RESEARCH ARTICLE

Global-Local Consistency Constrained Deep Embedded Clustering for Hyperspectral Band Selection

SHANGFENG NING AND **WENHONG WANG**
College of Computer Science, Liaocheng University, Liaocheng 252059, China
Corresponding author: Wenhong Wang (wangwenhong@lcu.edu.cn)

ABSTRACT Hyperspectral band selection plays a key role for overcoming the curse of dimensionality in the classification of hyperspectral remote sensing images (HSIs). Recently, clustering-based band selection methods have demonstrated great potential to select informative and representative bands for hyperspectral classification tasks. However, most clustering-based methods perform clustering directly on the original high-dimensional data, which reduces their performance. To address this problem, a novel band selection method called global-local consistency constrained deep embedded clustering (GLC-DEC) is proposed in this paper. In GLC-DEC, to simultaneously learn the low-dimensional embedded representation and cluster assignments of all bands in an HSI, the stacked autoencoder is integrated with the K-means method. In addition, to reduce the adverse impact of a limited number of training samples available in HSIs, local and global consistency constraints are imposed on the embedded representation so that discriminatively consistent representation of all bands is learned. Specifically, local graph regularization and global graph regularization are introduced into the GLC-DEC model, by which the strong correlation between neighboring bands and the manifold structure of all bands are fully exploited. Based on the clustering results provided by GLC-DEC, a group of representative bands are selected by using the minimum noise method. Experimental results on two real datasets demonstrate that the proposed GLC-DEC outperformed several state-of-the-art methods.

INDEX TERMS Hyperspectral band selection, deep embedded clustering, stacked autoencoder, representation learning, graph regularization.

I. INTRODUCTION

Hyperspectral remote sensing images (HSIs) usually consist of hundreds of narrow and continuous bands, and thus can provide rich spectral and spatial information of ground objects. Currently, HSIs are applied in a wide range of applications such as environmental protection [1], [2], [3], anomaly detection [4], [5], [6], land cover analysis [7], [8], image segmentation [9], and hyperspectral classification [10], [11], [12]. For these applications, hyperspectral classification is a vital task used to identify different land covers that have distinct spectral differences. However, the high-dimensionality of an HSI cube causes the Hughes' phenomenon [13], which may lead to performance degradation in hyperspectral classification applications. An effective method to alleviate this problem is band selection, with the aim of selecting a representative subset of bands for hyperspectral classification. In contrast to feature extraction, which may lose critical information of HSIs, band selection is more physically meaningful as it effectively preserves the original spectral information.

Band selection methods have been classified as supervised [14], [15], [16], semisupervised [17], [18] and unsupervised methods [19], [20], [21] according to the proportion of

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

labeled and unlabeled samples in the training set. Compared with supervised and semisupervised methods, which need a certain number of labeled samples, unsupervised methods present excellent prospects for application as they no longer need any labelled samples [22]. In recent years, a large number of unsupervised band selection methods have been proposed by researchers. These methods can be classified as three categories: searching-based, ranking-based, and clustering-based methods. Searching-based methods choose the target band subset by optimizing certain specific criteria. Ranking-based methods generally employ defined metrics to evaluate the significance of all bands, and then select a group of bands with high ranking. To obtain discriminative bands while reducing redundant information, clustering-based methods perform band selection by grouping highly similar bands into a set of clusters as well as selecting one representative band in each cluster.

Currently, clustering-based band selection methods have demonstrated great potential to solve the band selection problem. However, most clustering-based methods perform clustering directly on the original high-dimensional data, which reduces their performance because the distance measure becomes meaningless in high-dimensional space. This oriented researchers to find the interpretable (dis-)similarity between objects in embedded space [23]. Lately, deep embedded clustering (DEC) has attracted much attention as it can learn the embedded representation and cluster assignments simultaneously. According to the network structure being used, DEC can be divided into autoencoder-based, generative-adversarial-network-based, and variational-autoencoder-based methods [24]. Among these methods, autoencoder-based methods are the most common because they are easy to implement and have the ability to avoid trivial solutions. Yang et al. [25] proposed a deep clustering network (DCN), where a stacked autoencoder (SAE) is integrated with the K-means method, allowing the embedding representation to have a clustering-friendly structure. Although existing DEC methods have exhibited significant performance improvements, they are rarely used in band selection tasks and require further study. In addition, it remains challenging to obtain clustering results that are applicable to band selection by designing effective objective functions for DEC based on the prior information of HSIs.

To select a more representative band subset for hyperspectral classification by improving DCN, we propose a novel band selection method, namely global-local consistency constrained deep embedded clustering (GLC-DEC). As there are only a few hundred bands in a typical real HSI, the embedded representation learned by DCN may have significant difference for similar bands in the original data, which reduces the effectiveness of clustering for band selection. To maintain better consistency between the original data and the embedded representation, we introduce the local and global consistency constraints into the DCN model. These constraints are designed based on the assumption that

there exist a strong correlation between neighboring bands as well as a manifold structure between all bands of an HSI. Consequently, the clustering results obtained by the proposed GLC-DEC are expected to help select a more representative subset of bands. Specifically, in GLC-DEC, the prior correlation between neighboring bands is exploited by imposing local consistency constraints on the embedded representation via local graph regularization. In addition, to make full use of the manifold structure between all bands, global graph regularization is further introduced into GLC-DEC to impose global consistency constraints on the embedded representation, which contributes to achieving a globally consistent representation of all bands. Then, the minimum noise method is employed to select representative bands from the clustering results. In the experiments, our method is compared with seven representative methods on two real datasets. The experimental results validate the effectiveness of the proposed method. The main contributions of this paper are listed as follows.

- A deep embedded clustering-based band selection method is proposed. Our method can simultaneously learn the embedded representation and clustering assignments, which enable the embedded representation to have a clustering-friendly structure. To the best of our knowledge, this is the first time that deep embedded clustering is applied to band selection.

- We introduce the global and local consistency constraints into the deep embedded clustering. These constraints are designed to make full use of the manifold structure of HSIs and the prior information that there exists strong correlation between neighboring bands. Consequently, the connectivity from the original space to the representation space can be preserved, which is conducive to obtaining better embedded representation for clustering.

- We evaluated the proposed GLC-DEC method on two real hyperspectral datasets, and the performance of GLC-DEC is compared with seven representative band selection methods. The proposed method has been demonstrated to be effective by the results of our experiments.

II. RELATED WORK

The unsupervised band selection methods can be classified into three categories: clustering-based, ranking-based, and searching-based methods. Next, some representative band selection methods will be briefly reviewed.

A. CLUSTERING-BASED METHODS

To select representative bands, the clustering-based methods usually first conduct clustering on the target HSI, then select one representative band on each cluster. Typical methods include hypergraph spectral clustering [26], multiscale superpixel-level group-clustering [27], and the adaptive subspace partition strategy (ASPS) [28]. However,

these methods perform clustering directly on the original high-dimensional data, which reduces their performance because the distance measure becomes meaningless in high-dimensional space. Recently, several studies have worked on finding the embedded representation of HSIs to obtain effective cluster assignments for band selection. Sun et al. [29] adopted an L2-norm-based regularizer to generate sparse embedded representation of high dimensional hyperspectral bands under the linear assumption, by which the subsequent cluster task can be performed on the embedded representation. Wang et al. [30] proposed a region-aware latent features fusion based clustering method. This method employs superpixel segmentation to obtain multiple spatial regions and creates the Laplacian matrix of each region. Then, the low-dimensional feature representation of each spatial region is obtained based on the corresponding Laplacian matrix that reflects the band-wise similarity of the corresponding region. Next, the shared feature representation of an HSI can be generated by integrating all latent feature representations corresponding to each region. To fully utilize the spatial information, as well as jointly learn and fuse the latent feature representations in a unified framework, Wang et al. [31] employed a hierarchical strategy to learn the low-dimensional discriminative feature representation of each spatial region. Then, the information entropy is used as a metric to select the representative bands based on the unified feature representation. Recently, deep learning techniques have been applied to band selection applications to capture the inherent nonlinear structure of high-dimensional data [32]. For instance, [33] applied deep convolutional autoencoder to transform HSIs into embedded space in a nonlinear way. Their work regards the representation learning and clustering task as an independent phase, which may lead to unfavorable representation for the subsequent clustering task.

B. RANKING-BASED METHODS

This kind of method generally selects a group of representative bands with high ranking. To rank all bands, several metrics such as variance, entropy, and information divergence are used to assess the importance of all bands. For instance, Chang et al. [34] employed principal component analysis to evaluate the prioritization of bands. However, this method neglected the redundancy between the bands. To address this problem, the divergence-based band-decorrelation scheme [35] was designed by discarding bands with differences below a preset threshold. In addition, [36] designed a band selection network, which assumes that all bands can be sparsely reconstructed by some informative bands. Then, the final bands is selected by ranking the learned sparse weights of all bands. Sun et al. [37] proposed a concrete end-to-end autoencoder-based unsupervised band selection framework (CAE-UBS) by introducing concrete random variables into the autoencoder. The representative bands are obtained by the information entropy criterion. Since ranking-based methods

mainly concern the performance of individual bands and ignore the relationship between different bands, the band subsets given by these methods usually have a high level of redundancy.

C. SEARCHING-BASED METHODS

The searching-based methods choose the target band subset by optimizing certain specific criteria. Wang et al. [38] proposed a method that regards band selection as a column subset selection problem. It selects the column subset with the largest volume as the final band subset. Geng et al. [39] proposed a volume-gradient-based band selection method, in which the computational complexity can be reduced by avoiding explicitly calculating the volume. In [40], Zhu et al. employed a structure-aware metric to evaluate the significance of all bands, and then selected representative bands by using dominant set extraction. Liu et al. [41] proposed a band grouping-based sparse self-representation band selection method. To be specific, they first divided all bands into multiple non-overlapping band subsets. Next, the representative bands is determined by selecting a subset that has the smallest reconstruction error. In addition, Wan et al. [42] evaluated the redundancy between bands by using the fitness function based on information theory. Although the searching-based methods have certain global search capabilities, most searching-based methods usually have high computational complexity since they deal with a nonlinear optimization problem.

III. PROPOSED METHOD

Fig. 1 illustrates the framework of the proposed GLC-DEC method. The model of GLC-DEC mainly consists of the model of DCN, the local graph regularization, and the global graph regularization. In DCN, the model of SAE is integrated with that of K-means so that the embedded representation and cluster assignments of all bands in an HSI is simultaneously learned. In addition, to maintain better consistency between the original data and the embedded representation, we introduce the local and global consistency constraints into the model of DCN. Specifically, the prior correlation between neighboring bands is exploited by imposing local consistency constraints on the embedded representation using local graph regularization. Meanwhile, to fully utilize the manifold structure between all bands, global graph regularization is further introduced into DCN to impose the global consistency constraints on the embedded representation, which contributes to achieving globally consistent representation of all bands. According to the clustering results available from GLC-DEC, the minimum noise method is employed to select representative bands. Let $\mathcal{M} \subset \mathcal{R}^n$ be a d-dimensional manifold embedded in $\mathcal{R}^n$, where the original data are hypothesized to be located in. Given that the encoder of SAE is expressed as $f : \mathcal{M} \to \mathcal{R}^d$, and the decoder is represented by $g : \mathcal{R}^d \to \mathcal{M}$, the objective
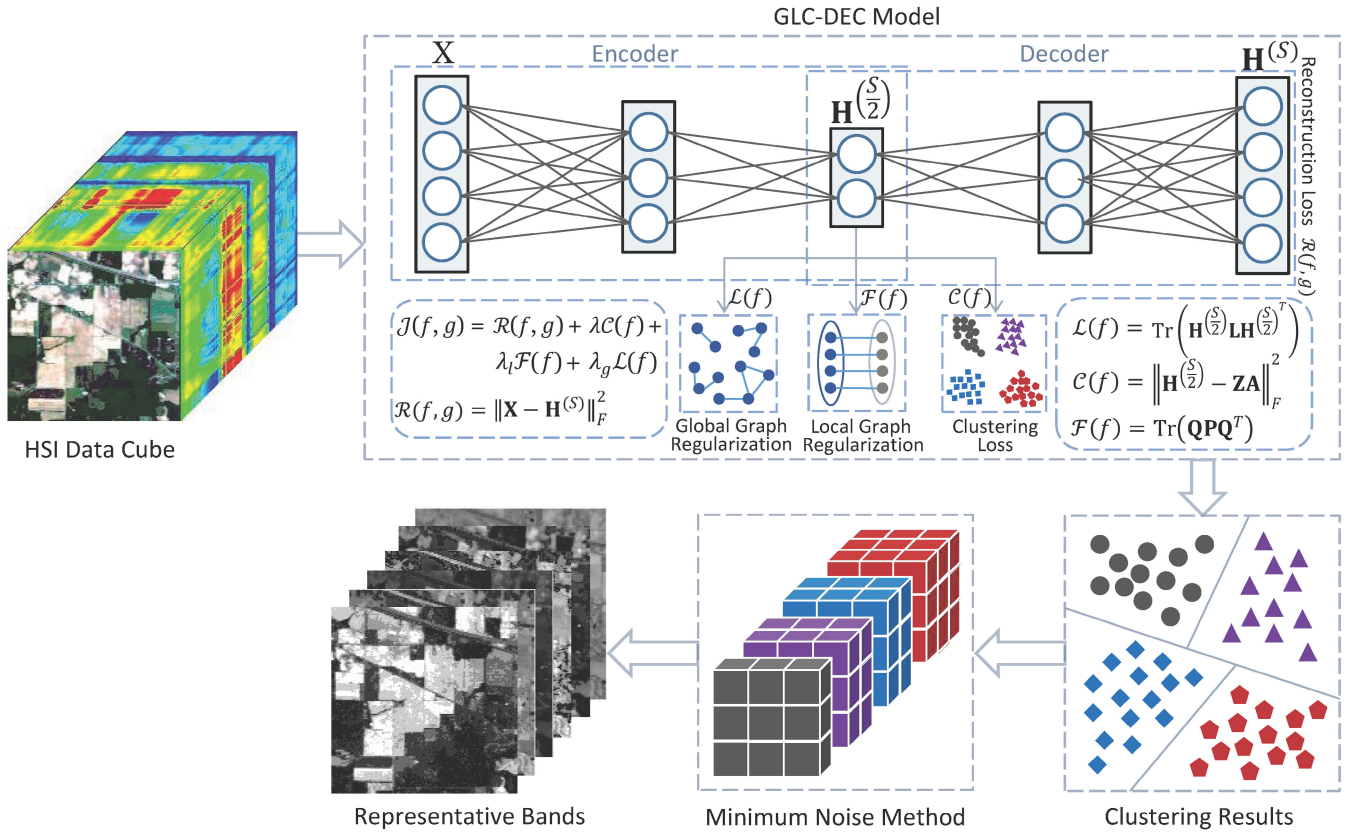
FIGURE 1. Illustration of the framework of GLC-DEC method. The model of GLC-DEC mainly consists of the model of DCN, the local graph regularization, and the global graph regularization. In DCN, the model of SAE is integrated with that of K-means so that the embedded representation and cluster assignments of all bands in an HSI is simultaneously learned. In addition, to maintain better consistency between the original data and the embedded representation, we introduce the local and global consistency constraints into the model of DCN. Specifically, the prior correlation between neighboring bands is exploited by imposing local consistency constraints on the embedded representation using local graph regularization. Meanwhile, to make full use of the manifold structure between all bands, global graph regularization is further introduced into DCN to impose the global consistency constraints on the embedded representation, which contributes to achieving globally consistent representation of all bands. According to the clustering results available from GLC-DEC, the minimum noise method is employed to select representative bands.

function of GLC-DEC can be defined as

$$\mathcal{J}(f, g) = \mathcal{R}(f, g) + \lambda \mathcal{C}(f) + \lambda_l \mathcal{F}(f) + \lambda_g \mathcal{L}(f), \quad (1)$$

where $\mathcal{R}(f, g) + \lambda \mathcal{C}(f)$ represents the loss of DCN; $\mathcal{F}(f)$ and $\mathcal{L}(f)$ represent the local graph regularization and the global graph regularization, respectively, used to impose consistency constraints on the space of function f; λ, λ_l and λ_g denote the regularization parameters. Next, we will introduce the model of GLC-DEC in detail.

A. DCN

The loss function of DCN [25] is composed of two parts, namely the SAE loss and the clustering loss. Among them, the SAE loss is utilized to acquire feature representations, while the clustering loss encourages these features to possess more discriminative properties for clustering. Specifically, the network structure of SAE includes $(S+1)$ fully connected layers, where S denotes a positive even number. The first $\frac{S}{2} + 1$ layers of SAE function as the encoder used to generate the embedded representation of the original data, while the last $\frac{S}{2} + 1$ layers form the decoder that reconstructs the

original data based on the output of the encoder. Given the original HSI data $\mathbf{X} \in \mathbb{R}^{M \times N}$, with M and N denoting the number of pixels and bands, respectively, the input of SAE is expressed as $\mathbf{H}^{(0)} = \mathbf{X}$, where $\mathbf{h}_i^{(0)} = \mathbf{x}_i$ denotes the i-th band. The output of each layer in SAE is represented as $\mathbf{H}^{(s)}$, $s = 0, 1, \ldots, S$, with

$$\mathbf{h}_i^{(s)} = \Phi^{(s)} \left(\mathbf{W}^{(s)} \mathbf{h}_i^{(s-1)} + \mathbf{b}^{(s)} \right), \quad (2)$$

where $\Phi^{(s)}(\cdot)$ indicates the activation function; $\mathbf{W}^{(s)}$ denotes the parameters between layers $(s-1)$ and s; $\mathbf{b}^{(s)}$ represents the bias parameters. To generate an embedded representation of the original HSI data, the encoder maps $\mathbf{X}$ into the potential space, and the obtained embedded representation matrix is expressed as

$$\mathbf{H}^{\left(\frac{S}{2}\right)} = [f(\mathbf{x}_1), f(\mathbf{x}_2), \cdots, f(\mathbf{x}_N)] \in \mathbb{R}^{B \times N}, \quad (3)$$

where B denotes the dimension of the embedded representation $\mathbf{h}_i^{\left(\frac{S}{2}\right)}$. Based on the output of the encoder, the output of the decoder provides the reconstructed data, which is formulated

as

$$\mathbf{H}^{(S)} = [g(f(\mathbf{x}_1)), g(f(\mathbf{x}_2)), \cdots, g(f(\mathbf{x}_N))] \in \mathbb{R}^{M \times M}. \quad (4)$$

To better reconstruct the original data, the loss function of the SAE is generally defined as

$$\mathcal{R}(f, g) = \|\mathbf{X} - \mathbf{H}^{(S)}\|_F^2, \quad (5)$$

where $\| \cdot \|_F$ represents the Frobenius norm of matrices.

Although SAE can learn the embedded representation of the original data, the generated embedded representation is not necessarily conducive to clustering. To jointly learn the cluster assignments and embedded representation, DCN integrates the loss function of K-means into SAE. More formally, the objective function of DCN can be given by [25]

$$\mathcal{J}(f, g)$$
$$= \mathcal{R}(f, g) + \lambda \mathcal{C}(f) = \|\mathbf{X} - \mathbf{H}^{(S)}\|_F^2 + \lambda \|\mathbf{H}^{\left(\frac{S}{2}\right)} - \mathbf{ZA}\|_F^2, \quad (6)$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_c, \ldots, \mathbf{z}_C] \in \mathbb{R}^{B \times C}$ represents the matrix formed by C centroids, with $\mathbf{z}_c$ denoting the centroid of the c-th cluster; and $\mathbf{A} \in \mathbb{R}^{C \times N}$ denotes the clustering assignments matrix in which each column has only one nonzero element. Note that to jointly optimize the objective function of SAE and K-means according to Equation (6), the embedded representation generated by the encoder is regarded as the input of K-means in each iteration of the optimization process of DCN.

B. CONSTRAINTS FOR CONSISTENT REPRESENTATION

Although the DCN model in Equation (6) can learn the embedded representation with a clustering-friendly structure, it neglects to exploit the prior information of HSIs. This may limit the effectiveness of clustering results for band selection. To address this problem, we impose two constraints on the embedded representation by designing effective regularization according to the assumption that there exist strong correlation between neighboring bands as well as manifold structure among all bands. Specifically, to promote the locally consistent representation between neighboring bands, we design an local graph regularization for DCN. Furthermore, aiming at learning the representation with global consistency based on the manifold structure of all bands, we introduce a global graph regularization into the model of DCN.

1) GLOBAL CONSISTENCY CONSTRAINT

To preserve the global consistency of representation based on the manifold structure of all bands, we design a global graph regularization for DCN. This regularization is proposed on the basis of the manifold assumption that the low-dimensional embedded representations of two data bands are similar if they are close to each other in the original geometric space. According to [43], the manifold structure of HSIs can be effectively captured by a K-nearest neighbor graph, in which

each vertex corresponds to a band. Accordingly, given an HSI matrix $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, we construct a K-nearest neighbor graph $G_1 = (V_1, E_1)$, where $V_1 = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ denotes the vertex set of G_1, while E_1 represents the edge set of G_1. Then, each edge $(\mathbf{x}_i, \mathbf{x}_j)$ in E_1 is constructed as follows: for each band $\mathbf{x}_i$, we find a band $\mathbf{x}_j$ from $\mathbf{x}_i$'s K-nearest neighbors based on the Euclidean distance in the original space. To measure the similarity between $\mathbf{x}_i$ and its neighbor $\mathbf{x}_j$, we calculated a weight for the edge $(\mathbf{x}_i, \mathbf{x}_j)$ by using the heat kernel weighting strategy

$$\mathbf{O}_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}}, \quad (7)$$

where $\mathbf{O} \in \mathbb{R}^{N \times N}$ represents the weight matrix [44]. Based on the matrix $\mathbf{O}$, the global graph regularization is expressed as

$$\mathcal{L}(f) = \sum_{i,j=1}^{N} \|\mathbf{h}_i^{\left(\frac{S}{2}\right)} - \mathbf{h}_j^{\left(\frac{S}{2}\right)}\|^2 \mathbf{O}_{ij} = \mathrm{Tr}\left(\mathbf{H}^{\left(\frac{S}{2}\right)} \mathbf{L} \mathbf{H}^{\left(\frac{S}{2}\right)^T}\right), \quad (8)$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of matrices, and $\mathbf{L} = \mathbf{D} - \mathbf{O}$ indicates the Laplace matrix, with $\mathbf{D}$ denoting the diagonal matrix whose diagonal entry is expressed by

$$\mathbf{D}_{ii} = \sum_j \mathbf{O}_{ij}. \quad (9)$$

With the introduced global graph regularization, the embedded representation learned by the encoder is expected to preserve better global consistency between each band and its neighbors.

2) LOCAL CONSISTENCY CONSTRAINT

In this study, to maintain the local consistency between the embedded representation of adjacent bands, we introduced the local graph regularization into DCN. This regularization is designed based on the assumption that neighbouring bands are strongly correlated with each other. Accordingly, given an HSI matrix $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, we construct a bipartite graph $G_2 = (V_2, E_2)$, in which $V_2 = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\}$ denotes the vertex collection of G_2, $E_2 = \{(\mathbf{x}_1, \mathbf{u}_1), (\mathbf{x}_1, \mathbf{u}_1), \ldots, (\mathbf{x}_N, \mathbf{u}_N)\}$ represents the set of edges of G_2. In the set V_2, $\mathbf{u}_i$ denotes the average vector of $\mathbf{x}_{i-1}$ and $\mathbf{x}_{i+1}$. To measure the similarity between $\mathbf{x}_i$ and $\mathbf{u}_i$, a weight ω_i is calculated for each edge $(\mathbf{x}_i, \mathbf{u}_i)$, where $i = 1, 2, \ldots, N$. Fig. 2 shows an illustration of the bipartite graph G_2. To facilitate the calculation of the weights of G_2, we constructed the matrix

$$\bar{\mathbf{X}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N]. \quad (10)$$

Based on $\bar{\mathbf{X}}$, the weight ω_i is defined by the heat kernel weighting strategy

$$\omega_i = e^{-\frac{\|\bar{\mathbf{x}}_{:,i} - \bar{\mathbf{x}}_{:,i+N}\|^2}{\sigma}}. \quad (11)$$
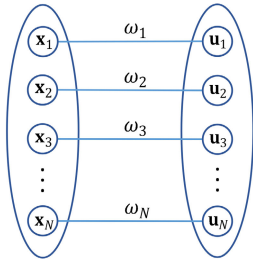
FIGURE 2. Illustration of the bipartite graph G_2.

Accordingly, the local graph regularization can be expressed as

$$\mathcal{F}(f) = \mathbf{Tr}\left(\mathbf{QPQ}^T\right), \qquad (12)$$

with

$$\mathbf{Q} = \left[\mathbf{h}_1^{\left(\frac{S}{2}\right)}, \mathbf{h}_2^{\left(\frac{S}{2}\right)}, \ldots, \mathbf{h}_N^{\left(\frac{S}{2}\right)}, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\right], \qquad (13)$$

where $\mathbf{h}_i^{\left(\frac{S}{2}\right)}$ represents the embedded representation of band $\mathbf{x}_i$, and $\mathbf{r}_i$ denotes the average vector of $\mathbf{h}_{i-1}^{\left(\frac{S}{2}\right)}$ and $\mathbf{h}_{i+1}^{\left(\frac{S}{2}\right)}$; $\mathbf{P} = \bar{\mathbf{D}} - \bar{\mathbf{O}}$ indicates the Laplace matrix, in which $\bar{\mathbf{O}} \in \mathbb{R}^{2N \times 2N}$ represents the weight matrix with $\bar{\mathbf{O}}_{i,i+N} = \omega_i$, $i = 1, 2, \ldots, N$, and $\bar{\mathbf{D}} \in \mathbb{R}^{2N \times 2N}$ denotes a diagonal matrix with $\bar{\mathbf{D}}_{nn} = \sum_m \bar{\mathbf{O}}_{nj}$, $m = 1, 2, \ldots, 2N$. Compared with global graph regularization, which is used to enforce the global consistency of embedded representation based on the manifold structure, the local graph regularization can impose the similarity constraint between the embedded representation of each band and the average of the embedded representation of its neighboring bands.

C. OVERALL MODEL

The overall objective function of GLC-DEC is expressed as

$$\begin{aligned}
\mathcal{J}(f, g) &= \mathcal{R}(f, g) + \lambda\mathcal{C}(f) + \lambda_l\mathcal{F}(f) + \lambda_g\mathcal{L}(f) \\
&= \|\mathbf{X} - \mathbf{H}^{(S)}\|_F^2 + \lambda\|\mathbf{H}^{\left(\frac{S}{2}\right)} - \mathbf{ZA}\|_F^2 \\
&\quad + \lambda_l\mathbf{Tr}\left(\mathbf{QPQ}^T\right) + \lambda_g\mathbf{Tr}\left(\mathbf{H}^{\left(\frac{S}{2}\right)}\mathbf{L}\mathbf{H}^{\left(\frac{S}{2}\right)T}\right),
\end{aligned}$$
$$(14)$$

where λ_l and λ_g denote the regularization parameters.

The advantage of GLC-DEC expressed in Equation (14) lies in that by introducing local and global consistency constraints, the strong correlation between neighboring bands and the manifold structure of all bands can be effectively exploited. Consequently, the connectivity between the original space and the representation space can be increased. This is conducive to obtaining better embedded representations for clustering. Based on the above analysis, the algorithm steps of GLC-DEC is listed in Algorithm 1.

Algorithm 1 GLC-DEC Algorithm for Band Selection

1: **Input:** HSI data $\mathbf{X} \in \mathbb{R}^{M \times N}$, the number of selected bands C, maximum iterations T, and regularization coefficients λ, λ_l, and λ_g.
2: **Output:** The band subset Υ.
3: Compute the Laplace matrix $\mathbf{L}$ of global graph regularization by Equation (7) and Equation (9);
4: Calculate the Laplace matrix $\mathbf{P}$ of local graph regularization by Equation (11);
5: Random initialize $\mathbf{W}$ and $\mathbf{b}$ according to the given network configure;
6: Pretrain SAE and obtain initial clustering centroids by running K-means in the embedded space;
7: **while** model is not convergent or maximum iterations T is not met **do**
8:     Obtain the embedded representation $\mathbf{H}^{\left(\frac{S}{2}\right)}$ by Equation (3);
9:     Update clustering assignments matrix $\mathbf{A}$ and cluster centroids matrix $\mathbf{Z}$;
10:     Compute the K-means clustering loss;
11:     Compute the local graph regularization loss by Equation (12);
12:     Calculate the global graph regularization loss by Equation (8);
13:     Obtain the reconstruction data $\mathbf{H}^{(S)}$ according to Equation (4);
14:     Calculate the reconstruction loss of SAE according to Equation (5);
15:     Update $\mathbf{W}$ and $\mathbf{b}$ via Equation (14);
16: **end while**
17: Select the representative band subset Υ using minimum noise method;
18: **return** Υ.

D. BAND SELECTION

Many conventional clustering-based methods perform band selection using criteria based on information divergence or band distance from the centroid to each band in a cluster. However, these methods do not consider noise interference, which limits the performance of hyperspectral classification when the selected bands contain significant noise. To address this problem, the minimum noise method [28] is applied to select representative bands based on the obtained clustering results. Specifically, each band in a cluster is split into η blocks, each of which has $R \times R$ pixels. Next, the noise level of each band is evaluated by calculating multiple local variances based on the blocks of the corresponding band. The local variance LV of each block is calculated by

$$\text{LV} = \frac{1}{(R^2 - 1)} \sum_{i=1}^{R^2} (T_i - \text{LM})^2, \qquad (15)$$

with

$$\text{LM} = \frac{1}{R^2} \sum_{i=1}^{R^2} T_i, \qquad (16)$$

TABLE 1. Detail information about two real HSI datasets.

Names	Size	Spatial Resolutions	Classes	Bands
Indian Pines	145 x 145 pixels	20 m/pixel	16	200
Pavia Centre	145 x 145 pixels	1.3 m/pixel	7	102

where LM indicates the average of all pixels in the current block and T_i indicates the value of the i-th pixel in this block. Next, based on the difference between the maximum and minimum variance of η blocks, k bins with equal width are divided by

$$k = (\max \mathrm{LV} - \min \mathrm{LV})/\beta, \qquad (17)$$

where $\max \mathrm{LV}$ and $\min \mathrm{LV}$ denote the maximum and minimum variance, respectively, while β denotes the partition granularity. Then, according to the value of the local variance, each block is allocated to one of these bins. The bin that contains the most blocks is an indicator of the approximate level of noise in the current band. Finally, according to the noise estimate in each band, a subset of bands is obtained by selecting a band with the minimum noise in each cluster.

IV. RESULTS

We tested the proposed GLC-DEC method on two real HSI datasets, and the performance of GLC-DEC is compared with seven representative band selection approaches: maximum-variance principal component analysis (MVPCA) [34], enhanced fast density-peak-based clustering (E-FDPC) [45], ASPS [28], deep subspace clustering (DSC) [33], region-aware hierarchical latent feature representation learning-guided clustering (HLFC) [31], band selection network (BS-Net) [36], CAE-UBS [37].

A. DATASETS

Two real HSI datasets listed in Table 1 were used to evaluate our method. Fig. 3 shows the pseudocolor images corresponding to these two datasets.

The Indian Pines dataset was collected by an airborne visible/infrared imaging spectrometer in 1992 at a test site in northwest Indiana, USA. This dataset contained 145 × 145 pixels, with each pixel comprising 220 spectral bands. Before the experiments, bands corresponded to the water absorption regions, including bands with numbers [104-108], [150-163], and 220, were deleted, and the remaining 200 bands were used. This scene contained 16 categories of objects. Table 2 shows the number of training and testing samples in each class of the Indian Pines dataset in our experiments.

The Pavia Centre dataset was acquired by the ROSIS sensor over Pavia in northern Italy. This dataset contained 102 spectral bands with an image size of 1096 × 715 pixels. The spatial resolution of this scene reaches 1.3 m. In our experiments, we selected a subscene located within [250-394, 400-544] of the Pavia Centre scene to evaluate our proposed method. This subscene contained 145 × 145 pixels and seven
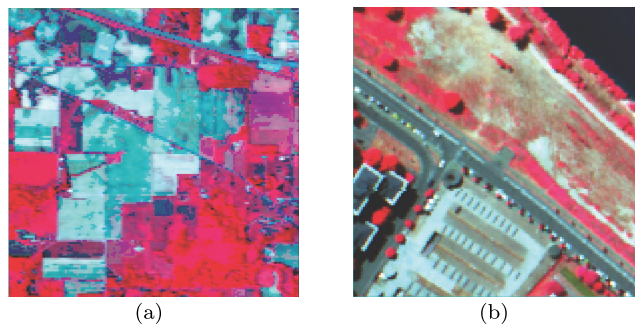
FIGURE 3. Pseudocolor images of two real HSI datasets. (a) Indian Pines. (b) Pavia Centre.

TABLE 2. The number of training and testing samples in each class of Indian Pines dataset.

Name	Training samples	Testing samples
Alfalfa	10	36
Wheat	42	163
Woods	254	1011
Buildings-Grass-Trees-Drives	78	308
Stone-Steel-Towers	18	75
Grass-pasture	96	387
Grass-trees	146	584
Corn-notill	286	1142
Corn-mintill	166	664
Corn	48	189
Oats	4	16
Soybean-notill	194	778
Soybean-mintill	492	1963
Soybean-clean	118	475
Grass-pasture-mowed	6	22
Hay-windrowed	96	382

TABLE 3. The number of training and testing samples in each class of Pavia Centre dataset.

Name	Training samples	Testing samples
Self-Blocking Bricks	223	892
Bitumen	683	2732
Tiles	175	702
Bare Soil	9	34
Trees	6	22
Asphalt	71	284
Water	61	246

classes. In our experiments, the number of training and testing samples in each class of the Pavia Centre dataset are shown in Table 3.

B. COMPARISON METHODS

Our proposed GLC-DEC is compared with seven benchmark methods, and their characteristics are summarized below:

1. E-FDPC [45]: E-FDPC belongs to the clustering-based band selection method. It can select representative bands by introducing the weighted local density and distance into FDPC, as well as adjust the representative bands via an exponential-based learning rule.

2. ASPS [28]: ASPS also belongs to the clustering-based band selection method. It first divided the 3D HSI cube into

multiple subcubes by conducting clustering on an HSI. After that, ASPS introduces local variance to evaluate the noise of bands. Lastly, the bands subset is generated by selecting a band with the lowest noise from each cluster.

3. DSC [33]: DSC is a clustering-based method with two phases. It first conducted clustering on the latent representation learned by the convolutional autoencoder. Then, the representative bands are obtained by selecting a band that is close to the cluster center in each cluster.

4. HLFC [31]: HLFC belongs to the clustering-based band selection method. It first employed superpixel segmentation to obtain multiple spatial regions. Then, Laplacian matrices that can learn latent features of all regions using the hierarchical strategy are generated based on the graphs that reflect the band-wise similarity of each region. The unified feature representation of an HSI can be generated by integrating all latent features corresponding to each region. Finally, the optimal band subset is formed by selecting a band that has maximum information entropy in each cluster given by k-means.

5. BS-Net [36]: BS-Net is a ranking-based method, which regards band selection as a band reconstruction task with the assumption that each band is sparsely reconstructed by some informative bands. Based on the learned sparse weights of bands, the final bands are selected by ranking the average band weight of each band.

6. CAE-UBS [37]: CAE-UBS is a ranking-based method, which proposed a concrete autoencoder by introducing concrete random variables into the autoencoder. Based on the weights of the concrete autoencoder, CAE-UBS can generate candidates for band selection. Then, the representative bands is chosen by the information entropy criterion.

7. MVPCA [34]: MVPCA belongs to the ranking-based method. It first estimates the priority of all bands via the definition of a variance-based band power ratio, and then regards the bands that have the high priority as representative bands.

C. EXPERIMENTAL SETUP

To compare the performance of all the methods, we used the classification accuracy given by support vector machine (SVM) and K-nearest-neighbor (KNN) classifiers as the performance metric. In our experiments, we select the appropriate proportion of training samples by cross-validation. According to the result of cross-validation, as shown in Tables 2–3, we randomly choose 20% of samples from each class as the training data, while keeping the remaining 80% samples as the testing data. To demonstrate the overall accuracy of all methods, we tested them on the two datasets while using different amounts of bands (5-50 with an interval of 5). Note that we evaluated the performance of all methods using three accuracy measures: overall accuracy (OA), average overall accuracy (AOA), and kappa coefficient (Kappa).

We implemented GLC-DEC method using the PyTorch platform. The encoder part of SAE has three hidden layers which include 700, 100 and 20 neurons, respectively. The

decoding network has a mirror structure. The learning rate was 1.0×10^{-4} and the maximum iterations T was 500. Three regularization parameters were set to $\lambda = 0.05$, $\lambda_g = 10,000$, and $\lambda_l = 100$, respectively. The network and parameter settings were the same for both the Indian Pines and Pavia Centre datasets. According to reference [28], the relevant parameters in the minimum noise method were set to $\eta = 841$, $R = 5$, $\beta = 3$, and $k = 5$ on the two datasets.

D. EXPERIMENTAL RESULTS

The AOA and Kappa values calculated based on the results given by SVM and KNN on Indian Pines and Pavia Centre datasets are compared in Fig. 4 and Fig. 5, respectively. It is obvious that GLC-DEC exhibits the best performance on both datasets. With the SVM classifier, HLFC achieved the second-best performance. Meanwhile, ASPS provided the second-best performance with the KNN classifier. In addition, Fig. 6 (a) and (b) show the OA values of SVM and KNN, respectively, when all methods were tested on the Indian Pines dataset. Concerning the Pavia Centre dataset, Fig. 7 (a) and (b) present the OA curves corresponding to SVM and KNN. Note that the OA curves in Fig. 6 and Fig. 7 show the performance of all the methods when they yielded 5-50 selected bands with an interval of 5.

To further evaluate the performance of GLC-DEC, the classification accuracy for each class on the two datasets using 20 selected bands is listed in Tables 4–5. The optimal values are indicated in red bold font, while the second-best values are indicated in blue italic font. According to Tables 4–5, it can be seen that some band selection methods are slightly unstable on different datasets. For instance, BS-Net performs better on the Indian Pines dataset than on the Pavia Centre dataset, the same trend is also observed for the DSC method. Moreover, GLC-DEC provides the best or second-best performance in most cases, which demonstrates that our method has superior performance on both datasets. In addition, detailed analysis of the OA values for all methods on the Indian Pines and Pavia Centre datasets is given as follows.

On the Indian Pines dataset, as illustrated in Fig. 6 (a) and (b), our proposed GLC-DEC method outperformed all the alternatives when using the SVM and KNN classifier. Especially, when selecting 5 and 15 bands, our method exhibited a significant advantage. According to Fig. 6 (a), we can see that the OA value of GLC-DEC improves by 6% over the second-best method when selecting 5 bands. As shown in Fig. 6 (b), with the KNN classifier, GLC-DEC outperforms the second-best method by 2% when 15 bands were selected.

As to the Pavia Centre dataset, we can see from Fig. 7 (a) that our proposed GLC-DEC method outperforms other methods with the SVM classifier except for the case that 10 and 15 bands are selected. Although HLFC performs similarly to GLC-DEC when selecting 10 and 15 bands, GEC-DEC performs slightly better at the other bands. As shown in Fig. 7 (b), MVPCA and DSC had lower

TABLE 4. The comparison of classification accuracy of all methods for each class of the Indian Pines dataset, where GLC-DEC is our method and larger values indicate better performance.

Class Name	GLC-DEC	E-FDPC	BS-Net	ASPS	CAE-UBS	MVPCA	DSC	HLFC
Alfalfa	83.78	0	64.86	51.35	40.54	72.97	59.46	81.08
Wheat	98.78	97.56	95.73	98.17	93.96	83.54	95.12	95.12
Woods	95.85	95.36	96.94	96.15	95.55	95.06	96.84	94.66
Buildings-Grass-Trees-Drives	55.99	8.74	44.98	55.34	46.93	30.42	61.81	53.81
Stone-Steel-Towers	95.95	78.38	82.43	93.24	94.59	93.24	91.89	91.89
Grass-pasture	94.82	53.63	88.34	88.08	94.36	84.29	89.64	91.97
Grass-trees	97.09	94.35	95.21	96.06	96.42	91.78	97.95	97.77
Corn-notill	76.36	53.15	73.82	68.04	71.45	27.06	71.63	76.15
Corn-mintill	54.82	43.83	56.63	50.67	49.85	51.05	51.27	53.96
Corn	71.05	16.32	54.74	53.68	57.37	52.63	65.26	60.02
Oats	87.54	0	43.75	75.01	50.83	31.25	56.25	68.75
Soybean-notill	75.19	34.32	68.89	69.79	68.25	54.63	64.14	75.12
Soybean-mintill	86.35	79.33	84.06	83.91	83.64	81.16	82.74	85.97
Soybean-clean	72.36	43.67	66.46	73.63	75.11	16.46	55.06	79.94
Grass-pasture-mowed	86.36	0	81.82	63.64	72.73	18.18	77.27	59.09
Hay-windrowed	99.48	98.95	98.43	99.21	98.69	97.91	8.17	99.45

TABLE 5. The comparison of classification accuracy of all methods for each class of the Pavia Centre dataset, where GLC-DEC is our method and larger values indicate better performance.

| Class Name	GLC-DEC	E-FDPC	BS-Net	ASPS	CAE-UBS	MVPCA	DSC	HLFC
Bricks	94.51	91.59	92.49	93.39	89.35	71.31	91.26	94.18
Bitumen	97.99	96.85	96.63	97.55	97.04	93.74	96.08	97.46
Tiles	97.15	96.36	96.44	95.29	96.72	72.51	95.87	96.93
Bare soil	98.91	98.25	97.06	98.54	97.06	67.65	98.07	98.42
Trees	86.36	59.09	76.09	77.27	63.64	4.55	45.45	86.27
Asphalt	95.03	91.28	93.48	89.08	88.38	66.98	85.42	94.15
Water	100.00	100.00	100.00	100.00	100.00	96.75	100.00	100.00
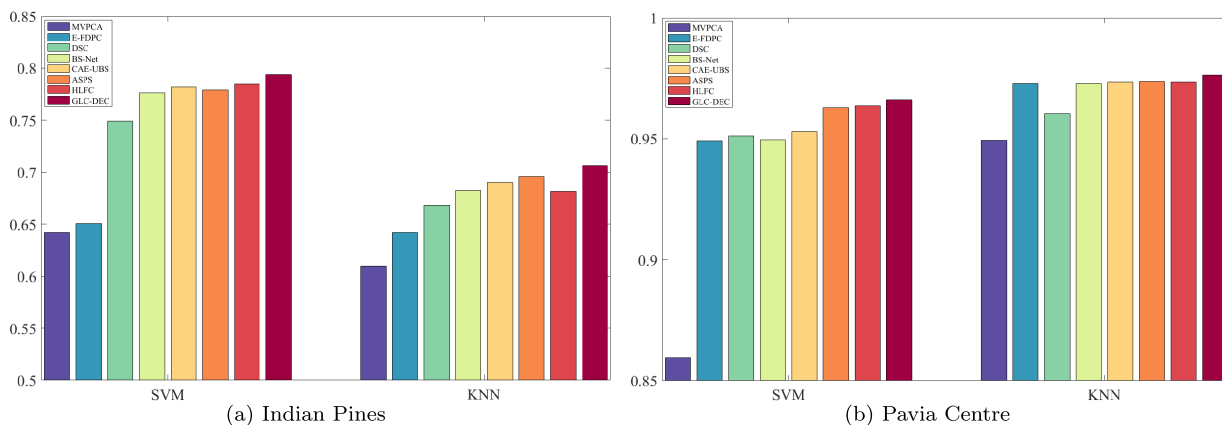
FIGURE 4. Average Overall accuracy (AOA) on two datasets for seven band selection methods.

performance, while GLC-DEC has higher OA values on all bands.

In addition, Fig. 8 and Fig. 9 compare the ground truth classification information to the classification maps based on GLC-DEC when 20 bands were selected. These figures demonstrate the superior classification of the GLC-DEC method for both datasets. It is worth noting that only 20 selected bands were used, which account for 10% and 20% of the original Indian Pines and Pavia Centre datasets, respectively. Thus, our proposed method can reduce a lot of

data redundancy while maintaining satisfactory classification results for HSIs.

V. DISCUSSION

According to the performance comparison as illustrated in Fig. 6 and Fig. 7, it can be observed that the amount of selected bands has a significant impact on the classification performance. Specifically, we can see form Fig. 6 and Fig. 7 that the OA values rise slowly or even decreases when the selected bands reached a certain number instead of keeping
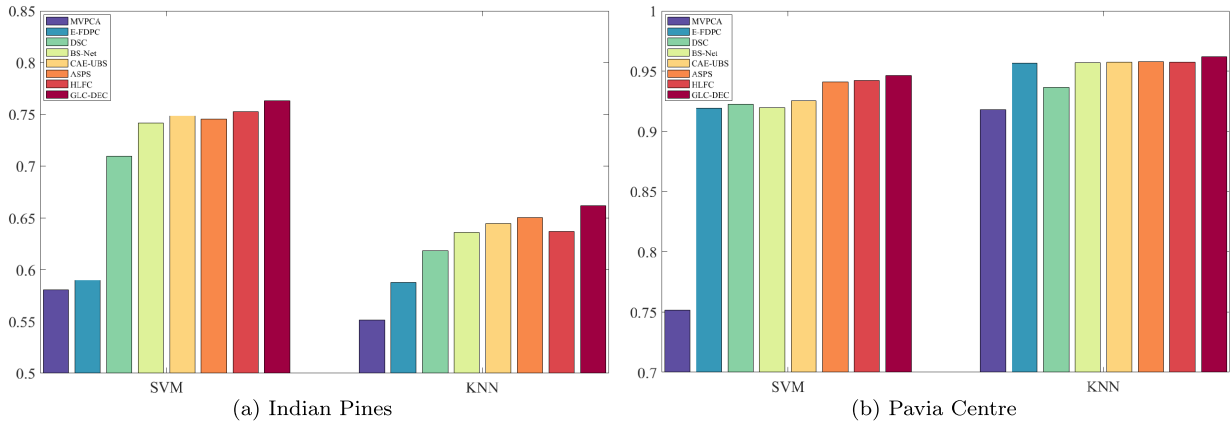
(a) Indian Pines

(b) Pavia Centre

FIGURE 5. Kappa coefficient (Kappa) on two datasets for seven band selection methods.
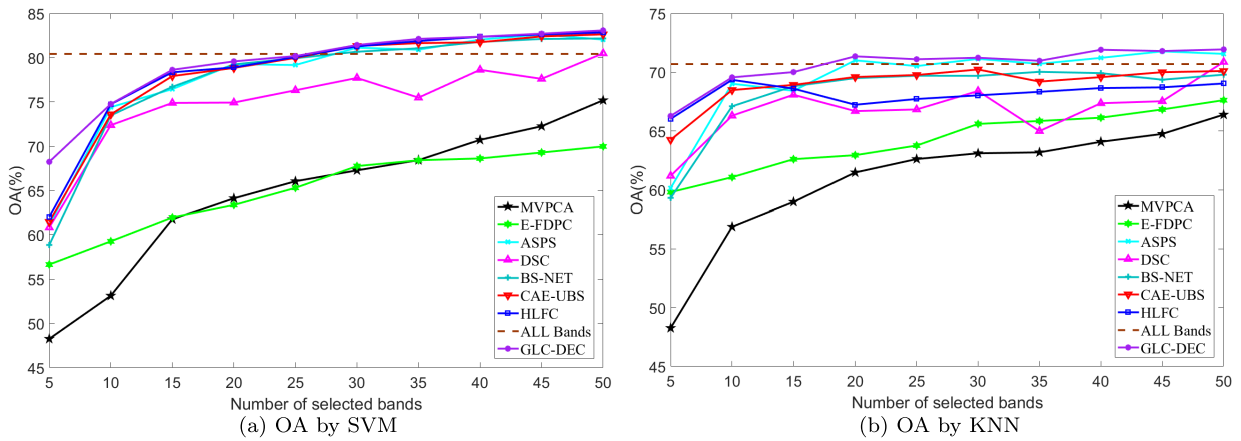

(a) OA by SVM

(b) OA by KNN

FIGURE 6. Overall accuracy (OA) for the support vector machine (SVM) and K-nearest-neighbor (KNN) classifiers on the Indian Pines dataset for different numbers of bands.
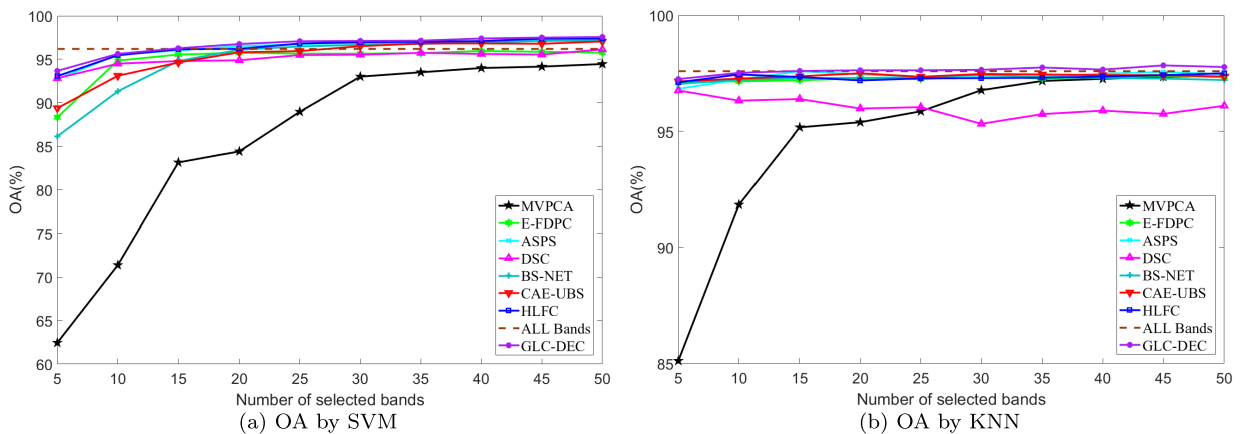

(a) OA by SVM

(b) OA by KNN

FIGURE 7. Overall accuracy (OA) for the support vector machine (SVM) and K-nearest-neighbor (KNN) classifiers on the Pavia Centre dataset for different numbers of bands.

significantly rising with the increasing of selected bands. Previous studies have also shown that the performance classification is not exactly proportional to the amount of bands [46].

This can be interpreted as more bands leading to more redundancy. Therefore, selecting a reasonable number of bands is necessary to achieve optimum classification performance.
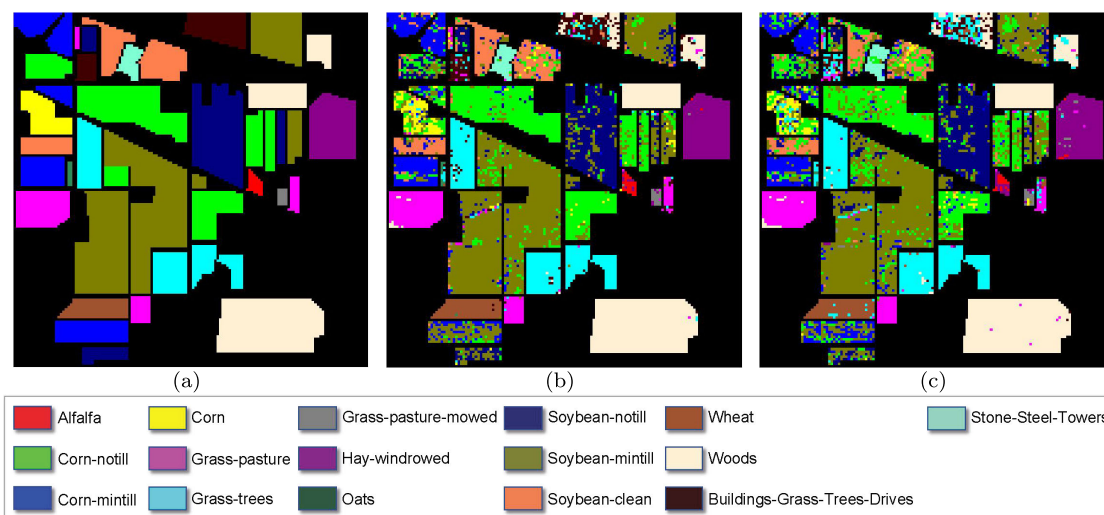
FIGURE 8. Ground truth map and classification maps generated by GLC-DEC using two classifiers on the Indian Pines dataset. (a) Ground truth. (b) GLC-DEC by SVM. (c) GLC-DEC by KNN.
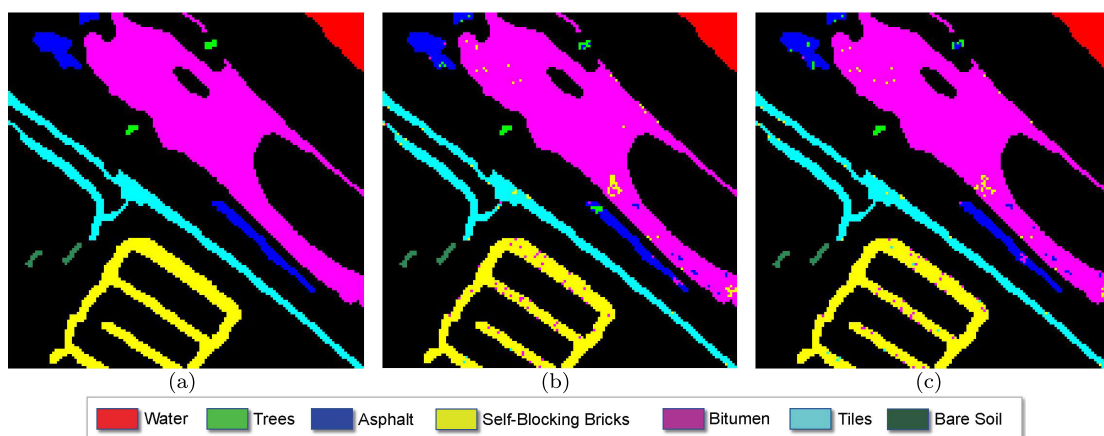

FIGURE 9. Ground truth map and classification maps generated by GLC-DEC using two classifiers on the Pavia Centre dataset. (a) Ground truth. (b) GLC-DEC by SVM. (c) GLC-DEC by KNN.

The advantage of our method lies in that the performance of GLC-DEC has superior performance on different classifiers, and is robust to the number of selected bands. As shown in Fig. 6 and Fig. 7, our method is optimal with SVM and KNN classifiers when the band is selected in the range 5-50 with an interval of 5. Compared with ranking-based methods which neglect to consider the similarity between the selected bands, our method chooses one band from each cluster, which can avoid high similarity among the selected bands. Compared with other clustering-based methods (i.e., E-FDPC, ASPS, DSC, and HLFC), our method has better performance because GLC-DEC can not only jointly learn embedded representation and clustering assignments of all bands, but also consider the similarity between neighboring bands and manifold structure of HSIs.

Our study also has some limitations. On the one hand, the conventional K-means clustering method is used in the GLC-DEC method, while the performance sensitivity of

our method to different clustering methods (e.g., spectral clustering and Gaussian mixture models) deserves further investigation. On the other hand, although the minimum noise method considers the noise interference to band selection, the affect of noise is not considered in our proposed model. In future studies, we will explore the deep embedded clustering based on other clustering methods and noise modeling to obtain more effective clustering results for band selection.

VI. CONCLUSION

In this paper, we proposed a novel hyperspectral band selection method based on the deep embedded clustering. Compared with conventional clustering-based band selection methods, our proposed method can achieve more effective clustering results for band selection by simultaneously learning the embedded representation and cluster assignments. Specifically, by introducing consistency-related constraints

into the DCN model, the strong correlation between neighboring bands and the manifold structure of all bands of HSIs are fully exploited. Consequently, the obtained clustering results achieved by our proposed method are more applicable to the task of band selection. The experimental results on two real datasets demonstrated that our method outperformed seven state-of-the-art methods.

REFERENCES

[1] A. C. Curcio, L. Barbero, and G. Peralta, "UAV-hyperspectral imaging to estimate species distribution in salt marshes: A case study in the Cadiz Bay (SW Spain)," *Remote Sens.*, vol. 15, no. 5, p. 1419, Mar. 2023.

[2] Z. Li et al., "In-orbit test of the polarized scanning atmospheric corrector (PSAC) onboard Chinese environmental protection and disaster monitoring satellite constellation HJ-2 A/B," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4108217.

[3] Z. Huang, W. Huang, S. Li, B. Ni, Y. Zhang, M. Wang, M. Chen, and F. Zhu, "Inversion evaluation of rare Earth elements in soil by visible-shortwave infrared spectroscopy," *Remote Sens.*, vol. 13, no. 23, p. 4886, Dec. 2021.

[4] Y. Gao, T. Cheng, and B. Wang, "Nonlinear anomaly detection based on spectral–spatial composite kernel for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 7, pp. 1269–1273, Jul. 2021.

[5] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500117.

[6] C.-I. Chang, "Target-to-anomaly conversion for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5540428.

[7] N. Dahiya, S. Singh, S. Gupta, A. Rajab, M. Hamdi, M. Elmagzoub, A. Sulaiman, and A. Shaikh, "Detection of multitemporal changes with artificial neural network-based change detection algorithm using hyperspectral dataset," *Remote Sens.*, vol. 15, no. 5, p. 1326, Feb. 2023.

[8] Z. Feng, S. Yang, M. Wang, and L. Jiao, "Learning dual geometric low-rank structure for semisupervised hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 346–358, Jan. 2021.

[9] J. Nalepa, M. Myller, Y. Imai, K.-I. Honda, T. Takeda, and M. Antoniak, "Unsupervised segmentation of hyperspectral images using 3-D convolutional autoencoders," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1948–1952, Nov. 2020.

[10] C. Zhao, B. Qin, S. Feng, W. Zhu, L. Zhang, and J. Ren, "An unsupervised domain adaptation method towards multi-level features and decision boundaries for cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5546216.

[11] T. Alipour-Fard, M. E. Paoletti, J. M. Haut, H. Arefi, J. Plaza, and A. Plaza, "Multibranch selective kernel networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1089–1093, Jun. 2021.

[12] W. Wang, W. Wang, and H. Liu, "Correlation-guided ensemble clustering for hyperspectral band selection," *Remote Sens.*, vol. 14, no. 5, p. 1156, Feb. 2022.

[13] Z. Dou, K. Gao, X. Zhang, H. Wang, and L. Han, "Band selection of hyperspectral images using attention-based autoencoders," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 147–151, Jan. 2021.

[14] H. Yang, M. Chen, G. Wu, J. Wang, Y. Wang, and Z. Hong, "Double deep Q-network for hyperspectral image band selection in land cover classification applications," *Remote Sens.*, vol. 15, no. 3, p. 682, Jan. 2023.

[15] C. He, Y. Zhang, and D. Gong, "A pseudo-label guided artificial bee colony algorithm for hyperspectral band selection," *Remote Sens.*, vol. 12, no. 20, p. 3456, Oct. 2020.

[16] M. Baisantry, A. K. Sao, and D. P. Shukla, "Discriminative spectral–spatial feature extraction-based band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518014.

[17] F. He, F. Nie, R. Wang, W. Jia, F. Zhang, and X. Li, "Semisupervised band selection with graph optimization for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10298–10311, Dec. 2021.

[18] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2956–2969, May 2015.

[19] M. Habermann, E. H. Shiguemori, and V. Frémont, "Unsupervised cluster-wise hyperspectral band selection for classification," *Remote Sens.*, vol. 14, no. 21, p. 5374, Oct. 2022.

[20] Y. Liu, X. Li, Z. Hua, and L. Zhao, "EBARec-BS: Effective band attention reconstruction network for hyperspectral imagery band selection," *Remote Sens.*, vol. 13, no. 18, p. 3602, Sep. 2021.

[21] Z. Zhang, D. Wang, X. Sun, L. Zhuang, R. Liu, and L. Ni, "Spatial sampling and grouping information entropy strategy based on kernel fuzzy C-means clustering method for hyperspectral band selection," *Remote Sens.*, vol. 14, no. 19, p. 5058, Oct. 2022.

[22] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.

[23] I. Assent, "Clustering high dimensional data," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 2, no. 4, pp. 340–350, Jun. 2012.

[24] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long, "A survey of clustering with deep learning: From the perspective of network architecture," *IEEE Access*, vol. 6, pp. 39501–39514, 2018.

[25] B. Yang, X. Fu, N. D. Sidiropoulos, and M. Hong, "Towards K-means-friendly spaces: Simultaneous deep learning and clustering," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3861–3870.

[26] J. Wang, H. Wang, Z. Ma, L. Wang, Q. Wang, and X. Li, "Unsupervised hyperspectral band selection based on hypergraph spectral clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[27] S. Jia, Y. Yuan, N. Li, J. Liao, Q. Huang, X. Jia, and M. Xu, "A multiscale superpixel-level group clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5523418.

[28] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, Dec. 2019.

[29] W. Sun, L. Zhang, B. Du, W. Li, and Y. Mark Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, Jun. 2015.

[30] J. Wang, C. Tang, Z. Li, X. Liu, W. Zhang, E. Zhu, and L. Wang, "Hyperspectral band selection via region-aware latent features fusion based clustering," *Inf. Fusion*, vol. 79, pp. 162–173, Mar. 2022.

[31] J. Wang, C. Tang, X. Liu, W. Zhang, W. Li, X. Zhu, L. Wang, and A. Y. Zomaya, "Region-aware hierarchical latent feature representation learning-guided clustering for hyperspectral band selection," *IEEE Trans. Cybern.*, vol. 53, no. 8, pp. 5250–5263, Aug. 2023.

[32] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 279–317, Dec. 2019.

[33] M. Zeng, Y. Cai, Z. Cai, X. Liu, P. Hu, and J. Ku, "Unsupervised hyperspectral image band selection based on deep subspace clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1889–1893, Dec. 2019.

[34] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.

[35] C.-I. Chang and K.-H. Liu, "Progressive band selection of spectral unmixing for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2002–2017, Apr. 2014.

[36] Y. Cai, X. Liu, and Z. Cai, "BS-Nets: An end-to-end framework for band selection of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1969–1984, Mar. 2020.

[37] H. Sun, J. Ren, H. Zhao, P. Yuen, and J. Tschannerl, "Novel Gumbel–Softmax trick enabled concrete autoencoder with entropy constraints for unsupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5506413.

[38] C. Wang, M. Gong, M. Zhang, and Y. Chan, "Unsupervised hyperspectral image band selection via column subset selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1411–1415, Jul. 2015.

[39] X. Geng, K. Sun, L. Ji, and Y. Zhao, "A fast volume-gradient-based band selection method for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7111–7119, Nov. 2014.

[40] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 227–239, Jan. 2016.

[41] K.-H. Liu, Y.-K. Chen, and T.-Y. Chen, "A band subset selection approach based on sparse self-representation and band grouping for hyperspectral image classification," *Remote Sens.*, vol. 14, no. 22, p. 5686, Nov. 2022.

[42] Y. Wan, C. Chen, A. Ma, L. Zhang, X. Gong, and Y. Zhong, "Adaptive multistrategy particle swarm optimization for hyperspectral remote sensing image band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5520115.

[43] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[44] Z. Xue, S. Yang, and L. Zhang, "Weighted sparse graph regularization for spectral–spatial classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1630–1634, Sep. 2021.

[45] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, Jan. 2016.

[46] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1431–1445, Mar. 2016.

WENHONG WANG received the B.S. degree in computer software from Shandong University, Jinan, China, in 1995, the M.S. degree in computer application technology from the Beijing University of Chemical Technology, Beijing, China, in 2003, and the Ph.D. degree in computer science and technology from Zhejiang University, China, in 2016.

Since 1998, he has been with the College of Computer Science, Liaocheng University, Liaocheng, China, where he is currently an Associate Professor of computer science and technology. His main research interests include hyperspectral image analysis, pattern recognition, and machine learning.

● ● ●

SHANGFENG NING received the B.S. degree from the Qingdao Institute of Technology, Qingdao, China, in 2020. He is currently pursuing the master's degree in software engineering with Liaocheng University, Liaocheng, China. His current research interests include machine learning and hyperspectral image processing.