

Dual-Scale Complementary Spatial-Spectral Joint Model for Hyperspectral Image Classification

Huayue Chen , Yue Sun, Xiang Li , Bochuan Zheng , and Tao Chen 

I. INTRODUCTION

Abstract—In the field of hyperspectral image classification, using spatial information as a supplement to spectral information has been widely applied. This article proposes a novel dual-scale complementary spatial-spectral joint classification model (DSCSM) to mitigate the issues of detail loss and insufficient utilization of spatial information, which traditionally lead to lower classification accuracy. In essence, the final classification result is obtained through decision fusion of two complementary feature extraction stages. In the preprocessing stage, a new dual-scale truncated filtering feature extraction method (DTFE) is proposed, which uses truncated filters with two different parameter settings to obtain two scales of smoothed patches, and then fuses them to obtain dual-scale structural features using Kernel principal component analysis. DTFE preserves edge information while smoothing details, effectively removing noise and retaining the dual-scale feature information. In the postprocessing stage, a sub-Markov random walk-based spatial probability optimization method is proposed, which models the spatial association of neighboring pixels, retaining complex textures as well as weak edge information to optimize the classification probability. Finally, the decision fusion strategy is employed to integrate the classification probabilities acquired from the aforementioned two stages. Comparative experiments on six different scene datasets with state-of-the-art classification methods validate that even with a small number of samples, DSCSM can achieve excellent object recognition performance. In addition, comprehensive parameter analysis proves the robustness and stability of the proposed method.

Index Terms—Dual-scale truncated filter, feature extraction, hyperspectral image (HSI) classification, sub-Markov random walk (subRW).

HYPERSPECTRAL images (HSIs) offer an exceptionally ultrahigh spectral resolution, providing hundreds of continuous spectral bands, which not only capture the spatial characteristics of surface objects but also extract wealthy spectral information from each pixel, facilitating more precise classification of target objects. Consequently, HSI classification holds significant application value across a range of fields such as mineral exploration [1], [2], [3], environmental monitoring [4], [5], [6], land cover mapping [7], [8], [9], and agricultural production [10], [11], [12], [13], thus garnering extensive attention [14], [15], [16].

Over the last several decades, substantial efforts have been devoted to developing precise pixel-level classifiers for HSI classification [17], [18], [19]. However, the scarcity of labeled HSI samples poses a challenge, commonly known as the “small sample” problem, which can lead to the Hughes phenomenon [20]. Furthermore, redundancy in spectral and spatial regions as well as noise from outside influences is introduced by the HSI data’s increased spectral and spatial resolution, making HSI classification tasks more difficult. Hence, increasing the accuracy of HSI classification requires dimensionality reduction and efficient feature extraction.

To reduce the spectral dimension of HSI, methods such as principal component analysis (PCA) [21] and independent component analysis (ICA) [22] are commonly employed, which map high-dimensional data into lower dimensional space. Considering the nonlinearity of HSI, kernel-based methods such as kernel principal component analysis (KPCA) [23] and kernel ICA [24] have been proposed. They utilize kernel functions to map data into high-dimensional feature spaces, enhancing data separability. However, HSI faces challenges such as “same thing different spectrum” and “foreign matter same spectrum” problems [10]. The techniques described above do not account for spatial connection among nearby pixels, which frequently results in incorrect labeling and “salt and pepper” appearances in classification results. Therefore, introducing spatial information is a crucial tactic to raise the HSI classification’s accuracy.

To exploit the spatial features of HSI, scholars have proposed numerous methods [25], [26], [27], [28], [29]. Among these, spatial filtering methods are commonly used to extract HSI spatial features because they can use pixels’ spatial proximity information efficiently, reduce spectral heterogeneity, and smooth the image, such as extended morphological profiles [30],

Received 4 September 2024; revised 24 October 2024 and 21 December 2024; accepted 18 January 2025. Date of publication 10 February 2025; date of current version 10 March 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62176217 and in part by the Innovation Team Funds of China West Normal University under Grant KCXTD2022-3. (Corresponding author: Tao Chen.)

Huayue Chen is with the School of Computer Science, China West Normal University, Nanchong 637002, China, and with the Institute of Artificial Intelligence, China West Normal University, Nanchong 637002, China, and also with the Key Laboratory of Optimization Theory and Applications, China West Normal University of Sichuan Province, Nanchong 637002, China (e-mail: sunnyxiaoyue20@cwnu.edu.cn).

Yue Sun and Bochuan Zheng are with the School of Computer Science, China West Normal University, Nanchong 637002, China (e-mail: sunnyue@stu.cwnu.edu.cn; zhengbc@cwnu.edu.cn).

Tao Chen is with the School of Geographical Science, China West Normal University, Nanchong 637002, China (e-mail: chentao@cwnu.edu.cn).

Xiang Li is with the School of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China (e-mail: 220940008@cauc.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2025.3540001

Gabor filters [31], and extended 3-D Gabor filter [32], which are good at extracting the morphological and textural features of HSI. However, the above-mentioned methods may blur and lose ground object edge information due to the lack of special edge processing. Kang et al. [33] applied the edge-preserving filter (EPF) to extract HSI spatial information; it smoothed the image while maintaining consistency with actual ground object boundaries, effectively preserving edge information and significantly improving ground object classification performance. However, EPFs fail to preserve the inherent structure of HSI and cannot remove useless textures [28]. In addition, due to the fixed smoothing characteristic in spatial filtering methods, balancing detail-smoothing and edge-preserving is challenging, leading to reduced robustness. Overall, the above-mentioned approaches generally acquire spatial-spectral information in advance for classification, but some texture information is lost during the feature extraction process, particularly for small objects that are easily overlooked.

To overcome the above-mentioned issues, spectral preclassification results are often optimized for postprocessing based on the spatial correlation between pixels [34]. Random Walker (RW) [35] treats pixels as graph nodes, with edges between nodes representing the adjacency of pixels. By simulating the random movement of walkers through nodes, it preserves spatial information in the image. Due to its robustness to noise and effective segmentation of weak boundaries, it is commonly used as a postprocessing method for optimizing classification probabilities. Traditional RW requires users to select seed nodes in each discrete target region, which is inefficient. Extended Random Walker (ERW) [36] improves efficiency by using node priors to obtain intensity models, thus avoiding user interaction. Kang et al. [37] utilized ERW to describe the spatial relationship and optimize the initial probability map. Although the postprocessing method based on ERW effectively avoids information loss during the preprocessing stage, ERW tends to access categories with lots of samples during propagation, failing to capture differences between similar categories and performing weakly in classifying HSI with complex textures. On the other hand, although RW can alleviate the issue of postprocessing methods being susceptible to noise in the original image, it still loses some detailed information, leading to misclassification.

Techniques for spectral-spatial classification based on deep learning have shown excellent performance lately [38], [39], [40], [41], [42], [43], [44]. Compared with typical machine learning methods, deep learning methods may suffer from overfitting when the number of training samples is limited. However, labeled samples in HSI are limited, representing a typical “small sample” data situation.

To address the bottlenecks mentioned above, considering the different information losses in the pre- and postprocessing stages, Duan et al. [45] presented a dual spatial information fusion classification approach (FDSI), which combined pre- and postprocessing techniques. In the preprocessing stage, a structural contour method (SP) with adaptive texture smoothing was used for feature extraction, and in the postprocessing stage, ERW was utilized to obtain spatial information from the original HSI to optimize classification probabilities. Although FDSI

effectively improves the HSI classification performance, it overlooks differential processing of spectral information, and ERW fails to adequately distinguish similar categories. Li et al. [46] proposed an efficient two-staged method (Two-Staged) based on various solutions of anisotropic and isotropic models to extract multiscale structural information and enhance detail smoothing in HSI, improving land cover discrimination while reducing computational costs. However, Two-Staged is sensitive to image noise and lacks robustness. Shi et al. [47] proposed an adaptive complementary spectral-spatial approach (CSSM), using an adaptive cubic total variation smoothing approach (ACTVSP) to extract coarse structural contours of HSI, followed by EPF to enhance details and edge contours. However, CSSM’s structural features obtained in the preprocessing stage are singular and have numerous parameters. Although the aforementioned methods effectively combine the results of the pre- and postprocessing stages, they still have the following three major issues.

- 1) In the preprocessing stage, smoothing methods cannot balance detail smoothing and edge preservation, leading to excessive smoothing and information loss. In addition, they only consider single-scale features, resulting in poor robustness.
- 2) In the postprocessing stage, RW performs poorly in segmenting complex textured images and has a weak distinguishing ability for similar categories, resulting in inferior optimization of classification probabilities.
- 3) The overall classification framework is complex, depends on an adequate quantity of training samples, and has limited application scenarios.

For the sake of handling the above problems, a dual-scale complementary spatial-spectral joint classification model (DSCSM) is proposed in this article, which integrates dual-scale feature extraction and spatially optimized classification probabilities, fully utilizing the spectral and spatial information of HSI to compensate for information loss in single-stage classification. In the preprocessing stage, a novel dual-scale truncated filtering feature extraction (DTFE) method that preserves edge information and smooths details is proposed. DTFE obtains smooth blocks of different scales through truncated filtering with different parameter settings; after overlaying them, KPCA is used to fuse them into structural features. In the postprocessing stage, the spatial probability optimization (SPO) method based on subRW [48] is proposed to refine the preclassification probabilities classified by SVM. Finally, the classification probabilities obtained from both stages are fused through decision fusion to obtain the final classification probabilities. The following are this article’s primary contributions.

- 1) In the preprocessing stage, a DTFE is proposed to achieve contradictory smoothing behavior. It prevents oversmoothing, minimizes the effect of noise on classification outcomes, preserves edge information while smoothing picture details, and keeps features at different HSI scales.
- 2) In the postprocessing stage, an SPO method based on subRW is proposed to improve RW’s ability to distinguish complex textured images and similar categories. It successfully compensates for information loss during feature

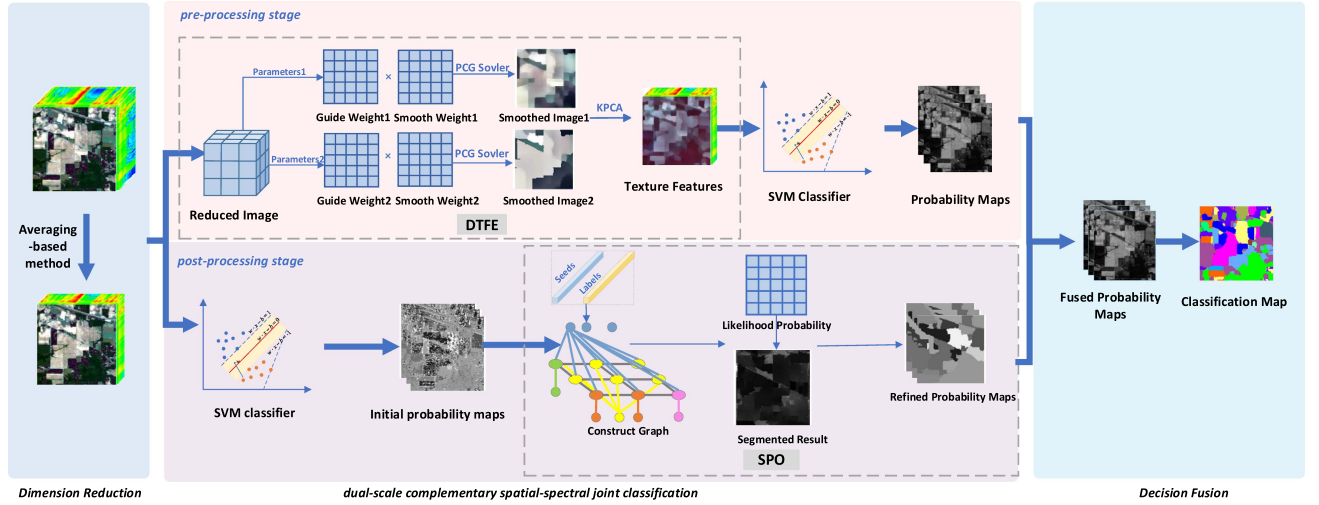


Fig. 1. Flowchart of DSCSM.

extraction and further improves the optimization of classification probabilities by modeling the spatial connections among neighboring pixels using SPO.

- 3) By combining preprocessing and postprocessing steps, a lightweight DSCSM is proposed. This method reduces the information loss of the two process methods and improves classification accuracy. Experiments show that the proposed method can also alleviate the “small sample” problem of HSI and demonstrates effectiveness, robustness, and stability in various application scenarios.

The rest of this article is organized as follows. Section II introduces the proposed DSCSM classification method. Section III discusses the detailed parameter settings. Section IV presents the experimental results and comparative analysis. Section V provides the conclusions and future work.

II. PROPOSED METHOD

To address the loss of information and insufficient utilization of spatial information, this article proposes a DSCSM. The flowchart is shown in Fig. 1.

First, HSI is dimensionally reduced using average dimensionality reduction [49], and the reduced data then enter the pre-and postprocessing stages separately. In the preprocessing stage, the proposed DTFE method mitigates the problem of losing image details. After feature extraction, classification probabilities are obtained through SVM. In the postprocessing stage, the proposed SPO method fully utilizes spatial information by optimizing the preliminary classification probabilities from SVM with SPO, yielding optimal probabilities. The classification probabilities from both stages are combined through decision fusion, compensating for the information loss in both stages and alleviating the “small sample” problem in HSI, resulting in the final classification outcome.

A. Dual-Scale Truncated Filtering Feature Extraction

Detail smoothing and edge preservation in HSI are key issues for improving classification accuracy. While detail smoothing

can reduce image noise, excessive smoothing leads to the loss of important edge information. Edge preservation can retain edge information but areas around edges are prone to retained noise due to improper handling. This article suggests a DTFE approach based on the truncated Huber penalty function, which preserves edge information while smoothing image details.

1) *Truncated Huber Penalty Function*: The smoothing operator’s property is fixed, making it difficult to satisfy the contradictory tasks of edge preservation and detail smoothing. To achieve different smoothing effects, penalty functions can be used to control whether to sharpen edges. Compared to widely used edge penalty functions such as Welsch [50] and Huber [51], the truncated Huber penalty function introduces a new parameter b to saturate the area, preventing the Huber’s value from continuously increasing and thus failing to sharpen edges. Therefore, b can serve as a threshold to adjust edge sharpening. For inputs greater than b , the function only penalizes weak edges and not strong edges. For inputs less than b , the truncated Huber penalty function behaves the same as Huber, acting as an edge-preserving penalty function without edge sharpening

$$h_T(x) = \begin{cases} \frac{1}{2a}x^2, & |x| < a \\ |x| - \frac{a}{2}, & a \leq |x| \leq b \text{ s.t. } a \leq b \\ b - \frac{a}{2}, & |x| > b \end{cases} \quad (1)$$

where a and b are constants.

The function images of $h_T(x)$ and commonly used penalty functions are depicted in Fig. 2, where δ is a tiny constant, set to 10^{-3} . I_m represents the maximum intensity value of the input image. It can be observed that by adjusting the value of b , one can achieve different smoothing effects.

2) *DTFE Model*: Due to the different smoothing behaviors of the Huber penalty function with varying parameters, incorporating it into feature extraction endows DTFE with strong flexibility and robustness, allowing for the realization of different smoothing effects. The pseudocode of the DTFE is shown in Algorithm 1.

Specifically, the algorithm assumes that the dimensionally reduced HSI is f , and sets the guidance image $g = f$. The output

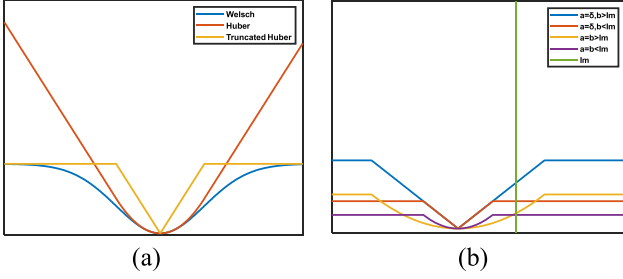


Fig. 2. Plots of (a) various penalty functions and (b) truncated Huber penalty functions with different parameter settings.

Algorithm 1: DTFE Model.

Input: Reduced dimension image f , max iteration $iter$, parameters r^* , b^* , where $*$ $\in \{d, s\}$, $u^0 = R$

- 1: **for** $l = 1$: 2 **do**
- 2: **for** $k = 0$: $iter$ **do**
- 3: Compute $(\nabla_{i,j}^*)^k$ and update $(l_{i,j}^*)^k$ according to (6)
- 4: Update $\nu_{i,j}^*$ according to (7)
- 5: Solve u^{k+1} according to (8)
- 6: **end for**
- 7: Get smooth blocks $F_l \leftarrow u^{N+1}$
- 8: **end for**
- 9: Overlay two scales of smooth blocks (F_1, F_2)
- 10: Using KPCA to obtain structural features

Output: structural features F

u after the truncated filtering smoothing process is the solution to the minimization of the subsequent objective function

$$E_u(u) = \sum_i \sum_{j \in N_d(i)} \omega_{i,j}^s h_T(u_i - f_j) + \frac{1}{2} \sum_i \sum_{j \in N_s(i)} \omega_{i,j}^s \omega_{i,j}^g h_T(u_i - u_j) \quad (2)$$

where $h_T(\cdot)$ is the truncated Huber penalty function defined in (1), $N_d(i)$ denotes the square block of size $(2r_d + 1) \times (2r_d + 1)$ focused on a pixel i , and $N_s(i)$ denotes the square block of size $(2r_s + 1) \times (2r_s + 1)$ centered at i but excluding the pixel i . $\omega_{i,j}^s$ is the Gaussian spatial kernel computed from (3), and $\omega_{i,j}^g$ represents the guidance weights, defined by (4). The first term of (2) is the data term, representing the penalty for assigning labels to pixels, which helps the model be robust to outliers in the input image. The second term is the smoothing term, representing the penalty for discontinuous neighboring pixels, which is beneficial for protecting edges. The constants a , b of the data item and the smooth item in (1) are represented by $\{a_d, b_d\}$ and $\{a_s, b_s\}$, respectively.

By applying the Gaussian spatial kernel to blur image details and reduce noise, the expression of $\omega_{i,j}^s$ is as follows:

$$\omega_{i,j}^s = \exp\left(-\frac{|i - j|^2}{2\sigma^2}\right). \quad (3)$$

The guidance weight $\omega_{i,j}^g$ is described as follows:

$$\omega_{i,j}^g = \frac{1}{|g_i - g_j|^\alpha + \delta} \quad (4)$$

where α determines the sensitivity of the edge of the guidance image g . $|\cdot|$ denotes the value in absolute terms.

Due to the use of $h_T(\cdot)$, (2) is not only nonconvex but also nonsmooth. By applying semidefinite optimization methods to solve the problem [52], [53], [54], by setting $\nabla_{i,j}^d = u_i - f_j$ and $\nabla_{i,j}^s = u_i - u_j$, the optimal conditions are as follows:

$$l_{i,j}^* = \begin{cases} 0, & |\nabla_{i,j}^*| \leq b_* \\ \nabla_{i,j}^*, & |\nabla_{i,j}^*| > b_* \end{cases} \quad (5)$$

$$\nu_{i,j}^* = \begin{cases} \frac{1}{2a_*}, & |\nabla_{i,j}^* - l_{i,j}^*| < a_* \\ \frac{1}{2|\nabla_{i,j}^* - l_{i,j}^*|}, & |\nabla_{i,j}^* - l_{i,j}^*| \geq a_* \end{cases}, * \in \{d, s\}. \quad (6)$$

The approximate solution for obtaining the final solution u^{k+1} can be derived as

$$u^{k+1} = (A^k - 2\lambda W^k)^{-1} (D^k + 2\lambda S^k) \quad (7)$$

where $W_{i,j}^k = \omega_{i,j}^s \omega_{i,j}^g (\nu_{i,j}^s)^k$ is an affinity matrix, $A_{ii}^k = \sum_{j \in N_d(i)} \omega_{i,j}^s (\nu_{i,j}^d)^k + 2\lambda \sum_{j \in N_s(i)} \omega_{i,j}^s \omega_{i,j}^g (\nu_{i,j}^s)^k$ is a diagonal matrix, $D_i^k = \sum_{j \in N_d(i)} \omega_{i,j}^s (\nu_{i,j}^d)^k (f_j + (l_{i,j}^d)^k)$, and $S_i^k = \sum_{j \in N_s(i)} \omega_{i,j}^s \omega_{i,j}^g (\nu_{i,j}^s)^k (l_{i,j}^s)^k$ are vectors.

The above optimization process iteratively executes $iter$ times, with each step monotonically decreasing the value of $E_u(u)$, theoretically ensuring its convergence, and ultimately obtaining the smoothed output u^{iter} .

DTFE uses the truncated filter with two parameter settings to construct dual-scale smoothing blocks. These two smoothing blocks are stacked together, and the structural features are obtained using KPCA to fuse two blocks. Let F_l denotes the smoothed output u^{iter} with the l th parameter setting

$$F_l = F_l(b_l^*, r_l^*), l \in \{1, 2\} \quad (8)$$

$$F = KPCA(F_1, F_2) \quad (9)$$

where b_l^* and r_l^* represent the parameter used in the l th structure extraction, and F denotes the extracted structural features. Then, F is input into SVM for classification, obtaining the probability distribution R_1 in the preprocessing stage.

B. Spatial Probability Optimization

To address the issues where traditional RW relies on user interaction and ERW cannot distinguish similar categories, resulting in poor segmentation of complex textures, this article proposes a postprocessing SPO method. The improved version of RW, subRW, is applied to optimize HSI classification results in the postprocessing stage for the first time. subRW adds auxiliary nodes to RW and initially estimates the probability of pixels belonging to each category, guiding image segmentation with reduced interaction, thereby reducing RW's misclassification of similar categories. It also uses sub-Markov transition probabilities to enhance the segmentation performance of complex textural images.

SubRW is used to model the nearby pixels' spatial correlation in SPO, enhancing the ability to distinguish complex textures

Algorithm 2: SPO Model.

Input: Reduced dimension image f , training samples TS , pre-classification probabilities obtained by SVM r

- 1: Generate seeds and labels based on TS
- 2: Define the adjacency matrix of the eight-neighbor graph structure according to (11)
- 3: Generate the GMM with five components from the seeds to obtain the probability density p_i^k
- 4: Set $c_i \leftarrow c_i^k$, where $c_i^k = p_i^k + \max_{t \neq k} p_i^t$
- 5: Calculate the transition probability matrix $\mathbf{F}$ according to (13)
- 6: Solve (14) to obtain optimization classification results $R_2 = \arg \max_{l_k} \bar{x}^{l_k}$

Output: Optimization classification results R_2

and similar categories, thereby retaining more complete spatial information. The pseudocode of SPO is shown in Algorithm 2.

Specifically, the first principal component of HSI is displayed as a weighted undirected graph $G = (V, E)$, $v \in V$, and N is the total number of nodes. v_i represents the pixel x_i in the first principal component. The multilabel seeds are defined as a set of labeled nodes $V_M = \{v_{l_1}, v_{l_2}, \dots, v_{l_K}\}$, and labels are represented as $L = \{l_1, l_2, \dots, l_K\}$, where K is the total label number. The nodes $V^{l_k} = \{v_1^{l_k}, v_2^{l_k}, \dots, v_{M_k}^{l_k}\}$ represent the set of seed nodes with label l_K , and M_k is the number of seed nodes with label l_K . An edge is defined as $e \in E \subseteq V \times V$ in the eight-connected neighborhood system, and e_{ij} connects two nodes v_i and v_j . The edge weight $w_{ij} \in \mathbf{W}$ represents the probability of an RW crossing this edge, i.e., the likelihood that adjacent pixels have the same label, and is described as follows:

$$w_{ij} = \exp \left(-\frac{\|C_i - C_j\|^2}{\varepsilon} \right) + \eta \quad (10)$$

where C_i and C_j represent the pixel color at v_i and v_j in the Lab color space, respectively, providing a numerical measure of the similarity between adjacent pixels. $\varepsilon = 1/60$ is a regularization parameter, and η is a small parameter set to 10^{-6} .

Assume each node has an intensity distribution S_k from the label l_k , where p_i^k represents the probability density of belonging to S_k at node v_i . Using GMM as the prior model, each prior distribution S_k can be regarded as a GMM learned from seed nodes labeled as l_k . With these previous distributions, a collection of prior auxiliary nodes $S_M = \{s_1, s_2, \dots, s_K\}$ is added to form an extended graph G_e ; these prior nodes are connected to all nodes, representing global label information, which helps establish connections between nodes that are not directly linked, thereby reducing the influence of seed nodes on the segmentation results. G_e is shown in Fig. 3.

When RW reaches the node Δ , the walk ends, meaning that an effective RW cannot reach it, and the corresponding probability is ignored. Stopping nodes Z , connected to labeled seed nodes, enable RW to make more effective use of prior information for segmentation, and when RW reaches a stopping node, it indicates that RW has arrived at a definite labeled region. Prior nodes S provide global label information, which helps RW to propagate

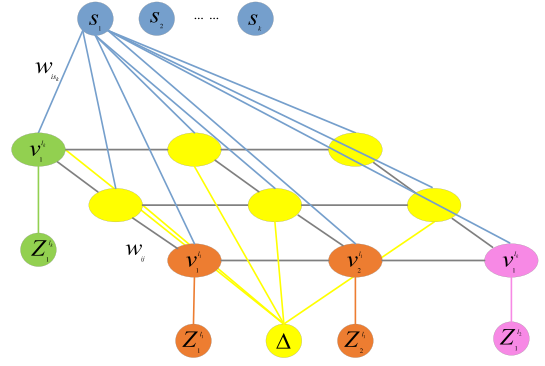


Fig. 3. Node graph G_e of SPO. Circular nodes are newly added auxiliary nodes, blue circles represent prior nodes, yellow ellipses represent nonseed nodes, and the remaining nodes are seed nodes. All prior nodes should be linked to the initial elliptical nodes. For simplicity, only the edges of one node are shown.

labels on a global scale, improving the accuracy of segmentation. Upon reaching the stopping node Z or the prior node S , the walk ends, and the corresponding label probability is obtained. The edge weight w_{is_k} between s_k and $v_i \in V$ is proportional to the probability density p_i^k . Set the weight w_{is_k} as follows:

$$w_{is_k} = (1 - g_i) \nu p_i^k \quad (11)$$

where g_i represents the leaving probability of a node in V , with its value empirically set according to [55]. ν is a regularization parameter used to measure the significance of the prior distribution.

The probability density p_i^k of the intensity distribution at v_i belonging to l_k is

$$\bar{p}_i^k = \frac{\lambda p_i^k}{d_i + \lambda c_i}. \quad (12)$$

The transition probability f_{ij} in V can be expressed as

$$f_{ij} = \frac{w_{ij}}{d_i + \lambda c_i} \quad (13)$$

where $d_i = \sum_{j \sim i} w_{ij}$ is the degree of node v_i , $j \sim i$ represents v_j in the neighborhood of v_i , and $c_i = \sum_{k=1}^K p_i^k$.

So the objective function for postprocessing optimization can be defined as

$$E_{Total}^{l_k}(x^{l_k}) = E_{Spatial}^{l_k}(x^{l_k}) + \mu E_{Prior}^{l_k}(x^{l_k}) \quad (14)$$

where μ controls the weight of the prior part in the energy function ($\mu = 1e - 5$), and $E_{Spatial}^{l_k}$ is the energy function that predicts class probabilities based on the spatial information of adjacent pixels in HSI. The spatial dependencies between neighboring pixels are represented through RW walks around nodes. $E_{Spatial}^{l_k}$ is expressed as follows:

$$E_{Spatial}^{l_k}(x^{l_k}) = (\mathbf{I} - \mathbf{D}_g) \mathbf{F} x^{l_k} + \frac{1}{Z_k} (\mathbf{I} - \mathbf{D}_g) \bar{\mathbf{p}}^k + \frac{1}{Z_k M_k} \mathbf{D}_g \mathbf{b}^{l_k} \quad (15)$$

where $\mathbf{I}$ is the identity matrix of size $N \times N$, $\mathbf{D}_g = \text{diag}(g_1, g_2, \dots, g_N)$ is a diagonal matrix, the transition probability matrix is $\mathbf{F} = [f_{ij}]_{N \times N}$, and the probability matrix is

$\bar{\mathbf{p}}^k = [\bar{p}_i^k]_{N \times 1}$. $\mathbf{b}^{l_k} = [b^{l_k}]_{N \times 1}$ is an indicator vector. If $v_i = v_m^{l_k}$, $b^{l_k} = 1$; otherwise, $b^{l_k} = 0$. Z_k is a normalization constant.

From the above-mentioned formula, the probability x^{l_k} that any unmarked node to be segmented reaches a stopping node or a priori node labeled as l_k can be calculated. Finally, each unmarked node is given to the category with the greatest probability, achieving image segmentation and spatial information preservation.

$E_{\text{Prior}}^{l_k}$ is the energy function used to predict class probabilities with the initial probabilities obtained from SVM as the prior information. After obtaining the initial probability $r_i^{l_k}$ for each pixel in the image, the probability that each pixel belongs to label l_k is calculated on the basis of simple Bayesian probability theory

$$\left(\sum_{k=1}^K \wedge^{l_k} \right) x^{l_k} = r^{l_k} \quad (16)$$

where $\wedge^{l_k}$ is a diagonal matrix composed of r^{l_k} , i.e., $\text{diag}(r^{l_k})$. The above-mentioned equation can be represented as the minimum energy distribution of the following objective function:

$$E_{\text{Prior}}^{l_k}(x^{l_k}) = \sum_{m=1, m \neq l_k}^K x^{mT} \wedge^m x^m + (x^{l_k} - 1)^T \wedge^{l_k} (x^{l_k} - 1). \quad (17)$$

The solution method for $E_{\text{Prior}}^{l_k}(x^{l_k})$ is the same as the traditional RW, with detailed solving steps provided in [36]. After obtaining $E_{\text{Prior}}^{l_k}(x^{l_k})$ and $E_{\text{Spatial}}^{l_k}(x^{l_k})$, (15) can be solved, with specific solving methods detailed in [48]. By selecting the maximum probability, the refined postprocessing optimized classification probabilities R_2 can be obtained.

C. Decision Fusion

After obtaining classification probabilities from two stages, decision fusion is employed to combine the probabilities obtained from each stage, ensuring that the class probability of the same target exceeds that of other classes, thus achieving more accurate recognition results. For each pixel, its final classification result is determined by selecting the class with the highest probability

$$R = \arg \max_i \{ \alpha R_1^i + (1 - \alpha) R_2^i \} \quad (18)$$

where α is a free parameter, R_1^i is the classification probability after feature extraction by DTFE, R_2^i is the classification probability optimized by SPO, and R is the final classification result.

III. EXPERIMENTAL DETAILS

Four hyperspectral datasets are used to demonstrate the effect of the proposed DSCSM in HSI classification, and eight commonly used and state-of-the-art classification approaches are compared. In addition, ablation experiments are conducted to confirm the effectiveness of combining the two proposed modules. All experiments are carried out using a simulation platform with an Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz, 24

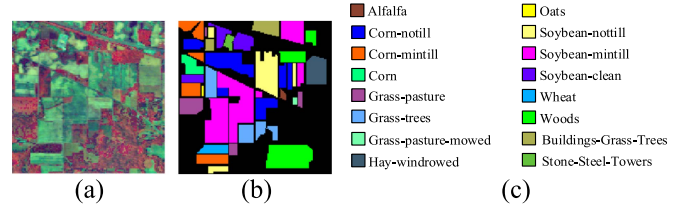


Fig. 4. Indian pines dataset. (a) False color image. (b) Ground truth. (c) Legends.

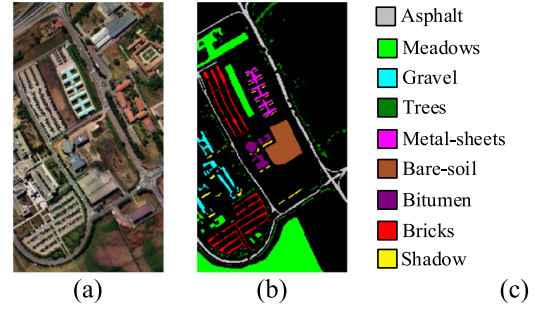


Fig. 5. Pavia University dataset. (a) False color image. (b) Ground truth. (c) Legends.

GB RAM, and MATLAB R2021a running on the Windows 11 operating system. In this section, the used datasets are first introduced. Then, the experimental parameter settings are presented, including evaluation metrics and comparison methods. Next, we discuss the parameters of the proposed method. Finally, we analyze the experimental results.

A. Datasets

Experiments are conducted using six datasets. Four classical hyperspectral datasets from different imaging scenes, Indian Pines, Pavia University, Houston 2013, and KSC, are used to verify the generality of the proposed DSCSM. In addition, the WHU-Hi-LongKou and WHU-Hi-HongHu datasets [56], [57], which are used for fine crop classification, are utilized to validate the practicality and generalizability capability of the proposed method. Ten samples from each class of each dataset are randomly selected as the training set, and the remaining samples are used as test sets. A brief description of the four datasets is given in Table I, and the false-color image, ground truth, and legends are shown in Figs. 4–9.

B. Experimental Parameter Setting

To evaluate the classification performance, DSCSM is compared with eight advanced HSI classification methods, and four quantitative evaluation metrics are used in experiments. Specifically, the details are as follows.

1) *Quantitative Evaluation Metrics*: OA, AA, Kappa, and ICA are used to comprehensively measure classification ability. OA is the correct classification rate of all samples, AA is the average accuracy for each class, and Kappa measures the consistency between predicted results and actual results. All experiments are repeated ten times to obtain the mean and standard deviation of

TABLE I
DATASET RELATED DESCRIPTION

Information	Dataset					
	Indian Pines	Pavia University	Houston 2013	KSC	WHU-Hi-LongKou	WHU-Hi-HongHu
Size (pixels)	145 × 145	610340	349×1905	512×614	550×400	940×475
Bands	220	103	144	176	270	270
Spatial resolution (m)	20	1.3	2.5	18	0.463	0.043
Spectral Wavelength (um)	0.40-2.50	0.43-0.86	0.38-1.05	0.40-2.50	0.40-1.00	0.40-1.00
Sensor	AVIRIS	ROSIS	CASI-1500	AVIRIS	Nano-Hyperspec	Nano-Hyperspec
Class	16	9	15	13	9	22

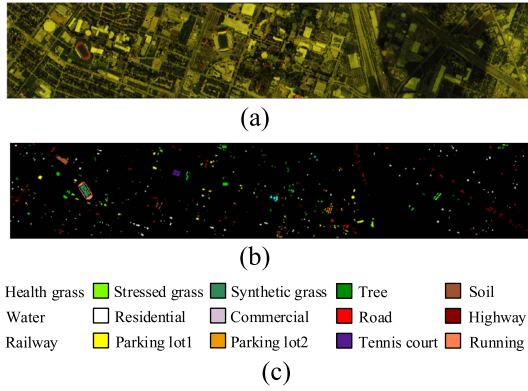


Fig. 6. Houston 2013 dataset. (a) False color image. (b) Ground truth. (c) Legends.

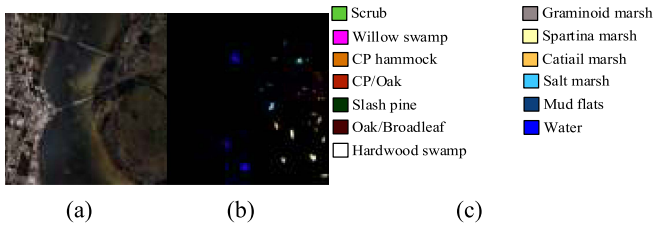


Fig. 7. KSC dataset. (a) False color image. (b) Ground truth. (c) Legends.

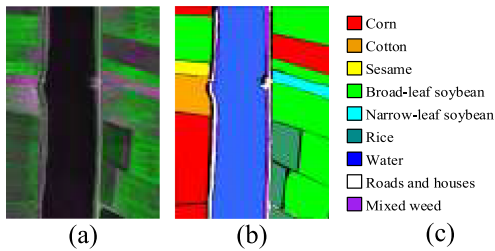


Fig. 8. WHU-Hi-LongKou dataset. (a) False color image. (b) Ground truth. (c) Legends.

ICA, OA, AA, and Kappa. The average values of these four evaluation metrics are used as the final results.

2) *Comparison Methods*: Eight advanced HSI classification techniques are compared with DSCSM, which are briefly detailed as follows.

SVM [58] uses the LIBSVM library and a radial basis function (RBF) kernel for classification. Fivefold cross validation is used

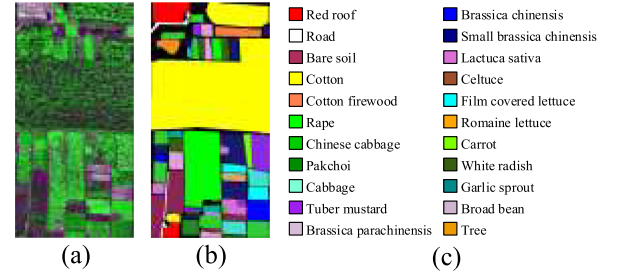


Fig. 9. WHU-Hi-HongHu dataset. (a) False color image. (b) Ground truth. (c) Legends.

to determine the optimal parameters through, with the RBF kernel width γ and the penalty factor c varying between $2^{-4} \sim 2^4$ and $10^{-2} \sim 10^4$.

RPNet [59] directly extracts random patches from images as convolution kernels and combines shallow and deep convolution features without any training.

GTR [60] extends multivariate label ridge regression to the tensor version and further enhances the model's classification ability using prior knowledge from different modalities.

FDSI [45] constructs the structural contour of HSI using an adaptive texture smoothing approach during the preprocessing step. In the postprocessing stage, it employs spatial optimization technology based on the ERW, effectively constructing dual spatial information.

Two-Staged [46] is based on the concept of total variation (TV). It combines the anisotropic TV model with different regularization parameter settings and the isotropic TV model based on the split Bregman algorithm to perform two-stage feature extraction and reclassification.

CSSM [47] is a complementary spectral-spatial method. The first-stage probability map is obtained by extracting and smoothing the geometric features of HSI using an ACTVSP. After that, the EPF is used to optimize the map in order to obtain the second-stage probability map.

PSAE-DMLSR [61] uses marginal principal component analysis (MP) and principal spatial local marginal principal component analysis to identify the best spectral subspace and optimal representation space (ORS) to obtain global spectral-spatial joint feature information. It then uses the discriminative

marginal least squares regression (DMLSR) method for HSI classification.

M^3FuNet [62] is an unsupervised multivariate feature fusion network. It uses multiscale super-vector matrix correction and multiscale random convolution dispersion as spectral and spatial feature extraction methods. After multivariate feature fusion, it uses M^3FuNet to achieve HSI classification.

$SPFormer$ [63] is a lightweight self-pooling transformer. It uses the multihead self-pooling channel shuffling module based on sparse mapping to extract distinctive features, and then the central token mixer is used to promote the information interaction between pixels. Finally, it uses the fully connected layer to realize the HSI classification in the case of small samples.

The parameter values for all comparison techniques are consistent with the default parameter settings specified in the related references. See the references for further information on the specific settings and explanations.

C. Parameter Analysis

In the proposed method, four fixed parameters are required: the dual-scale parameters $M(r1, b1, r2, b2)$, the average dimensionality reduction number D , the kernel principal component number K , and the fusion weight α . We categorize these parameters into two groups: dual-scale truncated filtering parameters and framework parameters. We conduct experiments on the Indian Pines dataset, initializing the aforementioned parameters and discussing them in groups as follows.

1) *Parameter Setting of Dual-Scale Truncated Filter*: In the dual-scale parameters $M(r1, b1, r2, b2)$, r represents the neighborhood radius of the data term and smoothing term, and b is the parameter controlling the truncated function. The subscripts 1 and 2 represent the values at two different scales. When discussing $r1$, the other parameters are fixed at $b1 = 0.6$, $r2 = 5$, $b2 = 0.1$. Similarly, when discussing the influence of $b1$, the other parameters are fixed at $r1 = 1$, $r2 = 5$, $b2 = 0.1$. Fig. 10 shows the impact of these parameters on classification performance. All optimal parameters are selected based on the highest OA.

It is evident that the suggested strategy provides the best classification accuracy when $r1 = 1$, $r2 = 5$, $b1 = 0.6$, $b2 = 0.1$, so these settings are set as the default values for the dual-scale parameters M .

2) *Parameter Setting of Framework*: The framework parameters include the average dimensionality reduction number D , the number of kernel principal components K , and the fusion weight α . Based on the obtained optimal $M(r1, b1, r2, b2)$, the framework parameters are initialized. While analyzing D and K , other parameters are fixed, with M and α set to (1, 0.6, 5, 0.1) and 0.7, respectively. Fig. 11 shows the impact of different K and D values on OA, AA, and Kappa.

As seen in the figure, when K and D with tiny values, the tendency for classification performance declines. This is because retaining fewer principal components during dimensionality reduction can cause the loss of important discriminating information. When K and D are set to 35 and 50, respectively, OA,

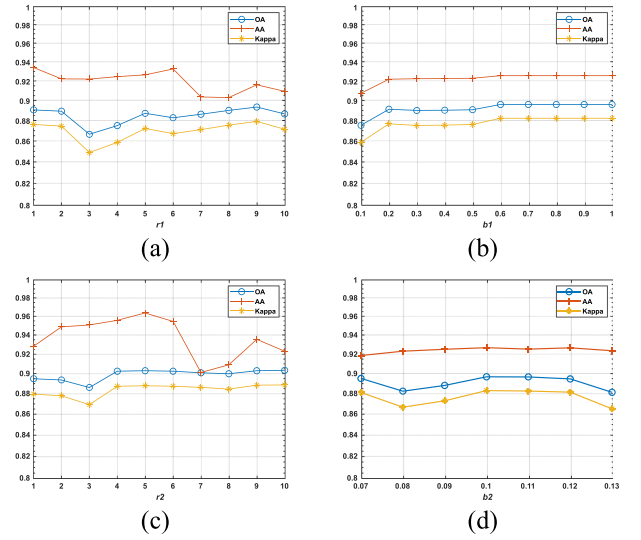


Fig. 10. Effect of dual-scale parameters M (a) $r1$. (b) $b1$. (c) $r2$. (d) $b2$ on classification performance.

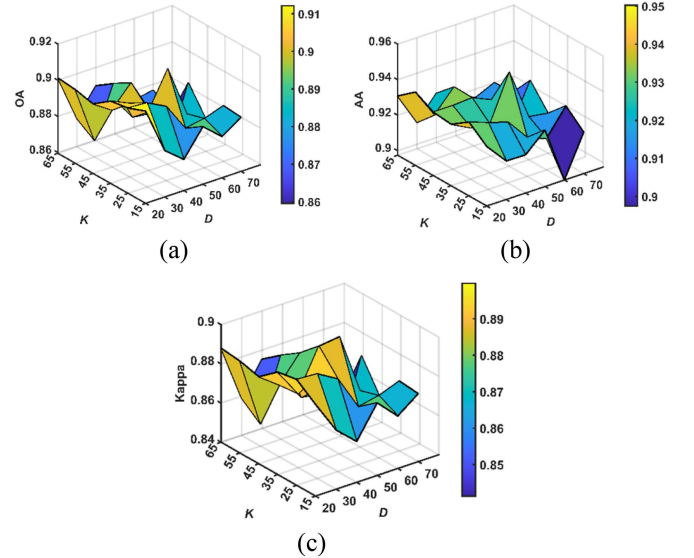


Fig. 11. Effect of parameters K and D on classification performance. (a) OA. (b) AA. (c) Kappa.

AA, and Kappa reach their best values. As K and D increase, classification accuracy gradually stabilizes. Therefore, these two parameters are initialized to $K = 35$ and $D = 50$.

In addition, based on the optimal values of K and D obtained previously, the value of α is initialized. Fig. 12 shows the effect of the parameter α on classification performance.

As seen in the figure, when the fusion weight α is set to 1 or 0, indicating that only the DTFE or SPO module is used for the classification task, the classification performance is not as good as when both modules are combined. The best results for the three evaluation metrics are achieved when $\alpha = 0.7$. Therefore, 0.7 is set as the default parameter for α .

In subsequent experiments, it was verified that the default parameters obtained above achieve good classification performance on different datasets. Therefore, in practice, there is no

TABLE II
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON INDIAN PINES DATASET (%)

Class	Train	Test	Baseline			State-of-the-art						Proposed		
			SVM	RPNNet	GTR	FDSI	Two- Staged	CSSM	PSAE- DMLSR	M ³ FuNet	SPFormer	DTFE	SPO	DSCSM
1	10	36	18.18	97.22	100.00	100.00	96.94	77.78	82.61	97.22	97.22	100.00	100.00	100.00
2	10	1418	36.33	54.72	64.74	75.47	80.28	68.40	75.24	73.84	70.80	82.19	85.54	99.58
3	10	820	28.16	70.61	84.39	97.37	83.35	69.12	73.02	87.20	59.27	83.72	98.58	68.39
4	10	227	23.30	82.82	99.12	86.82	97.53	88.84	89.19	100.00	98.68	94.15	100.00	89.02
5	10	473	69.35	72.94	89.22	91.60	88.27	95.62	95.76	91.75	83.30	81.00	85.20	99.78
6	10	720	81.43	94.86	96.67	99.43	97.04	98.44	96.23	92.36	99.86	100.00	98.07	100.00
7	10	18	16.35	88.89	100.00	54.55	100.00	75.00	100.00	100.00	100.00	100.00	94.74	100.00
8	10	468	97.82	95.73	99.36	100.00	100.00	100.00	100.00	99.79	100.00	100.00	100.00	100.00
9	10	10	2.63	100.00	100.00	71.43	100.00	83.33	80.00	100.00	100.00	83.33	100.00	100.00
10	10	962	36.15	71.41	76.61	76.38	84.78	71.26	82.51	85.14	85.34	80.27	89.43	80.38
11	10	2445	59.21	71.37	63.60	98.51	4.68	78.66	67.94	92.80	80.37	96.95	93.25	91.35
12	10	583	18.60	77.36	89.88	79.01	92.25	96.18	90.59	92.62	91.25	54.85	83.14	99.80
13	10	195	71.01	97.95	100.00	97.50	99.69	99.48	100.00	100.00	100.00	100.00	41.14	100.00
14	10	1255	87.44	78.65	95.54	97.34	99.90	98.64	88.92	91.39	99.04	99.92	95.94	95.95
15	10	376	35.22	83.78	94.15	90.56	96.76	94.21	88.71	99.20	99.47	100.00	100.00	100.00
16	10	83	76.19	95.18	90.36	67.23	97.71	98.81	87.23	89.16	100.00	97.59	97.62	96.51
OA			45.15	74.69	80.49	89.37	89.28	83.63	79.94	89.55	84.96	88.06	89.78	91.22
AA			47.34	83.34	90.23	86.45	93.70	87.11	87.37	93.28	91.54	90.87	91.42	95.03
Kappa			39.50	71.37	78.09	87.89	87.81	81.26	76.67	88.08	82.98	86.47	88.33	89.97

The best results are highlighted in bold.

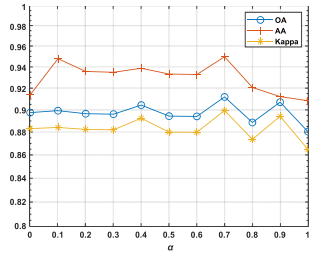


Fig. 12. Effect of parameter α on classification performance.

need to spend a significant amount of effort selecting parameters, using a fixed set of parameters that are applicable globally can yield satisfactory results. Since the scale of DTFE should vary according to different datasets and different resolutions, in order to achieve the best results, how to design an automatic method for selecting the optimal parameters will be a topic for further research in the future.

D. Quantitative Experimental Results

1) *Experimental Results of Indian Pines*: The best results are highlighted in bold.

The experiments are first conducted on the Indian Pines. The ICA, OA, AA, and Kappa obtained by different classification methods are given in Table II. It can be seen from Table II that the proposed method achieved the highest values of OA, AA, and Kappa and the highest ICA in ten categories, indicating the best classification performance on the Indian Pines. The classification maps of all methods are displayed in Fig. 13.

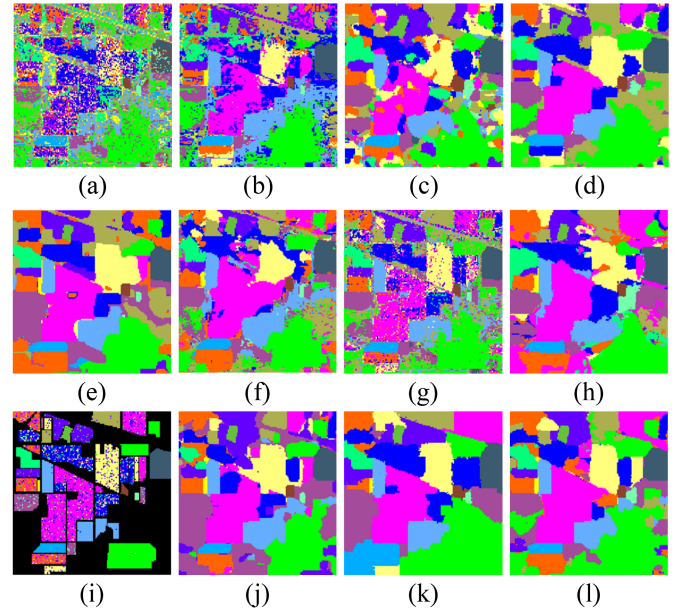


Fig. 13. Classification maps of different methods on Indian Pines. (a) SVM. (b) RPNNet. (c) GTR. (d) FDSI. (e) Two-Staged. (f) CSSM. (g) PSAE-DMLSR. (h) M³FuNet. (i) SPFormer. (j) DTFE. (k) SPO. (l) DSCSM.

As shown in Fig. 13, the map produced by SVM has many misclassifications, mainly because it only utilizes the spectral information of HSI, lacking consideration of spatial information. RPNNet, which takes spatial information into account, slightly improves misclassification compared to the SVM. However, the convolutional features of RPNNet are random, leading to significant variations in accuracy, resulting in a lower average accuracy

TABLE III
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON PAVIA UNIVERSITY DATASET (%)

Class	Train	Test	Baseline			State-of-the-art						Proposed		
			SVM	RNet	GTR	FDSI	Two- Staged	CSSM	PSAE- DMLSR	M ³ FuNet	SPFormer	DTFE	SPO	DSCSM
1	10	6621	92.41	74.63	52.86	97.66	80.44	98.42	64.63	63.75	92.46	98.13	72.58	98.31
2	10	18639	87.13	65.93	83.80	98.32	93.73	99.83	61.62	95.18	82.14	95.99	99.08	96.48
3	10	2089	37.93	81.62	77.21	99.24	95.43	42.25	85.56	98.66	94.26	99.95	84.29	100.00
4	10	3054	72.03	96.81	90.18	82.16	87.91	63.75	66.19	74.72	91.16	93.41	53.79	97.61
5	10	1335	97.63	100.00	99.10	98.82	99.40	93.68	96.47	93.93	100.00	96.32	87.22	96.36
6	10	5019	30.04	68.32	74.08	99.25	99.97	79.83	83.75	99.80	100.00	98.77	98.86	98.84
7	10	1320	46.38	95.38	98.64	100.00	98.82	78.56	97.34	97.42	100.00	99.62	89.43	96.35
8	10	3672	77.49	80.69	51.20	88.02	93.14	89.90	88.10	93.68	87.01	88.09	69.28	88.25
9	10	937	99.89	84.10	97.01	99.78	91.14	99.77	64.70	59.23	99.89	97.16	68.57	98.92
OA			64.31	74.24	76.41	96.34	92.30	84.37	70.87	88.67	89.00	96.14	86.47	96.44
AA			71.21	83.14	80.45	95.92	93.33	82.89	78.89	86.27	94.10	96.52	80.34	96.57
Kappa			56.25	67.99	69.70	95.15	89.90	80.20	64.21	85.18	85.95	94.85	82.31	95.25

The best results are highlighted in bold.

and many misclassified regions in the classification map. GTR treats pixels and their neighboring pixels as three-dimensional tensors, effectively preserving the inherent structural information of the HSI spatial domain and making its classification map relatively smooth. However, the spectral curves of related crops are not well fitted by the tensor regression model, resulting in low discrimination, such as in the Corn-notill and Corn-mintill regions.

FDSI, Two-Staged, and CSSM combine preprocessing and postprocessing by integrating spatial-spectral information, resulting in significantly improved classification performance. However, there are still many misclassified pixels due to the inadequate preservation of important spatial structures. PSAE-DMLSR, M³FuNet, and SPFormer are recently proposed spatial-spectral joint classification methods. PSAE-DMLSR can find the ORS and mitigate the impact of interfering pixels but it relies on a sufficient number of samples, resulting in negative classification performance on Indian Pines and the classification map with a lot of noise. M³FuNet addresses the issue of single-class feature antagonism, extracting features with strong retention and strong spectral-spatial dependencies. Although the classification map is smoother, the network structure is complex with too many parameters, making it prone to overfitting and neglecting small samples, resulting in misclassification. Although SPFormer is proposed for small sample problems, it does not make full use of the spatial information of HSI, resulting in a poor classification effect.

In addition, analyzing the classification results of the two suggested modules, DTFE and SPO, on Indian Pines, it is noted that because of the small number of samples per class, DTFE may remove these small sample classes. At this point, SPO is needed to clarify boundary information and optimize classification probabilities. Therefore, DSCSM achieves the best classification accuracy and visualization effects, validating the effectiveness of DSCSM.

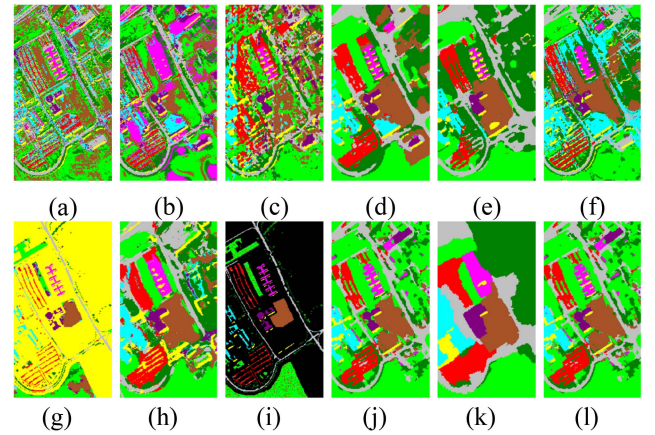


Fig. 14. Classification maps of different methods on Pavia University. (a) SVM. (b) RNet. (c) GTR. (d) FDSI. (e) Two-Staged. (f) CSSM. (g) PSAE-DMLSR. (h) M³FuNet. (i) SPFormer. (j) DTFE. (k) SPO. (l) DSCSM.

2) *Experimental Results of Pavia University:* The ICA, OA, AA, and Kappa obtained by different classification methods on the Pavia University dataset are given in Table III.

As given in Table III, DSCSM is superior to other comparative methods in terms of OA, AA, and Kappa. Specifically, for similar classes of Asphalt, Metal-sheets, and Bricks, DSCSM shows a superior recognition performance, indicating its effectiveness in leveraging spectral-spatial information to distinguish between similar categories. Although DSCSM has a lower number compared to SPFormer of the highest CAs, its CAs are at the forefront, and it exhibits high accuracy and stability in the overall classification task. Fig. 14 shows the classification maps of all methods.

It can be seen that most methods produce a lot of salt-and-pepper noise. While the classification map generated by the proposed SPO is overly smooth, the integration of the DTFE enhances the accuracy of the DSCSM classification results,

TABLE IV
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON HOUSTON 2013 DATASET (%)

Class	Train	Test	Baseline			State-of-the-art						Proposed		
			SVM	RPNet	GTR	FDSI	Two- Staged	CSSM	PSAE- DMLSR	M ³ FuNet	SPFormer	DTFE	SPO	DSCSM
1	10	1063	91.55	72.93	82.51	75.08	59.87	90.85	85.99	84.29	94.46	86.35	63.65	77.89
2	10	800	87.87	68.57	97.91	71.74	71.62	82.57	99.09	78.78	85.59	86.92	57.45	75.56
3	10	6777	99.71	100.00	99.71	100.00	97.96	100.00	99.45	90.98	99.42	100.00	100.00	100.00
4	10	1043	97.12	73.58	99.92	77.34	60.62	93.88	99.54	70.66	90.07	97.44	60.92	98.13
5	10	1232	88.57	94.24	100.00	100.00	86.36	94.48	97.53	89.45	98.71	99.84	100.00	100.00
6	10	315	73.24	96.83	98.10	91.50	75.56	95.70	99.43	96.19	88.44	93.64	84.15	90.09
7	10	968	78.68	80.37	73.77	90.40	64.94	82.59	26.40	70.51	85.43	89.58	98.20	91.55
8	10	614	70.31	54.70	52.11	68.50	85.41	48.35	90.04	69.37	38.50	68.86	49.21	71.10
9	10	1021	54.62	68.12	50.64	75.63	57.81	69.37	29.39	60.06	83.16	83.38	66.77	84.85
10	10	372	36.71	78.88	79.13	83.04	97.53	61.52	88.67	74.12	64.98	96.53	84.46	97.97
11	10	104	44.40	72.24	72.57	92.80	86.37	68.60	86.91	72.65	86.67	99.82	99.71	100.00
12	10	1223	56.64	59.12	60.67	82.98	68.11	66.60	87.17	62.55	55.54	90.41	89.78	90.66
13	10	439	32.72	76.25	91.50	58.94	93.90	52.52	72.41	77.56	90.73	58.40	70.63	62.28
14	10	418	88.52	95.45	100.00	92.27	99.76	96.76	99.28	99.52	100.00	100.00	95.43	100.00
15	10	650	99.06	92.46	100.00	100.00	100.00	99.84	99.61	94.31	100.00	100.00	100.00	100.00
OA			73.33	75.74	80.54	82.73	77.44	77.35	80.11	76.34	81.50	88.72	76.60	88.04
AA			73.32	78.92	83.90	84.01	80.39	80.24	84.06	79.40	84.11	90.08	81.36	89.34
Kappa			71.15	73.80	78.97	81.35	75.64	75.53	78.48	74.43	80.05	87.82	74.72	87.09

The best results are highlighted in bold.

particularly reducing misclassifications in the Bare-soil and Meadows areas. It indicates that DTFE and SPO provide complementary information, resulting in the best classification outcomes.

3) *Experimental Results of Houston 2013*: To validate the performance of DSCSM in classifying complex urban areas, experiments are conducted on the challenging Houston 2013 dataset. Since Houston 2013 is small, the region is cropped and scaled to better analyze the results. Table IV lists the ICA, OA, AA, and Kappa of all methods.

Compared with the other nine competing methods, DSCSM obtains the highest OA, AA, and Kappa. Among its two modules, DTFE exhibits better classification performance. The reason is that the Houston 2013 dataset has a cloud-covered area, and the class distribution is relatively scattered, posing a challenge for SPO to obtain edge contours through random walks. RW fails to sample distant pixels of the same class, resulting in lower accuracy for SPO. Consequently, the combined DSCSM accuracy decreases. However, considering the excellent performance of DSCSM in other types of scenarios, the 0.68% reduction in OA on Houston 2013 is acceptable. This also implies that the DTFE method performs well on details such as small-sized targets. In addition, the table illustrates that DSCSM has good recognition performance for elongated class distributions, such as Road, Highway, and Railway.

Fig. 15 displays the classification results achieved using all approaches. Because the Houston 2013 image is small, the region is cropped and scaled to better analyze the results.

As can be seen from the figure, SVM still suffers from severe salt-and-pepper noise. The performance of RPNet and PSAE-DMLSR is unsatisfactory on large-scale images with limited training samples, resulting in classification maps with substantial noise. GTR, CSSM, FDSI, and SPFormer also exhibit varying degrees of mislabeling. Two-Staged performs two stages of smoothing, leading to oversmoothing in the resulting classification map. M³FuNet also has many misclassifications, especially in the Commercial and Railway categories, indicating that unsupervised methods may not be well-suited for such challenging datasets. Therefore, DSCSM is the most robust compared to the methods for comparison and performs well in the classification of complex urban areas.

4) *Experimental Results of KSC*: KSC features a scattered distribution of classes, providing image details that are neither too coarse nor too fine, which can verify the effect of the proposed method on moderately challenging tasks. The ICA, OA, AA, and Kappa are given in Table V.

It can be seen from the table that the OA, AA, and Kappa of DSCSM are all above 99% , indicating that DSCSM is a very feasible classification method on KSC and shows superior performance. Moreover, the proposed DTFE and SPO can also achieve 98% accuracy in the case of a single stage, significantly the highest in all comparison methods, indicating that they can extract features well and retain important details. In terms of CA, most CAs of DSCSM achieve 100% classification accuracy. Particularly for the Graminoid marsh, which has a scattered distribution and limited spectral separability, only DSCSM achieves

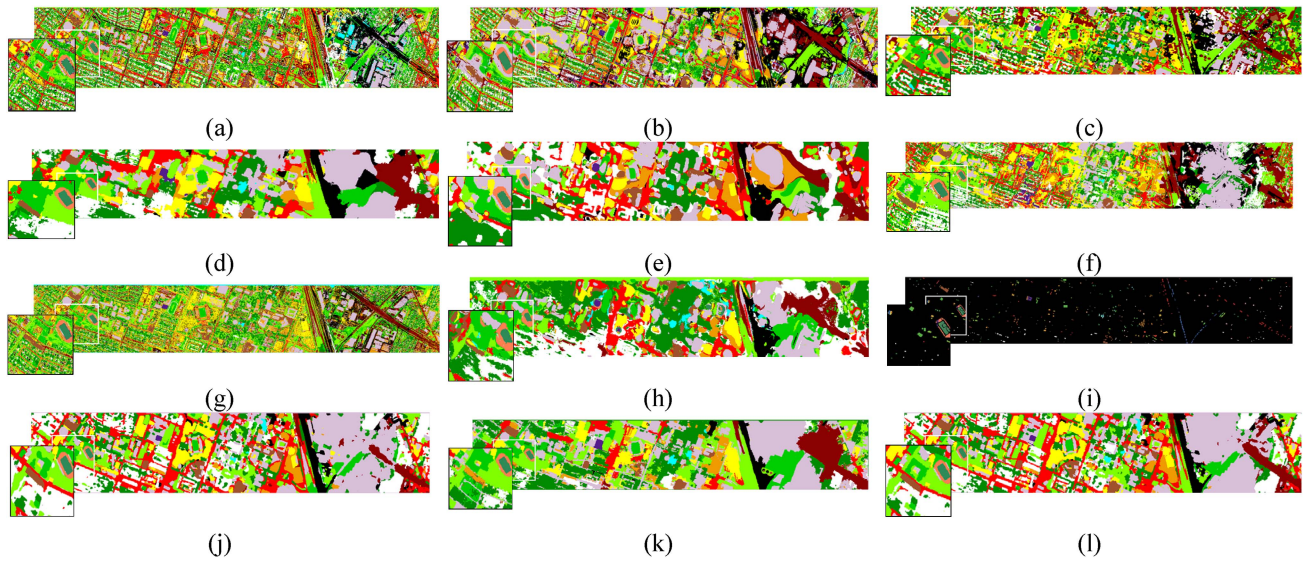


Fig. 15. Classification maps of different methods on Houston 2013. (a) SVM. (b) RpNet. (c) GTR. (d) FDSI. (e) Two-staged. (f) CSSM. (g) PSAE-DMLSR. (h) M³FuNet. (i) SPFormer. (j) DTfE. (k) SPO. (l) DSCSM.

TABLE V
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON KSC DATASET (%)

Class	Train	Test	Baseline			State-of-the-art						Proposed		
			SVM	RpNet	GTR	FDSI	Two- Staged	CSSM	PSAE- DMLSR	M³FuNet	SPFormer	DTfE	SPO	DSCSM
1	10	751	95.71	79.89	100.00	100.00	96.11	100.00	82.56	72.70	99.87	100.00	100.00	100.00
2	10	233	93.03	83.69	54.08	100.00	95.71	99.13	77.68	96.57	60.52	100.00	100.00	100.00
3	10	246	88.28	87.80	4.07	100.00	100.00	98.77	89.84	88.21	74.80	100.00	91.79	100.00
4	10	242	67.80	83.06	17.36	100.00	95.04	91.60	72.73	70.66	28.51	86.72	100.00	91.32
5	10	151	46.19	80.79	51.66	100.00	98.81	93.21	100.00	100.00	79.47	85.29	100.00	100.00
6	10	219	56.14	65.30	76.26	64.79	100.00	84.88	100.00	99.54	89.04	100.00	90.12	99.10
7	10	95	72.00	88.42	100.00	100.00	100.00	100.00	100.00	100.00	90.54	100.00	100.00	100.00
8	10	421	65.15	82.66	76.25	99.53	96.96	96.56	98.34	82.42	79.81	99.76	100.00	100.00
9	10	510	89.46	66.08	44.90	100.00	93.96	97.51	100.00	98.43	84.51	99.80	100.00	99.80
10	10	394	77.10	97.72	84.77	100.00	93.45	96.74	100.00	85.28	86.80	100.00	100.00	100.00
11	10	409	98.43	96.58	96.58	100.00	100.00	97.61	99.02	78.97	94.13	100.00	95.78	100.00
12	10	493	96.59	91.48	17.24	100.00	94.44	99.37	81.54	87.02	95.74	99.80	100.00	99.80
13	10	917	100.00	99.13	99.24	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
OA			84.77	86.32	69.73	97.62	97.05	97.46	92.60	88.11	87.15	98.88	98.74	99.45
AA			80.45	84.82	63.26	97.26	97.27	96.57	92.44	89.22	81.82	97.80	98.29	99.21
Kappa			83.06	84.78	66.17	97.35	96.72	97.17	91.77	86.82	85.67	98.75	98.60	99.39

The best results are highlighted in bold.

100% classification accuracy, demonstrating its extremely high recognition capability.

Fig. 16 displays the classification maps of all methods, from which it can be observed that FDSI, Two-Staged, DTfE, and DSCSM exhibit superior classification outcomes. In contrast, the remaining methods generate a significant amount of noise. Although the classification map produced by SPO is overly smooth, the edge information of the class it provides can be utilized to correct the results of DTfE, thereby enhancing the classification accuracy of DSCSM.

5) *Experimental Results of WHU-Hi-LongKou*: To validate the generalization performance of DSCSM in practical classification problems, experiments are conducted using the WHU-Hi-LongKou dataset for fine crop classification. The ICA, OA, AA, and Kappa are given in Table VI.

As can be seen from the table, although the categories in the WHU-Hi-LongKou are simple, there are still a few misclassifications. Particularly, with a smaller number of samples, the classification performance of RpNet and M³FuNet is

TABLE VI
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON WHU-HI-LONGKOU DATASET (%)

Class	Train	Test	Baseline			State-of-the-art						Proposed		
			SVM	RPNet	GTR	FDSI	Two- Staged	CSSM	PSAE- DMLSR	M³FuNet	SPFormer	DTFE	SPO	DSCSM
1	10	34501	93.45	98.74	98.78	99.06	100.00	95.39	97.72	89.17	99.87	99.63	97.89	99.64
2	10	8364	29.35	78.97	87.42	53.05	99.78	44.04	91.26	98.10	98.35	99.74	98.67	99.73
3	10	3021	29.44	82.39	99.44	99.70	98.25	59.56	97.64	90.37	99.97	99.96	57.47	100.00
4	10	63202	95.23	77.12	95.43	99.96	90.96	99.64	91.40	74.72	87.84	98.81	98.79	99.06
5	10	4141	15.86	80.80	98.09	91.81	100.00	51.11	94.23	99.44	99.86	99.22	41.84	99.23
6	10	11844	87.93	90.97	98.84	98.02	90.65	91.50	99.52	86.74	99.23	98.72	98.58	99.83
7	10	67046	100.00	96.60	99.69	99.63	99.99	99.99	99.87	92.98	99.88	99.97	98.96	99.98
8	10	7114	69.58	88.18	81.64	88.95	95.83	94.72	94.44	80.32	92.56	91.82	83.61	92.76
9	10	5219	43.47	88.10	84.92	78.65	61.43	79.06	94.07	78.65	96.43	68.41	70.60	74.62
OA			84.90	88.85	96.63	94.81	95.50	90.40	96.03	85.83	95.71	98.13	93.69	98.56
AA			62.70	86.88	93.81	89.87	92.99	79.45	95.57	87.83	97.11	95.14	82.94	96.09
Kappa			80.58	85.78	95.59	93.28	94.12	87.72	94.83	82.06	94.44	97.54	91.77	98.11

The best results are highlighted in bold.

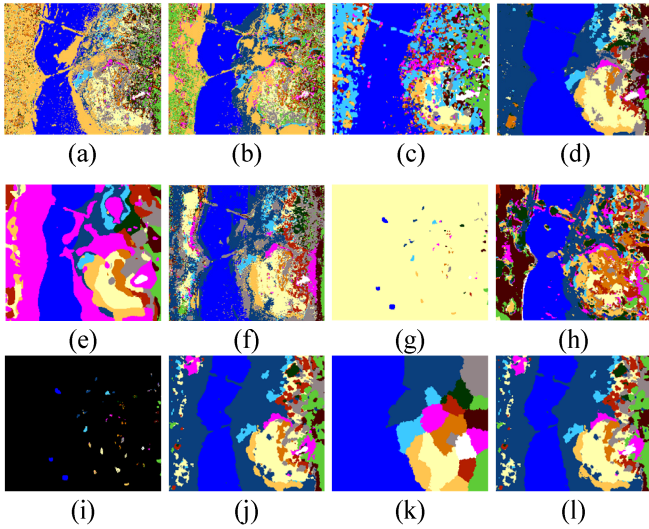


Fig. 16. Classification maps of different methods on KSC. (a) SVM. (b) RPNNet. (c) GTR. (d) FDSI. (e) Two-Stage. (f) CSSM. (g) PSAE-DMLSR. (h) M³FuNet. (i) SPFormer. (j) DTFE. (k) SPO. (l) DSCSM.

relatively poor, with OA below 90% . Due to the relatively simple scene, low spectral variability, and spatial heterogeneity, the remaining spectral-spatial joint methods achieve good performance, with OAs greater than 90% . Among all the comparison methods, DSCSM obtains the highest OA and Kappa. In terms of single-category classification accuracy, DSCSM has the highest CA for two categories, and for the remaining categories, except for Mixed weed, it ranks second highest in CA. The poor classification performance for Mixed weed is attributed to its narrow and scattered distribution, where SPO is less effective at predicting pixels of the same category that are far apart, which is also the reason why AA is lower than SPFormer.

The classification maps obtained by all methods are shown in Fig. 17. It can be seen from the figure that DSCSM has a strong ability to distinguish spectrally similar classes, such as

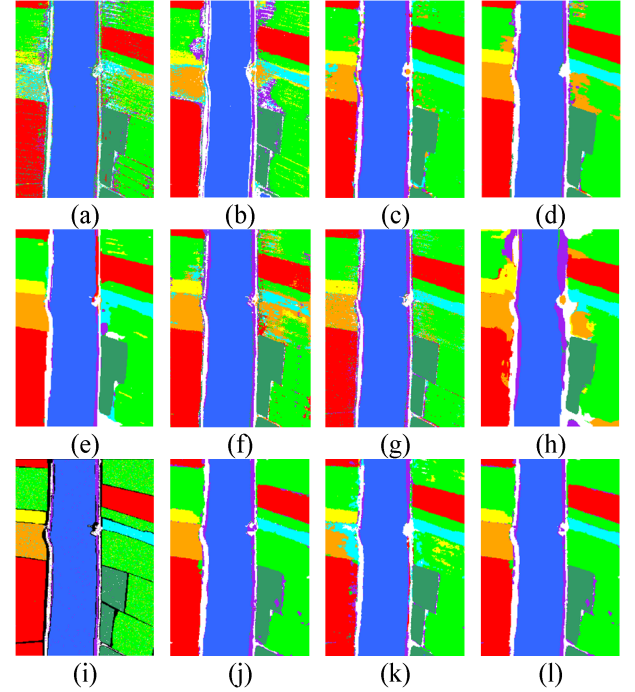


Fig. 17. Classification maps of different methods on WHU-Hi-LongKou. (a) SVM. (b) RPNNet. (c) GTR. (d) FDSI. (e) Two-Stage. (f) CSSM. (g) PSAE-DMLSR. (h) M³FuNet. (i) SPFormer. (j) DTFE. (k) SPO. (l) DSCSM.

Broad-leaf soybean and Narrow-leaf soybean, with smooth classification maps and clear boundaries. Two-Stage and M³FuNet exhibit over-smoothing, while the classification maps of other comparison methods contain more salt-and-pepper noise. Therefore, it can be inferred from the analysis above that DSCSM obtains excellent classification results on the WHU-Hi-LongKou dataset which with regular block distribution.

6) *Experimental Results of WHU-Hi-HongHu*: In addition, the WHU-Hi-HongHu dataset is fragmented and variably sized land parcels, with high spectral and spatial heterogeneity, representing a more complex agricultural scenario. In order to confirm

TABLE VII
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON WHU-HI-HONGHU DATASET (%)

Class	Train	Test	Baseline			State-of-the-art						Proposed		
			SVM	RPNNet	GTR	FDSI	Two- Staged	CSSM	PSAE- DMLSR	M ³ FuNet	SPFormer	DTFE	SPO	DSCSM
1	10	14031	76.85	83.39	73.20	98.97	96.89	97.25	86.20	87.32	93.51	99.74	90.24	99.49
2	10	3502	56.44	76.47	73.10	65.67	73.84	68.79	76.74	98.57	81.32	40.07	25.00	45.33
3	10	21811	87.14	81.34	55.94	99.47	93.57	95.19	66.67	82.22	79.19	99.87	86.45	99.87
4	10	163275	95.88	87.37	67.11	98.97	96.18	99.11	70.08	96.05	96.42	99.64	99.99	99.66
5	10	6208	22.20	96.54	71.65	44.77	99.53	38.11	66.07	96.88	84.20	84.72	15.31	84.83
6	10	44547	84.34	81.05	73.18	96.60	91.34	86.04	72.42	90.39	96.92	98.19	94.80	98.25
7	10	2493	70.52	38.92	26.01	89.58	72.20	65.07	52.69	70.97	77.33	92.49	98.43	90.27
8	10	4044	4.57	64.17	56.48	49.70	92.36	29.13	35.48	86.35	81.63	61.54	92.93	61.52
9	10	10809	99.40	72.39	79.13	99.36	93.28	99.76	89.83	94.25	94.55	99.61	86.43	99.45
10	10	12384	20.91	44.67	42.55	85.03	73.56	63.17	39.06	76.73	89.32	95.73	94.32	91.59
11	10	11005	15.72	39.44	26.66	80.58	85.02	52.93	39.44	71.79	80.50	96.67	91.55	99.43
12	10	8944	43.41	63.84	22.66	88.32	68.22	65.03	49.18	79.32	71.66	84.58	89.72	92.01
13	10	22497	54.89	56.83	28.55	72.84	80.67	44.76	42.57	68.67	68.40	67.79	89.15	72.05
14	10	7346	18.95	58.93	55.88	92.72	90.43	64.49	65.32	88.91	92.69	89.28	72.53	89.38
15	10	992	7.11	88.61	92.44	89.28	94.15	97.87	82.28	92.64	92.04	95.71	45.37	95.40
16	10	7252	69.80	62.45	72.48	98.50	92.21	97.86	74.76	96.35	92.68	99.18	87.14	98.89
17	10	3000	32.63	74.27	75.73	71.39	95.47	81.65	79.58	94.30	96.43	98.15	97.09	99.14
18	10	3207	12.63	68.88	89.02	90.84	99.91	83.09	66.25	83.19	99.28	84.11	51.26	83.92
19	10	8702	44.40	36.96	67.82	64.08	79.00	62.04	70.38	81.46	84.06	50.54	29.25	61.29
20	10	3476	14.00	80.47	44.65	87.78	99.40	60.10	74.70	88.81	97.99	98.93	55.57	97.30
21	10	1318	8.84	95.90	86.72	36.72	100.00	80.98	62.48	98.18	97.57	83.84	31.52	79.06
22	10	4030	12.45	66.67	71.66	43.88	100.00	54.24	65.40	99.23	92.63	86.05	74.75	81.20
OA			58.55	74.82	60.11	88.74	90.67	79.77	65.95	88.72	90.35	91.38	77.53	92.54
AA			43.32	69.07	61.48	79.32	89.42	72.12	64.89	87.39	88.20	86.67	72.67	87.24
Kappa			51.11	69.13	52.81	85.96	88.32	75.29	59.60	85.87	87.92	89.17	73.27	90.62

The best results are highlighted in bold.

the efficacy of the suggested approach in intricate real-world scenarios, the WHU-Hi-HongHu dataset is used for the experiment. Table VII displays the ICA, OA, AA, and Kappa of the experiment.

Among the comparison methods, all except Two-Staged and SPFormer achieve classification accuracies below 90%. This indicates that many classification methods do not perform well on the challenging WHU-Hi-HongHu dataset. However, the proposed DSCSM outperforms the best-performing Two-Staged by 1.87% in OA, achieving the best accuracy of 92.54%. This demonstrates that DSCSM is a feasible classification method for the WHU-Hi-HongHu dataset. Fig. 18 displays the classification maps for each approach.

From the figure, it can be seen that most methods exhibit many isolated classification error regions, resulting in noisy classification maps, and the classification map of SPO appears overly smooth. The classification maps of Two-Staged, SPFormer, DTFE, and DSCSM show better performance.

However, there are still some erroneous classification regions, such as similar crop types, Cotton and Cotton firewood, Brassica parachinensis, and Brassica chinensis, where misclassification occurs. The classification map of DSCSM achieves better classification in both the rapeseed and cotton areas, particularly for Brassica parachinensis and Brassica chinensis, indicating that

DSCSM can address high spatial-spectral heterogeneity issues. Overall, it is evident that DSCSM achieves the best performance with regard to visual effects and classification accuracy.

E. Validation of Sample Sensitivity

To compare the dependence of DSCSM and the comparison methods on the number of samples, the influence on classification performance is analyzed by randomly selecting different numbers of training samples for each class. Since the GMM in SPO has five components, the number of samples per class is at least 5. When 1%, 5%, and 10% of samples per class are selected as training samples, the number of training samples for each class is uniformly set to 5 if the sample size is less than 5. The total number of training samples for each dataset is shown in Fig. 19. All methods use the same sample number settings.

Fig. 20 shows the impact of the number of training samples on classification accuracy (OA). As illustrated in the figure, all approaches get greater precision as the size of the training sample rises, with DSCSM achieving the best results compared to other methods. When given 10% training samples per class, DSCSM can achieve over 99% accuracy across four datasets. Even when the sample size becomes small, such as randomly selecting 5 or 10 samples per class or selecting 1% and 5% of the samples per class as training samples, DSCSM still demonstrates excellent

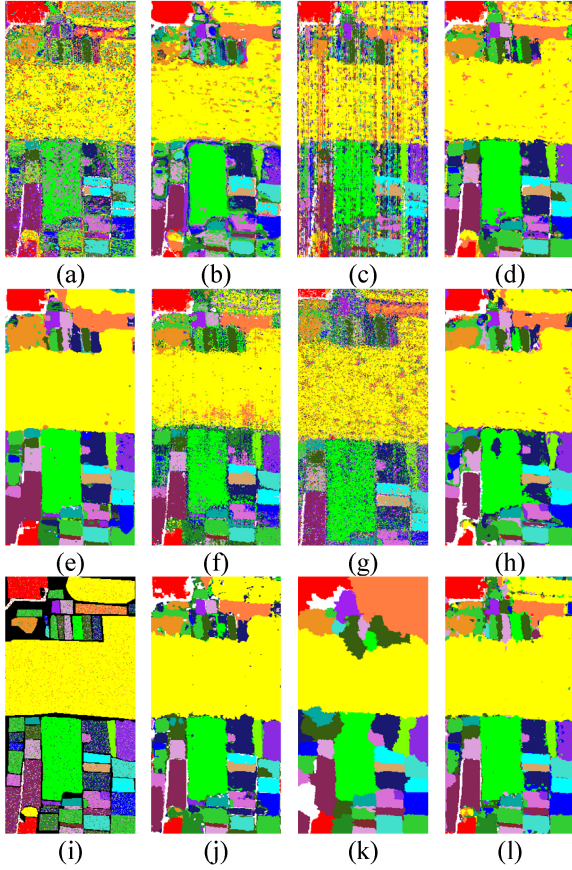


Fig. 18. Classification maps of different methods on WHU-Hi-HongHu. (a) SVM. (b) RPNNet. (c) GTR. (d) FDSI. (e) Two-staged. (f) CSSM. (g) PSAE-DMLSR. (h) M³FuNet. (i) SPFormer. (j) DTFE. (k) SPO. (l) DSCSM.

classification performance. It shows that DSCSM can extract high-quality features through DTFE and make full use of spatial information by SPO, reducing the loss of feature information and making full use of the spectral and spatial information of the original image, which can achieve high-quality ground object recognition in small sample scenarios.

F. Validation of the Effectiveness of DTFE

To intuitively demonstrate the effectiveness of the proposed DTFE method, Fig. 21 shows the separability of pixels from different categories on the Indian Pines dataset. For clarity, only the first eight classes are displayed, with 20 pixels selected from each category. Fig. 21(a), (c), and (e) shows the false-color images of the original HSI, the first three bands of the stacked truncated filtering features (DTF), and the first three principal components of the KPCA-fused features (DTFE), respectively. As seen in the figures, the DTF features enhance the contrast between categories and make the boundaries clearer but the spectral differences between classes are not very distinct. After KPCA feature fusion, the spectral differences between different categories in DTFE become more pronounced.

Fig. 21(b), (d), and (f) displays the scatter plots of the first three bands of the original HSI and DTF features, as well as the scatter plots of the first three principal components of DTFE,

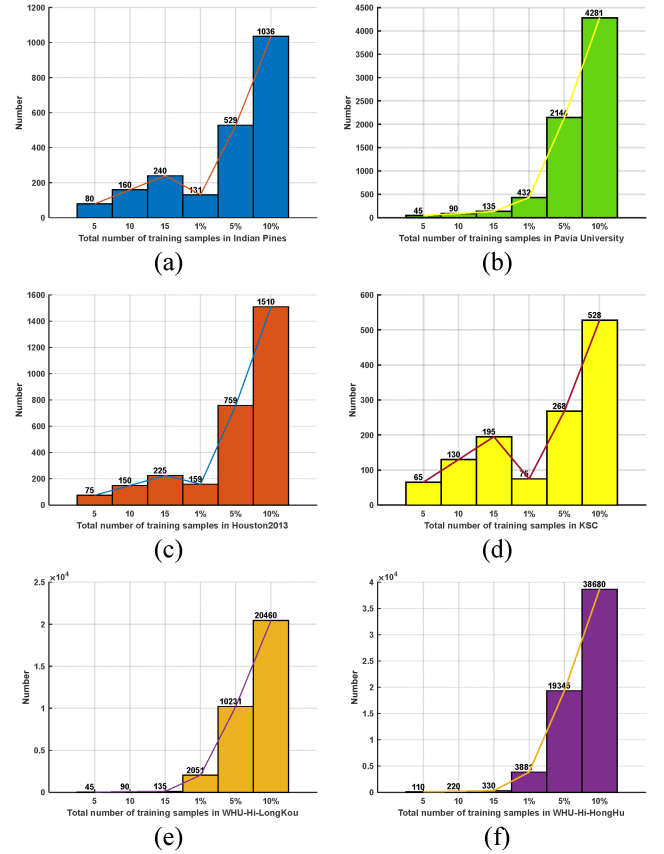


Fig. 19. Total number of training samples for each dataset. (a) Indian Pines. (b) Pavia University. (c) Houston 2013. (d) KSC. (e) WHU-Hi-LongKou. (f) WHU-Hi-HongHu.

TABLE VIII
CLASSIFICATION PERFORMANCE OF DIFFERENT DIMENSIONALITY REDUCTION METHODS ON INDIAN PINES DATASET (%)

Metric	DTF	DTF +PCA	DTF +ICA	DTF +FA	DTF +PPCA	DTF +KPCA (DTFE)
OA	88.15	89.87	90.50	84.95	90.36	91.22
AA	88.69	92.41	89.74	73.42	91.28	95.03
Kappa	86.47	88.30	89.14	82.88	88.88	89.97

The best results are highlighted in bold.

respectively. It can be observed that the pixel distribution in the original HSI is dispersed, with pixels from different categories overlapping. After DTF feature extraction, the distribution of pixels within the same category becomes more clustered but the distances between categories remain close. KPCA feature fusion further increases the interclass distances of the fused features, while the intraclass pixel distribution becomes more concentrated.

To intuitively demonstrate the role and significance of KPCA in the DTFE feature extraction process, four different dimensionality reduction methods are selected to combine with other modules of DSCSM for comparative classification performance analysis, which include PCA, ICA, factor analysis (FA) [64], and probabilistic PCA (PPCA) [65].

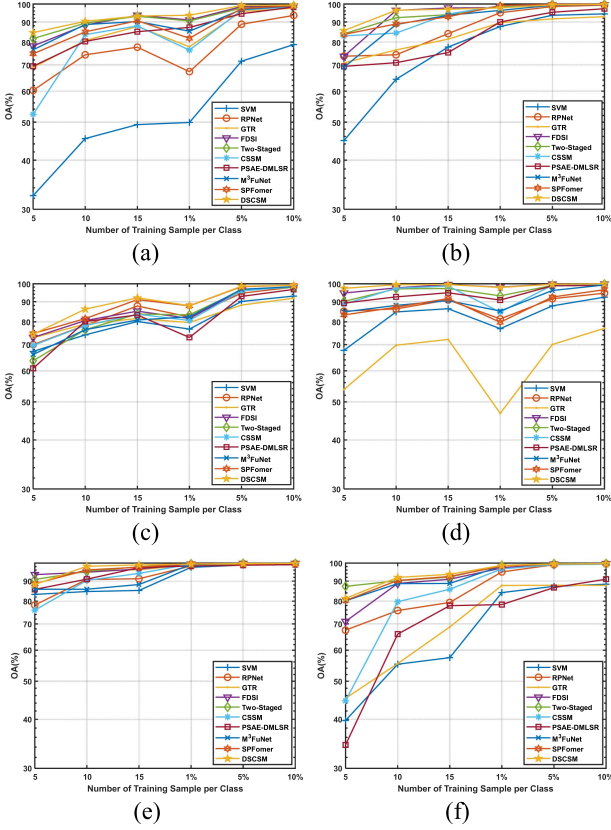


Fig. 20. Influence of training sample number of each dataset on OA. (a) Indian Pines. (b) Pavia University. (c) Houston 2013. (d) KSC. (e) WHU-Hi-LongKou. (f) WHU-Hi-HongHu.

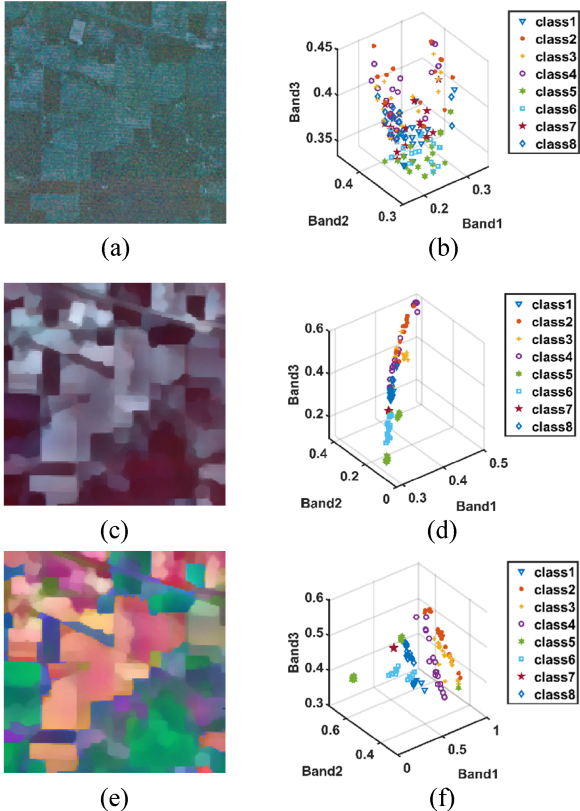


Fig. 21. Influence of DTF and DTFE on pixel separability on Indian pines.

TABLE IX
CLASSIFICATION PERFORMANCE OF DIFFERENT POSTPROCESSING METHODS ON INDIAN PINES DATASET (%)

Metric	MRF	ERW	EPF	STV	SPO
OA	87.55	88.37	89.98	89.08	91.22
AA	93.19	87.45	88.47	94.21	95.03
Kappa	85.90	86.84	88.64	87.61	89.97

The best results are highlighted in bold.

As can be observed from Table VIII, the combination of KPCA with other modules yields the highest classification accuracy. This is attributed to the fact that KPCA leverages kernel function to directly extract features in a high-dimensional space, which is more adept at handling nonlinear issues and conducive to the separation of data from different categories, thereby enhancing the discriminative power of the features and leading to superior classification performance.

G. Validation of the Effectiveness of SPO

To illustrate the suggested SPO's efficacy and ability of discrimination, several widely used postprocessing methods in HSI classification are used for comparison, including Markov random field (MRF) [34], ERW [45], EPF [47], and smoothed total variation (STV) [66], [67]. For a fair comparison, all settings other than the postprocessing methods are kept the same as those in the proposed classification framework, and the comparison techniques' parameter settings comply with the default values stated in the relevant literature. All experiments are conducted on the Indian Pines dataset. Table IX lists the average accuracy values obtained by different postprocessing methods over ten experiments.

As given in the table, SPO achieves the highest values in terms of OA, AA, and Kappa compared to other methods. In terms of OA, it improved by 1.24% compared to the best-performing EPF method and by 2.85% compared to ERW. This indicates that SPO is indeed effective in optimizing classification probabilities in HSI postprocessing.

IV. DISCUSSION

In the task of HSI classification, the quality of spatial and spectral information retention directly affects the accuracy of subsequent classification. In this article, the proposed method was validated for its superior performance across general, moderate, and complex challenge tasks through experiments conducted in six different scenarios. Ablation experiments were performed to systematically assess the impact of each module on overall performance using the method of controlled variables. The results demonstrate that the DTFE significantly enhanced the quality of feature extraction, while the SPO excels in preserving the spatial correlation between pixels. The combination of these two modules notably improved the accuracy and robustness of DSCSM in classification tasks. Although our method achieves good results for different classification tasks, the parameters of DTFE need manual adjustment for different classification

scenarios. Even with fixed parameters, good results can be obtained, but adaptive parameter adjustment remains an important direction for improvement. Moreover, since SPO is essentially a random walk-based image segmentation method, it may fail to sample pixels of the same category that are far apart, especially for small and complex-edged objects. Therefore, in our next step of work, we will consider optimizing this issue.

V. CONCLUSION

A DSCSM is proposed in this article, which mainly consists of DTFE and subRW-based SPO. DTFE smooths the HSI, preserving edge information and smoothing image details. KPCA is used to fuse dual-scale smooth blocks to obtain structural features, reducing spectral redundancy and improving interclass separability. SPO optimizes the initial probabilities obtained by SVM using the spatial relationships between pixels, overcoming DTFE's tendency to lose information and its need for numerous parameter adjustments while enhancing the ability to distinguish similar categories. Experiments on four real hyperspectral datasets from different imaging scenarios show that the proposed method increases classification accuracy by 2%–5% compared to advanced HSI classification methods, even with a small number of training samples. As the number of training samples rises, when 10% of each class's samples are chosen to train, the accuracy exceeds 99% on all four datasets. This indicates that the method has strong generalization and robustness.

REFERENCES

- [1] H. Li, B. Tu, B. Liu, J. Li, and A. Plaza, "Adaptive feature self-attention in spiking neural networks for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5500915, doi: [10.1109/TGRS.2024.3516742](https://doi.org/10.1109/TGRS.2024.3516742).
- [2] J. Zheng, P. Liang, H. Zhao, and W. Deng, "A broad sparse fine-grained image classification model based on dictionary selection strategy," *IEEE Trans. Rel.*, vol. 73, no. 1, pp. 576–588, Mar. 2024.
- [3] X. J. Ran et al., "A hybrid genetic-fuzzy ant colony optimization algorithm for automatic K-means clustering in urban global positioning system," *Eng. Appl. Artif. Intell.*, vol. 137, 2024, Art. no. 109237.
- [4] Y. He, B. Tu, P. Jiang, B. Liu, J. Li, and A. Plaza, "IGroupSS-mamba: Interval group spatial-Spectral Mamba for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5538817.
- [5] R. Yao, H. Zhao, Z. Zhao, C. Guo, and W. Deng, "Parallel convolutional transfer network for bearing fault diagnosis under varying operation states," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 3540713.
- [6] C. Huang, D. Q. Wu, X. B. Zhou, Y. J. Song, H. L. Chen, and W. Deng, "Competitive swarm optimizer with dynamic multi-competitions and convergence accelerator for large-scale optimization problems," *Appl. Soft Comput.*, vol. 167, 2024, Art. no. 112252.
- [7] H. Zhao, Y. Wu, and W. Deng, "Fuzzy broad neuroevolution networks via multi-objective evolutionary algorithms: Balancing structural simplification and performance," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–10, 2025.
- [8] C. Yu et al., "Distillation-constrained prototype representation network for hyperspectral image incremental classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 3359629.
- [9] H. Zhao, Y. Wu, and W. Deng, "An interpretable dynamic inference system based on fuzzy broad learning," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 2527412.
- [10] Y. He, B. Tu, B. Liu, J. Li, and A. Plaza, "3DSS-Mamba: 3D-spectral-spatial Mamba for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5534216.
- [11] Y. J. Song, L. H. Han, B. Zhang, and W. Deng, "A dual-time dual-population multi-objective evolutionary algorithm with application to the portfolio optimization problem," *Eng. Appl. Artif. Intell.*, vol. 133, 2024, Art. no. 108638.
- [12] Y. Lin, D. Guo, Y. Wu, L. Li, E. Q. Wu, and W. Ge, "Fuel consumption prediction for pre-departure flights using attention-based multi-modal fusion," *Inf. Fusion.*, vol. 101, 2024, Art. no. 101983.
- [13] Z. N. Zhu, X. Li, H. Y. Chen, X. B. Zhou, and W. Deng, "An effective and robust genetic algorithm with hybrid multi-strategy and mechanism for airport gate allocation," *Inf. Sci.*, vol. 654, 2024, Art. no. 119892.
- [14] L. Zhou, J. Y. Cai, and S. F. Ding, "The identification of ice floes and calculation of sea ice concentration based on a deep learning method," *Remote Sens.*, vol. 15, no. 10, 2023, Art. no. 2663.
- [15] H. Zhao, L. Wang, Z. Zhao, and W. Deng, "A new fault diagnosis approach using parameterized time-reassigned multisynchrosqueezing transform for rolling bearings," *IEEE Trans. Rel.*, to be published, doi: [10.1109/TR.2024.3371520](https://doi.org/10.1109/TR.2024.3371520).
- [16] H. Chen, T. Wang, T. Chen, and W. Deng, "Hyperspectral image classification based on fusing S3-PCA, 2D-SSA and random patch network," *Remote Sens.*, vol. 15, no. 13, Jul. 2023, Art. no. 3402.
- [17] W. Deng et al., "MOQEAD: Multi-objective QEA with decomposition mechanism and excellent global search and its application," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 9, pp. 12517–12527, Sep. 2024.
- [18] C. Yu, Y. Zhu, M. Song, Y. Wang, and Q. Zhang, "Unseen feature extraction: Spatial mapping expansion with spectral compression network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5521915.
- [19] M. Li, Z. Lu, Q. Cao, J. Gao, and B. Hu, "Automatic assessment method and device for depression symptom severity based on emotional facial expression and pupil-wave," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 2531215.
- [20] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.
- [21] M. D. Farrell and R. M. Mersereau, "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 192–195, Apr. 2005.
- [22] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.
- [23] J. Xia, N. Falco, J. A. Benediktsson, P. Du, and J. Chanussot, "Hyperspectral image classification with rotation random forest via KPCA," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1601–1609, Apr. 2017.
- [24] H. Shen, S. Jegelka, and A. Gretton, "Fast kernel-based independent component analysis," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3498–3511, Sep. 2009.
- [25] H. Chen, J. Ru, H. Long, J. He, T. Chen, and W. Deng, "Semi-supervised adaptive pseudo-label feature learning for hyperspectral image classification in Internet of Things," *IEEE Internet Things J.*, vol. 11, no. 19, pp. 30754–30768, Oct. 2024.
- [26] Q. Sun, J. Chen, L. Zhou, S. Ding, and S. Han, "A study on ice resistance prediction based on deep learning data generation method," *Ocean Eng.*, vol. 301, 2024, Art. no. 117467.
- [27] W. Li et al., "Fault diagnosis using variational autoencoder GAN and focal loss CNN under unbalanced data," *Struct. Health Monit.*, 2024, doi: [10.1177/14759217241254121](https://doi.org/10.1177/14759217241254121).
- [28] 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.
- [29] Z. Yan, H. Yang, D. Guo, and Y. Lin, "Improving airport arrival flow prediction considering heterogeneous and dynamic network dependencies," *Inf. Fusion.*, vol. 100, 2023, Art. no. 101924.
- [30] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
- [31] R. S. Majdar and H. Ghassemian, "A probabilistic SVM approach for hyperspectral image classification using spectral and texture features," *Int. J. Remote Sens.*, vol. 38, no. 15, pp. 4265–4284, Aug. 2017.
- [32] S. Jia et al., "Flexible gabor-based superpixel-level unsupervised LDA for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10394–10409, Dec. 2021.
- [33] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.

- [34] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.
- [35] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
- [36] L. Grady, "Multilabel random walker image segmentation using prior models," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 763–770.
- [37] X. D. Kang, S. T. Li, L. Fang, M. L. Li, and J. A. Benediktsson, "Extended random walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.
- [38] X. Liao, B. Tu, J. Li, and A. Plaza, "Class-wise graph embedding-based active learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5522813.
- [39] Y. Dong, Q. Liu, B. Du, and L. Zhang, "Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 31, pp. 1559–1572, 2022.
- [40] W. Deng, X. Li, J. Xu, W. Li, G. Zhu, and H. Zhao, "BFKD: Blockchain-based federated knowledge distillation for aviation Internet of Things," *IEEE Trans. Rel.*, to be published, doi: [10.1109/TR.2024.3474710](https://doi.org/10.1109/TR.2024.3474710).
- [41] B. Tu, Y. Peng, and A. Plaza, "A new context-aware framework for defending against adversarial attacks in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5505114.
- [42] Z. Zhang, D. Guo, S. Zhou, J. Zhang, and Y. Lin, "Flight trajectory prediction enabled by time-frequency wavelet transform," *Nature Commun.*, vol. 14, 2023, Art. no. 5258.
- [43] X. Li, H. Zhao, J. Xu, G. Zhu, and W. Deng, "APDPFL: Anti-poisoning attack decentralized privacy enhanced federated learning scheme for flight operation data sharing," *IEEE Trans. Wireless Commun.*, vol. 23, no. 12, pp. 19098–19109, Dec. 2024.
- [44] B. Tu, Q. Ren, Q. Li, W. He, and W. He, "Hyperspectral image classification using a superpixel-pixel-subpixel multilevel network," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 5013616.
- [45] P. Duan, P. Ghamisi, X. Kang, B. Rasti, S. Li, and R. Gloaguen, "Fusion of dual spatial information for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7726–7738, Sep. 2021.
- [46] C. Li, X. Tang, L. Shi, Y. Peng, and Y. Tang, "A two-staged feature extraction method based on total variation for hyperspectral images," *Remote Sens.*, vol. 14, no. 2, Jan. 2022, Art. no. 302.
- [47] L. Shi, C. Li, T. Li, and Y. Peng, "A complementary spectral-spatial method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5531017.
- [48] X. Dong, J. Shen, L. Shao, and L. Van Gool, "Sub-Markov random walk for image segmentation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 516–527, Feb. 2016.
- [49] P. Duan, X. Kang, S. Li, P. Ghamisi, and J. A. Benediktsson, "Fusion of multiple edge-preserving operations for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10336–10349, 2019.
- [50] W. Deng, J. R. Wang, A. B. Guo, and H. M. Zhao, "Quantum differential evolutionary algorithm with quantum-adaptive mutation strategy and population state evaluation framework for high-dimensional problems," *Inf. Sci.*, vol. 676, 2024, Art. no. 120787.
- [51] P. J. Huber et al., "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 1964.
- [52] D. Geman and C. Yang, "Nonlinear image recovery with halfquadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
- [53] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comput.*, vol. 27, no. 3, pp. 937–966, 2005.
- [54] W. Liu, P. Zhang, Y. Lei, X. Huang, J. Yang, and M. Ng, "A generalized framework for edge-preserving and structure-preserving image smoothing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6631–6648, Oct. 2022.
- [55] T. H. Kim, K. M. Lee, and S. U. Lee, "Generative image segmentation using random walks with restart," in *Proc. ECCV*, 2008, pp. 264–275.
- [56] Y. Zhong et al., "WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sens. Environ.*, vol. 250, Dec. 2020, Art. no. 112012.
- [57] Y. Zhong et al., "Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 4, pp. 46–62, Dec. 2018.
- [58] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [59] Y. Xu, B. Du, F. Zhang, and L. Zhang, "Hyperspectral image classification via a random patches network," *ISPRS J. Photogramm. Remote Sens.*, vol. 142, pp. 344–357, Aug. 2018.
- [60] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "Generalized tensor regression for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1244–1258, Feb. 2020.
- [61] H. Long, T. Chen, H. Chen, X. Zhou, and W. Deng, "Principal space approximation ensemble discriminative marginalized least-squares regression for hyperspectral image classification," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108031.
- [62] H. Chen et al., "M³FuNet: An unsupervised multivariate feature fusion network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5513015.
- [63] Z. Li, Z. Xue, Q. Xu, L. Zhang, T. Zhu, and M. Zhang, "SP-Former: Self-pooling transformer for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5502019.
- [64] A. Lavanya and S. Sanjeevi, "An improved band selection technique for hyperspectral data using factor analysis," *J. Indian Soc. Remote Sens.*, vol. 41, no. 2, pp. 199–211, 2013.
- [65] J. Xia, J. Chanussot, P. Du, and X. He, "(Semi-) Supervised probabilistic principal component analysis for hyperspectral remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2224–2236, Jun. 2014.
- [66] C. Huang, H. J. Ma, X. B. Zhou, and W. Deng, "Cooperative path planning of multiple unmanned aerial vehicles using cylinder vector particle swarm optimization with gene targeting," *IEEE Sensors J.*, to be published, doi: [10.1109/JSEN.2024.3516124](https://doi.org/10.1109/JSEN.2024.3516124).
- [67] R. Li, K. Cui, R. H. Chan, and R. J. Plemmons, "Classification of hyperspectral images using SVM with shape-adaptive reconstruction and smoothed total variation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 1368–1371.