Article

Hyperspectral Image Classification Based on Mutually Guided Image Filtering

Ying Zhan [1,*] , **Dan Hu** [2] , **Xianchuan Yu** [3] **and Yufeng Wang** [1]

1 School of Computer and Software, Nanyang Institute of Technology, Nanyang 473004, China; wangyufeng@whu.edu.cn
2 Department of Radiology and BRIC, University of North Carolina, Chapel Hill, NC 27514, USA; danhu0055@gmail.com
3 School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China; yuxianchuan@163.com
* Correspondence: zhanying@live.com

Abstract: Hyperspectral remote sensing images (HSIs) have both spectral and spatial characteristics. The adept exploitation of these attributes is central to enhancing the classification accuracy of HSIs. In order to effectively utilize spatial and spectral features to classify HSIs, this paper proposes a method for the spatial feature extraction of HSIs based on a mutually guided image filter (muGIF) and combined with the band-distance-grouped principal component. Firstly, aiming at the problem that previously guided image filtering cannot effectively deal with the inconsistent information structure between the guided and target information, a method for extracting spatial features using muGIF is proposed. Then, aiming at the problem of the information loss caused by a single principal component as a guided image in the traditional GIF-based spatial–spectral classification, a spatial feature-extraction framework based on the band-distance-grouped principal component is proposed. The method groups the bands according to the band distance and extracts the principal components of each set of band subsets as the guide map of the current band subset to filter the HSIs. A deep convolutional neural network model and a generative adversarial network model for the filtered HSIs are constructed and then trained using samples for HSIs' spatial–spectral classification. Experiments show that compared with the traditional methods and several popular spatial–spectral HSI classification methods based on a filter, the proposed methods based on muGIF can effectively extract the spatial–spectral features and improve the classification accuracy of HSIs.

Keywords: hyperspectral images classification; mutually guided image filter; band distance; deep learning

1. Introduction

Hyperspectral remote sensing image (HSI) classification is the process of attempting to allocate class labels to pixels within a set of hyperspectral images utilizing classification methods. It represents a primary application of hyperspectral remote sensing data and is a crucial technique for interpreting and analyzing HSIs [1].

Through years of research and exploration, numerous scholars have proposed an exceedingly diverse array of classification methods. Depending on the availability of class knowledge, the classification methods are classified as unsupervised, supervised, or semi-supervised learning approaches [2]. Unsupervised methods rely solely on the intrinsic separability of the data and include clustering algorithms such as K-Means [3], spectral clustering [4], Fuzzy C-Means (FCM) [5], hierarchical clustering [6], and density-based spatial clustering (DBSCAN) [7].

Supervised methods, using labeled samples for training, generally achieve better performance. Common supervised approaches include probability-based methods, kernel transformations, random forests, and sparse representation [1].

Semi-supervised learning methods can be categorized into generative and discriminative models. Generative models establish joint probability distribution between classes,

reflecting similarity within the same class. Examples include a Markov random field (MRF) model incorporating spatial and spectral features [8] and the S^2MLR (Soft sparse multinomial logistic regression) model [9]. Discriminative models directly learn conditional probability distributions, such as a graph-based framework describing the importance of labeled samples [10] and a transductive SVM method [11] implementing a weighting strategy for unlabeled samples. These methods enhance classification by leveraging both labeled and unlabeled data for improved performance.

In recent years, deep convolutional neural networks (CNNs) have achieved tremendous success in various tasks on natural images and have also quickly garnered significant attention from the hyperspectral remote sensing image-processing community [12,13]. Another highly significant representational learning model in deep learning is the generative adversarial networks (GANs), first proposed by Goodfellow and colleagues in 2014 [14]. This network combines the characteristics of both generative and discriminative models. Subsequently, a variety of GAN-based models have been continuously introduced [15–17], leading to a series of semi-supervised and unsupervised learning methods based on GANs [18]. For instance, Springenberg proposed Cat-GAN in 2015 [19], which is capable of handling image classification tasks and generating samples. This method offers an approach to learning discriminative classifiers from unlabeled or partially labeled data. In 2016, Odena et al. introduced Semi-GAN [20], which forces the discriminative model to output $k+1$ labels (where k represents the number of classes) to perform semi-supervised classification. He and co-authors presented a semi-supervised method for hyperspectral images based on GANs named 3DBF-GANs [21], which learns and classifies hyperspectral images after extracting spatial–spectral features using three-dimensional bilateral filtering. In 2023, Zhan et al. proposed a hyperspectral image semi-supervised classification method based on a one-dimensional GAN network with spectral angle distance (SADGAN), achieving commendable performance with limited labeled samples [22].

Due to the inherent phenomena of "same material different spectrum" and "different material same spectrum" in hyperspectral images [23], solely relying on spectral information is insufficient for the effective identification of ground objects. Simultaneously, pixels within hyperspectral images tend to belong to the same category as adjacent pixels with a high probability, forming the foundation for classification methods based on spatial features. Many scholars have proposed joint classification studies combining spectral and spatial features on this theoretical basis [24].

According to the literature summary [23], the spatial–spectral classification framework for hyperspectral images typically follows a series of steps. First, spectral and spatial features are extracted from the data. Then, these features are either selected or fused together. The selected or fused features are subsequently inputted into a classifier for the purpose of classification. Finally, spatial regularization is performed using spatial features to enhance the classification results. Depending on the stage at which the fusion of spectral and spatial information occurs, spatial–spectral classification methods can be categorized into three types [25]: preprocessing-based classification [26], integrated classification [27,28], and postprocessing-based classification [29,30].

In the preprocessing-based spatial–spectral classification of hyperspectral images, spatial-feature-extraction methods targeted at neighborhood pixels are commonly employed due to the converging characteristics between pixels and their surrounding neighborhoods [25]. Techniques based on Markov random fields [31], spectral correlation [32], morphological features [33], and superpixel segmentation [34] are prevalent in this context.

The utilization of filtering techniques for spatial feature extraction in preprocessing-based classification is garnering increasing attention from researchers. Commonly used filtering-based methods include the approach proposed by Bau et al. [35], which is based on Gabor filtering. This method extracts features such as orientation, scale, and wavelength of hyperspectral images through three-dimensional Gabor filtering. Another method, proposed by Chen et al. [36], combines Gabor filtering with CNNs techniques. It employs

Gabor filters for edge and texture feature extraction and, when combined with convolutional filters, can mitigate the issue of model overfitting.

At present, researchers predominantly employ filtering-based methods to extract spatial features [35,36], particularly the Guided Image Filter (GIF) proposed by He et al. [37], which has demonstrated promising results in hyperspectral image classification [38–41].

GIF operates based on the Local Linear Model, computing the filter output by considering the content of the guidance image, enabling image smoothing while preserving edges. However, while GIF can guide the filtering output by referencing the information of a guidance image, it overlooks the structural inconsistency between the reference signal and the target signal captured under different conditions, failing to adequately retain image edges in complex image structures. Recently, Guo et al. proposed the mutually guided image filter (muGIF) algorithm [42], which employs the concept of relative structure to measure the similarity between the reference image and the target image. A global optimization objective is designed on this basis to achieve high-quality image filtering.

Inspired by this, we propose a hyperspectral image classification model based on the muGIF algorithm in this paper. Initially, considering the two-dimensional spatial features and one-dimensional spectral features inherent in hyperspectral images, a spatial feature extraction method based on mutually guided image filtering and band-distance-grouped principal component is proposed. This method groups bands according to the band distance in hyperspectral images, extracting the principal components of each group of band subsets as the guidance map for the current band subset to carry out mutually guided image filtering. Subsequently, deep CNN- and GAN-classification models are constructed for the filtered hyperspectral images, selecting training samples for classifier training and performing spatial–spectral classification of the hyperspectral images.

The main contributions of this paper are as follows.

(1) Proposing a hyperspectral image classification model based on the muGIF algorithm. This model takes into account the two-dimensional spatial features and one-dimensional spectral features inherent in hyperspectral images.

(2) Introducing a spatial feature extraction method based on mutually guided image filtering and a band-distance-grouped principal component. This method groups bands according to the band distance in hyperspectral images, extracting the principal components of each group of band subsets as the guidance map for the current band subset to carry out mutually guided image filtering.

The rest of the paper is organized as follows. Section 2 describes the proposed muGIF-based classification method in detail. Section 3 describes the data sets used for the comparison experiments and the specific settings of the comparison methods in the experiments. Section 4 reports the results of the experiments and the related analysis. Conclusions are presented in Section 5.

2. Methodology

In this paper, a hyperspectral image-classification method based on mutually guided image filtering is proposed. On this basis, the filtered images are classified by combining deep learning models such as CNN and GAN.

Figure 1 describes the framework of the spatial–spectral classification method of muGIF and deep learning. Firstly, the original hyperspectral image is grouped based on band distance density, and principal component analysis is performed on each group of images after grouping. Each group extracts a principal component as the guided image of that group of band images; then, the muGIF filtering algorithm is used to filter all bands in each group of images; finally, the filtered data set is trained through CNN models or GAN models to obtain the corresponding classifiers for classification. Section 2.1 explains how muGIF works. Section 2.2 provides an overview of how the proposed methods in this paper work. Section 2.2.1 explains how to perform muGIF filtering after obtaining grouped principal components through band distance density. On this basis, Section 2.2.2 and

Section 2.2.3 present classification methods combining mutual conductive image filtering with CNNs and GANs, respectively.
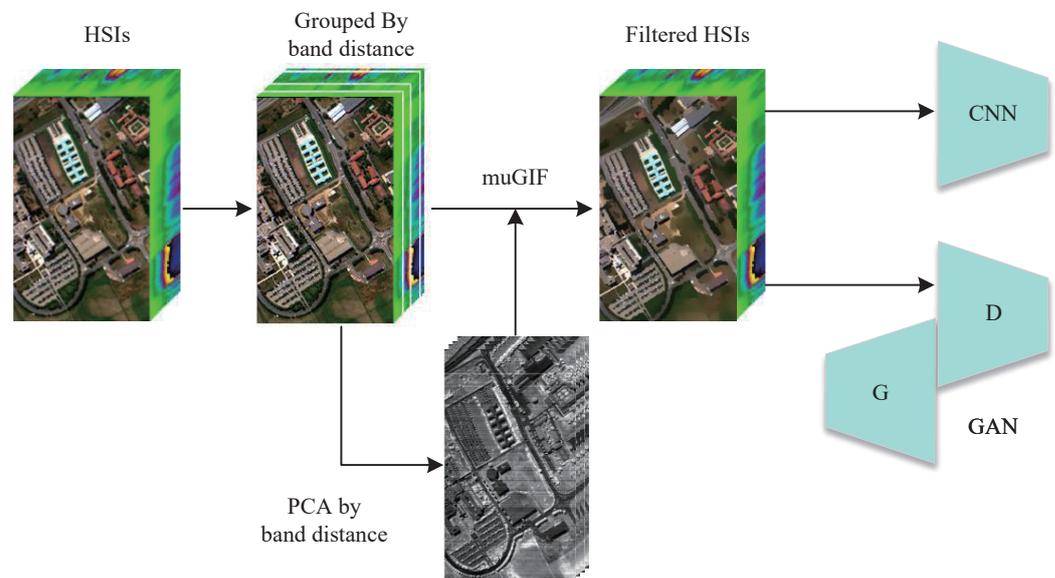

Figure 1. Framework of the spatial–spectral joint classification method based on MuGIF and deep learning.

2.1. Mutually Guided Image Filtering

Image filtering can generally be divided into two categories: frequency domain filtering and spatial filtering. Spatial filtering, in particular, is a technique employed to modify or enhance images according to specific rules. It can typically be formulated as follows:

$$\min_{T} \Psi(T, T_0) + \alpha \Phi(T) \tag{1}$$

T_0 and T denote the input and output signals respectively, $\Psi(T, T_0)$ represents the fidelity term, $\Phi(T)$ signifies the regularization term of the output, and α is a non-negative coefficient balancing these two terms.

To better utilize the information from the guided image and reflect the corresponding structural relationship between the reference image R and target image T, the relative structure of T with respect to R is defined as follows [42]:

$$\Re(T, R) = \sum_{i} \sum_{d \in \{h,v\}} \frac{|\nabla_d T_i|}{|\nabla_d R_i|}, \tag{2}$$

where i denotes a pixel on the image (x, y) and ∇_d represents a first-order derivative filter comprising both the horizontal (h) and vertical (v) directions. The relative structure $\Re(T, R)$ measures the structural difference of T with respect to R.

With the definition of relative structure, the optimization objective for muGIF can be constructed as follows:

$$\arg\min_{T,R} \ \alpha_t \Re(T, R) + \beta_t \|T - T_0\|_2^2 + \alpha_r \Re(R, T)$$
$$+ \beta_r \|R - R_0\|_2^2, \tag{3}$$

α_t, α_r, β_t, and β_r are non-negative constants used to balance the corresponding terms; $\| \cdot \|_2^2$ denotes the ℓ_2 norm, $|T - T_0|_2^2$ and $|R - R_0|_2^2$ are employed to constrain T and R not to deviate too much from T_0 and R_0.

Directly solving the aforementioned optimization problem is quite challenging. Guo et al. proposed an approximate solution method [42]: initially finding an approximate substitute for the relative structure $\Re(T, R)$:

$$\tilde{\Re}(T, R, \epsilon_t, \epsilon_r) = \sum_i \sum_{d \in \{h,v\}} \frac{(\nabla_d T_i)^2}{\max(|\nabla_d R_i|, \epsilon_r) \cdot \max(|\nabla_d T_i|, \epsilon_t)} \tag{4}$$

Herein, ϵ_t and ϵ_r are introduced to avoid division-by-zero errors. The corresponding optimization objective can be replaced with the following form:

$$\arg\min_{T,R} \alpha_t \tilde{\Re}(T, R, \epsilon_t, \epsilon_r) + \beta_t \|t - t_0\|_2^2$$
$$+\alpha_r \tilde{\Re}(T, R, \epsilon_t, \epsilon_r) + \beta_r \|r - r_0\|_2^2, \tag{5}$$

where t, t_0, r and r_0 are the vector forms of T, T_0, R and R_0 respectively. Let Q_d and P_d ($d \in \{h,v\}$) represent the diagonal matrices of the ith diagonal element $\frac{1}{\max(|\nabla_d Ti|, \epsilon t)}$ and $\frac{1}{\max(|\nabla_d Ri|, \epsilon r)}$ respectively. Correspondingly, the objective Function (5) is transformed into

$$\arg\min_{T,R} \alpha_t t^T \left(\sum_{d \in \{h,v\}} D_d^T Q_d P_d D_d \right) + \beta_t \|t - t_0\|_2^2$$
$$+\alpha_r r^T \left(\sum_{d \in \{h,v\}} D_d^T Q_d P_d D_d \right) + \beta_r \|r - r_0\|_2^2 \tag{6}$$

Herein, D_d is the Toeplitz matrix of the discrete gradient operator in the d direction. Equation (6) can be solved through Alternating Least Squares (ALS) to obtain the output after muGIF filtering

$$t = \frac{t_0}{\left(I + \frac{\alpha_t}{\beta_t} \left(\sum_{d \in \{h,v\}} D_d^T Q_d^{(k)} P_d^{(k)} D_d \right) \right)} \tag{7}$$

$$r = \frac{r_0}{\left(I + \frac{\alpha_r}{\beta_r} \left(\sum_{d \in \{h,v\}} D_d^T Q_d^{(k+1)} P_d^{(k)} D_d \right) \right)} \tag{8}$$

2.2. Mutually Guided Image Filtering for Hyperspectral Images Classification

2.2.1. Principal Components of Band Distance Density

Similar to the GIF algorithm, the muGIF algorithm also requires a guidance image as a reference to filter the target image. According to the concept of relative structure described in Equation (2), when the guidance image R is in the edge region, the first-order gradient is larger. Hence, the penalty term $\frac{1}{|\nabla_d R_i|}$ for $|\nabla_d T_i|$ is relatively more minor, achieving the effect of preserving boundaries. Conversely, when the guidance image R is in a smooth region with a smaller first-order gradient, the penalty term is relatively larger, resulting in a more pronounced smoothing effect. Therefore, the choice of the guidance image is of significant importance to the filtering effect of the muGIF algorithm.

In spatial–spectral classification methods of hyperspectral images using guided image filtering, the primary components obtained from PCA decomposition of all band images are usually used as the guidance images for filtering:

$$[c_1, c_2, \ldots, c_n] = PCA(H), \tag{9}$$

where H represents the hyperspectral image data, $PCA(H)$ denotes the PCA decomposition of H using principal component analysis [43,44], and c_i indicates the ith principal component, and n represents the number of bands.

Researchers often choose the first or the first three principal components as the guidance image for filtering [38,40,41].

However, since hyperspectral images often contain multiple bands, and there is a significant difference between different bands, using one or a few principal components simultaneously as the guidance image for multiple bands cannot provide a good filtering reference for different band images, which can easily lead to information distortion. This paper proposes a principal component-extraction method based on band distance density to address this question. By grouping all bands according to band distance density, one principal component is extracted for each group, and this principal component is used as the guidance image for filtering all band images in this group.

Band distance can be used to describe the sparsity level of features between bands:

$$d_i = |r_{i+1} - r_i|, i \in [1, \ldots, n-1], \tag{10}$$

where n is the total number of bands and r_i is the spectral response value of the ith band. The equation represents that d_i is the absolute value of the difference between the ith and the $i+1$th adjacent bands. From the Equation (10), we can derive that the total distance between all adjacent bands in the hyperspectral image is

$$d_{all} = \sum_j \sum_i |r_{j,i+1} - r_{j,i}|, \tag{11}$$

where j represents the pixel in the hyperspectral image. Next, the entire band needs to be divided into p subgroups. Let $pe_0 = 1$, pe_k be the band number of the ending position of the kth group, and let s_k denote the number of bands in each group. Then, we have

$$s_k = pe_k - pe_{k-1}, k \in [1, \ldots, p] \tag{12}$$

and

$$\sum_{k=1}^{p} s_k = n \tag{13}$$

By sequentially solving Equation (14) from top to bottom, we can determine the specific values of $pe_1, \ldots, pe_p$ and thus obtain the specific division of the p subintervals.

$$\sum_j \sum_{i=1}^{pe_1} |r_{j,i+1} - r_{j,i}| = \frac{d_{all}}{p}$$

$$\sum_j \sum_{i=pe_1}^{pe_2} |r_{j,i+1} - r_{j,i}| = \frac{d_{all}}{p} \tag{14}$$

$$\ldots$$

$$\sum_j \sum_{i=pe_{p-1}}^{pe_p} |r_{j,i+1} - r_{j,i}| = \frac{d_{all}}{p}$$

Once the partitioning of each group in the hyperspectral image is established, we can perform PCA decomposition on all bands of each group:

$$[c_{k,1}, c_{k,2}, \ldots, c_{k,s_k}] = PCA(H_k), k \in [1, \ldots, p] \tag{15}$$

$PCA(H_k)$ represents the PCA decomposition of the data set H_k composed of bands from the kth group. Finally, the first principal component $c_{k,1}$ obtained after decomposition for each group is used as the guide image for the muGIF filtering of that group:

$$T_i = muGIF(H_i, c_{k,1}), i \in [pe_{k-1}, pe_k], k \in [1, \ldots, p], \tag{16}$$

where H_i represents the image of the *i*th band, and T*i* is the corresponding output image after the muGIF filtering. All the filtered images corresponding to the bands are then reassembled into a new data set H̃, which is subsequently involved in the classification process.

2.2.2. Classification Algorithm Based on MuGIF and CNN with Spectral Angle Distance

After obtaining the data set H̃ that the algorithm above has filtered, a classifier can be employed to categorize this data set. In this context, the convolutional neural network is chosen as the classifier. We select training data for the training process, and ultimately, the model with the highest accuracy is picked as the final classifier for classification. This algorithm is named muGIF-CNN in this paper, as detailed in Algorithm 1.

Algorithm 1 muGIF-CNN

 1: **Data:** Hyperspectral datas et H
 2: **Result:** Classifier CNN
 3: Initialize group count p, CNN training epochs ng, training sample ratio sp
 4: Determine the starting position of each band grouping based on Equation (14)
 5: Decompose each band group separately using PCA according to Equation (15), obtaining the first principal component of each group as the guide image
 6: Use different guide images for different band groups, compute the filtered image for each band image based on Equation (16), and assemble them into the new data set H̃
 7: Extract training samples from the hyperspectral data based on the training sample ratio sp
 8: **for** $ng > 0$ **do**
 9:     Train the CNN model using the training samples
10:     Update the weights of the CNN using the gradient descent algorithm
11:     Update ng;
12: **end for**
13: After training, save the model CNN with the highest accuracy and output;

2.2.3. Classification Algorithm Based on MuGIF and GAN with Spectral Angle Distance

Similarly, we choose the SADGAN algorithm [22], combined with muGIF filtering, which will implement the spatial–spectral joint classification of hyperspectral images in a semi-supervised manner. This algorithm is named muGIF-SADGAN in this paper, and the specific description of the algorithm is shown in Algorithm 2.

Algorithm 2 muGIF-SADGAN

 1: **Data:** Hyperspectral data set H
 2: **Result:** Classifier finetuning CNN(FTCNN)
 3: Initialize group number p, SADGAN training epochs ng, FTCNN training epochs nc, extracted convolutional layer feature depth nl, training sample proportion sp
 4: Calculate the starting position of each band grouping according to Equation (14)
 5: Use PCA to decompose each band group separately based on Equation (15), and obtain the first principal component of each group as the guidance image
 6: Calculate the filtered image of each band based on Equation (16). Use different guidance images for different band groups and form the new data set H̃
 7: **for** $ng > 0$ **do**
 8:     Acquire m generated samples $g^{(1)}, \ldots, g^{(m)}$ and m real samples $x^{(1)}, \ldots, x^{(m)}$
 9:     Feed both the generated and real data into model D of SADGAN for training and update the weights of D using gradient ascent
10:     Take m noise samples and feed them into model G of SADGAN for training, update the weights of G using gradient descent
11:     Update ng;
12: **end for**

Algorithm 2 *Cont.*

13: After training, save models G and D
14: Select training samples from the hyperspectral data according to the proportion sp
15: Feed the training samples into model D, then extract features from nl convolutional layers, flatten them, and concatenate them as inputs for CNN
16: **for** $nc > 0$ **do**
17:     Train the FTCNN with the fused features of training samples
18:     Update the weights of FTCNN using the gradient descent algorithm
19:     Update nc ;
20: **end for**
21: After training, save and output the model with the highest accuracy

3. Datasets and Experimental Setting

This section discusses the details of the data sets chosen for the experiments and the setup of the various methods used in the experiments for comparison with the methods we propose.

3.1. Data Sets Description

To validate the effectiveness of the proposed muGIF-CNN and muGIF-SADGAN methods, we conducted comparative experiments on the Indian Pines, Pavia University, Salinas Valley, and Tianshan data sets. The first three datasets are widely used datasets used to validate the proposed method, and Tianshan is an actual application data set. We extracted 5%, 0.5%, 0.5%, and 10% of the data from them, respectively, as training data (further divided into training and validation data sets at an 8:2 ratio), with the remaining samples constituting the testing data set.

The first hyperspectral data set utilized in this study is the Indian Pines data set, which was acquired in 1992 using the airborne visible/infrared imaging spectrometer (AVIRIS) sensor over the Indian Pines region in northwestern Indiana. The image consists of 145×145 pixels and encompasses 220 spectral bands, covering a wavelength range from 400 to 2200 nm with a fine spectral resolution. The spatial resolution of the image is approximately 20 m. Prior to the experiments, bands affected by water absorption and noise were eliminated, resulting in a data set comprising 200 bands. For our experiments, we employed a total of 10,249 pixels representing sixteen distinct classes. Figure 2 showcases the color composite of the Indian Pines image alongside the corresponding ground truth data.
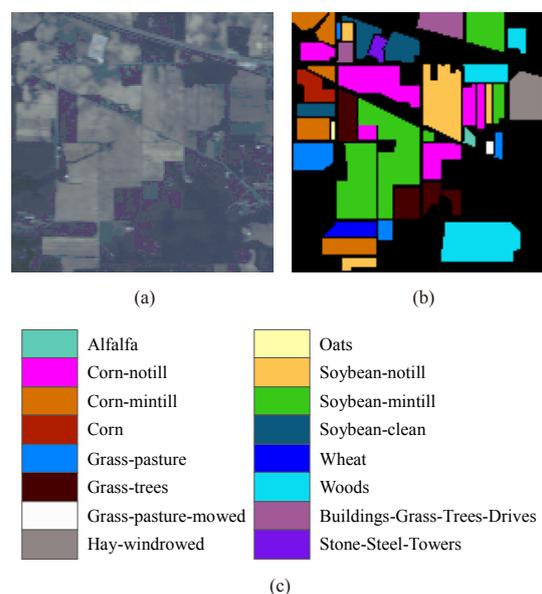

(a)   (b)

	Alfalfa		Oats
	Corn-notill		Soybean-notill
	Corn-mintill		Soybean-mintill
	Corn		Soybean-clean
	Grass-pasture		Wheat
	Grass-trees		Woods
	Grass-pasture-mowed		Buildings-Grass-Trees-Drives
	Hay-windrowed		Stone-Steel-Towers

(c)

Figure 2. Indian Pines data set. (**a**) Three-band color composite. (**b**) Reference data. (**c**) Class names.

The second data set used in this study was captured by the airborne Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the urban area of the University of Pavia, located in northern Italy, in 2002. The image has dimensions of 610×340 pixels. It possesses a high spatial resolution of 1.3 m and covers a spectral range from 430 to 860 nm. The acquired image consists of 103 bands. In total, there are 42,776 samples across nine distinct categories. Figure 3 displays the color composite of the Pavia University image along with the corresponding ground truth data.
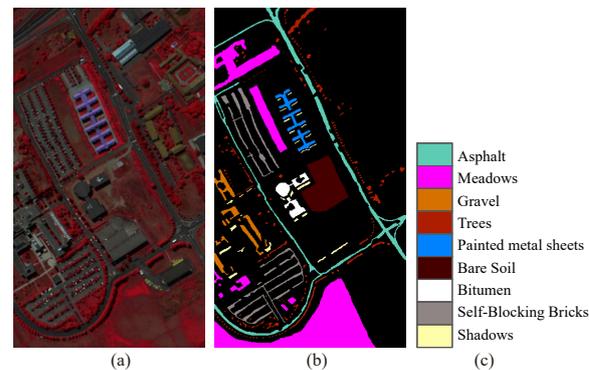

Figure 3. Pavia University data set. (**a**) Three-band color composite. (**b**) Reference data. (**c**) Class names.

The third data set was collected by the AVIRIS sensor over Salinas Valley in southern California, USA, in 1998. The image size is 512×217 pixels, the coverage is 400 to 2500 nm, and the spatial resolution is 3.7 m. After the noisy and water absorption bands were removed, the number of bands in the acquired image was 204. There are a total of 54,129 samples, including 16 classes. Figure 4 shows the color composite of the Salinas image and the corresponding ground truth data.
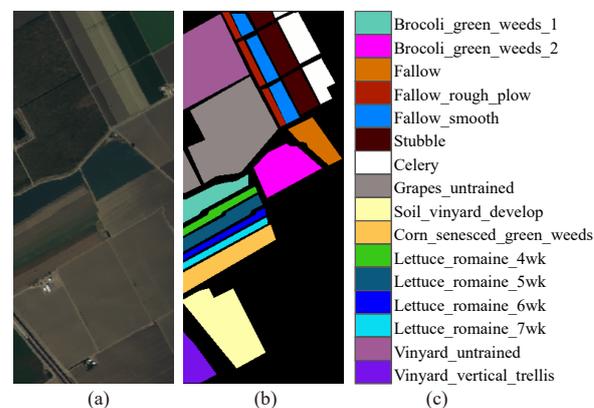

Figure 4. Salinas Valley data set. (**a**) Three-band color composite. (**b**) Reference data. (**c**) Class names.

The fourth data set used in this study is the airborne HyMap data collected over Tianshan, China. The HyMap imaging spectrometer operates within a spectral range of 400–2480 nm. The spectral bandwidth of the data is not fixed and typically falls between 15 nm and 18 nm, with an average bandwidth of approximately 16 nm. The data set possesses a spatial resolution of 9 m and a pixel resolution of 1090×1090. The spectral response values of the features span from 1 to 10,000. The experimental data were subjected to atmospheric and geometric correction. After removing bands affected by water absorption and noise, the data set was reduced to 123 bands. Figure 5a represents the color composite map of the Tianshan data. Based on the existing local geological map, the reference data are depicted in Figure 5b. Figure 5c provides the name of each class, accompanied by the color legend.
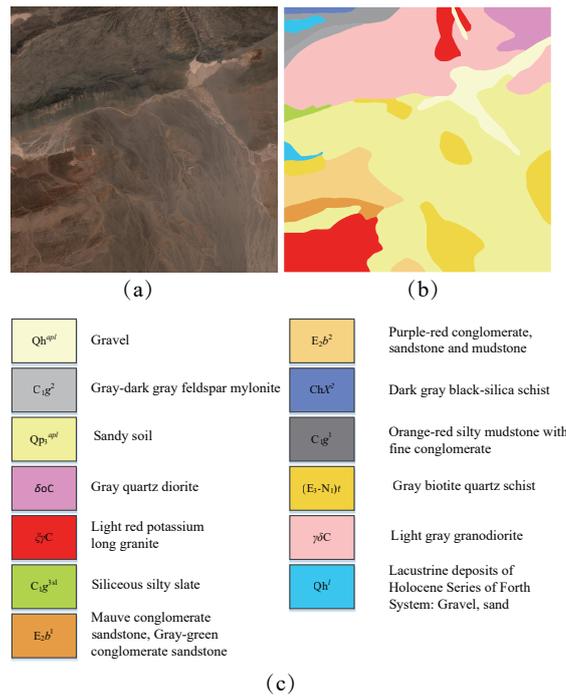
Figure 5. Tianshan data set. (**a**) Three-band color composite. (**b**) Reference data. (**c**) Class names.

Table 1 presents the sample sizes for the Indian Pines, Pavia University, Salinas, and Tianshan data sets. The Tianshan data set has a large number of data and small inter-class differences, which significantly differ from the first three datasets. This difference can verify the applicability of the proposed method on different sensor datasets in the experiments. This disparity also implies that the computational processing time for the same algorithm will be longer when applied to the Tianshan data set. To mitigate this issue, we employed a band selection algorithm to downscale the Tianshan data set, thereby reducing the processing time for subsequent experiments.

Table 1. Scene size, sample size, and number of bands for the Indian Pines, Pavia University, Salinas Valley, and Tianshan data sets.

Data Set	Scene Size	Sample Size	Number of Bands
Indian Pines	145×145	10,249	200
Pavia University	610×340	42,776	103
Salinas Valley	512×217	54,129	204
Tianshan	1090×1090	1,188,100	123

Figure 6 depicts the average spectral curves for each class within the four data sets mentioned. A notable observation is that the Tianshan data set showcases smoother intra-class spectral features and smaller inter-class differences when compared to the other publicly available data sets. This suggests that the Tianshan data set exhibits relatively limited variations between different classes of geologic bodies. Consequently, classification algorithms employed for this data set must possess enhanced performance to effectively differentiate between the various classes and facilitate efficient geologic mapping.

Moreover, obtaining corresponding training samples for hyperspectral images, particularly in geologic mapping regions characterized by complex environments, presents significant challenges. Therefore, reducing reliance on labeled samples in classification becomes crucial. To address this challenge, in addition to a supervised classification approach, we also adopt a semi-supervised classification approach utilizing GANs for spectral classification and the muGIF algorithm for spatial–spectral classification.
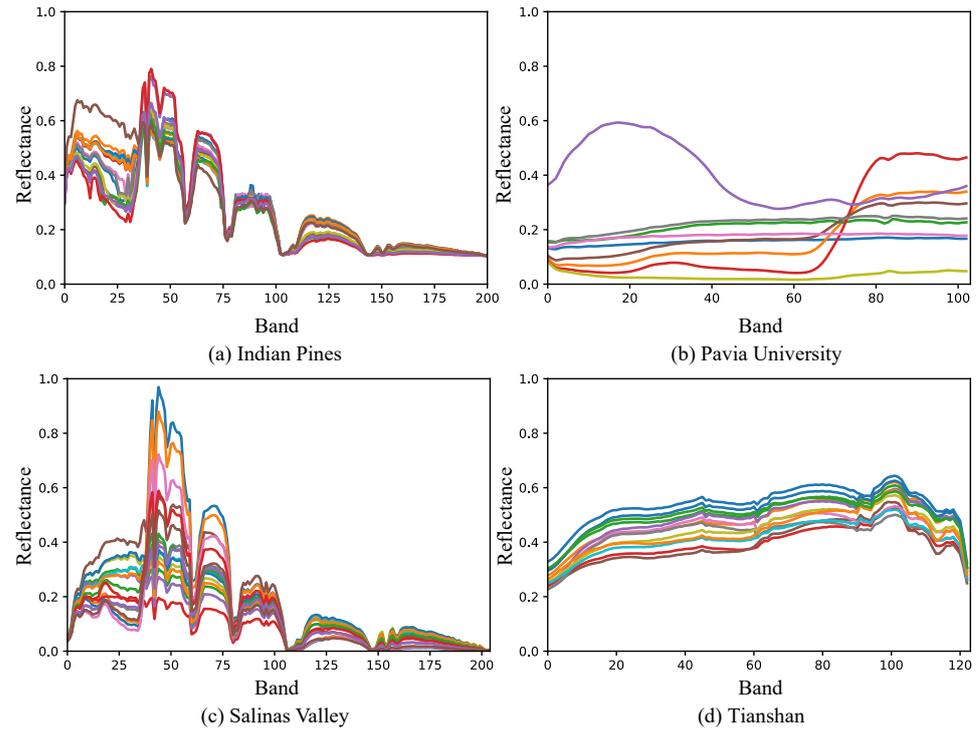
Figure 6. Average spectral curves for each class of the data sets. (**a**–**d**) represent the average curves for each class of Indian Pines, Pavia University, Salinas Valley, and Tianshan data sets respectively. Different colors represent the mean values of different classes.

3.2. Methods of Comparison and Experimental Setup

Methods compared with muGIF-CNN and muGIF-SADGAN include 1DCNN, EMP-SVM, GIF-FSAE, Gabor-CNN, SSGAT, and GIF-CNN, where those with CNN indicate that the method uses the convolutional neural network as the classifier. Unless expressly noted, relevant parameters in the methods are determined through five-fold cross-validation in the experiments.

Details of the setup of these methods in the experiment are described below:

(1) 1DCNN [45]: The method 1DCNN employs a one-dimensional deep convolutional neural network to classify spectral features. For consistency in comparison standards, unless otherwise specified, other methods in this experiment that use CNN as a classifier all adopt the structure of 1DCNN.

(2) EMP-SVM [33]: For the method, several principal components of the hyperspectral image are first extracted using the principal component analysis (PCA) method (four principal components in this experiment). Subsequently, the EMP algorithm is employed to extract spatial–spectral features, with a window radius size of 4. We performed four opening and closing operations on each principal component, resulting in $4 \times (2 \times 4 + 1) = 36$ spatial–spectral features derived from the four principal component images. Finally, we employed the Support Vector Machine (SVM) to classify these features.

(3) GIF-FSAE [46]: This is a method for hyperspectral image classification that combines guided image filtering and sparse autoencoders, integrating unsupervised and supervised feature learning. Initially, the raw hyperspectral image undergoes PCA decomposition. On one hand, the first three principal components are used as guide images, while on the other, the initial 30 principal components are stacked to serve as input images. Guided image filtering utilizes two window radius sizes for spatial feature extraction. These spatial features are then fused with spectral features and introduced to the sparse autoencoder for model training.

(4) Gabor-CNN [36]: This method combines Gabor filtering with CNNs. Initially, spatial features are extracted from the three principal components obtained through PCA decomposition of the hyperspectral image using Gabor filters. These Gabor filters consist

of four different orientations. The resulting spatial features are then stacked together to form a new data set, which is subsequently fed into the CNN for training.

(5) SSGAT: The SSGAT method [47] utilizes the aggregation of features from both labeled and unlabeled samples by computing attention coefficients between a node and its neighboring nodes. This approach effectively combines spectral and spatial information from all the samples. In this experiment, the weight parameters of the SSGAT method were optimized using the Adam optimizer. The learning rate was set to 0.01, and the number of training epochs was set to 300.

(6) GIF-CNN: The GIF-CNN method [41] is a hyperspectral image classification approach that combines multi-scale guided image filtering with CNNs.

(7) muGIF-CNN-1PC: This is a method proposed in this paper for hyperspectral image classification based on mutually guided image filtering and CNNs, using the first principal component as the guide image. As defined in Equation (5), the muGIF algorithm itself comprises six parameters: ϵ_r, ϵ_t, α_r, α_t, β_r, β_t. Among them, ϵ_r and ϵ_t are utilized to prevent division-by-zero errors and are set to $\epsilon_r = \epsilon_t = 0.01$ in this paper. As indicated by Equations (7) and (8), the performance of muGIF is determined by $\frac{\alpha_t}{\beta_t}$ and $\frac{\alpha_r}{\beta_r}$. Initially, we set $\beta_t = 1$ and $\beta_r = 1$, thus only considering the parameters α_t and α_r. According to Equation (6), α_t and α_r influence the input image and the guide image, respectively. Larger values result in smoother output images. Since our emphasis is on filtering the input image, we only consider the setting of the α_t parameter. The specific value of α_t is determined through five-fold cross-validation, and in our experiments, we set $\alpha_t = 0.01$.

(8) muGIF-CNN: In the muGIF-CNN method, the guiding image is obtained using a band distance grouping method. For each data set, the bands are divided into five groups. Each band group undergoes PCA decomposition, and the first principal component is selected as the guiding image for filtering that particular band group. The remaining parameters and settings of muGIF-CNN are consistent with the muGIF-CNN-1PC method.

(9) muGIF-SADGAN: Another hyperspectral image-classification method proposed in this paper is muGIF-SADGAN. This method combines mutually guided image filtering with the SADGAN algorithm [22]. The parameters used for muGIF in muGIF-SADGAN are consistent with those of the muGIF-CNN-1PC method.

The experimental hardware environment for this section consisted of an Intel Core i5-4590 3.3 GHz CPU, 20 GB of RAM, and a TitanX GPU. The computer operating system used was Ubuntu 16.04. The algorithms were developed using TensorFlow, the Keras deep learning library, and the Scikit-learn machine learning library. To quantitatively evaluate the experimental results, the confusion matrix was utilized as the basis. Key performance metrics such as Overall Accuracy (OA), Average Accuracy (AA), and the kappa coefficient κ (%) were calculated to compare the classification performance of different methods.

## 4. Experimental Results and Analysis

This section discusses the classification results of various comparative methods discussed in the Section 3 and the proposed methods on four datasets. Finally, an analysis and discussion are conducted on the impact of muGIF parameters' settings on the classification results.

4.1. Experimental Results for the Indian Pines Data Set

Table 2 presents the Overall Accuracy (OA), Average Accuracy (AA), and kappa coefficient κ (%) achieved by different comparison methods on the Indian Pines data set. For this evaluation, 5% of the samples from each class were randomly selected as training samples. Additionally, Figure 7 showcases the classification results obtained by each method. The experimental findings clearly demonstrate the superior performance of the methods proposed in this paper, which are based on the combination of muGIF and deep learning, for spatial–spectral classification of hyperspectral images.

Table 2. Comparison of the classification accuracies (%) of various methods for the Indian Pines data set (with 5% of training samples; bold numbers indicate the best results).

Class	1DCNN	EMP-SVM	GIF-FSAE	Gabor-CNN	SSGAT	GIF-CNN	muGIF-CNN-1PC	muGIF-CNN	muGIF-SADGAN
1	71.74	80.43	97.83	95.65	93.48	95.65	**99.55**	95.65	**99.55**
2	62.32	83.26	86.69	91.32	89.99	93.91	89.92	94.61	**97.27**
3	54.7	84.94	93.98	91.08	94.46	92.77	90.12	**99.04**	97.11
4	43.88	64.14	91.14	97.47	**100**	97.89	**100**	97.05	100
5	75.36	79.09	92.34	92.55	86.13	91.3	**95.45**	86.13	92.96
6	93.15	92.19	97.81	97.26	97.67	97.67	99.45	**99.86**	99.45
7	75.00	85.71	92.86	92.86	**100**	92.86	28.57	**100**	**100**
8	91.84	99.16	98.12	98.74	99.79	98.95	**100**	**100**	99.79
9	0	0	70	80	**100**	**100**	**100**	**100**	**100**
10	76.03	84.77	96.3	95.78	96.40	96.09	95.78	98.05	**98.87**
11	78.78	94.09	96.7	96.7	97.84	96.82	**98.29**	97.27	97.56
12	73.86	80.61	96.63	92.92	97.98	95.11	98.15	93.25	**98.48**
13	**99.51**	**99.51**	99.02	99.02	**99.51**	**99.51**	**99.51**	99.02	99.02
14	91.23	92.41	100	99.76	99.92	99.37	99.6	99.76	**100**
15	52.85	94.3	71.76	82.9	97.93	88.86	97.93	**98.45**	97.93
16	78.49	94.62	**97.85**	97.85	97.85	97.85	97.85	97.85	97.85
OA	75.43	88.53	94.32	95.01	96.20	95.91	96.25	97.06	**98.08**
AA	69.92	81.83	92.44	93.87	96.81	95.91	92.89	97.25	**98.11**
κ	71.94	86.91	93.51	94.31	95.67	95.34	95.72	96.65	**97.81**
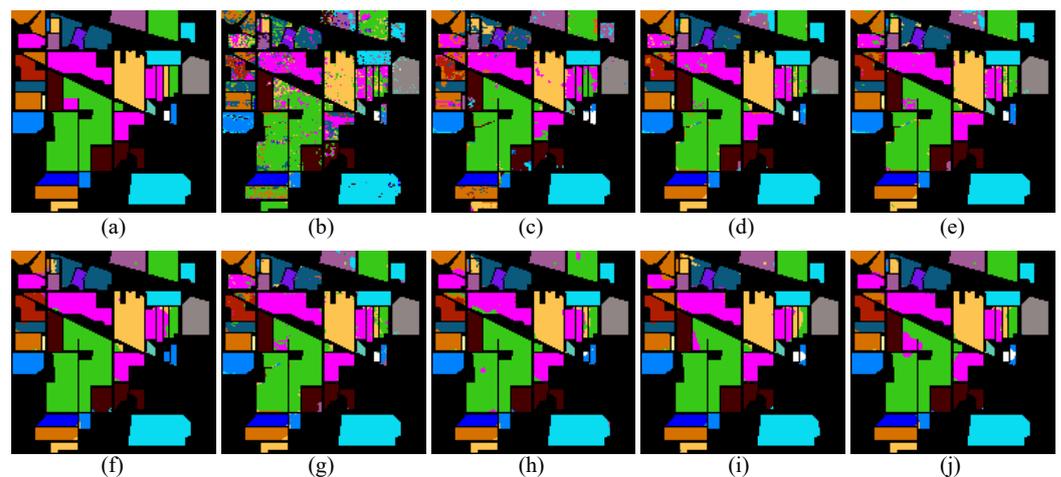
Figure 7. Classification results obtained by different methods on the Indian Pines data set. (**a**) Reference data. (**b**) 1DCNN. (**c**) EMP-SVM. (**d**) GIF-FSAE. (**e**) Gabor-CNN. (**f**) SSGAT. (**g**) GIF-CNN. (**h**) muGIF-CNN-1PC. (**i**) muGIF-CNN. (**j**) muGIF-SADGAN.

From Table 2, it is evident that methods integrating both spatial and spectral features exhibit significant improvements compared to the 1DCNN method, which relies solely on spectral features for classification. This indicates that models incorporating both spatial and spectral features are better suited for the spatial–spectral characteristics of hyperspectral images. Furthermore, deep-learning-based algorithms outperform the traditional classic spatial–spectral classification algorithm EMP-SVM, achieving higher accuracy. Specifically, the proposed muGIF-CNN and muGIF-SADGAN methods in this paper show an improvement of 8.53% and 9.55% in Overall Accuracy (OA) compared to EMP-SVM, respectively.

Methods such as Gabor-CNN and guided-image-based methods (GIF-FSAE, GIF-CNN, muGIF-CNN-1PC, muGIF-CNN, muGIF-SADGAN), thanks to their edge-preserving effect, generally outperform EMP-based methods. Among these, muGIF-based methods have an edge over GIF-based methods as muGIF can handle different structures between the guided

image and target image separately, leading to better edge preservation and homogenous area smoothing. Consequently, methods based on muGIF demonstrate superior classification accuracy. Specifically, compared to GIF-CNN methods, the muGIF-CNN-1PC method achieves an increase in OA of 0.34%, the muGIF-CNN method improves by 1.15%, and the muGIF-SADGAN method improves by 2.17%. Compared to the latest SSGAT method, muGIF-CNN-1PC, with only the first principal component, has mediocre performance, but muGIF-CNN and muGIF-SADGAN also achieve better classification results.

The muGIF-CNN method, which utilizes band distance grouping to obtain multiple principal components as guided images, shows a slight increase in accuracy compared to the muGIF-CNN-1PC method, which uses a single principal component as a guiding image. This suggests that guiding with grouped principal components is beneficial for feature extraction. Additionally, muGIF-SADGAN, which learns features from unlabeled samples in a semi-supervised manner, achieves slightly higher classification accuracy compared to muGIF-CNN. However, it is worth noting that as the spatial feature extraction of muGIF becomes more refined, the difference between semi-supervised and supervised classification methods becomes narrower.

Referring to the classification results depicted in Figure 7, it is evident that the muGIF-CNN-1PC, muGIF-CNN, and muGIF-SADGAN methods exhibit fewer noisy points compared to other classification methods. This observation highlights the notable filtering effect of the muGIF algorithm prior to classification. Specifically, when comparing the central region of Figure 7h–j, it becomes apparent that the muGIF-CNN and muGIF-SADGAN methods accurately identify class 2 objects surrounded by class 11 objects. This implies that utilizing the first principal component as the guiding image in muGIF-CNN-1PC results in excessive information loss from the original bands. In contrast, the muGIF-CNN and muGIF-SADGAN methods, which employ band distance grouping principal components as guiding images for each band group, can better preserve useful information during the filtering process, consequently enhancing classification accuracy.

4.2. Experimental Results for the Pavia University Data Set

Table 3 presents the overall classification accuracy (OA), average classification accuracy (AA), and kappa coefficient κ (%) achieved by different comparison methods on the Pavia University data set. For this evaluation, 0.5% of samples from each class were randomly selected as training samples. Additionally, Figure 8 showcases the classification result maps obtained using each method. The experimental findings unequivocally demonstrate the superior performance of the proposed muGIF-CNN and muGIF-SADGAN methods in the spatial–spectral classification of hyperspectral images.

Table 3. Comparison of the classification accuracies (%) of various methods for the Pavia University data set (with 0.5% training samples; bold numbers indicate the best results).

| Class	1DCNN	EMP-SVM	GIF-FSAE	Gabor-CNN	SSGAT	GIF-CNN	muGIF-CNN-1PC	muGIF-CNN	muGIF-SADGAN
1	65.86	93.45	94.21	93.71	99.49	95.17	**99.55**	**99.55**	**99.55**
2	95.7	98.74	98.54	99.27	99.98	99.67	99.44	**99.98**	99.92
3	44.69	86.52	62.17	59.6	99.67	62.6	**96.9**	80.23	83.28
4	55.71	**93.11**	88.74	86.65	86.00	78.62	84.73	83.91	86.52
5	96.58	98.59	98.59	**98.88**	96.73	98.88	95.32	98.74	98.88
6	72.38	63.97	92.78	95.11	93.02	93.08	96.48	99.44	**99.70**
7	58.5	92.26	38.87	61.95	93.76	77.67	76.54	90.53	**92.48**
8	89.05	79.82	95.06	**96.69**	74.93	95.87	86.01	91.99	93.37
9	**98.1**	80.99	85.85	93.45	38.86	94.51	55.86	74.97	82.05
OA	81.32	90.6	92.27	93.55	94.26	93.72	94.93	96.15	**96.84**
AA	75.17	87.49	83.87	87.26	86.94	88.45	87.84	91.03	**92.86**
κ	74.79	87.29	89.68	91.39	92.33	91.59	93.23	94.86	**95.79**
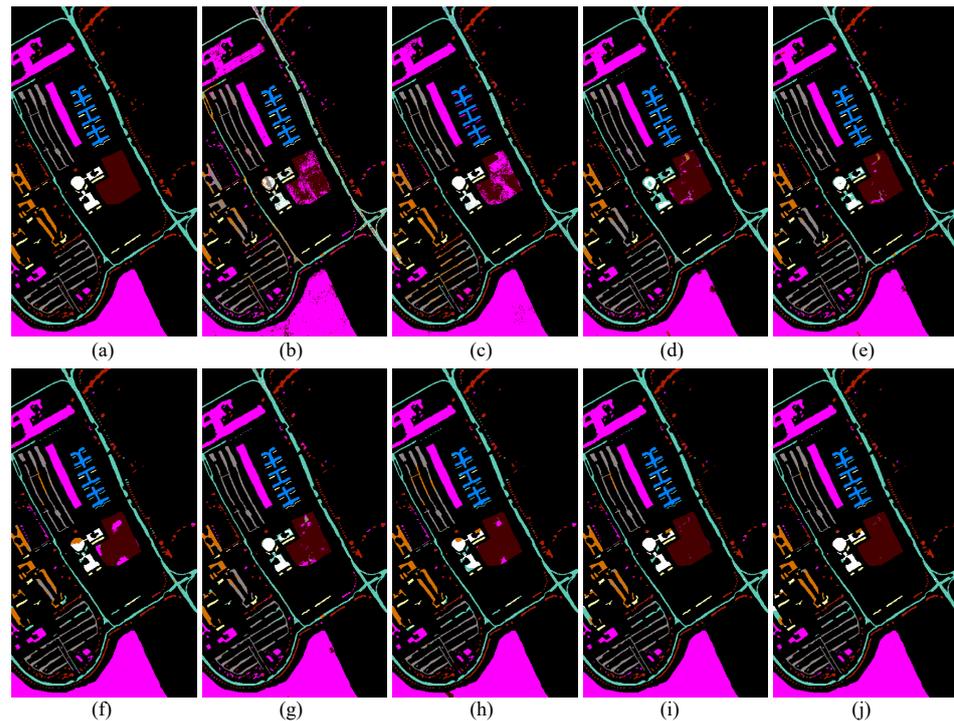
Figure 8. Classification results obtained by different methods on the Pavia University data set. (**a**) Reference data. (**b**) 1DCNN. (**c**) EMP-SVM. (**d**) GIF-FSAE. (**e**) Gabor-CNN. (**f**) SSGAT. (**g**) GIF-CNN. (**h**) muGIF-CNN-1PC. (**i**) muGIF-CNN. (**j**) muGIF-SADGAN.

Table 3 reveals that the performance of different methods on the Pavia University data set is similar to that on the Indian Pines data set. Notably, the 1DCNN method, relying solely on spectral features, exhibits significantly lower performance compared to methods integrating spatial–spectral features. The proposed muGIF-CNN and muGIF-SADGAN methods achieve improvements of 5.55% and 6.24% in OA accuracy, respectively, when compared to EMP-SVM.

Methods utilizing guided images, such as Gabor-CNN and those based on guided image filtering (GIF-FSAE, GIF-CNN, muGIF-CNN-1PC, muGIF-CNN, and muGIF-SADGAN), generally outperform EMP-SVM methods due to their edge-preserving effect. Among these, the muGIF-CNN and muGIF-SADGAN methods demonstrate notable advantages over GIF filtering methods GIF-FSAE and GIF-CNN.

The accuracy of the muGIF-CNN-1PC method, which uses the first principal component as a guiding image, is 1.22% lower than that of the muGIF-CNN method, which adopts band distance grouping principal components. Meanwhile, the semi-supervised muGIF-SADGAN method exhibits slightly higher accuracy than the muGIF-CNN and SSGAT.

Figure 8 shows that, similarly to the classification result maps on the Indian Pines data set, the muGIF-based methods do not exhibit as many noisy points as other classification methods. This observation further emphasizes the significant smoothing effect of muGIF prior to the classification process.

4.3. Experimental Results for the Salinas Valley Data Set

Table 4 presents the overall classification accuracy (OA), average classification accuracy (AA), and kappa coefficient κ (%) achieved by different comparison methods on the Salinas Valley data set. For this evaluation, 0.5% of samples from each class were randomly selected as training samples. Additionally, Figure 9 showcases the classification result maps obtained using each method. Consistent with the experiments conducted on the previous two data sets, the results once again highlight the superior performance of the proposed muGIF-CNN and muGIF-SADGAN methods in the spatial–spectral classification of hyperspectral images.

Table 4. Comparison of the classification accuracies (%) of various methods for the Salinas Valley data set (with 0.5% training samples; bold numbers indicate the best results).

| Class	1DCNN	EMP -SVM	GIF -FSAE	Gabor -CNN	SSGAT	GIF -CNN	muGIF -CNN-1PC	muGIF -CNN	muGIF -SADGAN
1	99.55	99.4	91.64	96.86	95.12	99.85	99.55	99.55	**99.9**
2	99.49	99.62	**100**	**100**	99.95	99.95	89.92	99.87	99.57
3	80.82	72.57	**100**	93.62	99.65	**100**	90.12	99.85	99.95
4	93.4	98.64	96.48	97.56	87.52	97.2	**100**	84.65	88.95
5	93.88	98.13	98.77	97.83	96.15	**99.74**	95.45	96.98	95.59
6	99.75	99.6	99.75	99.85	99.82	**100**	99.45	99.82	99.82
7	99.16	**99.69**	99.27	99.41	98.74	99.66	28.57	99.19	98.88
8	50.32	90.63	85.69	88.86	99.65	86.55	**100**	98.48	99.71
9	97.84	99.94	99.89	99.81	**100**	99.94	**100**	**100**	**100**
10	83.13	97.86	89.66	90.33	99.54	93.26	95.78	**99.76**	99.51
11	79.49	89.51	83.61	81.37	**98.78**	96.44	98.29	98.31	97.28
12	99.43	**100**	99.9	**100**	98.55	**100**	98.15	99.01	99.53
13	98.36	98.91	96.83	98.58	77.73	97.71	**99.51**	86.90	91.16
14	92.52	92.06	95.79	97.29	98.97	98.04	**99.6**	98.79	98.79
15	78.71	58.2	88.29	77.93	93.34	92.41	**97.93**	94.43	97.03
16	76.43	98.4	96.4	86.05	83.18	97.18	97.85	93.69	**99.45**
OA	82.52	90.64	93.7	92.63	97.16	95.41	96.25	97.62	**98.62**
AA	88.35	93.32	95.12	94.08	95.42	97.37	92.89	96.68	**97.82**
κ	80.62	89.54	92.98	91.77	96.84	94.89	95.72	97.39	**98.47**

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)
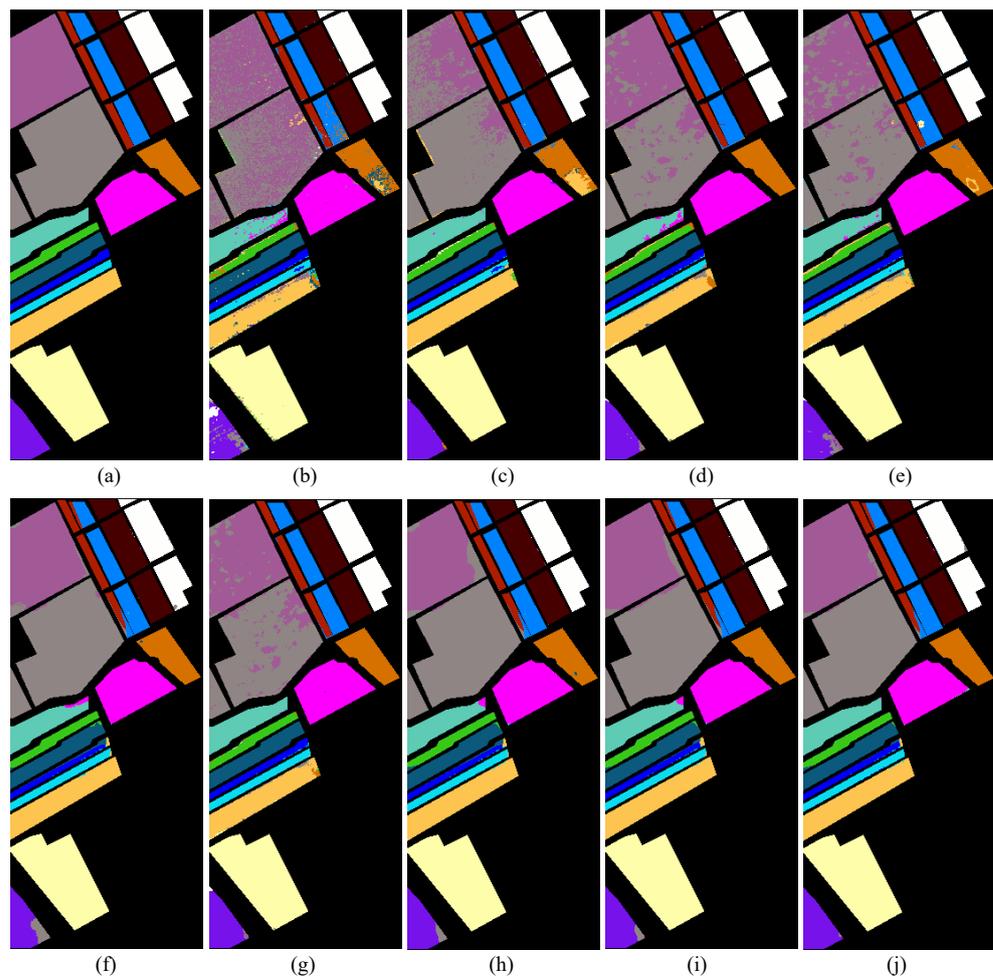
Figure 9. Classification results obtained using different methods on the Salinas Valley data set. (**a**) Reference data. (**b**) 1DCNN. (**c**) EMP-SVM. (**d**) GIF-FSAE. (**e**) Gabor-CNN. (**f**) SSGAT. (**g**) GIF-CNN. (**h**) muGIF-CNN-1PC. (**i**) muGIF-CNN. (**j**) muGIF-SADGAN.

Table 4 reveals that the classification accuracy of the 1DCNN method, which relies solely on spectral features, is generally lower compared to methods incorporating spatial–spectral features. The OA accuracy of the proposed muGIF-CNN and muGIF-SADGAN methods surpasses that of EMP-SVM by 6.98% and 7.98%, respectively. Furthermore, when compared to GIF-based method GIF-FSAE, the muGIF-SADGAN method demonstrates improvements in OA accuracy by 4.92%.

Furthermore, as depicted in Figure 9, both the muGIF-CNN and muGIF-SADGAN methods exhibit exceptional performance, achieving desirable accuracy results even with limited labeled samples. Notably, the result maps generated by these methods are free of excessive noise points.

4.4. Experimental Results for the Tianshan Data Set

The Tianshan data set offers insights into the classification characteristics of hyperspectral remote sensing specifically in geological bodies, setting it apart from the previous three hyperspectral data sets that focused on agriculture. In the experiments, we aim to evaluate various algorithms for classifying geological bodies in hyperspectral images, highlighting their performance in applications of the real-world classification of geological bodies.

For the task of hyperspectral image classification, we have selected several algorithms, namely SVM [33], 1DCNN [45], HSGAN [48], and SADGAN [22]. These algorithms primarily focus on spectral classification.

In addition to these spectral algorithms, we have also employed the SVM-EMP [33], GIF-CNN [41], SSGAT [47], muGIF-CNN, and muGIF-SADGAN algorithms for spatial–spectral classification. These algorithms incorporate both spectral and spatial information for enhanced classification accuracy.

Table 5 provides a comparison of the classification performances of various methods using only 10% of the training data. Upon analyzing the classification results of the first four spectral classification methods, it is observed that, among the deep learning methods, namely 1DCNN, HSGAN, and SADGAN, the semi-supervised learning methods HSGAN and SADGAN, which effectively utilize both labeled and unlabeled samples, outperform the 1DCNN method.

Table 5. Comparison of the classification accuracies (%) of various methods for the Tianshan data set (with 10% training samples; bold numbers indicate the best results).

Class	1DCNN	EMP -SVM	GIF -FSAE	Gabor -CNN	SSGAT	GIF -CNN	muGIF -CNN-1PC	muGIF -CNN	muGIF -SADGAN
1	16.98	71.23	70.58	59.31	41.72	75.16	82.92	81.75	**92.81**
2	0.39	39.13	26.21	27.49	44.07	79.26	83.47	80.02	**86.83**
3	93.29	93.67	92.28	90.13	93.2	96.35	96.68	**96.70**	96.30
4	57.95	85.16	77.62	89.1	80.94	91.93	93.44	**96.78**	96.35
5	75.38	84.04	79.73	81.23	85.35	89.94	90.33	**95.88**	95.15
6	0.24	52.11	45.07	57.4	48.79	88.90	86.59	88.15	**89.72**
7	7.56	62.93	44.77	50.8	48.76	75.07	82.18	76.18	**91.04**
8	69.33	82.6	82.5	67.68	80.09	85.84	86.77	**94.27**	91.38
9	6.17	68.83	59.33	61.5	62.55	84.11	87.18	**89.24**	88.66
10	60.23	72.33	71.72	67.78	69.54	81.55	75.78	82.95	**83.83**
11	39.67	74.96	73.53	72.83	78.3	87.05	89.16	91.82	**94.99**
12	92.03	93.11	92.71	88.27	94.75	94.42	94.79	95.38	**97.05**
13	15.68	61.56	39.28	42.84	55.47	71.26	87.73	86.25	**88.96**
OA	76.59	82.31	85.04	86.58	85.88	91.74	92.80	94.14	**95.19**
AA	41.15	65.87	65.8	70.91	67.97	84.68	87.46	**95.38**	91.77
κ	67.46	77.53	80.06	82.08	81.04	89.06	90.49	86.24	**93.7**

In terms of spatial–spectral joint classification, the non-deep-learning traditional classification method SVM-EMP leverages spatial features. However, its Overall Accuracy lags behind that of other deep learning-based spatial–spectral classification methods and is even slightly lower than that of the spectral-only classification method SADGAN.

Benefiting from the excellent filtering properties of muGIF, the muGIF-CNN method outperforms GIF-CNN and SSGAT, showcasing improvements in OA accuracy of 2.40% and 1.34%, respectively. The superior performance of muGIF-CNN highlights the effectiveness of muGIF in enhancing classification accuracy in comparison to other methods.

Additionally, the Kruskal–Wallis test [49] was used to analyze the statistical significance of differences in performance obtained by the different methods being compared. The results of the Kruskal–Wallis test are a test statistic value and a *p*-value. The test statistic indicates the degree of difference between the results, and the *p*-value represents the probability of obtaining such a difference by chance alone, assuming the null hypothesis is true.

The *p*-value obtained from the test of the eight methods in this experiment is 1.38×10^{-9}, which is significantly lower than the significance level of 0.05. This indicates a substantial difference in the median performance among the methods used in the experimental comparisons. Therefore, we can confidently conclude that the comparison methods are statistically distinct and can serve as reliable benchmarks for evaluating the performance of image classification.

Figure 10 showcases the maps comparing the classification results obtained from the experiments conducted on the Tianshan data set. In these experiments, 50 bands were selected as the final data set using the BSCNN band selection method [50].
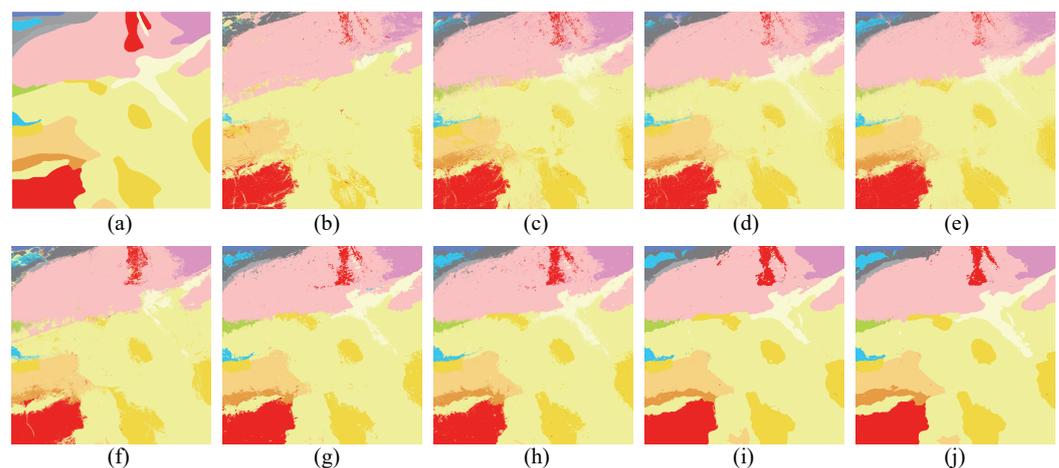
Figure 10. Classification results obtained using different methods on the Tianshan data set. (**a**) Reference data. (**b**) SVM. (**c**) CNN. (**d**) HSGAN. (**e**) SADGAN. (**f**) EMP-SVM. (**g**) GIF-CNN. (**h**) SSGAT. (**i**) muGIF-CNN. (**j**) muGIF-SADGAN.

The classification result maps of SADGAN, which benefit from the generator-generated (augmented) samples and multilayer convolutional features provided by GAN, demonstrate more detailed information. However, it is noticeable that many cluttered spots are present in homogeneous regions, as this method does not utilize spatial features.

Upon analyzing the classification results of the GIF-CNN, SSGAT, muGIF-CNN, and muGIF-SADGAN methods, it becomes evident that the integration of a spatial feature-extraction method aids in correctly categorizing most of the clutter in homogeneous regions, unlike SADGAN. This observation highlights the importance of incorporating spatial features in achieving accurate classification results.

Indeed, it is crucial to acknowledge that the ground truth data, although manually created, may still contain errors and uncertainties. These errors can impact the evaluation of classification methods and introduce challenges in accurately assessing their performance.

While spectral classification methods, such as 1DCNN and SADGAN, often exhibit lower accuracy compared to spatial–spectral joint classification methods, they have the advantage of providing a more detailed representation of ground features. These spectral methods can capture fine spectral variations and subtle differences in the data, resulting in a more granular classification output.

On the other hand, spatial–spectral joint classification algorithms tend to smooth out homogeneous regions by leveraging spatial information. This smoothing effect can potentially lead to the loss of some detailed information present in these regions.

To gain meaningful insights and perform a comprehensive analysis, it is recommended to conduct simultaneous spectral-based classification and spatial–spectral joint classification on the same hyperspectral image data within the same region. By comparing the classification results obtained from both approaches, a more holistic understanding of the data can be achieved, taking into account the strengths and limitations of each method.

4.5. Influence of Mutually Guided Image-Filtering Parameters

Figure 11 depicts the influence of variation in the parameter α_t on image filtering in muGIF-CNN. In Figure 11a, the original image of the 95th band of the Indian Pines data set is presented. As the value of α_t steadily increases, it becomes apparent that the original image progressively becomes smoother or more blurred.

Figure 12 demonstrates the influence of the α_t parameter on the final classification results of the Indian Pines data set. To clearly illustrate its impact on the classification performance, the *x*-axis in the graph is presented on a logarithmic scale.

From Figures 11 and 12, it can be inferred that, for hyperspectral image classification, the gradual smoothing of images while preserving edges is beneficial for extracting spatial features. This process helps to enhance the representation of spatial information in the classification task. However, it is important to note that excessive smoothing can cause hyperspectral image samples to converge, ultimately affecting the overall classification performance.

In the comparative experiment conducted in this paper, a value of $\alpha_t = 0.01$ corresponds to the filtering effect depicted in Figure 11f. At this specific parameter setting, the edge portions in the original image are well preserved, while the interior becomes as smooth as possible. This optimal balance between smoothing and edge preservation results in the highest classification accuracy among the different parameter values tested.

These findings highlight the significance of selecting an appropriate value for the α_t parameter in muGIF-CNN. It is crucial to strike a balance between smoothing and edge preservation to achieve the best classification performance for hyperspectral images.
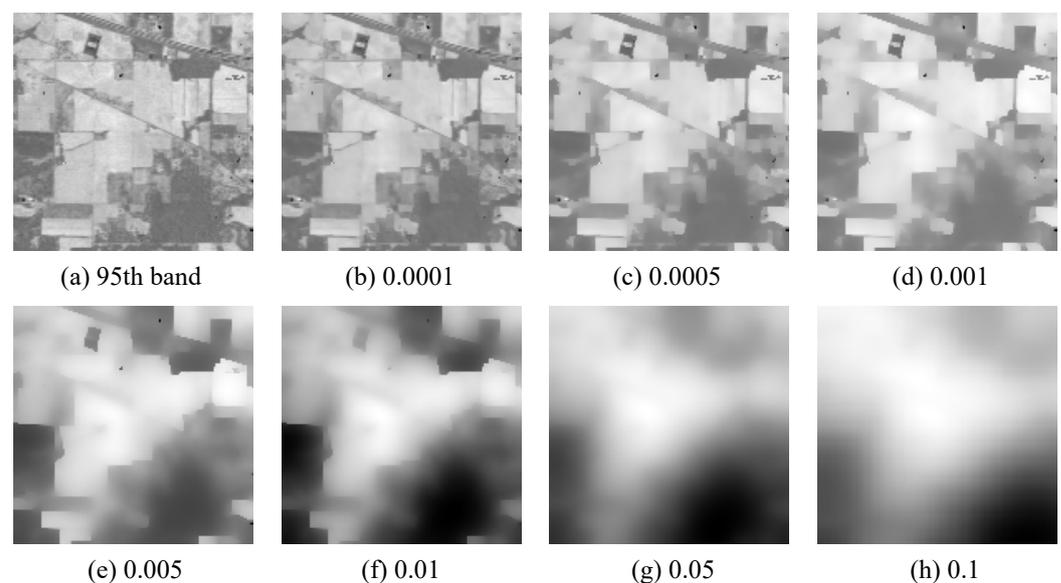

| (a) 95th band | (b) 0.0001 | (c) 0.0005 | (d) 0.001 |
| (e) 0.005 | (f) 0.01 | (g) 0.05 | (h) 0.1 |

Figure 11. Different filtering effects of different α_t parameters. (**a**) represents the image of the 95th band of Indian Pines data set. (**b**–**h**) represent the filtered images of the 95th band with α_t parameters ranging from 0.0001 to 0.1.
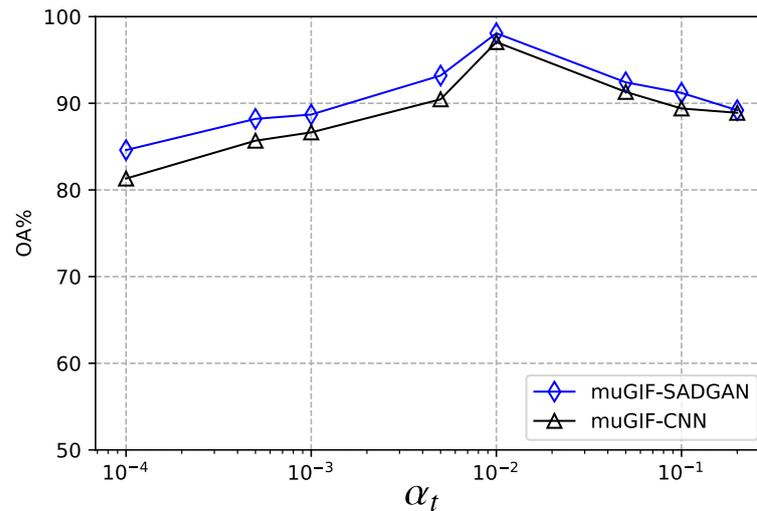
Figure 12. Impact of different α_t parameters on classification accuracy. For clarity, the *x*-axis in the graph is set to a logarithmic scale. There are a total of 8 values for α_t. From left to right, they are 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, and 0.2.

5. Conclusions

In this study, we propose a method for HSI spatial feature extraction based on a muGIF and combined with the band-distance-grouped principal component, which aims to fully exploit the spatial–spectral features of hyperspectral images and enhance their classification performance.

The method extracts the principal components of hyperspectral image band groups based on band distance and uses them as guided images for the muGIF algorithm. Each band group of the hyperspectral image is individually filtered to extract spatial features from the data. Finally, deep learning models, specifically CNNs and GANs, are employed to classify the filtered hyperspectral images.

The key innovation of our approach lies in the application of muGIF with gouping principal components in hyperspectral image classification. Unlike traditional methods that use a single principal component as the guided image, we introduce a novel approach for principal-component segmentation extraction based on band-distance density. This method enriches the theoretical content of hyperspectral image spatial filtering, leading to improved classification results.

Experimental results demonstrate that our proposed muGIF-based approach effectively extracts the spatial–spectral joint features of hyperspectral images, thereby elevating the classification accuracy of hyperspectral imagery. Comparative evaluations against conventional techniques and several recently popular spatial–spectral classification methods validate the superior performance of muGIF-CNN and muGIF-SADGAN.

References

1. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–32. [CrossRef]
2. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [CrossRef]
3. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 27 December 1965; pp. 281–297.
4. Ng, A.; Jordan, M.; Weiss, Y. *Advances in Neural Information Processing Systems: On Spectral Clustering: Analysis and an Algorithm*; MIT Press: Cambridge, MA, USA, 2002.
5. Bezdek, J.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [CrossRef]
6. Guha, S.; Rastogi, R.; Shim, K. Cure: An efficient clustering algorithm for large databases. *Inf. Syst.* **2001**, *26*, 35–58. [CrossRef]
7. Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; pp. 226–231.
8. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semi-supervised hyperspectral image classification based on a Markov random field and sparse multinomial logistic regression. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009; pp. 817–820.
9. Li, J.; Bioucas-Dias, J.; Plaza, A. Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 318–322.
10. Camps-Valls, G.; Bandos, T.; Zhou, D. Semi-supervised graph-based hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3044–3054. [CrossRef]
11. Bruzzone, L.; Chi, M.; Marconcini, M. A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3363–3373. [CrossRef]
12. Zou, J.; He, W.; Zhang, H. LESSFormer: Local-Enhanced Spectral-Spatial Transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5535416. [CrossRef]
13. Liang, H.; Bao, W.; Shen, X.; Zhang, X. HSI-mixer: Hyperspectral image classification using the spectral–spatial mixer representation from convolutions. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6013005. [CrossRef]
14. 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 International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
15. Zhu, J.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
16. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 105–114.
17. Xu, T.; Zhang, P.; Huang, Q.; Zhang, H.; Gan, Z.; Huang, X.; He, X. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1316–1324.
18. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
19. Springenberg, J. Unsupervised and semi-supervised learning with categorical generative adversarial networks. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
20. Odena, A. Semi-supervised learning with generative adversarial networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
21. He, Z.; Liu, H.; Wang, Y.; Hu, J. Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 1042. [CrossRef]
22. Zhan, Y.; Wang, Y.; Yu, X. Semisupervised hyperspectral image classification based on generative adversarial networks and spectral angle distance. *Sci. Rep.* **2023**, *13*, 22019. [CrossRef] [PubMed]
23. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.; Chanussot, J.; Tilton, J. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE* **2013**, *101*, 652–675. [CrossRef]
24. Plaza, A.; Benediktsson, J.; Boardman, J.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **2009**, *113*, 110–122. [CrossRef]
25. He, L.; Li, J.; Liu, C.; Li, S. Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 1579–1597. [CrossRef]

26. Plaza, A.; Martinez, P.; Perez, R.; Plaza, J. A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles. *Pattern Recognit.* **2004**, *37*, 1097–1116. [CrossRef]

27. Li, J.; Bioucas-Dias, J.; Plaza, A. Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 844–856. [CrossRef]

28. Hong, D.; Wu, X.; Ghamisi, P.; Chanussot, J.; Yokoya, N.; Zhu, X. Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3791–3808. [CrossRef]

29. Li, J.; Bioucas-Dias, J.; Plaza, A. Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 809–823. [CrossRef]

30. Xia, J.; Chanussot, J.; Du, P.; He, X. Spectral–spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2532–2546. [CrossRef]

31. Ghamisi, P.; Benediktsson, J.; Ulfarsson, M. Spectral–spatial classification of hyperspectral images based on hidden Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2565–2574. [CrossRef]

32. Shen, X.; Bao, W.; Liang, H.; Zhang, X.; Ma, X. Grouped collaborative representation for hyperspectral image classification using a two-phase strategy. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5505305. [CrossRef]

33. Benediktsson, J.; Palmason, J.; Sveinsson, J. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 480–491. [CrossRef]

34. Liu, Y.; Cao, G.; Sun, Q.; Siegel, M. Hyperspectral classification via deep networks and superpixel segmentation. *Int. J. Remote Sens.* **2015**, *36*, 3459–3482. [CrossRef]

35. Bau, T.; Sarkar, S.; Healey, G. Hyperspectral region classification using a three-dimensional Gabor filterbank. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3457–3464. [CrossRef]

36. Chen, Y.; Zhu, L.; Ghamisi, P.; Jia, X.; Li, G.; Tang, L. Hyperspectral images classification with Gabor filtering and convolutional neural network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2355–2359. [CrossRef]

37. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409. [CrossRef] [PubMed]

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

39. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the IEEE International Conference on Computer Vision, Mumbai, India, 4–7 January 1998; pp. 839–846.

40. Guo, Y.; Cao, H.; Han, S.; Sun, Y.; Bai, Y. Spectral-spatial hyperspectralimage classification with k-nearest neighbor and guided filter. *IEEE Access* **2018**, *6*, 18582–18591. [CrossRef]

41. Guo, Y.; Cao, H.; Bai, J.; Bai, Y. High efficient deep feature extraction and classification of spectral-spatial hyperspectral image using cross domain convolutional neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 345–356. [CrossRef]

42. Guo, X.; Li, Y.; Ma, J.; Ling, H. Mutually Guided Image Filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 694–707. [CrossRef]

43. Rodarmel, C.; Shan, J. Principal component analysis for hyperspectral image classification. *Surv. Land Inf. Sci.* **2002**, *62*, 115–122.

44. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]

45. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [CrossRef]

46. Wang, L.; Zhang, J.; Liu, P.; Choo, K.; Huang, F. Spectral-spatial multi-feature-based deep learning for hyperspectral remote sensing image classification. *Soft Comput.* **2017**, *21*, 213–221. [CrossRef]

47. Zhao, Z.; Wang, H.; Yu, X. Spectral–Spatial Graph Attention Network for Semisupervised Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5503905. [CrossRef]

48. Zhan, Y.; Hu, D.; Wang, Y.; Yu, X. Semisupervised Hyperspectral Image Classification Based on Generative Adversarial Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 212–216. [CrossRef]

49. Vargha, A.; Delaney, H.D. The Kruskal–Wallis test and stochastic homogeneity. *J. Educ. Behav. Stat.* **1998**, *23*, 170–192. [CrossRef]

50. Zhan, Y.; Hu, D.; Xing, H.; Yu, X. Hyperspectral band selection based on deep convolutional neural network and distance density. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2365–2369. [CrossRef]