A Novel Band Selection and Spatial Noise Reduction Method for Hyperspectral Image Classification

Abstract

Dimensionality reduction (DR) is a crucial technique for enhancing hyperspectral image (HSI) classification by minimizing data redundancy. We've introduced an innovative unsupervised DR framework that merges band selection (BS) and spatial noise reduction methods. Our approach, employing a novel Neighborhood Grouping Normalized Matched Filter (NGNMF) for BS, effectively reduces data dimensions while preserving spectral details. Additionally, we've developed an Enhanced 2-D Singular Spectrum Analysis (E2DSSA) method to capture spatial context and structural information, mitigating intra-class variability and spatial noise effects. Evaluating these features using the Support Vector Machine (SVM) classifier demonstrates their effectiveness. Across three publicly available HSI datasets, our NGNMF-E2DSSA method outperforms several state-of-the-art DR techniques, validating its superior efficacy.

Keywords: Band selection (BS), dimensionality reduction (DR), enhanced 2-D singular spectrum analysis (E2DSSA), hyperspectral image (HSI), image classification.

1 Introduction

When it comes to the challenges of utilizing hyperspectral imaging (HSI) in remote sensing applications, the passage zeroes in on two core issues and their corresponding approaches. Firstly, it mentions a prevalent problem: the potential mismatch between the number of samples available and the number of spectral bands in HSI, affecting classification accuracy. This "Hughes' phenomenon" frequently poses a challenge in classification tasks due to insufficient samples to cover all spectral dimensions. To address this, the article introduces two dimensionality reduction methods: feature extraction (FE) and band selection (BS). FE aims to extract the most significant information from high-dimensional data using methods like Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA). However, these methods may lose interpretability regarding the physical or chemical properties of the objects while preserving spectral characteristics. On the other hand, BS focuses more on selecting the most relevant spectral bands but traditional methods might suffer from issues like high band correlation or sensitivity to noise.

Secondly, the article addresses the challenge of increased spatial resolution. Despite high-resolution images providing more information, they might increase intra-class variability and reduce inter-class variability, resulting in decreased classification performance. To combat this, a series of spatial feature extraction methods have been proposed, attempting to improve classification performance by leveraging contextual and structural information in the spatial domain. These methods include Extended Morphological Attribute Profiles

(EMAPs), Gabor filters, edge-preserving filtering, among others, aiming to smooth textures and noise while retaining spatial details and context. However, these methods also exhibit limitations such as feature ambiguity and loss of structural information.

Lastly, the article points out the emergence of deep learning techniques in recent years for HSI classification, yet highlights their challenges in interpretability, computational costs, and hyperparameter tuning. While deep learning excels in extracting high-level spatially invariant features, the interpretability of these features is limited, demanding substantial computational resources and parameter adjustments.

To address these challenges, the article proposes an unsupervised dimensionality reduction framework that combines band selection and spatial feature extraction. It introduces a method called Neighborhood Grouping Normalized Matched Filter (NGNMF) for band selection, integrating clustering and ranking methods to comprehensively reduce spectral redundancy. Additionally, it proposes an improved Enhanced 2D Singular Spectrum Analysis (E2DSSA) to extract significant spatial context features from each selected band while mitigating the impact of noise within the scene. Finally, the article evaluates the efficacy of the extracted features using a Support Vector Machine (SVM) classifier. This framework aims to extract prominent features from HSI that possess high interpretability, low redundancy, robustness to noise, and strong discriminative power in land cover classification.

This article introduces significant contributions:

A novel DR framework, NGNMF-E2DSSA, is presented for effective low-dimensional spectral-spatial feature extraction in HSIs. This framework combines NGNMF-based band selection (BS) with E2DSSA-based spatial processing, enabling simultaneous DR and feature extraction. This approach offers high efficiency, minimal spectral redundancy, noise resilience, and strong feature effectiveness. NGNMF employs a new band grouping strategy to reduce spectral redundancy in raw data and an enhanced MF-based ranking strategy to mitigate the impact of noisy bands. This method selects the most representative spectral bands with assured quality in BS. E2DSSA introduces a local spatial similarity-based adaptive embedding to construct a low-rank trajectory matrix. This process enhances the extraction of spatial context and structural features, improving intraclass similarity and interclass variability of land covers, surpassing many existing techniques. Experimental validation on three datasets demonstrates that NGNMF-E2DSSA outperforms several state-of-the-art DR methods and deep learning approaches in classification performance, even when working with a small number of training samples.

2 Related works

In this section, a literature review of BS methods for HSIs is presented, including clustering-based, ranking-based, and search-based FS methods (i.e., EA-based and EMT-based methods).

2.1 Ranking-based BS methods

Ranking-based BS methods quantify the importance of each spectral band based on predefined band prioritization criteria and then select the top-ranked bands in the ranking sequence. For example, Rodriguez et al. proposed a fast density-peak-based clustering method (FDPC) [10], which identifies cluster centers by estimating the local density and intra-cluster distance of each point. To make FDPC more applicable to the

hyperspectral BS problem, Jia et al. [7] proposed an improved variant of FDPC that automatically determines the appropriate number of clusters to be selected by using a ranked sample component analysis method. In addition, Varade et al. [15] applied the image denoising method to rank the importance of bands and then selected a subset of bands based on the match with the denoising error of the reference image. Xu et al. [18] proposed a similarity-based ranking strategy, which uses structural similarity to measure the relationship between bands and then selects a subset of bands based on their ability to become cluster centers. This type of BS methods has low complexity and high execution efficiency, but most of them tend to consider only the importance of each band and ignore the correlation between bands, thus the subset of bands may contain a large amount of redundancy [5].

2.2 Clustering-based BS methods

Clustering-based BS methods divide the original bands into several different clusters based on some specific metrics, and then select representative bands from each cluster to form the final subset of bands. For example, Sun et al. [13] proposed a sparse spectral clustering method based on correlation entropy to cluster the bands and then selected high-quality bands from each cluster on the connected graph. Ji et al. [6] proposed a divisive hierarchical clustering approach, which aims to obtain an arbitrary number of subsets of bands and capture the intrinsic hierarchical nature of hyperspectral bands. This method first suppresses outlier clustering by introducing average dissimilarity through local density, and then selects representative frequency bands in each cluster. Sun et al. [14] proposed a fast and latent low-rank subspace clustering method, which conducts spectral clustering on the connected graph denoted by the affinity matrix and then selects the bands that are closest to their corresponding cluster centroids to form the final band subset. This type of BS methods shows efficient performance in terms of redundancy elimination, but they select the representative bands from each cluster, which may ignore the overall performance after all the selected bands are combined. In addition, it is also difficult to determine an appropriate number of clusters in advance [5].

Table 1. Summary of various BS methods regarding the type, task, fitness function, obtaining multiple results in one run and considering class preference.

Algorithms	Year	Type	Task type	Fitness function
WDOMCS [11]	2021	supervised	Single task	Accuracy
BLNPK2021 [9]	2021	supervised	Single task	Accuracy
MABC-BS [19]	2021	unsupervised	Single task	Variation coefficient and correlation measure
HGWO [17]	2022	supervised	Single task	Inter-class dispersion matrix and intra-class dispersion matrix
MOCSA/D_BS [2]	2022	unsupervised	Single task	Information entropy, mutual information and the ratio of the relative scatter value
LFGA [1]	2022	semi-supervised	Single task	Accuracy
MTBS [12]	2022	semi-supervised	Multitask	Information entropy, mutual information and accuracy
MBBS-VC [5]	2022	unsupervised	Multitask	Variation coefficient

2.3 Search-based BS methods

Search-based BS methods treat the BS problem as a target optimization problem and then apply some nature-inspired heuristic search methods to explore the optimal subset of features. Among them, EAs have

become one of the mainstream methods in this category due to their good search capability, ease of implementation, and scalability. For example, Sawant et al. [11] proposed a hybrid EA that applies wind-driven optimization and cuckoo search to co-evolve two subpopulations, which can avoid premature convergence and obtain a good subset of features. Phaneendra et al. [9] proposed a whale optimization-based algorithm with a hybrid filter to extract intrinsic and edge-preserving spatial features, which applies a nonlinear support vector machine trained on smoothed bands to achieve effective classification. Chen et al. [2] proposed a decomposition-based multiobjective clone selection algorithm. This method first decomposes the target problem into several subproblems by a set of weight vectors, and then employs a clone selection operator to perform an efficient local search on each subproblem to obtain a high-quality subset of features.

Recently, EMT, an emerging paradigm of EAs, has received much attention because it exhibits good performance with low computational cost. Therefore, a few research studies have implemented the idea of EMT for BS of HSIs. For example, Shi et al. [12] proposed a multi-criteria semi-supervised EMT-based BS method. This method accelerates the search for promising bands by merging the band information, and then selects bands with high information content, high discrimination and low redundancy based on fully mining the numerical properties of labelled and unlabelled samples. He et al. [5] proposed an unsupervised multitasking BS algorithm based on artificial bee colony and variable-size clustering. This method uses a variable-size band clustering method to transform the BS problem as a multitasking optimization problem. Then, it applies a multitasking multi-micro-group bee colony algorithm with variable coding length to search for multiple optimal subsets of bands of different sizes simultaneously.

3 Method

3.1 Overview

The workflow presented in Figure 1 encapsulates the NGNMF-E2DSSA method, comprising three sequential steps. Initially, NGNMF handles dimensionality reduction and spectral feature extraction from the original Hyperspectral Imagery (HSI) dataset. Following this, E2DSSA is engaged to enhance spatial contextual features within each selected band, contributing to the creation of low-dimensional spectral-spatial representations. Finally, the obtained spectral-spatial features undergo classification through the Support Vector Machine (SVM) classifier. This stage involves training the SVM model to classify the enhanced spectral-spatial features accurately. Notational conventions are observed, using italicized letters for scalars, boldface lowercase for vectors, boldface uppercase for matrices, and calligraphic fonts for tensors. The input HSI data is denoted as $\mathcal{X} \in \mathbb{R}^{W \times H \times B}$, where W, H, and B represent the row, column, and band dimensions of the HSI dataset, respectively. This methodological framework integrates dimensionality reduction, spectral-spatial feature extraction, and classification, aiming to extract robust low-dimensional spectral-spatial representations crucial for accurate hyperspectral data analysis and classification.

3.2 NGNMF for BS

1) Grouping of Neighboring Bands: The B bands within dataset X are initially divided into $K(K < B)$ equal groups. Each band group, denoted as $\mathbf{X}_k(k = 1, ..., K)$, is defined as follows:

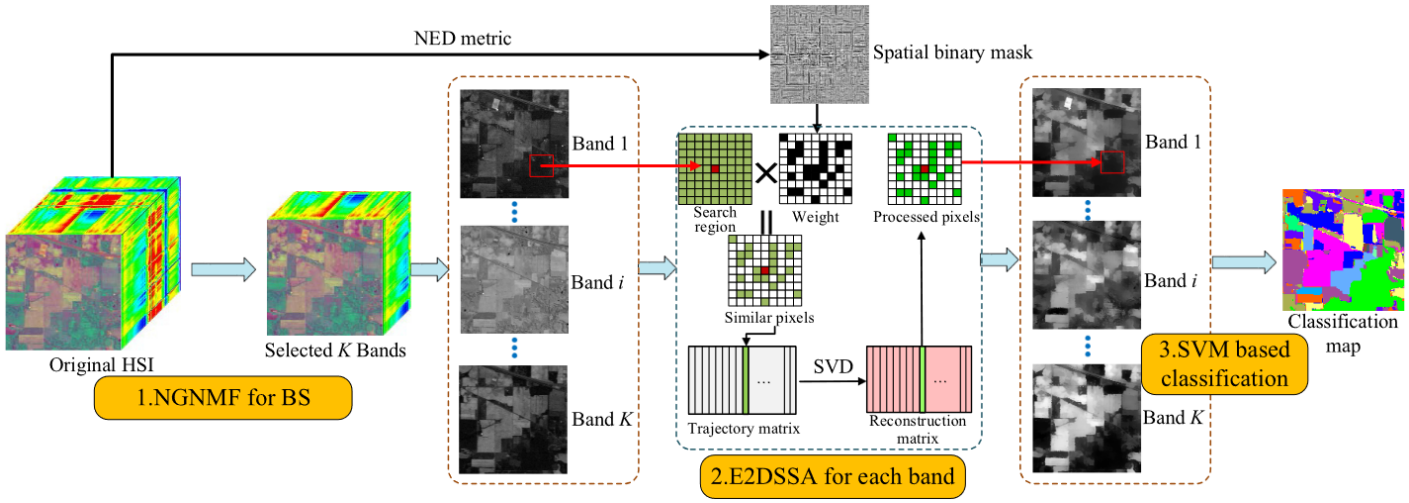


Figure 1. Flowchart of the proposed NGNMF-E2DSSA method for HSI classification.

$$\mathbf{X}_k = \begin{cases} \{\mathbf{X}^{(k-1)\left[\frac{B}{K}\right]+1}, \ldots, \mathbf{X}^{k\left[\frac{B}{K}\right]}\} & \text{if } k\left[\frac{B}{K}\right] \leqslant B \\ \{\mathbf{X}^{(k-1)\left[\frac{B}{K}\right]+1}, \ldots, \mathbf{X}^{B}\} & \text{if } k\left[\frac{B}{K}\right] > B \end{cases} \tag{1}$$

where $\mathbf{X}^{(i)}$ denotes the ith band image within $\mathcal{X}$, and $\left[\frac{B}{K}\right]$ represents the smallest integer greater than or equal to B/K.

$$p_k = \left\lfloor k \times \frac{B}{K} - \frac{B}{2K} \right\rceil, \quad k = 1, 2, \ldots, K \tag{2}$$

$$(a, b) = \begin{cases} [1, p_{k+1}), & k = 1, \\ (p_{k-1}, p_{k+1}), & 2 \leq k \leq K-1, \\ (p_{k-1}, L], & k = K. \end{cases} \tag{3}$$

The fine partition algorithm [16] is then applied to each initial band group $\text{group } \mathrm{X}_k (k = 1, ..., K)$ to obtain a new band group $\mathbf{X}_k$ $(k = 1, ..., K)$, where the number of bands in each group might vary:

$$\mathcal{X} = \{\mathbf{X}'_1, \ldots, \mathbf{X}'_k, \ldots, \mathbf{X}'_K\}, \quad \mathbf{X}'_k \in \mathbb{R}^{W \times H \times B_k} \tag{4}$$

where B_k represents the number of bands in the new group $\mathbf{X}'_k$. Post the fine band partitioning, highly correlated spectral bands are grouped together, resulting in lower inter-group correlations. Consequently, the selection of representative bands from each band group becomes more reasonable.

2) Band Prioritization Using Normalization MF: In the process of selecting representative bands, there's a risk of mistakenly choosing bands affected by water absorption or exhibiting low signal-to-noise levels. Drawing inspiration from the MF-based approach for eliminating poor bands [**?**], a modified band normalization MF ranking technique has been devised to enhance the efficacy of selecting representative bands.

Initially, each band group $\mathbb{X}'_k$ undergoes a normalization process within the range [0, 1]. Unlike the normalization using a matrix in [**?**], which minimizes band differences significantly, this normalization method fosters better band differentiation. Subsequently, each subset $\mathbf{X}'_k$ is transformed into a 2-D matrix $\{\mathrm{x}_1^k, \ldots, \mathrm{x}_i^k, \ldots, \mathrm{x}_{WH}^k\} \in$

$\mathbb{R}^{WH\times B_k}$, where x_i^k denotes the ith pixel vector with a dimension of B_k. Every pixel x_i^k is regarded as the target pixel, and its weight w_i^k is determined via the MF detector as follows:

$$\mathrm{w}_i^k = \kappa \mathbf{C}^{-1}\left(\mathrm{x}_i^k - \mathrm{m}^k\right) \in \mathbb{R}^{B_k\times 1}, \quad (i = 1, \ldots, WH; k = 1, \ldots, K) \tag{5}$$

where m^k represents the mean of all x_i^k, $\mathcal{C} \in \mathbb{R}^{B_k\times B_k}$ denotes the covariance matrix, and κ is a normalization constant.

Finally, the average of the absolute weight vectors is calculated to derive the weights of each band:

$$\left|\mathrm{w}^k\right|_{\text{mean}} = \frac{1}{WH}\sum_{i=1}^{WH}\left|\mathrm{w}_i^k\right| \in \mathbb{R}^{B_k\times 1} \tag{6}$$

where $\left|\mathrm{w}^k\right|_{\text{mean}}$ represents the weight vector corresponding to the band subset $\mathbf{X}_k'$. Bands with higher weights are assumed to possess superior signal-to-noise ratios and image quality, while lower weights imply potential problematic bands. Thus, the band with the highest weight is chosen as the most representative band.

By selecting one representative band from each group, a spectral band set $\mathcal{Y}$ with a dimension of K is obtained:

$$\mathcal{Y} = \left\{\mathbf{Y}^1, \ldots, \mathbf{Y}^i, \ldots, \mathbf{Y}^K\right\} \in \mathbb{R}^{W\times H\times K} \tag{7}$$

where $\mathbf{Y}^i$ signifies the chosen bands from the ith band group. The low-dimensional spectral feature γ is then used as input for E2DSSA for further refinement.

3.3 E2DSSA for Spatial FE

After band selection (BS), the remaining K bands preserve essential spectral features while operating at a reduced dimensionality. To efficiently extract spectral-spatial joint features, the Enhanced 2D Spectral-Spatial Association (E2DSSA) method is applied to $\mathcal{Y}$, involving two primary steps:

1) Spatial Binary Mask Generation: This step involves creating a spatial binary mask based on spectral similarity within the hypercube. For each spectral pixel $\mathbf{x}_i \in \mathbb{R}^{B\times 1}(i = 1, ..., WH)$ considered as the central pixel, a local search window of size $w \times w$ is defined. The similarity between the pixels in this search window and the central pixel is computed using the Normalized Euclidean Distance (NED) [**?**]. The similarity between any two spectral pixel vectors is given by:

$$s_{ij} = \sqrt{\sum\left(\frac{\mathbf{x}_j}{\|\mathbf{x}_j\|_2} - \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2}\right)^2} \tag{8}$$

where s_{ij} denotes the similarity between pixels x_i and $\mathbf{x}_j$, while $||\cdot||_2$ represents the 2-norm of the spectral vector.

The process involves initializing a mask of size $w \times w$ and identifying $L\,(L < w \times w)$ highly similar pixels based on s_{ij}, including the center pixel, marking them as 1, and the rest as 0. This procedure generates a spatial binary mask for each pixel, as illustrated in Figure 2. Notably, the search window size is set to 17×17 to balance efficiency and accuracy in selecting qualified similar pixels [**?**]. Since search windows may overlap, pixel mirroring of the original image is necessary to construct search windows for edge pixels.

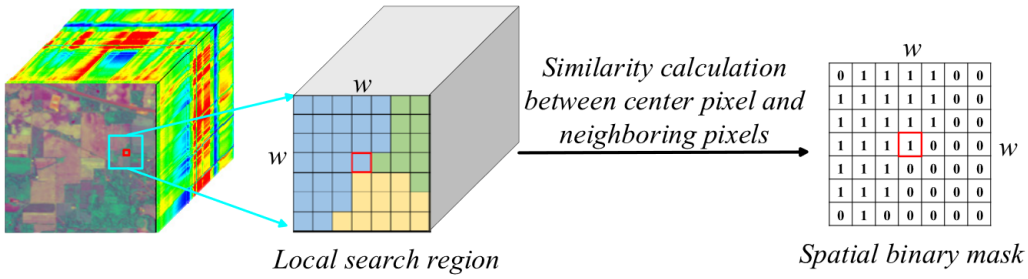


Figure 2. Spatial binary mask generation using NED metrics. Each pixel corresponds to a local search area, and these search regions overlap.

2) Band-Based Spatial Processing:

a) Adaptive Embedding: For each band image $\mathbf{Y}^i$, the spatial binary mask guides the selection of each pixel and its $(L-1)$ similar neighbors to construct a trajectory matrix. Unlike 2DSSA, our method adapts by effectively utilizing local spatial information. The L pixels around each central pixel form a column vector, creating a trajectory matrix $\mathbf{M}^i \in \mathbb{R}^{L\times WH}$. This matrix encapsulates global and local spatial features, enhancing image characterization [**?**]. Notably, $\mathbf{M}^i$ possesses certain characteristics—it's not strictly a Hankel-block-Hankel (HbH) matrix, with elements along the antidiagonal line differing slightly. Additionally, $\mathbf{M}^i$ exhibits pixel similarity in the column direction and high correlation in the row direction, showcasing a low-rank nature.

b) Singular Value Decomposition (SVD): SVD is performed on the trajectory matrix $\mathbf{M}^i$, yielding eigenvalues $(\lambda_1 \geq \lambda_2 \geq, \cdots, \geq \lambda_L)$ and their corresponding eigenvectors $(u_1, u_2, \ldots, u_L)$ of $\mathbf{M}^i\mathbf{M}^{iT}$. This decomposition splits $\mathbf{M}^i$ into submatrices $\mathbf{M}^i_j$, defined as:

$$\mathbf{M}^i_j = \sqrt{\lambda_j}\mathrm{u}_j\mathrm{v}_j, \quad \mathbf{v}_j = \left(\mathbf{M}^i\right)^T \mathrm{u}_j/\sqrt{\lambda_j}, \quad j = 1, \ldots, L \tag{9}$$

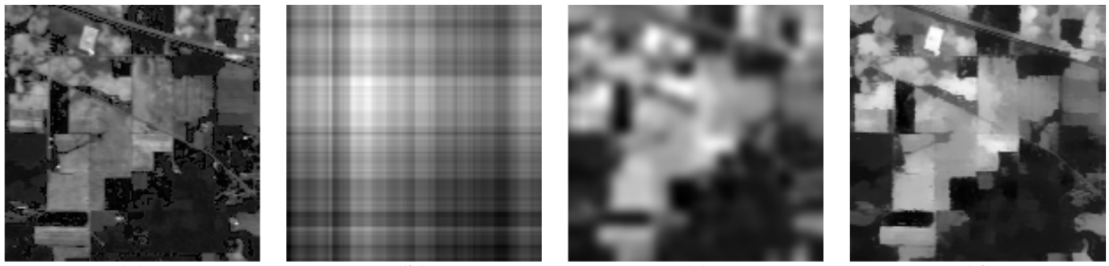
where, μ_i and υ_i represent the left and right singular vectors, respectively, while $\sqrt{\lambda_i}$ denotes the singular value of the trajectory matrix. Different $\mathbf{M}^i_i$ submatrices contain distinct spatial structures—such as trends, textures, edges, and noise.

c) Grouping: To retain key information and remove noise or insignificant components, we select the submatrix corresponding to the largest singular value for reconstruction. Leveraging the low-rank characteristic of the trajectory matrix $\mathbf{M}^i$, $\mathbf{M}^i_1$ approximates $\mathbf{M}^i$ efficiently.

d) Reprojection: The selected submatrix $\mathbf{M}^i_1$ is reprojected to the original image size, forming the refined feature image. This involves returning the processed pixels to their initial positions, necessitating the tracking of similar pixels from the prior steps. For multiple pixel values in the same position, the averaged value is used. This process enhances regional consistency in homogeneous areas while preserving pixel differences in heterogeneous regions (e.g., edges).

Compared to 2DSSA, E2DSSA's main advantage lies in constructing a low-rank trajectory matrix via adaptive embedding. The low-rank nature concentrates most spatial information in the initial component post SVD decomposition, crucial for ground object discrimination.

Spatial feature maps from E2DSSA, SVD, and 2DSSA are compared in Figure 3. Contrasting SVD, feature maps generated by 2DSSA and E2DSSA exhibit improved land cover continuity owing to the retention of local spatial information in the trajectory matrix. Moreover, E2DSSA's feature map highlights more edges and textures, enhancing differentiation among various land covers.



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

Figure 3. (a) Test image. Feature image corresponding to maximum singular value of (b) SVD, (c) 2DSSA, and (d) E2DSSA.

3.4 SVM-Based Classification

The obtained low-dimensional spectral-spatial features $\mathbb{Z}$ after E2DSSA can be represented as

$$\mathbf{Z} = \left\{\mathbf{Z}^1, \ldots, \mathbf{Z}^i, \ldots, \mathbf{Z}^K\right\} \in \mathbb{R}^{W \times H \times K} \tag{10}$$

where $\mathbf{Z}^i$ represents the feature band image of $\mathbf{Y}^i$. The spectral direction of feature $\mathbb{Z}$ contains the main spectral identification information for distinguishing the ground objects, while in the spatial direction, it retains the spatial context and structure information of the original band. where the effects of noise are removed while improving the intraclass consistency. As a result, the low-dimensional feature Z has an improved interpretability.

In addition, the choice of a classifier is crucial for the performance evaluation of the obtained features, especially with a limited number of labeled training samples in HSI. Among many classifiers, the SVM has achieved excellent performance in HSI classification due to its applicability to small samples and robustness to the data dimensionality [53],[54]. Therefore, we chose the SVM to perform the classification using the obtained low-dimensional features as input, and the results are given in Section III.

4 Implementation details

4.1 Comparing with the released source codes

The code for this paper has been entirely replicated following the original logic. Many thanks to the author of the original article, Hang Fu, for his support and guidance on the code. In addition to reproducing the classifiers mentioned in the paper such as SVM, we expanded the scope by including two new classifiers, namely KNN and RF. These additional classifiers offer increased flexibility and options for the algorithm. To enhance computational efficiency, we optimized the code by leveraging GPU acceleration, resulting in significantly faster execution times.

Ultimately, beyond the datasets mentioned in the original paper, we applied the algorithm code to several new unmanned aerial vehicle (UAV) hyperspectral datasets. This extension to new datasets demonstrates the algorithm's versatility and applicability across various scenarios. These new datasets provide ample opportunities for validation and practical application, further validating the effectiveness and robustness of the algorithm.

4.2 Datasets

This section presents the details of all the adopted datasets in our experiments, introduced as follows:

- **Indian Pines:** The Indian Pines dataset is considered the earliest test data for hyperspectral image classification, which was imaged by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) in 1992 on a patch of Indian Pines in Indiana, USA, and then intercepted at a size of 145 $\times$ 145 for annotation for hyperspectral image classification test purposes. The AVIRIS imaging spectrometer images wavelengths in the range of 0.4-2.5μm, and the Indian pine landscape contains two-thirds agriculture and one-third forest or other natural perennial vegetation. There are two major two-lane highways, a railway line, and some low-density housing, other built structures, and smaller roads. Since the scene is shot in June, some crops, corn, and soybeans are in the early stages of growth. The cover is less than 5%, and the available basic facts are assigned to 16 classes. The number of bands is reduced to 200 after removing the bands covering the water absorption area.

- **Pavia University:** The Pavia University dataset was acquired by the ROSIS sensor in 2002 and is often used for hyperspectral image classification. The sensor has a total of 115 bands and a size of 610 $\times$ 610 pixels. This dataset discarded some uninformative samples before further analysis, and after processing, the Pavia University dataset has 103 bands with a size of 610 $\times$ 340 pixels. The geometric resolution is 1.3 meters and the images have 9 classes.

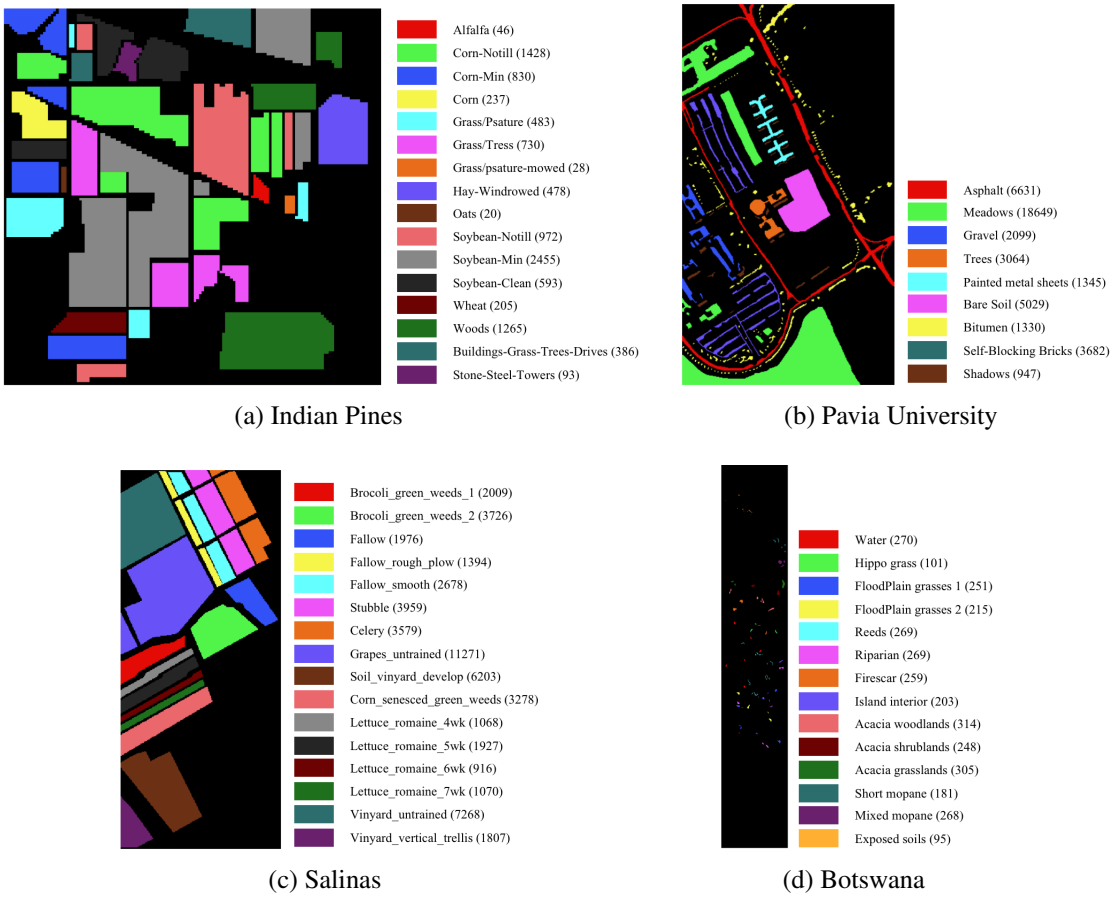
- **Salinas:** The Salinas dataset was acquired by the VIRIS sensor at 224-band AVIRIS over the Salinas Valley, California, USA, and is characterized by a high spatial resolution (3.7 meter pixels). The area covered includes 512 lines by 217 samples. As with the Indian pine scene, 20 absorption bands were discarded. The image is available only as radiometric data at the sensor. It includes 16 classes of vegetables, bare soil and vineyards, among others.

- **Botswana:** The Botswana dataset is a series of data acquired by the Hyperion sensor on board the NASA EO-1 satellite over the Okavango Delta in Botswana in 2001-2004 at 30 m pixel resolution over a 7.7 km strip. 242 bands in the 10 nm window covering the 400-2500 nm portion of the spectrum. Preprocessing of the data was performed by the UT Space Research Center to mitigate the effects of bad detectors, inter-detector calibration errors, and intermittent anomalies. The remaining 145 bands are included as candidate features by removing the uncalibrated and noisy bands covering the absorption features. Observations from 14 classes are included.

Furthermore, Fig. 4 provides the ground-truth maps and the feature classes contained in the Indian Pines, Pavia University, Salinace, and Botswana datasets, respectively.

4.3 Experimental environment setup

- ***Classification settings:*** Three commonly used classifiers, including k-nearest neighborhood (KNN) [20] , support vector machine (SVM) [8] and random forest (RF) [4], are used to classify samples on four public HSI datasets and verify the quality of the obtained feature subset. In this paper, as suggested in [16], k is set to 5 for the KNN classifier. For the SVM classifier, the RBF kernel is used and the



(a) Indian Pines (b) Pavia University

(c) Salinas (d) Botswana

Figure 4. The ground-truth maps of the Indian Pines, Pavia University, Salinas and Botswana datasets.

penalties C and gamma are initialized to 1×10^4 and 0.5, respectively. For the RF classifier, the number of tree is set to 10. In addition, the context of each dataset is not considered to simplify the classification problem. In our experiments, 20% of the samples are randomly selected for each experiment to form the training set, and the remaining 80% of the samples are used for testing.

- ***Performance metrics:*** Two well-known classification accuracy metrics are chosen to evaluate the performance of the algorithm, including the overall accuracy (OA) and KAPPA coefficient [3], introduced as follows:

  1) Overall accuracy (OA): The formula for OA is defined as follows:

$$OA = \frac{CP}{TP}, \tag{11}$$

  where CP denotes the number of correctly classified pixels and TP denotes the total number of pixels.

  2) KAPPA coefficient: The formula for KAPPA is defined as follows:

$$KAPPA = \frac{Pr(o) - Pr(e)}{1 - Pr(e)}, \tag{12}$$

  with

$$Pr(o) = OA, \tag{13}$$

$$Pr(e) = \frac{a_1 \times b_1 + a_2 \times b_2 + ... + a_n \times b_n}{n \times n}, \tag{14}$$

where $Pr(e)$ is the hypothetical probability of chance agreement. It is assumed that there are n classes of feature information, $a_1, a_2, ..., a_n$ denote the number of real samples in each category of feature information, $b_1, b_2, ..., b_n$ denote the number of samples in each type of feature information in the prediction result.

5 Results and analysis

Comparative Analysis of E2DSSA Against Classical Spatial Methods: The effectiveness of the proposed E2DSSA method was evaluated in comparison to several classical spatial methods—EMAP [25], Gabor[26], DTRF[28], and 2DSSA[29]—utilizing the Indian Pines dataset. For parameter exploration, varying embedding window sizes were employed: 3×3 to 9×9 for 2DSSA and 9 to 81 for E2DSSA.

EMAP employed $\lambda_a = 500$ for the area attribute and $\lambda_s = 15$ for the standard deviation attribute during thinning and thickening operations. The Gabor filter considered four orientations $\theta = [\pi/8, 3\pi/8, 5\pi/8, 7\pi/8]$. DTRF adjusted δ_{s} and δ_r within ranges (50 to 200 and 0.3 to 0.9, respectively) to modulate smoothness. Resulting spatial feature images on band-20 were compared visually in Fig. 10.

The analysis reveals that 2DSSA, particularly with larger embedded windows, tends to smooth out spatial details, notably edges. EMAP and Gabor, limited in feature representation under single parameters, relied on multiscale feature fusion to enhance classification. DTRF smoothed areas within classes while preserving class edges but suffered from sensitivity to noise, leading to artifacts. In contrast, E2DSSA demonstrated the ability to maintain shape and edge features of land covers while mitigating noise influence. Additionally, as the embedding window expanded, E2DSSA notably improved intraclass similarity and retained interclass differences, favoring ground object classification.

Further, quantitative assessment using SVM with only 2% labeled samples evaluated the classification accuracy. 2DSSA, DTRF, and E2DSSA were applied to each band of the HSI, employing specific parameters (7×7 window for 2DSSA, $\delta_s = 50$, $\delta_r = 0.3$, and $L = 49$ for DTRF and E2DSSA). EMAP and Gabor utilized multiple parameters from Figure 5, operating on the first three principal components.

6 Conclusion and future work

This article introduces a new approach, NGNMFE2DSSA, designed to extract easily understandable, lower-dimensional features from HSI. NGNMF works to reduce dimensions while retaining crucial spectral information, while E2DSSA enhances spatial context and structural details within each chosen band. These resulting features significantly boost HSI classification accuracy. Comparative analysis across three publicly available HSI datasets strongly supports NGNMF-E2DSSA's superiority, showcasing smoother classification maps and enhanced performance over existing DR methods.

Moreover, these extracted features stem directly from the original data space, without requiring data transformation or projection. This characteristic heightens feature interpretability. However, E2DSSA's ability to selectively extract local object features while considering global characteristics may increase computational overhead, especially on larger images. This issue parallels conventional 2DSSA challenges.

Our forthcoming efforts will focus on integrating additional techniques like image partitioning to enhance

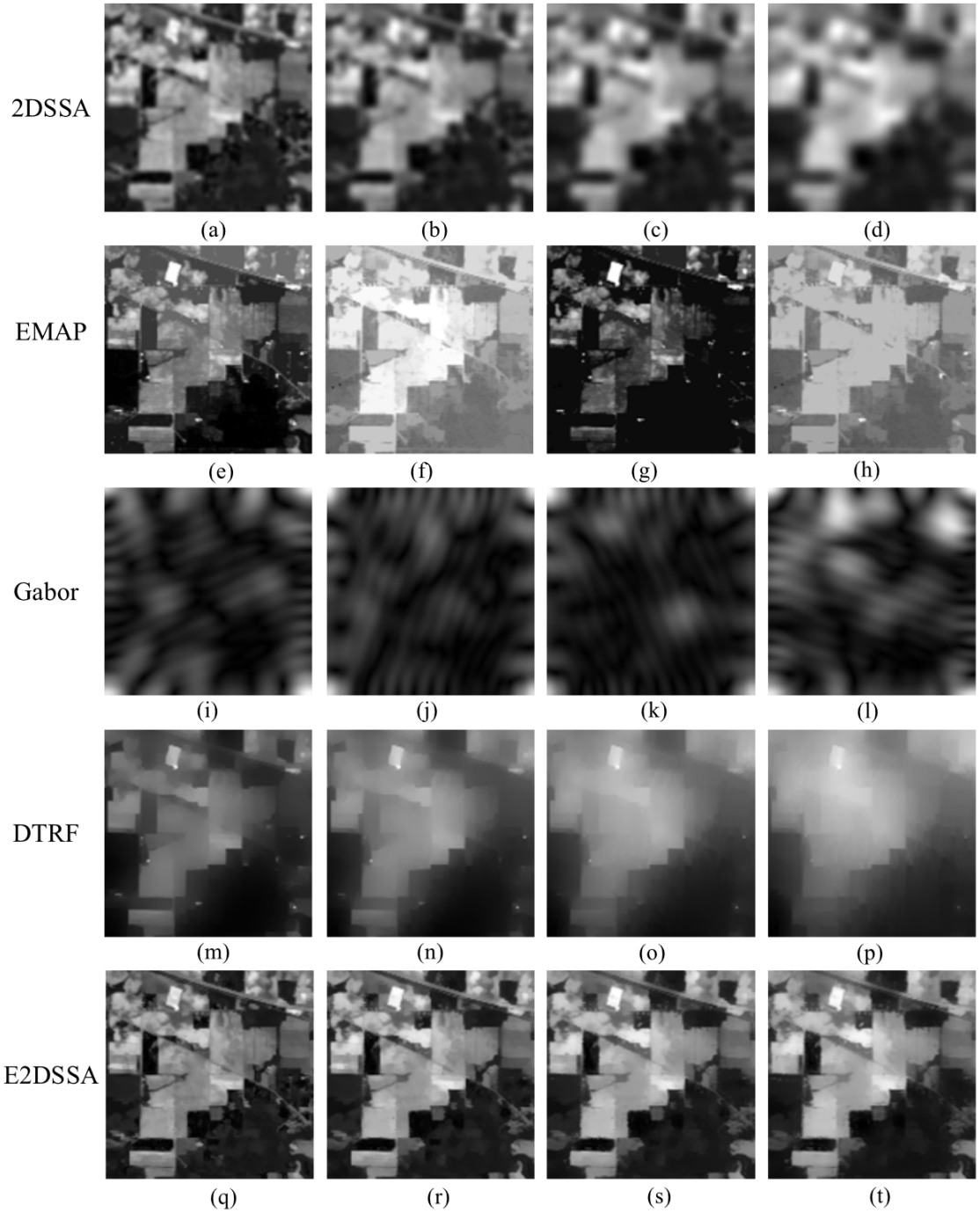


Figure 5. Feature maps obtained by (a)–(d) 2DSSA, (e)–(h) EMAP, (i)–(l) Gabor, (m)–(p) DTRF, and (q)–(t) E2DSSA with different parameters on band-20 of Indian Pines.

E2DSSA's computational efficiency. This strategic approach aims to further optimize HSI classification, making it more efficient and practical for real-world applications.

References

[1] Reza Aghaee, Mehdi Momeni, and Payman Moallem. Semisupervised band selection from hyperspectral images using levy flight-based genetic algorithm. *IEEE Geosci Remote Sens Lett*, 19:1–5, 2022.

[2] Chao Chen, Yuting Wan, Ailong Ma, Liangpei Zhang, and Yanfei Zhong. A decomposition-based multiobjective clonal selection algorithm for hyperspectral image feature selection. *IEEE Trans Geosci Remote Sens*, 60:1–16, 2022. doi:10.1109/TGRS.2022.3216685.

[3] Jacob Cohen. A coefficient of agreement for nominal scales. *Educ Psychol Meas*, 20(1):37–46, 1960.

[4] Wei Feng, Yinghui Quan, Gabriel Dauphin, Qiang Li, Lianru Gao, Wenjiang Huang, Junshi Xia, Wentao Zhu, and Mengdao Xing. Semi-supervised rotation forest based on ensemble margin theory for the classification of hyperspectral image with limited training data. *Inf Sci*, 575:611–638, 2021.

[5] Chunlin He, Yong Zhang, Dunwei Gong, Xianfang Song, and Xiaoyan Sun. A multi-task bee colony band selection algorithm with variable-size clustering for hyperspectral images. *IEEE Trans Evol Comput*, 26(6):1566–1580, 2022.

[6] Haochen Ji, Zongyu Zuo, and Qing-Long Han. A divisive hierarchical clustering approach to hyperspectral band selection. *IEEE Trans Instrum Meas*, 71:1–12, 2022.

[7] Sen Jia, Guihua Tang, Jiasong Zhu, and Qingquan Li. A novel ranking-based clustering approach for hyperspectral band selection. *IEEE Trans Geosci Remote Sens*, 54(1):88–102, 2015.

[8] Wenping Ma, Haoxiang Ma, Hao Zhu, Yating Li, Longwei Li, Licheng Jiao, and Biao Hou. Hyperspectral image classification based on spatial and spectral kernels generation network. *Inf Sci*, 578:435–456, 2021.

[9] Boggavarapu LN Phaneendra Kumar and Prabukumar Manoharan. Whale optimization-based band selection technique for hyperspectral image classification. *Int J Remote Sens*, 42(13):5105–5143, 2021.

[10] Alex Rodriguez and Alessandro Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014.

[11] Shrutika Sawant and Prabukumar Manoharan. A hybrid optimization approach for hyperspectral band selection based on wind driven optimization and modified cuckoo search optimization. *Multimed Tools Appl*, 80:1725–1748, 2021.

[12] Jiao Shi, Xi Zhang, Xiaodong Liu, Yu Lei, and Gwanggil Jeon. Multicriteria semi-supervised hyperspectral band selection based on evolutionary multitask optimization. *Knowl Based Syst*, 240:107934, 2022.

[13] Weiwei Sun, Jiangtao Peng, Gang Yang, and Qian Du. Correntropy-based sparse spectral clustering for hyperspectral band selection. *IEEE Geosci Remote Sens Lett*, 17(3):484–488, 2019.

[14] Weiwei Sun, Jiangtao Peng, Gang Yang, and Qian Du. Fast and latent low-rank subspace clustering for hyperspectral band selection. *IEEE Trans Geosci Remote Sens*, 58(6):3906–3915, 2020.

[15] Divyesh Varade, Ajay K Maurya, and Onkar Dikshit. Unsupervised hyperspectral band selection using ranking based on a denoising error matching approach. *Int J Remote Sens*, 40(20):8031–8053, 2019.

[16] Qi Wang, Qiang Li, and Xuelong Li. A fast neighborhood grouping method for hyperspectral band selection. *IEEE Trans Geosci Remote Sens*, 59(6):5028–5039, 2020.

[17] Yulei Wang, Qingyu Zhu, Haipeng Ma, and Haoyang Yu. A hybrid gray wolf optimizer for hyperspectral image band selection. *IEEE Trans Geosci Remote Sens*, 60:1–13, 2022.

[18] Buyun Xu, Xihai Li, Weijun Hou, Yiting Wang, and Yiwei Wei. A similarity-based ranking method for hyperspectral band selection. *IEEE Trans Geosci Remote Sens*, 59(11):9585–9599, 2021.

[19] Zhang Yong, He Chun-lin, Song Xian-fang, and Sun Xiao-yan. A multi-strategy integrated multi-objective artificial bee colony for unsupervised band selection of hyperspectral images. *Swarm Evol. Comput.*, 60:100806, 2021.

[20] Yuanshu Zhang, Yong Ma, Xiaobing Dai, Hao Li, Xiaoguang Mei, and Jiayi Ma. Locality-constrained sparse representation for hyperspectral image classification. *Inf Sci*, 546:858–870, 2021.