remote sensing

MDPI

Article

EBARec-BS: Effective Band Attention Reconstruction Network for Hyperspectral Imagery Band Selection

Yufei Liu [1], Xiaorun Li [1,*], Ziqiang Hua [1] and Liaoying Zhao [2]

[1] College of Electrical Engineering, Zhejiang University, No. 38, Zheda Road, Xihu District, Hangzhou 310027, China; 11910076@zju.edu.cn (Y.L.); ziqhua@zju.edu.cn (Z.H.)
[2] Department of Computer Science, Hangzhou Dianzi University, Hangzhou 310027, China; zhaoly@hdu.edu.cn
* Correspondence: lxrly@zju.edu.cn

Abstract: Hyperspectral band selection (BS) is an effective means to avoid the Hughes phenomenon and heavy computational burden in hyperspectral image processing. However, most of the existing BS methods fail to fully consider the interaction between spectral bands and cannot comprehensively consider the representativeness and redundancy of the selected band subset. To solve these problems, we propose an unsupervised effective band attention reconstruction framework for band selection (EBARec-BS) in this article. The framework utilizes the EBARec network to learn the representativeness of each band to the original band set and measures the redundancy between the bands by calculating the distance of each unselected band to the selected band subset. Subsequently, by designing an adaptive weight to balance the influence of the representativeness metric and redundancy metric on the band evaluation, a final band scoring function is obtained to select a band subset that well represents the original hyperspectral image and has low redundancy. Experiments on three well-known hyperspectral data sets indicate that compared with the existing BS methods, the proposed EBARec-BS is robust to noise bands and can effectively select the band subset with higher classification accuracy and less redundant information.

Keywords: hyperspectral image (HSI); unsupervised band selection; convolutional autoencoder (CAE); band attention

check for updates

Citation: Liu, Y.; Li, X.; Hua, Z.; Zhao, L. EBARec-BS: Effective Band Attention Reconstruction Network for Hyperspectral Imagery Band Selection. *Remote Sens.* **2021**, *13*, 3602. https://doi.org/10.3390/rs13183602

Academic Editor: Jaime Zabalza

Received: 6 July 2021
Accepted: 7 September 2021
Published: 9 September 2021

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

Copyright: © 2021 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

Hyperspectral images (HSIs) are composed of hundreds of contiguous bands containing rich spatial and spectral information, making it possible to identify objects of interest accurately. However, in practical applications, the data redundancy brought about by a large number of bands causes the Hughes phenomenon [1] and heavy computation burden. Thus, effective dimensionality reduction (DR) methods are of great significance to the subsequent tasks of HSIs.

Generally, DR methods can be divided into band selection (BS) and feature extraction. BS is to select a band subset that contains as much effective information as possible from the original band set. Compared with feature extraction methods [2,3], which utilize the complex feature transformation to obtain the reduced-dimensional HSIs, BS methods [4,5] can retain the physical information of the original HSI. In this sense, we focus mainly on BS methods.

BS methods can basically be summarized as supervised [6] and unsupervised [7] methods according to whether prior knowledge is required. Since prior knowledge is often difficult to obtain in practice, unsupervised BS methods have attracted extensive attention in recent decades. Unsupervised BS methods can be further divided into four categories: point-wise methods, group-wise methods, ranking-based methods, and advanced machine learning-based methods. The point-wise unsupervised BS methods, such as volume-gradient-based BS (VGBS) [8] and orthogonal-projection-based BS (OPBS) [9],

are generally based on greedy algorithms. Specifically, VGBS is based on sequential backward search (SBS), and OPBS is based on sequential forward search (SFS). The point-wise BS methods utilize specific subset evaluation criteria to add or remove bands one by one until the required number of bands is obtained. The design of the subset evaluation criteria has a great influence on the performance of selected bands. The group-wise unsupervised BS methods are commonly based on evolutionary algorithms, e.g., particle swarm optimization-based BS [10] and ant colony optimization-based BS [11]. The ranking-based unsupervised BS methods sort the importance of each band through certain evaluation indicators and then directly select the bands ranked in the front with a required number. This kind of method includes maximum-variance principal component analysis (MVPCA) [12], covariance-based method [13], and linearly constraint minimum variance (LCMV) [14]. The advanced machine learning-based unsupervised BS methods have received extensive attention with the development of machine learning algorithms. This kind of method includes clustering-based BS [15,16], sparsity learning-based BS [17], manifold learning-based BS [18], and graph theory-based BS methods [19].

Nevertheless, most of the existing unsupervised BS methods cannot sufficiently consider the relationship between spectral bands. For instance, clustering-based methods usually treat each spectral band as an independent entity and evaluate it, making a great deal of hidden information of the original HSI lost [15,20]. In addition, most ranking-based methods mainly consider the information of each band while ignoring the redundancy existing between the selected bands [12,14]. Moreover, most of the BS methods only consider the linear correlation between the bands or simply the nonlinear correlation based on the predefined kernel function and cannot analyze the inherent nonlinear correlation between the bands well [4,9]. In this context, some deep learning-based BS methods are proposed to consider the underlying nonlinear relationship between the bands. However, most of the existing deep learning-based BS methods ignore the redundant information between the bands. For example, the state-of-the-art BS network using convolutional neural networks (BS-Net-Conv) [21] mainly considers the representativeness of the selected bands to the original band set. Because interdependent bands often have similar effects, this BS network cannot guarantee that the selected bands contain less redundant information, which is not conducive to the implementation of downstream tasks, such as classification [22], object detection [23], and unmixing [24]. Furthermore, BS-Net-Conv calculates the evaluation index corresponding to each band by first projecting the bands to a low-dimensional space and then mapping them back, making the correspondence between the band and its evaluation index indirect. That is, the band evaluation index cannot accurately reflect the original spectral band information. In addition, the existing BS methods based on reconstructed networks commonly employ the mean square error (MSE) as the reconstruction criterion [21,25]. However, the MSE means the complete reconstruction. That is, when only the MSE is utilized as the criterion for spectrum reconstruction performance, scalable reconstruction cannot be achieved, limiting the applicability of the model. To address these shortcomings of the existing BS methods, a better BS architecture, which comprehensively considers the redundancy and representativeness of the band and reveals the inherent nonlinear relationship between bands, should be designed.

To achieve it, in this article, we propose an effective band attention reconstruction BS (EBARec-BS) network. Specifically, we first constructed an effective band attention reconstruction (EBARec) network to explore the underlying nonlinear relationship between the bands. It is worth noting that, to make each band directly correspond to its weight, the proposed method utilizes effective band attention (EBA) to calculate the weight of each band. Then, the reweighted spectral bands are utilized to reconstruct the original HSI using a convolutional autoencoder (CAE). In addition, to improve the applicability of the model, the reconstruction criterion of the EBARec-BS network not only adopts the traditional MSE constraint item but also adds a spectral angle error constraint item. After the network training is completed, to obtain the band subset with low redundancy information and high representativeness, we design a BS scoring function that can comprehensively consider

the band attention weight and the redundancy between the selected bands. The main contributions of this article can be summarized as follows:

1. We propose a novel BS scoring function that can consider the redundancy and representativeness of the bands simultaneously. To the best of our knowledge, this is the first time that the redundancy and representativeness of bands are explored simultaneously in the attention and reconstruction network-based BS method. Specifically, we design an adaptive balance coefficient that can balance the representativeness metric and the redundancy metric to solve the problem that the scoring function has different sensitivities to these two metrics. According to the proposed BS scoring function, a band subset with a good representation of the original band set and less redundant information can be selected, which is conducive to downstream tasks.
2. The proposed attention reconstruction network-based BS architecture adds the spectral angle error as one of the evaluation criteria of the reconstruction effect, which is proposed for the first time. As a result, unlike the traditional reconstruction network that only uses MSE as the reconstruction criterion, our attention reconstruction network-based BS architecture combines MSE and spectral angle error to improve the applicability of the model.
3. A novel unsupervised BS framework in which attention weights and bands are closely connected is proposed, which helps to resolve the problem that correspondence between the band and its weight is indirect in current attention mechanism-based methods.

The remainder of this article is organized as follows: Section 2 explains some related works of this article. Section 3 specifically introduces the proposed EBARec-BS network. Section 4 shows experiments and results on different real-world HSIs. Finally, Section 5 presents concluding remarks.

2. Related Works

2.1. Attention Mechanism

Attention was initially designed for machine translation [26]. Recently, attention has developed rapidly in the fields of speech [27], natural language processing [28,29], and computer vision [30] because of its ability to improve the interpretability of neural networks. The expression of an attention module is

$$\boldsymbol{a} = f(\boldsymbol{x}_{\mathrm{a}}; \Theta) \tag{1}$$

where f denotes the attention module, Θ is the parameters of the module, $\boldsymbol{x}_{\mathrm{a}}$ is the input of attention module, and $\boldsymbol{a}$ is the attention map.

Attention modules can be summarized as the channel, spatial, and joint attention according to different domains of interest. The spatial attention is to focus more attention on the spatial location worthy of attention; the channel attention learns the weight of each channel through the attention module, thereby generating the attention of the channel domain; and the joint attention mechanism is a combination of the previous two.

The basic idea of unsupervised band selection is to find the most valuable bands, which can be abstracted as the most noteworthy channels. Thus, we expect to utilize channel attention to calculate the importance of each spectral band and reflect the inherent relationship between the bands. The diagram of channel attention is illustrated in Figure 1.

However, in practical applications, the band subset selected by the channel attention method is not always the optimal combination. This is because the existing channel attention-based BS methods [21,25] first map the band features to a low-dimensional space and then map them back so that the correspondence between the band and its weight is indirect. That is, the band weight cannot accurately reflect the original spectral band information.

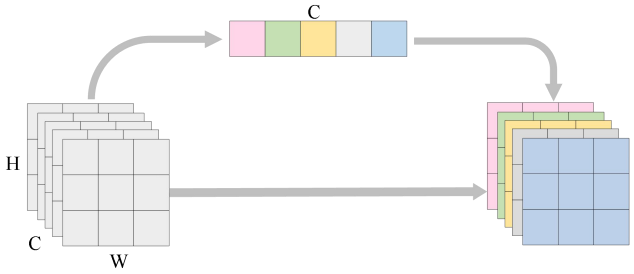


Figure 1. Channel attention diagram.

2.2. Autoencoder

The autoencoder is a neural network that reproduces the input vector to the output through certain transformations. Specifically, an autoencoder consists of two parts, namely the encoder and the decoder. The encoder compresses the input data into a latent space representation; the decoder uses the features of the latent space to reconstruct the original input data. Mathematically, the encoder and decoder of a single-layer autoencoder with input $\boldsymbol{x}_1 \in \mathbb{R}^q$ can be, respectively, denoted as follows:

$$\boldsymbol{y} = \sigma(\mathbf{W}_1 \boldsymbol{x}_1 + \boldsymbol{b}_1) \tag{2}$$

$$\hat{\boldsymbol{x}}_1 = \sigma(\mathbf{W}_2 \boldsymbol{y} + \boldsymbol{b}_2) \tag{3}$$

where $\hat{\boldsymbol{x}}_1 \in \mathbb{R}^q$ denotes the reconstruction of the input data, $\sigma(\cdot)$ is a nonlinear activation function (such as ReLU and Sigmoid, etc.), $\mathbf{W}_1 \in \mathbb{R}^{m \times q}$ and $\mathbf{W}_2 \in \mathbb{R}^{q \times m}$ are wight parameters, $\boldsymbol{b}_1 \in \mathbb{R}^m$ and $\boldsymbol{b}_2 \in \mathbb{R}^q$ indicate bias vectors.

However, the autoencoder is originally designed to process one-dimensional data. Therefore, when reconstructing image data, the traditional autoencoder has the problem of input size processing and the drawbacks that the features are forced to be global. The existing research on the object recognition direction [31] shows that the model whose input is local features is better than the model whose input is global features.

In order to overcome the shortcomings of a traditional autoencoder, the formula of convolutional autoencoder(CAE) [32] for high-dimensional input is proposed. The latent representation of the kth feature map and the reconstruction of input, for a single-channel input $\boldsymbol{x}_\mathrm{h}$, are respectively expressed as follows:

$$\boldsymbol{y}^k = \sigma(\boldsymbol{x}_\mathrm{h} * \mathbf{W}^k + \boldsymbol{b}^k) \tag{4}$$

$$\hat{\boldsymbol{x}}_\mathrm{h} = \sigma(\sum_{k \in H} \boldsymbol{y}^k * \tilde{\mathbf{W}}^k + \boldsymbol{c}) \tag{5}$$

where $*$ represents the two-dimensional convolution operator, $\hat{\boldsymbol{x}}_\mathrm{h}$ denotes the reconstruction of the input data, $\mathbf{W}$ is a convolution kernel, $\tilde{\mathbf{W}}$ represents the flip operation on the two dimensions of the weight, $\boldsymbol{b}$ and $\boldsymbol{c}$ represent bias, and H denotes the potential feature map group.

Just as for traditional autoencoders, the definition of the cost function is to minimize the MSE. Specifically, given a set of data $\mathcal{X} = \{\boldsymbol{x}_\mathrm{h}^{(1)}, \boldsymbol{x}_\mathrm{h}^{(2)}, \ldots, \boldsymbol{x}_\mathrm{h}^{(n)}\}$, the MSE is defined as

$$E(\Theta) = \frac{1}{2n} \sum_{i=1}^{n} (\boldsymbol{x}_\mathrm{h}^{(i)} - \hat{\boldsymbol{x}}_\mathrm{h}^{(i)})^2. \tag{6}$$

where Θ represents the trainable parameters. In the error backpropagation phase, the gradient descent method is used to update the parameters.

3. The Proposed Method

This section introduces the proposed EBARec-BS network in detail. As shown in Figure 2, the proposed BS framework adopts EBA to make each band directly correspond

to its weight and reconstructs the original hyperspectral data through a CAE. Moreover, compared with the existing attention reconstruction network-based BS methods that only use MSE when calculating the loss function, our proposed method also constrains the spectral angle error after considering the characteristics of the hyperspectral data. Furthermore, to solve the problem that the existing methods do not consider the redundancy between the bands when selecting the band subset, the band selection strategy of the EBARec-BS network comprehensively considers the attention weights and the redundancy between bands. A detailed description of each step of the proposed method is given as follows.

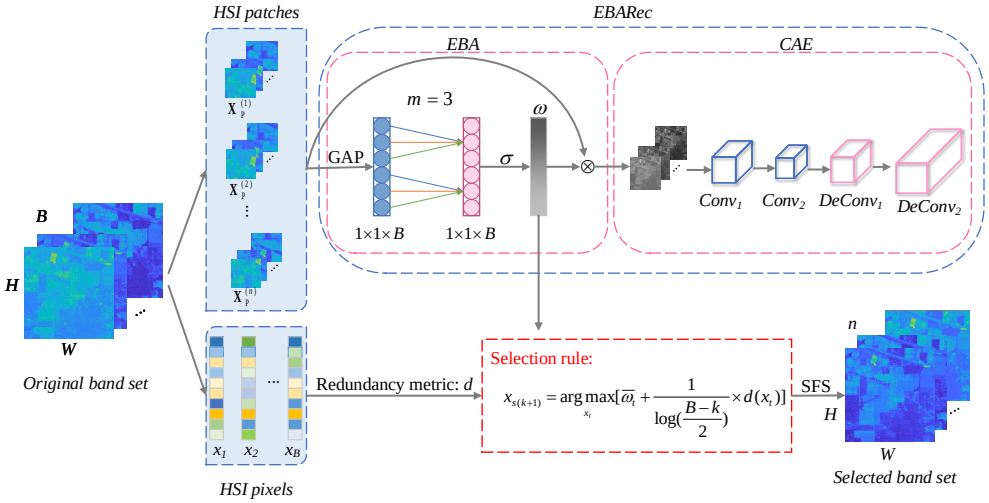


Figure 2. Overview of the proposed EBARec-BS network.

3.1. EBARec

Let $\mathbf{X} \in \mathbb{R}^{W \times H \times B}$ be a spatial-spectral HSI, where $W \times H$ is the number of pixels and B represents the number of bands. In order to ensure the quantity and quality of training samples, our EBARec module utilizes an $a \times a$-sized squared window to slide across the original HSI with a step length of t to obtain three-dimensional patches containing spatial and spectral information as input.

To select the most meaningful bands from the original band set, it is important to analyze the cross-band interaction. This can be achieved through the feature attention mechanism. The proposed BS method utilizes an efficient channel attention [33] to recalibrate each band.

As illustrated in Figure 2, EBA takes HSI cubes $\mathbf{X}_\mathrm{p} \in \mathbb{R}^{a \times a \times B}$ as input and produces a band attention vector $\boldsymbol{\omega}$ as output, i.e.,

$$\boldsymbol{\omega} = \mathrm{F}_{\mathrm{EBA}}(\mathbf{X}_\mathrm{p}; \Theta_\mathrm{p}) \tag{7}$$

where Θ_p is the trainable parameters of the EBA module.

To be more specific, first, we input the HSI patches into a global average pooling (GAP) to obtain aggregated features, that is

$$\boldsymbol{y} = \mathrm{G}(\mathbf{X}_\mathrm{p}) = \frac{1}{a \times a} \sum_{i,j=1}^{a} \mathbf{X}_{\mathrm{p}i,j} \tag{8}$$

where $\mathrm{G}(\cdot)$ denotes channel-wise GAP and $\boldsymbol{y} \in \mathbb{R}^{1 \times 1 \times B}$ contains aggregated features.

Then, in order to avoid obtaining the indirect correspondence between the band and its weight, the correlation between bands is captured by a one-dimensional convolution with a kernel size of m, that is

$$\boldsymbol{\omega} = \sigma(\mathrm{C1D}_m(\boldsymbol{y})) \tag{9}$$

where $\sigma(\cdot)$ indicates a Sigmoid function and C1D$(\cdot)$ denotes one-dimensional convolution. Specifically, the weight of each aggregated feature $y_i \in \boldsymbol{y}(i = 1, \ldots, B)$ can only consider the interaction between itself and its m neighbors, and all bands have the same learning parameters. It can be explicitly represented as follows:

$$\omega_i = \sigma(\sum_{j=1}^{m} \beta^j y_i^j), \qquad y_i^j \in \mathbf{\Omega}_i^m \tag{10}$$

where β^j denotes the shareable learning parameter associated with each y_i^j, and $\mathbf{\Omega}_i^m$ is the collection of m neighbors of y_i.

As shown in Figure 2, we perform the band-wise production operation on the original input HSI block and the weight obtained by EBA, and the reweighted spectral inputs can be computed as follows:

$$\mathbf{Z} = \mathbf{X}_{\mathrm{p}} \otimes \boldsymbol{\omega} \tag{11}$$

where $\otimes$ is the band-wise production.

In the next step, in order to obtain the representativeness of the reweighted spectral bands to the original data set, we use a CAE to reconstruct the original input HSI block. The reconstruction network with the reweighted spectral bands $\mathbf{Z}$ as input and the predicted value $\hat{\mathbf{X}}_{\mathrm{p}}$ of the original image as output can be defined as follows:

$$\hat{\mathbf{X}}_{\mathrm{p}} = \mathrm{F}_{\mathrm{Rec}}(\mathbf{Z}; \Theta_{\mathrm{c}}) \tag{12}$$

where Θ_{c} is the trainable parameters of the reconstruction network.

The existing reconstruction network can reflect a certain degree of the relationship between input and output through the MSE. However, for HSIs, the spectral similarity measurement based on the MSE is not suitable for scalable reconstruction. In order to establish an effective reconstruction network for HSI band selection, an architecture with wide applicability should be proposed. Therefore, the proposed EBARec uses MSE and spectral angle similarity to minimize the reconstruction error. We define the cost function as follows:

$$\mathcal{L} = \frac{1}{2n} \sum_{i=1}^{n} \|\hat{\mathbf{X}}_{\mathrm{p}}^{(i)} - \mathbf{X}_{\mathrm{p}}^{(i)}\|_F^2 + \eta \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{a \times a} \arccos \frac{\hat{\mathbf{X}}_{\mathrm{p}j}^{(i)\mathrm{T}} \cdot \mathbf{X}_{\mathrm{p}j}^{(i)}}{\|\hat{\mathbf{X}}_{\mathrm{p}j}^{(i)}\| \|\mathbf{X}_{\mathrm{p}j}^{(i)}\|} \tag{13}$$

where n denotes the number of samples, $\|\cdot\|_F$ indicates the Frobenius norm for matrices, $\mathbf{X}_{\mathrm{p}}^{(i)}(i = 1, 2, \ldots, n)$ denotes the ith input HSI cube, η is a balance parameter, $\mathbf{X}_{\mathrm{p}j}^{(i)} \in \mathbf{X}_{\mathrm{p}}^{(i)}(j = 1, 2, \ldots, a \times a)$ denotes the jth pixel of the ith input HSI cube, and the superscript T represents the transpose operation.

Furthermore, in order to make the weight of each band easier to interpret to facilitate BS, we impose sparse constraints on them. From this point of view, the band weights are constrained by the ℓ_1-norm, i.e.,

$$\mathcal{R} = \sum_{i=1}^{n} \|\boldsymbol{\omega}^{(i)}\|_1. \tag{14}$$

Therefore, the final objective function of the proposed EBARec includes three parts: an MSE term for the complete reconstruction of HSIs, a spectral similarity error term for the scalable reconstruction of HSIs, and a sparse constraint term for band weights. Mathematically, the objective function of the proposed EBARec can be given as follows:

$$\mathcal{L}(\Theta_{\mathrm{p}}, \Theta_{\mathrm{c}}) = \frac{1}{2n} \sum_{i=1}^{n} \|\hat{\mathbf{X}}_{\mathrm{p}}^{(i)} - \mathbf{X}_{\mathrm{p}}^{(i)}\|_F^2 + \eta \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{a \times a} \arccos \frac{\hat{\mathbf{X}}_{\mathrm{p}j}^{(i)\mathrm{T}} \cdot \mathbf{X}_{\mathrm{p}j}^{(i)}}{\|\hat{\mathbf{X}}_{\mathrm{p}j}^{(i)}\| \|\mathbf{X}_{\mathrm{p}j}^{(i)}\|} + \gamma \sum_{i=1}^{n} \|\boldsymbol{\omega}^{(i)}\|_1 \tag{15}$$

where γ denotes a penalty parameter.

We utilize adaptive moment estimation (Adam) to optimize the proposed EBARec model. After training, the representativeness of a certain band to the original band set can be obtained by averaging the weights of this band for all training samples. The average attention weight of the tth band is formulated as

$$\bar{\omega}_t = \frac{1}{n}\sum_{i=1}^{n}\omega_t^{(i)}. \tag{16}$$

The obtained average band weights can be used as the representative metric in the proposed BS method.

3.2. Band Selection Module Based on Representativeness and Redundancy

The band attention weights can reflect the contribution of each band to the original HSI reconstruction. The larger weight represents the more significant contribution of the corresponding band to the reconstruction, which means that the band can better represent the original band set. However, simply selecting bands based on weights will ignore the amount of redundant information between the selected bands and affect the implementation of downstream tasks (such as classification). Therefore, how to make the best use of the band attention weights to guide the selection of bands and how to construct a BS framework to weigh the attention weights of bands and the redundancy between bands are the challenges. In order to solve these problems, we design a BS scoring function that comprehensively considers the band attention weight and the redundancy between the bands.

In the process of BS, if a candidate band with plenty of redundant information with the selected band subset is selected, it will affect the implementation of downstream tasks. For this reason, we try to avoid selecting the band that has much redundant information with the selected band subset. To achieve it, we measure the redundancy of the candidate band by calculating the distance of this candidate band to the hyperplane spanned by the selected bands. The greater the distance is, the less redundant information the candidate band contains. This article utilizes the orthogonal subspace projection (OSP), which was originally designed for linear spectral mixture analysis, to measure the distance between bands. It is worth mentioning that through OSP, the distance between the candidate band and the selected band subset can be measured jointly rather than in pairs, which ensures the efficiency of the proposed method. Next, we introduce how to use OSP to calculate the redundancy constraint in our BS scoring function.

Suppose that the hyperspectral data set is denoted as $\mathbf{X}_{2D} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_B] \in \mathbb{R}^{N\times B}$, where B and N are the number of bands and pixels, respectively. Assuming that n bands need to be selected from the total bands, when k bands have been selected, we use $\mathbf{X}_S = [\boldsymbol{x}_{s(1)}, \boldsymbol{x}_{s(2)}, \cdots, \boldsymbol{x}_{s(k)}] \in \mathbb{R}^{N\times k}$ $(k < n)$ to represent the matrix composed of the selected bands. Then, the subspace W spanned by column vectors of the matrix $\mathbf{X}_S$ is expressed as

$$W = \mathrm{Span}\{\mathbf{X}_S\} = \{\boldsymbol{x} : \boldsymbol{x} = \sum_{i=1}^{k} a_i \times \boldsymbol{x}_{s(i)}\} \tag{17}$$

where $\mathrm{Span}\{\mathbf{X}_S\}$ denotes the set consisting of all linear combinations of the column vectors of the matrix $\mathbf{X}_S$ and a_i represents a scalar.

Assuming that a candidate band is denoted by $\boldsymbol{x}_t$ (the tth band in $\mathbf{X}_{2D}$), the relationship between the candidate band and the selected band subset can be measured by the distance from the candidate band $\boldsymbol{x}_t$ to the subspace spanned by vector set $\mathbf{X}_S$, that is, the orthogonal projection of $\boldsymbol{x}_t$ on the band vector space W. Mathematically, by introducing the orthogonal projection operator:

$$\mathbf{P} = \mathbf{X}_S(\mathbf{X}_S^T\mathbf{X}_S)^{-1}\mathbf{X}_S^T, \tag{18}$$

the projection of $\boldsymbol{x}_t$ onto W can be expressed as

$$\hat{\boldsymbol{x}}_t = \mathbf{P}\boldsymbol{x}_t = \mathbf{X}_S(\mathbf{X}_S^T\mathbf{X}_S)^{-1}\mathbf{X}_S^T\boldsymbol{x}_t. \tag{19}$$

Then, the redundancy between the candidate band $\boldsymbol{x}_t$ and the set of selected bands $\mathbf{X}_S$ is measured by calculating the distance from $\boldsymbol{x}_t$ to $\mathbf{X}_S$, which can be given as

$$\begin{aligned} d(\boldsymbol{x}_t) &= \|\boldsymbol{x}_t - \hat{\boldsymbol{x}}_t\| \\ &= [\boldsymbol{x}_t^T\boldsymbol{x}_t - (\boldsymbol{x}_t^T\mathbf{X}_S)(\mathbf{X}_S^T\mathbf{X}_S)^{-1}(\boldsymbol{x}_t^T\mathbf{X}_S)^T]^{\frac{1}{2}} \end{aligned} \tag{20}$$

where $d(\boldsymbol{x}_t)$ denotes the redundancy metric of candidate band $\boldsymbol{x}_t$. The smaller the distance $d(\boldsymbol{x}_t)$ is, the more the redundancy of $\boldsymbol{x}_t$ is.

In order to construct a comprehensive consideration of the contribution of the selected band subset to the original band set and the redundancy between selected bands, the above two metrics, i.e., $\bar{\omega}$ and d, are used to construct the proposed BS criterion. Since our objective is to make the selected bands better represent the original band set while containing a small amount of redundant information, we have to find the band with high $\bar{\omega}$ and d. To achieve it, the proposed EBARec-BS scoring function that can comprehensively consider these two factors is defined as

$$\begin{cases} S(\boldsymbol{x}_t) = \bar{\omega}_t + r \times d(\boldsymbol{x}_t) \\ r = \dfrac{1}{\log(\dfrac{B-k}{2})} \end{cases} \tag{21}$$

where $S(\boldsymbol{x}_t)$ represents the scoring function of candidate band $\boldsymbol{x}_t$, and the band with a higher score is more important. r denotes a coefficient that balances the two constraints.

It is worth noting that we design the balance coefficient r to be $\dfrac{1}{\log[(B-k)/2]}$. The reason is that as the number of selected bands increases, the distance from the candidate band to the selected band subset gradually decreases. That is, the redundancy metric gradually declines, which means that the BS scoring function will mainly depend on the representativeness of the candidate band to the original HSI but is not sensitive to the redundancy metric. Therefore, to balance these two metrics, we have to appropriately amplify the influence of the redundancy indicator, which is decreasing as the number of selected bands increases. To this end, we design the weight $r = \dfrac{1}{\log[(B-k)/2]}$ that augments as the number of selected bands increases.

Based on the proposed selection criterion, we use a sequential forward search (SFS) to iteratively add the optimal band into the set of selected bands. Specifically, when selecting the $(k+1)$th band, the EBARec-BS method adds the candidate band with the highest score to the selected band set, that is

$$\boldsymbol{x}_{s(k+1)} = \arg\max_{\boldsymbol{x}_t} [S(\boldsymbol{x}_t)]. \tag{22}$$

Then, $\mathbf{X}_S$ is updated, and we repeat the process of adding the current optimal band calculated according to Equation (22) to the selected band subset $\mathbf{X}_S$ until $\mathbf{X}_S$ contains the required number of bands. Therefore, the proposed BS strategy can consider the contribution of the selected band to the original HSI and the redundancy among bands simultaneously. Note that when selecting the first band, $\mathbf{X}_S$ does not contain any bands. Thus, the scoring function of the candidate band only depends on the contribution of the candidate band to the original HSI. The procedures of EBARec-BS are given in Algorithm 1.

Algorithm 1 The EBARec-BS Algorithm

Input: HSI cube $\mathbf{X} \in \mathbb{R}^{W \times H \times B}$, the number of selected bands n, and EBARec-BS hyper-parameters.
Step1: Preprocess HSI and generate training samples.
Step2: Train EBARec network.
while *Model is convergent or maximum iteration is met* **do**
1: Sample a batch of training samples $\mathbf{X}_{\mathrm{p}}$.
2: Calculate bands weights: $\boldsymbol{\omega} = \mathrm{F}_{\mathrm{EBA}}(\mathbf{X}_{\mathrm{p}}; \Theta_{\mathrm{p}})$.
3: Reweight spectral bands: $\mathbf{Z} = \mathbf{X}_{\mathrm{p}} \otimes \boldsymbol{\omega}$.
4: Reconstruct spectral bands: $\hat{\mathbf{X}}_{\mathrm{p}} = \mathrm{F}_{\mathrm{Rec}}(\mathbf{Z}; \Theta_{\mathrm{c}})$.
5: Update Θ_{p} and Θ_{c} by minimizing Equation (15) using Adam algorithm.
end while
Step3: Calculate average attention weight of each band according to Equation (16).
Step4: Set counter $k = 0$.
Step5: Band selection.
while $k < n$ **do**
1: For the ith band $\boldsymbol{x}_i$, $(i = 1, 2, \ldots, B)$, calculate its score according to Equation (21). Note that if the ith band $\boldsymbol{x}_i$ has already been selected, its score would not be calculated and compared.
2: Find the band with the highest score and add it to the selected band subset.
3: $k \leftarrow k + 1$.
end while
Output: n selected bands.

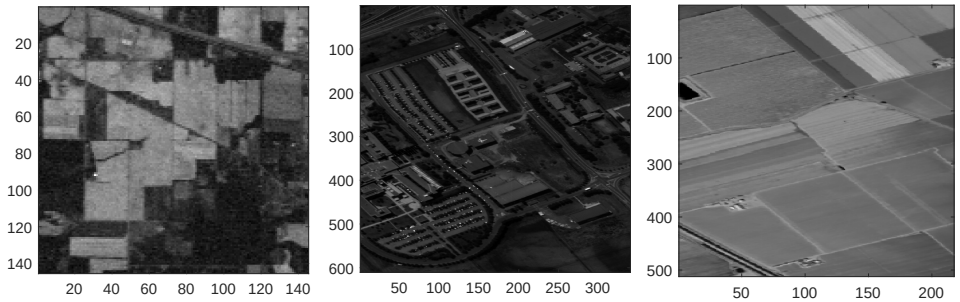
4. Experiments

In this section, the proposed EBARec-BS method and six existing unsupervised BS methods, namely MVPCA [12], BS-Net-Conv [21], OPBS [9], exemplar component analysis (ECA) [15], LCMV band correlation minimization (LCMVBCM) [14], and LCMV band correlation constraint (LCMVBCC) [14], are compared on three real-world HSIs. Among these methods, MVPCA is a classical BS method; BS-Net-Conv is a newly proposed state-of-the-art method; OPBS is a point-wise BS method; LCMVBCM and LCMVBCC are ranking-based BS methods; ECA is based on the density-based clustering method. To comprehensively evaluate the effect of each BS method, the classification effect, band correlation, and robustness of different BS methods are compared through specific analysis. Furthermore, to facilitate the experimental results to be clearly understood, we conduct an in-depth analysis of the selected band subsets from the two aspects of quantification and visualization.

The Indian Pines data set (Figure 3a) contains 220 bands and 145 $\times$ 145 pixels. The low signal-to-noise ratio and atmospheric water vapor absorption bands (i.e., bands 1–3, 103–112, 148–165, and 217–220) are removed, and the remaining 185 bands were utilized in our experiments. This data set has 16 land-cover categories. Although the number of samples in each category is not balanced [34], they are all used in the verification experiment to evaluate the classification performance of the selected band subsets. The Pavia University data set (Figure 3b) has 103 bands, 610 $\times$ 340 pixels, and 9 categories. All bands and categories in the Pavia University data set are utilized in our experiments. The Salinas data set (Figure 3c) was acquired by the AVIRIS sensor in Salinas Valley, CA, USA. This data set includes 512 $\times$ 217 pixels, 224 bands, and 16 classes. Similar to the Pavia University data set, all bands and classes are utilized in our experiments. The details of these three data sets are listed in Table 1.

4.1. Datasets and Experimental Setup

The experiments utilize three well-known HSIs, i.e., Indian Pines, Pavia University, and Salinas.



(**a**) (**b**) (**c**)

Figure 3. Grayscale images of some bands of the three datasets. (**a**) Band 170 of Indian Pines, (**b**) Band 50 of Pavia University, (**c**) Band 100 of Salinas.

Table 1. The descriptions of three hyperspectral datasets in our experiments.

Dataset	Indian Pines	Pavia University	Salinas
Pixel	145 × 145	610 × 340	512 × 217
Band	185	103	224
Used class	16	9	16

For the pixel classification of the selected band subsets, two different classifiers, i.e., support vector machine (SVM) [35] and edge-preserving filtering (EPF) [36], are respectively utilized in our experiments. The widely used SVM classifier has good performance under a small sample size [37]. The kernel function of this classifier adopts a Gaussian radial basis function (RBF) [38]; moreover, the parameters of SVM are set by the cross-validation and grid search; furthermore, the one-against-all scheme [39] is adopted for multi-class classification. For the EPF method, we adopt it because of the availability and superiority of this classification method. The EPF-G-g classifier among the four EPF-based methods [36] is utilized in our experiment to evaluate the classification performance of different BS methods. The abbreviation G represents that the edge-preserving filtering is a guided filter, and the abbreviation g stands for the first principal component to be used as the guidance image.

As for the hyper-parameter settings of the proposed EBARec-BS method, the mini-batch size is set to 32. Moreover, the initial value of the learning rate is set to 1×10^{-3}, which is reduced by 10 times every 8 epochs. The kernel size m of the one-dimensional convolution in the EBARec module is set to 3, and the balance coefficients η and γ in the objective function are set to 3.14 and 1×10^{-2}, respectively. For the comparison method BS-Net-Conv, we use the same hyper-parameter settings in [21].

4.2. Classification Results

In this experiment, overall accuracy (OA) and average accuracy (AA) are utilized as quantitative evaluation indicators for classification. For the sake of fairness, for each hyperspectral data set, we randomly select 10% of the labeled samples from each type of ground object as the training set and the rest as the test set. Moreover, to minimize the instability caused by random selection, the final result is attained by averaging five individual runs. Figure 4 shows the OA curves of using different BS methods to select different numbers of bands on the three data sets. The number of the selected bands ranges from 5 to 30, and the performance of all bands is also drawn in Figure 4 as an important reference. Additionally, Table 2 lists the OAs and AAs when a fixed number of bands are selected for different BS methods in different data sets. Moreover, Figures 5–7 show the SVM classification maps of the band subsets obtained by different BS methods under three

HSIs. The results illustrate that the proposed EBARec-BS method obtains the best overall classification performances.

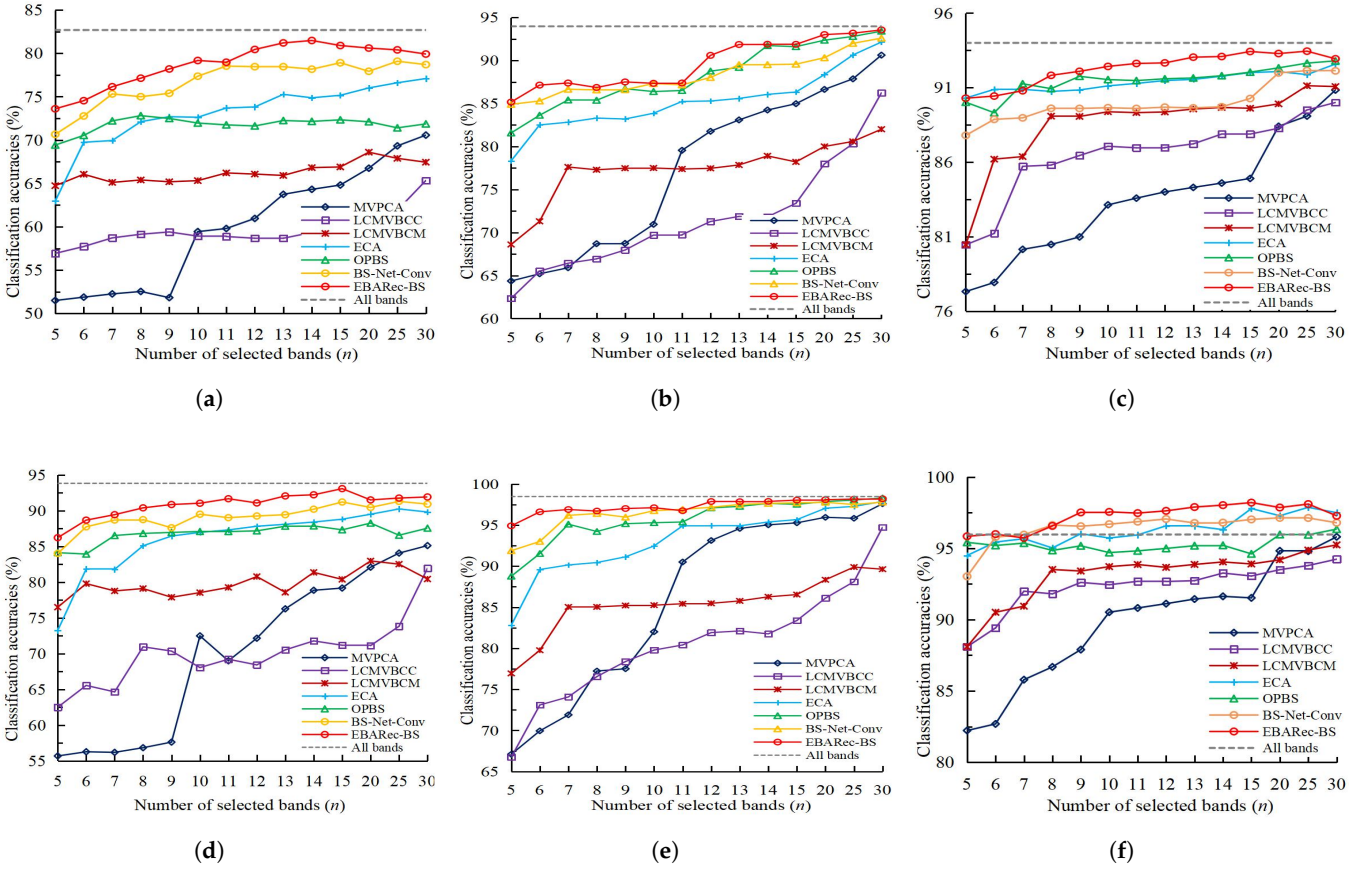


Figure 4. Overall classification accuracies of using the band subset selected by different BS methods from three different data sets. (**a**) Indian Pines-SVM, (**b**) PaviaU-SVM, (**c**) Salinas-SVM, (**d**) Indian Pines-EPF-G-g, (**e**) PaviaU-EPF-G-g, (**f**) Salinas-EPF-G-g.

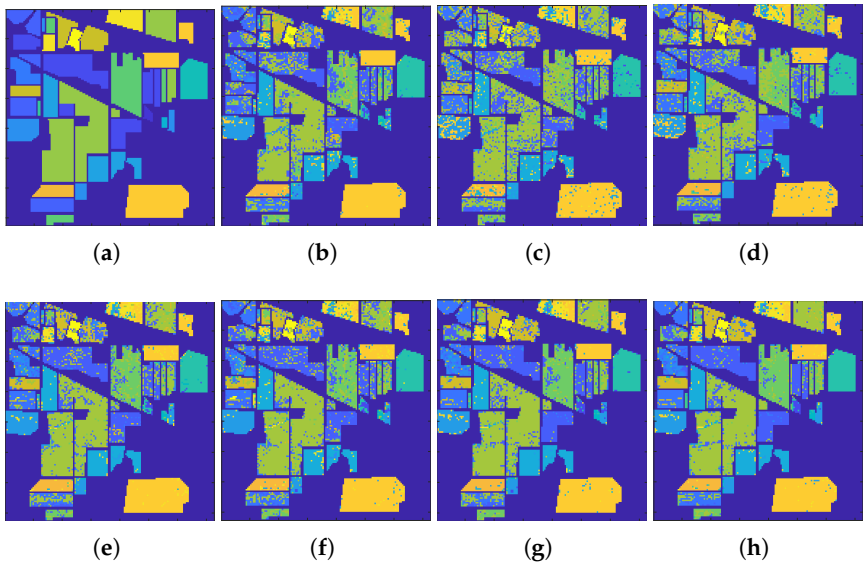


Figure 5. SVM classification maps of using the fifteen bands selected by different methods from the Indian Pines data set. (**a**) Ground truth, (**b**) MVPCA, (**c**) LCMVBCC, (**d**) LCMVBCM, (**e**) ECA, (**f**) OPBS, (**g**) BS-Net-Conv, (**h**) EBARec-BS.

Table 2. Overall accuracies (OA) (%) and average accuracies (AA) (%) of using the fifteen/ten bands selected from different datasets. (The bold denotes the best result achieved by BS methods.)

Indian (15 Bands)	**SVM**		**EPF-G-g**	
	OA (%)	**AA (%)**	**OA (%)**	**AA (%)**
1. MVPCA	64.81	50.83	79.17	67.51
2. LCMVBCC	58.95	49.74	71.17	62.01
3. LCMVBCM	66.90	60.98	80.38	73.33
4. ECA	75.16	65.25	88.80	80.03
5. OPBS	72.33	62.97	87.31	80.38
6. BS-Net-Conv	78.91	72.27	91.20	85.25
7. EBARec-BS	**80.90**	**74.30**	**93.07**	**88.60**
PaviaU (10 Bands)	**SVM**		**EPF-G-g**	
	OA (%)	**AA (%)**	**OA (%)**	**AA (%)**
1. MVPCA	70.95	55.99	82.01	70.92
2. LCMVBCC	69.70	63.76	79.79	80.29
3. LCMVBCM	77.50	67.97	85.25	83.25
4. ECA	83.86	71.88	92.46	83.87
5. OPBS	86.39	76.28	95.29	86.80
6. BS-Net-Conv	87.31	77.11	96.76	87.57
7. EBARec-BS	**87.34**	**77.15**	**97.12**	**87.59**
Salinas (15 Bands)	**SVM**		**EPF-G-g**	
	OA (%)	**AA (%)**	**OA (%)**	**AA (%)**
1. MVPCA	84.91	84.10	91.53	90.14
2. LCMVBCC	87.88	87.82	93.06	91.87
3. LCMVBCM	89.62	89.21	93.91	91.98
4. ECA	92.01	90.23	97.79	93.24
5. OPBS	92.04	90.10	94.61	92.13
6. BS-Net-Conv	90.27	89.07	97.03	92.93
7. EBARec-BS	**93.42**	**90.97**	**98.21**	**93.35**

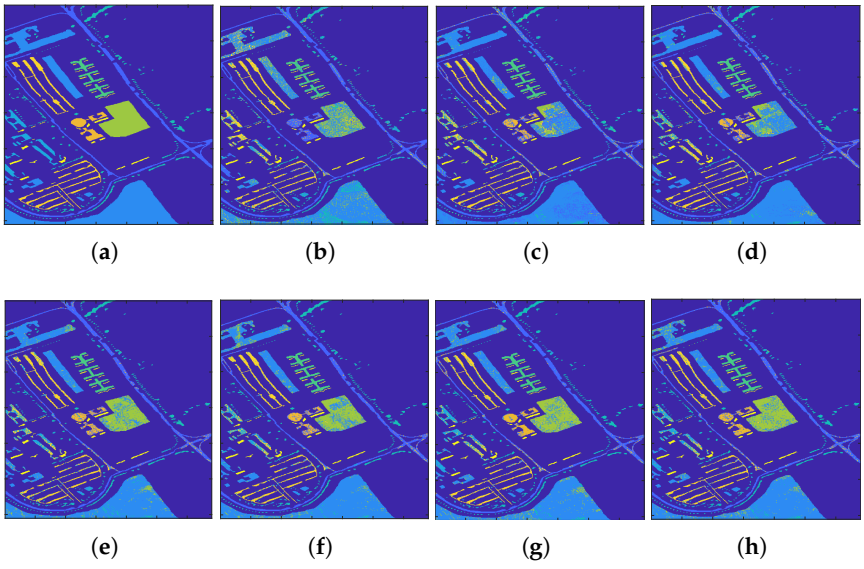


Figure 6. SVM classification maps of using the ten bands selected by different methods from the Pavia University data set. (**a**) Ground truth, (**b**) MVPCA, (**c**) LCMVBCC, (**d**) LCMVBCM, (**e**) ECA, (**f**) OPBS, (**g**) BS-Net-Conv, (**h**) EBARec-BS.

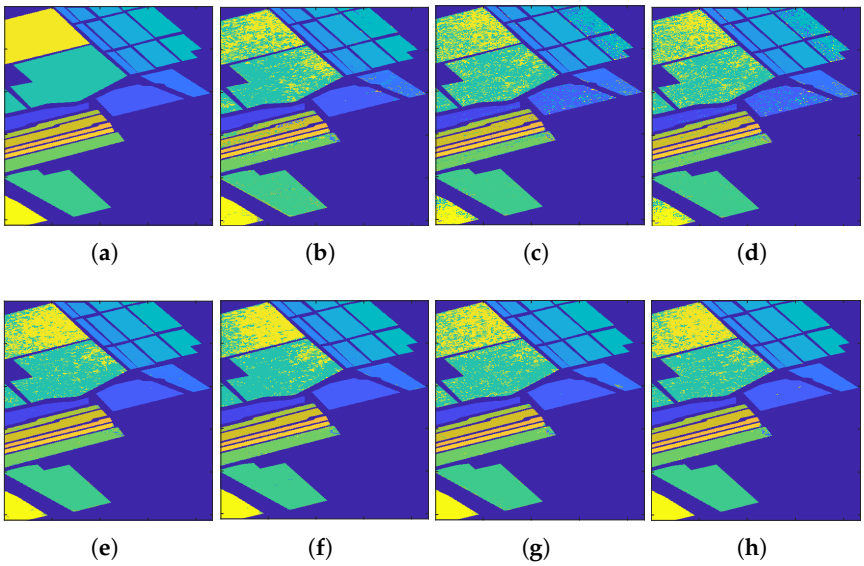


Figure 7. SVM classification maps of using the fifteen bands selected by different methods from the Salinas data set. (**a**) Ground truth, (**b**) MVPCA, (**c**) LCMVBCC, (**d**) LCMVBCM, (**e**) ECA, (**f**) OPBS, (**g**) BS-Net-Conv, (**h**) EBARec-BS.

The accuracy curves in Figure 4 show the average value of OAs of five independent running classification experiments of different BS methods in different data sets. The training set and the test set of each experiment are re-divided.

For the Indian Pines data set (Figure 4a,d), the proposed method has obvious superiority when compared with the other BS methods concerning the performances of both classifiers. For the SVM classifier, as shown in Figure 4a, the EBARec-BS method consistently achieves the best OA under different numbers of selected bands. For example, when the number of selected bands is equal to 14, the classification accuracy of the EBARec-BS method is 3.31% higher than that of the state-of-the-art BS-Net-Conv. Additionally, it can be found from the results that an increase in the number of selected bands does not always mean an improvement in classification accuracy. For example, when the number of selected bands is greater than eight, the OA of the OPBS method shows a downward trend. This can be explained by the Hughes phenomenon [1], i.e., in the case of a small sample, when the data dimension increases to a certain height, increasing the dimension will actually decrease the classification accuracy. For the EPF-G-g classifier, as shown in Figure 4d, the EBARec-BS method consistently holds the highest classification accuracy under different numbers of selected bands. The classification accuracy of the EBARec-BS method reaches 90.38% when the number of selected bands is equal to 8, while the best competitor, i.e., the BS-Net-Conv method, obtains approximate accuracy only when the number of selected bands is greater than 15. This result indicates that the EBARec-BS method can achieve excellent classification performance in a limited number of selected bands. It is worth noting that when the number of selected bands is equal to 15, the classification accuracy of the proposed EBARec-BS method is higher than the ones of compared methods and approximates the classification accuracy of all bands. Moreover, since spatial information is utilized in the EBARec-BS method and BS-Net-Conv method, these two methods are significantly better than the other comparison BS methods (i.e., OPBS, ECA, LCMVBCC, LCMVBCM, and MVPCA). Furthermore, the classification accuracy of the proposed EBARec-BS is significantly higher than that of the state-of-the-art BS-Net-Conv, which illustrates the importance of considering the characteristics of HSI and the redundancy among bands when selecting band subset.

For the Pavia University data set (Figure 4b,e), although OPBS and BS-Net-Conv obtain relatively good classification results, the proposed EBARec-BS still achieves the best overall classification performance. As shown in Figure 4b, for the SVM classifier,

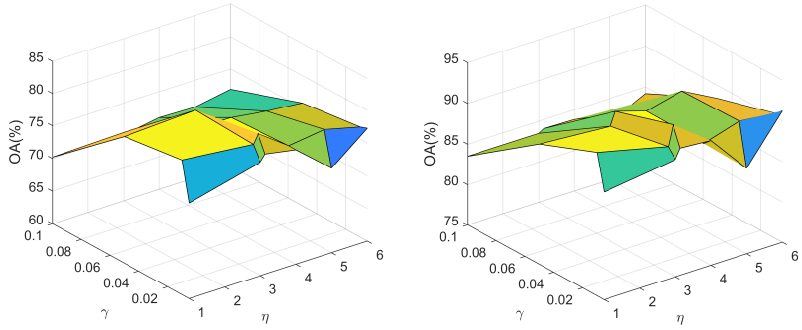
when the number of selected bands is equal to five, the proposed EBARec-BS method and the advanced BS-Net-Conv achieve similar classification performance, whereas when the number of selected bands is greater than five, the proposed EBARec-BS method achieves higher classification accuracy than BS-Net-Conv. From Figure 4e, when the number of selected bands is higher than 12, the classification accuracy of the proposed EBARec-BS method using EPF-G-g classifier is higher than that of the compared methods and approximates the classification accuracy of all bands.

For the Salinas data set (Figures 4c,f), EBARec-BS obtains the best classification results when the size of the selected band subset is between 8 and 25. For the SVM classifier, as shown in Figure 4c, EBARec-BS achieves higher OAs than BS-Net-Conv, LCMVBCM, LCMVBCC, and MVPCA. When the size of selected band subset is greater than eight, the EBARec-BS method achieves the best classification performance. From Figure 4f, for the EPF-G-g classifier, EBARec-BS, BS-Net-Conv, and ECA achieve better classification results than all bands. This phenomenon can also be explained by the Hughes phenomenon [1], that is, the classification accuracy will first increase and then decrease as the number of selected bands increases. Nevertheless, when the number of selected bands is greater than 9 and less than 25, the proposed method has obvious advantages over the comparative methods. Moreover, the EBARec-BS method achieves higher classification accuracy than the state-of-the-art BS-Net-Conv when the number of selected bands is greater than nine, indicating the superiority of the proposed method and the importance of well-considering representativeness and redundancy when selecting the optimal band subset.

The OAs and AAs when a fixed number of bands are selected for different BS methods in different data sets are listed in Table 2. To avoid the contingency of the experiment, the results in Table 2 are the average of five independent runs. It can be found from the results that the proposed EBARec-BS method consistently obtains the best OAs and AAs for three different data sets and two classifiers. For the Indian Pines data set, the proposed EBARec-BS method obtains the AAs of 74.30% and 88.60% when using the SVM classifier and the EPF-G-g classifier, respectively, which are at least 2.03% and 3.35% higher than the comparison methods. For the Pavia University data set, the proposed EBARec-BS method consistently achieves the highest OAs and AAs for two classifiers. For the Salinas data set, although most comparison methods (such as ECA, OPBS, and BS-Net-Conv) obtain relatively high OAs and AAs, the proposed EBARec-BS method is still superior to all comparison methods. Moreover, when using the SVM classifier, the OA of the proposed EBARec-BS method is at least 1.38% higher than that of the comparison methods.

To visually observe the classification performance of the band subsets selected by different BS methods, the classification accuracy diagrams of the SVM classifier used on the three data sets are shown in Figures 5–7. Specifically, the SVM classification maps and ground truth on the Indian Pines data set containing 16 feature categories are shown in Figure 5. Moreover, Figures 6 and 7 show the ground truth and SVM classification maps of the Pavia University data set and Salinas data set, respectively. As shown in Figures 5–7, the EBARec-BS method achieves better classification results than other BS methods on three different data sets.

To analyze the parameter sensitivity of the proposed model (15), the OA change trend of different combinations of balance parameters η and γ on the Indian Pines dataset is shown in Figure 8. The value of parameter η is set as $\{1, 3, 3.14, 4, 5, 6\}$, and the value of parameter γ is set as $\{5 \times 10^{-3}, 1 \times 10^{-2}, 5 \times 10^{-2}, 1 \times 10^{-1}\}$. The grid in Figure 8a shows the OA results on the SVM classifier under different combinations of parameters η and γ. It can be seen from Figure 8a that when γ is set to 1×10^{-2} or 5×10^{-2}, the classification performance is better, but when the value of γ is too large or too small, the performance is significantly degraded. For parameter η, better classification performance is achieved when the value is 3.14. For the EPF-G-g classifier, as shown in Figure 8b, the best classification performance is obtained when η and γ are 3.14 and 1×10^{-2}, respectively. Hence, we set η to 3.14 and γ to 1×10^{-2} through all the experiments.



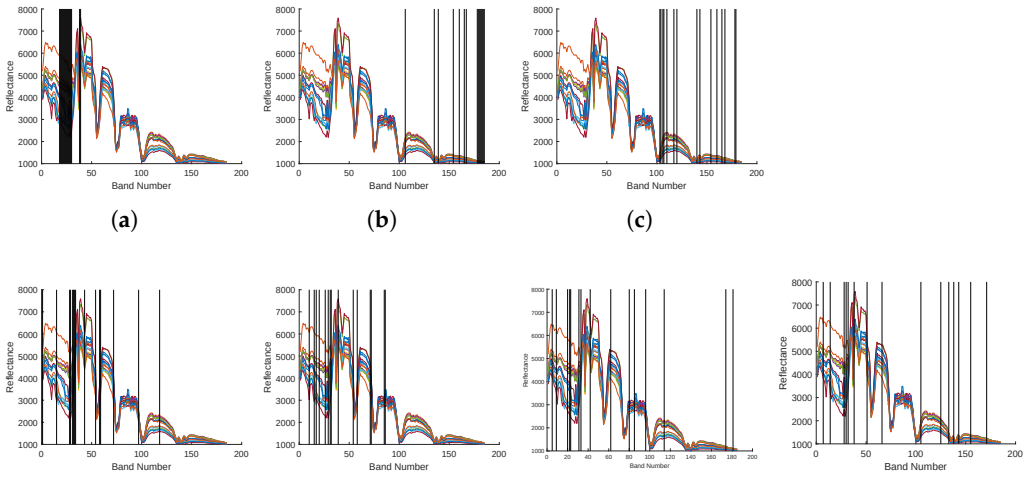
(**a**) (**b**)

Figure 8. Parameter sensitivity analysis of the proposed EBARec-BS method in terms of η and γ on the Indian Pines dataset. (**a**) SVM classification, (**b**) EPF-G-g classification.

In summary, the proposed EBARec-BS framework achieves the best overall classification performance on three different data sets, demonstrating that EBARec-BS can select the band subset that best represents the original band set and contains less redundant information. The results confirm the effectiveness of the proposed BS method.

4.3. Band Correlation Comparison

If the selected bands contain much redundant information, it is not conducive to subsequent classification tasks. To analyze the redundant information contained in the bands selected by different BS methods, we plot the distribution of the bands selected by different BS methods and the reflectance spectrum curves of different ground feature types on three different data sets in Figures 9–11, respectively. Each vertical line in the figure represents the position of each selected band. The results in Figures 9–11 show that the bands selected by the proposed EBARec-BS method are more widely and evenly distributed than those selected by other BS methods. Since adjacent bands in HSIs often contain redundant information, based on this fact, experimental results verify that the proposed BS method can select bands with little redundant information.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**) (**g**)

Figure 9. Spectrum curves of the categories on the Indian Pines data set. The vertical lines denote the fifteen bands selected by the different BS methods. (**a**) MVPCA, (**b**) LCMVBCC, (**c**) LCMVBCM, (**d**) ECA, (**e**) OPBS, (**f**) BS-Net-Conv, (**g**) EBARec-BS.

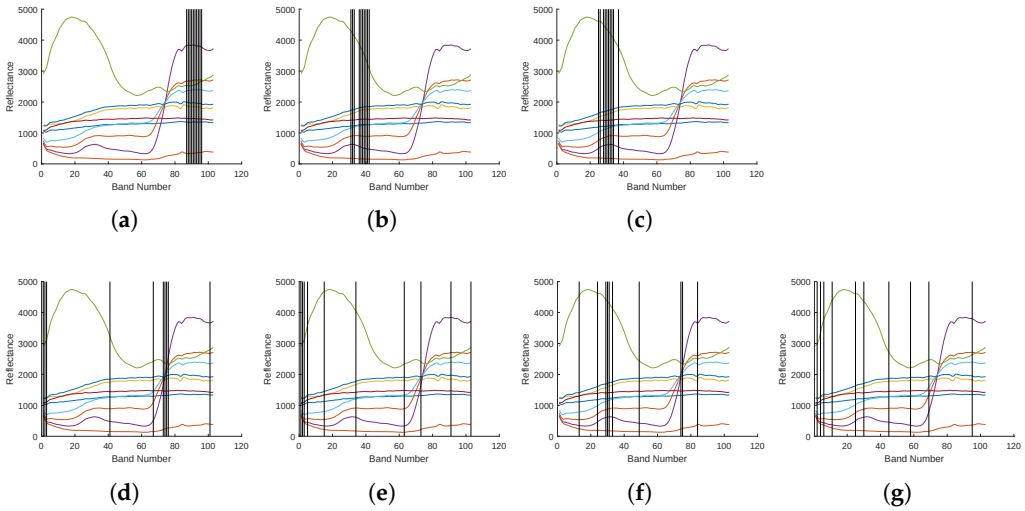


Figure 10. Spectrum curves of the categories on the Pavia University data set. The vertical lines denote the ten bands selected by the different BS methods. (**a**) MVPCA, (**b**) LCMVBCC, (**c**) LCMVBCM, (**d**) ECA, (**e**) OPBS, (**f**) BS-Net-Conv, (**g**) EBARec-BS.

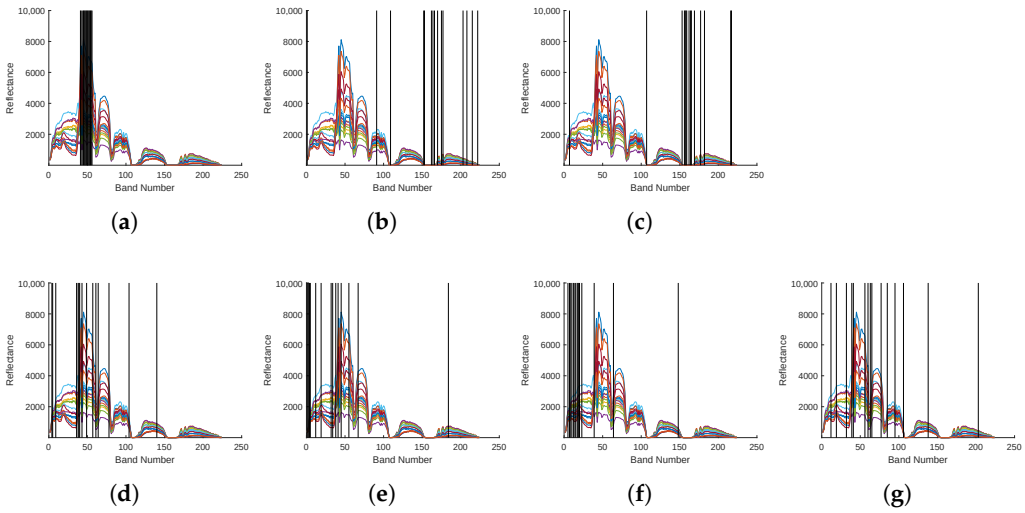


Figure 11. Spectrum curves of the categories on the Salinas data set. The vertical lines denote the fifteen bands selected by the different BS methods. (**a**) MVPCA, (**b**) LCMVBCC, (**c**) LCMVBCM, (**d**) ECA, (**e**) OPBS, (**f**) BS-Net-Conv, (**g**) EBARec-BS.

As shown in Figure 9, on the Indian Pines data set, the bands selected by the EBARec-BS method have the most extensive and uniform distribution. For the Pavia University data set, as shown in Figure 10, the bands selected by the MVPCA method are concentrated between band 85 and band 100, and the bands selected by the LCMV-based methods are mainly distributed between band 20 and band 40. Although the bands selected by the ECA method are widely distributed, they are mainly concentrated between bands 1 to 5 and bands 75 to 80. The OPBS method selects four bands between sequence numbers 1 to 5, and the bands selected by the BS-Net-Conv method are concentrated between band 25 and band 35. The EBARec-BS method selects the least adjacent bands. Similarly, the result on the Salinas data set (Figure 11) is that the bands selected by the EBARec-BS method are the most widely distributed, and the adjacent bands are the least selected, while twelve of the fifteen bands selected by the BS-Net-Conv method are distributed between band 5 and band 23. These results demonstrate that the proposed EBARec-BS method is able to select

bands with less redundant information than the comparison BS methods, which verifies the effectiveness of the proposed method.

Subsequently, we found a specific connection between the classification results and the redundancy results through a comprehensive analysis of these two types of results. Taking the Salinas data set as an example, it can be seen from Figure 11f that the state-of-the-art BS-Net-Conv does not consider the redundancy between the bands, resulting in a large number of adjacent bands being selected, so the redundancy between the selected bands is relatively high. Moreover, it can be seen from Figure 4 and Table 2 that the classification accuracy of BS-Net-Conv is not as good as that of the EBARec-BS method. Since OPBS considers the correlation between bands, the selected bands, as shown in Figures 4 and 11e, have low redundancy and high classification accuracy. However, OPBS does not consider the contribution of the selected band to the original HSI and the complex nonlinear relationship between the bands, and thus the classification performance of OPBS is not as good as that of the EBARec-BS method. As shown in Figure 11d, the redundancy between the bands selected by ECA, which is based on clustering, is not very high. However, the ECA method evaluates each spectral band as an independent point, so the classification accuracy is also lower than the proposed EBARec-BS method. As shown in Figure 11a–c, the distributions of the bands selected by the LCMV-based methods and the MVPCA method are relatively concentrated. That is, the redundancy is relatively high, and the corresponding classification effect is poor. The proposed EBARec-BS method has the highest classification accuracy and the lowest redundancy due to the consideration of redundant information and nonlinear relationships between bands and the representativeness of each band to the original band set. Similar results can be found in the Indian Pines data set and the Pavia University data set.

In conclusion, the proposed EBARec-BS method can accurately select the bands that are important to the original band set and ensure that redundant information is relatively small. Moreover, through the comprehensive analysis of the classification results and the redundancy results, it can be known that an effective BS method needs to be able to take into account the redundancy between the bands and the representativeness of each band to the original HSI simultaneously.

4.4. Robustness to Noisy Bands

To test the robustness of different BS methods to noise bands, as shown in Table 3, we select fifteen bands from the Indian Pines data set with all bands, that is, without removing noise bands. If a specific BS method selects fewer noise bands, it means that this BS method has strong robustness to noise bands.

Table 3. Fifteen bands selected by different methods from the Indian Pines dataset (the bold denotes noisy bands).

	Fifteen Selected Bands							
MVPCA	21	22	23	24	25	26	27	28
	29	30	31	32	33	41	42	
LCMVBCC	**108**	119	**152**	**154**	**155**	**156**	**158**	**159**
	160	**161**	**162**	**165**	196	**218**	**220**	
LCMVBCM	**119**	120	123	130	**153**	**155**	**159**	**160**
	165	171	174	185	196	199	209	
ECA	**1**	**2**	18	31	32	35	36	37
	46	57	61	62	75	100	101	
OPBS	**1**	18	20	23	29	32	34	35
	42	57	61	74	75	88	89	
BS-Net-Conv	**1**	6	42	68	99	**105**	**106**	**107**
	108	123	**150**	**153**	**162**	194	203	
EBARec-BS	17	18	19	20	27	33	53	130
	141	167	168	169	173	182	202	

As shown in Table 3, the EBARec-BS and MVPCA methods do not select any noise band, whereas the band subsets selected by the other BS methods all contain some noise bands. In particular, the band subsets selected by the state-of-the-art BS-Net-Conv method and the LCMV-based methods all contain more than five noise bands. Experimental results show that the proposed EBARec-BS method can select a subset of bands that represent the original HSI and is robust to noise bands, which confirms the effectiveness of the proposed BS method.

4.5. Summary

From all the experiments, some significant results can be summarized. The unsupervised BS method needs to consider the representativeness of each band to the original HSI and the correlation between bands simultaneously. Moreover, from the experimental results, it can be seen that the high correlation of the band subset often corresponds to the low classification accuracy. The proposed EBARec-BS method comprehensively considers representativeness and redundancy when selecting the band subset, so the selected band subset has the best overall classification performances and relatively low correlations on three different data sets. The classification performances of the EBARec-BS method are even better than that of the state-of-the-art BS-Net-Conv method. These results demonstrate the rationality and superiority of the proposed EBARec-BS method. In addition, EBARec-BS achieves stable and excellent classification performances on two different classifiers, which indicates the strong robustness of our proposed method. Additionally, the proposed EBARec-BS method has good robustness to noise bands. In conclusion, the experimental results verify the effectiveness of the proposed EBARec-BS method.

5. Conclusions

This article proposes a novel unsupervised EBARec-BS network for HSI. The main idea of the proposed architecture is to learn the contribution of each band to the original HSI by considering the inherent nonlinear relationship between the bands and consider the correlation among the bands by measuring the distance of a candidate band to the hyperplane consisting of the selected bands. Subsequently, we design the BS scoring function that comprehensively considers the redundancy between the selected bands and the contribution of the selected band subset to the original band set. The obtained framework can select a band subset that is not only well representative of the original band set but also has low redundancy. The experimental results demonstrate that the band subset selected by the implemented EBARec-BS method obtains significantly better classification performance and lower correlation than the band subsets selected by other BS methods. At the same time, the EBARec-BS method has good robustness to noise bands. In the future, we will explore other suitable ways to integrate the two measures of representativeness and redundancy.

Author Contributions: All the authors made significant contributions to the work. Y.L. designed the research and analyzed the results. X.L. provided advice for the preparation and revision of the paper. Z.H. and L.Z. assisted in the preparation work and validation work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Nature Science Foundation of China under Grant 61671408 and in part by the Joint Fund of the Ministry of Education of China under Grant 6141A02022362.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available at https://github.com/Liuyufei-bs/hyperspectral-images (accessed on 1 March 2021).

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

References

1. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [CrossRef]
2. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [CrossRef]
3. Fang, Y.; Li, H.; Ma, Y.; Liang, K.; Hu, Y.; Zhang, S.; Wang, H. Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1712–1716. [CrossRef]
4. Liu, Y.; Li, X.; Feng, Y.; Zhao, L.; Zhang, W. Representativeness and Redundancy-Based Band Selection for Hyperspectral Image Classification. *Int. J. Remote Sens.* **2021**, *42*, 3534–3562. [CrossRef]
5. Song, M.; Shang, X.; Wang, Y.; Yu, C.; Chang, C.I. Class Information-Based Band Selection for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8394–8416. [CrossRef]
6. Yang, H.; Du, Q.; Su, H.; Sheng, Y. An efficient method for supervised hyperspectral band selection. *IEEE Geosci. Remote Sens. Lett.* **2010**, *8*, 138–142. [CrossRef]
7. Zhang, W.; Li, X.; Zhao, L. Discovering the Representative Subset with Low Redundancy for Hyperspectral Feature Selection. *Remote Sens.* **2019**, *11*, 1341. [CrossRef]
8. Geng, X.; Sun, K.; Ji, L.; Zhao, Y. A fast volume-gradient-based band selection method for hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7111–7119. [CrossRef]
9. Zhang, W.; Li, X.; Dou, Y.; Zhao, L. A geometry-based band selection approach for hyperspectral image analysis. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4318–4333. [CrossRef]
10. Xu, Y.; Du, Q.; Younan, N.H. Particle swarm optimization-based band selection for hyperspectral target detection. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 554–558. [CrossRef]
11. Gao, J.; Du, Q.; Gao, L.; Sun, X.; Zhang, B. Ant colony optimization-based supervised and unsupervised band selections for hyperspectral urban data classification. *J. Appl. Remote Sens.* **2014**, *8*, 085094. [CrossRef]
12. 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]
13. Kim, J.H.; Kim, J.; Yang, Y.; Kim, S.; Kim, H.S. Covariance-based band selection and its application to near-real-time hyperspectral target detection. *Opt. Eng.* **2017**, *56*, 053101. [CrossRef]
14. Chang, C.I.; Wang, S. Constrained band selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1575–1585. [CrossRef]
15. Sun, K.; Geng, X.; Ji, L. Exemplar component analysis: A fast band selection method for hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 998–1002.
16. Sun, W.; Peng, J.; Yang, G.; Du, Q. Fast and Latent Low-Rank Subspace Clustering for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3906–3915. [CrossRef]
17. Feng, Y.; Yuan, Y.; Lu, X. A non-negative low-rank representation for hyperspectral band selection. *Int. J. Remote Sens.* **2016**, *37*, 4590–4609. [CrossRef]
18. 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]
19. Yuan, Y.; Zheng, X.; Lu, X. Discovering diverse subset for unsupervised hyperspectral band selection. *IEEE Trans. Image Process.* **2016**, *26*, 51–64. [CrossRef]
20. Sun, K.; Geng, X.; Ji, L. A New Sparsity-Based Band Selection Method for Target Detection of Hyperspectral Image. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 329–333. [CrossRef]
21. 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]
22. Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Plaza, A.; Chanussot, J. Graph Convolutional Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5966–5978. [CrossRef]
23. Zhao, B.; Wu, Y.; Guan, X.; Gao, L.; Zhang, B. An Improved Aggregated-Mosaic Method for the Sparse Object Detection of Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 2602. [CrossRef]
24. Yao, J.; Meng, D.; Zhao, Q.; Cao, W.; Xu, Z. Nonconvex-sparsity and nonlocal-smoothness-based blind hyperspectral unmixing. *IEEE Trans. Image Process.* **2019**, *28*, 2991–3006. [CrossRef]
25. Dou, Z.; Gao, K.; Zhang, X.; Wang, H.; Han, L. Band Selection of Hyperspectral Images Using Attention-Based Autoencoders. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 147–151. [CrossRef]
26. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* **2014**, arXiv:1409.0473.
27. Chorowski, J.; Bahdanau, D.; Serdyuk, D.; Cho, K.; Bengio, Y. Attention-based models for speech recognition. *arXiv* **2015**, arXiv:1506.07503.
28. Galassi, A.; Lippi, M.; Torroni, P. Attention in natural language processing. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**. [CrossRef]
29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.
30. Wang, F.; Tax, D.M. Survey on the attention based RNN model and its applications in computer vision. *arXiv* **2016**, arXiv:1601.06823.

31. Serre, T.; Wolf, L.; Poggio, T. Object recognition with features inspired by visual cortex. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 994–1000.
32. Masci, J.; Meier, U.; Cireşan, D.; Schmidhuber, J. Stacked convolutional auto-encoders for hierarchical feature extraction. In *Artificial Neural Networks and Machine Learning—ICANN 2011, Proceedings of the International Conference on Artificial Neural Networks, Espoo, Finland, 14–17 June 2011*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 52–59.
33. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–18 June 2020.
34. Liu, C.; Li, J.; He, L.; Plaza, A.; Li, S.; Li, B. Naive Gabor Networks for Hyperspectral Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 376–390. [CrossRef]
35. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
36. Kang, X.; Li, S.; Benediktsson, J.A. Spectral–Spatial Hyperspectral Image Classification With Edge-Preserving Filtering. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2666–2677. [CrossRef]
37. Mitra, P.; Murthy, C.; Pal, S. Unsupervised feature selection using feature similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 301–312. [CrossRef]
38. Tarabalka, Y.; Chanussot, J.; Benediktsson, J.A. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognit.* **2010**, *43*, 2367–2379. [CrossRef]
39. Rifkin, R.; Klautau, A. In Defense of One-Vs-All Classification. *J. Mach. Learn. Res.* **2004**, *5*, 101–141.