Article

A Band Subset Selection Approach Based on Sparse Self-Representation and Band Grouping for Hyperspectral Image Classification

Keng-Hao Liu *, Yu-Kai Chen and Tsun-Yang Chen

Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-sen University,
Kaohsiung 80424, Taiwan
* Correspondence: keng3@mail.nsysu.edu.tw

Abstract: Band subset selection (BSS) is one of the ways to implement band selection (BS) for a hyperspectral image (HSI). Different from conventional BS methods, which select bands one by one, BSS selects a band subset each time and preserves the best one from the collection of the band subsets. This paper proposes a BSS method, called band grouping-based sparse self-representation BSS (BG-SSRBSS), for hyperspectral image classification. It formulates BS as a sparse self-representation (SSR) problem in which the entire bands can be represented by a set of informatively complementary bands. The BG-SSRBSS consists of two steps. To tackle the issue of selecting redundant bands, it first applies band grouping (BG) techniques to pre-group the entire bands to form multiple band groups, and then performs band group subset selection (BGSS) to find the optimal band group subset. The corresponding representative bands are taken as the BS result. To efficiently find the nearly global optimal subset among all possible band group subsets, sequential and successive iterative search algorithms are adopted. Land cover classification experiments conducted on three real HSI datasets show that BG-SSRBSS can improve classification accuracy by 4–20% compared to the existing BSS methods and requires less computation time.

Keywords: band subset selection (BSS); band selection (BS); hyperspectral image (HSI); sparse self-representation (SSR); band grouping (BG)

1. Introduction

Band selection is one of the most important techniques to deal with the dimension issue for hyperspectral images (HSI) [1,2] in the remote sensing community. With the abundant spectral information captured by the HSI image sensor, a HSI image cube usually has enormous data volume and contains a large amount of inter-band redundancy. As a consequence, using the entire bands for data transmission, storage, and image analysis is sometimes impracticable. The objective of BS is to select a set of bands that can best represent the information of the original cube. In the past, many different types of BS methods were proposed for various kinds of applications, such as classification, spectral unmixing, endmember extraction, and target detection. There are many ways to achieve BS. According to [1], they can roughly be classified as six divisions: ranking- based [3–6], searching-based [7–18], clustering-based [19–23], sparsity-based [24–29], embedding-based [30–34], and hybrid scheme based [35–37].

Suppose a HSI cube contains L bands. Let p denotes the number of bands required to be selected. Given an objective or evaluation function, the BS aims to find the best band group from all $C_p^L = \frac{L!}{p!(L-p)!}$ possible combinations. Since it is almost impossible to exhaust all combinations in practice, using an efficient search strategy to select the band is necessary. From the point of view of algorithm design, some BS methods adopt sequential forward selection (SFS) to select bands one by one in a sequential manner, referred to as

sequential multiple band selection (SQMBS). Many existing ranking-based and searching-based approaches belong to SQMBS. However, the drawbacks of SQMBS are apparent: 1. The selected bands may no longer be able to be replaced or removed. It reduces the probability of finding the global optimal or nearly global optimal solution among all possible band subsets; 2. The initial condition (i.e., the first-selected band or the given ranking criterion) may greatly affect the BS result. As a result, the SQMBS may be implemented efficiently; however, it could only find the locally optimal band subset. In order to tackle this dilemma, recent BS works towards developing an approach that can select multiple bands each time, called simultaneous multiple band selection (SMMBS) [18]. Since there are $\frac{L!}{p!(L-p)!}$ band subsets that can be considered by SMMBS in total, the main issue is located on the design of the search strategy. For instance, clonal selection algorithm (CSA) [9], particle swarm optimization (PSO) [10], evolution-based methods [8,12], multigraph determinantal point processes (MDPP) [22], and other clustering algorithms [19–21,23] were proposed to achieve SMMBS. Although those works have shown great success on BS, the searching algorithms may rely on a random initialization process that creates the irreproducibility and inconsistency in BS results. In addition, the performance of some search algorithms heavily relies on the initial parameter setting. These innate characteristics limit the practicality and stability of such methods.

Recently, a new branch of SMMBS was developed to resolve this issue, called band subset selection (BSS) [16–18]. The BSS formulates a BS problem as an endmember finding a procedure in which each band is considered an endmember and it is desired to select an optimal set of p endmembers (bands) from the entire L endmembers (bands). The main difference between BSS and the above-mentioned SMMBS methods is that BSS utilizes a double-nested iterative algorithm originating from N-FINDER algorithms [38,39] to search the optimal band subset in an efficient way. More concretely, given an initial band subset with p bands, the BSS can gradually find the optimal band subset by continuously updating/replacing its members. The first advantage of BSS is that it avoids performing an exhausted search while it can find a nearly global optimal solution. The second advantage of BSS is that using a random initial condition is unnecessary. The initial band subset is purely determined by uniform band selection (UBS) instead of random selection to ensure the consistency of the final BS result while keeping high information complementary of the selected bands. Additionally, in BSS's framework, the p is suggested to be determined by virtual dimensionality (VD) [40,41] algorithms, which provides a clear guidance for p. The third advantage of BSS is the flexibility of choosing evaluation function. We can adopt the most suitable objective functions to evaluate the band subsets according to the back-end application.

On the other hand, sparse representation-based techniques have shown promising results in hyperspectral image classification [42,43] and spectral unmixing [44,45]. Recently, the sparse representation was applied to solve the BS problem [24–29]. Most of them adopted a so-called sparse self-representation (SSR) model [46], which is regarded as an extended version of the traditional sparse representation. The SSR-based BS assumes that all bands of a HSI image can be linearly represented by a set of a few bands. More specifically, the SSR-based BS treats each band as a band vector, and all band vectors play the roles of both dictionary matrix and input matrix. The objective is to find a set of band vectors that can minimize the sparse error (reconstructed error) with sparsity p. Based on the characteristics of the complementary in vector space, the SSR-based BS is able to select the band groups with highly informative complementarity. Those approaches especially show effectiveness on the topic of HSI image classification. Motivated by this idea, we can employ the SSR model as the evaluation criterion for BSS and develop a new BSS approach called sparse self-representation BSS (SSRBSS). The optimization of SSRBSS is to find a band subset that minimizes the reconstruction error of the SSR model. Similar to the existing BSS methods, two search strategies, referred to as successive (SC) and sequential (SQ) iterative algorithms, are used for the efficient search of the optimal band subset. However, according to the results of the existing BSS approaches [16–18], the selected bands may still be very

adjacent or clustered in some specific spectral regions. This is mainly due to the inherent characteristics of the evaluation function or the interference of the noise in band images. It would lead to the high information redundancy in the selected bands, which may not be favorable to subsequent image analysis. This issue can also occur on SSRBSS.

In order to tackle this dilemma, this paper proposes band grouping-based SSRBSS (BG-SSRBSS) by integrating band grouping (BG) techniques into an SSRBSS framework. In BG-SSRBSS, BG is employed as a pre-processing step that divides all the bands into g band groups ($g \geq p$) in the direction of the spectrum. Such a process is the so-called neighborhood band grouping (NBG) [47]. The NBG can group the adjacent bands together to achieve the effect of removing redundancy. After all the bands are pre-grouped, the BG-SSRBSS performs band group subset selection (BGSS), which finds an optimal band group subset (composed of p groups) that minimizes the reconstruction error of the SSR model. Finally, the representative bands of the selected band group subset are taken as the BS result. It is noted that all the band members of the selected band groups are involved in the SSR optimization. Such a concept originated from group orthogonal matching pursuit [48], which can select groups of variables for the best regression performance. Based on the above, the concept of BG-SSRBSS can be viewed as a combination of a band decorrelation process (BD) [49] with BSS, where the redundant bands are predefined in advance and removed after BSS's optimization is done. In other words, it possesses the capabilities of efficiently searching an optimal subset and removing redundant bands. In BG-SSRBSS, the parameter g controls the degree of redundancy removal. Finally, it is worth mentioning that SSRBSS belongs to a special case of BG-SSRBSS when $g = p$.

The experiments conducted on three real HSI datasets verify the effectiveness of the proposed BG-SSRBSS in BS as well as the classification performance. The main contributions of this paper are as follows.

1. We integrate the SSR model with a BSS framework and propose the SSRBSS method for hyperspectral image classification. The SSRBSS adopts the reconstructed error of the SSR model as the objective function to evaluate the quality of band subset. In our SSRBSS, the search for the optimal band subset can be efficiently done via SC or SQ search scheme, and the model error can be efficiently calculated via least square equation.
2. In order to reduce the information redundancy in the final selected bands, a novel two-stage BS method is proposed, called band grouping-based SSRBSS (BG-SSRBSS). In BG-SSRBSS, the BG is adopted as preprocessing to partition all the bands into several non-overlapping band groups; then, BGSS is performed to find the most influential band group subset via SC or SQ search. The representative bands of the optimal band group are taken as the BS result. It is worth mentioning that BG-SSRBSS integrates SSRBSS into one framework.
3. The proposed BG-SSRBSS enlarges the existing BSS framework to a new branch, called band grouping-based BSS (BG-BSS). In BG-BSS, any type of subset evaluation criteria, as well as BG methods, can be used.

Despite the success provided by BG-SSRBSS, it also suffers from an obvious issue occurring in all BSS methods: when p changes, the entire iteration process must be re-run to find a new optimal subset. Another issue is the difficulty in setting an appropriate g value. In BG-SSRBSS, g controls the degree of removing redundant bands. Setting the g value too large or too small would result in poor performance of BG-SSRBSS. We leave this part of investigating the relationship between g and p to our future work.

The remainder of this paper is organized as follows. Section 2 introduces the related works. Section 3 gives a brief review of BSS. Section 4 introduces the SSR Model and SSS-RBSS method. Section 5 presents the proposed BG-SSRBSS method. Section 6 describes the real HSI datasets for experiment, experimental settings, and results. Finally, the conclusion is drawn in Section 7.

2. Related Works

In the past, many SMMBS methods were proposed to solve the BS problem. For example, Feng, J. et al. [9] proposed a new clonal selection algorithm (CSA) that can search an appropriate band subset by the trivariate mutual information (TMI) or semi-supervised TMI (STMI) criteria for hyperspectral image classification. Su, H. et al. [10] proposed a particle swarm optimization (PSO)-based system to select bands. Two PSOs are incorporated in the system in which an outer one is responsible for selecting the optimal number of bands, while the inner one is in charge of searching for the optimal bands. Yuan et al. [19] proposed a novel technique named the dual-clustering-based BS method to select the most discriminative bands. They proposed a new descriptor to reveal the contextual information of HSI that can take both spatial and spectral information in the clustering process. Yuan et al. [22] proposed a multigraph determinantal point process (MDPP) model to find a diverse band combination that contains discriminative and informative bands for hyperspectral applications. They adopted a graph framework to exploit the intrinsic structure of different bands and used Mix-DPP to find the most diverse band combinations. Zeng et al. [23] proposed a novel clustering-based approach using deep subspace clustering (DSC). This approach combines the subspace clustering task into a convolutional auto-encoder by treating it as a self-expressive layer that can learn a nonlinear spectral–spatial relationship during the training process. Those works demonstrated the great success in hyperspectral image classification. However, their performance relies on a random initialization process or parameter setting.

BSS [16–18] is a new type of SMMBS methods to achieve BS. In BSS, the evaluation function plays the key role to select bands. Wang et al. [16] proposed a constrain band subset selection (CBSS) that uses the measure proposed in CBS [4] to evaluate the quality of multiple band subsets. Chang et al. [17] formulated BSS as a channel capacity problem in an information theorem and proposed channel capacity BSS (CCBSS), in which the band subset represents the channel input and the entire set of bands represents the channel output. These works demonstrate that the BSS-based approaches have great potential in finding a band subset suitable for endmember extraction and land cover classification. Followed by this trend, Yu et al. [18] developed LCMV-BSS, which makes use of linearly constrained minimum variance (LCMV) as the objective function to find the optimal band subset for hyperspectral image classification. This work also demonstrates the strength of BSS against a few existing SMMBS approaches. Based on the selectivity of the objective function, apparently, there is still a lot of space for BSS-based approaches.

On the other hands, sparse representation is a popular technique to solve the BS problem [24–29]. For example, Li et al. [24] proposed a sparse representation-based band selection (SpaBS) algorithm. They adopted K-SVD to decompose the image data into the multiplication of the signature matrix and the coefficient matrix. The bands are selected based on the histogram of the coefficient matrix. Lai et al. [27] proposed an efficient BS method, called SpaBS-OMP, to select the representative bands in an efficient way. They formulated the BS problem as the optimization of the SSR model and used the orthogonal matching pursuit (OMP) algorithm to select bands one-by-one in a sequential fashion from the dictionary matrix. Sun et al. [28] proposed a dissimilarity-weighted sparse self-representation (DWSSR) method to select a proper band subset for HSI classification. The DWSSR improves from the traditional SSR model by integrating a new dissimilarity-weighted regularization term. The experiment proves that using DWSSR's selected bands could reach higher accuracy and lower computing time than the compared methods.

3. Band Subset Selection (BSS)

The BSS were originally proposed in [16–18]. Let B_i ($1 \leq i \leq L$) represent a band in a HSI cube, $\Omega = \{B_1, B_2, \ldots, B_L\}$ represents the full band set, and $\Omega_p = \left\{B_{l_1}, B_{l_2}, \ldots, B_{l_p}\right\}$ represent a band subset selected from Ω. Given a specific objec-

tive function $J(\cdot)$, BSS aims to find an optimal band subset $\Omega_p^* = \left\{ B_1^{(*)}, B_2^{(*)}, \ldots, B_p^{(*)} \right\}$ that can maximize or minimize $J(\cdot)$. Thus, the problem of BSS can be formulated by

$$\Omega_p^* = \arg\left\{ \max/\min_{\Omega_p \subset \Omega} J(\Omega_p) \right\} \tag{1}$$

In the existing BSS approaches, the initial band subset $\Omega_p^{(0)}$ is suggested to be determined by UBS, and p is determined by VD [40]. After the initial condition $\Omega_p^{(0)}$ is set, BSS finds a better band subset $\Omega_p^{(t)}$ by iteratively exchanging the members between previous band subsets $\Omega_p^{(t-1)}$ and the corresponding complementary band set $\left\{ \Omega - \Omega_p^{(t-1)} \right\}$, and retaining the one that maximizes or minimizes the objective function until the iteration ends. The iteration process is controlled by a double-nested for loop. Two search methods, referred to as successive (SC) and sequential (SQ) algorithms, can be used in the iteration. Thus, BSS can be subdivided into two methods: SC BSS and SQ BSS.

3.1. SC BSS

In SC BSS, the outer loop controls the index of the current band subset, and the inner loop controls the index of the complementary band set. The procedure of SC-BSS is described as follows:

1. Initialization:
    Set $\Omega_p^{(0)} = \{ B_1^{(0)}, B_2^{(0)}, \ldots, B_p^{(0)} \}$ by UBS or random band selection, and calculate $E^{(0)} = J\left\{ \Omega_p^{(0)} \right\}$.

2. Outer Loop:
    Use index j as a counter to check all the bands in $\Omega_p^{(j)}$ ($1 \leq j \leq p$). If $j > p$, the algorithm terminates.

3. Inner Loop:
    Use index l as a counter to track the lth band B_l ($1 \leq l \leq L$). If $B_l \notin \Omega_p^{(j-1)}$, set candidate band $B^* = B_l$, and calculate $E_l^{(j)} = J\left\{ B_1^{(j-1)}, \ldots, B_{j-1}^{(j-1)}, B^*, B_{j+1}^{(j-1)}, \ldots, B_p^{(j-1)} \right\}$. Inner loop ends.

4. Suppose $E_k^{(j)}$ denotes the smallest (or largest) value found in Inner Loop. If $E_k^{(j)} < E^{(j-1)}$ (or $E_k^{(j)} > E^{(j-1)}$), $B_j^{(j)}$ will be replaced by $B^* = B_k$. Then update $\Omega_p^{(j)} = \left\{ B_1^{(j-1)}, \ldots, B_{j-1}^{(j-1)}, B_k, B_{j-1}^{(j-1)}, \ldots, B_p^{(j-1)} \right\}$ and set $E^{(j)} = E_k^{(j)}$. Otherwise, set $\Omega_p^{(j)} = \Omega_p^{(j-1)}$ and $E^{(j)} = E^{(j-1)}$. Finally, let $j \leftarrow j + 1$ and go to Step 2. Outer loop ends.

5. Step 2–4 will be repeated until the termination condition is reached. The final output is $\Omega_p^* = \Omega_p^{(p)}$.

3.2. SQ BSS

Conversely, the outer loop of SQ BSS controls the index of the complementary band set while the inner loop controls the index of the current band subset. The procedure of SQ BSS is described as follow.

1. Initialization:
    Set $\Omega_p^{(0)} = \{ B_1^{(0)}, B_2^{(0)}, \ldots, B_p^{(0)} \}$ by UBS or random band selection, and calculate $E^{(0)} = J\left\{ \Omega_p^{(0)} \right\}$.

2. Outer Loop:

Use index l as a counter to check if band $B_l \in \Omega$ for $1 \leq l \leq L$. If $l > L$, the algorithm terminates. If $B_l \notin \Omega_p^{(l-1)}$, set candidate band $B^* = B_l$, then go to Inner Loop. Otherwise, set $\Omega_p^{(l)} = \Omega_p^{(l-1)}$ and $E^{(l)} = E^{(l-1)}$.

3. Inner Loop:
Use index j be a counter to track the jth band of Ω_p for $1 \leq j \leq p$. Then calculate

$$J\left\{B^*, \ldots, B_j^{(l-1)}, \ldots, B_p^{(l-1)}\right\}, \qquad J\left\{B_1^{(l-1)}, \ldots, B^*, \ldots, B_p^{(l-1)}\right\}, \qquad \ldots,$$
$$J\left\{B_1^{(l-1)}, \ldots, B_j^{(l-1)}, \ldots, B^*\right\}. \text{ Inner loop ends.}$$

4. Suppose $E_k^{(l)}$ denotes the smallest (or largest) value found in Inner Loop. If $E_k^{(l)} < E^{(l-1)}$ (or $E_k^{(l)} > E^{(l-1)}$), $B_k^{(l-1)}$ will be replaced by B^*, that is $\Omega_p^{(l)} = \left\{B_1^{(l-1)}, \ldots, B_{k-1}^{(l-1)}, B^*, B_{k+1}^{(l-1)}, \ldots, B_p^{(l-1)}\right\}$. Finally, set $E^{(l)} = E_k^{(l)}$ and $l \leftarrow l + 1$. Go to Step 2. Outer loop ends.

5. Step 2–5 will be repeated until the termination criterion is reached. The final output is $\Omega_p^* = \Omega_p^{(L)}$.

4. SSR Model and SSRBSS

This section briefly explains the concept of SSR model and introduces the SSRBSS method, which is a combination of the SSR model and BSS.

4.1. The SSR Model for BS

The SSR model was used to solve BS problems in the past [24–29]. Let L and N be the total numbers of bands and spectral pixels in a HSI cube, respectively ($N >> L$). Suppose $\mathbf{b}_i \in \mathcal{R}^{N \times 1}$ represents the band vector of band B_i, and $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_L] \in \mathcal{R}^{N \times L}$ represents the HSI matrix. The optimization of SSR-based BS is formulated as:

$$\mathbf{C}^* = \underset{\mathbf{C}}{\arg\min} ||\mathbf{B} - \mathbf{BC}||_F^2 \ s.t. ||\mathbf{C}||_{0,2} \leq p \tag{2}$$

where $\mathbf{C} \in \mathcal{R}^{L \times L}$ is the sparse coefficient matrix and p is sparsity. The objective of (2) is to find an optimal coefficient matrix $\mathbf{C}^*$ with only p non-zero rows that minimizes the reconstructed error $||\mathbf{B} - \mathbf{BC}||_F^2$ From another point of view, solving (2) is equivalent to finding an optimal band subset Ω_p^* with p band members. Figure 1 shows the concept of the SSR-based BS.
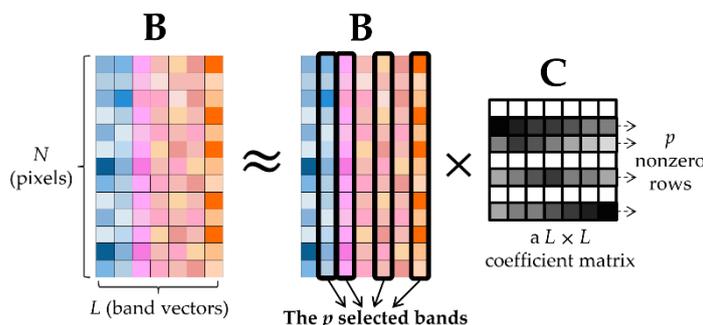

Figure 1. The concept of SSR model for BS.

The past works adopt several optimization algorithms such as K-SVD [24], LSR [26], OMP [27], or ADMM [28] to search the optimal band subset to fulfill the minimum reconstruction error of the SSR model. Additional constraints can be imposed during the optimization.

4.2. SSRBSS

If we combine the BSS framework introduced in Section 3 with the SSR model, a new BSS approach can be developed, referred to as SSR-based BS (SSRBSS). The difference from the existing SSR-based BS approaches is that we adopt BSS's SC/SQ search strategy to find the optimal band vector subset from the dictionary matrix **B**, and thus, do not rely on the above-mentioned optimization algorithms. More concretely, the SSRBSS designates multiple band subsets, evaluates them via the error of the SSR model, and leaves the best one at the end. Suppose $\Omega_p^{(t)}$ is a band subset designated by BSS's at time *t*, and **P** denotes its band vector matrix, the problem (2) can be re-formulated as:

$$\hat{\mathbf{Q}} = \underset{Q}{\mathrm{argmin}} ||\mathbf{B} - \mathbf{PQ}||_F^2 \tag{3}$$

where $\mathbf{Q} \in \mathcal{R}^{p \times N}$ denotes the coefficient matrix respect to **P**. If there is no constraint imposed for **Q**, the problem (3) can be efficiently solved by the least square formula:

$$\hat{\mathbf{Q}} \left(\mathbf{P}^T\mathbf{P}\right)^{-1} \mathbf{P}^T\mathbf{B} \tag{4}$$

Furthermore, the reconstruction error of **B** given **P** can be expressed by

$$E(\mathbf{P}) = ||\mathbf{B} - \mathbf{P}\hat{\mathbf{Q}}||_F^2 \tag{5}$$

Equations (4) and (5) can be used to measure the quality of any band subset. As a result, the SSRBSS can be regarded as a BSS method using Equation (3) as the objective function and can also be viewed as a solution to solve SSR-based BS problems in Equation (2) via the SC or SQ search method.

5. BG-SSRBSS

This section introduces the details of the proposed BG-SSRBSS method, which include band grouping (BG) methods, the band group subset selection (BGSS) process, and BG-SSRBSS algorithms.

5.1. Band Grouping and Representative Bands

It is known that the adjacent bands in an HSI cube are highly similar because of the nature of continuous spectrum. When performing BS, according to the characteristics of the algorithm, similar or adjacent bands are inevitably selected, which may produce poor BS results. Therefore, it is necessary to take some strategies to avoid selecting redundant bands.

Band grouping (BG) [47,50] is a technique that can cluster similar bands together into the same group. Suppose Ω present the full band set. The BG method can segment Ω to a set of non-overlapping band group set $\Phi_g = \{G_1, G_2, \ldots, G_g\}$, where *g* denotes the number of band groups, and G_i ($1 \leq i \leq g$) means the *i*-th band group with n_i band members $\left\{ \mathrm{B}_{G_{i1}}, \mathrm{B}_{G_{i2}}, \mathrm{B}_{G_{i3}}, \ldots, \mathrm{B}_{G_{in_i}} \right\}$. Once all the *L* bands are grouped, it is able to limit the number of candidate bands in each group that will be picked by BS at the same time. Doing so can ensure that the selected bands will be more informatively complementary or less informatively redundant. As we know, the similarity between the non-adjacent bands is usually lower than that between adjacent bands. If BG is operated in the direction of spectrum, theoretically, it will produce the best decorrelation result. Such a method is called neighbor band grouping (NBG). The simplest BG method is uniform band grouping (UBG), which segments the spectrum uniformly. However, it does not take into account the image/spectral information at all. Theoretically, it is impossible to obtain a good grouping performance. In this paper, we adopt two NBG methods to be a preprocessing step for our BG-SSRBSS. One is fast neighborhood grouping (FNG) [47], and the other is the band decorrelation (BD) [6,50] process.

The FNG is a coarse-to-fine NGB method. It first partitions all bands uniformly into g band groups, called coarse band grouping, and then performs a fine neighborhood grouping scheme to repartition the initial groups to obtain finer band grouping result. The BD is a decorrelation process proposed in PBS [6]. The basic idea of PBS is to select an information criterion to prioritize all bands and select the first p bands with highest prioritization scores. However, it often occurs that one band is selected and its adjacent bands are also selected because their priority scores are similar. In this situation, it is able to use BD to remove those redundant bands by imposing a limit that the newly incoming band must be certainly different form the selected bands via spectral angle mapper (SAM) or spectral information divergence (SID) measure with a threshold ϵ. If we directly apply BD to all the bands B_1, B_2, ..., B_L in original band order, we can obtain a NBG result.

Except for the decrease in information redundancy of the selected bands, another advantage of using BG is to reduce the number of iterations required in searching the optimal subset under SC/SQ search so as to save the computing time.

5.2. Band Group Subset Selection (BGSS) and BG-SSRBSS

After the BG is done, the next step is to find the optimal band group subset that can maximize or minimize $J(\cdot)$. Such a step is called band group subset selection (BGSS). If the SSR model is adopted for $J(\cdot)$, a new BS approach, called BG-SSRBSS, is proposed.

More specifically, Ω is first divided into g band groups via BG, denoted as $\Phi_g = \{G_1, G_2, \ldots, G_g\}$. The next step is to select a band group subset with p band groups from Φ_g, denoted as $\Psi_p^t = \left\{ G_1^{(t)}, G_2^{(t)}, \ldots, G_p^{(t)} \right\} \subset \Phi_g$, and calculate Equation (5) with $\mathbf{P} = \left[\mathbf{G}_1^{(t)} \ \mathbf{G}_2^{(t)} \cdots \mathbf{G}_p^{(t)} \right]$, the concatenation of the band matrices of the p band groups in $\Psi_p^{(t)}$, to measure the quality of the current band group subset. This action continues until the optimal band group subset $\Psi_p^* = \left\{ G_1^{(*)}, G_2^{(*)}, \ldots, G_p^{(*)} \right\}$ is found. The concept of BGSS is demonstrated in Figure 2.
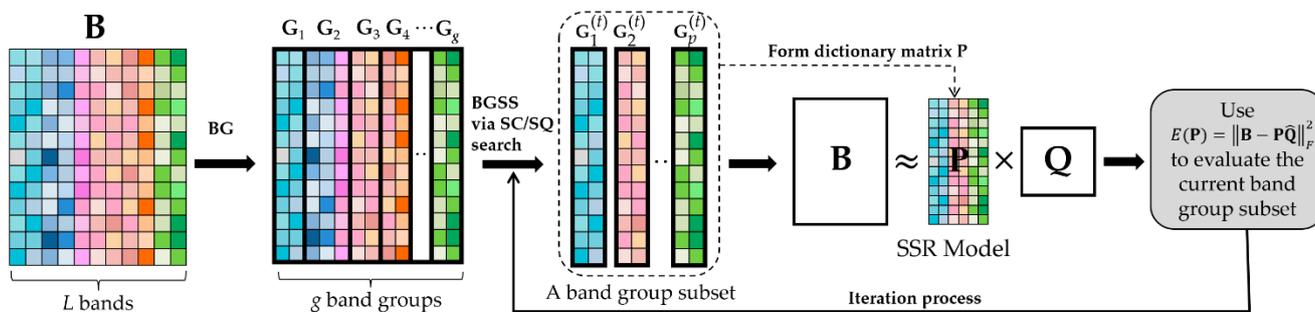

Figure 2. The mechanism of BG-SSRBSS.

Theoretically, Ψ_p^* represents the band groups that contribute substantially to the SSR model. The final step is to generate the representative band of each band group in Ψ_p^* as the BS result. There are several ways to generate the representative band. For instance, it can be assigned by the band with maximum information entropy. In this paper, we select the band that is closest to the centroid of the bands in $G_i^{(*)}$ as the representative band.

Let $B_{G_{ik}^{(*)}}$ present a band in $G_i^{(*)}$ and $\mathbf{b}_{G_{ik}^{(*)}}$ be the corresponding band vector $(1 \leq k \leq n_i)$. The representative band of $G_i^{(*)}$ can be obtained with the formula

$$B^r_{G_i^{(*)}} = \arg \min_{B_{G_{ik}^{(*)}} \in G_i^{(*)}} || \mathbf{b}_{G_{ik}^{(*)}} - \overline{\mathbf{b}}_{G_i^{(*)}} || \qquad (6)$$

where $\|\cdot\|$ denotes the Euclidean norm and $\bar{\mathbf{b}}_{G_i^{(*)}}$ denotes the averaged vector of all the band vectors in $G_i^{(*)}$.

According to the choices of search strategy and BG technique, there are six BG-SSRBSS algorithms in total that can be implemented. If BG-SSRBSS is implemented with SC/SQ search strategy, the method is called SC/SQ BG-SSRBSS, respectively. In the case of $g < L$, if FNG is adopted for BG, the algorithm is called SC/SQ FNG-SSRBSS. Similarly, if BD is adopted for BG, the algorithm is called SC/SQ BD-SSRBSS. It is worth noting that if g is set to L, it implies that no BG is carried out. In this case, the SC/SQ BG-SSRBSS degenerates to SC/SQ SSRBSS. In other words, SSRBSS is a special case of BG-SSRBSS.

Table 1 summarizes the names and detailed settings of the six BG-SSRBSS methods. The block diagram of performing BG-SSRBSS is shown in Figure 3.

Table 1. The names and definition of six BG-SSRBSS methods.

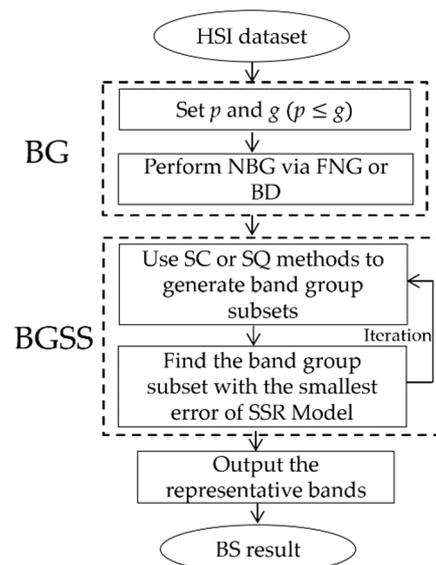| Method's Name	BG Method	Search Method for BGSS	Relationship of g, p, and L
SC FNG-SSRBSS	FNG	SC	$p \leq g < L$ (w/BG)
SC BD-SSRBSS	BD	SC	
SQ FNG-SSRBSS	FNG	SQ	
SQ BD-SSRBSS	BD	SQ	
SC SSRBSS	n/a	SC	$p < g = L$ (w/o BG)
SQ SSRBSS	n/a	SQ	

Figure 3. The block diagram of BG-SSRBSS.

5.3. Parameter Selection for p and g

Performing BG-SSRBSS only requires two parameters: p the number of bands to be selected) and g (the number of band groups to be partitioned). The parameter p can either be set manually or pre-determined by an auxiliary algorithm such as VD [40,41]. The parameter g can also be freely specified as long as the relationship $p \leq g < L$ holds. When FNG is used for BG, g can be directly specified by any desired value. It is noted that when BD is used for BG, g will be indirectly generated from the BD results performed with BD's parameter ϵ.

The value of g plays a vital role in our BG-SSRBSS. If g is set too small, it will significantly reduce the number of iterations required in BGSS so as to save computing time.

However, it may also lower the probability of finding the nearly global solution to the band subset due to the reduced number of band subset combinations. On the contrary, if g is set too large, this will result in too few bands in each group. It would reduce the performance of removing redundancy between the final selected bands and be unable to reduce the number of iterations required in BGSS. As a result, g must be set to a suitable value for each HSI dataset to maximize the performance of BS. Based on our experience, we will set the value of g to be 2 ~ 4 times the value of p in the experiment to produce the best results.

5.4. BG-SSRBSS Algorithms

We summarize the initialization and optimization procedure for the proposed SC/SQ BG-SSRBSS as Algorithms 1 and 2.

Algorithm 1 SC BG-SSRBSS

Input: A HSI cube with L bands $\Omega = \{B_1, B_2, \ldots, B_L\}$ with band matrix $\mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2 \ldots \mathbf{b}_L]$
Step 1: Initialization
1. Determine p and g. It must satisfy $p \leq g$.
2. Perform FNG or BD on Ω to generate band groups $\Phi_g = \{G_1, G_2, \ldots, G_g\}$.
3. Let $\Psi_p^{(0)} = \left\{ G_1^{(0)}, G_2^{(0)}, \cdots, G_p^{(0)} \right\}$ be the initial band group subset uniformly selected from Φ_g.
$\quad$ Set $\mathbf{P} = \left[\mathbf{G}_1^{(0)}\ \mathbf{G}_2^{(0)} \cdots \mathbf{G}_p^{(0)} \right]$ and calculate $E^{(0)} = E(\mathbf{P})$ via Equations (4) and (5).
Step 2: Outer loop
$\quad$ For $j = 1, \cdots, p$ do
$\quad\quad$ Set $\Psi_p^{(j)} = \Psi_p^{(j-1)}$
$\quad\quad$ Step 3: Inner Loop
$\quad\quad\quad$ For $l = 1, \cdots, g$ do
$\quad\quad\quad\quad$ If $G_l \notin \Psi_p^{(j-1)}$, set $\Psi_{temp} = \Psi_p^{(j-1)}$, $\Psi_{temp}(j) = G_l$
$\quad\quad\quad\quad\quad$ Set $\mathbf{P} = \Psi_{temp}$ and calculate $E(\mathbf{P})$ with Equations (4) and (5)
$\quad\quad\quad\quad\quad$ If $E(\mathbf{P}) < E^{(l-1)}$, set $\Psi_p^{(j)} = \Psi_{temp}$, and $E^{(l)} = E(\mathbf{P})$
$\quad\quad\quad\quad$ Else
$\quad\quad\quad\quad\quad$ $E^{(l)} = E^{(l-1)}$
Step 4: Set $\Psi_p^* = \Psi_p^{(p)}$ and calculate the p representative bands of the band groups in Ψ_p^* with Equation (5)
Output: Band subset $\Omega_p^* = \left\{ B_{G_1^{(p)}}^r, B_{G_2^{(p)}}^r, \cdots, B_{G_p^{(p)}}^r \right\}$

Algorithm 2 SQ BG-SSRBSS

Input: A HSI cube with L bands $\Omega = \{B_1, B_2, \ldots, B_L\}$ with band matrix B $= [\mathbf{b}_1\ \mathbf{b}_2 \ldots \mathbf{b}_L]$
Step 1: Same as Step 1 of Algorithm 1
Step 2: Outer loop
$\quad$ For $l = 1, \cdots, g$ do
$\quad$ If $G_l \notin \Psi_p^{(l-1)}$, set $G^* = G_l$
$\quad$ Step 3: Inner Loop
$\quad$ For $j = 1, \cdots, p$ do
$\quad$ Set $\Psi_{temp} = \Psi_p^{(l-1)}$, $\Psi_{temp}(j) = G^*$, $P = \Psi_{temp}$
$\quad$ Calculate $E(\mathbf{P})$ with Equations (4) and (5)
If $E(\mathbf{P}) < E^{(l-1)}$ set $\Psi_p^{(l)} = \Psi_{temp}$ and $E^{(l)} = E(\mathbf{P})$
$\quad$ Else
$\quad$ Set $E^{(l)} = E^{(l-1)}$ and $\Psi_p^{(l)} = \Psi_p^{(l-1)}$
Step 4: Set $\Psi_p^* = \Psi_p^{(g)}$ and calculate the p representative bands of the band groups in Ψ_p^* with Equation (5)
Output: Band subset $\Omega_p^* = \left\{ B_{G_1^{(g)}}^r, B_{G_2^{(g)}}^r, \cdots, B_{G_p^{(g)}}^r \right\}$

---

6. Experiments

This section first describes the real HSI datasets for experiments: University of Pavia, Purdue Indian Pines, and Salinas scene. Those datasets are available on the website [51]. Then, the experimental settings are described. Finally, the experimental results and discussion are presented.

6.1. HSI Datasets

The first dataset used for experiments is a real hyperspectral image that was collected by the ROSIS optical sensor over an urban area of the University of Pavia. This image is of the size 610×340 with very high spatial resolution, approximately 1.3 m per ground pixel. The original data contains 115 spectral bands with a spectral range from 0.43 μm to 0.86 μm. After removing the noisy bands, the remaining 103 bands are used for the experiments. This image includes nine classes of interest and one background class. Figure 4a–c show the image scene of band 80, the ground truth map, and the class labels of Pavia image, respectively.
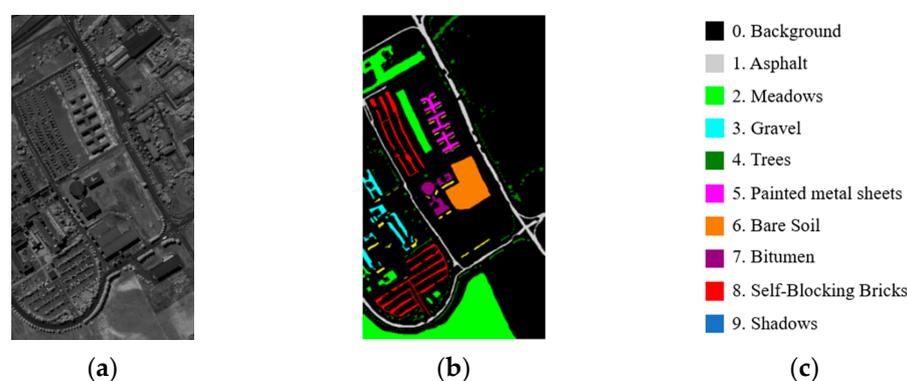

Figure 4. Pavia image scene. (**a**) Band 80th; (**b**) Ground truth map for nine classes; (**c**) Ground truth class label.

The second dataset used for experiments is the Purdue Indiana Indian Pine test site, which was collected by the AVIRIS sensor on 12 June 1992. It has 20 m spatial resolution and 10 nm spectral resolution in the range of 0.4–2.5 μm with a size of 145×145 pixel vectors taken from an area of mixed agriculture and forestry in Northwestern Indiana, USA. It was recorded on June 1992 with 220 bands. After removing water absorption bands (bands 104–108, 150–162), 202 bands are retained for analysis. Figure 5a–c show the image scene of band 20, the ground truth map, and the class labels of the Purdue image, respectively.
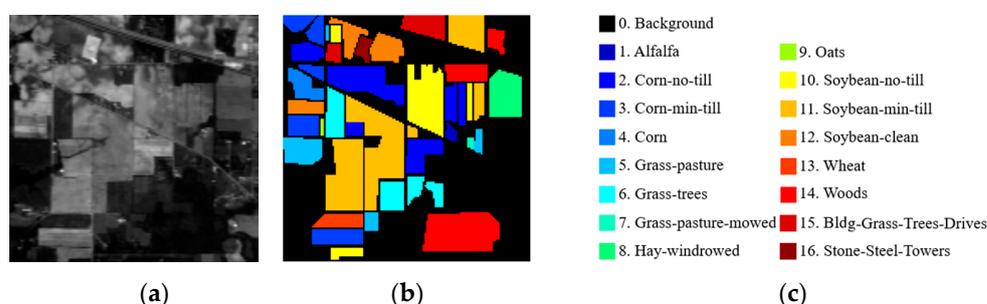
Figure 5. Purdue image scene. (**a**) Band 20th; (**b**) Ground truth map; (**c**) Ground truth class labels.

The third dataset used for experiments is the Salina scene, which was collected by the AVIRIS sensor at Salinas Valley, CA, USA, Its spatial resolution is 3.7 m and spectral resolution is 10 nm. The image size is 512×217 with 224 bands. There are 16 classes of plants or crops and 1 background class in the scene. Figure 6a–c show the image scene of

band 170, the ground truth map, and the class labels of Salinas image, respectively. In the following, we simply use Pavia, Purdue, and Salinas to refer to these images.
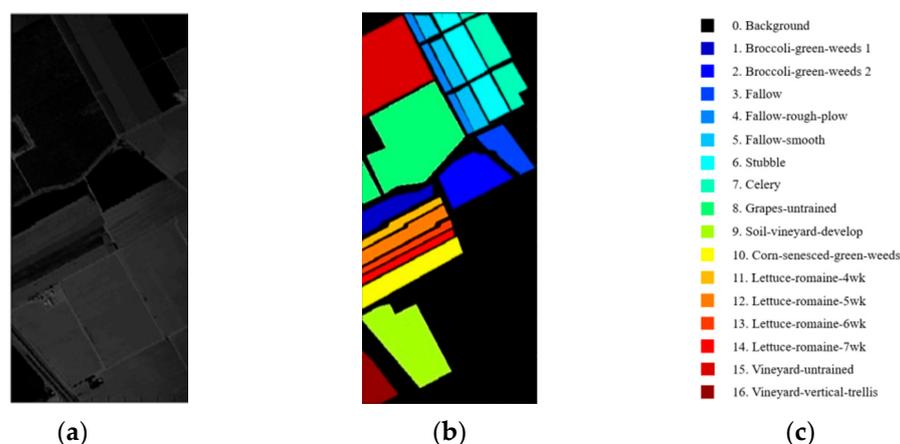

Figure 6. Salinas image scene. (**a**) Band 170th; (**b**) Ground truth map; (**c**) Ground truth class labels.

6.2. Expermental Setting

6.2.1. Parameter Setting

For all the BS methods used in the experiment, the p values for the three HSI datasets were determined by a VD algorithm, called noise whitened Harsanyi–Farrand–Chang (NWHFC) [40], with $P_F = 10^{-3}$. The p values for Pavia, Purdue, and Salinas datasets were 17, 18, and 21, respectively. For the BG-SSRBSS implemented with FNG, the g values for Pavia and Salinas data were set to be 3 times their p values, and the g value of the Purdue dataset was set to be 2 times its p value. When implementing BD, the SAM criterion was used, and the threshold ϵ was empirically set to 0.0227, 0.0132, and 0.054 to produce the corresponding g values for the three datasets. These g values were approximately 3–4 times the p values. Table 2 summarizes the parameters adopted in the experiments.

Table 2. The parameters of BG-SSRBSS used in the experiment.

	P	g (FNG)	g (BD)
Pavia data	17	51	68
Purdue data	18	54	72
Salinas data	21	42	63

6.2.2. Classifiers and Quantitative Metrics

After BS was done, the p selected bands were used for land cover classification to evaluate the information integrity. In the experiment, two classifiers were used for the evaluation of different BS algorithms: Linear SVM [52] and HybridSN [53]. The SVM is a pixel-wise classifier that classifies each spectral pixel independently without using spatial information. The HybridSN is a joint spatial–spectral classifier that utilizes the modules of both 3D-CNN and 2D-CNN simultaneously for better prediction. In our experiment, 10% samples of all classes were randomly selected from the image scene for training purposes, and the other 90% were used for test purposes.

Three quantitative metrics were used to evaluate the test results: overall accuracy (OA), average accuracy (AA), and Cohen's Kappa coefficient (Kappa). Those metrics are the typical quantitative measures used in the literature for HSI classification. The OA is calculated by dividing the number of correctly classified pixels by the number of total pixels. The AA is obtained by averaging the classification rate per class. The Kappa is calculated by both OA's value and the probability of random agreement. It is a measure

that can handle imbalanced data and multi-class problems. Theoretically, Kappa is more informative than OA when dealing with the topic of HSI classification.

6.2.3. State-of-the-Arts Methods for Comparative Study

Three types of BS algorithms were selected for the comparative study. The first one was uniform band selection (UBS), which is a widely used BS method since no prior knowledge is required. The bands were sampled with equal intervals $\left\lfloor \frac{L}{p} \right\rfloor$ or $\left\lceil \frac{L}{p} \right\rceil$ from 1st band to the last band. The second type was SQMBS, including OMP-BS [27] and PBS [6]. The PBS was implemented with the variance criterion for better classification performance. The third type was SMMBS, including two BSS methods, CCBSS [17] and LCMV-BSS [18], and FNGBS [47]. The CCBSS was implemented with SAM criterion. Additionally, the result of using full bands (i.e., no BS is implemented) was also considered in the experiments.

6.2.4. Computing Environment

In the experiments, the generation of training and test samples is repeated ten times, and the averaged results are reported. All the experiments were implemented on the hardware with Intel i7-10700 CPU, RTX 3080 GPU, and 80 GB RAM. The BS algorithms and SVM were implemented on Matlab 2021b, and HybridSN was implemented on Jupyter-notebook 6.3.0 with Python 3.6.

6.3. BS Results

Tables 3–5 list the bands selected by different BS methods: UBS, OMP-BS, PBS, FNGBS, SC/SQ CCBSS, SC/SQ LCMV-BSS, and the proposed SC/SQ SSRBSS, SC/SQ FNG-SSRBSS, and SC/SQ BD-SSRBSS, for three HSI datasets, respectively. In each table, the upper part shows the results of the compared methods, and the lower part shows the results of the proposed methods. In the results of four BG-SSRBSS methods, the brace {a–b} followed by each band index indicates the range of the corresponding band group generated by BG. Based on the tables, there are several key observations:

1. The results of CCBSS, PBS, and LCMV-BSS have two issues: selecting adjacent bands and selecting the bands in specific spectral regions. For example, in Table 3, bands 98–103 were selected by PBS, bands 50–57 and 88–96 were selected by SC CCBSS, and bands 1–5, 60–61, and 88–96 were selected by SQ LCMV-BSS. Similar phenomenon can also be found in Tables 4 and 5.
2. The first issue was significantly alleviated in SC/SQ SSRBSS and OMP-BS. Their selected bands were distributed more uniformly in the whole spectrum. However, there are still cases where the bands in a certain range were ignored to be selected. For example, in Table 4, bands 125–188 were missing from the SC SSRBSS result, while bands 123–190 were missing from the SQ SSRBSS result. A similar phenomenon can also be found in Table 5.
3. The FNG-SSRBSS and BD-SSRBSS methods seemed to overcome both issues. Not only did they reduce the probability of selecting adjacent bands, but they ensured that each segment of the spectrum could generate at least one band for better information integrity.
4. The BG results of FNG-SSRBSS and BD-SSRBSS were obviously different. The group size of the band groups produced by FNG tended to be consistent, while the group size of the BD-generated band groups varied greatly. For instance, in the results for SQ BD-SSRBSS in Table 5, there are three band groups consisting of only 1 band: {3}, {41}, and {222}, and one group containing 29 bands: {190–218}. On the contrary, the group size produced by FNG consistently ranged from 2 to 6. This is due to the inherent nature of each BG algorithm.

Table 3. The bands selected by the proposed and the compared methods for the Pavia dataset.

Data	Method	Selected Bands
	UBS	1, 7, 13, 19, 25, 31, 37, 43, 49, 55, 61, 67, 73, 79, 85, 91, 103
	OMP-BS [27]	1, 2, 4, 6, 9, 13, 24, 33, 42, 54, 66, 74, 81, 83, 86, 94, 100
	PBS [6]	1, 27, 37, 43, 51, 52, 64, 83, 89, 91, 94, 95, 98, 100, 101, 102, 103
	FNGBS [47]	5, 12, 19, 22, 30, 32, 42, 49, 56, 61, 63, 74, 79, 81, 88, 92, 99
	SC CCBSS [17]	50, 51, 52, 53, 54, 55, 56, 57, 88, 89, 90, 91, 92, 93, 94, 95, 96
	SQ CCBSS [17]	15, 16, 17, 18, 19, 20, 21, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96
	SC LCMV-BSS [18]	1, 2, 3, 4, 5, 6, 48, 55, 68, 69, 81, 87, 89, 91, 98, 100, 103
	SQ LCMV-BSS [18]	1, 2, 3, 4, 5, 59, 60, 61, 88, 89, 90, 91, 92, 93, 94, 95, 96
Pavia (16 bands)	SC SSRBSS	1, 2, 4, 5, 7, 10, 13, 24, 33, 44, 54, 64, 71, 81, 85, 97, 102
	SQ SSRBSS	1, 2, 3, 5, 8, 12, 20, 31, 44, 54, 64, 73, 82, 83, 85, 94, 101
	SC FNG-SSRBSS	2{1–3}, 4{4,5}, 6{6,7}, 10{10,11}, 15{14–17}, 22{22,23}, 29{28–30}, 39{38–40}, 48{48,49}, 54{54,55}, 62{62,63}, 72{72,73}, 76{76,77}, 83{82–84}, 89{88–90}, 97{97,98}, 102{102,103}
	SQ FNG-SSRBSS	2{1–3}, 4{4,5}, 6{6,7}, 8{8,9}, 15{14–17}, 31{31}, 44{44,45}, 54{54,55}, 60{60,61}, 68{68,69}, 74{74,75}, 83{82–84}, 85{85}, 86{86,87}, 95{94–96}, 99{99}, 102{102,103}
	SC BD-SSRBSS	1{1,2}, 4{4}, 6{6}, 9{9}, 13{13}, 19{19,20}, 31{31,32}, 43{43,44}, 47{47,48}, 55{55,56}, 66{66}, 73{73}, 79{79}, 82{82,83}, 85{85}, 94{92–96}, 101{100–102}
	SQ BD-SSRBSS	1{1,2}, 4{4}, 6{6}, 8{8}, 11{11}, 19{19,20}, 31{31,32}, 41{41,42}, 53{53,54}, 66{66}, 74{74}, 80{80,81}, 82{82,83}, 85{85}, 94{92–96}, 101{100–102}, 103{103}

Table 4. The bands selected by the proposed and the compared methods for the Purdue dataset.

Data	Method	Selected Bands
	UBS	1, 13, 25, 37, 49, 61, 73, 85, 97, 109, 121, 133, 145, 157, 169, 181, 193, 202
	OMP-BS [27]	1, 2, 3, 4, 6, 9, 19, 29, 34, 38, 42, 50, 68, 81, 97, 113, 131, 189
	PBS [6]	1, 7, 10, 26, 34, 44, 46, 48, 57, 65, 66, 85, 87, 92, 107, 157, 166, 196
	FNGBS [47]	9, 15, 28, 43, 49, 59, 66, 83, 97, 107, 118, 129, 138, 157, 165, 173, 181, 190
	SC CCBSS [17]	45, 46, 47, 48, 49, 50, 51, 52, 53, 155, 156, 160, 161, 162, 163, 164, 165, 166
	SQ CCBSS [17]	10, 11, 12, 13, 14, 15, 16, 17, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53
	SC LCMV-BSS [18]	3, 6, 11, 24, 41, 70, 103, 143, 144, 145, 166, 190, 192, 193, 194, 195, 198, 200
	SQ LCMV-BSS [18]	13, 63, 65, 66, 69, 70, 72, 75, 89, 103, 119, 123, 168, 169, 171, 195, 197, 198
Purdue (17 bands)	SC SSRBSS	1, 2, 3, 4, 12, 15, 25, 32, 34, 37, 41, 42, 52, 68, 90, 98, 124, 189
	SQ SSRBSS	1, 2, 3, 4, 7, 8, 18, 30, 33, 35, 38, 42, 53, 62, 72, 90, 122, 191
	SC FNG-SSRBSS	5{1–10}, 14{13–15}, 18{16–19}, 21{20–23}, 26{24–27}, 33{31–34}, 36{35–37}, 39{38–41}, 44{42–45}, 47{46–48}, 54{49,58}, 65{64–68}, 74{72–76}, 86{83–87}, 88{88–92}, 95{95–99}, 123{121–125}, 154{149–155}
	SQ FNG-SSRBSS	5{1–10}, 14{13–15}, 26{24–27}, 33{31–34}, 36{35–37}, 39{38–41}, 44{42–45}, 47{46–48}, 62{62–63}, 74{72–76}, 93{93,94}, 119{117–120}, 123{121–125}, 126{126,127}, 147{146–148}, 156{156,157}, 159{158–160}, 192{190–193}
	SC BD-SSRBSS	1{1,2}, 3{3}, 4{4}, 5{5}, 9{9}, 12{12,13}, 16{16,17}, 27{27,28}, 32{32}, 35{35}, 38{38}, 44{42–45}, 51{46–51}, 69{66–71}, 85{83–93}, 97{96–98}, 123{121–124}, 193{187–202}
	SQ BD-SSRBSS	1{1,2}, 3{3}, 4{4}, 5{5}, 12{12,13}, 16{16,17}, 27{27,28}, 33{33}, 35{35}, 38{38}, 44{42–45}, 53{52–54}, 74{72–76}, 85{83–93}, 97{96–98}, 123{121–124}, 149{149}, 193{187–202}

Table 5. The bands selected by the proposed and the compared methods for the Salinas dataset.

Data	Method	Selected Bands
	UBS	1, 12, 23, 34, 45, 56, 67, 78, 89, 100, 111, 122, 133, 144, 155, 166, 177, 188, 199, 210, 224
	OMP-BS [27]	1, 2, 3, 4, 8, 14, 19, 23, 31, 34, 37, 39, 42, 50, 66, 72, 104, 121, 126, 152, 198
	PBS [6]	1, 20, 60, 197, 203, 204, 207, 208, 209, 210, 211, 213, 214, 215, 216, 217, 219, 221, 222, 223, 224
	FNGBS [47]	7, 15, 31, 38, 55, 62, 67, 83, 92, 96, 118, 122, 137, 147, 152, 166, 172, 187, 196, 212, 217
	SC CCBSS [17]	1, 2, 3, 4, 5, 10, 18, 26, 34, 37, 39, 42, 48, 57, 70, 76, 85, 134, 152, 170, 184
	SQ CCBSS [17]	26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52
	SC LCMV-BSS [18]	27, 44, 54, 61, 64, 74, 80, 88, 94, 104, 124, 134, 144, 154, 164, 174, 184, 194, 204, 214, 216
	SQ LCMV-BSS [18]	1, 2, 6, 24, 27, 28, 34, 44, 46, 54, 74, 84, 104, 114, 134, 144, 164, 174, 184, 194, 204
Salinas (21 bands)	SC SSRBSS	1, 2, 3, 4, 5, 10, 18, 26, 34, 37, 39, 42, 48, 57, 70, 76, 85, 134, 152, 170, 184
	SQ SSRBSS	1, 2, 3, 4, 5, 9, 14, 20, 28, 35, 38, 41, 46, 55, 67, 76, 83, 92, 134, 170, 184
	SC FNG-SSRBSS	3{1–4}, 6{5–9}, 12{10–17}, 21{18–22}, 33{28–35}, 37{36–39}, 41{40–42}, 47{43–48}, 58{55–61}, 68{66–70}, 78{76–80}, 94{92–95}, 133{130–136}, 142{141–144}, 152{151–153}, 165{162–167}, 175{171–176}, 185{182–188}, 193{193–198}, 201{199–203}, 215{215–218}
	SQ FNG-SSRBSS	3{1–4}, 6{5–9}, 12{10–17}, 21{18–22}, 33{28–35}, 37{36–39}, 41{40–42}, 47{43–48}, 53{49–54}, 58{55–61}, 68{66–70}, 78{76–80}, 94{92–95}, 133{130–136}, 152{151–153}, 169{168–170}, 175{171–176}, 178{177–181}, 185{182–188}, 193{193–198}, 215{215–218}
	SC BD-SSRBSS	1{1,2}, 3{3}, 4{4}, 7{5–7}, 9{8–11}, 21{19–22}, 33{32–37}, 40{40}, 51{44–55}, 75{67–83}, 89{88–103}, 115{115,116}, 130{127–132}, 152{152}, 162{162}, 165{165}, 168{168}, 170{170}, 173{171–174}, 203{190–218}, 222{222}
	SQ BD-SSRBSS	1{1,2}, 3{3}, 4{4}, 7{5–7}, 15{12–18}, 21{19–22}, 29{27–31}, 33{32–37}, 38{38}, 41{41}, 51{44–55}, 60{56–62}, 75{67–83}, 89{88–103}, 130{127–132}, 152{152}, 162{162}, 165{165}, 173{171–174}, 203{190–218}, 222{222}

6.4. Classification Results

This section shows the HSI classification results of using two different classifiers performed on the selected bands of various BS methods.

6.4.1. SVM Results

In this section, we adopted the bands selected by UBS, OMP-BS, PBS, FNGBS, SQ CCBSS, SC LCMV-BSS, and the proposed six BG-SSRBSS methods for image classification on the three HSI datasets. Both quantitative and qualitative analysis is presented.

Tables 6–8 present the SVM classification results of using the bands listed in Table 3, Table 4, and Table 5, respectively, for Pavia, Purdue, and Salinas datasets. The bold values in each table represents the highest accuracies in the corresponding class. From Table 6, it can be observed that in addition to FNGBS, SQ CCBSS, and SC LCMV-BSS, which only produced 75.66% and 84.28% in OA, the other BS methods, UBS, OMP-BS, PBS, and SC/SQ SSRBSSs could produce 88–90% in OA. We speculate that this is due to their selected bands being more complementary in the spectrum. Among all BS methods, the proposed SQ BS-SSRBSS-2 produced the highest accuracies in OA, AA, and Kappa: 90.57%, 86.3%, and 87.17%. In general, the Pavia data are relatively simpler data for classification. Using the selected bands provided by most BS methods could provide satisfactory classification performance.

Compared to the Pavia dataset, the Purdue dataset is a heavily-mixed image and was considered much more difficult to be classified in the literature. The performance difference between various BS methods can be observed more through the experiments classifying Purdue data. From Table 7, it is found that the OA range is from 60% to 80%. The OA values generated by UBS, PBS, FNGBS, CQ CCBSS, and SC LCMV-BSS are 74.8%, 74.45%, 78.52%, 60.83%, and 62.82%, respectively. Using the bands selected by the proposed SC/SQ SSRBSS, SC/SQ FNG-SSRBSS, and SC/SQ BD-SSRBSS achieved 76–80% in OA. The highest OA, AA, and Kappa values were generated by SQ FNG-SSRBSS: 80.18%, 73.73%, and 77.4%. Additionally, most of the best single-class values were produced by our SSRBSS methods. Similar to Pavia's classification results shown in Table 6, the performance of SQ CCBSS and SC LCMV-BSS was relatively lower in the comparison. It verifies that the bands selected

by the proposed BG-SSRBSS methods contain a larger amount of spectral information to identify different substances.

Table 6. SVM classification results of using the bands described in Table 3 for the Pavia dataset.

Class	Full Bands (Ref)	UBS	OMP-BS	PBS	FNGBS	SQ CCBSS	SC LCMV-BSS	SC/SQ SSRBSS	SC/SQ FNG-SSRBSS	SQ BD-SSRBSS
1	91.11	**88.05**	85.79	81.66	85.7	74.95	75.73	84.18/85.38	87.18/87.82	87.99/87.24
2	94.45	90.35	91.74	88.99	85.9	70.66	86.06	91.53/90.85	90.64/87.79	**92.1/92.1**
3	88.75	84.08	**85.51**	84.18	84.18	76.56	68.69	85.27/85.08	84.8/**85.51**	83.99/84.65
4	97.35	96.5	**97.06**	**97.06**	96.05	92.95	95.98	96.6/96.6	96.54/96.08	96.21/96.89
5	99.92	**99.92**	**99.92**	**99.92**	99.85	99.7	**99.92**	**99.92/99.2**	**99.92/99.92**	**99.92/99.92**
6	94.81	92.5	90.69	87.27	88	79.26	87.67	92.1/91.8	90.69/89.83	91.01/90.89
7	95.18	93.98	94.13	94.13	94.66	93.3	92.63	93.45/93.9	93.75/94.51	94.73/93.98
8	89.08	84.27	85.06	83.81	84.43	74.33	79.79	84.51/86.04	85.14/84.76	84.54/85.63
9	99.89	99.89	**100**	**100**	99.89	**100**	99.89	99.89/100	99.89/99.89	**100/100**
OA	93.36	89.98	90.31	87.86	86.76	75.66	84.28	89.95/89.95	89.68/88.29	90.53/**90.57**
AA	89.77	86.06	86.01	83.63	83.79	74.86	79.22	85.02/85.42	85.75/85.1	86.28/**86.3**
Kappa	90.93	86.43	86.83	83.59	82.21	68.35	78.93	86.37/86.39	86/84.21	87.11/**87.17**

Table 7. SVM classification results of using the bands described in Table 4 for the Purdue dataset.

Class	Full Bands (Ref)	UBS	OMP-BS	PBS	FNGBS	SQ CCBSS	SC LCMV-BSS	SC/SQ SSRBSS	SC/SQ FNG-SSRBSS	SC/SQ BD-SSRBSS
1	92.59	83.33	87.03	87.03	85.18	74.07	81.48	87.03/87.03	88.88/**90.74**	88.33/88.88
2	74.68	64.43	70.22	62.62	70.92	35.56	50.76	72.66/69.87	72.45/**73.7**	71.74/71.68
3	76.61	60.19	63.3	64.02	69.9	38.6	51.31	**73.98**/68.58	67.74/73.14	65.1/68.22
4	93.58	88.88	92.3	90.59	92.73	73.93	83.33	88.03/88.46	90.17/**94.01**	91.45/91.88
5	87.92	83.09	88.73	87.92	84.5	79.67	75.65	90.34/80.34	90.34/90.94	89.53/90.14
6	93.17	87.28	87.28	89.82	91.96	83.8	72.42	89.69/89.82	90.36/**91.56**	89.15/89.02
7	92.3	**92.3**	**92.3**	**92.3**	**92.3**	88.46	80.76	**92.3/92.3**	**92.3**/88.46	**92.3/92.3**
8	96.31	95.7	**96.31**	96.11	95.5	91.41	94.68	92.84/**96.31**	**96.31**/95.5	94.88/94.88
9	100	95	**100**	90	**100**	75	65	95/95	**100/100**	85/90
10	83.57	73.76	76.44	74.27	82.64	66.52	67.97	75.72/76.44	77.27/83.57	69.73/71.69
11	72.08	68.51	71.51	69.85	69.12	57.86	59.35	66.08/67.098	67.94/73.98	65.35/66.93
12	79.47	82.08	77.85	65.79	**85.17**	52.76	53.09	73.61/76.22	83.38/80.29	78.5/78.5
13	99.05	96.22	97.64	96.22	96.69	92.45	88.67	97.16/96.69	98.58/99.52	97.64/98.58
14	93.04	92.73	94.35	93.81	93.74	87.17	81.83	92.73/94.12	92.96/92.58	**94.51**/93.43
15	73.42	66.84	57.36	48.68	68.94	34.21	37.1	66.05/68.42	**72.1**/68.94	58.42/69.47
16	96.84	95.78	**97.89**	**97.89**	96.84	**97.89**	96.84	**97.89/97.89**	96.84/**97.89**	**97.89/97.89**
OA	79.85	74.8	77.15	74.45	78.52	60.83	62.82	76.1/76.53	77.72/**80.18**	75.56/76.34
AA	74.73	69.02	68.59	68.19	71.54	55.35	55.02	68.04/69.02	72.31/**73.73**	67.69/68.84
Kappa	77.05	71.26	73.89	70.86	75.52	55.52	57.85	72.82/73.3	74.62/**77.4**	72.19/73.1

Among three datasets, the Salinas dataset is the one relatively easier to be classified. From Table 8, it can be seen that except for class 8 and class 15, the accuracy values of the other 14 classes are as high as 98–99%. In addition to PBS and SQ CCBSS, which only provided 85.71% and 87.25% in OA, the other BS methods achieved 90–91% in OA. Similarly, the proposed SQ FNG-SSRBSS achieved the highest overall accuracy of 91.7%. At the same time, the proposed SSRBSS methods produced the best single-class accuracies in classes 1, 2, 4–8, and 10–17.

Table 8. SVM classification results of using the bands described in Table 5 for the Salinas dataset.

Class	Full Bands (Ref)	UBS	OMP-BS	PBS	FNGBS	SQ CCBSS	SC LCMV-BSS	SC/SQ SSRBSS	SC/SQ FNG-SSRBSS	SC/SQ BD-SSRBSS
1	99.5	99.3	99.5	99.35	99.95	98.5	99.3	99.4/99.6	99.4/**99.7**	99.05/99.4
2	99.7	99.4	99.67	99.59	100	98.87	99.73	99.51/99.51	99.89/**99.97**	99.89/99.46
3	99.64	99.49	99.54	98.27	**99.84**	97.06	99.79	99.19/99.54	99.69/99.29	98.98/98.98
4	99.42	99.21	99.49	99.56	99.56	99.35	98.99	**99.64**/99.42	99.21/99.35	99.28/99.35
5	98.99	98.73	98.35	98.58	98.91	96.37	98.91	97.57/98.05	**99.17**/98.87	98.99/98.35
6	99.77	99.82	99.82	99.82	99.82	99.41	99.82	99.87/99.82	99.84/99.84	**99.92**/99.82
7	99.88	99.66	**99.91**	99.13	99.88	99.55	99.8	99.91/99.88	**99.91/99.91**	99.83/99.88
8	79.55	78.71	79.59	63.34	79.28	71.2	79.3	79.38/80.33	78.22/79.51	**80.25/80.71**
9	99.04	99.32	99.01	98.51	99.17	96.01	99	99.24/**99.38**	99/99.82	99.17/99.11
10	94.81	96.21	94.96	88.74	95.63	91.36	95.33	95.02/95.3	**96.06**/95.79	95.14/94.56
11	99.71	99.53	99.06	98.4	**99.9**	98.31	99.71	99.43/99.81	99.53/99.62	99.34/99.53
12	99.58	99.37	99.63	99.06	99.79	99.16	**99.89**	99.84/99.68	99.74/**99.89**	99.68/99.58
13	99.56	99.67	**99.89**	98.47	99.67	99.67	99.45	99.78/**99.89**	99.78/**99.89**	99.67/99.56
14	97.75	97.66	99.15	92.99	**99.71**	98.31	98.13	99.53/99.62	99.25/99.34	97.85/99.43
15	74.18	73.21	76.85	67.94	75.85	68.94	75.49	76.47/75.56	78.9/**79.26**	68.87/72.53
16	99.44	99.33	99.39	97.5	**99.5**	98.83	99.39	99.39/99.33	99.44/99.44	99.39/99.39
OA	90.8	90.61	91.26	85.71	91.26	87.25	91.05	91.15/91.34	91.4/**91.7**	90.53/90.57
AA	94.98	95.23	95.02	90.61	95.64	91.44	95.24	95.12/95.66	**95.69/95.69**	86.28/86.3
Kappa	89.67	89.45	90.19	84.01	90.18	85.7	89.94	90.07/90.27	90.34/**90.69**	87.11/87.17

From the above-mentioned observations, it is concluded that the proposed BG-SSRBSS methods outperformed the compared BS methods. It is worth noting that using BG-SSRBSS's selected bands could even produce a better classification performance than using full bands in the Purdue and Salinas experiments.

6.4.2. HybridSN Results

The classification performance via a spatial–spectral classifier for various BS methods is another focus of observation. Tables 9–11 present the HybridSN classification results of using the bands listed in Table 3, Table 4, and Table 5, respectively, for the Pavia, Purdue, and Salinas datasets. By virtue of the information integration capability through spatial and spectral dimensions of CNN, the classification performance on the Pavia and Salinas datasets was close to 100%. Simultaneously, the classification results of the Purdue data were also greatly improved to 98%. Almost all BS methods reached nearly the same classification performance as using full bands. Despite the powerful classification capabilities provided by HybridSN, the proposed SSRBSS methods still slightly outperform the other ones. It was found that the best OA in Table 11 was obtained by using SC FNG-SSRBSS (99.77%), the best OA in Table 10 was obtained by using SQ SSRBSS, and the best OA in Table 11 was obtained by SC BD-SSRBSS.

Table 9. HybridSN classification results of using the bands described in Table 3 for the Pavia dataset.

	Full Bands (Ref)	UBS	OMP-BS	PBS	FNGBS	SQ CCBSS	SC LCMV-BSS	SC/SQ SSRBSS	SC/SQ FNG-SSRBSS	SC/SQ BD-SSRBSS
OA	99.46	99.71	99.71	99.63	99.7	99.45	99.53	99.58/99.69	**99.77**/99.67	99.49/99.69
AA	99.18	99.61	99.54	99.43	99.57	99.18	99.32	99.43/99.52	**99.96**/99.59	99.3/99.53
Kappa	99.29	99.62	99.62	99.52	99.61	99.28	99.38	99.45/99.59	**99.69**/99.57	99.33/99.6

Table 10. HybridSN classification results of using the bands described in Table 4 for the Purdue dataset.

	Full Bands (Ref)	UBS	OMP-BS	PBS	FNG-BS	SQ CCBSS	SC LCMV-BSS	SC/SQ SSRBSS	SC/SQ FNG-SSRBSS	SC/SQ BD-SSRBSS
OA	96.77	96.49	98.48	96.57	97.82	95.12	97.67	98.27/**98.63**	98.12/98.28	98.36/98.31
AA	95.42	95.17	97.51	94.25	95.73	93.59	97.37	**97.82**/97.46	96.22/96.29	96.81/96.35
Kappa	96.32	96	98.27	96.09	97.52	94.43	97.35	98.03/**98.44**	97.86/98.04	98.13/98.08

Table 11. HybridSN classification results of using the bands described in Table 5 for Salinas dataset.

	Full Bands (Ref)	UBS	OMP-BS	PBS	FNG-BS	SQ CCBSS	SC LCMV-BSS	SC/SQ SSRBSS	SC/SQ FNG-SSRBSS	SC/SQ BD-SSRBSS
OA	99.94	99.95	99.94	99.9	99.96	99.9	99.84	99.95/99.83	99.94/99.96	**99.97**/99.96
AA	99.92	99.87	99.88	99.89	99.94	99.88	99.85	99.92/99.74	99.85/99.92	**99.97**/99.91
Kappa	99.93	99.94	99.94	99.89	99.96	99.89	99.82	99.94/99.48	99.94/99.95	**99.97**/99.96

In order to see the difference in using various BS approaches by visual assessment, Figure 7 shows the HybridSN classification maps of the Pavia dataset by using full bands and the bands selected by various BS methods. Although those maps are similar, the advantages of BG-SSRBSS can be seen in tiny details. For instance, in the maps of UBS and SQ CCBSS, Figure 7b,e, several pixels in the prairie below the scene were misclassified as bare soil or trees instead of meadows. In the cropped image, a few pixels of gravel (cyan) were misclassified as self-blocking bricks (red) in the maps of SQ CCBSS and SC LCMV-BSS. These tiny classification defects were relatively absent from the results of the SQ SSRBSS and SC FNG-SSRBSS methods shown in Figure 7h,i. Similarly, Figure 8 shows the HybridSN classification maps of Purdue image by using full bands and the bands selected by various BS methods. There are several obvious classification flaws could be found in the maps of the compared BS methods, for example, classes 3, 12 in Figure 8a; classes 2, 4, 5 in Figure 8b; classes 3, 4, 11 in Figure 8d; classes 2, 3, 11, 14, in Figure 8e; and classes 2, 4, 11 in Figure 8f. In contrast, there are significantly fewer misclassification pixels in BG-SSRBSS's maps shown in Figure 8g–i. Finally, Figure 9 shows the HybridSN classification maps of the Salinas dataset by using full bands and the bands selected by various BS methods. In the maps of the compared BS methods (Figure 9b–f), it can be observed that a few pixels located in class 8 (Grape-untrained) and class 15 (Vineyard-untrained) were misclassified. Surprisingly, such an issue is milder in the proposed BG-SSRBSS's maps shown in Figure 9g–i. In conclusion, the result of visual assessment also verifies the strength of the proposed BG-SSRBSS.

6.5. Discussion

In addition to the quantitative analysis and visual comparison presented in Section 6.4, there are additional findings from Tables 3–11 worthy of mentioning:

1. According to Tables 3–8, the BS methods that utilize sparse self-representation as objective function, OMP-BS and SSRBSSs, significantly outperformed the other ones. It implies that the SSR model is indeed an ideal objective function for BS in hyperspectral image classification.

2. According to Tables 3–8, the BS methods that utilize sparse self-representation as objective function, OMP-BS and SSRBSSs, significantly outperformed the other ones. It implies that the SSR model is indeed an ideal objective function for BS in hyperspectral image classification.

3. The classification performance of using full bands was usually not the best. This shows that the excessive redundant information in full bands could interfere with the performance of the classifier due to the curse of dimensionality.

4. Among all BSS methods, the proposed BG-SSRBSSs significantly outperformed SQ CCBSS and SC LCMV-BSS, particularly in Purdue's experiment. This implies that SSR

is a more suitable objective function than CC or LCMV for selecting the bands useful for classification.

5. According to the results of six BG-SSRBSSs, the classification accuracy of using SC and SQ search methods are quite similar, even though their final selected band groups are slightly different. This suggests that both of them could find good local optimal solutions.
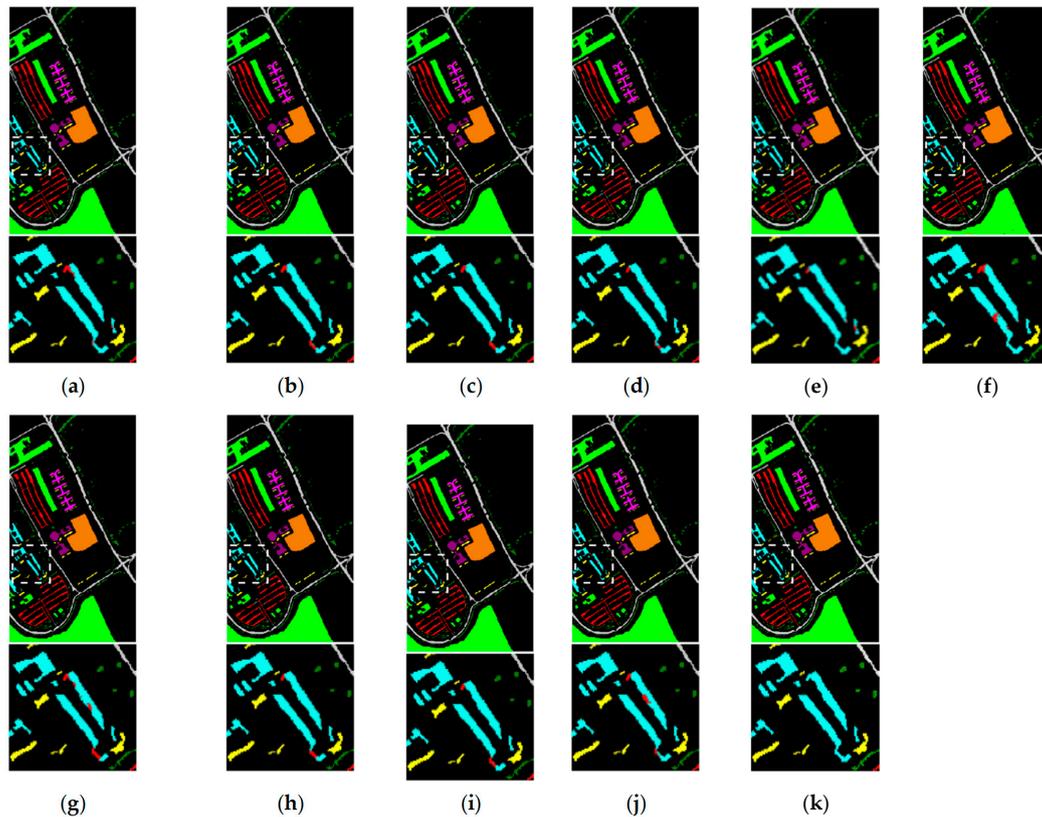

Figure 7. Classification maps generated by HybridSN using different BS methods for the Pavia dataset: (**a**) Full bands; (**b**) UBS; (**c**) OMP-BS; (**d**) PBS; (**e**) FNGBS; (**f**) SQ CCBSS; (**g**) LCMV-BSS; (**h**) SQ SSRBSS; (**i**) SC FNG-SSRBSS; (**j**) SQ BD-SSRBSS; (**k**) Ground truth.
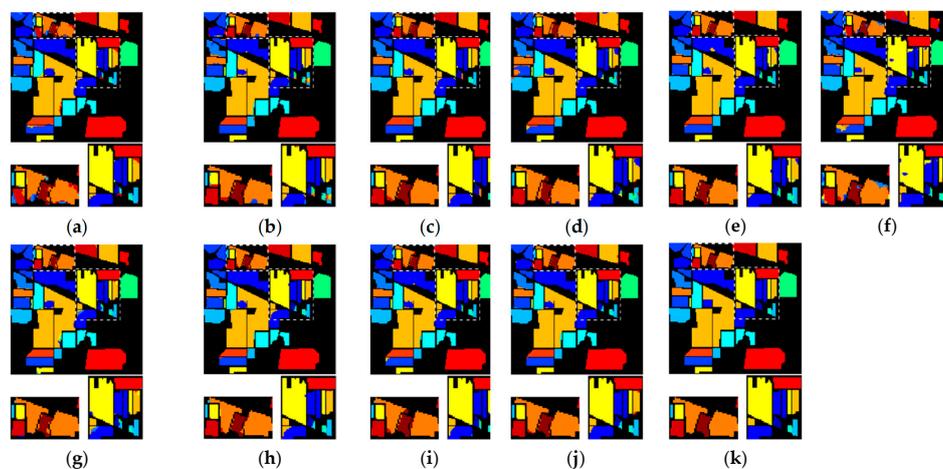
Figure 8. Classification maps generated by HybridSN using different BS methods for the Purdue dataset: (**a**) Full bands; (**b**) UBS; (**c**) OMP-BS; (**d**) PBS; (**e**) FNGBS; (**f**) SQ CCBSS; (**g**) LCMV-BSS; (**h**) SQ SSRBSS; (**i**) SQ FNG-SSRBSS; (**j**) SC BD-SSRBSS; (**k**) Ground truth.
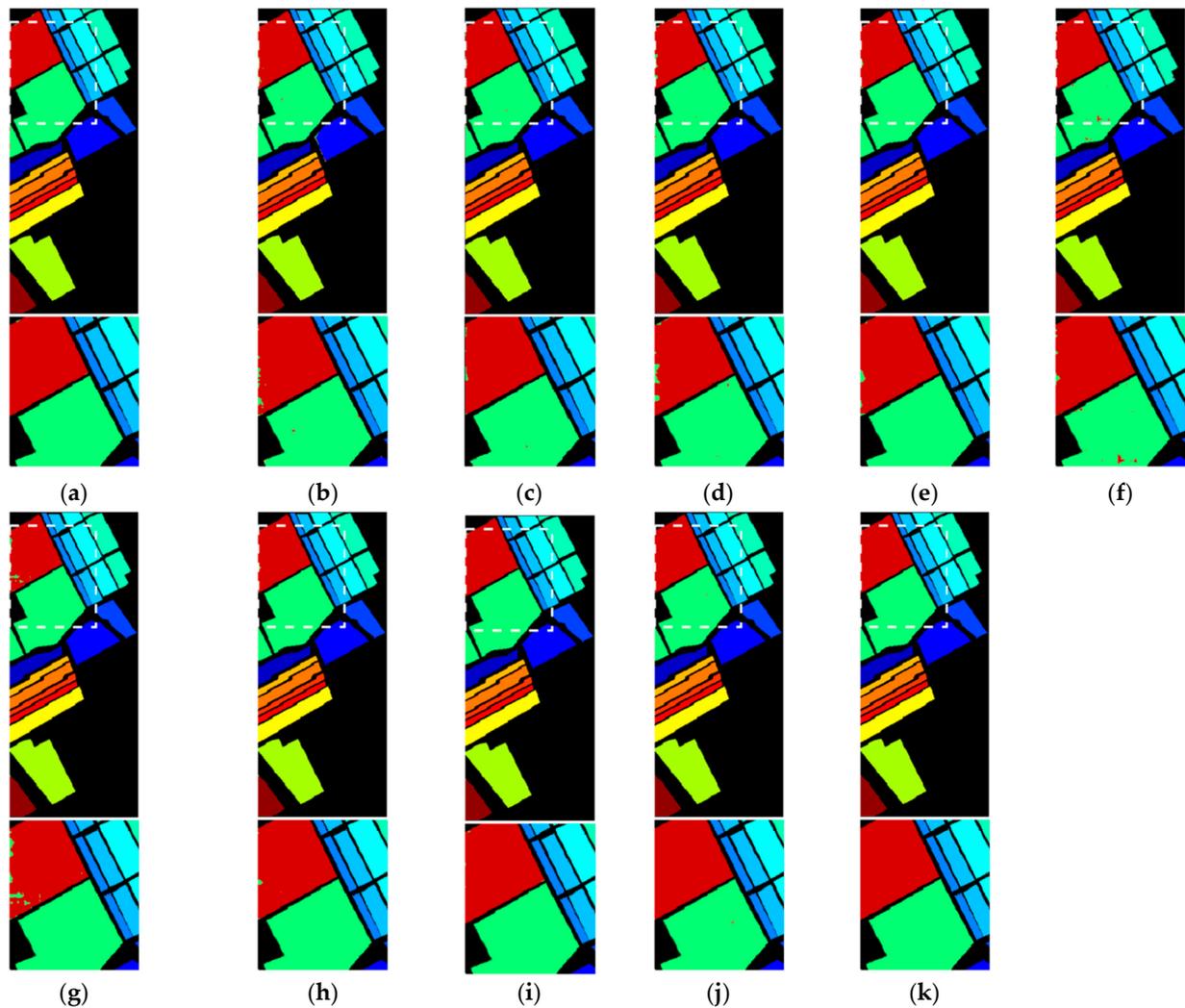
Figure 9. Classification maps generated by HybridSN using different BS methods for the Salinas dataset: (**a**) Full bands; (**b**) UBS; (**c**) OMP-BS; (**d**) PBS; (**e**) FNGBS; (**f**) SQ CCBSS; (**g**) LCMV-BSS; (**h**) SC SSRBSS; (**i**) SQ FNG-SSRBSS; (**j**) SQ BD-SSRBSS; (**k**) Ground truth.

As we know, the BG plays an important role for BG-SSRBSS. Here we further investigate the NBG performance of two BG methods. Figures 10–12 demonstrate the FNG and BD results of the Pavia, Purdue, and Salinas datasets, respectively. In each figure, the *x*-axis denotes band index, the *y*-axis denotes spectral reflectance, and the vertical blue line indicates the boundary of two adjacent band groups. Each spectral curve represents the spectral signature of the class with the same color in the corresponding ground truth map shown in Figures 4–6. As we can see, the BG results of using FNG and BD are quite different. The FNG seems to segment the spectrum more uniformly, while the BD tends to segment more in the regions whose spectral values of different classes are lower or closed. However, after the selection done by BGSS, the resulting bands made no obvious difference in classification performance. This shows that the given *g* values are large enough so that all BG-SSRBSSs could find the ideal band group combinations for effective classification.
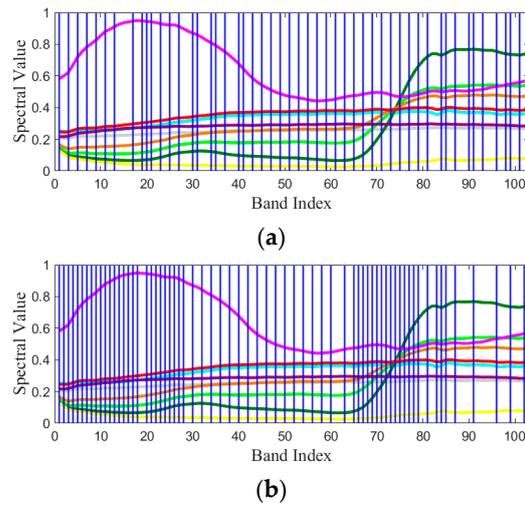
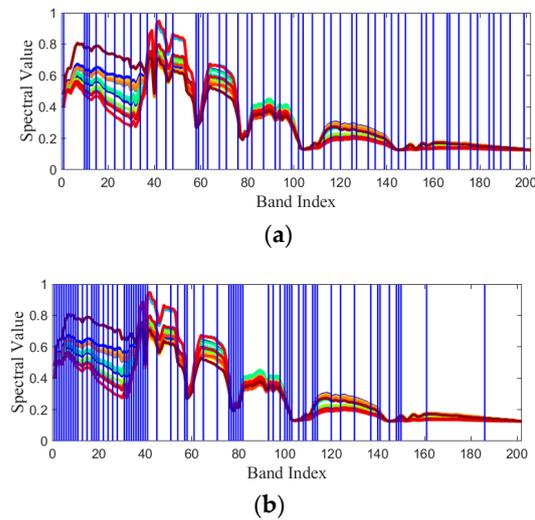**Figure 10.** The BG results for the Pavia experiment. (**a**) FNG; (**b**) BD.

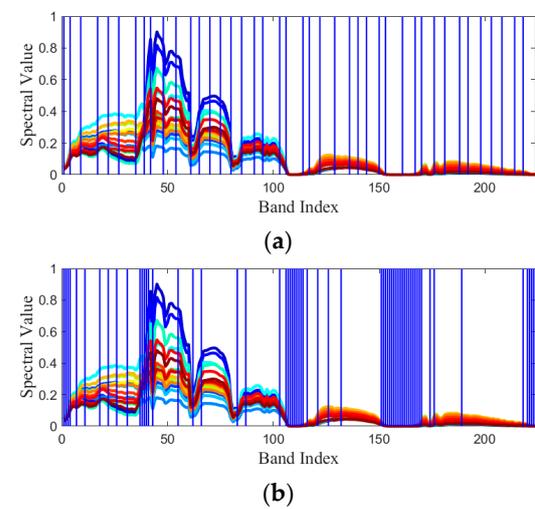**Figure 11.** The BG results for the Purdue experiment. (**a**) FNG; (**b**) BD.

Figure 12. The BG results for the Salinas experiment. (**a**) FNG; (**b**) BD.

Since BG-SSRBSS belongs to a search-based BS method, its computation speed is worthy of attention. Table 12 tabulates the computing times in seconds for OMP-BS, PBS, SC/SQ CCBSS (SAM), SC/SQ LCMV-BSS, and the proposed six BG-SSRBSS methods.

Table 12. Averaged computing time in seconds for various BS methods.

	Pavia	Purdue	Salinas
OMP-BS [27]	98.31	44.73	359.58
SC CCBSS [17]	51.88	214.57	332.19
SQ CCBSS [17]	57.32	204.44	338.11
SC LCMV-BSS [18]	247.61	39.94	414.78
SQ LCMV-BSS [18]	275.7	43.9	420.51
SC SSRBSS	119.55	51.31	426.48
SQ SSRBSS	125.56	53.66	395.21
SC FNG-SSRBSS	0.47 + 59.88	0.14 + 14.01	0.61 + 26.69
SQ FNG-SSRBSS	0.47 + 66.04	0.14 + 18.59	0.61 + 33.77
SC BD-SSRBSS	2.64 + 80.13	0.48 + 18.08	1.36 + 52.3
SQ BD-SSRBSS	2.64 + 86.83	0.48 + 21.32	1.36 + 64.14

The computing time of each BG-SSRBSS method is expressed as the summation of the time required by BG and BGSS. It is apparent that the proposed BG-SSRBSS (FNG-SSRBSS and BD-SSRBSS) require less time than CCBSS and LCMV-BSS in most cases, particularly for the Purdue and Salinas datasets. Another interesting point is that the FNG-SSRBSS and BD-SSRBSS require less computing time than the SSRBSS. This is simply because the number of iterations required for SC/SQ searches was drastically reduced. More specifically, the SSRBSS roughly requires a total of $(L - p) \times p$ computations of the optimizations of the SSR model, while the BGSSRBS only requires a total of $(g - p) \times p$ computations. Although the size of the matrix **P** formed in BG-SSRBSS is larger than that in SSRBSS, the overall computation time did not increase dramatically because the optimization of the SSR model could be efficiently calculated with Matlab. Overall, compared with the existing BSS methods, the proposed BG-SSRBSS can strike a balance between classification accuracy and computational complexity.

7. Conclusions

This paper developed a new BS method, called BG-SSRBSS, which can select an informative band subset for hyperspectral image classification. The BG-SSRBSS consists of two parts: band grouping (BG) and band group subset selection (BGSS). The BG divides entire bands into several non-overlapping band groups in the direction of the spectrum. It can be viewed as a pre-decorrelation process. Using BG not only reduces the similarity between selected bands, but also decreases the number of iterations required in BGSS's search so as to reduce computing time. The BGSS aims to find the optimal band group subset with minimal reconstruction error of the SSR model via SC or SQ search. Finally, the representative bands of the optimal band group subset are taken as the BS result. The experiments conducted on three HSI datasets show that using the bands selected by BG-SRBSSS could achieve better classification accuracy than state-of-the-art methods either with two different classifiers.

It is worth noted that if BG is not used, the BG-SSRBSS will degenerate to a typical BSS approach, called SSRBSS. From this point of view, the proposed BG-SSRBSS enlarges the existing BSS framework to a brand new approach: Band-grouping-based BSS (BG-BSS). How to produce better band groups as the input of BG-BSS would be an interesting topic. On the other hand, the current BSS approaches, as well as our BG-SSRBSS, can be run with multiple cycles. That is, it is able to use the result of the first run as the input of the next run. Doing so may prompt BSS to find better band subsets. This part will be reserved for our future work. Furthermore, there are several points that can be improved under the existing BSS framework. For instance, how to select a more appropriate initial band subset to reach

the global optima, and how to provide a guidance that can automatically determine the value of parameter *g* in BG-BSS. Those parts will also be investigated in the future.

References

1. Sun, W.; Du, Q. Hyperspectral Band Selection: A Review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 118–139. [CrossRef]
2. Patro, R.N.; Subudhi, S.; Biswal, P.K.; Dell'Acqua, F. A Review of Unsupervised Band Selection Techniques: Land Cover Classification for Hyperspectral Earth Observation Data. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 72–111. [CrossRef]
3. 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]
4. Chang, C.-I.; Wang, S. Constrained band selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1575–1585. [CrossRef]
5. Huang, R.; He, M. Band selection based on feature weighting for classification of hyperspectral data. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 156–159. [CrossRef]
6. Chang, C.-I.; Liu, K.-H. Progressive Band Selection of Spectral Unmixing for Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2002–2017. [CrossRef]
7. Du, Q.; Yang, H. Similarity-based unsupervised band selection for Hyperspectral Image Analysis. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 564–568. [CrossRef]
8. Yin, J.; Wang, Y.; Hu, J. A New Dimensionality Reduction Algorithm for Hyperspectral Image Using Evolutionary Strategy, *IEEE Trans. Ind. Informat.* **2012**, *8*, 935–943. [CrossRef]
9. Feng, J.; Jiao, L.; Zhang, X.; Sun, T. Hyperspectral band selection based on trivariate mutual information and clonal selection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4092–4105. [CrossRef]
10. Su, H.; Du, Q.; Chen, G.; Du, P. Optimized Hyperspectral Band Selection Using Particle Swarm Optimization. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2014**, *7*, 2659–2670. [CrossRef]
11. Ghamisi, P.; Couceiro, M.S.; Benediktsson, J.A. A Novel Feature Selection Approach Based on FODPSO and SVM. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2935–2947. [CrossRef]
12. Su, H.; Yong, B.; Du, Q. Hyperspectral Band Selection Using Improved Firefly Algorithm. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 68–72. [CrossRef]
13. Medjahed, S.A.; Ait Saadi, T.; Benyettou, A.; Ouali, M. Gray Wolf Optimizer for Hyperspectral Band Selection. *Appl. Soft Comput.* **2016**, *40*, 178–186. [CrossRef]
14. Imbiriba, T.; Bermudez, J.C.M.; Richard, C. Band Selection for Nonlinear Unmixing of Hyperspectral Images as a Maximal Clique Problem. *IEEE Trans. Image Process.* **2017**, *26*, 2179–2191. [CrossRef]
15. Wang, C.; Gong, M.; Zhang, M.; Chan, Y. Unsupervised Hyperspectral Image Band Selection via Column Subset Selection. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1411–1415. [CrossRef]
16. Wang, L.; Li, H.; Xue, B.; Chang, C. Constrained Band Subset Selection for Hyperspectral Imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2032–2036. [CrossRef]
17. Chang, C.; Lee, L.; Xue, B.; Song, M.; Chen, J. Channel Capacity Approach to Hyperspectral Band Subset Selection. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2017**, *10*, 4630–4644. [CrossRef]
18. Yu, C.; Song, M.; Chang, C.-I. Band Subset Selection for Hyperspectral Image Classification. *Remote Sens.* **2018**, *10*, 113. [CrossRef]
19. Yuan, Y.; Lin, J.; Wang, Q. Dual-clustering-based hyperspectral band selection by contextual analysis. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1431–1445. [CrossRef]
20. Zhu, G.; Huang, Y.; Lei, J.; Bi, Z.; Xu, F. Unsupervised hyperspectral band selection by dominant set extraction. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 227–239. [CrossRef]
21. Yang, C.; Tan, Y.; Bruzzone, L.; Lu, L.; Guan, R. Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images. *Remote Sens.* **2017**, *9*, 782. [CrossRef]
22. Yuan, Y.; Zheng, X.; Lu, X. Discovering Diverse Subset for Unsupervised Hyperspectral Band Selection. *IEEE Trans. Image Process.* **2017**, *26*, 51–64. [CrossRef] [PubMed]
23. Zeng, M.; Cai, Y.; Cai, Z.; Liu, X.; Hu, P.; Ku, J. Unsupervised Hyperspectral Image Band Selection Based on Deep Subspace Clustering. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1889–1893. [CrossRef]

24. Li, S.; Qi, H. Sparse representation based band selection for hyperspectral images. In Proceedings of the 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 2693–2696. [CrossRef]
25. Li, H.; Wang, Y.; Duan, J.; Xiang, S.; Pan, C. Group sparsitybased semi-supervised band selection for hyperspectral images. In Proceedings of the IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; IEEE: Piscataway, NJ, USA; pp. 3225–3229. [CrossRef]
26. Sun, W.; Zhang, L.; Du, B.; Li, W.; Mark Lai, Y. Band Selection Using Improved Sparse Subspace Clustering for Hyperspectral Imagery Classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2015**, *8*, 2784–2797. [CrossRef]
27. Lai, C.-H.; Chen, C.-S.; Chen, S.-Y.; Liu, K.-H. Sequential band selection method based on group orthogonal matching pursuit. In Proceedings of the 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA, 21–24 August 2016; pp. 1–4. [CrossRef]
28. Sun, W.; Zhang, L.; Zhang, L.; Lai, Y.M. A Dissimilarity-Weighted Sparse Self-Representation Method for Band Selection in Hyperspectral Imagery Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4374–4388. [CrossRef]
29. Sun, W.; Tian, L.; Xu, Y.; Zhang, D.; Du, Q. Fast and Robust Self-Representation Method for Hyperspectral Band Selection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5087–5098. [CrossRef]
30. Kuo, B.-C.; Ho, H.-H.; Li, C.-H.; Hung, C.-C.; Taur, J.-S. A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* **2014**, *7*, 317–326. [CrossRef]
31. 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]
32. Cai, R.; Yuan, Y.; Lu, X. Hyperspectral band selection with convolutional neural network. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Guangzhou, China, 23–26 November 2018; pp. 396–408. [CrossRef]
33. Cai, Y.; Liu, X.; Cai, Z. BS-Nets: An end-to-end framework for band selection of hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1969–1984. [CrossRef]
34. Feng, J.; Li, D.; Gu, J.; Cao, X.; Shang, R.; Zhang, X.; Jiao, L. Deep reinforcement learning for semisupervised hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5501719. [CrossRef]
35. Chang, Y.-L.; Liu, J.-N.; Chen, Y.-L.; Chang, W.-Y.; Hsieh, T.-J.; Huang, B. Hyperspectral band selection based on parallelparticle swarm optimization and impurity function band prioritization schemes. *J. Appl. Remote Sens.* **2014**, *8*, 084798. [CrossRef]
36. Paul, A.; Bhattacharya, S.; Dutta, D.; Sharma, J.R.; Dadhwal, V.K. Band selection in hyperspectral imagery using spatial cluster mean and genetic algorithms. *GISci. Remote Sens.* **2015**, *52*, 643–659. [CrossRef]
37. 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]
38. Xiong, W.; Chang, C.-I.; Wu, C.-C.; Kalpakis, K.; Chen, H.M. Fast Algorithms to Implement N-FINDR for Hyperspectral Endmember Extraction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 545–564. [CrossRef]
39. Chang, C.-I. *Real Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection*; Springer: New York, NY, USA, 2016. [CrossRef]
40. Chang, C.-I. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2003. [CrossRef]
41. Chang, C.-I.; Du, Q. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 608–619. [CrossRef]
42. Yu, H.; Gao, L.; Liao, W.; Zhang, B. Group Sparse Representation Based on Nonlocal Spatial and Local Spectral Similarity for Hyperspectral Imagery Classification. *Remote Sens.* **2018**, *18*, 1695. [CrossRef]
43. Sun, W.; Jiang, M.; Li, W.; Liu, Y. A Symmetric Sparse Representation Based Band Selection Method for Hyperspectral Imagery Classification. *Remote Sens.* **2016**, *8*, 238. [CrossRef]
44. Iordache, M.-D.; Bioucas-Dias, J.M.; Plaza, A. Collaborative Sparse Regression for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 341–354. [CrossRef]
45. Li, C.; Ma, Y.; Mei, X.; Liu, C.; Ma, J. Hyperspectral Unmixing with Robust Collaborative Sparse Regression. *Remote Sens.* **2016**, *8*, 588. [CrossRef]
46. Elhamifar, E.; Sapiro, G.; Vidal, R. See all by looking at a few: Sparse modeling for finding representative objects. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1600–1607. [CrossRef]
47. Wang, Q.; Li, Q.; Li, X. A Fast Neighborhood Grouping Method for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5028–5039. [CrossRef]
48. Lozano, A.C.; Świrszcz, G.; Abe, N. Group Orthogonal Matching Pursuit for variable selection and prediction. In Proceedings of the 22nd International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 6–14 December 2009; pp. 1150–1158.
49. Chang, C.-I. *Hyperspectral Data Processing: Algorithm Design and Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2013. [CrossRef]
50. Bigdeli, B.; Samadzadegan, F.; Reinartz, P. Band Grouping versus Band Clustering in SVM Ensemble Classification of Hyperspectral Imagery. *Photogramm. Eng. Remote Sens.* **2013**, *79*, 523–533. [CrossRef]

51. Hyperspectral Remote Sensing Scenes. Available online: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes (accessed on 6 October 2022).
52. Chang, C.-C.; Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [CrossRef]
53. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 277–281. [CrossRef]