



## Article

# Hyperspectral Band Selection via Band Grouping and Adaptive Multi-Graph Constraint

Mengbo You <sup>†</sup>, Xiancheng Meng, Yishu Wang, Hongyuan Jin, Chunting Zhai and Aihong Yuan <sup>\*,†</sup>

College of Information Engineering, Northwest A&amp;F University, Yangling, Xianyang 712100, China

\* Correspondence: ahyuan@nwafu.edu.cn

† These authors contributed equally to this work.

**Abstract:** Unsupervised band selection has gained increasing attention recently since massive unlabeled high-dimensional data often need to be processed in the domains of machine learning and data mining. This paper presents a novel unsupervised HSI band selection method via band grouping and adaptive multi-graph constraint. A band grouping strategy that assigns each group different weights to construct a global similarity matrix is applied to address the problem of overlooking strong correlations among adjacent bands. Different from previous studies that are limited to fixed graph constraints, we adjust the weight of the local similarity matrix dynamically to construct a global similarity matrix. By partitioning the HSI cube into several groups, the model is built with a combination of significance ranking and band selection. After establishing the model, we addressed the optimization problem by an iterative algorithm, which updates the global similarity matrix, its corresponding reconstruction weights matrix, the projection, and the pseudo-label matrix to ameliorate each of them synergistically. Extensive experimental results indicate our method outperforms the other five state-of-the-art band selection methods in the publicly available datasets.

**Keywords:** hyperspectral band selection; band grouping; adaptive multi-graph



**Citation:** You, M.; Meng, X.; Wang, Y.; Jin, H.; Zhai, C.; Yuan, A.

Hyperspectral Band Selection via Band Grouping and Adaptive Multi-Graph Constraint. *Remote Sens.* **2022**, *14*, 4379. <https://doi.org/10.3390/rs14174379>

Academic Editors: Qi Wang, Xiangtao Zheng and Fulin Luo

Received: 26 July 2022

Accepted: 30 August 2022

Published: 3 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Unlike traditional RGB three-channel digital images, hyperspectral images (HSI) have multiple channels in spectral dimensions. Because HSI can distinguish land-cover details with high spectral diagnosis ability [1], they are widely applied in city planning, agricultural and forestry detection, topographic map updating, mineral exploration, and many other fields [2,3]. However, HSI provides spectral image information with enormous bands, which is vulnerable to noise [4]. Furthermore, the high dimensionality also leads to great redundancy [5] in hyperspectral data due to the high correlation among bands, posing obstacles to image processing, transmission, storage, and analysis [6,7].

Band selection is one of the most popular dimensionality reduction techniques, it aims to search for a subset of all bands containing as few bands as possible, which has enough content to represent the overall spectral information [8]. Like many other machine learning problems, band selection can be divided into supervised [9], semi-supervised [10,11], and unsupervised [12,13] according to the availability of prior information [14]. Supervised band selection usually sets a standard function to evaluate the similarity between the selected band and the marked image [15]. Unsupervised band selection expects to find a representative subset of bands unrelated to the labeled sample [16]. Supervised or semi-supervised methods need label information [17], but in practical application, the unsupervised method will cover a wider range in utilization because the prior conditions of the objects are difficult to obtain. Therefore, the research on unsupervised band selection technology has gained much attention.

### 1.1. Overview and Motivation

Previous research associated with unsupervised band selection focuses on four categories: ranking-based, clustering-based, searching-based, and embedding learning-based methods. Ranking-based methods quantify the importance of bands based on some indicators to select the top-ranked bands, while the selected subset usually suffers from information redundancy since they ignore the correlation of bands. The clustering-based band selection methods intend to obtain representative bands from each cluster by grouping original data [18], and these selected bands form the subset. This method can minimize interclass variance and maximize interclass variance to avoid redundancy. The searching-based methods select a subset by searching band combinations according to a given criterion function, which transforms band selection into an optimization problem. The embedding-based methods select bands by optimizing specific application models such as classification, target detection, and spectral separation [19].

To address the problem of the ranking-based method and inspired by robust unsupervised feature selection via multi-group adaptive graph representation (MGAGR), we propose an unsupervised band selection method for hyperspectral images based on band grouping and adaptive multi-graph constraint in this paper. We put forward the band grouping strategy to fully mine the effective information and avoid strong correlations among adjacent bands. Considering that band groups with high definition and abundant information should have higher significance, different weights are assigned to each group to construct the global similarity matrix. Spatially, hyperspectral images can be divided into various regions. For example, hyperspectral images used for vegetation damage recognition can be divided into non-vegetation (background), healthy vegetation, pest vegetation, water-deficient vegetation, etc. Obviously, non-vegetation is not the key information, while the others are important and useful. We construct a graph matrix through the correlation between pixels to save the information on the spatial dimension. The local similarity matrix is defined as the graph matrix constructed by each spectral group. In order to describe the spatial information more accurately and comprehensively, we acquire the global similarity matrix using the linear combination of multiple local similarity matrices. On this basis, we use the regularization constraints to ensure the accuracy of classification. Our method with high-class separability takes sufficient account of spatial and spectral correlation, while utilizing pseudo-label matrix concurrently, and has high classification separability. After establishing the model, we solve the optimization problem by an iterative algorithm that updates the global similarity matrix, its corresponding reconstruction weights matrix, and the projection and pseudo-label matrix to synergistically ameliorate each of them.

### 1.2. Contributions

The key contributions can be summarized as follows:

1. The method of band grouping is used originally to process hyperspectral data, which mines the context information of the whole spectral dimension and avoids redundancy in order to obtain the more accurate selected subset.
2. An unsupervised adaptive graph constraint is introduced into the hyperspectral band selection model. The global similarity matrix is reconstructed by the linear combination of the similarity matrix of all groups with adaptive weighting.
3. An iterative optimization algorithm is proposed for obtaining the optimal weights of the proposed model. Moreover, the objective function is solved by the algorithm to select the optimal subset of bands. Through several experiments, the results are compared with the results of previous methods to verify the efficiency of our algorithm.

### 1.3. Organization

The rest of this paper is organized as follows. In Section 2, some related works are briefly introduced. The detailed model of our method is presented in Section 3. Section 3.2 provides a skillful optimization algorithm for the proposed model. To validate the proposed

method, the experimental results are shown and analyzed in Section 4. At last, Section 5 makes a brief conclusion for this paper.

## 2. Related Works

In this section, we would like to briefly review some representative unsupervised band selection methods. UBS is usually implemented with four classical schemes, ranking-, clustering-, searching- [20], and embedding-based methods [21]. As introduced earlier, ranking-based methods quantify the importance of bands based on some indicators to gather the top-ranked bands, and the selected subset may have high information redundancy since the method ignores the internal correlation between bands. For example, maximum-variance principal component analysis (MVPCA) [22] and the manifold ranking-based band selection algorithm [23] are typical algorithms. The clustering-based band selection methods aim to obtain representative bands from each cluster through grouping original data [18], which can minimize inter-class variance and avoid redundancy. By conducting the local density and the intra-cluster distance of each point, fast density-peak-based clustering (FDPC) [24] identifies cluster centers as points with anomalously large scores. The adaptive subspace partition strategy [25] regards the attained sub-cube as a framework, which means that other criteria can be applied to the selection strategy. Meanwhile, this method estimates band noise level to get high-quality images. Although these clustering-based methods have achieved great success, they do not take the global information and spatial distribution of different objects into consideration well, while being sensitive to initial conditions.

The searching-based methods, such as the firefly algorithm (FA) [26] and particle swarm optimization (PSO) [27], transform band selection into an optimization problem using an objective function. The FA can automatically adjust the induction radius and search for multiple peaks at the same time. PSO is always used to optimize parameters for classification methods such as kernel-based fuzzy c-means and support vector machine (SVM) [28]. However, existing searching-based strategies almost always consider only one score, leading to poor effect in the application of multi-graph structure for band selection [29]. The embedding-based methods select bands by optimizing specific application models. These models include classification, target detection, and spectral separation [19]. Spectral analysis and sparse constraint are commonly used in typical this method. For instance, recursive support vector machines [30] and sparse multinomial logistic regression algorithms [31] are commonly used typical methods. In order to reveal the geometric structure embedding in original high-dimensional data, many manifold learning methods are introduced, such as locally linear embedding (LLE) [32], Laplacian eigenmaps (LE) [33], and neighborhood preserving embedding (NPE) [34]. The above methods can be unified under the graph embedding framework (GE) [35].

In addition to the methods mentioned above, the wide application of deep learning (DL) [36] and deep neural networks (DNN) [37] has demonstrated remarkable achievements in HSI processing. Generally speaking, HSI classification algorithms include traditional machine learning techniques and DL methods that require feature processing [38]. Compared with typical classifier technologies such as SVM and k-nearest neighbor (KNN) [39], DNN can minimize the dimension of data representation and effectively identify targets. For example, the authors of [40] propose two versions of BS-Nets (band selection networks), which are implemented using fully connected networks and convolutional networks, showing less redundant results and competitive time cost. Deep reinforcement learning is used with Q-network in [41] and its validity has been verified extensively with multiple datasets and classifiers. Moreover, an increasing number of researchers have attempted to use a convolutional neural network (CNN) to exploit deep features of HSI classification [20].

we analyzed the above classical methods, and the corresponding advantages and disadvantages are shown in Table 1.

**Table 1.** The advantages and disadvantages of classical methods.

Method	Pros	Cons
MVPCA	All bands are ranked by the variance of band capacity.	Redundancy of band information is not considered.
UBS	Using divergence based on the analysis of band features, it tries to solve the redundancy problem caused by sorting algorithm.	The spatial information of HSI is not considered.
FDPC	It is a clustering method based on weighted normalized local density and ranking.	The selected bands do not necessarily contain the most information, and different metrics will affect the results. Moreover, the random initialization of the clustering algorithm is uncertain.
FA	FA can reduce the complexity of the ELM (extreme learning machine) network, and is suitable for optimizing the parameters in the network. It converges faster compared with PSO.	It is sensitive to parameters and will be less attractive when the dimension is high, affecting the result update.
PSO	It is a probabilistic global optimization algorithm that is relatively simple and easier to implement.	Its peak seeking rate and solution accuracy are low.
ABA (Attention-Based Autoencoders)	This method presents an automatic encoder based on an attention mechanism to realize the non-linear relationship between bands.	The optimization process of its hyperparameters is random, which leads to the instability of the model.
ABCNN (Attention-Based Convolutional Neural Networks)	This method attains the optimal subset of bands by coupling attention-based CNNs with anomaly detection.	Deep learning incorporating an attention mechanism is prone to over-fitting.
DRL (Deep Reinforcement Learning)	It is a deep learning method for environment simulation that makes full use of hyperspectral sequence to select bands.	Since this is an algorithm based on deep learning, it takes more time to train.
BS-Nets	It proposes a deep learning method combined with an attention mechanism and rebuilding RecNet. The framework is flexible and can adapt to more existing networks.	Models need a long time to be trained.

### 3. Methods

This part introduces our unsupervised band selection algorithm based on band grouping and adaptive multi-graph constraints. Each step of the proposed method is detailed below.

#### 3.1. Model Construction

Let  $\mathbf{X} = (x_1 \ x_2 \ \dots \ x_{mn}) \in \mathbb{R}^{r \times mn}$  be the high-dimensional data, which are reconstructed from the original  $m \times n \times r$  data set.  $mn$  is the number of pixel points and  $r$  is the number of the spectral bands.  $x_i \in \mathbb{R}^{r \times 1}$  represents the information of the  $i$ -th pixel. We want to project the pixel into the label space, assuming that all the data consists of  $c$  labels. Self-expression matrix  $\mathbf{Z} \in \mathbb{R}^{r \times c}$  maps each pixel point to the label space. Let  $\mathbf{Y}_p \in \mathbb{R}^{mn \times c}$

denote the pseudo-label matrix. The error between  $\mathbf{X}^T \mathbf{Z}$  and  $\mathbf{Y}_p$  should be minimized on this condition:

$$\min \left\| \mathbf{X}^T \mathbf{Z} - \mathbf{Y}_p \right\|_{F'}^2, \quad (1)$$

where  $\mathbf{Z}$  can be optimized to obtain the best linear combination of  $r$  bands to approximate the pseudo-label information. Maintaining the sparsity of the matrix reduces the number of bands selected and as a result, the retained bands contain more valid information on account of being less affected by noise. Consequently, the sparsity of the representation matrix is maintained by the sparsity rule constraint, meanwhile, the selected bands are guaranteed to better reflect the representative information. In this paper,  $l_{2,1}$ -norm is exploited as the sparse regular constraint:

$$\min \alpha \|\mathbf{Z}\|_{2,1}, \quad (2)$$

where  $\alpha$  is a weight factor to control the scale of the sparse regular constraint.

To improve the performance of the self-expression model, the manifold regularization method should be utilized to ensure that spatial information is maintained during optimization. For instance, if the similarity between two pixels is large, this property should be reflected after projecting them into the label space and the distance between their pseudo label should be optimized to a small value. The global similarity matrix can be noted as  $\mathbf{S} = [\mathbf{s}_{ij}]_{mn \times mn}$ , and  $\mathbf{s}_{ij}$  denotes the similarity between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Then, the manifold regularization can be expressed as the following formula:

$$\min \frac{\beta}{2} \left\| \mathbf{Y}_p^i - \mathbf{Y}_p^j \right\|_{\mathbf{s}_{ij}}, \quad (3)$$

where  $\beta$  is used to adjust the weight of the manifold regularization term. We set that the similarity matrix  $\mathbf{S}$  should satisfy the following conditions:

1.  $\mathbf{s}_{ij} = \mathbf{s}_{ji}$ , i.e.,  $\mathbf{S}$  is a real symmetric matrix;
2. For any sample  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , the similarity value should be between 0 and 1, i.e.,  $0 \leq \mathbf{s}_{ij} \leq 1$ . It means that the closer the similarity is to 1, the more similar the two columns of data;
3. The sum of each row (or each column) of  $\mathbf{S}$  equals to 1, i.e.,  $\sum_{j=1}^{mn} \mathbf{s}_{ij} = 1$  and  $\sum_{i=1}^{mn} \mathbf{s}_{ij} = 1$ .

Considering the above constraints, and to further improve the classification effect, we constrain the  $\mathbf{Y}_p$  matrix to be orthogonal. Equation (3) can be rewritten as follows:

$$\min \beta \text{Tr} \left( \mathbf{Y}_p^T \mathbf{L} \mathbf{Y}_p \right), \text{ s.t. } \mathbf{S}_1^T \mathbf{1}_{mn} = 1, \mathbf{Y}_p^T \mathbf{Y}_p = \mathbf{I}_c \quad (4)$$

where  $\text{Tr}()$  denotes trace of matrix and  $\mathbf{1}_{mn}$  is a  $m \times n$ -dimensional all-one column vector. Let  $\mathbf{L} \in \mathbb{R}^{mn \times mn}$  be the Laplacian matrix used to keep the geometric structure of the data, that is, to maintain the information of bands in the spatial dimension.  $\mathbf{L}$  is calculated as the following formula:

$$\mathbf{L} = \mathbf{D} - \mathbf{S}, \quad (5)$$

Here  $\mathbf{S}$  is the global similarity matrix mentioned above. Additionally, define a diagonal matrix  $\mathbf{D} \in \mathbb{R}^{mn \times mn}$ .  $\mathbf{d}_{ii}$  is the  $i$ -th diagonal element of  $\mathbf{D}$  calculated by  $\mathbf{d}_{ii} = \sum_{j=1}^{mn} \mathbf{s}_{ij}$ , which means  $\mathbf{d}_{ii}$  is the sum of  $i$ -th column of  $\mathbf{S}$ .

In previous experiments,  $\mathbf{S}$  is computed from  $\mathbf{X}$  representing the original data and unchanged, so  $\mathbf{S}$  will not be updated and optimized. However, in this paper, the pseudo-label matrix, self-expression matrix, and other matrices mentioned later are updated step by step using the idea of iterative optimization.

We think that other matrices will also affect the calculation of  $\mathbf{S}$  during the iteration process. For example,  $\mathbf{Y}_p$  should have an impact on it. When the categories of two pixels are classified into one category and their pseudo labels are determined to be consistent, we believe that this information should be fed back to  $\mathbf{S}$  to update the similarity of the two

pixels, indicating that they are more closely related. The result is that the same substance in space tends to be grouped into the same class, which enhances the accuracy of classification by mining spatial information. However, if the similarity of two pixels is misjudged due to the influence of noise and continues to be iterative updated, the result will be that it is difficult to reclassify correctly.

To solve the above problem, we divide the whole spectral band into five groups to avoid the information redundancy caused by ignoring the strong correlation between adjacent bands and construct a local similarity matrix in each group, then construct a global similarity matrix  $\mathbf{S}$  through their linear combination. This method can find out more comprehensive similarity information in the spatial dimension and effectively select a more informative subset of bands.

Specifically, we divide all the bands into five groups, where  $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times mn}$  represents the data of group  $v$  and  $\sum_{v=1}^V d_v = r$ . Then the local similarity matrix formed by group  $v$  is  $\mathbf{S}^{(v)} \in \mathbb{R}^{mn \times mn}$ , where  $s_{ij}^{(v)}$  describes the similarity between the  $i$ -th pixel point and the  $j$ -th pixel point in group, that is, the similarity between  $x_i^{(v)}$  and  $x_j^{(v)}$ .  $s_{ij}^{(v)}$  is calculated as the following formula:

$$s_{ij}^{(v)} = \frac{\exp\left\{-\frac{\|x_i^{(v)} - x_j^{(v)}\|^2}{\sigma}\right\}}{\sum_{k=1}^{mn} \exp\left\{-\frac{\|x_i^{(v)} - x_k^{(v)}\|^2}{\sigma}\right\}}, \tag{6}$$

where  $s_{ij}^{(v)}$  represents the  $(i, j)$ -th element of  $\mathbf{S}^{(v)}$ ,  $\sigma$  is a hyperparameter.

Next,  $\mathbf{S}$  is constructed.  $\mathbf{S}_i$  in column  $i$  of it and should be a linear combination of  $(\mathbf{S}_i^{(1)} \ \mathbf{S}_i^{(2)} \ \dots \ \mathbf{S}_i^{(V)})$ . Then the global similarity matrix can be constructed as follows:

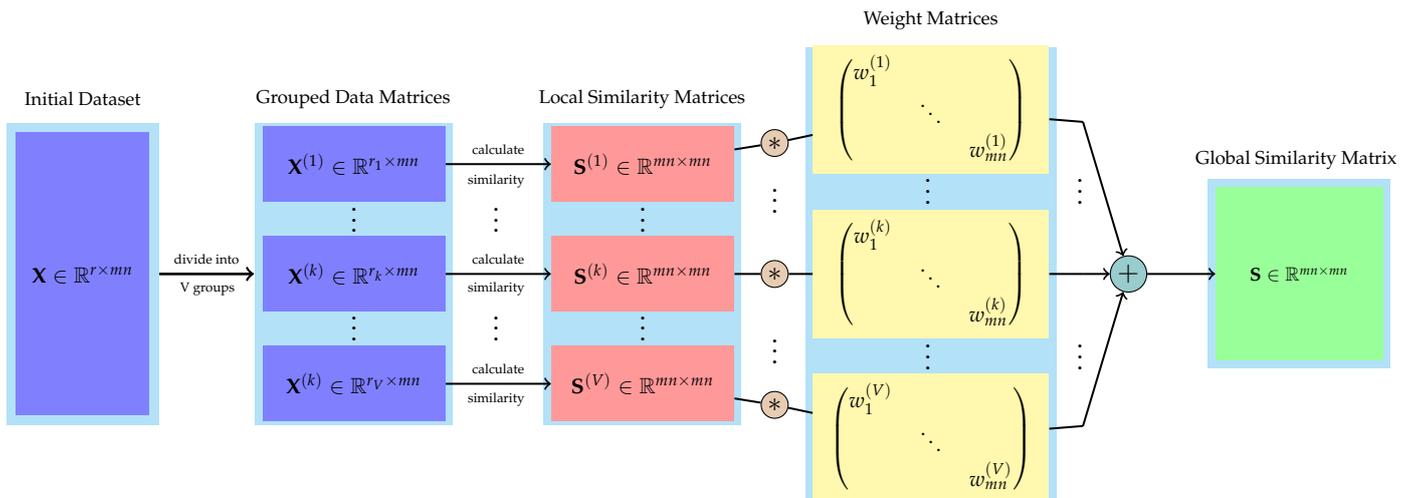
$$\min \sum_{i=1}^{mn} \left\| \mathbf{S}_i - \sum_{v=1}^V w_i^{(v)} \mathbf{S}_i^{(v)} \right\|_F^2, \text{ s.t. } \mathbf{S}_i^T \mathbf{1}_{mn} = 1, \mathbf{W}_i^T \mathbf{1}_V = 1, \tag{7}$$

$w_i^{(v)}$  represents the weight of column  $i$  of the local similarity matrix for the  $v$ -th spectral group. Moreover,  $\mathbf{w}_i = (w_i^{(1)} \ w_i^{(2)} \ \dots \ w_i^{(V)})^T \in \mathbb{R}^{V \times 1}$  represents the weight of all the local similarity matrices in  $i$ -th column, and the entire reconstructed weight matrix is  $\mathbf{W} = (\mathbf{W}_1 \ \mathbf{W}_2 \ \dots \ \mathbf{W}_{mn}) \in \mathbb{R}^{V \times mn}$ .

To sum up, according to Equations (1)–(7), the loss function is constructed as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{Z}, \mathbf{Y}_p, \mathbf{S}, \mathbf{W}) &= \left\| \mathbf{X}^T \mathbf{Z} - \mathbf{Y}_p \right\|_F^2 + \alpha \|\mathbf{Z}\|_{2,1} + \beta \text{Tr}(\mathbf{Y}_p^T \mathbf{L} \mathbf{Y}_p) \\ &+ \sum_{i=1}^{mn} \left\| \mathbf{S}_i - \sum_{v=1}^V w_i^{(v)} \mathbf{S}_i^{(v)} \right\|_F^2, \\ \text{s.t. } \mathbf{S} &\geq \mathbf{0}, \mathbf{S}_i^T \mathbf{1}_{mn} = 1, \mathbf{Y}_p^T \mathbf{Y}_p = 1, \mathbf{W}_i^T \mathbf{1}_V = 1, \end{aligned} \tag{8}$$

Hence, the model construction for band selection has been completed. In order to have a more intuitive understanding of our proposed method, we constructed a workflow, which is shown in Figure 1.



**Figure 1.** The workflow of the idea of band grouping of the global similarity matrix reconstructed by the local similarity matrix.

### 3.2. Model Optimization

In this paper, the strategy of iterative optimization is used to obtain the optimal values of the four parameters ( $\mathbf{W}$ ,  $\mathbf{S}$ ,  $\mathbf{Y}_p$ ,  $\mathbf{Z}$ ) involved in Equation (8). Specifically, it is to fix the values of three parameters to optimize the other parameter. We update  $\mathbf{W}$ ,  $\mathbf{S}$ ,  $\mathbf{Y}_p$ ,  $\mathbf{Z}$  in each iteration process and then calculate the loss function. The process will be ended in advance when the loss function converges.

#### 3.2.1. Fix $\mathbf{S}$ , $\mathbf{Z}$ , and $\mathbf{Y}_p$ : Update $\mathbf{W}$

Since  $\mathbf{S}$ ,  $\mathbf{Z}$ , and  $\mathbf{Y}_p$  are fixed, we only need to care about the expressions and conditions containing  $\mathbf{W}$  when solving the optimal value of the parameters by the gradient. Therefore, we should evaluate the following expression:

$$\mathcal{L}_1(\mathbf{W}) = \sum_{i=1}^{mn} \left\| \mathbf{S}_i - \sum_{v=1}^V \mathbf{W}_i^{(v)} \mathbf{S}_i^{(v)} \right\|_F^2, \text{ s.t. } \mathbf{W}_i^T \mathbf{1}_V = 1, \quad (9)$$

For this equation, using the augmented Lagrangian multiplier, it can be written as:

$$\begin{aligned} \mathcal{L}_1(\mathbf{W}) &= \sum_{i=1}^{mn} \left\| \mathbf{S}_i - \sum_{v=1}^V \mathbf{W}_i^{(v)} \mathbf{S}_i^{(v)} \right\|_F^2 + \varphi_i (1 - \mathbf{W}_i^T \mathbf{1}_V) \\ &= \sum_{i=1}^{mn} \left\{ \text{Tr} \left[ \left( \mathbf{A}^{(i)} \mathbf{W}_i \right)^T \left( \mathbf{A}^{(i)} \mathbf{W} \right) \right] + \varphi_i (1 - \mathbf{W}_i^T \mathbf{1}_V) \right\}, \end{aligned} \quad (10)$$

where  $\varphi = [\varphi_1, \varphi_2, \dots, \varphi_{mn}]^T$  is the Lagrangian multiplier vector. We define  $\mathbf{A}^{(i)} = [\mathbf{S}_i - \mathbf{S}_i^{(1)}, \mathbf{S}_i - \mathbf{S}_i^{(2)}, \dots, \mathbf{S}_i - \mathbf{S}_i^{(V)}] \in \mathbb{R}^{mn \times V}, i = 1, 2, \dots, mn$ . Next, we calculate the partial derivatives of  $\mathcal{L}_1(\mathbf{W})$ , with respect to  $\mathbf{W}_i$  and  $\varphi_i$ . Then we set the partial derivatives equal to 0 to get the following system of equations:

$$\begin{cases} \frac{\partial \mathcal{L}_1(\mathbf{W})}{\partial \mathbf{W}_i} = 2 \left( \mathbf{A}^{(i)} \right)^T \mathbf{A}^{(i)} \mathbf{W}_i - \varphi_i \mathbf{1}_V = \mathbf{0} \\ \frac{\partial \mathcal{L}_1(\mathbf{W})}{\partial \varphi_i} = 1 - \mathbf{W}_i^T \mathbf{1}_V = 0 \end{cases}$$

The results of solving this system of equations are as follows:

$$\mathbf{W}_i = \frac{\left[ (\mathbf{A}^{(i)})^T \mathbf{A}^{(i)} \right]^{-1} \mathbf{1}_V}{\mathbf{1}_V^T \left[ (\mathbf{A}^{(i)})^T \mathbf{A}^{(i)} \right]^{-1} \mathbf{1}_V}, \tag{11}$$

### 3.2.2. Fix $\mathbf{W}$ , $\mathbf{Z}$ , and $\mathbf{Y}_p$ : Update $\mathbf{S}$

When  $\mathbf{W}$ ,  $\mathbf{Z}$ , and  $\mathbf{Y}_p$  are fixed, only items related to  $\mathbf{S}$  have the effect on the optimization results, for others could be considered as constants. Therefore, at this moment, Equation (8) has the identical resolvent with the following equation:

$$\begin{aligned} \mathcal{L}_2(\mathbf{S}) &= \sum_{i=1}^{mn} \left\| \mathbf{S}_i - \sum_{v=1}^V w_i^{(v)} \mathbf{S}_i^{(v)} \right\|_F^2 + \beta \text{Tr}(\mathbf{Y}_p^T \mathbf{L} \mathbf{Y}_p) \\ &= \sum_{i=1}^{mn} \text{tr} \left[ \left( \mathbf{S}_i - \sum_{v=1}^V w_i^{(v)} \mathbf{S}_i^{(v)} \right)^T \left( \mathbf{S}_i - \sum_{v=1}^V w_i^{(v)} \mathbf{S}_i^{(v)} \right) \right] + \frac{\beta}{2} \sum_{i=1}^{mn} \sum_{j=1}^{mn} \left\| \mathbf{Y}_{pi}^T - \mathbf{Y}_{pj}^T \right\|_F^2 s_{ij} \\ &= \sum_{i=1}^{mn} \text{tr} \left[ \mathbf{S}_i^T \mathbf{S}_i - 2b_i^T \mathbf{S}_i + b_i^T b_i \right] + a_i^T \mathbf{S}_i \\ &\text{s.t. } s_{ij} \geq 0, \mathbf{S}_i^T \mathbf{1}_{mn} = 1 \end{aligned} \tag{12}$$

where  $\mathbf{Y}_{pi}^T$  means the  $i$ -th column of transpose of  $\mathbf{Y}_p$ ,

$$a_i = \frac{\beta}{2} \left[ \left\| \mathbf{Y}_{pi}^T - \mathbf{Y}_{p1}^T \right\|_F^2, \left\| \mathbf{Y}_{pi}^T - \mathbf{Y}_{p2}^T \right\|_F^2, \dots, \left\| \mathbf{Y}_{pi}^T - \mathbf{Y}_{pmn}^T \right\|_F^2 \right]^T \in \mathbb{R}^{mn \times 1}$$

and  $b_i = \sum_{v=1}^V w_i^{(v)} \mathbf{S}_i^{(v)} \in \mathbb{R}^{mn \times 1}$ . We use Lagrange multipliers to solve the problem of (12) as follows:

$$\mathcal{L}_2(\mathbf{S}, \lambda, \mathbf{\Pi}) = \sum_{i=1}^{mn} \text{tr} \left[ \mathbf{S}_i^T \mathbf{S}_i - 2b_i^T \mathbf{S}_i + b_i^T b_i \right] + a_i^T \mathbf{S}_i + \lambda (1 - \mathbf{S}_i^T \mathbf{1}_{mn}) - \mathbf{\Pi}^T \mathbf{S}_i \tag{13}$$

where  $\lambda$  and  $\mathbf{\Pi} = [\Pi_1, \Pi_2, \dots, \Pi_{mn}]^T$  is the Lagrangian multiplier. In order to solve the above inequalities and equations, we need to satisfy the KKT condition while using the Lagrange multiplier:

$$\begin{cases} \frac{\partial \mathcal{L}_2}{\partial \mathbf{S}_i} = 2\mathbf{S}_i - 2b_i + a_i - \lambda \mathbf{1}_{mn} - \mathbf{\Pi} = 0, \\ \mathbf{\Pi}_j \geq 0, j = 1, 2, \dots, mn, \\ \mathbf{\Pi}_j s_{ij} = 0, j = 1, 2, \dots, mn, \\ \frac{\partial \mathcal{L}_2}{\partial \lambda} = 1 - \mathbf{S}_i^T \mathbf{1}_{mn} = 0. \end{cases}$$

By simplifying the above formula, the updated formula can be written as follows:

$$\begin{cases} \bar{\mathbf{\Pi}} = \frac{1}{mn} \sum_{j=1}^{mn} \mathbf{\Pi}_j \\ c_i = 2b_i - a_i \\ s_{ij} = \max\left(\left(\frac{c_i}{2} \left(I_{mn} - \frac{1_{mn \times mn}}{mn}\right) + \frac{1_{mn}}{mn} - \frac{1}{2} \bar{\mathbf{\Pi}}_i \mathbf{1}_{mn}\right), 0\right) \\ \mathbf{\Pi}_{ij} = \max\left(-2\left(\frac{c_i}{2} \left(I_{mn} - \frac{1_{mn \times mn}}{mn}\right) + \frac{1_{mn}}{mn} - \frac{1}{2} \bar{\mathbf{\Pi}}_i \mathbf{1}_{mn}\right), 0\right) \end{cases} \tag{14}$$

### 3.2.3. Fix $\mathbf{W}$ , $\mathbf{S}$ , and $\mathbf{Z}$ : Update $\mathbf{Y}_p$

When  $\mathbf{W}$ ,  $\mathbf{S}$ , and  $\mathbf{Z}$  are fixed, the terms that are not related to  $\mathbf{Y}_p$  can be regarded as constant terms. Consequently, Equation (8) is equivalent to:

$$\begin{aligned} \mathcal{L}_3(\mathbf{Y}_p) &= \|\mathbf{X}^T\mathbf{Z} - \mathbf{Y}_p\|_F^2 + \beta\text{Tr}(\mathbf{Y}_p^T\mathbf{L}\mathbf{Y}_p) \\ &= \text{Tr}(\mathbf{X}^T\mathbf{Z} - \mathbf{Y}_p)^T(\mathbf{X}^T\mathbf{Z} - \mathbf{Y}_p) + \beta\text{Tr}(\mathbf{Y}_p^T\mathbf{L}\mathbf{Y}_p) \\ &= \text{Tr}(\mathbf{Y}_p^T(\mathbf{I} + \beta\mathbf{L})\mathbf{Y}_p) - 2\text{Tr}(\mathbf{Y}_p^T\mathbf{X}^T\mathbf{Z}) \\ &= \text{Tr}(\mathbf{Y}_p^T(\mathbf{A}\mathbf{Y}_p - 2\mathbf{X}^T\mathbf{Z})) \text{ s.t. } \mathbf{Y}_p^T\mathbf{Y}_p = \mathbf{I}_c \end{aligned} \tag{15}$$

where  $\mathbf{A} = \mathbf{I} + \beta\mathbf{L} \in \mathbb{R}^{mn \times mn}$  is the symmetric matrix. This is an optimization problem constrained by an orthonormal matrix, the orthogonal Procrustes problem. Here we take the generalized power iteration method (GPI) [42] to solve the quadratic problem on the Stiefel manifold (QPSM), instead of the Lagrange multipliers to solve this problem. So Equation (15) can be further relaxed to:

$$\arg \max \text{Tr}(\mathbf{Y}_p^T(\gamma_A\mathbf{I} - \mathbf{A})\mathbf{Y}_p + 2\mathbf{X}^T\mathbf{Z}), \text{ s.t. } \mathbf{Y}_p^T\mathbf{Y}_p = \mathbf{I}_c \tag{16}$$

where  $\gamma_A$  is the maximum singular value of the matrix  $\mathbf{A}$  and it can be calculated by the singular value decomposition (SVD) method. Then further simplify the above formula:

$$\arg \max \text{Tr}(\mathbf{F}^T\mathbf{M}), \text{ s.t. } \mathbf{F}^T\mathbf{F} = \mathbf{I}_c \tag{17}$$

where  $\mathbf{F} = \mathbf{Y}_p \in \mathbb{R}^{mn \times c}$  and  $\mathbf{M} = (\gamma_A\mathbf{I} - \mathbf{A})\mathbf{Y}_p + 2\mathbf{X}^T\mathbf{Z} \in \mathbb{R}^{mn \times c}$ . We perform singular value decomposition for  $\mathbf{M}$ , assume that  $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  and  $\mathbf{G} = \mathbf{V}^T\mathbf{F}^T\mathbf{U}$ ,  $\mathbf{G}\mathbf{G}^T = \mathbf{I}_c$ . The solution steps are as follows:

$$\begin{aligned} \text{Tr}(\mathbf{F}^T\mathbf{M}) &= \text{Tr}(\mathbf{\Sigma}\mathbf{V}^T\mathbf{F}^T\mathbf{U}) = \text{Tr}(\mathbf{\Sigma}\mathbf{G}) \\ &= \sigma_{11}g_{11} + \sigma_{22}g_{22} + \dots + \sigma_{cc}g_{cc} \\ &= \sum_{i=1}^c \sigma_{ii}g_{ii} \leq \sum_{i=1}^c \sigma_{ii} \end{aligned} \tag{18}$$

If and only if all  $g_{ii} = 1$ , Equation (18) takes the equal sign, this moment  $\mathbf{G} = [\mathbf{I}_c | \mathbf{0}]$ . Ultimately,  $\mathbf{Y}_p$  can be indicated by

$$\mathbf{Y}_p = \mathbf{U}[\mathbf{I}_c | \mathbf{0}]^T \mathbf{V}^T. \tag{19}$$

### 3.2.4. Fix $\mathbf{W}$ , $\mathbf{S}$ , and $\mathbf{Y}_p$ : Update $\mathbf{Z}$

When fixing  $\mathbf{W}$ ,  $\mathbf{S}$ , and  $\mathbf{Y}_p$  and removing items that do not contain  $\mathbf{Z}$ , Equation (8) is equivalent to the following formula:

$$\begin{aligned} \mathcal{L}_4(\mathbf{Z}) &= \|\mathbf{X}^T\mathbf{Z} - \mathbf{Y}_p\|_F^2 + \alpha\|\mathbf{Z}\|_{2,1} \\ &= \text{Tr}[(\mathbf{X}^T\mathbf{Z} - \mathbf{Y}_p)^T(\mathbf{X}^T\mathbf{Z} - \mathbf{Y}_p)] + \alpha\|\mathbf{Z}\|_{2,1} \end{aligned} \tag{20}$$

Set the partial derivative of  $\mathcal{L}_4(\mathbf{Z})$  over  $\mathbf{Z}$  to 0, then we gain:

$$\frac{\partial \mathcal{L}_4(\mathbf{Z})}{\partial \mathbf{Z}} = 2(\mathbf{X}\mathbf{X}^T\mathbf{Z} - \mathbf{X}\mathbf{Y}_p + \alpha\mathbf{\Lambda}\mathbf{Z}) \tag{21}$$

where  $\mathbf{\Lambda}$  is a diagonal matrix and its diagonal elements can be computed as follows:

$$\Lambda_{ii} = \frac{1}{2\sqrt{\sum_{j=1}^c z_{ij}^2}}, (i = 1, 2, \dots, r), \tag{22}$$

To ensure that the denominator meets the definition, a small enough positive constant  $\Delta$  is added and  $\Lambda$  is transformed into  $\Lambda = \text{diag}(\Lambda'_{11}, \Lambda'_{22}, \dots, \Lambda'_{dd})$  that the diagonal element can be calculated by:

$$\Lambda'_{ii} = \frac{1}{2\sqrt{\sum_{j=1}^c z_{ij}^2 + \Delta}}, (i = 1, 2, \dots, r), \quad (23)$$

When replacing  $\Lambda'$  with  $\Lambda$ , Equation (21) can be rewritten as follows:

$$\frac{\partial \mathcal{L}_4(\mathbf{Z})}{\partial \mathbf{Z}} = 2(\mathbf{X}\mathbf{X}^T\mathbf{Z} - \mathbf{X}\mathbf{Y}_p + \alpha\Lambda'\mathbf{Z}) = 0, \quad (24)$$

According to Equation (24),  $\mathbf{Z}$  can be expressed as follows:

$$\mathbf{Z} = (\mathbf{X}\mathbf{X}^T + \alpha\Lambda')^{-1}\mathbf{X}\mathbf{Y}_p, \quad (25)$$

Owing to  $\Lambda'$  being bound up with  $\mathbf{Z}$ , Equation (25) is not an analytical solution for  $\mathbf{Z}$ . As a result, we need to utilize an iterative update algorithm to solve optimal  $\Lambda'$  and  $\mathbf{Z}$ . That means that after each round of updating  $\mathbf{W}$ ,  $\mathbf{S}$ ,  $\mathbf{Y}_p$ , and  $\mathbf{Z}$ , we will also update  $\Lambda'$  according to Equation (23). The overall algorithm optimization is shown in Algorithm 1.

---

**Algorithm 1:** Alternative iterative algorithm to solve Equation (8).

---

**Input:** The data matrix  $\mathbf{X} \in \mathbb{R}^{r \times mn}$ ,  $\mathbf{X}^{(v)} \in \mathbb{R}^{r_v \times mn}$ , and the hyperparameters  $\alpha$ .

**Output:**  $K$  selected bands.

```

1 while not convergence do
2   a).Update  $\mathbf{W}$ :
3   for  $i = 1; i \leq mn$  do
4      $\mathbf{A}^{(i)} = [\mathbf{S}_i - \mathbf{S}_i^{(1)}, \dots, \mathbf{S}_i - \mathbf{S}_i^{(V)}]$ ;
5      $\mathbf{W}_i = \frac{[(\mathbf{A}^{(i)})^T \mathbf{A}^{(i)}]^{-1} \mathbf{1}_V}{\mathbf{1}_V^T [(\mathbf{A}^{(i)})^T \mathbf{A}^{(i)}]^{-1} \mathbf{1}_V}$ ;
6   b).Update  $\mathbf{Z}$ :
7   while not convergence do
8      $\mathbf{Z} \leftarrow (\mathbf{X}\mathbf{X}^T + \alpha\Lambda')^{-1}\mathbf{X}\mathbf{Y}_p$ ;
9      $\Lambda'_{ii} \leftarrow \frac{1}{2\sqrt{\sum_{j=1}^c z_{ij}^2 + \Delta}}, (i = 1, 2, \dots, r)$ ;
10  c).Update  $\mathbf{S}$ :
11  for  $i = 1; i \leq mn$  do
12     $a_i = \frac{\beta}{2} [ \|\mathbf{Y}_p^i - \mathbf{Y}_p^1\|_F^2 \dots \|\mathbf{Y}_p^i - \mathbf{Y}_p^{mn}\|_F^2 ]^T$ ;
13     $\mathbf{C}_i = 2 * [\mathbf{S}_i^{(1)} \mathbf{S}_i^{(2)} \dots \mathbf{S}_i^{(V)}] \mathbf{W}_i - a_i$ ;
14    for  $j = 1; j \leq mn$  do
15       $\phi = \frac{1}{2} (\mathbf{I}_{mn} - \frac{\mathbf{1}_{mn} \mathbf{1}_{mn}^T}{mn}) \mathbf{C}_i + \frac{\mathbf{1}_{mn}}{mn} - \frac{\bar{\pi}_i}{2} \mathbf{1}_{mn}$ ;
16       $\mathbf{S}_{ij} \leftarrow \max(\phi, 0)$ ;
17       $\pi_{ij} \leftarrow \max(\phi, 0)$ ;
18  d).Update  $\mathbf{Y}_p$ :
19   $\mathbf{U}\Sigma\mathbf{V}^T = \text{SVD}((\gamma_A \mathbf{I} - \mathbf{A})\mathbf{Y}_p + 2\mathbf{X}^T\mathbf{Z})$ ;
20   $\mathbf{Y}_p = \mathbf{U}\mathbf{I}_c^T \mathbf{V}^T$ 

```

---

## 4. Experiments

### 4.1. Dataset Descriptions

The proposed method is tested in the experiments on five publicly available benchmark datasets. These publicly available datasets are presented below:

#### 4.1.1. ROSIS Pavia University Image

The Pavia University (PaviaU) scene has a spatial resolution of  $640 \times 340$  pixels and 115 bands, which covers a spectral range from 430 to 860 nm and contains nine classes. After removing 12 noisy bands, the remaining 103 bands are utilized for our experiment. It was acquired by the Reflective Optics System Imaging Spectrometer from the University of Pavia.

#### 4.1.2. AVIRIS Indian Pines Image

Another dataset, Indian Pines (IndianP), has a size of  $145 \times 145$  pixels. It has 20 m spatial resolutions and 10 nm spectral resolutions covering a spectrum range of 200–2400 nm. After removing 20 water absorption bands, the remaining 200 bands have 13 different classes of ground truth. The data set was acquired by NASA in 1992 using the AVIRIS sensor from JPL.

#### 4.1.3. AVIRIS Salinas Scene

The Salinas scene is characterized by high spatial resolution (3.7-meter pixels), which is collected by the 224-band AVIRIS sensor over Salinas Valley, California. After removing water absorption bands (i.e., bands 108–112, 154–167, and 224) are removed. The remaining 204 bands are utilized for our experiment. The Salinas ground truth contains 16 classes and comprises  $512 \times 127$  pixels.

#### 4.1.4. Botswana Image

The image size of the dataset is  $1476 \times 256$ . Pre-processing of the data was performed by the UT Center for Space Research to mitigate the effects of bad detectors, inter-detector miscalibration, and intermittent anomalies. Water absorption bands and noise-affected bands are discarded, and the remaining 145 bands have 14 identified classes. These datasets are all public available ([https://www.ehu.eus/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes)).

#### 4.1.5. University of Houston

The data were acquired by the NSF-funded Center for Airborne Laser Mapping (NCALM) over the University of Houston campus and the neighboring urban area, which was provided by the 2018 IEEE GRSS Data Fusion Contest. The data size is  $601 \times 2384$ , including 144 bands with a spectral range from 380 to 1050 nm. The image contains 20 kinds of different feature information ([https://hyperspectral.ee.uh.edu/?page\\_id=1075](https://hyperspectral.ee.uh.edu/?page_id=1075)).

### 4.2. Methods Taken for Comparison

We accessed the classification performance of our algorithm by comparing it with nine popular band selection techniques, including NC-OC-IE [43], NC-OC-MVPCA [44], TRC-OC-FDPC [24], UBS [45], ONR [46], LvaHAI [47], SOR-SRL [48], PCA (principal components analysis), and PCAS (principal components analysis based on manifold structure) [49]. Our band selection strategy is denoted as the BAMGC method.

Normalized cut-based optimal clustering (NC-OC) is a group-wise selection method. It searches for optimal clustering results through an optimization method of dynamic programming and proposes a ranking-based strategy to select representative bands in each group. We combine NC with information entropy (IE) and the maximum-variance criterion (MVPCA) as criteria. Fast density-peak-based clustering (TRC-OC-FDPC) employs a ranking-based as well as a clustering-based scheme, which can automatically identify the cluster centers and reduce redundancy. The band subset usually has a certain uniformity in earlier studies, and the index of the selected band should have a certain degree of uniform distribution. Therefore, UBS simply selects the band uniformly. In optimal neighborhood reconstruction (ONR), band selection is considered a combinatorial optimization problem. It chooses a better band combination by evaluating its performance to reconstruct the original data and applies a noise reducer to minimize the influence of noisy bands. In the local-view-assisted discriminative band selection method with hypergraph autolearning (LvaHAI), the whole band space is first randomly divided into several subspaces (LVs) of different dimensions. For different LVs, a robust hinge loss function for isolated pixels regularized by the row-sparsity is adopted to measure the importance of the corresponding bands. Scalable one-pass self-representation learning (SOP-SRL) is a ranking-based scheme that is proposed to address this problem by processing data in a streaming fashion without storing the entire data ranking-based. PCA is a widely used dimension reduction method.

Its main idea is to project the  $n$ -dimensional features to the low-dimensional space to make new orthogonal features. In addition, based on PCA, PCAS is an improved method by adding manifold regular term constraints, which can preserve spatial information to a greater extent.

#### 4.3. Experimental Setting

Considering the complexity and efficiency of the algorithm, we first select a  $20 \times 20 \times r$  sub-block for preliminary experiments, corresponding to the label matrix size of  $m \times 1$ . The BAMGC algorithm is used to process the extracted  $400 \times r$  sub-blocks to get the rank of each band. In this experiment, we use support vector machine (SVM) and k-nearest neighbor (KNN) classifiers. In order to fairly compare the performance of various algorithms, the band subsets obtained by different algorithms are processed by the classifier with the same parameters, and the performance of the algorithm is evaluated by the same criteria. The kernel function of SVM uses the radial basis function (RBF), and the parameter settings for different datasets are shown in Table 2. The parameter  $K$  employed by KNN is set to 3 on all datasets (the code of this article can be found on <https://github.com/misteru/BAMGC>). Meanwhile, the accuracy (ACC), overall accuracy (OA), and Kappa coefficient ( $\kappa$ ) are employed to demonstrate the performance of different methods, which are obtained by nesting loops for all parameters. The larger values of the three indexes represent a better effect. Since the optimal number of bands suitable for different datasets is unpredictable, we evaluate the performance of algorithms through traversal search. Specifically, it is to test the effect of different algorithms when selecting the same number of bands on the same dataset. Furthermore, in order to minimize the number of result subsets, we finally decide to select the number of bands to test the performance of these algorithms within the range of  $\{3, 6, 9, 12, 15, 18, 21, 24, 27, 30\}$ . Thus, 10 trials were conducted with different training samples for reducing the random effect of the results.

**Table 2.** The table illustrates the SVM parameter settings for each dataset.

Parameter	Pavia University	Indian Pines	Salinas	Botswana	University of Houston
C	10,000.0	100.0	100.0	10,000.0	10,000.0
gamma	0.5	4.0	16.0	0.5	0.5

#### 4.4. Result Analysis

Next, we analyze the results of all 10 algorithms on 5 different datasets separately.

##### 4.4.1. Experimental Results on Pavia University dataset

It can be seen from Figure 2 that the overall effect of the BAMGC method is relatively consistent. For the SVM, BAMGC has a similar effect to NC-OC-MVPCA when the number of bands is greater than 12, but before that, its effect is relatively better than others. Compared with LvaHAI and SOR-SRL methods based on sorting strategy, BAMGC shows better performance, as well as other methods. For the KNN, the OA and  $\kappa$  obtained with BAMGC increased steadily by increasing the number of selected bands. When the number of bands is less than 18, its effect is superior to other algorithms, which is similar to ONR when the number of bands is greater than 18. Although the ONR and LvaHAI occasionally achieved a higher classification accuracy than the proposed techniques, their performances are not as consistent as that of BAMGC. Overall, BAMGC shows higher performance on this dataset, and the visualization of classification is shown in Figure 3. There are ten algorithms in total, and the visualized images from (b) to (h) are sorted according to the algorithms selected by increasing OA. Here, we show the seven best algorithms.

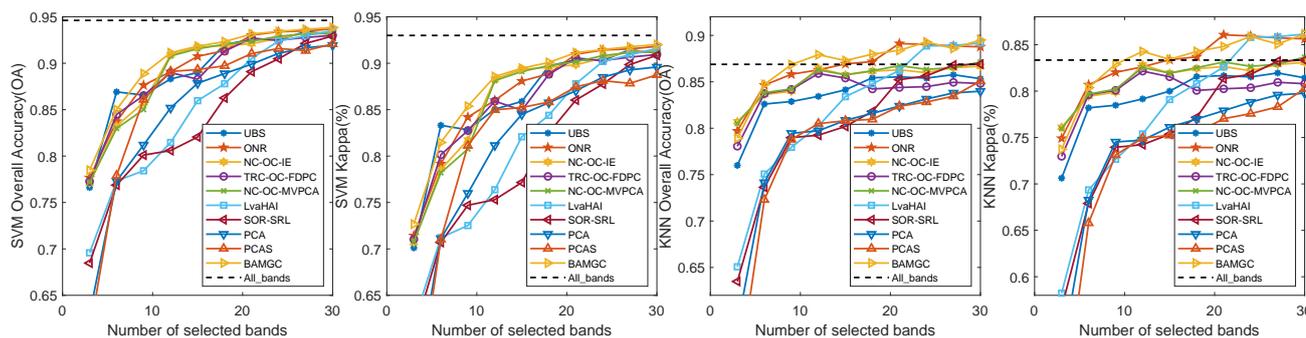


Figure 2. The comparison of OA and  $\kappa$  produced by SVM and KNN on the Pavia University dataset.

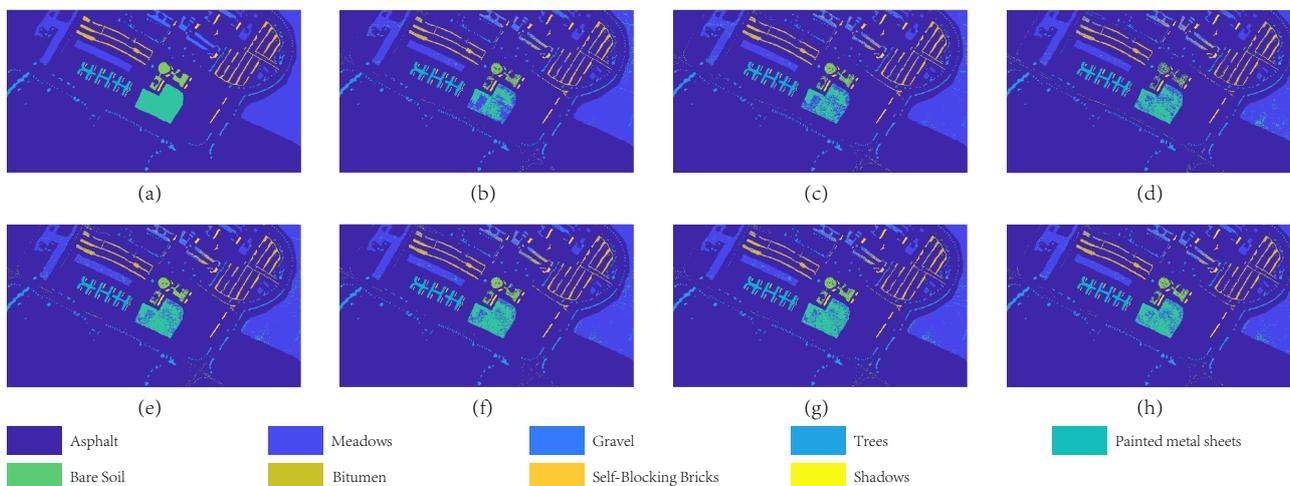


Figure 3. The visualization of classification on the Pavia University dataset. (a) Ground truth; (b) TRC-OC-FDPC; (c) UBS; (d) PCAS; (e) ONR; (f) NC-OC-MVPCA; (g) NC-OC-IE; (h) BAMGC.

#### 4.4.2. Experimental Results on Indian Pines

Figure 4 shows the result of each algorithm on the Indian Pines dataset. It can be found that the NC-OC-MVPCA algorithm has an advantage when the number of bands is 3, 6, and 9. When the number of bands is greater than 12, the performance of BAMGC exceeds that of the other algorithms, and the classification effect is outstanding when 21 and 27 bands are selected. Especially in the case of KNN, BAMGC obtained an OA of 71.61% with 27 bands, followed by SOR-SRL (68.26%). In the case of computing all bands, the OA reaches 66.10%, indicating that BAMGC uses fewer bands to express clearer information than the full band, reflecting its superior performance. The visualization of the classification of the Indian Pines dataset is shown in Figure 5.

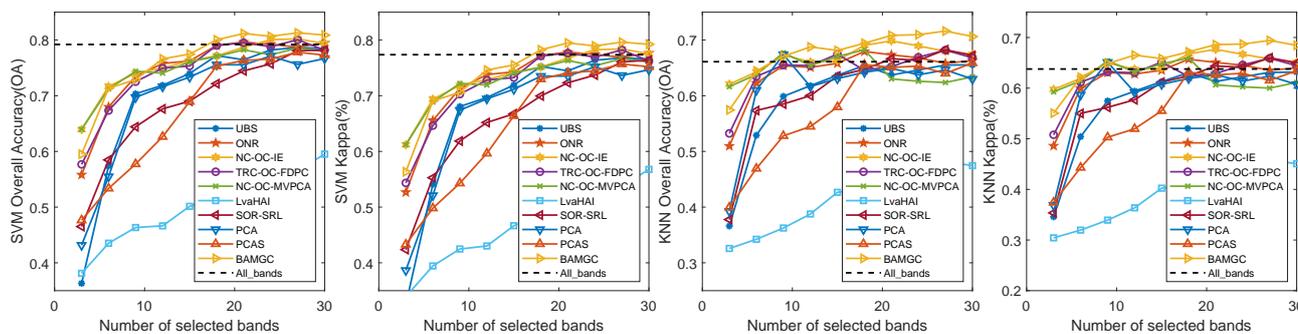
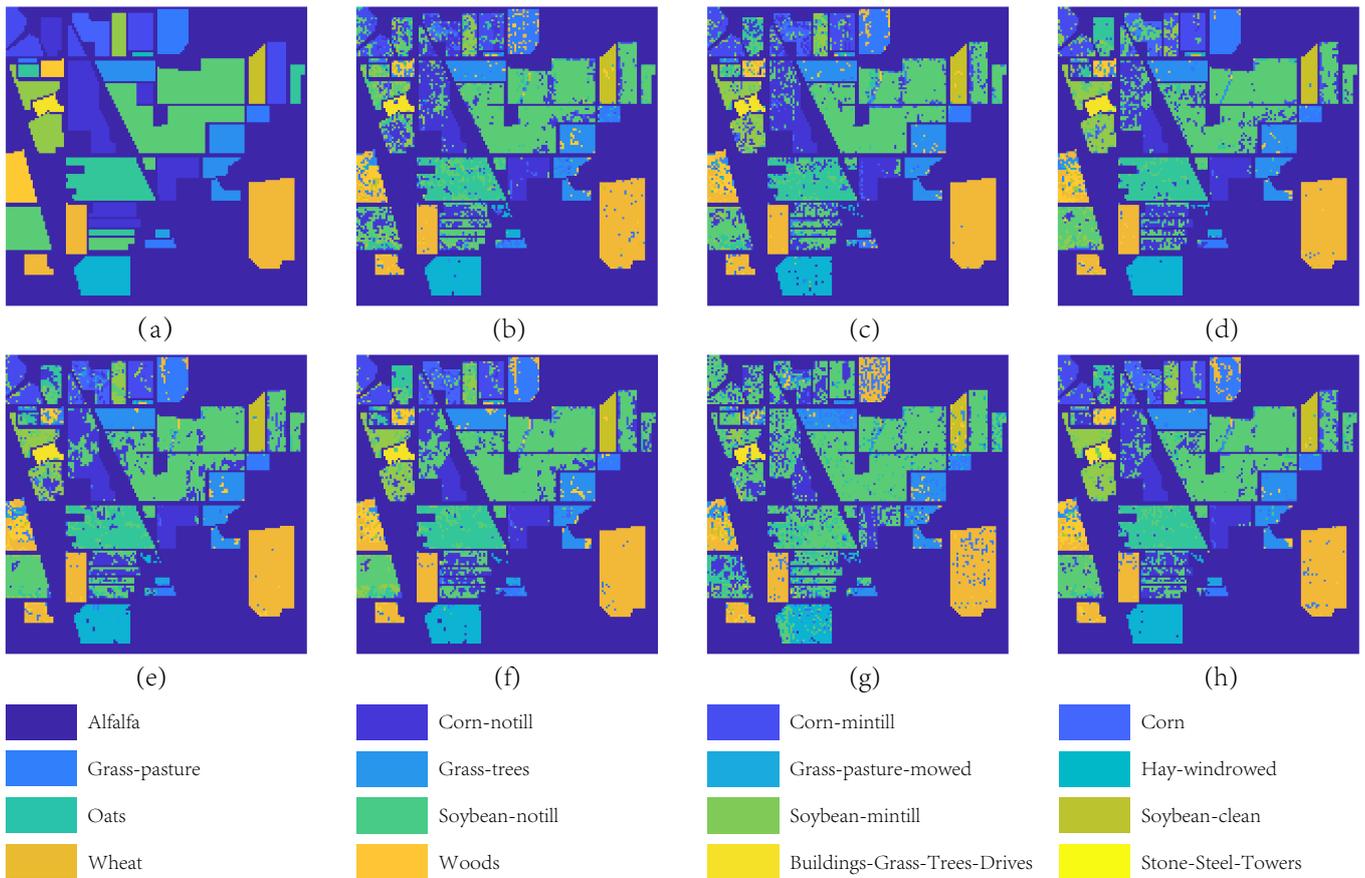


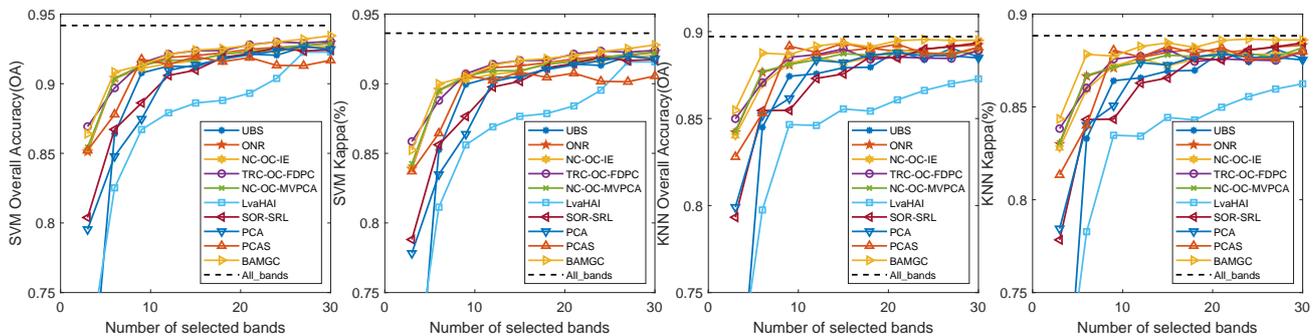
Figure 4. The comparison of OA and  $\kappa$  produced by SVM and KNN on the Indian Pines dataset.



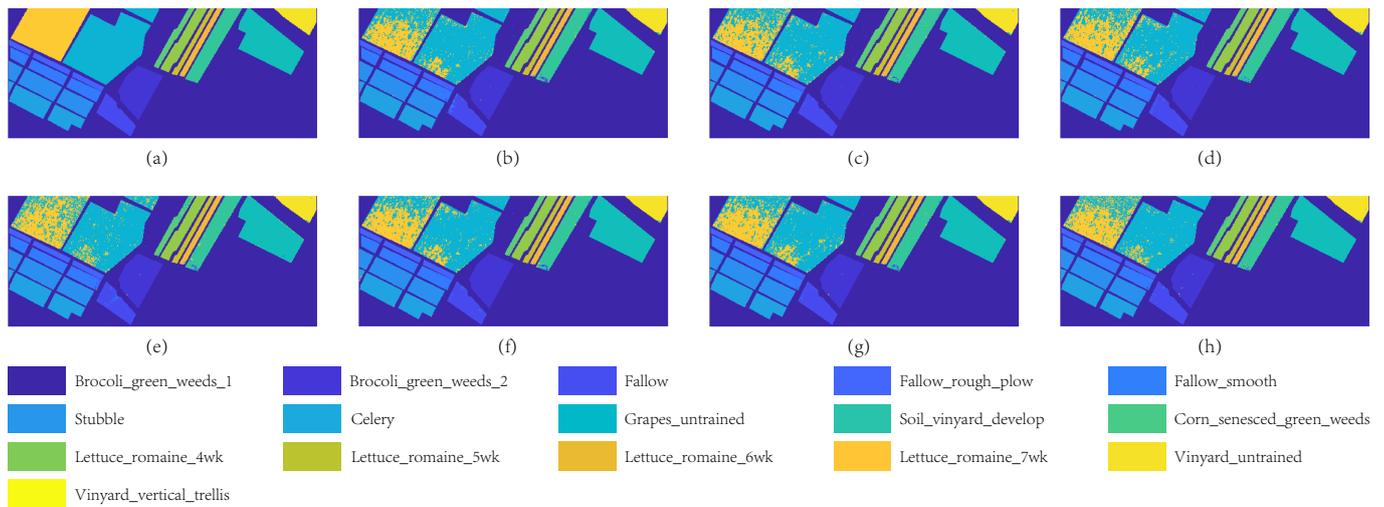
**Figure 5.** The visualization of classification on the Indian Pine dataset. (a) Ground truth; (b) PCAS; (c) PCA; (d) ONR; (e) NC-OC-MVPA; (f) NC-OC-IE; (g) LvaHAI; (h) BAMGC.

4.4.3. Experimental Results on Salinas

Due to the advantages of SVM in dealing with small samples and high-dimensional data classification, ONR, NC-OC-IE, TRC-OC-FDPC, NC-OC-MVPCA, and BAMGC have achieved similar results (Figure 6). Nevertheless, under the condition of the KNN that has weak classification ability, BAMGC has the best result, followed by ONR, NC-OC-IE, and NC-OC-MVPCA. As shown in Figure 6, the OA and  $\kappa$  obtained with BAMGC firstly increased sharply by increasing the number of selected bands from 3 to 6. Then, it slowly grows when the number of selected bands is larger than 15. This also illustrates that BAMGC has higher accuracy on different classifiers of various datasets. The visualization of the effect is shown in Figure 7.



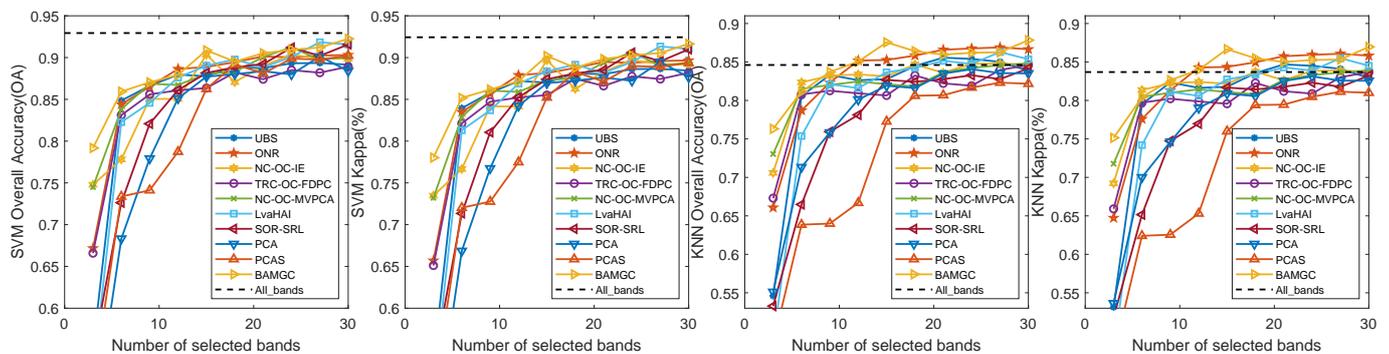
**Figure 6.** The comparison of OA and  $\kappa$  produced by SVM and KNN on the Salinas dataset.



**Figure 7.** The visualization of classification on the Salinas dataset. (a) Ground truth; (b) PCA; (c) NC-OC-IE; (d) NC-OC-MVPCA; (e) PCAS; (f) ONR; (g) TRC-OC-FDPC; (h) BAMGC.

#### 4.4.4. Experimental Results on Botswana Image

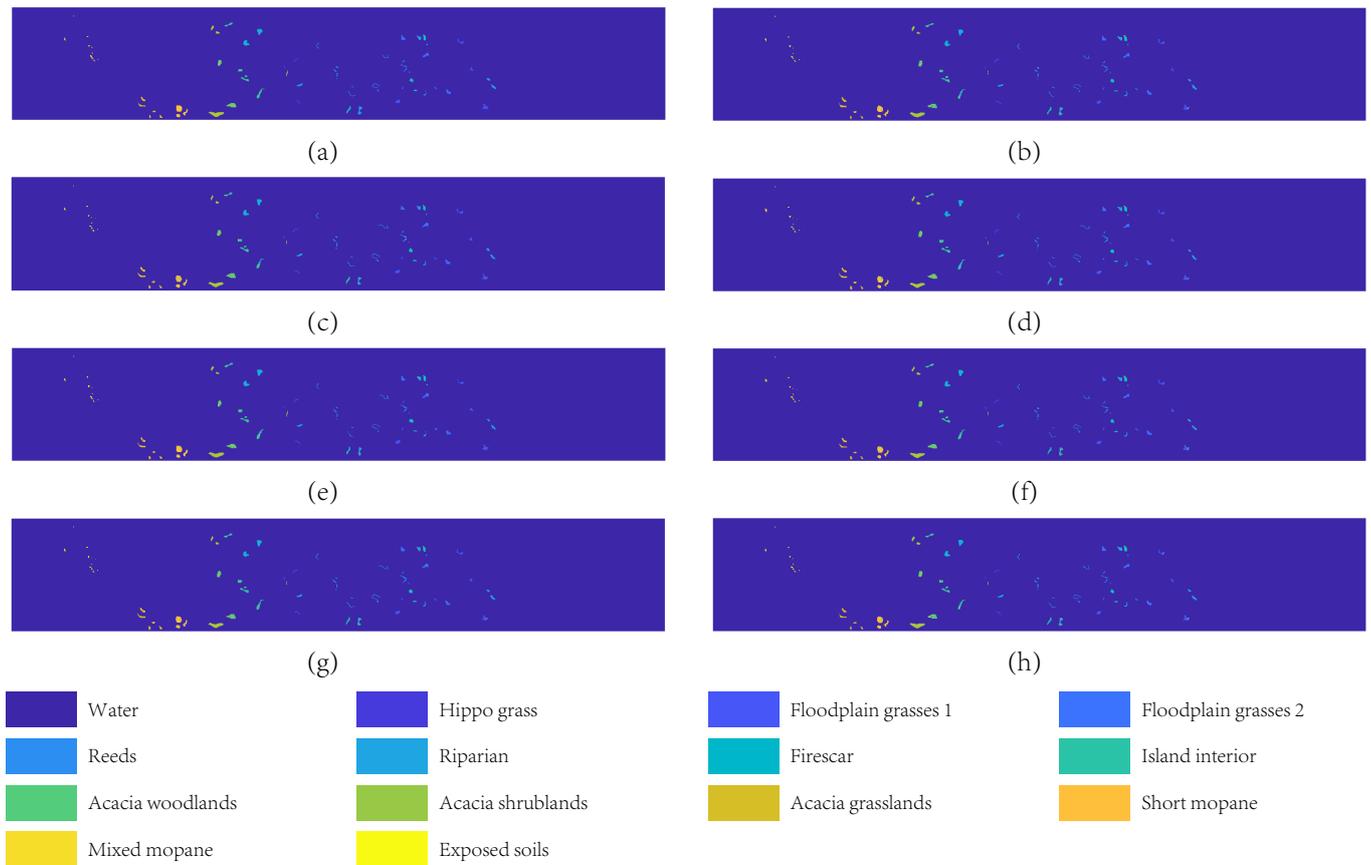
In Figure 8, it can be seen that BAMGC has an advantage when using the SVM and KNN with the number of bands ranging from 3 to 15. In particular, our algorithm performs better than others by selecting 15 bands, achieving 87.55% accuracy, especially in the KNN classifier. However, when the band number is greater than 15, its OA and  $\kappa$  decrease slightly and show similar results with ONR. Additionally, BAMGC remains relatively consistent with small fluctuations by increasing the number of selected bands, especially when the number is greater than 21. In general, the effect of BAMGC on the Botswana dataset is superior to the other nine comparison algorithms. The visualization of the effect is shown in Figure 9.



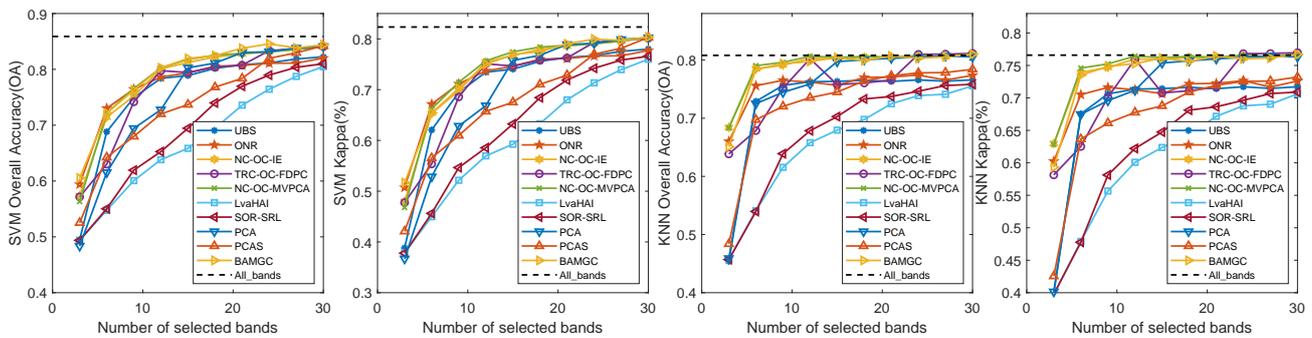
**Figure 8.** The comparison of OA and  $\kappa$  produced by SVM and KNN on the Botswana dataset.

#### 4.4.5. Experimental Results on University of Houston

Figure 10 records the statistical results of the ten algorithms on the University of Houston dataset. It can be seen from the figure that BAMGC, NC-OC-IE, and NC-OC-MVPCA have similar effects in the SVM classifier, especially when the number of selected bands exceeds 21. Additionally, the classification effect of these three methods is obviously better than other algorithms when the number of bands is small. When the KNN classifier is used, BAMGC, NC-OC-IE, and NC-OC-MVPCA show satisfactory results for various numbers of bands. It is worth noting that the PCA method has achieved good accuracy in 12 bands, but it fluctuates with the increase in bands and the effect is unstable. Thus, when the number of bands is more than 24, the classification accuracy of TRC-OC-FDPC and BAMGC is relatively close. The visualization of the effect is shown in Figure 11, which has been stretched to make it clearer.



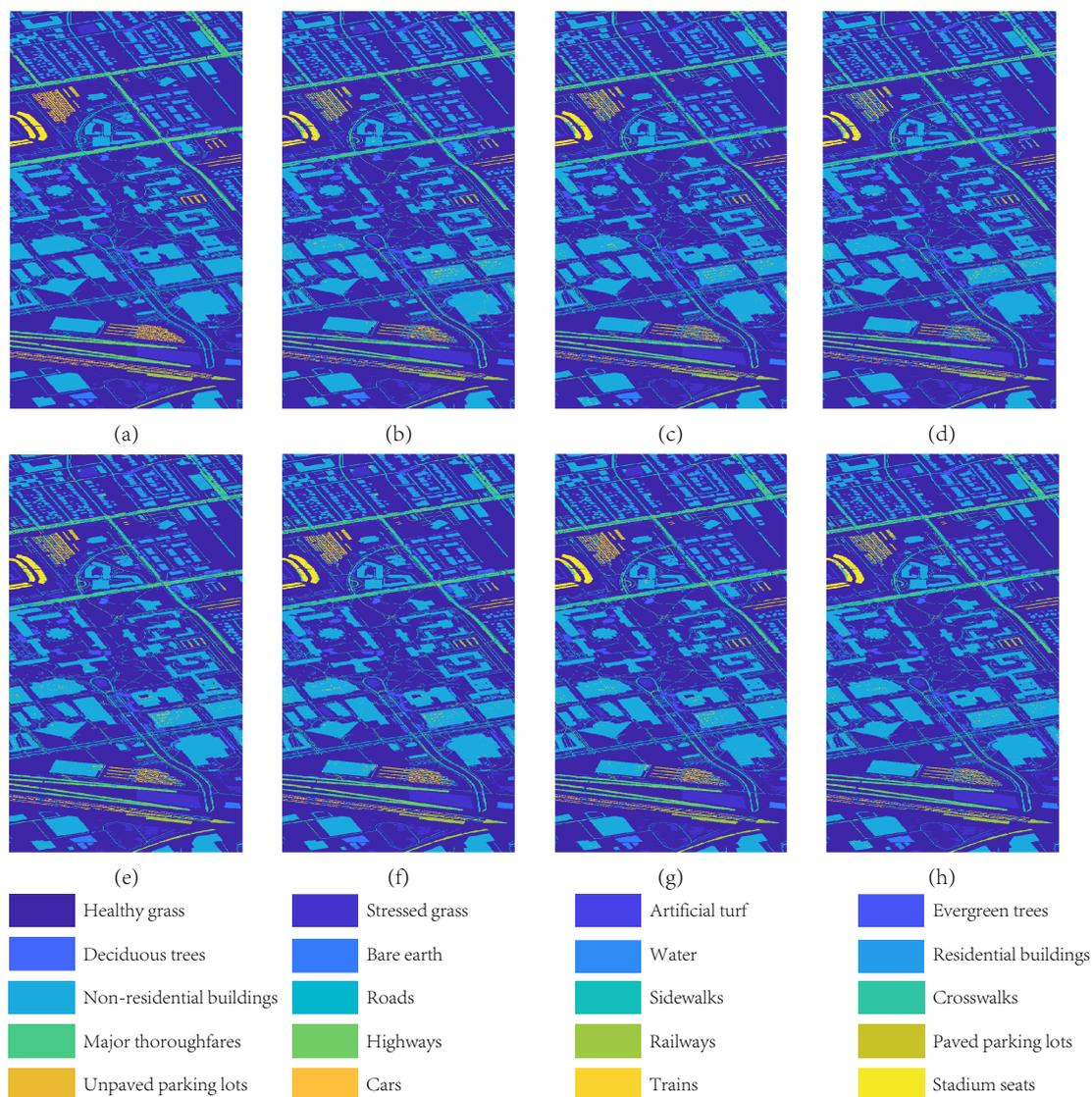
**Figure 9.** The visualization of classification on the Botswana dataset. (a) Ground truth; (b) UBS; (c) NC-OC-MVPCA; (d) SOR-SRL; (e) ONR; (f) LvaHAI; (g) SORSRL; (h) BAMGC.



**Figure 10.** The comparison of OA and  $\kappa$  produced by SVM and KNN on the University of Houston dataset.

4.5. Experimental Result Summary

The experimental result performance of 10 algorithms is analyzed on 5 datasets, respectively. For all of the data sets with 15 bands selected, BAMGC provided the highest values whether classified by SVM or KNN. We have marked these values in bold in Table 3. This demonstrates that our proposed BAMGC method has consistent performance using the different classifiers.



**Figure 11.** The visualization of classification on the University of Houston dataset. (a) Ground truth; (b) UBS; (c) ONR; (d) TRC-OC-FDPC; (e) PCA; (f) NC-OC-IE; (g) NC-OC-MVPCA; (h) TRC-OC-FDPC.

**Table 3.** Results of each algorithm in different datasets with 15 bands selected.

Dataset	Method	SVM		KNN	
		OA	$\kappa$	OA	$\kappa$
Pavia University	UBS	89.00	85.89	84.14	80.00
	ONR	90.75	88.08	86.93	83.38
	NC-OC-IE	91.73	89.34	85.74	81.92
	TRC-OC-FDPC	88.28	84.93	85.40	81.54
	NC-OC-MVPCA	91.55	89.11	85.75	81.94
	LvaHAI	85.97	82.08	83.40	79.09
	SOR-SRL	82.05	77.16	80.19	75.35
	PCA	87.85	84.49	80.85	76.11
	PCAS	89.32	85.23	80.80	75.23
	BAMGC	<b>91.83</b>	<b>89.46</b>	<b>87.29</b>	<b>83.41</b>

Table 3. Cont.

Dataset	Method	SVM		KNN	
		OA	$\kappa$	OA	$\kappa$
Indian Pines	UBS	73.99	71.91	63.67	61.29
	ONR	76.24	74.29	65.81	63.53
	NC-OC-IE	76.26	74.20	66.31	63.96
	TRC-OC-FDPC	75.34	73.26	67.47	65.14
	NC-OC-MVPCA	76.10	74.10	67.02	64.76
	LvaHAI	50.14	46.66	42.66	40.24
	SOR-SRL	69.03	66.76	63.54	61.17
	PCA	73.34	71.17	63.08	60.76
	PCAS	68.95	66.39	57.97	55.51
	<b>BAMGC</b>	<b>77.47</b>	<b>75.52</b>	<b>68.13</b>	<b>65.84</b>
Salinas	UBS	91.20	90.42	87.92	86.93
	ONR	92.01	91.28	88.86	87.94
	NC-OC-IE	91.63	90.88	88.19	87.22
	TRC-OC-FDPC	92.37	<b>91.67</b>	88.99	88.08
	NC-OC-MVPCA	91.68	90.93	88.72	87.79
	LvaHAI	88.61	87.66	85.55	84.43
	SOR-SRL	90.98	90.18	87.56	86.56
	PCA	91.35	90.57	88.21	87.25
	PCAS	91.94	90.83	89.33	88.25
	<b>BAMGC</b>	<b>92.41</b>	<b>91.67</b>	<b>89.35</b>	<b>88.45</b>
Botswana	UBS	87.78	87.01	82.76	81.78
	ONR	88.87	88.15	85.26	84.35
	NC-OC-IE	90.34	89.69	83.14	82.16
	TRC-OC-FDPC	86.35	85.51	80.65	79.57
	NC-OC-MVPCA	88.05	87.29	82.12	81.09
	LvaHAI	89.03	88.37	83.66	82.74
	SOR-SRL	88.16	87.40	82.66	81.67
	PCA	87.71	86.93	81.91	80.89
	PCAS	86.25	85.22	77.27	75.99
	<b>BAMGC</b>	<b>90.87</b>	<b>90.19</b>	<b>87.55</b>	<b>86.66</b>
Houston	UBS	78.89	74.04	76.34	71.44
	ONR	79.38	74.59	75.66	70.66
	NC-OC-IE	81.22	76.76	<b>80.21</b>	<b>75.91</b>
	TRC-OC-FDPC	79.43	74.66	75.77	70.77
	NC-OC-MVPCA	81.77	77.45	80.46	76.20
	LvaHAI	65.85	59.25	67.94	62.37
	SOR-SRL	69.45	63.25	70.28	64.75
	PCA	80.26	75.74	79.73	75.37
	PCAS	73.69	67.52	74.49	68.77
	<b>BAMGC</b>	<b>81.95</b>	<b>76.91</b>	80.43	76.14

On the Pavla University dataset, NC-OC-IE and NC-OC-MVPCA were almost as good when classified by SVM, and ONR was almost as good when classified by KNN. We have marked these values in italics in Table 3. This illustrates that our proposed method has consistent performance on different classifiers. On the Indian Pines dataset, BAMGC achieves a more accurate classification effect than the whole band when using 18 bands on SVM and when using 9 bands for the KNN. In addition, on the Salinas dataset, BAMGC shows similar optimal performance together with TRC-OC-FDPC. Because the dataset is a wheat field image, the data has the characteristics of high redundancy, demonstrating that BAMGC also has consistent performance in removing redundant bands. On the Botswana dataset, BAMGC mined the spatial information, measured the difference among different pixels, and further improved the accuracy by band grouping, so it gained an outstanding result. Therefore, it can be concluded that BAMGC is applicable to various datasets and different classifiers, reflecting strong stability and high efficiency.

Overall, when the band subsets include  $\{3, 6, 9, 12, 18, 21, 24, 27, 30\}$  bands, BAMGC attains the best results. From Table 4, it can be seen that when using 30 bands, BAMGC consistently gives the best results (Table 4: bold). With the Pavla University dataset, ONR is nearly as good as BAMGC when classified with SVM and LavHAI is nearly as good when classified with KNN (Table 4: italics). On the Salinas data, the performance of BAMGC and TRC-OC-FDPC algorithms are more effective than others under SVM, and that of BAMGC and SOR-SRL are better under KNN. On the Botswana dataset, BAMGC and LavHAI have a better effect on the SVM.

We should note that the Salinas scene only has six distinctive classes, which are easier to discriminate than other datasets. Therefore, the OA and  $\kappa$  on this image are higher than those on the Indian Pines, Pavia University, and Botswana images, as shown in Table 4. In order to further demonstrate the effect difference between the above algorithms, we verify their classification result of them by selecting 15 bands from 5 datasets as band subsets.

By changing the sizes of training samples, the following conclusions can be drawn from all the experiments, which compare the classification accuracy of all the above methods. Firstly, the performance of the 10 methods is quite different for 5 datasets. LvaHAI and SORSRL, for instance, can select a representative subset of bands on the Pavia University dataset, whereas they perform much worse when applied to the Indian Pines dataset. On the contrary, our proposed algorithm has stable and excellent performance on each dataset and different classifiers, especially for Salinas with redundant band information and Pavia University with large spans of spatially adjacent pixels, reflecting the advantages of BAMGC in considering both band and spatial conditions. Additionally, we achieve similar performance to NC-OC-MVPCA on the University of Houston with large size and complex terrain, and better performance than other algorithms. Additionally, an increase in the number of bands does not necessarily lead to better performance, because adding bands that are heavily affected by noise will reduce the accuracy of classification. BAMGC performs satisfactorily in selecting 15 bands, but after exceeding 15 bands, the efficiency reaches a plateau or even declines. We calculate the OA of these algorithms when the number of bands on each dataset is in the range of  $\{3, 6, 9, 12, 18, 21, 24, 27, 30\}$  by SVM, and then calculate their weighted mean. The mean values, sorted from low to high, are shown in Table 5. BAMGC achieves the best results (Table 5: bold).

Table 4. Optimal results of each method in 30 bands.

Dataset	Method	SVM		KNN	
		OA	$\kappa$	OA	$\kappa$
Pavia University	UBS	93.41	91.46	85.77	81.95
	ONR	93.69	91.81	89.12	86.06
	NC-OC-IE	93.33	91.36	86.67	83.08
	TRC-OC-FDPC	93.00	90.93	85.89	82.15
	NC-OC-MVPCA	93.19	91.20	87.14	83.66
	LvaHAI	93.52	91.60	89.18	86.14
	SOR-SRL	92.93	90.86	86.90	83.34
	PCA	91.93	89.59	83.99	79.76
	PCAS	92.08	88.75	85.08	80.27
	BAMGC	<b>93.84</b>	<b>92.02</b>	<b>89.51</b>	<b>86.16</b>
Indian Pines	UBS	78.66	76.79	65.54	63.16
	ONR	79.63	77.83	67.94	65.71
	NC-OC-IE	80.21	78.40	69.85	67.66
	TRC-OC-FDPC	80.05	78.22	68.19	65.89
	NC-OC-MVPCA	78.55	76.66	68.25	65.97
	LvaHAI	59.51	56.77	48.15	45.82
	SOR-SRL	78.19	76.39	68.26	66.00
	PCA	77.11	75.18	67.38	65.10
	PCAS	77.82	75.69	65.89	63.44
	BAMGC	<b>81.29</b>	<b>79.61</b>	<b>71.61</b>	<b>69.42</b>

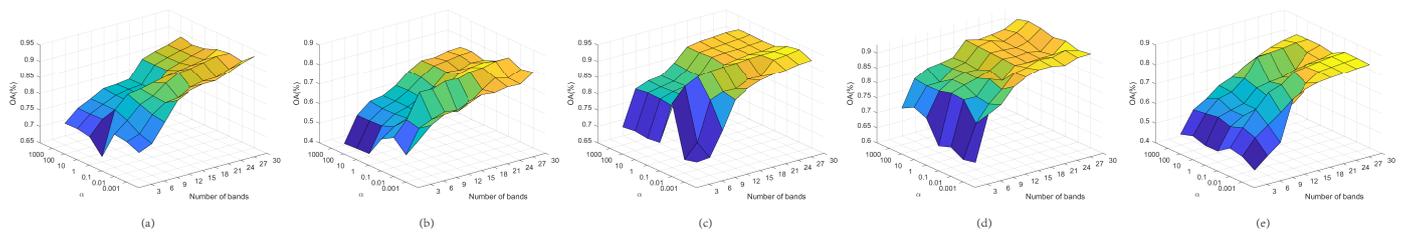
Table 4. Cont.

Dataset	Method	SVM		KNN	
		OA	$\kappa$	OA	$\kappa$
Salinas	UBS	92.64	91.96	89.02	88.11
	ONR	92.95	92.29	89.23	88.33
	NC-OC-IE	92.70	92.02	88.96	88.04
	TRC-OC-FDPC	93.04	92.39	88.99	88.08
	NC-OC-MVPCA	92.88	92.23	89.10	88.20
	LvaHAI	92.27	91.57	87.28	86.25
	SOR-SRL	92.57	91.88	89.35	88.45
	PCA	92.67	92.00	88.78	87.85
	PCAS	91.94	90.83	89.33	88.25
	<b>BAMGC</b>	<b>93.44</b>	<b>92.78</b>	<b>89.54</b>	<b>88.66</b>
Botswana	UBS	89.35	88.65	85.56	84.68
	ONR	90.31	89.66	86.89	86.07
	NC-OC-IE	90.34	89.69	84.78	83.87
	TRC-OC-FDPC	88.91	88.18	84.61	83.69
	NC-OC-MVPCA	91.23	90.63	84.81	83.91
	LvaHAI	91.84	91.33	86.30	85.49
	SOR-SRL	91.57	90.99	84.16	83.22
	PCA	90.14	89.48	84.06	83.12
	PCAS	90.23	89.37	82.31	81.14
	<b>BAMGC</b>	<b>92.24</b>	<b>91.64</b>	<b>87.83</b>	<b>86.96</b>
Houston	UBS	82.24	78.00	76.59	71.72
	ONR	82.01	77.72	77.42	72.65
	NC-OC-IE	84.16	80.27	80.96	76.75
	TRC-OC-FDPC	84.06	80.15	<b>81.15</b>	<b>76.97</b>
	NC-OC-MVPCA	84.14	80.25	80.99	76.79
	LvaHAI	80.49	75.99	75.56	70.62
	SOR-SRL	81.00	76.58	75.82	70.90
	PCA	84.13	80.23	80.70	76.48
	PCAS	84.20	80.29	78.40	73.22
	<b>BAMGC</b>	<b>84.56</b>	<b>80.34</b>	81.00	76.69

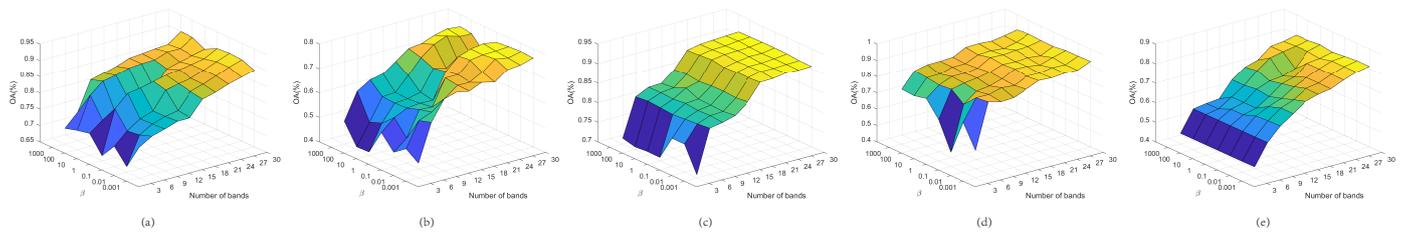
Table 5. The table illustrates the weighted mean of OA when SVM is used.

	Pavia University	Indian Pines	Salinas	Botswana	University of Houston	Mean Value
LvaHAI	84.99	50.26	86.87	84.74	67.26	74.83
SOR-SRL	83.90	68.47	89.87	83.46	69.24	78.99
PCAS	85.91	67.37	90.58	81.03	73.48	79.67
PCA	84.79	69.38	89.54	80.75	74.76	79.85
UBS	89.10	69.89	88.50	84.77	75.68	81.59
TRC-OC-FDPC	88.78	74.34	91.67	84.89	76.58	83.25
ONR	89.38	74.54	91.36	86.58	77.23	83.82
NC-OC-IE	89.20	75.63	91.19	85.92	78.12	84.01
NC-OC-MVPCA	89.12	74.89	91.31	86.94	78.33	84.12
	<b>90.17</b>	<b>76.19</b>	<b>91.80</b>	<b>88.54</b>	<b>79.02</b>	<b>85.14</b>

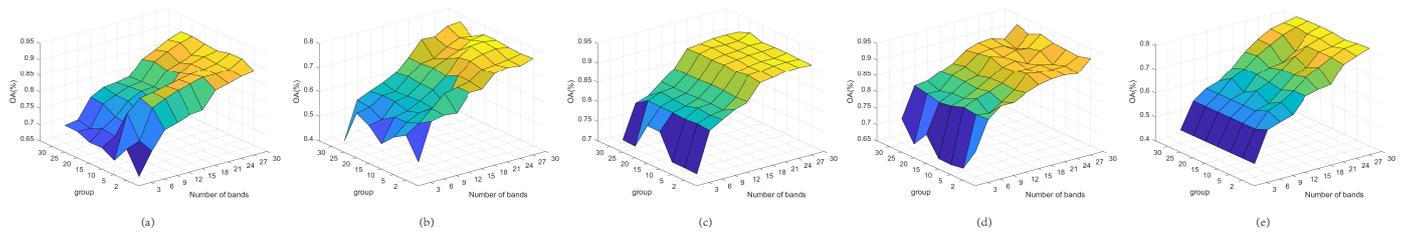
Yet this does not affect the algorithm completing the task of band selection, and we can infer from these facts that our method can obtain superior classification results with fewer bands, which indicates that it is sensitive to discriminative bands. Moreover, the sensitivity of hyperparameters involved in the proposed method is also tested to validate the effect, as shown in Figure 12 for  $\alpha$ , Figure 13 for  $\beta$ , Figure 14 for the number of groups, and Figure 15 for  $\sigma$ . To sum up, the experimental results verify the validity of the proposed method.



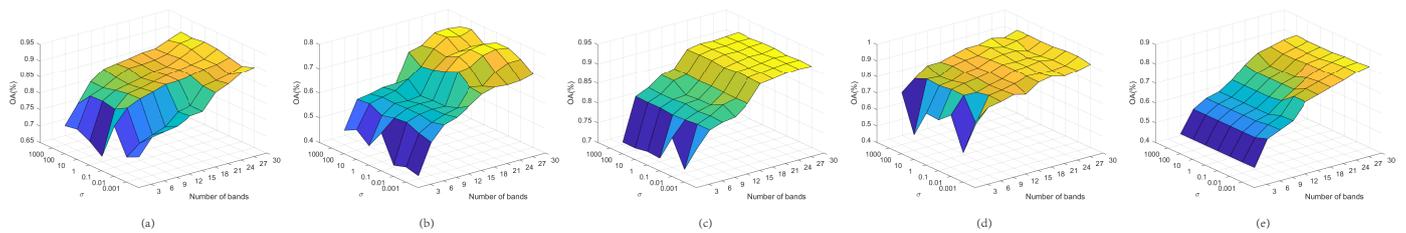
**Figure 12.** The sensitivity of hyperparameter  $\alpha$  on the five datasets. (a) Pavia University; (b) Indian Pines; (c) Salinas; (d) Botswana; (e) Houston University.



**Figure 13.** The sensitivity of hyperparameter  $\beta$  on the five datasets. (a) Pavia University; (b) Indian Pines; (c) Salinas; (d) Botswana; (e) Houston University.



**Figure 14.** The sensitivity of hyperparameter, the number of groups, on the five datasets. (a) Pavia University; (b) Indian Pines; (c) Salinas; (d) Botswana; (e) Houston University.



**Figure 15.** The sensitivity of hyperparameter  $\sigma$  on the five datasets. (a) Pavia University; (b) Indian Pines; (c) Salinas; (d) Botswana; (e) Houston University.

## 5. Conclusions

In this article, we propose an unsupervised band selection method for hyperspectral images based on band grouping and adaptive multi-graph constraint (BAMGC). This method solves the problem of different significance and redundancy of bands by grouping original data and then combining them with weights. The global similarity matrix is reconstructed by local similarity and the weight matrix to preserve the spatial structure information. Furthermore, unsupervised adaptive graph constraints are introduced in order to further optimize the model. We address the model optimization problem by the iterative algorithm and then obtain the best parameters for it. A large number of experimental results show that BAMGC gives consistent results in a wide range of situations and is superior compared to the other nine advanced band selection methods.

**Author Contributions:** Conceptualization, M.Y. and A.Y.; methodology, M.Y. and A.Y.; software, M.Y. and A.Y. and X.M.; validation, X.M.; formal analysis, X.M.; investigation, M.Y. and A.Y.; resources, M.Y. and A.Y.; data curation, X.M.; writing—original draft preparation, Y.W. and X.M.; writing—review and editing, Y.W.; visualization, H.J.; supervision, C.Z.; project administration, M.Y. and A.Y.; funding acquisition, M.Y. and A.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the Natural Science Foundation of Shaanxi Province under Grant 2020JQ-279, in part by the National College Students Innovation and Entrepreneurship Training Program under Grant S202110712609.

**Data Availability Statement:** All data in this paper are open source data, and the corresponding link has been given in the paper.

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

## References

1. Wei, Y.; Zhu, X.; Li, C.; Guo, X.; Yu, X.; Chang, C.; Sun, H. Applications of hyperspectral remote sensing in ground object identification and classification. *Adv. Remote Sens.* **2017**, *6*, 201. [[CrossRef](#)]
2. Liu, S.; Marinelli, D.; Bruzzone, L.; Bovolo, F. A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 140–158. [[CrossRef](#)]
3. Zheng, X.; Gong, T.; Li, X.; Lu, X. Generalized Scene Classification From Small-Scale Datasets With Multitask Learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [[CrossRef](#)]
4. Zhang, Q.; Yuan, Q.; Li, J.; Sun, F.; Zhang, L. Deep spatio-spectral Bayesian posterior for hyperspectral image non-iid noise removal. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 125–137. [[CrossRef](#)]
5. Xie, F.; Li, F.; Lei, C.; Yang, J.; Zhang, Y. Unsupervised band selection based on artificial bee colony algorithm for hyperspectral image classification. *Appl. Soft Comput.* **2019**, *75*, 428–440. [[CrossRef](#)]
6. Yang, H.; Du, Q.; Chen, G. Unsupervised hyperspectral band selection using graphics processing units. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 660–668. [[CrossRef](#)]
7. Zheng, X.; Sun, H.; Lu, X.; Xie, W. Rotation-Invariant Attention Network for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2022**, *31*, 4251–4265. [[CrossRef](#)]
8. Feng, J.; Ye, Z.; Liu, S.; Zhang, X.; Chen, J.; Shang, R.; Jiao, L. Dual-graph convolutional network based on band attention and sparse constraint for hyperspectral band selection. *Knowl.-Based Syst.* **2021**, *231*, 107428. [[CrossRef](#)]
9. Habermann, M.; Fremont, V.; Shiguemori, E.H. Supervised band selection in hyperspectral images using single-layer neural networks. *Int. J. Remote Sens.* **2019**, *40*, 3900–3926. [[CrossRef](#)]
10. Guo, Z.; Bai, X.; Zhang, Z.; Zhou, J. A hypergraph based semi-supervised band selection method for hyperspectral image classification. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia, 15–18 September 2013; pp. 3137–3141.
11. Zheng, X.; Wang, B.; Du, X.; Lu, X. Mutual Attention Inception Network for Remote Sensing Visual Question Answering. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [[CrossRef](#)]
12. Zhu, G.; Huang, Y.; Lei, J.; Bi, Z.; Xu, F. Unsupervised hyperspectral band selection by dominant set extraction. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 227–239. [[CrossRef](#)]
13. Zheng, X.; Chen, X.; Lu, X.; Sun, B. Unsupervised Change Detection by Cross-Resolution Difference Learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. doi: 10.1109/TGRS.2021.3079907. [[CrossRef](#)]
14. Yang, C.; Bruzzone, L.; Zhao, H.; Tan, Y.; Guan, R. Superpixel-based unsupervised band selection for classification of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 7230–7245. [[CrossRef](#)]
15. Gong, M.; Zhang, M.; Yuan, Y. Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 544–557. [[CrossRef](#)]
16. Jia, S.; Ji, Z.; Qian, Y.; Shen, L. Unsupervised band selection for hyperspectral imagery classification without manual band removal. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 531–543. [[CrossRef](#)]
17. Tarabalka, Y.; Fauvel, M.; Chanussot, J.; Benediktsson, J.A. SVM-and MRF-based method for accurate classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 736–740. [[CrossRef](#)]
18. Tang, C.; Liu, X.; Zhu, E.; Wang, L.; Zomaya, A. Hyperspectral Band Selection via Spatial-Spectral Weighted Region-wise Multiple Graph Fusion-Based Spectral Clustering. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Montreal-themed Virtual Reality, 19–27 August 2021; pp. 3038–3344.
19. Beirami, B.A.; Mokhtarzade, M. An Automatic Method for Unsupervised Feature Selection of Hyperspectral Images Based on Fuzzy Clustering of Bands. *Trait. Signal* **2020**, *37*, 319–324. [[CrossRef](#)]
20. Sun, W.; Du, Q. Hyperspectral band selection: A review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 118–139. [[CrossRef](#)]
21. Sun, W.; Du, Q. Graph-regularized fast and robust principal component analysis for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3185–3195. [[CrossRef](#)]

22. Chang, C.I.; Du, Q. Interference and noise-adjusted principal components analysis. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2387–2396. [[CrossRef](#)]
23. Wang, Q.; Lin, J.; Yuan, Y. Salient band selection for hyperspectral image classification via manifold ranking. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 1279–1289. [[CrossRef](#)]
24. Jia, S.; Tang, G.; Zhu, J.; Li, Q. A novel ranking-based clustering approach for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 88–102. [[CrossRef](#)]
25. Wang, Q.; Li, Q.; Li, X. Hyperspectral band selection via adaptive subspace partition strategy. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4940–4950. [[CrossRef](#)]
26. Su, H.; Cai, Y.; Du, Q. Firefly-Algorithm-Inspired Framework With Band Selection and Extreme Learning Machine for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 309–320. [[CrossRef](#)]
27. Yang, H.; Du, Q.; Chen, G. Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 544–554. [[CrossRef](#)]
28. Su, H.; Du, Q.; Chen, G.; Du, P. Optimized hyperspectral band selection using particle swarm optimization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2659–2670. [[CrossRef](#)]
29. Yuan, Y.; Zheng, X.; Lu, X. Discovering diverse subset for unsupervised hyperspectral band selection. *IEEE Trans. Image Process.* **2016**, *26*, 51–64. [[CrossRef](#)]
30. Zhang, R.; Ma, J. Feature selection for hyperspectral data based on recursive support vector machines. *Int. J. Remote Sens.* **2009**, *30*, 3669–3677. [[CrossRef](#)]
31. Zhong, P.; Zhang, P.; Wang, R. Dynamic learning of SMLR for feature selection and classification of hyperspectral data. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 280–284. [[CrossRef](#)]
32. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)] [[PubMed](#)]
33. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396. [[CrossRef](#)]
34. He, X.; Cai, D.; Yan, S.; Zhang, H.J. Neighborhood preserving embedding. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Beijing, China, 17–21 October 2005; Volume 2, pp. 1208–1213.
35. Li, K.; Luo, G.; Ye, Y.; Li, W.; Ji, S.; Cai, Z. Adversarial privacy-preserving graph embedding against inference attack. *IEEE Internet Things J.* **2020**, *8*, 6904–6915. [[CrossRef](#)]
36. Zhong, Z.; Li, J.; Luo, Z.; Chapman, M. Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 847–858. [[CrossRef](#)]
37. Islam, M.; Sohaib, M.; Kim, J.; Kim, J.M. Crack Classification of a Pressure Vessel Using Feature Selection and Deep Learning Methods. *Sensors* **2018**, *18*, 4379. [[CrossRef](#)]
38. Ribalta Lorenzo, P.; Tulczyjew, L.; Marcinkiewicz, M.; Nalepa, J. Hyperspectral band selection using attention-based convolutional neural networks. *IEEE Access* **2020**, *8*, 42384–42403. [[CrossRef](#)]
39. Kang, M.; Kim, J.; Wills, L.M.; Kim, J.M. Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2015**, *62*, 7749–7761. [[CrossRef](#)]
40. Cai, Y.; Liu, X.; Cai, Z. BS-Nets: An end-to-end framework for band selection of hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 1969–1984. [[CrossRef](#)]
41. Mou, L.; Saha, S.; Hua, Y.; Bovolo, F.; Bruzzone, L.; Zhu, X.X. Deep reinforcement learning for band selection in hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [[CrossRef](#)]
42. Nie, F.; Zhang, R.; Li, X. A generalized power iteration method for solving quadratic problem on the Stiefel manifold. *Sci. Inf. Sci.* **2017**, *60*, 112101. [[CrossRef](#)]
43. Wang, Q.; Zhang, F.; Li, X. Optimal clustering framework for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5910–5922. [[CrossRef](#)]
44. Stella, X.Y.; Shi, J. Multiclass spectral clustering. In Proceedings of the Computer Vision, IEEE International Conference, Nice, France, 13–16 October 2003; pp. 313–313.
45. Chang, C.I.; Du, Q.; Sun, T.L.; Althouse, M.L. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2631–2641. [[CrossRef](#)]
46. Wang, Q.; Zhang, F.; Li, X. Hyperspectral band selection via optimal neighborhood reconstruction. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8465–8476. [[CrossRef](#)]
47. Wei, X.; Cai, L.; Liao, B.; Lu, T. Local-View-Assisted Discriminative Band Selection With Hypergraph Autolearning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2042–2055. [[CrossRef](#)]
48. Wei, X.; Zhu, W.; Liao, B.; Cai, L. Scalable one-pass self-representation learning for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4360–4374. [[CrossRef](#)]
49. Zabalza, J.; Ren, J.; Ren, J.; Liu, Z.; Marshall, S. Structured covariance principal component analysis for real-time onsite feature extraction and dimensionality reduction in hyperspectral imaging. *Appl. Opt.* **2014**, *53*, 4440–4449. [[CrossRef](#)] [[PubMed](#)]