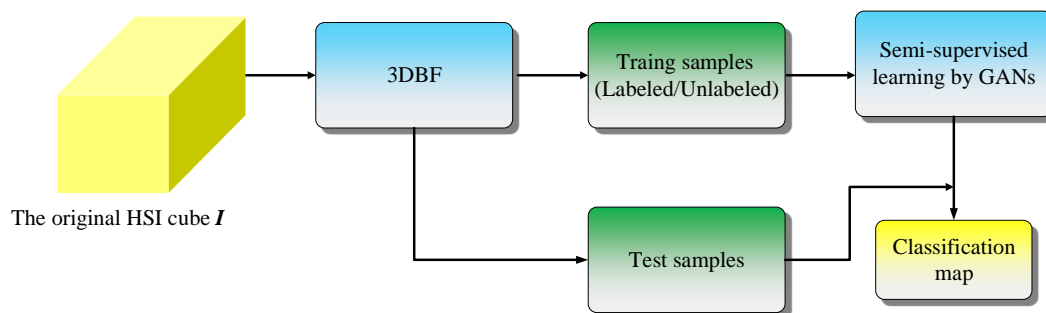

Figure 1. Flowchart of the proposed method.

It is noteworthy that both 3DBF and GANs are of great importance for semi-supervised learning of HSI classification. On the one hand, 3DBF is adopted for extracting the spectral-spatial features of the HSI. As emphasized in Section 1, incorporating spatial information into the hyperspectral classification helps to improve the performance the classifiers, and thus, exploring spectral-spatial feature extraction methods has become an important research topic in the hyperspectral community. In addition, since the HSI data is naturally a 3D cube, the 3D/tensor-based methods are more effective to extract the joint spectral-spatial structure information than the vector/matrix-based methods. As will be shown in Section 3.3, the GANs with the original spectral features (i.e., Spec-GANs) provide much worse performance than the GANs with 3DBF features (i.e., 3DBF-GANs), which further highlights the significance of the 3DBF. On the other hand, GANs are utilized for semi-supervised classification of the HSI. The recent development of deep learning has opened up new opportunities for hyperspectral classification. GANs, which are newly proposed deep architectures for training deep generative

models by a minimax game, have shown promising results in unsupervised/semi-supervised learning. Although the GANs have been successfully employed in various areas and demonstrated remarkable success, the application of GANs in semi-supervised hyperspectral classification has never been addressed in the literature to the best of our knowledge. Therefore, it is valuable for us to represent the first attempt to develop a semi-supervised hyperspectral classification framework based on GANs. In this section, we introduce the detailed procedure of the proposed semi-supervised classification method, elaborating on the spectral-spatial feature extraction based on 3DBF and semi-supervised classification of HSI by GANs.

2.1. Spectral-Spatial Features Extracted by 3D Bilateral Filter

The bilateral filter was originally introduced by [63] under the name "SUSAN". It was then rediscovered by [67] termed as "bilateral filter", which is now the widely used name in the literature. Over the past few years, the bilateral filter has emerged as a powerful tool for several applications, such as image denoising [64] and hyperspectral classification [36]. The great success of the bilateral filter stems from several properties. It is a local, non-iterative and simple filter, which smooths images while preserving edges in terms of a nonlinear combination of the neighboring pixels. Although the bilateral filter has announced impressive results in hyperspectral classification, it is performed in each two-dimensional probability map, and thus ignoring the 3D nature of the HSI cube.

In this paper, we extend the bilateral filter to 3DBF for spectral-spatial feature extraction of the HSI volumetric data. Suppose the original HSI cube can be represented as $I \in \mathbb{R}^{m \times n \times b}$, where m, n and b indicate the number of rows, columns and spectral bands, respectively, the result I^{bf} of the 3DBF, which replaces each pixel in the I by a weighted average of its neighbors, can be defined by

$$I^{bf}(p) = \frac{1}{W^{bf}(p)} \sum_{q \in \mathcal{S}} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I(p) - I(q)|) I(q) \tag{1}$$

with

$$W^{bf}(p) = \sum_{q \in \mathcal{S}} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I(p) - I(q)|) \tag{2}$$

where p refers to the coordinate of the HSI cube I, i.e., $p = (x, y, z), x = 1, \ldots, m, y = 1, 2, \ldots, n,$ $z = 1, 2, \ldots, b$, q indicates the index of the neighborhoods centered at p, W^{bf} denotes the normalizing term of the neighborhood pixels q, $G_{\sigma_s}(\|p - q\|) = \exp(-\|p - q\|^2/2\sigma_s^2)$ and $G_{\sigma_r}(|I(p) - I(q)|) = \exp(-|I(p) - I(q)|^2/2\sigma_r^2)$ are the Gaussian filters measuring the distance in the 3D image domain (i.e., the spectral-spatial domain $\mathcal{S}$) and the distance on the intensity axis (i.e., the range domain $\mathcal{R}$), respectively.

To speed up the implementation, we decompose the 3DBF into a convolution followed by two nonlinearities based on signal processing grounds. Note that the nonlinearity of the 3DBF (see Equation (1)) originates from the division by W^{bf} and the dependency on the intensities by $G_{\sigma_r}(|I(p) - I(q)|)$, we study each point separately and isolate them during computation. Multiplying both sides of Equation (1) by W^{bf}, Equations (1) and (2) can be rewritten as

$$\begin{pmatrix} W^{bf}(p) I^{bf}(p) \\ W^{bf}(p) \end{pmatrix} = \sum_{q \in \mathcal{S}} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I(p) - I(q)|) \begin{pmatrix} I(q) \\ 1 \end{pmatrix} \tag{3}$$

We then define a function W whose value is 1 everywhere (W is a function whose value is 1 everywhere, i.e., $W((x, y, z)) = 1, x = 1, \ldots, m, y = 1, 2, \ldots, n, z = 1, 2, \ldots, b$. Therefore, the size of W

is the same as that of the original HSI cube) to maintain the weighted mean property of the 3DBF and represent Equation (3) as

$$
\begin{pmatrix} W^{bf}(\boldsymbol{p})\boldsymbol{I}^{bf}(\boldsymbol{p}) \\ W^{bf}(\boldsymbol{p}) \end{pmatrix} = \sum_{\boldsymbol{q}\in\mathcal{S}} G_{\sigma_s}(\|\boldsymbol{p}-\boldsymbol{q}\|)G_{\sigma_r}(|\boldsymbol{I}(\boldsymbol{p})-\boldsymbol{I}(\boldsymbol{q})|) \begin{pmatrix} W(\boldsymbol{q})\boldsymbol{I}(\boldsymbol{q}) \\ W(\boldsymbol{q}) \end{pmatrix} \tag{4}
$$

The above-mentioned Equation (4) can be equivalently expressed as

$$
\begin{pmatrix} W^{bf}(\boldsymbol{p})\boldsymbol{I}^{bf}(\boldsymbol{p}) \\ W^{bf}(\boldsymbol{p}) \end{pmatrix} = \sum_{\boldsymbol{q}\in\mathcal{S}}\sum_{\zeta\in\mathcal{R}} G_{\sigma_s}(\|\boldsymbol{p}-\boldsymbol{q}\|)G_{\sigma_r}(|\boldsymbol{I}(\boldsymbol{p})-\zeta|)\delta(\zeta-\boldsymbol{I}(\boldsymbol{q})) \begin{pmatrix} W(\boldsymbol{q})\boldsymbol{I}(\boldsymbol{q}) \\ W(\boldsymbol{q}) \end{pmatrix} \tag{5}
$$

where $\mathcal{R}$ denotes the intensity interval, $\delta(\zeta)$ is the Kronecker symbol with $\delta(\zeta) = 1$ if $\zeta = 0$, and $\delta = 0$ otherwise. Specifically, $\delta(\zeta - \boldsymbol{I}(\boldsymbol{q})) = 1$ if and only if $\zeta = \boldsymbol{I}(\boldsymbol{q})$. The sum in Equation (5) is over the product space $\mathcal{S} \times \mathcal{R}$, on which we express the functions by lowercases. That means, g_{δ_s,δ_r} represents a Gaussian kernel given by

$$
g_{\delta_s,\delta_r} : (\boldsymbol{x} \in \mathcal{S}, \zeta \in \mathcal{R}) \longmapsto G_{\delta_s}(\|\boldsymbol{x}\|)G_{\delta_r}(|\zeta|) \tag{6}
$$

Based on Equation (5), two functions i and w can be build on $\mathcal{S} \times \mathcal{R}$ by

$$
i : (\boldsymbol{x} \in \mathcal{S}, \zeta \in \mathcal{R}) \longmapsto \boldsymbol{I}(\boldsymbol{x}) \tag{7}
$$

and

$$
w : (\boldsymbol{x} \in \mathcal{S}, \zeta \in \mathcal{R}) \longmapsto \delta(\zeta - \boldsymbol{I}(\boldsymbol{x}))W(\boldsymbol{x}) \tag{8}
$$

Observed from the definitions of i and w in Equations (7) and (8), we have

$$
\begin{aligned}
\boldsymbol{I}(\boldsymbol{x}) &= i(\boldsymbol{x}, \boldsymbol{I}(\boldsymbol{x})) & (9) \\
W(\boldsymbol{x}) &= w(\boldsymbol{x}, \boldsymbol{I}(\boldsymbol{x})) & (10) \\
w(\boldsymbol{x}, \zeta) &= 0, \forall \zeta \neq \boldsymbol{I}(\boldsymbol{x}) & (11)
\end{aligned}
$$

Let the input of g_{δ_s,δ_r} be $(\boldsymbol{p} - \boldsymbol{q}, \boldsymbol{I}(\boldsymbol{p}) - \zeta)$, the input of i and w be $(\boldsymbol{q}, \zeta)$, Equation (5) becomes

$$
\begin{aligned}
\begin{pmatrix} W^{bf}(\boldsymbol{p})\boldsymbol{I}^{bf}(\boldsymbol{p}) \\ W^{bf}(\boldsymbol{p}) \end{pmatrix} &= \sum_{(\boldsymbol{q},\zeta)\in\mathcal{S}\times\mathcal{R}} g_{\delta_s,\delta_r}(\boldsymbol{p}-\boldsymbol{q}, \boldsymbol{I}(\boldsymbol{p})-\zeta) \begin{pmatrix} w(\boldsymbol{q},\zeta)i(\boldsymbol{q},\zeta) \\ w(\boldsymbol{q},\zeta) \end{pmatrix} \\
&= \left[g_{\delta_s,\delta_r} \otimes \begin{pmatrix} wi \\ w \end{pmatrix} \right](\boldsymbol{p}, \boldsymbol{I}(\boldsymbol{p}))
\end{aligned} \tag{12}
$$

where "$\otimes$" indicates the convolution operator.

Therefore, the 3DBF can be modeled by

$$
\boldsymbol{I}^{bf}(\boldsymbol{p}) = \frac{w^{bf}(\boldsymbol{p}, \boldsymbol{I}(\boldsymbol{p}))i^{bf}(\boldsymbol{p}, \boldsymbol{I}(\boldsymbol{p}))}{w^{bf}(\boldsymbol{p}, \boldsymbol{I}(\boldsymbol{p}))} \tag{13}
$$

where the functions w^{bf} and i^{bf} are defined as $(w^{bf}i^{bf}, w^{bf}) = g_{\delta_s,\delta_r} \otimes (wi, w)$.

In hyperspectral analysis, the 3D image domain (i.e., the spectral-spatial domain $\mathcal{S}$) is a *xyz* volume and the range domain $\mathcal{R}$ is a simple axis labelled ζ. As described in Equation (13), the 3DBF can be achieved by the following three steps:

- Convolve *wi* and *w* with a Gaussian defined on *xyzζ*. In this step, *wi* and *w* are "blurred" into $w^{bf}(x, y, z, \zeta)i^{bf}(x, y, z, \zeta)$ and $w^{bf}(x, y, z, \zeta)$, respectively.
- Obtain $i^{bf}(x, y, z, \zeta)$ by dividing $w^{bf}(x, y, z, \zeta)i^{bf}(x, y, z, \zeta)$ by $w^{bf}(x, y, z, \zeta)$;

- Compute the value of i^{bf} at (x, y, z, ζ) to get the filtered result $I^{bf}(x, y, z)$.

Moreover, the 3DBF can be accelerated by downsample and upsample without changing the major steps of the implementation. That is, we downsample (wi, w) to obtain $(w_\downarrow i_\downarrow, w_\downarrow)$, perform the convolution to generate $(w_\downarrow^{bf} i_\downarrow^{bf}, w_\downarrow)^{bf}$, followed by upsample $(w_\downarrow^{bf} i_\downarrow^{bf}, w_\downarrow^{bf})$ to get $(w_{\downarrow\uparrow}^{bf} i_{\downarrow\uparrow}^{bf}, w_{\downarrow\uparrow}^{bf})$. The remaining steps are the same as the above-mentioned steps 2 and 3. To sum up, the schematic diagram of the 3DBF can be depicted in Figure 2, by which the original HSI cube I is filtered and the spectral-spatial feature cube I^{bf} is obtained. It is worth underlining that the dimension of the 3DBF cube I^{bf} is the same as that of the original HSI cube, i.e., $I^{bf} \in \mathbb{R}^{m \times n \times b}$. As will be shown in Figures 9 and 10, the spectral and spatial profiles of the 3DBF smooth the original data while still preserving edges.
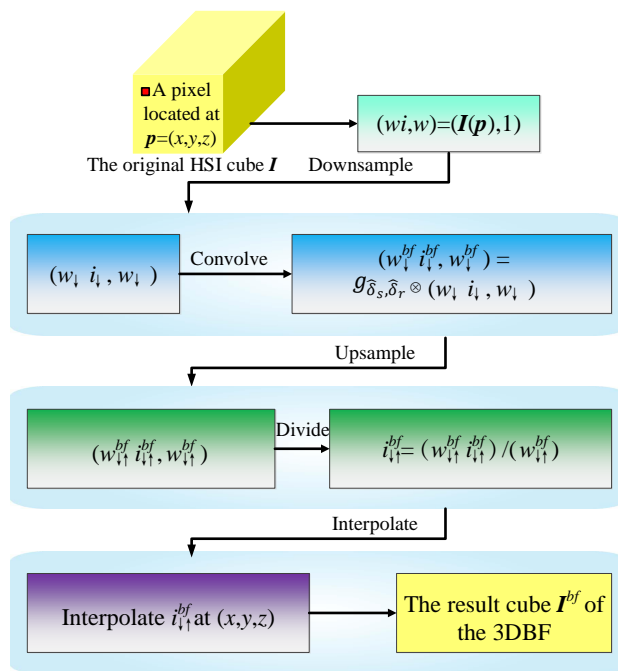

Figure 2. Schematic diagram of the 3DBF.

2.2. Semi-Supervised Classification of HSI by Generative Adversarial Networks

2.2.1. Brief of Generative Adversarial Networks

GANs are newly proposed deep architectures based on adversarial nets to train the model in an adversarial fashion to generate data mimicking certain distributions. Unlike the other deep learning methods, a GAN is an architecture around two functions (see Figure 3), i.e., a generator G, which can map a sample from a random uniform distribution to the data distribution, and a discriminator D, which is trained to distinguish whether a sample belongs to the real data distribution. In GANs, the generator and discriminator are learned jointly based on game theory. The generator G and the discriminator D can be trained in an alternating manner. In each step, G produces a sample from the random noise z that may fool D, and D is then presented the real data samples as well as the samples generated by G to classify the samples as "real" or "fake". Subsequently, G is rewarded for producing samples that can "fool" D and D for correct classification. Both functions are updated and the iteration stops until a Nash equilibrium is achieved. In greater detail, let $D(s)$ be the probability

that **s** comes from the real data rather than the generator, *G* and *D* play a minimax game with the following value function

$$\min_{G} \max_{D} V(D, G) = \mathbb{E}_{\boldsymbol{s} \sim p_{\text{data}}(\boldsymbol{s})} \Big[\log D(\boldsymbol{s}) \Big] + \mathbb{E}_{\boldsymbol{z} \sim p_z(\boldsymbol{z})} \Big[\log(1 - D(G(\boldsymbol{z}))) \Big] \tag{14}$$

Much work has been carried out to improve the GAN since it was pioneered by Goodfellow et al. [65] in 2014. Two remarkable aspects can be highlighted: theory and application. On the one hand, several improved versions of GANs in aspects of stability of training, perceptual quality, etc., have been proposed in recent literature, including the well-known deep convolutional GAN (DC-GAN) [68], conditional GAN (C-GAN) [69], Laplacian pyramid GAN (LAP-GAN) [70], information-theoretic extension to the GAN (Info-GAN) [71], unrolled GAN [72] and Wasserstein GAN (W-GAN) [73]. On the other hand, recent work has also shown that GANs can provide very successful results in image generation [74], image super resolution [75], image inpainting [76] and semi-supervised learning [77].
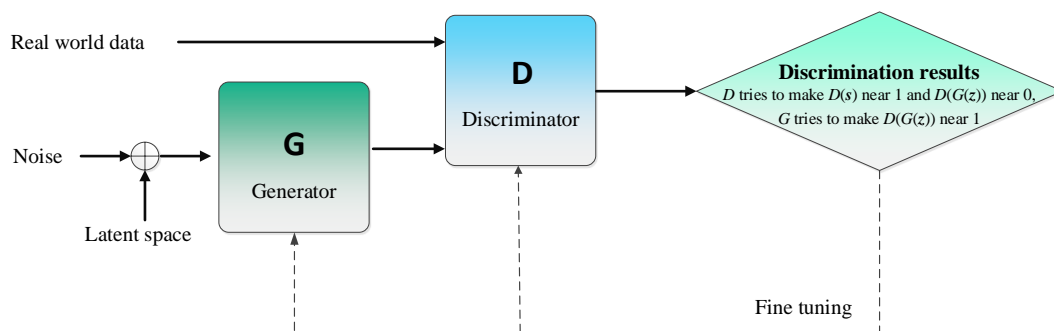

Figure 3. The general GANs architectures.

2.2.2. Generative Adversarial Networks for Classification

GANs, which can train deep generative models with a minimax game, have recently emerged as powerful tools for unsupervised and semi-supervised classification. Several unsupervised/ semi-supervised techniques motivated by the GANs have sprung up over the past few years to overcome the difficulties of labeling large amounts of training samples. For instance, DC-GAN is proposed in [68] to bridge the gap between the success of the CNN for supervised and unsupervised learning. Several constraints are evaluated to make the convolutional GANs stable to train, and the trained discriminators are applied for image classification tasks, resulting in competitive performance with other unsupervised methods. Info-GAN is proposed in [71] to learn disentangled representations in a completely unsupervised manner. As an information-theoretic extension to the GAN, the Info-GAN maximizes the mutual information between a small subset of the latent variables and the observation, and therefore, interpretable and disentangled representations can be learned. Categorical GAN (CatGAN) [77], which is a framework for robust unsupervised and semi-supervised learning, combines ANN classifiers with an adversarial generative model that regularizes a discriminatively trained classifier. By heuristically understanding the non-convergence problem, an improved semi-supervised learning method is proposed in [66], which can be regarded as a continuation and refinement of the effort in [77]. Moreover, Premachandran and Yuille [78] learns a deep network by generative adversarial training. Features learned by adversarial training is fused with a traditional unsupervised classification approach, i.e., k-means clustering, and the combination produces better results than direct prediction. In situation of semi-supervised classification, the adversarial training has the potential to outperform supervised learning.

Note that different versions of GANs have different objective functions and procedures, it is hard to obtain a unified architecture for describing the unsupervised/semi-supervised techniques. In this

section, we try to give a schematic illustration of the procedure for unsupervised/semi-supervised learning in Figure 4, which contains the main steps in most of the scenarios but not all of them. It is noteworthy that the logistic regression classifier based on the soft-max function is employed to discriminate different classes in Figure 4. That means, by applying the soft-max function, the class probabilities of *s* can be expressed as

$$p_{\text{model}}(c = j|s) = \frac{\exp(l_j)}{\sum\limits_{c=1}^{C} \exp(l_c)}, j = 1, 2, \ldots, C \tag{15}$$

and the class label of *s* can be determined by

$$\text{class}(s) = \arg\max_j p_{\text{model}}(c = j|s) \tag{16}$$

In addition, despite remarkable success of GANs, their applications in semi-supervised classification of HSI are surprisingly unstudied to the best of our knowledge. Therefore, this study represents the first attempt to develop a semi-supervised classification framework for the HSI.
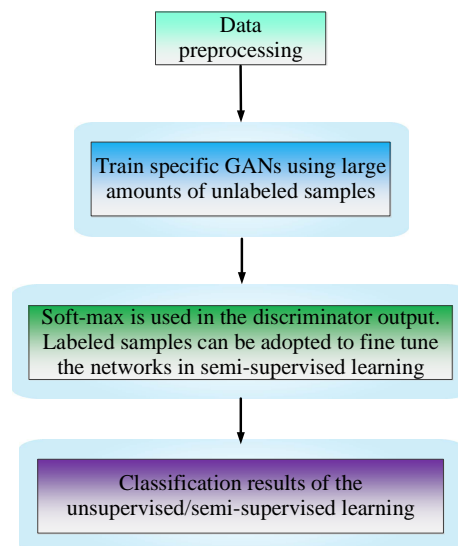

Figure 4. Schematic illustration of the procedure for unsupervised/semi-supervised learning based on GANs.

2.2.3. Hyperspectral Classification Framework Using Generative Adversarial Networks

In hyperspectral classification, a standard classifier assigns each sample *s* to one of the *C* possible classes based on the training samples available for each class. For instance, a logistic regression classifier takes *s* as input and outputs a *C*-dimensional vector, which can be turned into the class probabilities by soft-max $p_{\text{model}}(c = j|s) = \frac{\exp(l_j)}{\sum\limits_{c=1}^{C} \exp(l_c)}$. Classifiers like this usually have a cross-entropy objective function in supervised scenario. That means, a discriminative model can be trained by minimizing the objective function between observed labels and the model predictive distribution $p_{\text{model}}(c|s)$. However, the supervised learning usually needs enough labeled training samples to guarantee the representativeness and prevent the classifier from overfitting, especially for a deep discriminative model with huge parameter volume such as CNN. The strong demand for abundant training samples conflicts with the fact that the labels of the samples are extremely difficult and expensive to identify. At the same time, there are vast of unlabeled samples in the HSI. Therefore,

we propose a GANs-based classification method [65,66] to simultaneously utilize both the limited labeled samples and the sufficient unlabeled samples in a semi-supervised fashion.

To establish a new semi-supervised hyperspectral classification framework based on GANs, we add the generated samples to the HSI dataset and denote them as the $(C+1)$th class. The dimension of the classifier output is correspondingly increased from C to $(C+1)$. The probability when s comes from G can be represented as $p_{\text{model}}(c=C+1|s)$, which is a substitution of $1-D(s)$ in the objective function $V(D,G)$ of the original GANs [65]. Note that the unlabeled training samples belong to the former C classes, we can learn from those unlabeled samples to improve the classification performance by maximizing $\log p_{\text{model}}(c \in 1,2,\ldots,C|s)$.

Without loss of generality, assuming half of the dataset consists of real data and half is the generated data, the loss function L of the classifier yields

$$
\begin{aligned}
L &= -\mathbb{E}_{s,c\sim p_{\text{data}(s,c)}}\Big[\log p_{\text{model}}(c|s)\Big] - \mathbb{E}_{s\sim G}\Big[\log p_{\text{model}}(c=C+1|s)\Big] \\
&= L_{\text{supervised}} + L_{\text{unsupervised}}
\end{aligned}
\tag{17}
$$

$$
L_{\text{supervised}} = -\mathbb{E}_{s,c\sim p_{\text{data}(s,c)}}\log p_{\text{model}}(c|s,c<C+1)
\tag{18}
$$

$$
\begin{aligned}
L_{\text{unsupervised}} = -\Big\{ &\mathbb{E}_{s\sim p_{\text{data}(s)}}\log\Big[1-p_{\text{model}}(c=C+1|s)\Big] \\
+ &\mathbb{E}_{s\sim G}\log\Big[p_{\text{model}}(c=C+1|s)\Big]\Big\}
\end{aligned}
\tag{19}
$$

where $L_{\text{supervised}}$ represents the negative log probability of the label with the data is from the real HSI features, $L_{\text{unsupervised}}$ equals the standard GAN game-value function in case we substitute $D(s)=1-p_{\text{model}}(c=C+1|s)$ into Equation (19)

$$
L_{\text{unsupervised}} = -\Big\{\mathbb{E}_{s\sim p_{\text{data}(s)}}\log D(s) + \mathbb{E}_{z\sim\text{noise}}\log\Big[1-D(G(z))\Big]\Big\}
\tag{20}
$$

According to the Output Distribution Matching (ODM) cost theory of [79], if we have $\exp[l_j(s)] = f(s)p(c=j,s), \forall j < C+1$ and $\exp[l_{C+1}(s)] = f(s)p_G(s)$ for some undetermined scaling function $f(s)$, the unsupervised loss will be consistent with the supervised loss. As such, by combining $L_{\text{supervised}}$ and $L_{\text{unsupervised}}$, we can get the total cross entropy loss L, whose optimal solution can be estimated by minimizing both loss functions jointly.

Moreover, to address the instability of the unsupervised optimization part related to the GANs, we adopt a strategy called feature matching to substitute the traditional way of training the generator G by requiring it to match the statistics characteristics of the real data. In greater detail, the generator G is trained to match the expected value of the output $d(s)$ on an intermediate layer in the discriminator D. By optimizing an alternative objective function defined as $\left\|\mathbb{E}_{s\sim p_{\text{data}(s)}}d(s) - \mathbb{E}_{z\sim p_z(z)}d(G(z))\right\|_2^2$, we obtain a fixed point where G matches the distribution of training data. Based on the above analysis, we show a visual illustration of the semi-supervised hyperspectral classification method by GANs in Figure 5. The network parameters of the generator G and the discriminator D in Figure 5 are trained by optimizing the loss function in Equation (17). The unlabeled data is taken as the true data $s \sim p_{\text{data}}$ in Equation (19) to train both generator G and discriminator D. Moreover, the latent space of the generator G is chosen from the unlabeled data (To be exact, the latent space can also be chosen from the labeled data by ignoring the class labels), the noise follows the uniform distribution, and the output of the generator G is the fake data. By jointly minimizing the loss functions in Equation (17), the parameters of the generator G are updated to fool the discriminator D, and the fake examples are generated accordingly. The logistic regression classifier based on the soft-max function is adopted to perform the multi-class classification in the GANs. It is notable that the actual differences between the traditional GANs and the modified GANs used in this paper lie in threefold: (1) the objective functions are changed to make full use of both labeled and unlabeled samples; (2) the output layer

of the discriminator is modified from binary classification to multi-class semi-supervised learning; (3) feature matching is adopted to improve the stability of the traditional GANs.
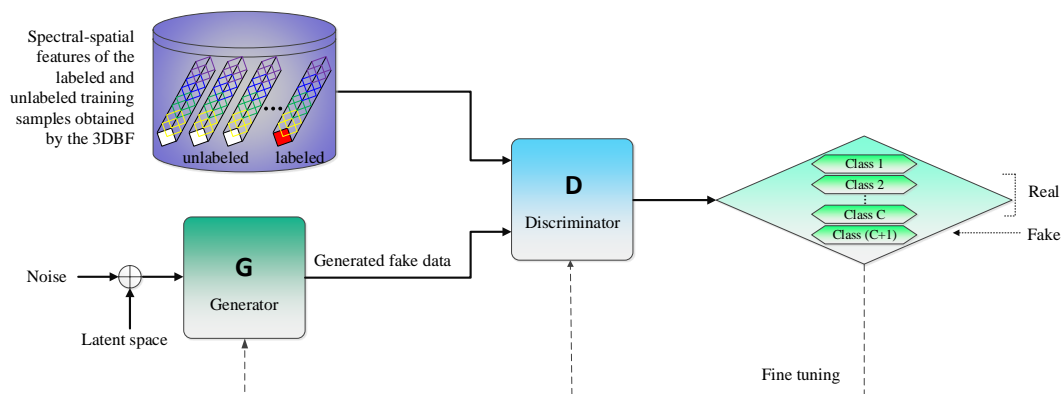

Figure 5. A visual illustration of the semi-supervised hyperspectral classification method by GANs.

3. Experimental Section

In this section, we investigate the performance of the proposed method (abbreviated as 3DBF-GANs for simplicity) on three benchmark HSI datasets. A series of experiments are conducted to perform a comprehensive comparison with other state-of-the-art methods, including 2DBF [64], SVM [12], Laplacian SVM (LapSVM) [22,24] and CDL-MD-L [62]. 2DBF and 3DBF are feature extraction methods, SVM is a widely-used supervised classifier, while LapSVM, GANs and CDL-MD-L are classifiers based on semi-supervised learning. Moreover, the original spectral features are also considered as a baseline for comparison.

3.1. Dataset Description

In the experiments, three publicly available hyperspectral datasets (i.e., Indian Pines data, University of Pavia data and Salinas data) are employed as benchmark datasets. What follows are details of the three hyperspectral datasets.

1. **Indian Pines data**: the first dataset was captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the agricultural Indian Pines test site in the Northwestern Indiana, USA, on 12 June 1992. The original image contains 224 spectral bands. After removing 4 bands full of zero and 20 bands affected by noise and water-vapor absorption, 200 bands are left for experiments. It consists of 145×145 pixels with a spatial resolution of 20 m per pixel, and the spectral coverage ranging from 0.4 to 2.5 μm. Figure 6 depicts the color composite of the image as well as the ground truth map. There are 16 classes of interest and the number of samples in each class is displayed in Table 1, whose background color denotes different classes of land-covers. Since the number of samples is unbalanced and the spatial resolution is relatively low, it poses a big challenging to the classification task.

2. **University of Pavia data**: the second dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over an urban area surrounding the University of Pavia, northern Italy, on 8 July 2002. The original data contains 115 spectral bands ranging from 0.43 to 0.86 μm and the size of each band is 610×340 with a spatial resolution of 1.3 m per pixel. After removing 12 noisiest channels, 103 bands remained for experiments. The dataset contains 9 classes with various types of land-covers. The color composite image together with the ground truth data are shown in Figure 7. The detailed number of samples in each class is listed in Table 2, whose background color also corresponds to the color in Figure 7.

3. **Salinas data**: the third dataset was collected by the AVIRIS sensor over the Salinas Valley, Southern California, USA, on 8 October 1998. The original dataset contains 224 spectral bands

covering from the visible to short-wave infrared light. After discording 20 water absorption bands, 204 bands are preserved for experiments. This dataset consists of 512×217 pixels with a spatial resolution of 3.7 m per pixel. The color composite of the image and the ground truth are plotted in Figure 8, which contains 16 classes of interest. The detailed number of classes in each class is shown in Table 3, whose background color represents different classes of land-covers.

(a)   (b)
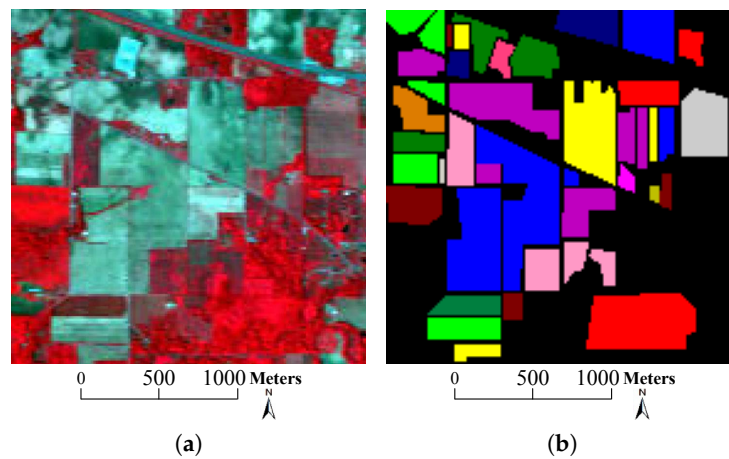
Figure 6. Indian Pines data. (**a**) Three-band false color composite and (**b**) ground truth data with 16 classes.

Table 1. Number of samples (NoS) used in the Indian Pines data.

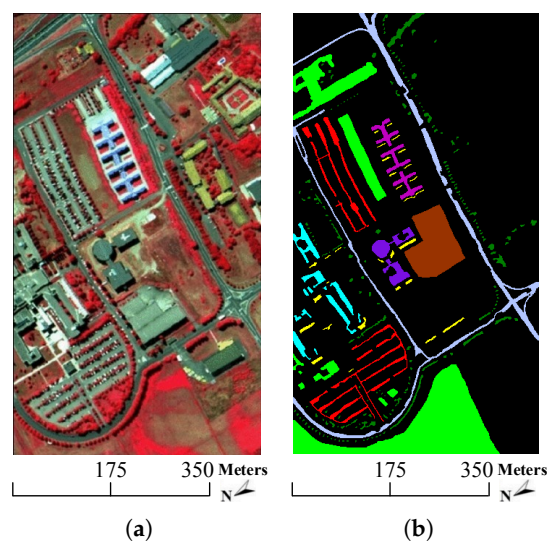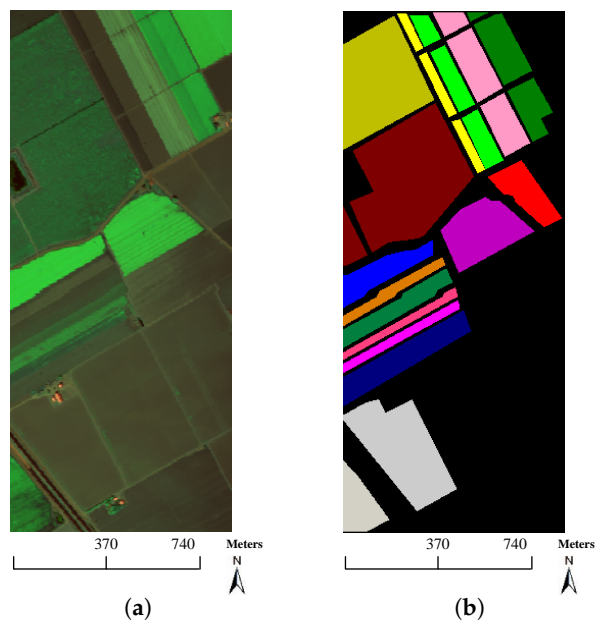| Class	Name	NoS	Class	Name	NoS
1	alfalfa	54	9	oats	20
2	corn-no till	1434	10	soybean-no till	968
3	corn-min till	834	11	soybean-min till	2468
4	corn	234	12	soybean-clean till	614
5	grass/pasture	497	13	wheat	212
6	grass/trees	747	14	woods	1294
7	grass/pasture-mowed	26	15	bldg-grass-tree-drives	380
8	hay-windrowed	489	16	stone-steel towers	95
Total					10,366

(a)   (b)

Figure 7. University of Pavia data. (**a**) Three-band false color composite and (**b**) ground truth data with 9 classes.

Table 2. NoS used in the University of Pavia data.

Class	Name	NoS	Class	Name	NoS
1	asphalt	6631	6	bare soil	5029
2	meadows	18,649	7	bitumen	1330
3	gravel	2099	8	bricks	3682
4	trees	3064	9	shadows	947
5	metal sheets	1345	Total		42,776

Figure 8. Salinas data. (**a**) Three-band false color composite and (**b**) ground truth data with 16 classes.

Table 3. NoS used in the Salinas data.

Class	Name	NoS	Class	Name	NoS
1	brocoli-green-weeds-1	2009	9	soil-vinyard-develop	6203
2	brocoli-green-weeds-2	3726	10	corn-senesced-green-weeds	3278
3	fallow	1976	11	lettuce-romaine-4wk	1068
4	fallow-rough-plow	1394	12	lettuce-romaine-5wk	1927
5	fallow-smooth	2678	13	lettuce-romaine-6wk	916
6	stubble	3959	14	lettuce-romaine-7wk	1070
7	celery	3579	15	vinyard-untrained	7268
8	grapes-untrained	11,271	16	vinyard-vertical-trellis	1807
Total					54,129

3.2. Experimental Setup

In order to evaluate the performance of the proposed 3DBF-GANs method, we compare it with some other algorithms, i.e., 2DBF, SVM, LapSVM, and CDL-MD-L. The original spectral features (abbreviated as "Spec") are also considered in the experiments. Specifically, the "Spec", 2DBF and 3DBF are feature extraction methods, while SVM, LapSVM and CDL-MD-L are supervised/semi-supervised classifiers. The LapSVM, which is a graph-based semi-supervised learning method, introduces an additional manifold regularizer on the geometry of both unlabeled and labeled data in terms of the Graph Laplacian. It has been applied to hyperspectral classification and the results have demonstrated the advantage of this graph-based method in semi-supervised classification of the HSI. As to the GANs, the standard framework is used, except for adding a softmax classifier in the output of the discriminator and adopting feature matching to improve the stability of

the original GANs. By combining the feature extraction and classification methods in pairs, 12 methods (i.e., Spec-SVM, Spec-LapSVM, Spec-GANs, Spec-CDL-MD-L, 2DBF-SVM, 2DBF-LapSVM, 2DBF-CDL-MD-L, 2DBF-GANs, 3DBF-SVM, 3DBF-LapSVM, 3DBF-CDL-MD-L and 3DBF-GANs) are obtained for comparison. Since the spectral-spatial information is used in the original CDL-MD-L, Spec-CDL-MD-L and 3DBF-CDL-MD-L denote the input of the CDL-MD-L is the original HSI and the dataset given by 3DBF, respectively.

In the experiments, a training/test sample is a single pixel, whose size is $1 \times b$. Each pixel can be taken as the feature of a certain class and classified by the discriminator of the GANs or other classifiers. Each pixel corresponds to a unique label. The whole cube contains many pixels and therefore, has lots of labels. All the HSI datasets are normalized between zero and one at the beginning of the experiments. All the experiments are implemented on the normalized hyperspectral datasets, whose available data is randomly divided into two parts, i.e., about 60% for training and the rest for testing. In all the datasets, very limited labeled samples, i.e., 5 samples per class, are randomly selected from the training samples as labeled samples, and the remaining ones are used as unlabeled samples. The experiments are repeated ten times using random selection of training and test sets, and the average accuracies are reported. To assess the experimental results quantitatively, we compare the aforementioned methods by three popular indexes, i.e., overall accuracy (OA), average accuracy (AA) and kappa coefficient (κ). Moreover, the F-Measure of various methods is also compared.

For the parameter settings, since the number of labeled samples is limited, leave one out cross validation is adopted in this paper. The range of the filtering size σ_s and blur degree σ_r in the 2DBF are selected in the range of $[1, 2, \ldots, 9]$ and $[0.1, 0.2, \ldots, 0.5]$, respectively, whereas both σ_s and σ_r in the 3DBF are chosen from $[5, 10, \ldots, 50]$. In the SVM and LapSVM, radial basis function (RBF) kernels are adopted. The RBF parameter γ is obtained from the range $[2^{-2}, 2^{-1}, \ldots, 2^{10}]$ and the penalty term is set to 60. 4 spectral neighbors are adopted to calculate the Laplacian graph in the LapSVM. Three layers are used in the CDL-MD-L, whose window size and the number of hidden units are set to the same as [62]. The generator in the GANs has two hidden layers, and the number of units is set to 500 and 300, respectively. In the discriminator, three hidden layers are adopted, and the number of units is set to 300, 200 and 150, respectively. Gaussian noise is added to the output of each layer of the discriminator. Moreover, the learning rate and training epoch are set to 0.001 and 100, respectively.

3.3. Experimental Results

To demonstrate the effectiveness of the 3DBF for spectral-spatial feature extraction, we compare the spectral profiles of the pixel (18,6) from the original Indian Pines data, and the features obtained by the 2DBF and the 3DBF in Figure 9. Moreover, the spatial scenes of the 4th, 22nd, 34th bands are compared in Figure 10. As can be seen, the profiles of 3DBF preserve the trend of the original data while provide smoother features in both spectral and spatial domains.
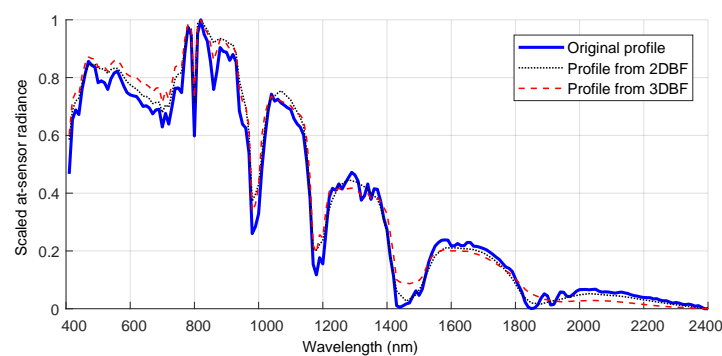

Figure 9. The spectral profiles of the pixel (18,6) from the original Indian Pines data, the 2DBF and the 3DBF.
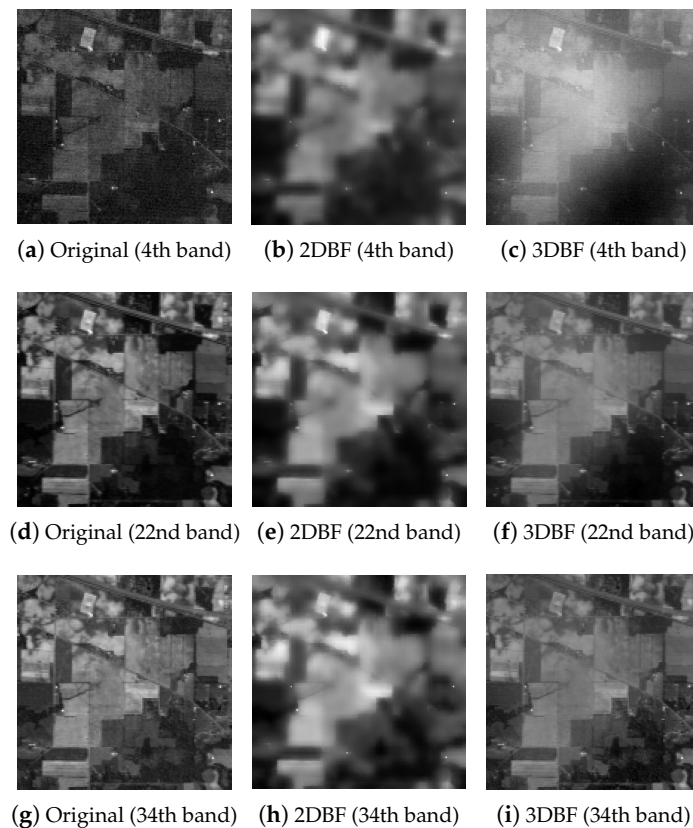
(**a**) Original (4th band) (**b**) 2DBF (4th band) (**c**) 3DBF (4th band)

(**d**) Original (22nd band) (**e**) 2DBF (22nd band) (**f**) 3DBF (22nd band)

(**g**) Original (34th band) (**h**) 2DBF (34th band) (**i**) 3DBF (34th band)

Figure 10. Spatial scenes of the 4th, 22nd, 34th bands. (**a**,**d**,**g**) are chosen from the original Indian Pines data, (**b**,**e**,**h**) are obtained by the 2DBF, and (**c**,**f**,**i**) are obtained by the 3DBF.

The qualitative evaluations of various methods are shown in Tables 4–6, and the classification maps are also visually compared in Figures 11–13. Based on the above-mentioned experimental results, a few observations and discussions can be highlighted. It can be first seen that, the methods (i.e., Spec-SVM, 2DBF-SVM, and 3DBF-SVM) using only the limited labeled training samples provide worse classification performance than the semi-supervised methods that take the unlabeled training samples into consideration. This stresses yet again the importance of unlabeled samples for HSI classification. For instance, it is observed from Table 4 that the SVM leads to lower classification accuracies than other classifiers (i.e., LapSVM, CDL-MD-L and GANs). Taking the same original "Spec" features as inputs, the OA of SVM is 2.15%, 23.28% and 9.49% lower than those of the LapSVM, CDL-MD-L and GANs, respectively. Similar properties can also be found in Tables 5 and 6. The above-mentioned phenomena demonstrate the effectiveness of utilizing the abundant unlabeled samples for the HSI data.

Second, the "Spec"-based features provide higher classification errors than the 2DBF/3DBF-based features. As shown in Table 5, the OA, AA, κ and F-Measure of Spec-SVM are lower than those of the 2DBF-SVM and 3DBF-SVM. Similarly, the OA, AA, κ and F-Measure of Spec-LapSVM/ CDL-MD-L/GANs are also lower than 2DBF-LapSVM/CDL-MD-L/GANs and 3DBF-LapSVM/ CDL-MD-L/GANs. It is also clearly visible that more scattered noise is generated in Figure 12a than in Figure 12e,i. This is due to the fact that the "Spec" features based only on spectral characteristics, while 2DBF and 3DBF methods can effectively incorporate the spatial information. Since the CDL-MD-L can make use of both spectral and spatial information in the classification process, the classification accuracies of Spec-CDL-MD-L are much higher than those of the Spec-SVM, Spec-LapSVM and Spec-GANs. As shown in Table 5, the OA of Spec-CDL-MD-L is at least 8% higher than other classifiers. Moreover, with the same classifiers, the 3DBF performs much better than 2DBF. For instance, the OA

of 3DBF-GANs in Table 5 is about 4% higher than that of the 2DBF-GANs. The reason for good results of 3DBF is that it exploits the spectral-spatial features by obeying the 3D nature of the HSI cube.

Finally, as to different classifiers, the GANs with 2DBF or 3DBF features provides better or comparable classification results as compared with SVM, LapSVM and CDL-MD-L. It is observed from Table 4 that the OA of 2DBF-GANs is 19.65%, 17.02% and 0.25% higher than those of the 2DBF-SVM, 2DBF-LapSVM and 2DBF-CDL-MD-L, respectively, the OA of 3DBF-GANs is also much higher than 3DBF-SVM and 3DBF-LapSVM, and slightly higher than 3DBF-CDL-MD-L. Classification results of the University of Pavia data (see Table 5) and the Salinas data (see Table 6) also yield similar properties. Specifically, it is noteworthy that the "meadows" (i.e., class 2) and "bare soil" (i.e., class 6) in the University of Pavia are difficult to be separated, and the classification accuracies of those two classes obtained by the 3DBF-GANs outperform other methods (see Table 5). Moreover, the GANs with the original spectral features are much inferior to the CDL-MD-L. As shown in Table 4, the OA of Spec-GANs is 13.79% less than that of the Spec-CDL-MD-L. In Table 5 (or Table 6), the OA of Spec-GANs is also 8.17% (or 5.55%) lower than the Spec-CDL-MD-L. The main reason why Spec-GANs obtains poor results is the ignorance of spatial information. In a nutshell, the afore-mentioned analysis validates the effectiveness of the proposed 3DBF-GANs method in semi-supervised hyperspectral classification.
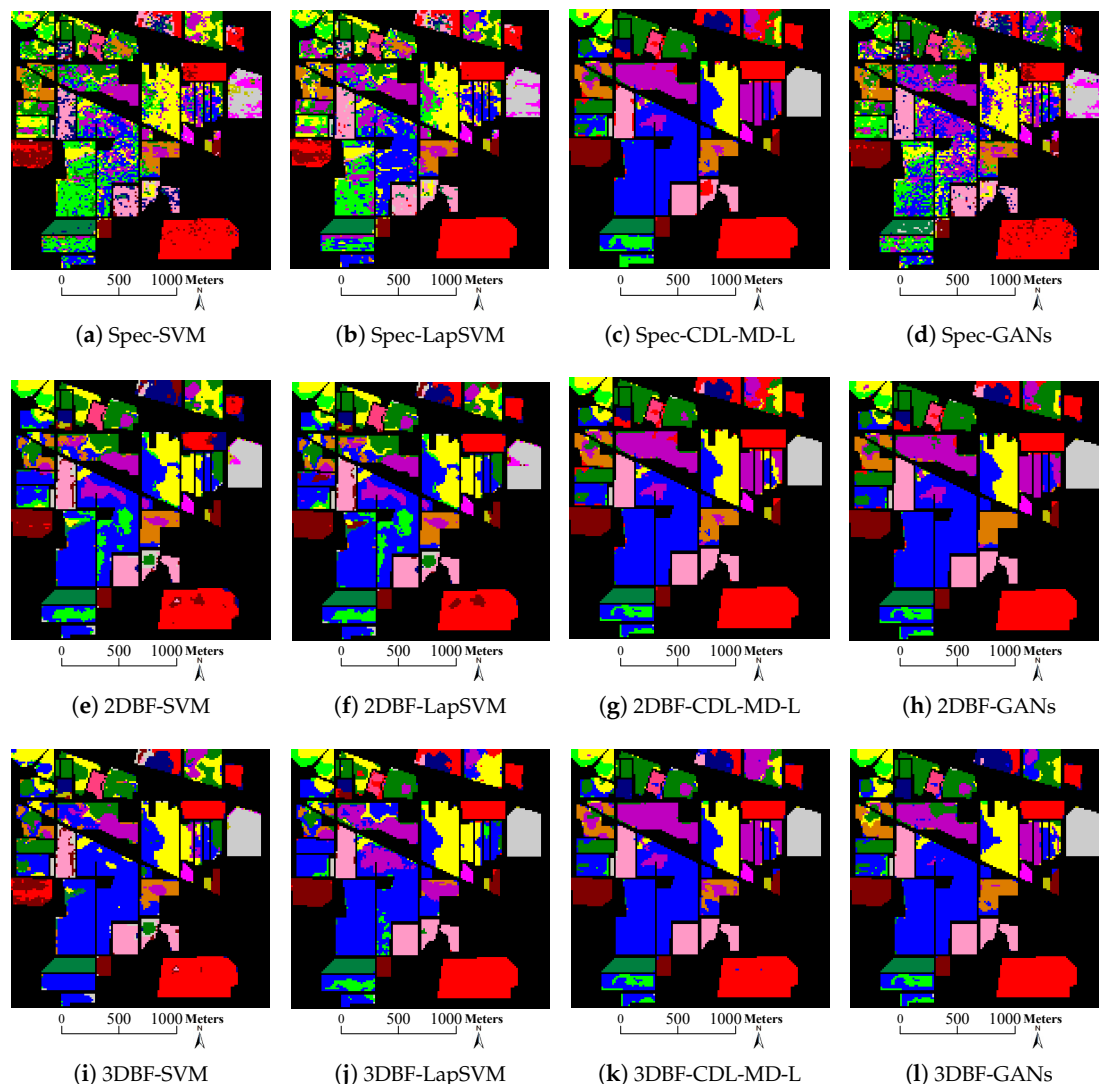

Figure 11. Classification maps of the Indian Pines data with 5 samples per class.
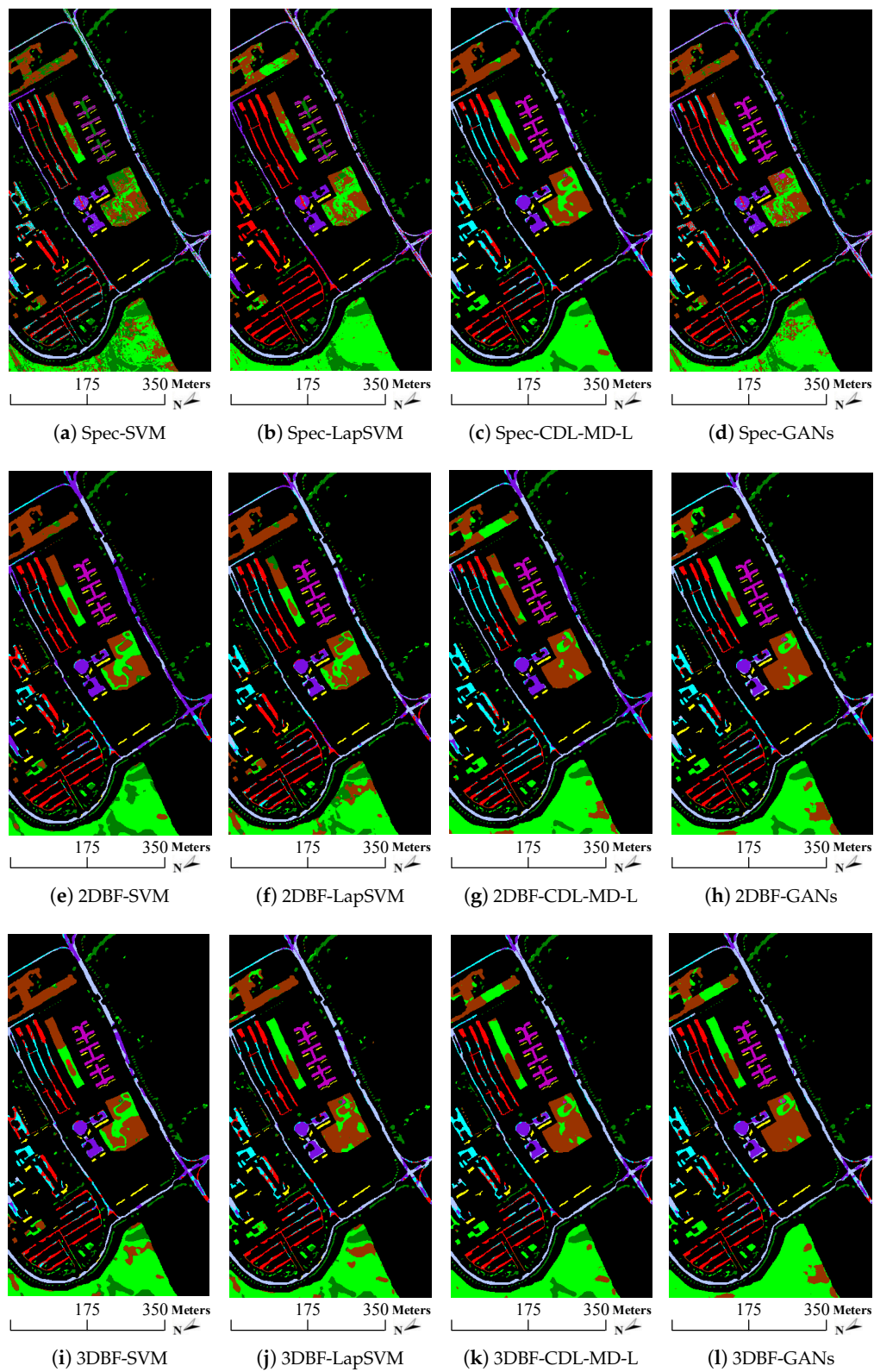
Figure 12. Classification maps of the University of Pavia data with 5 samples per class.
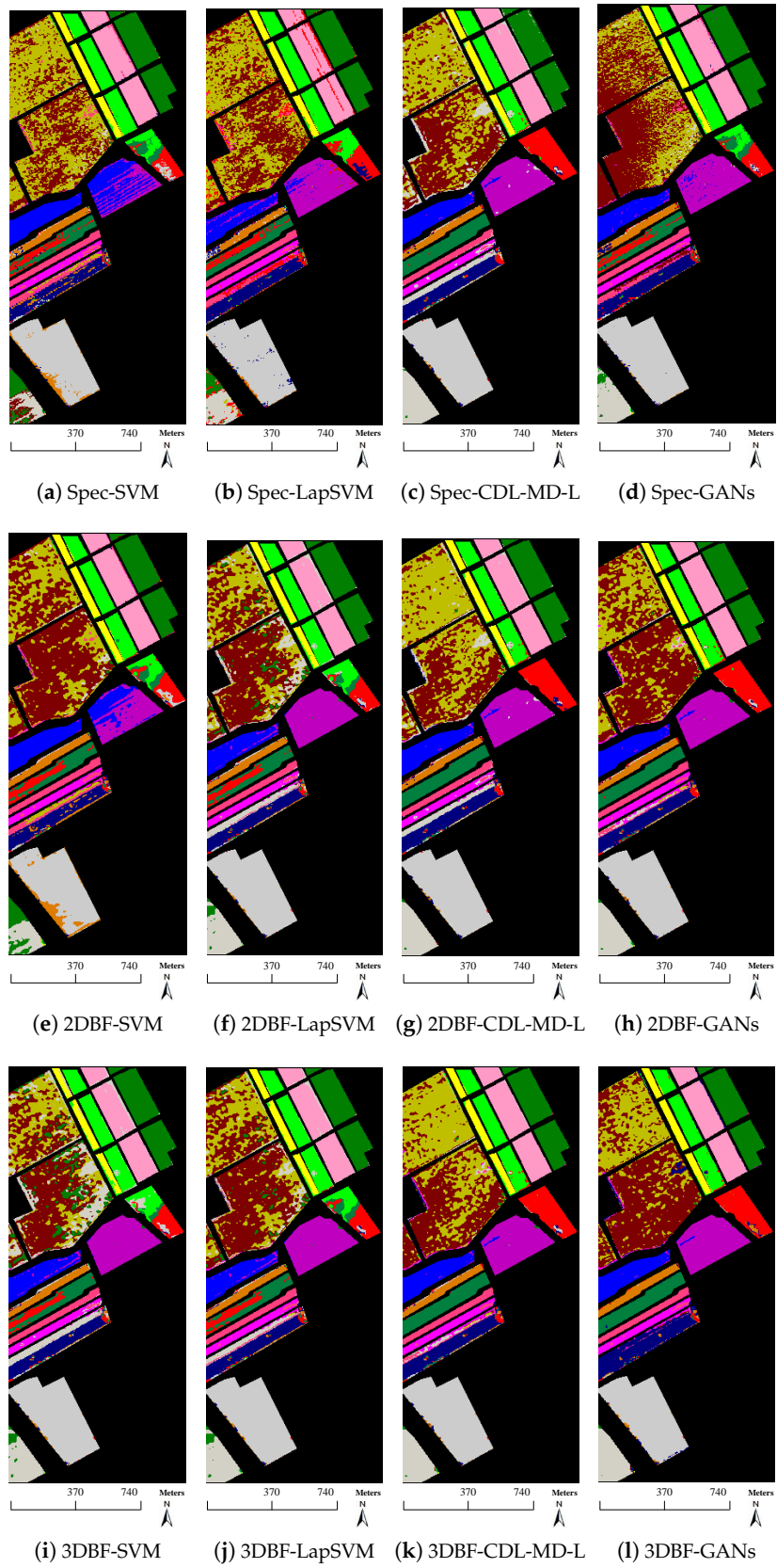
Figure 13. Classification maps of the Salinas data with 5 samples per class.

Table 4. Classification accuracy (%) of various methods for the Indian Pines data with 5 labeled training samples per class, bold values indicate the best result for a row.

Class	Spec				2DBF				3DBF			
	SVM	LapSVM	CDL-MD-L	GANs	SVM	LapSVM	CDL-MD-L	GANs	SVM	LapSVM	CDL-MD-L	GANs
1 [a]	40.17	46.15	**96.58**	48.42	77.65	84.48	96.47	95.46	95.27	94.27	96.51	96.08
2	33.07	32.98	65.31	48.02	39.67	39.93	64.93	63.71	46.01	45.05	63.70	**66.31**
3	40.06	38.02	51.35	44.96	30.16	29.03	52.21	**53.14**	39.72	36.61	49.89	49.95
4	25.93	27.73	50.79	37.30	26.89	22.67	50.25	48.09	45.35	39.42	46.41	**53.98**
5	61.38	55.02	85.21	74.23	69.10	74.06	89.35	88.73	74.51	77.23	**92.42**	91.75
6	68.90	73.71	97.18	80.31	92.52	94.23	97.59	97.62	97.04	97.90	97.96	**98.19**
7	38.11	56.80	64.58	60.39	28.52	29.10	61.23	61.04	56.22	53.82	**81.33**	71.89
8	77.12	87.39	99.84	89.10	96.84	98.83	**99.86**	99.81	99.81	99.83	99.83	99.79
9	24.62	20.01	81.08	39.60	25.03	26.61	**81.20**	79.68	73.88	73.93	80.34	79.03
10	46.77	43.29	57.60	57.66	34.88	34.35	57.81	**60.08**	45.63	46.59	56.66	58.37
11	48.73	50.98	66.44	55.36	41.81	53.06	66.17	67.86	52.04	60.77	69.22	**71.61**
12	26.28	28.31	61.31	36.92	37.96	38.06	60.56	57.50	46.19	51.53	**64.85**	64.34
13	85.68	83.23	99.23	88.16	96.76	96.99	99.15	98.90	98.91	99.27	**99.60**	99.41
14	76.75	78.87	89.22	82.70	85.81	88.12	91.57	90.95	85.47	88.30	93.18	**95.11**
15	30.73	19.03	81.53	38.46	60.38	62.74	83.55	82.73	68.39	69.76	82.06	**85.36**
16	73.24	75.93	83.13	87.75	95.58	**97.26**	82.94	83.16	66.14	69.46	87.97	84.49
OA	49.60	51.75	72.88	59.09	53.88	56.51	73.28	73.53	62.56	65.33	74.12	**75.62**
AA	60.93	59.62	79.29	70.12	68.85	70.04	79.99	79.48	70.64	70.92	80.47	**81.05**
κ	43.84	45.70	69.18	54.36	48.75	51.44	69.71	69.89	57.42	60.09	70.59	**72.23**
F-Measure	49.85	51.09	76.90	60.58	58.72	60.60	77.18	76.78	68.16	68.98	78.87	**79.10**

[a] Lines 3 to 18 are the F-Measure per class.

Table 5. Classification accuracy (%) of various methods for the University of Pavia data with 5 labeled training samples per class, bold values indicate the best result for a row.

Class	Spec				2DBF				3DBF			
	SVM	LapSVM	CDL-MD-L	GANs	SVM	LapSVM	CDL-MD-L	GANs	SVM	LapSVM	CDL-MD-L	GANs
1 [a]	71.43	77.56	79.24	73.88	71.35	72.65	79.44	79.00	61.97	76.91	80.30	**81.21**
2	58.68	59.85	78.14	69.26	66.55	66.76	78.96	79.95	59.44	75.99	81.83	**84.45**
3	39.99	20.16	52.27	42.49	36.52	48.97	51.24	53.92	46.36	49.88	**60.12**	58.56
4	48.83	62.14	65.65	67.27	60.17	57.39	68.42	71.53	79.67	70.29	79.52	**84.57**
5	53.20	89.41	96.23	93.09	94.85	82.61	96.11	96.45	95.76	96.64	97.10	**97.29**
6	32.20	36.39	56.88	38.01	44.92	47.43	56.64	58.43	60.71	52.91	60.21	**62.60**
7	63.78	51.27	52.65	54.75	46.10	47.63	53.33	53.44	**76.52**	51.07	56.45	59.25
8	57.80	64.92	67.92	64.05	60.78	63.34	68.28	68.38	60.60	66.01	**71.60**	71.54
9	95.61	**99.92**	95.91	99.90	94.55	93.84	95.44	95.63	96.98	94.52	96.31	96.28
OA	53.62	60.17	71.83	63.66	61.80	62.62	72.32	73.29	63.39	69.87	75.78	**77.94**
AA	63.76	69.53	76.08	72.85	70.21	70.66	76.37	77.33	70.89	75.43	80.43	**81.36**
κ	44.18	51.33	64.24	54.62	52.81	53.81	64.85	66.04	54.47	62.07	69.26	**71.82**
F-Measure	57.95	62.40	71.65	66.97	63.98	64.51	71.98	72.97	64.82	70.47	75.94	**77.30**

[a] Lines 3 to 11 are the F-Measure per class.

Table 6. Classification accuracy (%) of various methods for the Salinas data with 5 labeled training samples per class, bold values indicate the best result for a row.

Class	Spec				2DBF				3DBF			
	SVM	LapSVM	CDL-MD-L	GANs	SVM	LapSVM	CDL-MD-L	GANs	SVM	LapSVM	CDL-MD-L	GANs
1 [a]	82.66	87.42	97.35	92.06	82.37	93.65	97.81	97.89	94.24	93.82	97.75	**98.18**
2	86.98	92.44	98.01	92.98	88.92	94.71	98.25	98.30	95.93	95.70	98.37	**98.63**
3	46.08	32.97	89.44	66.85	46.42	65.43	89.60	90.14	67.57	68.24	90.88	**92.86**
4	**96.56**	96.00	95.04	96.47	88.05	93.57	95.95	96.05	95.04	95.10	96.18	94.63
5	87.68	84.84	93.79	90.16	81.58	87.39	94.35	94.48	88.76	89.75	**95.52**	94.10
6	98.88	94.60	98.08	98.72	98.84	97.25	98.19	98.91	96.97	97.04	98.94	**99.64**
7	91.97	93.90	97.20	93.71	93.05	93.39	97.19	97.17	93.43	93.46	97.22	**98.18**
8	56.12	59.05	66.00	58.39	68.87	58.12	69.10	68.05	59.30	61.64	73.27	**76.40**
9	93.92	96.45	98.04	96.49	97.91	97.02	98.20	98.20	97.46	97.49	98.33	**98.45**
10	52.58	69.94	76.58	71.22	43.17	65.56	77.37	79.16	65.01	64.47	81.58	**83.23**
11	61.79	71.97	83.86	73.16	55.93	74.06	84.04	86.07	74.48	74.72	84.05	**88.22**
12	72.09	75.79	97.52	83.88	72.09	82.98	97.61	97.70	83.76	82.06	96.02	**98.18**
13	72.10	77.04	84.87	78.53	75.46	78.43	88.99	89.04	79.28	79.97	85.15	**89.56**
14	79.22	81.93	81.64	79.74	74.10	80.11	82.54	81.24	81.75	80.82	82.18	**85.53**
15	55.90	56.25	56.73	51.64	59.81	50.32	55.05	61.31	55.52	54.94	56.34	**67.11**
16	62.32	76.79	88.19	73.46	78.34	71.90	87.53	91.80	69.92	71.45	92.57	**94.58**
OA	73.22	74.23	82.72	77.17	75.15	76.47	83.53	84.38	77.78	78.12	85.11	**87.63**
AA	77.92	78.89	89.48	83.18	77.45	82.75	89.88	90.71	83.57	83.90	90.61	**92.30**
κ	70.40	71.47	80.84	74.70	72.44	73.93	81.71	82.69	75.39	75.77	83.44	**86.26**
F-Measure	74.80	77.96	87.65	81.09	75.31	80.24	88.24	89.09	81.15	81.29	89.02	**91.09**

[a] Lines 3 to 18 are the F-Measure per class.

4. Discussions

4.1. Statistical Significance Analysis of the Results

The statistical significance of the classification differences between various methods is assessed by the McNemar's test, which is based upon the standardized normal test statistic

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \qquad (21)$$

where f_{ij} refers to the number of samples classified correctly by the classifier i but incorrectly by classifier j and Z indicates the pairwise statistical significance of the classification difference between the ith and jth classifiers. In case the test statistic $|Z| > 1.96$, the difference of classification accuracies between the ith and jth classifiers is regarded as statistical significant at the 5% level of significance. For comparison purpose, the results of the McNemar's test on the 3DBF-GANs and other methods are listed in Table 7, which shows that the proposed 3DBF-GANs is superior ($Z > 1.96$) to Spec-SVM, Spec-LapSVM, Spec-GANs, Spec-CDL-MD-L, 2DBF-SVM, 2DBF-LapSVM, 2DBF-CDL-MD-L, 2DBF-GANs, 3DBF-SVM, 3DBF-LapSVM, or comparable ($|Z| < 1.96$ in the Indian Pines data) with 3DBF-CDL-MD-L. According to the McNemar's test, both the 3DBF and GANs are helpful for improving the classification performance since the test statistic is statistical significant, which further confirms the effectiveness of the proposed method.

Table 7. McNemar's test between 3DBF-GANs and other classifiers.

Methods	Z (Indian Pines Data)	Z (University of Pavia Data)	Z (Salinas Data)
3DBF-GANs vs. Spec-SVM	28.04	42.21	40.21
3DBF-GANs vs. Spec-LapSVM	26.93	41.45	38.54
3DBF-GANs vs. Spec-CDL-MD-L	5.91	18.52	17.24
3DBF-GANs vs. Spec-GANs	18.62	30.72	18.38
3DBF-GANs vs. 2DBF-SVM	25.01	41.32	29.31
3DBF-GANs vs. 2DBF-LapSVM	19.63	39.79	24.52
3DBF-GANs vs. 2DBF-CDL-MD-L	4.72	17.92	20.11
3DBF-GANs vs. 2DBF-GANs	4.45	17.63	19.54
3DBF-GANs vs. 3DBF-SVM	16.71	38.22	18.91
3DBF-GANs vs. 3DBF-LapSVM	15.57	25.07	18.57
3DBF-GANs vs. 3DBF-CDL-MD-L	1.64	10.37	16.81

4.2. Sensitivity Analysis of the Parameters

There are four important parameters in the proposed 3DBF-GANs method: the filtering size σ_s, the blur degree σ_r, the training epoch and the learning rate. The influence of these parameters on the classification performance (e.g., OA) is analyzed in Figures 14 and 15. In Figure 14, the effect of σ_s and σ_r is plotted with the training epoch is fixed to 100. It can be seen from Figure 14 that, if the filtering size σ_s and the blur degree σ_r are too small or too large, the OA of 3DBF-GANs is not satisfactory. This is due to the fact that very little spatial information is considered in case σ_s and σ_r are too small, while too large σ_s and σ_r will cause oversmooth. Furthermore, the influence of the training epoch is depicted in Figure 15, from which one can observe that, the OA rapidly increases at first, then slowly increases and finally trends to a certain stable value with the increasing of training epoch. The influence of the learning rate is shown in Figure 16, from which one can find that the OA with a large learning rate (e.g., 0.1) is much lower than that with a smaller learning rate. The reason is that too large learning rate can cause the loss function to fluctuate around the minimum, or even worse, to diverge. Note that too small (e.g., 0.00001) will lead to slow convergence, it is better to set the learning rate to 0.001 or 0.0001. In analogy to the other comparison methods, the appropriate parameters are of importance to the classification performance our proposed 3DBF-GANs method. It is highlighted from the above

analysis that we are able to gain satisfying classification results for different hyperspectral datasets with the provided parameter settings.
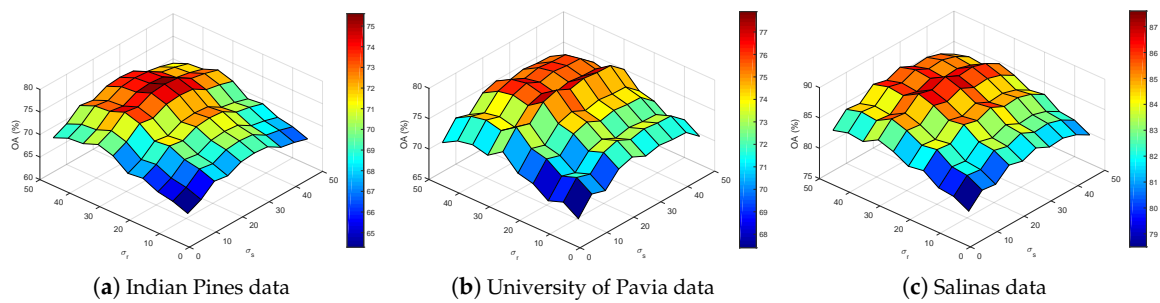

Figure 14. The impact of parameters σ_s and σ_r on the OA in (**a**) Indian Pines data, (**b**) University of Pavia data and (**c**) Salinas data.
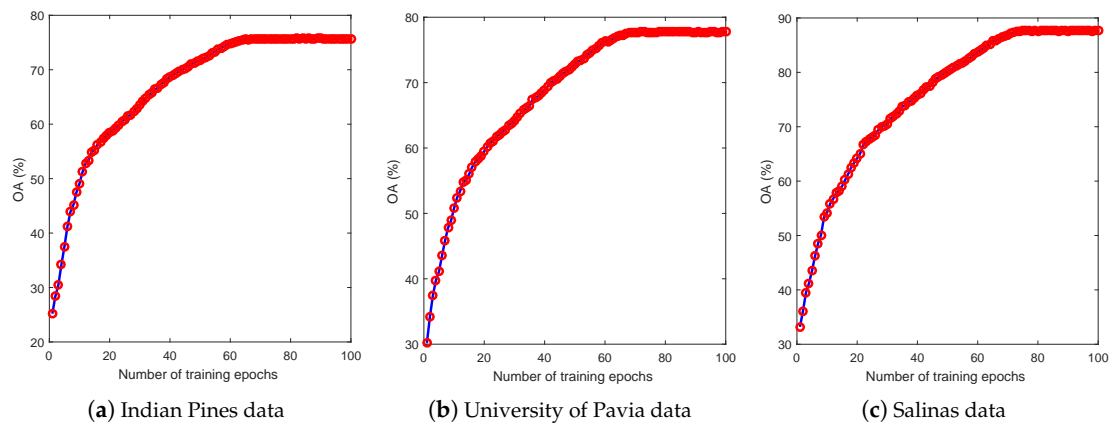

Figure 15. The impact of training epoch on the OA in (**a**) Indian Pines data, (**b**) University of Pavia data and (**c**) Salinas data.
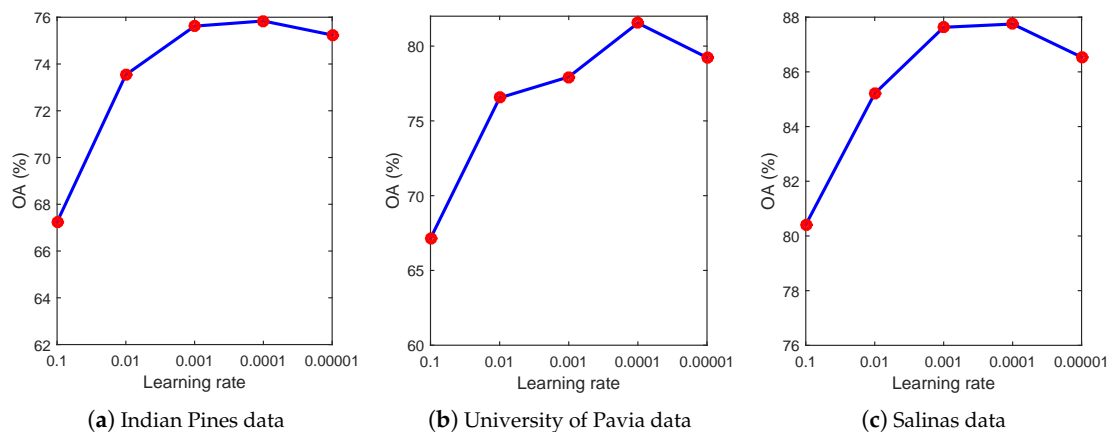

Figure 16. The impact of learning rate on the OA in (**a**) Indian Pines data, (**b**) University of Pavia data and (**c**) Salinas data.

Moreover, the impact of the number of labeled training samples is also evaluated in this section. We randomly choose 5, 10, 15 and 20 samples from each class as the labeled training samples and the OA of various methods is plotted in Figure 17, which shows that the classification accuracy increases as the number of labeled training samples goes up and the 3DBF-GANs method is superior to other

methods when the same number of labeled training samples is chosen. Although the performance of different methods changes as the number of training samples changes, the 3DBF-GANs provides higher classification accuracies than other methods. In addition, it should be pointed out here that the number of synthetic examples generated by the 3DBF-GANs method equals to the total number of training samples (including labeled samples and unlabeled samples), and therefore, is not varied with the number of labeled samples.
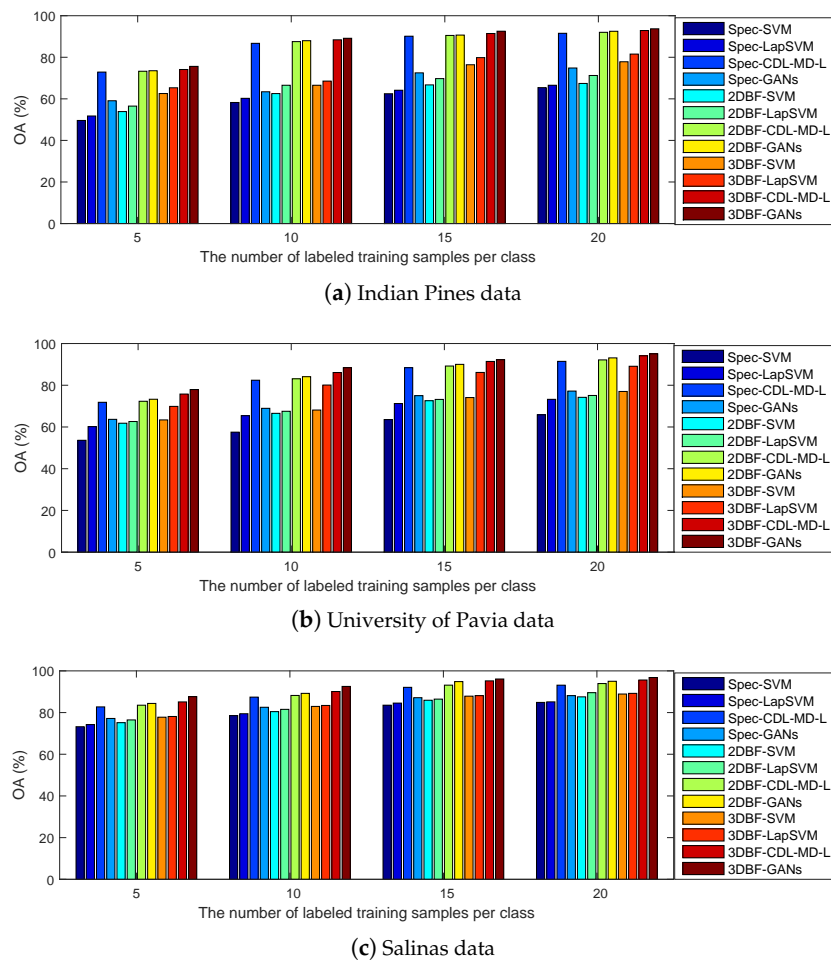

(**a**) Indian Pines data

(**b**) University of Pavia data

(**c**) Salinas data

Figure 17. The impact of the number of labeled training samples per class on the OA in (**a**) Indian Pines data, (**b**) University of Pavia data and (**c**) Salinas data.

5. Conclusions

In this paper, we have proposed a semi-supervised learning method based on 3DBF and GANs for hyperspectral classification. The proposed 3DBF-GANs method relies on two aspects. The first is the extraction of spectral-spatial features by the 3DBF. The main advantage of the 3DBF is that it is able to smooth the hyperspectral cube while preserving the edges by naturally treating the HSI as a volumetric dataset. The second one is the semi-supervised classification by the GANs. The GANs can make full use of both the limited labeled samples and the abundance of unlabeled samples for significant semi-supervised learning. Compared to the shallow learning and non-adversarial networks, the GANs are one of the effective deep learning methods which can take advantage of the discriminative models to train the generative network in an adversarial fashion. The proposed method has been tested on AVIRIS and ROSIS datasets with very limited labeled samples, and the comparison with other state-of-the-art methods (i.e., 2DBF, SVM, LapSVM, and CDL-MD-L) has confirmed the effectiveness of the proposed method. Quantitatively, the OA of 3DBF-GANs improves about 1% to 25% compared

to other state-of-the-art methods. Note that the GANs have a complex structure, a future research topic is to investigate how to determine the network parameters in a more effective and automatic way. Introducing the graph-based semi-supervised learning method [26] from other areas to hyperspectral classification is also a probable future research direction.

Author Contributions: All coauthors made significant contributions to the manuscript. Zhi He and Han Liu designed the research framework, analyzed the results and wrote the manuscript. Jie Hu and Yiwen Wang assisted in the prepared work and validation work. Moreover, all coauthors contributed to the editing and review of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sun, W.; Jiang, M.; Li, W.; Liu, Y. A symmetric sparse representation based band selection method for hyperspectral imagery classification. *Remote Sens.* **2016**, *8*, 238, doi:10.3390/rs8030238.
2. Sun, W.; Zhang, D.; Xu, Y.; Tian, L.; Yang, G.; Li, W. A probabilistic weighted archetypal analysis method with earth mover's distance for endmember extraction from hyperspectral imagery. *Remote Sens.* **2017**, *9*, 841, doi:10.3390/rs9080841.
3. Pan, L.; Li, H.C.; Deng, Y.J.; Zhang, F.; Chen, X.D.; Du, Q. Hyperspectral dimensionality reduction by tensor sparse and low-rank graph-based discriminant analysis. *Remote Sens.* **2017**, *9*, 452, doi:10.3390/rs9050452.
4. Feng, F.; Li, W.; Du, Q.; Zhang, B. Dimensionality reduction of hyperspectral image with graph-based discriminant analysis considering spectral similarity. *Remote Sens.* **2017**, *9*, 323.
5. Gao, L.; Zhao, B.; Jia, X.; Liao, W.; Zhang, B. Optimized kernel minimum noise fraction transformation for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 548.
6. Sun, B.; Kang, X.; Li, S.; Benediktsson, J.A. Random-walker-based collaborative learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 212–222.
7. Yang, L.; Wang, M.; Yang, S.; Zhang, R.; Zhang, P. Sparse spatio-spectral LapSVM with semisupervised kernel propagation for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2046–2054.
8. Zhong, Y.; Ma, A.; Zhang, L. An adaptive memetic Fuzzy clustering algorithm with spatial information for remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1235–1248.
9. Niazmardi, S.; Homayouni, S.; Safari, A. An improved FCM algorithm based on the SVDD for unsupervised hyperspectral data classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 831–839.
10. Zhong, Y.; Zhang, L.; Huang, B.; Li, P. An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 420–431.
11. Zhu, W.; Chayes, V.; Tiard, A.; Sanchez, S.; Dahlberg, D.; Bertozzi, A.L.; Osher, S.; Zosso, D.; Kuang, D. Unsupervised classification in hyperspectral imagery with nonlocal total variation and primal-dual hybrid gradient algorithm. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2786–2798.
12. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: New York, NY, USA, 2013.
13. Kuo, B.C.; Ho, H.H.; Li, C.H.; Hung, C.C.; Taur, J.S. A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 317–326.
14. Adep, R.N.; Shetty, A.; Ramesh, H. EXhype: A tool for mineral classification using hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 106–118.
15. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral image classification via kernel sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 217–231.
16. Zhang, Y.; Du, B.; Zhang, L.; Liu, T. Joint sparse representation and multitask learning for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 894–906.

17. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. *IEEE Geosci. Remote. Sens. Lett.* **2013**, *10*, 318–322.

18. Chapel, L.; Burger, T.; Courty, N.; Lefevre, S. PerTurbo manifold learning algorithm for weakly labeled hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1070–1078.

19. Joachims, T. Transductive inference for text classification using support vector machines. In Proceedings of the Sixteenth International Conference on Machine Learning, San Francisco, CA, USA, 27–30 June 1999; pp. 200–209.

20. Maulik, U.; Chakraborty, D. Learning with transductive SVM for semisupervised pixel classification of remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2013**, *77*, 66–78.

21. Wang, L.; Hao, S.; Wang, Q.; Wang, Y. Semi-supervised classification for hyperspectral imagery based on spatial-spectral label propagation. *ISPRS J. Photogramm. Remote Sens.* **2014**, *97*, 123–137.

22. Belkin, M.; Niyogi, P.; Sindhwani, V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* **2006**, *7*, 2399–2434.

23. Camps-Valls, G.; Marsheva, T.V.B.; Zhou, D. Semi-supervised graph-based hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3044–3054.

24. Melacci, S.; Belkin, M. Laplacian support vector machines trained in the primal. *J. Mach. Learn. Res.* **2011**, *12*, 1149–1184.

25. De Morsier, F.; Borgeaud, M.; Gass, V.; Thiran, J.P.; Tuia, D. Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3410–3420.

26. Yamaguchi, Y.; Faloutsos, C.; Kitagawa, H. Camlp: Confidence-aware modulated label propagation. In Proceedings of the 2016 SIAM International Conference on Data Mining, SIAM, Miami, FlL, USA, 5–7 May 2016; pp. 513–521.

27. Dopido, I.; Li, J.; Marpu, P.R.; Plaza, A.; Dias, J.M.B.; Benediktsson, J.A. Semisupervised self-learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4032–4044.

28. Aydemir, M.S.; Bilgin, G. Semisupervised hyperspectral image classification using small sample sizes. *IEEE Geosci. Remote. Sens. Lett.* **2017**, *14*, 621–625.

29. Zhang, X.; Song, Q.; Liu, R.; Wang, W.; Jiao, L. Modified co-training with spectral and spatial views for semisupervised hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2044–2055.

30. Romaszewski, M.; Głomb, P.; Cholewa, M. Semi-supervised hyperspectral classification from a small number of training samples using a co-training approach. *ISPRS J. Photogramm. Remote Sens.* **2016**, *121*, 60–76.

31. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE* **2013**, *101*, 652–675.

32. Cavallaro, G.; Mura, M.D.; Benediktsson, J.A.; Bruzzone, L. Extended self-dual attribute profiles for the classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1690–1694.

33. Bao, R.; Xia, J.; Mura, M.D.; Du, P.; Chanussot, J.; Ren, J. Combining morphological attribute profiles via an ensemble method for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 359–363.

34. Jia, S.; Shen, L.; Li, Q. Gabor feature-based collaborative representation for hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1118–1129.

35. He, L.; Li, J.; Plaza, A.; Li, Y. Discriminative low-rank Gabor filtering for spectral–spatial hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1381–1395.

36. Kang, X.; Li, S.; Benediktsson, J.A. Spectral-spatial hyperspectral image classification with edge-preserving filtering. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2666–2677.

37. Demir, B.; Erturk, S. Empirical mode decomposition of hyperspectral images for support vector machine classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4071–4084.

38. He, Z.; Wang, Q.; Shen, Y.; Jin, J.; Wang, Y. Multivariate gray model-based BEMD for hyperspectral image classification. *IEEE Trans. Instrum. Meas.* **2013**, *62*, 889–904.

39. Zabalza, J.; Ren, J.; Zheng, J.; Han, J.; Zhao, H.; Li, S.; Marshall, S. Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4418–4433.

40. Tarabalka, Y.; Chanussot, J.; Benediktsson, J. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognit.* **2010**, *43*, 2367–2379.

41. Li, J.; Zhang, H.; Zhang, L. Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5338–5351.

42. He, Z.; Liu, L.; Zhou, S.; Shen, Y. Learning group-based sparse and low-rank representation for hyperspectral image classification. *Pattern Recognit.* **2016**, *60*, 1041–1056.

43. Camps-Valls, G.; Gomez-Chova, L.; Muñoz-Marí, J.; Vila-Francés, J.; Calpe-Maravilla, J. Composite kernels for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 93–97.

44. Gu, Y.; Liu, T.; Jia, X.; Benediktsson, J.A.; Chanussot, J. Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3235–3247.

45. Niazmardi, S.; Safari, A.; Homayouni, S. A novel multiple kernel learning framework for multiple feature classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3734–3743.

46. Sun, L.; Wu, Z.; Liu, J.; Xiao, L.; Wei, Z. Supervised spectral-spatial hyperspectral image classification with weighted markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1490–1503.

47. Bai, J.; Xiang, S.; Pan, C. A graph-based classification method for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 803–817.

48. Sun, X.; Qu, Q.; Nasrabadi, N.M.; Tran, T.D. Structured Priors for Sparse-Representation-Based Hyperspectral Image Classification. *IEEE Geosci. Remote. Sens. Lett.* **2014**, *11*, 1235–1239.

49. Xu, Y.; Fang, F.; Zhang, G. Similarity-guided and l_p-regularized sparse unmixing of hyperspectral data. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2311–2315.

50. Liu, C.; Zhou, J.; Liang, J.; Qian, Y.; Li, H.; Gao, Y. Exploring structural consistency in graph regularized joint spectral-spatial sparse coding for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1151–1164.

51. Soltani-Farani, A.; Rabiee, H.R.; Hosseini, S.A. Spatial-aware dictionary learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 527–541.

52. Sumarsono, A.; Du, Q. Low-rank subspace representation for estimating the number of signal subspaces in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6286–6292.

53. Sun, W.; Yang, G.; Du, B.; Zhang, L.; Zhang, L. A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4032–4046.

54. Qian, Y.; Ye, M.; Zhou, J. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2276–2291.

55. Tsai, F.; Lai, J.S. Feature extraction of hyperspectral image cubes using three-dimensional gray-level cooccurrence. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3504–3513.

56. Zhang, L.; Zhang, L.; Tao, D.; Huang, X. Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 242–256.

57. He, Z.; Liu, L. Robust multitask learning with three-dimensional empirical mode decomposition-based features for hyperspectral classification. *ISPRS J. Photogramm. Remote Sens.* **2016**, *121*, 11–27.

58. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40.

59. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107.

60. Chen, Y.; Zhao, X.; Jia, X. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2381–2392.

61. Li, Y.; Zhang, H.; Shen, Q. Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sens.* **2017**, *9*, 67, doi:10.3390/rs9010067.

62. Ma, X.; Wang, H.; Wang, J. Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning. *ISPRS J. Photogramm. Remote Sens.* **2016**, *120*, 99–107.

63. Smith, S.M.; Brady, J.M. SUSAN—A new approach to low level image processing. *Int. J. Comput. Vis.* **1997**, *23*, 45–78.

64. Paris, S.; Durand, F. A fast approximation of the bilateral filter using a signal processing approach. In Proceedings of the 9th European Conference on Computer Vision—ECCV, Graz, Austria, 7–13 May 2006; pp. 568–580.

65.　Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.

66.　Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training GANs. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 2226–2234.

67.　Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), Bombay, India, 7 January 1998; Narosa Publishing House: Delhi, India, 1998; pp. 839–846.

68.　Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.

69.　Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.

70.　Denton, E.L.; Chintala, S.; Szlam, A.; Fergus, R. Deep generative image models using a Laplacian pyramid of adversarial networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 7–12 December 2015; pp. 1486–1494.

71.　Chen, X.; Duan, Y.; Houthooft, R.; Schulman, J.; Sutskever, I.; Abbeel, P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Barcelona, Spain, 5–10 December 2016; pp. 2172–2180.

72.　Metz, L.; Poole, B.; Pfau, D.; Sohl-Dickstein, J. Unrolled generative adversarial networks. *arXiv* **2016**, arXiv:1611.02163.

73.　Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein gan. *arXiv* **2017**, arXiv:1701.07875.

74.　Wang, X.; Gupta, A. Generative image modeling using style and structure adversarial networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 318–335.

75.　Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv* **2016**, arXiv:1609.04802.

76.　Yeh, R.; Chen, C.; Lim, T.Y.; Hasegawa-Johnson, M.; Do, M.N. Semantic image inpainting with perceptual and contextual losses. *arXiv* **2016**, arXiv:1607.07539.

77.　Springenberg, J.T. Unsupervised and semi-supervised learning with categorical generative adversarial networks. *arXiv* **2015**, arXiv:1511.06390.

78.　Premachandran, V.; Yuille, A.L. Unsupervised learning using generative adversarial training and clustering. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.

79.　Sutskever, I.; Jozefowicz, R.; Gregor, K.; Rezende, D.; Lillicrap, T.; Vinyals, O. Towards principled unsupervised learning. *arXiv* **2015**, arXiv:1511.06440.