Hyperspectral Band Selection with Dynamic Graph Enhancement and Hierarchical Feature Fusion

Jiahua Zhang, Jingyuan Wang, Dejie Li, Jun Zhang, Xiaodi Shang, *Member, IEEE*

***Abstract*—Graph Convolutional Networks (GCNs) have been widely used in hyperspectral band selection (BS) and exhibits great potential. However, the substantial noise inherent in hyperspectral imagery and the fixed graph structure hinder the ability of GCNs to effectively capture the complex inter-band relationships. Moreover, most existing methods rely on a single criterion during the BS phase, often leading to biased results. To address these issues, this paper proposes a dynamic graph enhancement and hierarchical feature fusion (DEHF) model for efficient and robust BS. Specifically, the model first utilizes an autoencoder (AE) to extract latent features of key samples, thereby enhancing the contribution of high-quality samples in the BS process. To overcome the limitations of conventional graph construction, we introduce a two-stage graph optimization strategy comprising initial graph optimization and dynamic enhanced GCN (DEGCN). This strategy adaptively updates both the graph structure and node representations, enabling more accurate modeling of the complex local-global relationships among spectral bands. In order to further enhance feature integration of the model, a hierarchical fusion mechanism is designed to fuse multi-source features extracted by the AE and DEGCN across different spatial scales, thereby emphasizing the responses of salient bands. Finally, to comprehensively evaluate band importance, a multi-dimensional priority criterion is developed by integrating the information entropy and class discriminability, which guides spectral clustering to yield a more informative and representative band subset. Experimental results on three real hyperspectral datasets demonstrate the superior performance of DEHF compared to state-of-the-art methods.**

***Index Terms*—Hyperspectral band selection, essential feature extraction, dynamic graph enhancement, multi-dimensional priority.**

I. Introduction

HYPERSPECTRAL image (HSI), consisting of hundreds of narrow and contiguous spectral bands, provides rich spectral-spatial information [1], and has shown great potential in various fields such as target recognition [2], mineral exploration [3], environmental monitoring [4], [5], and precision agriculture [6]. However, its high dimensionality often introduces considerable redundancy and noisy bands, leading to heavy computational and storage burdens that limit practical applications [7]. Moreover, the scarcity of labeled samples combined with high dimensionality causes the "Hughes phenomenon" [8], thereby degrading classification performance. Consequently, dimensionality reduction (DR) is crucial, as it not only mitigates redundancy and reduces computational overhead, but also improves the efficiency and accuracy of downstream tasks.

Feature extraction (FE) and band selection (BS) are two primary techniques for hyperspectral DR [9]. FE projects the high-dimensional data into a lower-dimensional feature space through an optimal transformation, effectively compressing the data while retaining the most discriminative features for subsequent tasks [10], [11]. Typical methods include principal component analysis (PCA) [12], local linear embedding (LLE) [13], and isometric mapping (Isomap) [14]. However, these methods often involve complex matrix operations and produce transformed features that lack physical interpretability, limiting their practical applicability [15]. In contrast, BS directly selects a subset of representative bands from the original HSI, preserving the physical meaning of the data [16] and offering greater practicality for real-world applications.

In recent years, machine learning has brought new advances to BS. Approaches such as maximum variance principal component analysis (MVPCA) [17] and similarity-based ranking structural similarity (SR-SSIM) [18] evaluate the band importance through priority criteria. Additionally, clustering-based BS methods including enhanced fast density-peak-based clustering (E-FDPC) [19] and hyperbolic clustering-based band hierarchy (HCBH) [20], group similar bands and select the most representative ones to reduce redundancy. Sparse representation-based BS methods, like tensorial global-local graph self-representation (TGSR) [21] and hypergraph regularized self-representation (HyGSR) [22], assume that the original data can be linearly represented by a small number of representative bands. Furthermore, the optimization-based strategies [23], [24], [25], treat BS as an optimization process, combining the predefined search strategies to find the optimal band combination. Despite their effectiveness, most of these methods rely on manually defined, rigid assumptions to model the internal interactions between bands [26], making it difficult to address the complexities of real-world scenarios.

Additionally, a series of deep learning-based BS approaches have emerged. Attention-based neural networks, such as BS network (BSNet) [27], nonlocal band attention network (NBAN) [28], dual-attention reconstruction network (DARecNet-BS) [29], global-local spectral weight network based on attention (GLSWA) [30], and spectral-spatial cross-dimensional attention network (SSANet-BS) [26], require nei-

This work was supported in part by the National Natural Science Foundation of China under Grant 42301380, in part by China Postdoctoral Science Foundation Grant 2023M731843, in part by the Science and Technology Support Plan for Youth Innovation of Colleges and Universities of Shandong Province of China under Grant 2023KJ232, and in part by the Postdoctoral Applied Research Foundation of Qingdao under Grant QDBSH20230101012, in part by the Qingdao Natural Science Foundation Grant 23-2-1-64-zyydjch. *(Jiahua Zhang and Jingyuan Wang contributed equally to this work. Corresponding author: Xiaodi Shang.)*

Jingyuan Wang, Dejie Li, Jun Zhang and Xiaodi Shang are with the College of Computer Science and Technology, Qingdao University, Qingdao 266071, China (e-mail: wannng722@gmail.com; jeremiahcn985@gmail.com; jun.zhang.edu@gmail.com; shangxd@qdu.edu.cn).

Jiahua Zhang is with Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: zhangjh@radi.ac.cn).

ther manual feature engineering nor strong prior assumptions, and can automatically discovering and learning the latent interdependencies within HSI, thereby enabling the extraction of highly salient bands. Although these methods successfully capture the complex interdependencies among spectral bands, they often fall short in modeling the local structural relationships between bands, which can lead to redundancy in the selected subsets [31].

Graph convolutional networks (GCNs) effectively handle the non-Euclidean structure data and overcome the fixed receptive field limitation of convolutional neural networks (CNNs), has been widely used in BS. For example, methods such as the dual-GCN based on band attention and sparsity constraint (BSD-GCN) [32] and the diversity learning guided dual graph autoencoder (DLG-DGAE) [33], can partially capture inter-band correlations. However, the presence of substantial background and noise information can interfere with the accurate construction of the graph, and these methods typically rely on manually designed, fixed graph structures, which limits their ability to model complex inter-band relationships. In summary, a critical challenge remains in effectively mitigating the impact of noise on graph construction and dynamically learning the band adjacency structure to more precisely model both local and global spectral correlations.

Motivated by these issues, we propose a dynamic graph enhancement and hierarchical feature fusion framework for hyperspectral BS, named DEHF. To suppress noise interference on GCN learning, DEHF first employs an autoencoder (AE) for data cleansing, guiding BS toward essential object features and enhancing the discriminability of the selected bands. Next, to overcome the limitations of fixed graph structure, we designed a spectral-spatial joint similarity-based initialization strategy, followed by a dynamic enhanced GCN that collaboratively refines the band local structures and node representations for more accurate graph construction. On this basis, an attention mechanism is incorporated for adaptive weighting and hierarchical fusion of sample-band features, facilitating the exploration of complex spectral-spatial interactions in HSI and reinforcing the feature representation of salient bands. Finally, a multi-dimensional band priority criterion is introduced to steer spectral clustering to select the informative and representative bands. Experimental results on three real hyperspectral datasets confirm the superiority of DEHF over state-of-the-art methods. In summary, the main contributions of this paper are as follows:

1) We propose a novel DEHF framework for BS, which hierarchically fuses the key sample features with the band adjacency structures. This approach effectively models the complex spectral-spatial interactions in HSI while suppressing the influence of nterfering factors, thus enhancing band discriminability and reducing redundancy.

2) We construct a GCN combining fine-grained band graph initialization with dynamic graph enhancement. By refining the graph construction mechanism and adopting a dynamic learning strategy, the model adaptively optimizes the topological representation of band adjacency structure.

3) We design a multi-dimensional band priority criterion that comprehensively evaluates band importance based on both information content and class separability. Experiments show that, compared to single-metric approaches, it selects band subset with higher information and stronger discriminability.

The rest of this paper is structured as follows. Section II reviews the previous studies for BS based on AE and GCN. Section III introduces the DEHF framework, including essential sample feature extraction, dynamic graph enhancement, hierarchical feature fusion, and multi-dimensional priority-based BS. Section IV compares the performance of DEHF with other state-of-the-art methods. Finally, Section V concludes the paper.

II. Related Work

In this section, we reviewed previous studies related to the proposed DEHF, including AE-based and GCN-based BS methods. Furthermore, the motivations for this work are presented to emphasize the novelty and necessity of the proposed approach.

A. AE-based BS Methods

Unlike traditional methods, which typically consider only linear relationships or rely on predefined kernels to model nonlinearities, AE-based BS learns the latent representations of hyperspectral data through nonlinear mapping, enabling unsupervised dimensionality reduction and data reconstruction while effectively reducing redundancy and preserving spectral information.

Most AE-based BS studies aim to leverage the AE's ability to learn latent HSI representations and identify key bands. For instance, Sun et al. [34] proposed the concrete autoencoder unsupervised BS (CAE-UBS) framework, which employs the Gumbel-Softmax trick and entropy constraints for differentiable discrete BS. To overcome the limited parameterization of discrete variables, Xu et al. [35] further developed the dropout CAE, combining concrete relaxation with a dropout feature ranking strategy to select informative bands directly. For noise suppression, Liu et al. [36] employed a convolutional autoencoder to evaluate band representativeness while balancing redundancy and information content to optimize the band subset. The SAE-1DCNN [37] integrated a stacked autoencoder with CNN to identify and remove noisy bands.

To enhance sparsity and interpretability, Dou et al. [38] assumed that HSI can be sparsely reconstructed using a small number of informative bands and proposed the attention-AE, which explicitly models nonlinear inter-band dependencies through an attention mechanism. Ahishali et al. [39] combined self-representation learning with a sparse 1D-operational autoencoder (SOA), using 1D operational layers with nonlinear neurons to enhance sparse feature representation. The stochastic gate-based autoencoder (SGAE) [40] introduced a learnable stochastic gating layer to generate band subsets. Additionally, some studies incorporated intelligent optimization or physical modeling to improve global optimality and plausibility. For example, Singh et al. [41] combined AE with genetic algorithms, using reconstruction-error-driven evolutionary search to achieve globally optimal BS. Sangeetha et al. [42] embedded spectral physics into the AE via a reflectance recovery

mechanism and employed a grouping strategy to obtain effective low-dimensional representations. However, while AE-based BS methods effectively handle nonlinearity and noise, they often overlook spatial correlations among bands, limiting their overall performance.

B. GCN-based BS Methods

To address the limitations of traditional methods in modeling complex inter-band dependencies, graph convolutional networks (GCNs) and their variants have been gradually introduced into the field of hyperspectral BS. By constructing graph structures, these methods effectively capture the high-order dependencies between spectral and spatial features, providing a new paradigm for BS.

Specifically, Feng et al. [32] first proposed a dual-GCN based on band attention and sparse constraint (BSD-GCN), effectively capturing both local and long-range correlations among bands and jointly modeling spectral-spatial information. Yu et al. [43] introduced a hybrid GCN (GCN-CNN-BS), which combines CNN and GCN features within an unsupervised framework and uses an auxiliary spectral reconstruction task to guide BS. Wang et al. [44] developed a dual-branch Mamba-based BS enhancement model, which leverages Mamba for long-range dependency modeling and GCN for local spatial feature extraction, achieving superior performance in spectral-spatial feature learning. Zhao et al. [45] further applied GNN to thermal infrared hyperspectral BS for land surface temperature retrieval, extracting cross-band correlations via graph attention and multi-scale convolutional kernels while incorporating brightness temperature constraints to guide feature learning. To further explore deep spectral-spatial interactions, Chen et al. [46] proposed an enhanced self-representation heterogeneous GCN (ESR-HGCN), which constructs pixel-level and band-level dual graphs and employs heterogeneous graph convolutions to capture multi-level feature relationships, enabling robust BS.

C. Motivations

Although AE-based and GCN-based BS methods have achieved remarkable progress, several critical challenges remain. AE-based methods are effective in capturing nonlinear relationships and suppressing noisebut often overlook spatial correlations among bands, limiting their ability to fully exploit the spectral–spatial information of HSI. In contrast, GCN-based methods can model complex inter-band dependencies through graph structures, but most rely on manually designed static graphs that fail to adaptively represent local and global spectral relationships. Moreover, they are easily affected by noise and background interference in raw HSI data, resulting in suboptimal band representations. In addition, relying on a single evaluation criterion in subsequent band clustering further restricts the quality of selected subsets.

To overcome these issues, this study proposes a robust and adaptive BS framework that integrates AE and GCN to jointly mitigate data interference, enhance dynamic graph modeling, and improve band evaluation. Specifically, The AE extracts key spectral features to suppress noise and purify the data, while a dynamically enhanced GCN adaptively captures spectral–spatial dependencies through optimized graph structures. Meanwhile, a multi-dimensional priority criterion combining information entropy and band separability is further employed to select the most informative and representative bands.

Overall, this work innovatively integrates AE and GCN within a unified framework. The AE-extracted features purify the data and enhance graph learning, while the dynamic graph optimization mechanism adaptively updates graph structures and node representations to accurately model inter-band relationships. By combining hierarchical feature fusion with a multi-criteria prioritization strategy, the proposed DEHF efficiently selects representative and discriminative band subsets, significantly improving classification performance while preserving spectral information.

III. METHOD

This paper proposes a novel hyperspectral BS framework, named dynamic graph enhancement and hierarchical feature fusion (DEHF), as shown in Fig. 1. First, the AE is utilized to capture the nonlinear relationships among pixels, extract essential sample features, and suppress the influence of background and noise on the BS process. Subsequently, a dynamically enhanced GCN (DEGCN) is constructed to adaptively model neighborhood dependencies and learn complex inter-band correlations, thereby refining the band feature representation. Then, a hierarchical feature fusion mechanism is introduced to integrate sample-band features, so as to emphasize informative bands. Finally, spectral clustering is applied to partition the spectral bands, and a multi-dimensional priority (MP) criterion based on information entropy (IE) and intra-/inter-cluster similarity (I^2cS) is designed to select the most representative bands.

A. Essential Sample Feature Extraction

HSI usually contains a large number of background and noise pixels. Performing BS directly on the entire image may result in the selection of bands biased toward such interference, thus diminishing their ability to characterize the target objects. The AE, through nonlinear dimensionality reduction, can effectively remove the redundant information from the high-dimensional space and compact low-dimensional latent features, which can significantly enhance the data quality. The extracted essential sample features can deeply capture the spectral characteristics of target objects, suppress the noise and background interference, and provide more accurate sample-level guidance for BS.

In hyperspectral imaging, the original data are typically represented as a 3-D cube $\mathbf{X} \in \mathbb{R}^{H\times W\times d}$, where H and W represent the height and width of the image, and d denotes the number of spectral bands. To facilitate subsequent modeling, the 3-D HSI cube is reshaped into a 2-D matrix. Specifically, the spectral vector of each pixel is unfolded into a column, and all pixels are arranged sequentially to form $\mathbf{X} \in \mathbb{R}^{d\times N}$, where $N = H \times W$ is the total number of pixels. $\mathbf{x}_i \in \mathbb{R}^{1\times N}$ represents the i-th band vector.

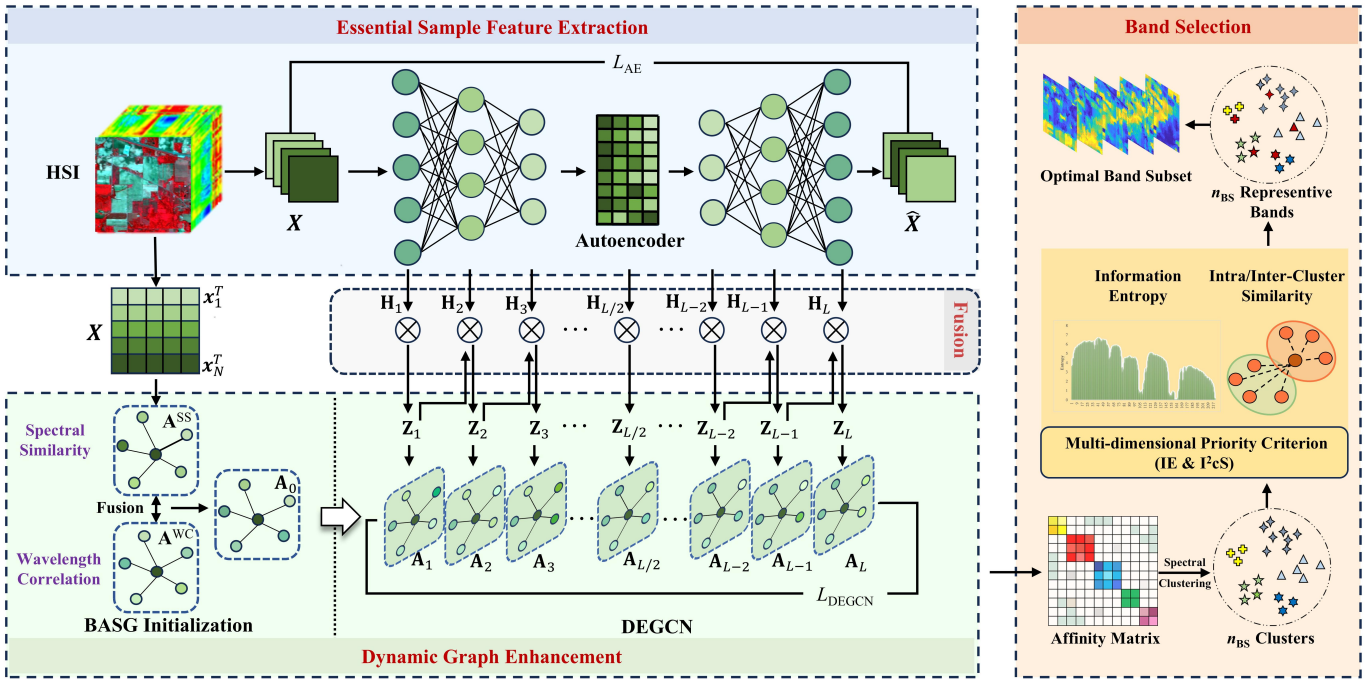


Fig. 1. Framework of DEHF.

The AE comprises two components: an encoder and a decoder. The encoder projects the high-dimensional data into a low-dimensional latent space through multi-layer nonlinear transformations, which effectively removes the noise and background interference, thereby enhancing the representation capability of the latent features for target objects. The decoder reconstructs the data via inverse nonlinear mapping to recover the high-fidelity spatial structure while further suppressing the residual noise. Both the encoder and decoder are constructed from multiple stacked modules, each consisting of a linear layer, batch normalization, and an activation function. Let the input of the l-th layer be $\mathbf{H}_{l-1}$, then he output $\mathbf{H}_l$ can be expressed as:

$$\mathbf{H}_l = \phi\left(\mathbf{W}_l \mathbf{H}_{l-1} + \mathbf{b}_l\right) \tag{1}$$

Here, $\phi(\cdot)$ denotes the ReLU activation function. $\mathbf{W}_l$ represents the weight matrix of the l-th layer, and $\mathbf{b}_l$ denotes the corresponding bias vector. Specifically, the input to the first layer is the original HSI, $\mathbf{H}_0 = \mathbf{X}$. The model first performs hierarchical feature abstraction to filter out irrelevant components that are not associated with the target objects, ensuring that the output features of each layer encode only the key information of the target classes. Then, the feature maps are progressively refined to yield the final output, $\mathbf{H}_L = \widehat{\mathbf{X}}$, which not only preserves the spatial structure but also accurately characterizes the spectral features of the target objects.

In AE, the residual between the original data $\mathbf{X}$ and the reconstructed data $\widehat{\mathbf{X}}$ is utilized as the loss function:

$$L_{AE} = \frac{1}{2N} ||\mathbf{X} - \widehat{\mathbf{X}}||_F^2 \tag{2}$$

The loss function L_{AE} directs AE to concentrate on the key features of the target objects, gradually suppressing background and noise interference during the feature extraction. This ensures that the feature representations at each layer more accurately capture the intrinsic characteristics of the objects, thereby enhancing the purity of feature expression and providing more discriminative and representative information for subsequent BS.

B. Dynamic Graph Enhancement

Although AE are effective in extracting discriminative spectral features from HSI, it often neglects the adjacency relationships between bands, leading to redundant band subsets. To address this issue, this paper introduces a dynamic graph enhancement module, comprising the band adjacency structure graph (BASG) initialization and the dynamic enhanced GCN (DEGCN). This module aims to construct a more accurate BASG during initialization and to adaptively refine its topology throughout learning via DEGCN. By jointly optimizing local band structures and node feature representations, DEGCN can obtain a more comprehensive characterization of band adjacency relationships.

1) BASG Initialization: Studies have demonstrated that the quality of the initial graph directly affects the performance of GCN in feature learning [47]. Unlike conventional approaches that rely on a single metric to model band relationships, this paper designs a spectral-spatial joint similarity metric that simultaneously considers spectral similarity and wavelength correlation. It facilitates the precise construction of the initial BASG, and can provide a more reliable guidance for the subsequent adaptive learning of band dependencies via GCN.

Specifically, the initial BASG is denoted as $G = (V, E)$, where V and E represent the vertex and edge sets, respectively. Each band vector $\mathbf{x}_i$ is treated as a node in G. Then, the spectral similarity (SS) between bands is defined as:

$$\mathbf{A}_{ij}^{SS} = \begin{cases} \frac{d_{i,k+1} - d_{ij}}{\sum_{m=1}^{k}(d_{i,k+1} - d_{i,m})} & , \mathbf{x}_j \in N_k(\mathbf{x}_i) \\ 0 & , \text{otherwise} \end{cases} \tag{3}$$

Here, N_k is the set of k nearest neighbors of $\mathbf{x}_i$, and d_{ij} represents the Euclidean distance between bands $\mathbf{x}_i$ and $\mathbf{x}_j$. $d_{i,k+1}$ denotes the $(k+1)$-th smallest distance between $\mathbf{x}_i$ and all other bands. Evidently, a smaller d_{ij} indicates a higher spectral similarity $\mathbf{A}_{ij}^{SS}$, suggesting that bands in closer spectral distance exhibit stronger correlation and more consistent representation capacity for the target objects.

Additionally, spectral bands are typically arranged in ascending order of wavelength [48], and bands at different wavelengths interact with ground objects through distinct physical mechanisms. For instance, the red band (630–690 nm) exhibits an absorption valley in the spectral curve due to vegetation's selective absorption, whereas the green band (520–570 nm) forms a reflectance peak as a result of lower absorption efficiency. This indicates that adjacent bands within a continuous spectral range often exhibit strong correlation and locally consistent representation capacity. Therefore, when evaluating band adjacency, it is crucial to incorporate wavelength characteristics to avoid mistakenly linking spectrally similar bands that differ significantly in their ability to represent ground objects, thereby preserving the rationality of the band structure. Accordingly, the wavelength correlation (WC) between $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as:

$$\mathbf{A}_{ij}^{WC} = \begin{cases} \exp\left(-\frac{|I_{\mathbf{x}_i} - I_{\mathbf{x}_j}|}{2\omega^2}\right) & , \quad \mathbf{x}_j \in N_k(\mathbf{x}_i) \\ 0 & , \quad \text{otherwise} \end{cases} \tag{4}$$

Here, $\omega = 1/k \sum_{\mathbf{x}_j \in N_k(\mathbf{x}_i)} \left| I_{\mathbf{x}_i} - I_{\mathbf{x}_j} \right|$, where $I_{\mathbf{x}_i}$ denotes the positional index of band $\mathbf{x}_i$. The operator $|\cdot|$ indicates the absolute value. The wavelength correlation $\mathbf{A}_{ij}^{WC}$ assigns a higher weight to spectrally adjacent bands in physical order to enhance the compactness of local band clusters. Specifically, the closer the positional indices of $\mathbf{x}_i$ and $\mathbf{x}_j$, the more similar their physical properties, resulting in a higher $\mathbf{A}_{ij}^{WC}$. Finally, by integrating the spectral similarity $\mathbf{A}_{ij}^{SS}$ defined in Eq. (3), the spectral-spatial joint similarity is formulated as:

$$\mathbf{A}_{ij} = \mathbf{A}_{ij}^{SS} \mathbf{A}_{ij}^{WC} \tag{5}$$

It can effectively capture both global and local structural characteristics of spectral bands by integrating their statistical and physical properties. This enables more comprehensive modeling of band adjacency relationships, with the constructed BASG providing more accurate guidance for subsequent GCN-based learning.

2) DEGCN: As HSI are inherently non-graph structured, the initially constructed band graph is often suboptimal, which may lead to inaccurate clustering results if used directly. To do so, we propose a dynamic enhanced GCN, named DEGCN, which iteratively optimizes both the node feature representation and the associated adjacency matrix of BASG to learn a more reliable graph structure. Each layer of DEGCN takes the adjacency matrix $\mathbf{A}$ and node feature matrix $\mathbf{Z}$ as input and performs convolutional operations to aggregate information from neighboring nodes, thereby generating more discriminative hidden representations.

Specifically, since the initial adjacency matrix $\mathbf{A}$ is asymmetric, it is first symmetrized before being fed into the DEGCN, i.e., $\mathbf{A}_0 = 1/2(\mathbf{A} + \mathbf{A}^T)$. Assuming that the DEGCN consists of L layers, with the input to the l-th layer denoted as $\mathbf{Z}_{l-1}$, the output of this layer can be formulated as:

$$\mathbf{Z}_l = \phi\left(\mathbf{D}_{l-1}^{-\frac{1}{2}}\left(\mathbf{A}_{l-1} + \mathbf{I}\right)\mathbf{D}_{l-1}^{-\frac{1}{2}}\mathbf{Z}_{l-1}\mathbf{U}_l\right) \tag{6}$$

where $\mathbf{A}_{l-1}$ denotes the adjacency matrix at the $(l-1)$-th layer, derived from $\mathbf{Z}_{l-1}$. $\mathbf{D}_{l-1}$ is the degree matrix of $\mathbf{A}_{l-1} + \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. $\mathbf{U}_l$ represents the trainable parameters of the l-th layer.

To enable DEGCN to adaptively adjust the band adjacency structure during each iteration and enhance its generalization capability across diverse datasets, we design a dynamic update mechanism for the adjacency matrix of BASG, formulated as: $\widehat{\mathbf{A}}_{ij} = \widehat{\mathbf{A}}_{ij}^{SS} \widehat{\mathbf{A}}_{ij}^{WC}$. Specifically, $\widehat{\mathbf{A}}_{ij}^{SS} = e^{-d_{ij}} / \sum_{j=1}^{d} e^{-d_{ij}}$ quantifies the spectral similarity between bands $\mathbf{x}_i$ and $\mathbf{x}_j$, where $d_{ij} = \|\mathbf{z}_i - \mathbf{z}_j\|_2^2$ denotes the Euclidean distance between the feature vectors $\mathbf{z}_i$ and $\mathbf{z}_j$. Compared with Eq. (3), $\widehat{\mathbf{A}}_{ij}^{SS}$ can effectively capture the long-range dependencies among distant yet correlated bands, thereby improving the smoothness of the adjacency matrix and enhancing the information propagation efficiency of graph convolution. Meanwhile, $\widehat{\mathbf{A}}_{ij}^{WC}$ assesses the proximity between bands $\mathbf{x}_i$ and $\mathbf{x}_j$ in terms of their physical wavelengths, ensuring the learned band adjacency structure to more effectively preserve the local distribution characteristics. Additionally, to avoid the model instability caused by abrupt changes in the graph structure during training, we apply the following linear fusion strategy of the new and old graph to ensure the stability of DEGCN training.

$$\mathbf{A}_l = \mu \widehat{\mathbf{A}}_l + (1 - \mu)\mathbf{A}_{l-1} \tag{7}$$

Here, μ denotes the smoothing factor. By incorporating historical information, the new graph inherits part of the structure from the previous graph, thereby improving the model stability and promoting convergence. After L iterations, a more robust and reliable graph structure $\mathbf{A}^{Final}$ is obtained.

To preserve the fundamental topological properties of the learned graph structure, an adaptive graph reconstruction loss is formulated as:

$$L_{DEGCN} = H(\mathbf{A}, \mathbf{A}^{Final}) = -\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{A}_{ij} log \mathbf{A}_{ij}^{Final} \tag{8}$$

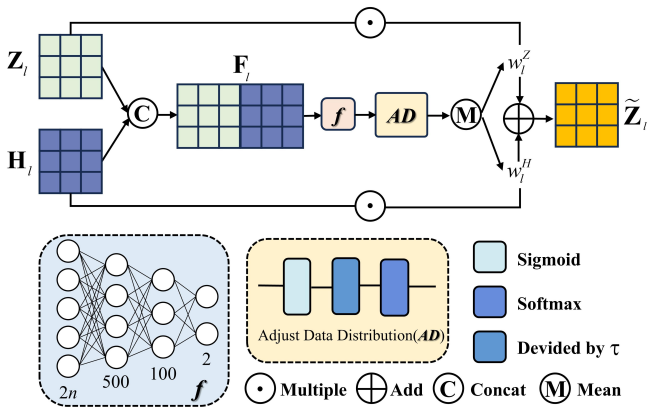


Fig. 2. Illustration of feature hierarchical fusion module.

Accordingly, by incorporating the autoencoder loss L_{AE}, the overall loss function is defined as follows:

$$L_{DEHF} = L_{DEGCN} + \lambda L_{AE} \tag{9}$$

In Eq. (9), λ serves as a balancing coefficient that regulates the contribution of each loss term, ensuring that the model can simultaneously maintain the accuracy of key sample features extracted by the AE and the rationality of the graph structure learned by DEGCN. The proposed DEGCN overcomes the static limitations of traditional GCN by introducing a dynamic graph structure optimization mechanism. This allows the model to progressively optimize the connection weights between bands, thereby learning more accurate spatial adjacency relationships and facilitating BS.

C. Hierarchical Feature Fusion

To fully exploit the complex spectral-spatial interactions in HSI, an attention mechanism is introduced to adaptively balance the contributions of the key sample feature $\mathbf{H}_l$ extracted by AE and the band spatial feature $\mathbf{Z}_l$ learned by DEGCN. This facilitates the collaborative utilization of spectral and spatial information. Ultimately, this mechanism guides DEGCN to progressively enhance the representation of salient bands, thereby improving the structural plausibility of the learned adjacency graph. In detail, a new feature matrix $\mathbf{F}_l = concat(\mathbf{H}_l, \mathbf{Z}_l)$ is first constructed by concatenating the spectral and spatial features. Then, the weight coefficients w_l^H and w_l^Z corresponding to $\mathbf{H}_l$ and $\mathbf{Z}_l$ are learned. To this end, a three-layer fully connected network is employed to perform a nonlinear transformation on $\mathbf{F}_l$, generating an initial attention score $\mathbf{G}_l$ that fuses the spectral-spatial information.

$$\mathbf{G}_l = f(\mathbf{F}_l) \tag{10}$$

Here, $f(\cdot)$ denotes the three-layer fully connected network, with its detailed architecture presented in Table I.

Then, the softmax function is applied to obtain the attention weight $\mathbf{E}_l$ to quantify the significance of the two types of information in the collaborative utilization of spectral-spatial features:

$$\mathbf{E}_l = softmax\left(sigmoid\left(\mathbf{G}_l\right)/\tau\right) \tag{11}$$

TABLE I
DETAILED ARCHITECTURE

Module	Layer
f	FC1: in = $2n$, out = 500
	ReLU
	FC2: in = 500, out = 100
	ReLU
	FC3: in = 100, out = 2

It is important to note that directly applying the softmax function to the attention score $\mathbf{G}_l$ may result in certain weights converging towards 0 or 1, which could lead the model to disproportionately emphasize or neglect specific information. To address this issue, we first adjust the data distribution using the sigmoid function prior to the softmax operation. Additionally, a predefined calibration factor τ is introduced to compress the value range, thereby reducing the impact of extreme weights and ensuring the stability of the weight allocation.

Furthermore, the global weights w_l^H and w_l^Z are derived by averaging $\mathbf{E}_l$ along the sample dimension, and subsequently concatenated to form the attention weight vector w_l, as defined below:

$$w_l = Mean\left(\mathbf{E}_l, dim = 0\right) = [w_l^H, w_l^Z] \tag{12}$$

Here, w_l^H and w_l^Z respectively measure the contributions of $\mathbf{H}_l$ and $\mathbf{Z}_l$ during the fusion process. Accordingly, the fused feature representation $\tilde{\mathbf{Z}}_l$ is given by:

$$\tilde{\mathbf{Z}}_l = w_l^H \mathbf{H}_l + w_l^Z \mathbf{Z}_l \tag{13}$$

The entire process is illustrated in Fig. 2. It effectively integrates the spectral and spatial information, adaptively capturing the key elements within heterogeneous data and the complex dependencies between bands, thus providing a more precise feature representation for the subsequent update of the BASG.

D. Multi-dimensional Priority-Based Band Selection

In theory, due to the high local similarity among spectral bands, the adjacency matrix learned by DEGCN typically exhibits a block-diagonal structure [49], indicating that the band correlations are stronger within local regions. This property provides a solid theoretical foundation for band clustering. Accordingly, spectral clustering is adopted in this work to group spectrally similar bands. Specifically, a Laplacian matrix is constructed based on the DEGCN-derived adjacency matrix, followed by eigen-decomposition to obtain low-dimensional embeddings. In this embedded space, feature vectors of similar bands tend to aggregate. Finally, the most representative band from each cluster is selected to form the final subset, effectively reducing band redundancy.

However, most clustering-based BS methods rely on a single criterion to identify the representative bands, which typically reflects only one aspect of the band characteristics and fails to provide a comprehensive evaluation. To address this, we propose a multi-dimensional priority (MP) strategy

that integrates the information entropy (IE) and the intra/inter-cluster similarity (I^2cS) to select bands with high information content and strong discriminative capability.

On one hand, IE quantifies the information richness of spectral bands, and it is defined as follows:

$$IE_i = -\sum_{z \in GRAY} p(z) \log p(z) \tag{14}$$

Here, IE_i is the information entropy of band $\mathbf{x}_i$. $GRAY$ represents the grayscale space, encompassing all possible grayscale intensity levels within the band. $p(z)$ denotes the probability of grayscale value z in the i-th band. A higher IE_i suggests that the band $\mathbf{x}_i$ contains richer and more informative information, facilitating a more accurate and discriminative characterization of land cover types.

On the other hand, a desired band should exhibit higher similarity to intra-cluster bands and greater dissimilarity to inter-cluster bands. To this end, the Wasserstein distance is introduced to quantify the intra/inter-cluster similarity of $\mathbf{x}_i$.

$$I^2cS_i = \inf_{\pi \in \Gamma(X_i, Y_i)} \mathbb{E}_{(x_i, y_i) \sim \pi} \left[\|x_i - y_i\| \right] \tag{15}$$

Here, X_i and Y_i are the intra-cluster and inter-cluster similarity distributions, respectively. $\Gamma(X_i, Y_i)$ refers to the set of all joint distributions $\pi(x_i, y_i)$, whose marginals are X_i and Y_i. $\mathbb{E}$ denotes the statistical expectation, and $\inf$ signifies the infimum. Eq. (15) can be interpreted as finding an optimal distribution π in $\Gamma(X_i, Y_i)$ that minimizes the expectation distances between x_i and y_i. Thus, if the similarity between band $\mathbf{x}_i$ and other intra-cluster bands is greater while the disparity with inter-cluster bands is larger, its I^2cS_i value will be larger, indicating that this band is more representative.

Finally, the multi-dimensional priority criterion is defined as:

$$MP_i = \delta\left(IE_i\right) + \delta\left(I^2cS_i\right) \tag{16}$$

where $\delta(\cdot)$ denotes the normalization operation. A higher MP_i indicates that band $\mathbf{x}_i$ possesses greater advantages in terms of information richness and land-cover discriminability. Accordingly, after spectral clustering, the band with the highest MP value is sequentially selected from each cluster to form the representative band subset.

IV. Experiments

To validate the effectiveness of the proposed DEHF, extensive experiments were conducted on three public datasets, systematically comparing it with nine state-of-the-art BS methods in terms of classification accuracy, ablation studies, and computational complexity.

A. Datasets

Three publicly available datasets—Indian Pines, Pavia University, and Washington DC Mall—are selected for the experiments. Fig. 3 illustrates their false-color images, ground truth maps, and class labels, respectively.

1) **Indian Pines (IP220)**: The IP220 dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in 1992 in Indiana, USA. It consists of 220 spectral bands covering wavelengths from 0.4 to 2.5 μm, with a spatial resolution of 145×145 pixels. The dataset includes 16 classes and a total of 10249 labeled pixels.
2) **Pavia University (PU103)**: The PU103 dataset was obtained by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor in 2003 over the University of Pavia, Italy. It consists of 610×340 pixels and 115 spectral bands, with 12 noisy bands removed, leaving 103 bands for analysis. The dataset includes 42776 labeled samples across 9 classes.
3) **Washington DC Mall (DC191)**: The DC191 dataset is an airborne HSI acquired by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor, covering a region of Washington, D.C., USA. It has a size of 280×307×191 pixels and includes 6 classes, such as roof, grass, and road. It is primarily used for urban analysis and land cover classification.

B. Experimental Setup

Nine state-of-the-art methods are compared in our experiments, including maximum-variance principal component analysis (MVPCA) [17], enhanced fast density-peak clustering (E-FDPC) [19], adaptive subspace partitioning strategy (ASPS) [48], optimal clustering framework (OCF) [50], spectral structure preserved self-representation (S^4P) [51], tensorial global-local graph self-representation (TGSR) [21], spectral-spatial cross-dimensional attention network (SSANet-BS) [26], dropout concrete autoencoder (CAE) [35] and capsule network-based BS (CapUBS) [52] to validate the effectiveness of the proposed method.

Support vector machine (SVM), K-nearest neighbors (KNN) and convolutional neural network (CNN) are employed to classify the band subsets selected by each BS method, with the overall accuracy (OA), average accuracy (AA) and Kappa coefficient serving as performance metrics. For the IP220, PU103 and DC191 datasets, 10%, 1%, and 5% of the samples are randomly selected as the training set, and the remaining samples are used for testing. To reduce the uncertainty introduced by random sample selection, the OA for each band subset is averaged over five independent trials.

C. Parameter Analysis

DEHF comprises three hyperparameters: the number of neighbors k in the BASG initialization, the fusion coefficient μ for integrating new and old graph, and the weight coefficient λ for balancing the losses L_{AE} and L_{DEGCN}. In order to determine the optimal combination, a grid search is conducted over the ranges $\{5, 10, 15, 20\}$, $\{0.2, 0.4, 0.6, 0.8\}$, and $\{1, 10, 100, 1000\}$ for k, μ, and λ, respectively. As a result, the optimal parameter sets $\{k, \mu, \lambda\}$ for the IP220, PU103, and DC191 datasets are determined as $\{10, 0.8, 1000\}$, $\{15, 0.2, 1000\}$, and $\{10, 0.8, 1000\}$, respectively.

Moreover, a univariate analysis is performed to assess the impact of individual parameters on model performance. In each experiment, only one parameter is varied while the others are fixed at their optimal values to evaluate its effect

TABLE II
CLASSIFICATION RESULTS OF VARIOUS BS METHODS USING SVM ON THREE DATASETS

Method	IP220			PU103			DC191		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
MVPCA	61.67±7.38	60.30±13.79	55.95±9.23	67.92±11.28	75.48±10.42	60.23±13.11	89.97±1.93	91.97±1.47	87.67±2.35
E-FDPC	51.62±5.04	46.81±5.36	44.52±5.82	70.71±1.62	82.15±1.16	63.59±1.88	88.25±1.52	91.01±0.91	85.62±1.83
ASPS	**74.55±5.01**	**78.83±6.44**	**70.95±5.69**	73.99±4.21	82.95±3.27	66.64±4.94	92.35±1.99	94.26±1.56	90.60±2.42
OCF	64.45±7.23	65.64±10.55	59.27±8.36	74.28±3.94	83.73±2.30	67.75±4.64	92.22±2.25	94.13±1.77	90.44±2.74
S^4P	68.65±6.01	70.77±9.77	64.11±7.07	**77.70±3.99**	**85.21±1.89**	**71.81±4.64**	92.43±1.83	94.22±1.44	90.70±2.24
TGSR	73.70±4.14	77.75±4.05	69.93±4.78	75.30±3.68	83.53±3.31	68.86±4.40	92.80±1.35	94.55±1.13	91.14±1.65
SSANet-BS	66.34±16.81	64.47±19.35	61.58±19.09	73.75±2.12	82.32±2.84	67.03±2.54	90.49±2.43	92.90±1.92	88.35±2.94
CAE	73.49±7.05	**78.66±8.12**	69.77±7.97	74.21±5.49	84.00±3.27	68.87±8.28	92.78±1.73	94.37±1.21	90.88±1.81
CapUBS	70.93±8.85	73.26±11.97	66.68±10.33	76.87±3.59	84.15±2.67	70.75±4.29	**92.85±1.23**	**94.56±1.03**	**91.21±1.51**
DEHF	**75.30±4.39**	77.93±5.34	**71.80±4.98**	**78.24±4.75**	**85.27±3.41**	**72.43±5.68**	**93.13±0.85**	**94.80±0.74**	**91.55±1.04**

TABLE III
CLASSIFICATION RESULTS OF VARIOUS BS METHODS USING KNN ON THREE DATASETS

Method	IP220			PU103			DC191		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
MVPCA	47.58±1.89	52.51±2.48	41.18±2.10	60.15±8.27	68.77±9.09	50.93±9.64	87.99±2.25	90.39±1.97	85.25±2.74
E-FDPC	50.25±4.58	55.23±5.50	43.96±5.01	69.51±0.29	**79.39±0.29**	61.75±0.36	85.81±0.18	88.62±0.10	82.62±0.22
ASPS	53.01±4.97	58.30±4.73	47.07±5.47	69.67±1.15	**79.37±1.69**	61.94±1.37	89.22±0.68	91.74±0.60	86.77±0.83
OCF	51.71±2.85	57.51±3.34	45.60±3.14	69.58±0.47	78.89±0.53	61.79±0.56	89.52±0.80	91.94±0.69	87.14±0.97
S^4P	54.45±9.21	59.45±9.98	48.72±10.18	69.34±0.64	78.73±0.62	61.62±0.73	**89.92±0.20**	**92.20±0.28**	**87.63±0.25**
TGSR	56.55±2.28	61.01±3.91	50.92±2.54	**69.83±1.02**	79.13±1.38	**62.12±1.24**	89.68±0.11	**92.14±0.08**	87.34±0.14
SSANet-BS	54.16±16.14	59.34±17.14	48.65±17.36	68.91±1.50	77.98±2.39	61.03±1.80	87.72±1.31	90.38±1.14	84.94±0.16
CAE	**59.88±4.59**	**66.11±5.22**	**54.71±5.08**	69.35±1.49	79.29±1.54	61.62±1.78	87.43±1.25	90.13±1.39	84.60±1.54
CapUBS	54.57±6.01	59.61±7.10	48.81±6.56	69.15±0.54	78.08±0.60	61.30±0.64	89.41±0.42	91.91±0.32	87.01±0.51
DEHF	**58.30±1.91**	**62.94±2.88**	**52.94±2.11**	**70.12±0.86**	79.31±0.87	**62.47±1.04**	**89.72±0.10**	92.13±0.09	**87.38±0.12**

TABLE IV
CLASSIFICATION RESULTS OF VARIOUS BS METHODS USING CNN ON THREE DATASETS

Method	IP220			PU103			DC191		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
MVPCA	57.39±1.57	57.42±3.19	51.24±1.86	66.55±12.49	74.90±9.87	59.01±13.65	89.99±1.48	92.08±1.02	87.70±1.78
E-FDPC	64.48±4.69	69.09±6.21	59.48±5.37	70.52±1.59	80.98±0.50	63.20±1.72	88.73±0.56	91.49±0.31	86.21±0.67
ASPS	64.96±5.27	68.24±7.26	59.99±6.07	71.98±2.39	81.35±1.11	64.89±2.85	91.78±1.43	93.62±0.89	89.90±1.54
OCF	63.48±3.46	68.83±5.29	58.36±4.06	72.24±2.45	81.70±1.11	65.15±2.82	91.73±1.43	93.59±1.03	89.84±1.74
S^4P	66.69±4.87	70.19±7.45	62.05±5.60	**75.69±1.68**	**83.60±0.65**	**69.31±1.90**	91.58±0.89	93.40±0.67	89.66±1.08
TGSR	**70.88±2.97**	**74.79±4.12**	**66.75±3.41**	73.38±2.26	81.63±2.31	66.50±2.72	92.25±0.67	93.95±0.55	90.46±0.82
SSANet-BS	63.50±13.49	65.14±16.63	58.40±15.26	71.89±1.26	80.42±2.10	64.73±1.52	89.90±2.05	92.36±1.40	87.62±2.48
CAE	63.36±5.88	**75.55±6.40**	65.19±6.59	72.87±4.61	82.08±2.22	66.03±5.28	92.29±0.83	94.05±0.64	90.52±1.01
CapUBS	67.19±7.40	70.46±1.03	62.56±8.54	72.42±2.20	81.46±1.12	65.47±2.49	**92.44±0.64**	**94.13±0.52**	**90.70±0.79**
DEHF	**70.85±2.74**	74.50±4.68	**66.85±3.11**	**75.02±2.64**	**82.74±2.06**	**68.44±3.17**	**92.64±0.23**	**94.24±0.15**	**90.95±0.27**

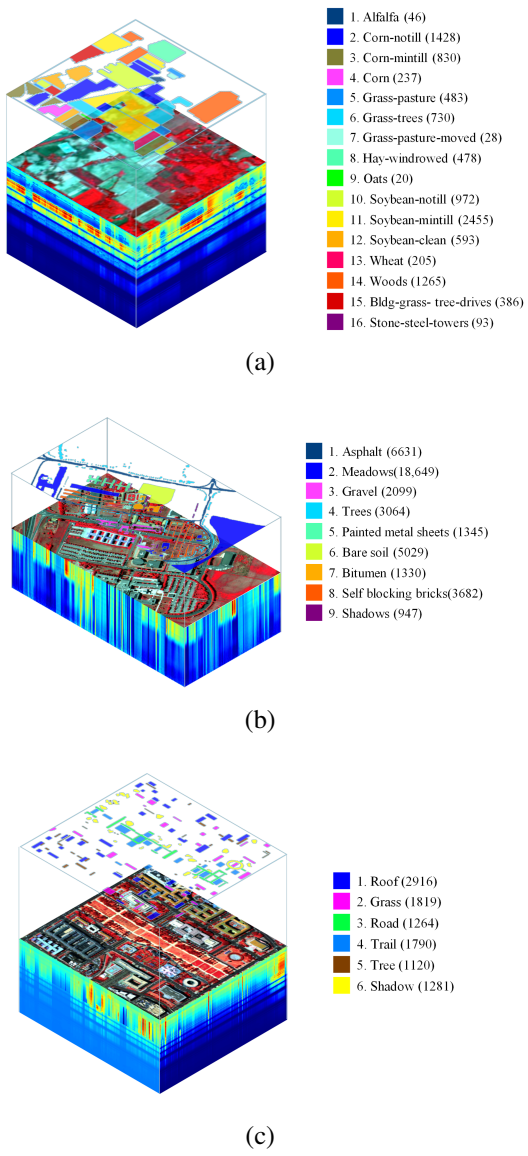


Fig. 3. The datasets used in the experiment. The land cover types and the number of samples for each dataset are indicated, respectively. (a) IP220. (b) PU103. (c) DC191.

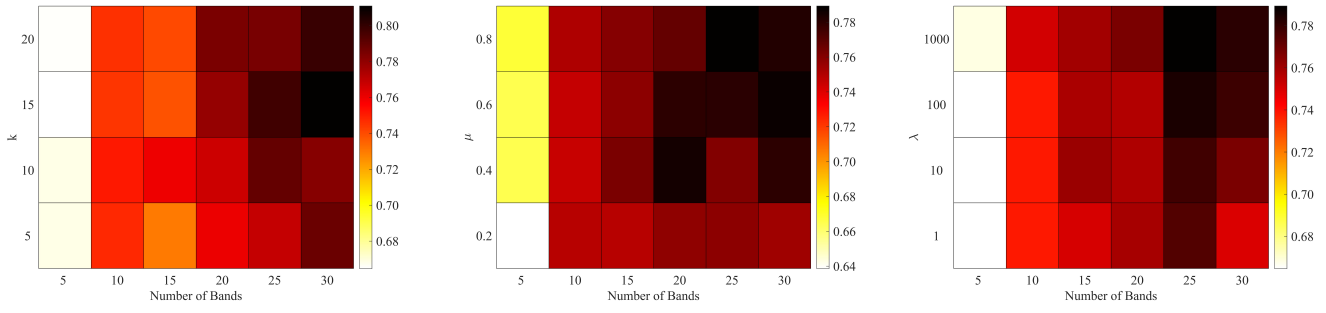
on classification accuracy. Taking the IP220 dataset as an example, Fig. 4 shows the OA values under varying parameter settings across different band subsets. The results indicate that both excessively large and small k values reduce model accuracy. When $k = 10$ or $k = 15$, DEHF achieves superior performance across different band subsets. This is because too few neighbors hinder the model from capturing the complete graph structure, while too many may lead to similar features across nodes, thereby neglecting local information. Additionally, as μ increases, the model's accuracy improves. This is attributed to the greater contribution of the new graph in the updating process, which introduces a higher quality band local structure for subsequent model learning, thereby enhancing the precise expression of band adjacency relationships in the graph. Similarly, increasing the loss weight λ also enhances the model's performance. This indicates that the reconstruction loss L_{AE} plays a dominant role in the objective function, introducing more discriminative features during the fusion process, and ultimately improving the quality of the learned graph.

D. Results

To assess the overall effectiveness of each method under varying subset sizes and datasets, the average OA, AA and Kappa results obtained by various BS methods on the three datasets are reported in Tables II-IV, corresponding to the SVM-, KNN-, and CNN-based classifiers, respectively. These metrics are averaged over six band subset sizes, ranging from 5 to 30 with a step size of 5. The optimal and suboptimal results are highlighted in bold red and black, respectively.

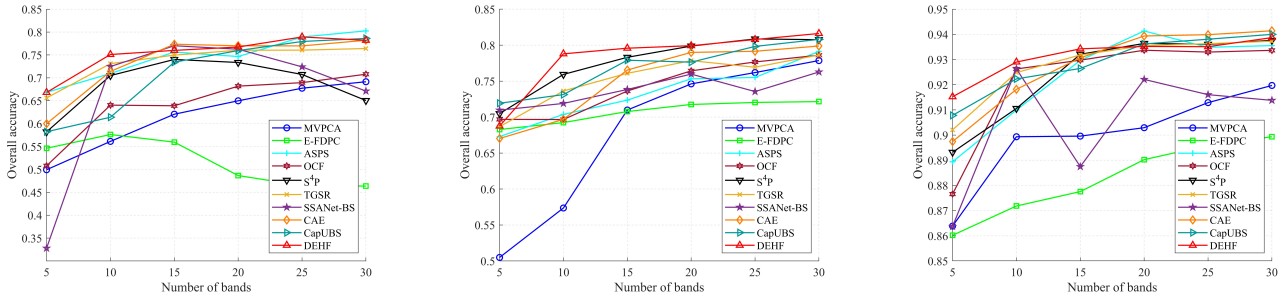
Table II clearly shows that in terms of average OA, DEHF ranks first in all datasets—IP220, PU103, and DC191—surpassing the suboptimal methods ASPS, S^4P, and CapUBS by approximately 0.75%, 0.54%, and 0.28%, respectively. The Kappa values follow a similar trend, with even larger margins, further highlighting DEHF's advantage. It should be noted that, although the suboptimal methods perform well on the specific dataset, their performance fluctuates significantly across different datasets, indicating poor generalizability. Moreover, except for the IP220 dataset where DEHF's average AA is slightly lower than that of ASPS, it achieves the highest AA on PU103 and DC191. Table III further verifies the robustness of DEHF with the KNN classifier. On the PU103 dataset, DEHF attains the highest OA and Kappa values, with its AA being only 0.08% lower than that of the optimal method, E-FDPC. On the IP220 and DC191 datasets, it also maintains strong competitiveness, with most metrics typically ranking at suboptimal position. Although CAE achieves the best result on the IP220 dataset, its performance on DC191 is nearly 2% lower than that of the optimal method, indicating poor stability. In contrast, DEHF exhibits stable and reliable performance across all datasets, consistently ranking first or second. Similarly, Table IV shows that when using the CNN classifier, DEHF achieves the highest OA, AA, and Kappa values on the DC191 dataset. On the IP220 and PU103 datasets, DEHF also demonstrates competitive results, with all three metrics remaining at suboptimal levels. These findings demonstrate that DEHF can effectively capture both key object features and band spatial structure, demonstrating its stable and superior performance across various datasets and subset sizes. In summary, the experimental results on the three classifiers demonstrate that the proposed DEHF consistently achieves top or near-top performance on all datasets, thereby confirming its superiority and robustness across diverse data scenarios.

Furthermore, to evaluate each BS method's efficacy across different subset sizes, Fig. 5 illustrates the OA values for band numbers from 5 to 30 bands. It is evident that MVPCA and E-FDPC exhibit relatively poor performance across all three datasets. Although ASPS, OCF, SSANet-BS, and CapUBS achieve competitive classification results on certain datasets, their overall performance lacks consistency. For example, OCF and CapUBS demonstrate strong competitiveness on the DC191 dataset; however, their performance on IP220 and PU103 is unsatisfactory. CAE performs well on IP220 and



(a) (b) (c)

Fig. 4. Sensitivity test of parameters in terms of OA by SVM on the IP220 dataset. (a) k. (b) μ. (c) λ.



(a) (b) (c)

Fig. 5. OA values of SVM classifier with various BS approaches on the three datasets. (a) IP220. (b) PU103. (c) DC191.

DC191, but its performance on PU103 is noticeably inferior. Similarly, SSANet-BS performs strongly on IP220 and PU103 but fails to achieve high classification accuracy on DC191. In contrast, S^{4}P and TGSR exhibit relatively stable performance across all datasets, though their classification accuracy remains suboptimal. Compared to these methods, the proposed DEHF reveals its clear advantages. On the IP220 dataset, DEHF achieves higher OA values than other BS methods in most cases. Notably, while some methods such as E-FDPC and S^{4}P display a declining OA trend as the number of selected bands increases, DEHF maintains a generally upward trajectory. This suggests that the band subset selected by DEHF has lower redundancy and provides more discriminative information for classification. Additionally, on the PU103 dataset, DEHF performs comparably to S^{4}P, while on DC191, it significantly outperforms all competing methods in most cases.

Finally, Figs. 6-8 present a visual comparison of the classification maps generated with 15 bands selected by different BS methods across three datasets. Consistent with the previous findings, DEHF performs well on all datasets, except on the IP220 dataset, where DEHF is slightly lower than CAE and SSANet-BS. Overall, DEHF performs superiorly on different datasets with good stability, verifying its effectiveness in the BS task.

E. Ablation Study

To validate the effectiveness of the DEGCN module, hierarchical feature fusion module, and representative BS module in DEHF, we conduct a series of ablation studies. Initially, a simple AE network is used as the baseline framework, with its results provided in row (a) of Table V. Furthermore, DEGCN is incorporated into the AE framework, and a linear fusion strategy is employed to examine the impact of spectral-spatial feature interaction, as shown in row (b) of Table V. In order to assess the advantages of the multi-dimensional priority strategy within the representative BS module, we individually apply the IE-based criterion and the I^{2}cS-based criterion to select the final band subset, with results listed in rows (c) and (d) of Table V, respectively. The results of the proposed DEHF are reported in row (e). In this context, (a), (b), and (e) constitute a comparative analysis to investigate the effectiveness of spectral structure information and the hierarchical feature fusion. Meanwhile, (c), (d), and (e) form another comparison for the purpose of evaluating the rationale of integrating both criteria within the representative BS module.

Comparing rows (a) and (b) of Table V, it is evident that AE-DEGCN combined model outperforms the baseline AE on all three datasets, particularly on PU103, where the AOA improves by more than 4%. This indicates that incorporating DEGCN enhances the model's ability to capture the spatial structural information of bands, thereby significantly improving performance. Furthermore, the AOA of DEHF surpasses that of the linear fusion model in row (b) for all datasets, indicating that the proposed hierarchical feature fusion strategy can capture the correlation between the key sample features and the band features considering the spatial information more effectively, thus improving the stability of the model. Additionally, the last three rows of the table show that the

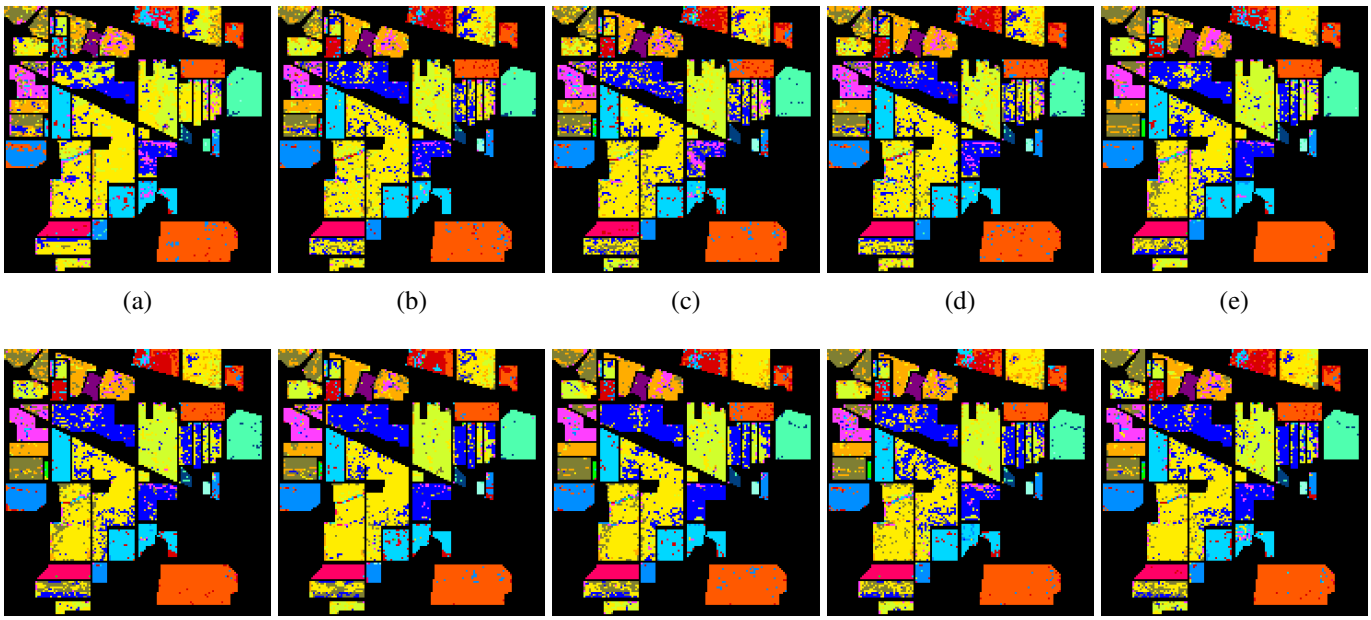


Fig. 6. Classification maps for IP220 dataset under 15 bands. (a) MVPCA (63.17%). (b) E-FDPC (56.53%). (c) ASPS (76.44%). (d) OCF (64.29%). (e) S^{4}P (74.43%). (f) TGSR (75.91%). (g) SSANet-BS (78.30%). (h) CAE (78.87%). (i) CapUBS (74.38%). (j) DEHF (76.73%).

TABLE V
AOA VALUES (%) OF FIVE MODELS FOR THREE DATASETS

Model	HF	IE	I^{2}cS	IP220	PU103	DC191
(a) Base				69.70	73.46	92.93
(b)		✓	✓	71.54	77.98	92.98
(c)	✓	✓		73.99	77.23	93.07
(d)	✓		✓	73.80	76.93	92.98
(e) Ours	✓	✓	✓	**75.30**	**78.24**	**93.13**

AOA values of the BS models based solely on the IE or I^{2}cS criterion are similar but still fall short of DEHF. In particular, for the IP220 and PU103 datasets, models using only IE or I^{2}cS perform substantially worse than DEHF. This confirms that the proposed MP criteria, which integrate both information content and intra/inter-cluster similarity, offer a more comprehensive assessment of band prioritization than a single metric.

F. Model Complexity

To comprehensively evaluate the efficiency of DEHF, we assess its complexity from theoretical analysis, practical metrics, and computational time perspectives, as detailed below.

*1) **Theoretical Analysis**:* The model's computational complexity is mainly determined by three stages: initial BASG construction, dynamic graph convolutional propagation, and hierarchical feature fusion. The cost of each stage is detailed below.

(1) Initial BASG Construction: The pairwise Euclidean distances between all d spectral bands is first computed, requiring $O(N)$ operations per pair, resulting in a total of $O(d^2N)$ for all pairs. Then, finding the k nearest neighbors for each band adds a sorting cost of $O(d^2log_2d)$. Thus, the overall complexity of this stage is $O(d^2N + d^2log_2d)$.

(2) Dynamic Graph Convolutional Propagation: For the i-th DEGCN layer, the linear transformation has a complexity of $O(dh_{i-1}h_i)$, while neighborhood aggregation introduces an additional cost of $O(d^2h_i)$. Consequently, For an L-layer DEGCN, the total cost can be approximately as $O(\sum_{i=1}^{L} dh_{i-1}h_i + d^2h_i)$.

(3) Hierarchical Feature Fusion: The i-th layer output features of size dh_i, which are concatenated with the AE output to form $2dh_i$ feature vector, then processed by fully connected layers. The complexity of this stage is roughly $O(dh_i)$.

Combining all stages, the overall theoretical complexity of DEHF is $O(d^2N + d^2log_2d + \sum_{i=1}^{L} dh_{i-1}h_i + d^2h_i)$. Since h_i and d are much smaller than N, this can be approximated as $O(d^2N)$.

*2) **Practical Indicator**:* Table VI summarizes practical metrics, including FLOPs, MACs, number of parameters, and memory usage, for SSANet-BS, CAE, CapUBS, and the proposed DEHF, as these indicators mainly assess the computational complexity of deep learning models. DEHF has 9.94G FLOPs, which is higher than the lightweight CAE (0.67G) but much lower than SSANet-BS (40.47G) and CapUBS (2001.95G), indicating a moderate forward computation cost. Regarding MACs, DEHF's 4.97G is reasonable, supporting efficient computation. Moreover, DEHF contains 45.19M parameters, exceeding CAE but well below CapUBS (1284M), demonstrating a good balance between model capacity and computational feasibility. For memory usage, DEHF consumes 255.82MB, substantially less than CapUBS while maintaining strong performance, reflecting good hardware adaptability.

Unlike CAE, which requires retraining for each band subset,

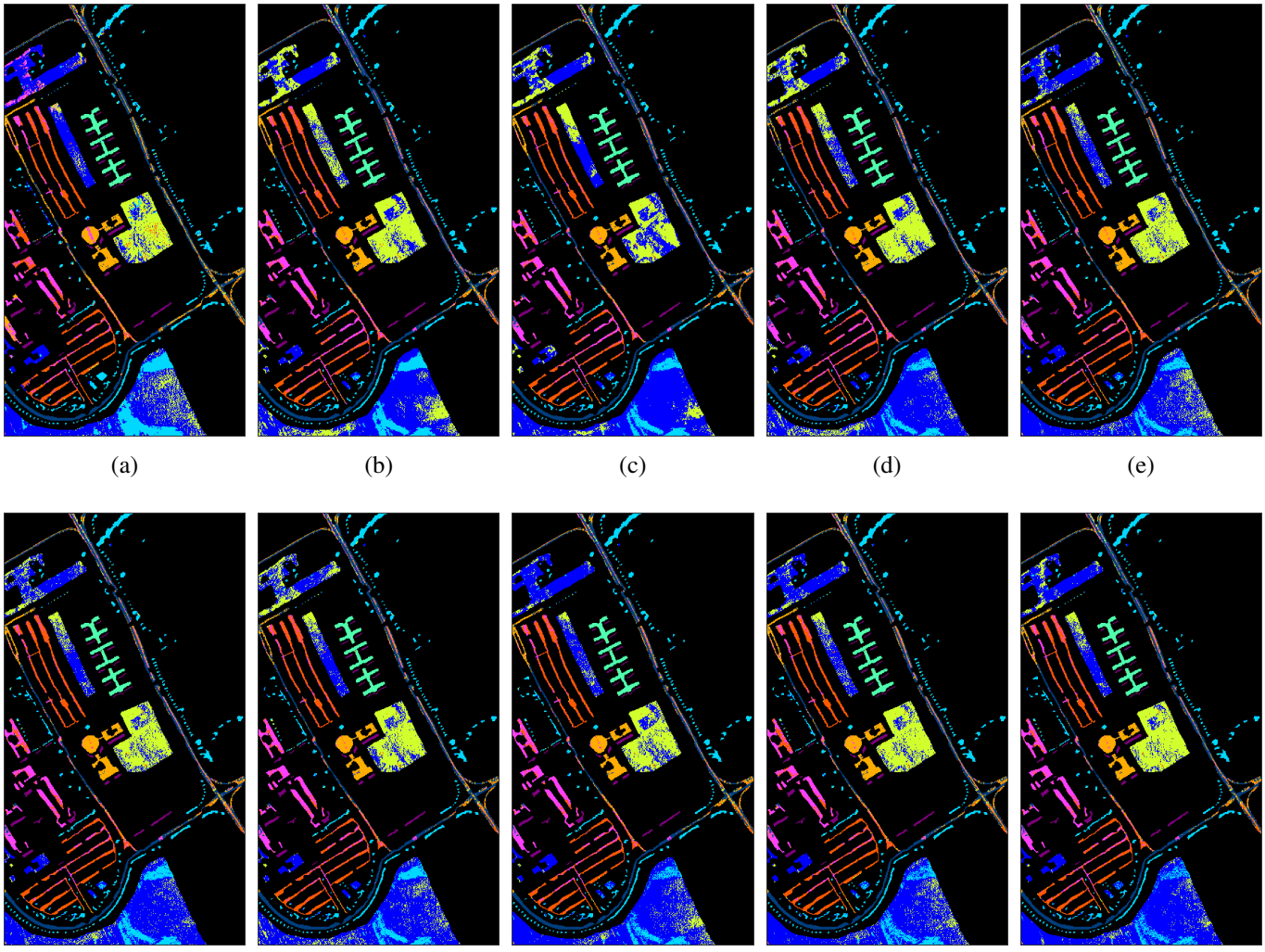


Fig. 7. Classification maps for PU103 dataset under 15 bands. (a) MVPCA (71.47%). (b) E-FDPC (72.27%). (c) ASPS (75.41%). (d) OCF (76.84%). (e) S^4P (78.82%). (f) TGSR (78.16%). (g) SSANet-BS (77.26%). (h) CAE (79.78%). (i) CapUBS (79.36%). (j) DEHF (82.27%).

TABLE VI
COMPLEX METRICS OF DEEP LEARNING-BASED BS METHODS

Method	FLOPs (G)	MACs (G)	Parameters (M)	Memory (MB)
SSANet-BS	40.47	20.23	0.68	3.45
CAE	0.67	0.34	0.03	0.12
CapUBS	2001.95	1000.97	1284	5555.55
DEHF	9.94	4.97	45.19	255.82

DEHF only needs a single training run and can flexibly select subsets of arbitrary size based on feature importance, greatly improving efficiency. Although the dynamic graph refinement module slightly increases computation, the significant gain in classification accuracy offsets this overhead. Overall, DEHF achieves a well-balanced trade-off between classification performance and computational efficiency, demonstrating its superior effectiveness and broad applicability in hyperspectral data analysis.

*3) **Computational Time**:* Table VII presents the runtime (s) for selecting 15 bands on the IP220 dataset using different BS methods. Deep learning-based methods, including SSANet-BS, CAE, CapUBS, and the proposed DEHF, are run on GPUs, while the others run on CPUs. The first four machine learning-based methods are computationally efficient with short runtimes. In contrast, TGSR and S^4P, which rely on iterative optimization, exhibit higher computational complexity. Although deep learning-based methods generally have higher computational costs due to parameter training and feature learning, DEHF achieves notably lower runtime than other deep learning BS methods, thanks to its lightweight design with a shallow autoencoder and DEGCN network. In conclusion, despite the typically higher computational cost of deep learning methods, DEHF significantly enhances computational efficiency while maintaining robust BS performance, outperforming other deep learning methods in speed and showcasing strong application potential.

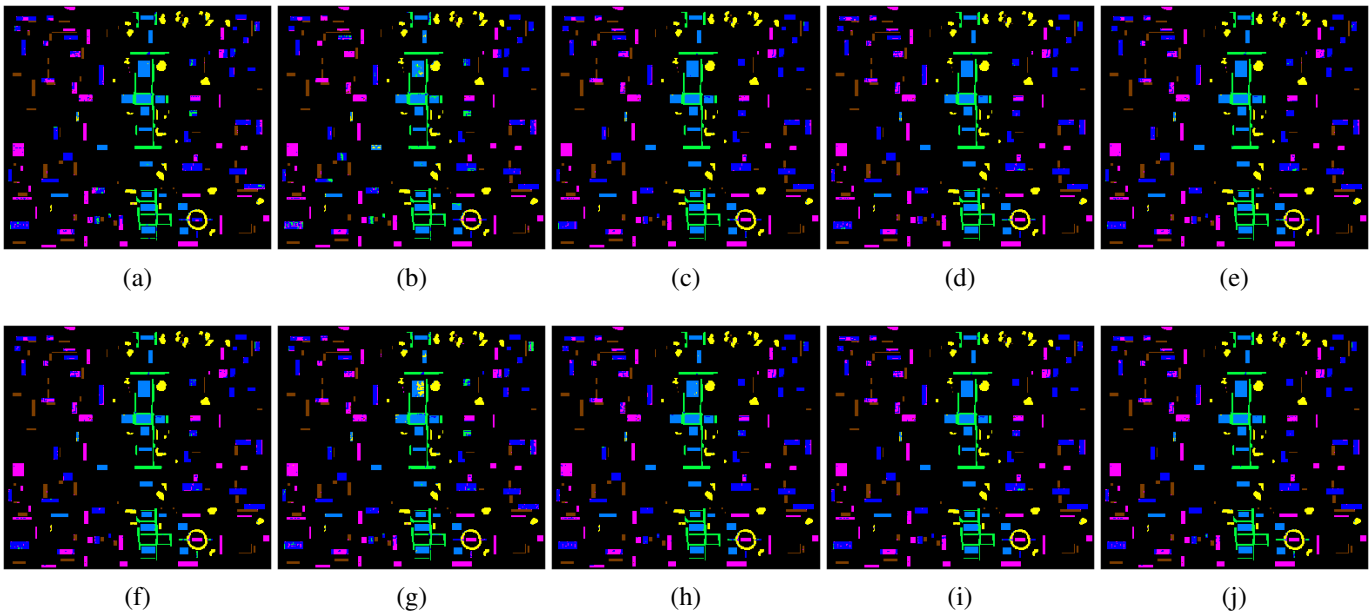


Fig. 8. Classification maps for DC191 dataset under 15 bands. (a) MVPCA (89.59%). (b) E-FDPC (87.68%). (c) ASPS (93.25%). (d) OCF (92.76%). (e) S^4P (93.27%). (f) TGSR (93.34%). (g) SSANet-BS (88.61%). (h) CAE (93.13%). (i) CapUBS (92.88%). (j) DEHF (93.86%).

TABLE VII
RUNTIME(S) OF SELECTING 15 BANDS BY DIFFERENT BS METHODS ON THE IP220 DATASET

MVPCA	E-FDPC	ASPS	OCF	S^4P
0.04	0.05	0.16	0.17	16.31
TGSR	SSANet-BS	CAE	CapUBS	DEHF
26.88	317.73	590.60	7801.49	27.51

TABLE VIII
THE WILCOXON SIGNED-RANK TEST FOR DIFFERENT BS METHODS ($\alpha = 0.05$)

Method	**IP220**	**PU103**	**DC191**
MVPCA	+	+	+
E-FDPC	+	+	+
ASPS	≈	+	≈
OCF	+	≈	+
S^4P	+	≈	≈
TGSR	+	+	≈
SSANet-BS	≈	≈	+
CAE	≈	+	≈
CapUBS	≈	≈	≈
DEHF	/	/	/

V. DISCUSSION

A. Significance Analysis

To further identify whether there is a significant difference between the proposed DEHF and the comparative methods, we conducted a Wilcoxon signed-rank test. The analysis is based on the OA values with six different band subset sizes (ranging from 5 to 30 with a step size of 5) across multiple datasets, using a significance level threshold $\alpha = 0.05$. The results of Wilcoxon signed-rank test are summarized in Table VIII, where the symbols are interpreted as follows: "+" indicates that DEHF significantly outperforms the comparison method, "≈" denotes no significant difference, "–" means that DEHF underperforms, and "/" represents a self-comparison.

Among the 30 valid comparisons, DEHF achieves 14 "+" and 13 "≈", with no instances of "–". These results clearly demonstrate the superior performance and strong competitiveness of DEHF. Notably, DEHF demonstrates a significant performance advantage over most competing methods on the IP220 and PU103 datasets. Although its performance on the DC191 dataset is slightly lower than on the first two, it remains highly competitive. In comparison with the newly deep learning-based method CapUBS, there is no statistically significant difference according to the significance analysis; however, DEHF achieves higher classification accuracy. Overall, the Wilcoxon signed-rank test provides strong evidence supporting the effectiveness and superiority of DEHF over the baseline methods.

B. Band Quality Evaluation

In order to evaluate the quality of the selected bands, Fig. 9 (a) and (b) illustrate the correlation and information entropy of the band subsets selected by various BS methods. Generally, lower correlation and higher information entropy indicate stronger complementarity and richer information representation, leading to better discriminative capability.

As shown in Fig. 9 (a), CapUBS and CAE performs best on IP220 and PU103 datasets, respectively. S^4P, which relies on a sparse constraint, yields low-redundancy band subsets on DC191 datasets. In comparison, the DEHF-selected band subset achieves near-optimal performance on DC191 and outperforms more than half of the BS methods on IP220. Although DEHF is slightly less competitive on PU103, it still achieves a relatively high OA value (Fig. 5 (b)). This confirms that the DEGCN within DEHF effectively models the band adjacency relationships, while the integration of the

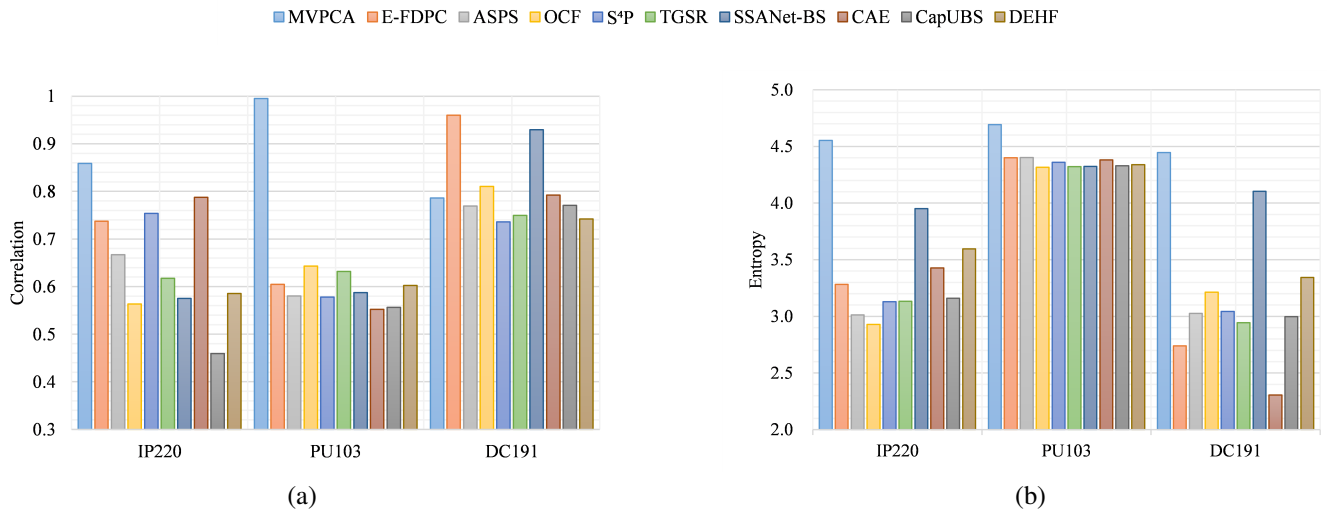


Fig. 9. Band quality of 15 bands selected by different BS methods. (a) Correlation (↓). (b) Entropy (↑).

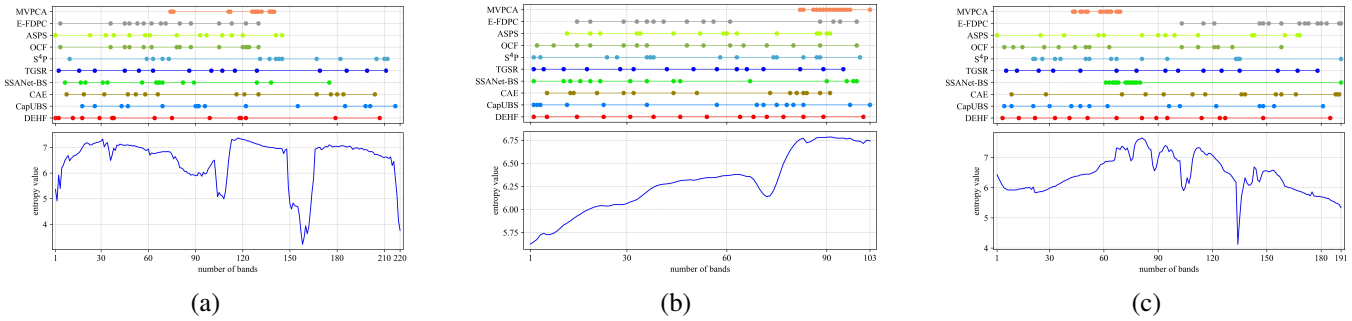


Fig. 10. The distribution of the 15 bands selected by each BS method (top) and the entropy of each band (bottom). (a) IP220. (b) PU103. (c) DC191.

I^2cS-based criterion in spectral clustering significantly reduces band redundancy. Furthermore, as shown by the entropy values in Fig. 9 (b), MVPCA selects bands via a weighted sum of the principal component transformation vectors and their corresponding information content, effectively capturing high-information bands but often introducing significant redundancy. Among all methods, DEHF consistently attains high entropy levels across all three datasets, demonstrating that the IE-based criterion can effectively guide the selection of informative and representative bands.

To intuitively illustrate the selected bands, Fig. 10 shows the 15 bands selected by different methods along with their entropy values across three datasets. As shown in Fig. 10 (a), for the IP220 dataset, MVPCA, E-FDPC, ASPS, OCF, and SSANet-BS select relatively concentrated bands, while S^4P, TGSR, CAE, CapUBS and DEHF produce more dispersed distributions, enabling broader spectral feature coverage. However, according to the entropy curves, most bands selected by S^4P are concentrated within three narrow intervals (59-75, 130-145, and 205-215), limiting the coverage of the full spectral information and potentially reducing the ability to discriminate different land-cover types. Although TGSR and CapUBS achieve wider and more uniform distributions, they still include several low-entropy bands that may degrade performance. Similarly, CAE yields relatively uniform band distributions and reduces redundancy but may miss spectrally important bands, leading to less ideal classification results.

In contrast, DEHF employs a dynamic graph enhancement module and multi-dimensional band priority strategy to select bands spanning a wider spectral range while avoiding low-entropy regions, providing richer discriminative information and superior classification performance. Furthermore, as illustrated in Figs. 10 (b) and 10 (c), DEHF maintains a similarly uniform and dispersed band distribution on the PU103 and DC191 datasets, effectively avoiding low-information bands. This demonstrates its robustness and strong generalizability across datasets with complex and diverse spectral characteristics. Overall, considering both spectral redundancy and image information, the proposed DEHF achieves outstanding performance. The selected bands exhibit strong discriminative capability for ground objects, further validating DEHF's effectiveness in BS tasks.

VI. Conclusion

In this paper, we propose a novel dynamic graph enhancement and hierarchical feature fusion (DEHF) model for hyperspectral BS. DEHF first performs feature cleansing on raw hyperspectral data to suppress noise interference and focus on the key latent features. This process provides clearer learning objectives for subsequent BS. Subsequently,

a dual-stage optimization strategy is used to model the local structural relationships between bands, in which a dynamic graph enhancement module is introduced to adaptively capture the complex nonlinear inter-band correlations. Furthermore, a hierarchical feature fusion mechanism is incorporated to effectively integrate the information of different spatial scales and semantic levels, thereby enhancing both the robustness and discriminativity of the feature representation. Furthermore, a multi-dimensional priority strategy is designed, which takes into account information content, inter-band complementarity, and redundancy, thus ensuring that the final selected subset of bands is representative. Extensive experiments conducted on three widely used hyperspectral datasets demonstrate that DEHF outperforms the existing state-of-the-art methods in terms of band quality, method saliency, and computational efficiency.

Although DEHF utilizes the AE to effectively reduce the influence of noise on BS, its robustness remains limited when faced with complex noise distributions across different sensors or scenes. Moreover, determining the optimal number of selected bands remains a challenging task. These issues will be further investigated in future work.

REFERENCES

[1] H. Yu, Z. Ling, K. Zheng, L. Gao, J. Li, and J. Chanussot, "Unsupervised Hyperspectral and Multispectral Image Fusion With Deep Spectral-Spatial Collaborative Constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1-14, 2024, Art. no. 5534114, doi: 10.1109/TGRS.2024.3472226.

[2] B. Xue, N. Li, Q. Du, and W. Li, "A Subpixel Target Detection Approach to Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5093–5114, Sept. 2017, doi: 10.1109/TGRS.2017.2702197.

[3] G. Shaw and D. Manolakis, "Signal Processing for Hyperspectral Image Exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002, doi: 10.1109/79.974715.

[4] Y. Chen, X. Wang, J. Zhang, X. Shang, Y. Hu, S. Zhang, and J. Wang, "A New Dual-Branch Embedded Multivariate Attention Network for Hyperspectral Remote Sensing Classification," *Remote Sens.*, vol. 16, p. 2029, 2024, doi: 10.3390/rs16112029.

[5] D. Zhang, L. Lu, X. Li, J. Zhang, S. Zhang, and S. Yang, "Spatial Downscaling of ESA CCI Soil Moisture Data Based on Deep Learning with an Attention Mechanism," *Remote Sens.*, vol. 16, p. 1394, 2024, doi: 10.3390/rs16081394

[6] J. Jia et al., "Tradeoffs in the Spatial and Spectral Resolution of Airborne Hyperspectral Imaging Systems: A Crop Identification Case Study," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5510918, doi: 10.1109/TGRS.2021.3096999.

[7] Y. Guo, X. Zhao, X. Sun, J. Zhang, and X. Shang, "Sample Latent Feature-Associated Low-Rank Subspace Clustering for Hyperspectral Band Selection," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 14050–14063, 2024, doi: 10.1109/JSTARS.2024.3435846.

[8] K. Abend, T. Harley, B. Chandrasekaran, and G. Hughes, "Comments on 'On the Mean Accuracy of Statistical Pattern Recognizers' by Hughes, G. F.," *IEEE Trans. Inf. Theory*, vol. 15, no. 3, pp. 420–423, May 1969, doi: 10.1109/TIT.1969.1054314.

[9] B. Xu, X. Li, W. Hou, Y. Wang, and Y. Wei, "A Similarity-Based Ranking Method for Hyperspectral Band Selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9585–9599, Nov. 2021, doi: 10.1109/TGRS.2020.3048138.

[10] Y.-J. Deng, H.-C. Li, S.-Q. Tan, J. Hou, Q. Du, and A. Plaza, "T-Linear Tensor Subspace Learning for Robust Feature Extraction of Hyperspectral Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501015, doi: 10.1109/TGRS.2023.3233945.

[11] Y. Zhang et al., "Structural and Textural-Aware Feature Extraction for Hyperspectral Image Classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 5502305, doi: 10.1109/LGRS.2024.3357201.

[12] A. Maćkiewicz and W. Ratajczak, "Principal Components Analysis (PCA)," *Comput. Geosci.*, vol. 19, no. 3, pp. 303–342, 1993, doi: 10.1016/0098-3004(93)90090-R.

[13] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000, doi: 10.1126/science.290.5500.2323.

[14] J. B. Tenenbaum et al., "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000, doi: 10.1126/science.290.5500.2319.

[15] B. Fu, X. Sun, C. Cui, J. Zhang, and X. Shang, "Structure-Preserved and Weakly Redundant Band Selection for Hyperspectral Imagery," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 12490–12504, 2024, doi: 10.1109/JSTARS.2024.3425906.

[16] X. Sun, P. Lin, X. Shang, H. Pang, and X. Fu, "MOBS-TD: Multiobjective Band Selection with Ideal Solution Optimization Strategy for Hyperspectral Target Detection," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 10032–10050, 2024, doi: 10.1109/JSTARS.2024.3402381.

[17] 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, doi: 10.1109/36.803411.

[18] T. Hu, X. Zhang, Z. Sun, S. Ye, and P. Gao, "Contrasting Effects of Structural Similarity and Entropic Metrics on Band Selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 5505205, doi: 10.1109/LGRS.2024.3389057.

[19] 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, doi: 10.1109/TGRS.2015.2450759.

[20] H. Sun, L. Zhang, J. Ren, and H. Huang, "Novel Hyperbolic Clustering-Based Band Hierarchy (HCBH) for Effective Unsupervised Band Selection of Hyperspectral Images," *Pattern Recognit.*, vol. 130, 2022, Art. no. 108788, doi: 10.1016/j.patcog.2022.108788.

[21] Y. Zhang, J. Qi, X. Wang, Z. Cai, J. Peng, and Y. Zhou, "Tensorial Global-Local Graph Self-Representation for Hyperspectral Band Selection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 12, pp. 13213–13225, Dec. 2024, doi: 10.1109/TCSVT.2024.3439369.

[22] X. Shang, C. Cui, and X. Sun, "Spectral-Spatial Hypergraph-Regularized Self-Representation for Hyperspectral Band Selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5504405, doi: 10.1109/LGRS.2023.3276055.

[23] Q. Wang, Y. Liu, K. Xu, Y. Dong, F. Cheng, Y. Tian, B. Du, and X. Zhang, "Multi-Objective Evolutionary Multi-Tasking Band Selection Algorithm for Hyperspectral Image Classification," *Swarm Evol. Comput.*, vol. 90, 2024, Art. no. 101665, doi: 10.1016/j.swevo.2024.101665.

[24] 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, doi: 10.1109/TGRS.2023.3305545.

[25] Y. Wang, Q. Zhu, H. Ma, and H. Yu, "A Hybrid Gray Wolf Optimizer for Hyperspectral Image Band Selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527713, doi: 10.1109/TGRS.2022.3167888.

[26] C. Cui, X. Sun, B. Fu, and X. Shang, "SSANet-BS: Spectral–Spatial Cross-Dimensional Attention Network for Hyperspectral Band Selection," *Remote Sens.*, vol. 16, no. 15, 2024, Art. no. 2848, doi: 10.3390/rs16152848.

[27] 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, doi: 10.1109/TGRS.2019.2951433.

[28] T. Li, Y. Cai, Z. Cai, X. Liu, and Q. Hu, "Nonlocal Band Attention Network for Hyperspectral Image Band Selection," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 3462–3474, 2021, doi: 10.1109/JSTARS.2021.3065687.

[29] S. K. Roy, S. Das, T. Song, and B. Chanda, "DARecNet-BS: Unsupervised Dual-Attention Reconstruction Network for Hyperspectral Band Selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 12, pp. 2152–2156, Dec. 2021, doi: 10.1109/LGRS.2020.3013235.

[30] H. Zhang, X. Sun, Y. Zhu, F. Xu, and X. Fu, "A Global-Local Spectral Weight Network Based on Attention for Hyperspectral Band Selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art. no. 6004905, doi: 10.1109/LGRS.2021.3130625.

[31] D. Yan, X. Sun, J. Zhang and X. Shang, "Dual-View Structural Similarity Subspace Clustering for Hyperspectral Band Selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 22, pp. 1–5, 2025, Art. no. 5503205, doi: 10.1109/LGRS.2025.3554356.

[32] J. Feng, Z. Ye, S. Liu, X. Zhang, J. Chen, R. Shang and L. Jiao, "Dual-Graph Convolutional Network Based on Band Attention and Sparse

Constraint for Hyperspectral Band Selection," *Knowl.-Based Syst.*, vol. 231, 2021, Art. no. 107428, doi: 10.1016/j.knosys.2021.107428.

[33] W. Feng, Y. Li, Z. Wang, and X. Chen, "Diversity Learning Guided Dual Graph Autoencoder for Unsupervised Hyperspectral Band Selection," *IEEE Trans. Circuits Syst. Video Technol.*, early access, doi: 10.1109/TCSVT.2025.3547137.

[34] 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, doi: 10.1109/TGRS.2021.3075663.

[35] L. Xu, M. Ahishali and M. Gabbouj, "Dropout Concrete Autoencoder for Band Selection on Hyperspectral Image Scenes," *IEEE Geosci. Remote Sens. Lett*, vol. 22, 2025, Art no. 5504705, doi: 10.1109/LGRS.2025.3564478.

[36] Y. Liu, X. Li, Z. Hua, C. Xia and L. Zhao, "A Band Selection Method With Masked Convolutional Autoencoder for Hyperspectral Image," *IEEE Geosci. Remote Sens. Lett*, vol. 19, pp. 1-5, 2022, Art no. 6010005, doi: 10.1109/LGRS.2022.3178824.

[37] M. Jijón-Palma, C. Amisse, J. Macuácua and J. Centeno, "Noisy band selection based on the integration of the Stacked-Autoencoder and Convolutional Neural Network approaches for hyperspectral data," *Geosciences (Geociências)*, vol. 42, pp. 269-280, 2023, doi: 10.5016/geociencias.v42i2.16976.

[38] 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, 2021, pp. 147-151, doi: 10.1109/LGRS.2020.2967815.

[39] M. Ahishali, S. Kiranyaz, I. Ahmad and M. Gabbouj, "SRL-SOA: Self-Representation Learning with Sparse 1D-Operational Autoencoder for Hyperspectral Image Band Selection," *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2022, pp. 2296-2300, doi: 10.1109/ICIP46576.2022.9897863.

[40] H. Sun, L. Zhang, L. Wang and H. Huang, "Stochastic gate-based autoencoder for unsupervised hyperspectral band selection," *Pattern Recognition*, vol. 132, 2022, Art no. 108969, doi: 10.1016/j.patcog.2022.108969.

[41] P. S. Singh and S. Karthikeyan,"Enhanced classification of remotely sensed hyperspectral images through efficient band selection using autoencoders and genetic algorithm,"*Neural Computing and Applications*,vol. 34, 2022, pp. 21539-21550, doi: 10.1007/s00521-021-06121-4.

[42] V. Sangeetha and L. Agilandeeswari,"A New Band Selection Approach Integrated with Physical Reflectance Autoencoders and Albedo Recovery for Hyperspectral Image Classification,"*Scientific Reports*, vol. 15, Art. no. 27964, 2025, doi: 10.1038/s41598-025-09355-7.

[43] C. Yu, S. Zhou, M. Song, B. Gong, E. Zhao and C. -I. Chang, "Unsupervised Hyperspectral Band Selection via Hybrid Graph Convolutional Network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1-15, 2022, Art no. 5530515, doi: 10.1109/TGRS.2022.3179513.

[44] H. Wang, P. Zhuang, X. Zhang and J. Li, "DBMGNet: A Dual-Branch Mamba-GCN Network for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, pp. 1-17, 2025, Art no. 4410517, doi: 10.1109/TGRS.2025.3564364.

[45] E. Zhao et al., "Thermal Infrared Hyperspectral Band Selection via Graph Neural Network for Land Surface Temperature Retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1-14, 2024, Art no. 5003414, doi: 10.1109/TGRS.2024.3391008.

[46] J. Chen, W. Li, S. Maharjan and H. El-Askary, "Hyperspectral Band Selection via Heterogeneous Graph Convolutional Self-Representation Network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 18, pp. 18543-18560, 2025, doi: 10.1109/JSTARS.2025.3589866.

[47] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang and J. Yang, "Multiscale Dynamic Graph Convolutional Network for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3162-3177, May 2020, doi: 10.1109/TGRS.2019.2949180.

[48] 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, doi: 10.1109/JSTARS.2019.2941454

[49] S. Li, Z. Liu, L. Fang and Q. Li, "Block Diagonal Representation Learning for Hyperspectral Band Selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1-13, 2023, Art no. 5509213, doi: 10.1109/TGRS.2023.3266811.

[50] 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, 2018, doi: 10.1109/TGRS.2018.2828161.

[51] C. Tang et al., "Spatial and Spectral Structure Preserved Self-Representation for Unsupervised Hyperspectral Band Selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1-13, 2023, Art no. 5531413, doi: 10.1109/TGRS.2023.3331236.

[52] Q. Li, J. Wang, X. Zu, D. Chen, K. Zhang and J. Li, "CapUBS: Capsule Network-Based Band Selection for Underwater Hyperspectral Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, pp. 1-16, 2025, Art no. 4206116, doi: 10.1109/TGRS.2025.3566399.