Mean-Weighted Collaborative Representation-Based Spatial-Spectral Joint Classification for Hyperspectral Images

Hongjun Su, *Senior Member, IEEE*, Dezhong Shi, Zhaohui Xue, *Member, IEEE*, and Qian Du, *Fellow, IEEE*

***Abstract*—Collaborative representation (CR) models have been widely used in hyperspectral image (HSI) classification tasks. However, most CR classification models lack stability and generalization when targeting small samples as well as spatial homogeneity and heterogeneity problems. Therefore, this article proposes a mean-weighted CR classification model (MWCRC) based on the joint spatial-spectral data. It imposes mean and weighted constraints on the representation coefficients based on CR, which attenuates the noise effect and increases the distinguishability between classes. Second, a sample augmentation method based on the principle of minimizing the representation residuals is proposed. Sample augmentation is realized through initial classification and calculation of representation residuals to achieve the objective of consolidating model stability and improving classification accuracy. Meanwhile, in order to alleviate the problem of spatial homogeneity and heterogeneity, the extended morphological profile (EMP) and the stacking approach are utilized to construct the joint spatial-spectral data for the classification of MWCRC. The superiority of the proposed method is demonstrated by experimental validation using a small number of training samples in three real datasets.**

***Index Terms*—Collaborative representation (CR), hyperspectral image (HSI) classification, sample augmentation, spatial-spectral joint.**

I. Introduction

HYPERSPECTRAL remote sensing has been widely used in many industries [1], [2], [3], [4], [5], [6], [7], [8], and the main challenges still faced by hyperspectral remote sensing image classification [9], [10], [11], [12], [13] as one of the main applications are the high dimensionality of the data [14], the small number of labeled samples and the difficulty of their collection [15], [16], [17], and the problem of spatial homogeneity and heterogeneity of classification results [18], [19], [20], [21], [22], [23].

In order to solve these problems, there are three main hyperspectral image (HSI) classification methods as follows.

1) Statistical learning classification algorithms based on kernel transform technology. This type of method utilizes a kernel transform to map a linearly indivisible problem in the original space, into a high-dimensional space where classes can be separable. The most typical classification method is support vector machine [24] with a specific kernel function. However, it is still difficult for them to select optimal kernel functions and parameter combinations.
2) Deep learning based classification algorithms [25]. They realize accurate classification by building network models in multiple layers and extracting features layer by layer. They include stacked autoencoder, deep belief network, convolutional neural network (CNN), etc. [26], [27], [28]. Although this class of algorithms can achieve better classification performance, they tend to have complex tuning parameters, time-consuming training, and poor physical interpretability.
3) Classification algorithms based on representation learning. The methods of this type are all based on a known dictionary to represent testing pixels, which in turn determines the predicted class labels based on the discriminative criterion of representation residuals. Sparse representation classification (SRC) [29] and collaborative representation classification (CRC) [30] are the two most classical approaches. They impose sparse constraints on the representation coefficients through different l_p norm, so that the coefficients mainly contribute values from the true class. These algorithms are computationally simple and interpretable, but the strength of the constraint on sparsity affects the flexibility of the algorithms.

The analysis shows that classification methods based on representation models have been rapidly developed in the hyperspectral field with its advantages. SRC [31], [32] is the most primitive classification model for representation learning. Chen et al. [29] proposed a joint sparse representation classification (JSRC) using simultaneous orthogonal matching tracking solving, which mainly incorporates the spatial context information of the testing pixels into the solving process. Zhang et al. [33] proposed a nonlocally weighted joint sparse representation classification method (NLW-JSRC) based on JSRC by considering the similarity problem of neighboring pixels to testing pixels. However, SRC, while exhibiting strong robustness, has high computational complexity and is difficult to solve. Zhang et al. [34] discussed

Manuscript received 18 February 2024; revised 22 April 2024 and 29 April 2024; accepted 30 April 2024. Date of publication 7 May 2024; date of current version 30 May 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42122008 and Grant 42371327. *(Corresponding author: Hongjun Su.)*

Hongjun Su, Dezhong Shi, and Zhaohui Xue are with the School of Earth Sciences and Engineering, Hohai University, Jiangning 211100, China (e-mail: hjsu@hhu.edu.cn).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA.

Digital Object Identifier 10.1109/JSTARS.2024.3397803

identified that the main contribution in the representation model comes from the collaborative representation (CR) of the training samples and thus proposed the l_2 norm-constrained collaborative representation classification (CRC). Various improved models based on collaborative representation classification emerged. Li et al. [35] proposed a nearest regularized subspace classification method based on CRC, which utilizes Tikhonov regularization to construct distance weighting matrices between testing pixels and each class to enhance the similarity discrimination between samples. Li et al. [36] considered incorporating the idea of joint representation into collaborative representation classification, thus suggesting an adaptive dictionary-based nonlocal joint collaborative representation classifier. Li et al. [37] proposed a Tikhonov regularized kernel collaborative representation classifier, which utilizes the kernel idea to project the original data into a high-dimensional kernel space, thus improving the class separability. Su et al. [38] proposed a band-weighted relaxation collaborative classification method based on superpixel segmentation on the basis of RCR, which fully considered the differences and similarities between features, and at the same time incorporated spatial information using superpixel segmentation. However, with the above multiple variants based on CR, the models tend to lack flexibility, especially under small sample conditions, where the classification results are susceptible to noise, and when the types of features are more complex, the generalization performance of the models is poor; and the enhancement of the generalization has to be at the expense of algorithm complexity.

In order to improve the generalization capability, many methods impose constraints on representation coefficients in the field of face recognition, e.g., competitive collaborative representation classification [39], probabilistic collaborative representation classification (ProCRC) [40], etc.; the main purpose of all these methods is to make representation coefficients to be clustered in the correct class. Recently Gou et al. [41] proposed a CR model based on mean vector and weighted competition for face recognition, which adds competition, mean vector, and weighting three items as constraints of the objective function to the CRC, and considers the similarity between testing samples and the dictionary in many aspects, thus showing better classification results. Considering that too many constraints may bring a large number of too many model hyperparameters, a new classification model can be constructed for the classification of HSIs by creating suitable constraint terms.

When the feature types in the classification scenario are very complex, the classification results often have the problem of spatial homogeneity and heterogeneity, for which many solutions have been proposed to incorporate spatial information in the classification [42], [43], [44]. Su et al. [45] proposed a collaborative representational classification model with multifeature fusion dictionary learning, which simultaneously considers the spectral, local, global, and morphological features of the data and morphological features, which in turn yields the representation coefficients to determine the predicted class. Gao et al. [46] combined the method of multifeature dictionary learning with relaxation collaborative representation (RCR) and suggested a RCR method based on self-balancing dictionary learning. Su et al. [47] proposed a kernel probabilistic CR integrated learning method based on a spatial augmentation dictionary, which incorporates spatial information into the integrated learning classifier by means of a morphological augmentation dictionary. Hu et al. [48] utilized the method of combining local and global texture information to achieve sample augmentation and proposed extended subspace projection classification based on the augmentation of samples with global spatial and local spectral similarity. Although all of the above-mentioned methods consider spatial information, when the extracted spatial features are more complex, it may include information redundancy imposed; when the extracted spatial features are relatively simple, it cannot solve the problem of spatial homogeneity and heterogeneity very well.

The above-mentioned analysis shows that the classification algorithm based on representation learning is simple in method, efficient in operation, and meaningful in interpretation. However, under small sample conditions, affected by the complex spectral and spatial information, the model often lacks a certain degree of generalization, and the enhancement of the generalization capability often increases model complexity. Therefore, under small sample conditions, how to alleviate the problems of spatial homogeneity and heterogeneity based on the representation model while making the model have ideal flexibility and computational complexity is a key issue to be solved.

Therefore, this article first proposes the mean-weighted constraint-based CRC (MWCRC) for HSI classification, so as to alleviate the problem that the classification accuracy is affected by noise under small sample conditions. Second, to further improve the classification accuracy and alleviate the classification problem caused by small samples, a sample augmentation method based on the principle of minimizing the representation residual is suggested. Meanwhile, to solve the problem of spatial homogeneity and heterogeneity, extended morphological profile (EMP) and direct stacking are used to construct joint spatial-spectral data for MWCRC classification. The technical contributions can be summarized as follows.

1) The constructed MWCRC employs a mean vector term with a representation weighting term to constrain the coefficients. Among them, the representation weighting term is different from the common distance weighting, the weights are not easily affected by the sample noise, and it works together with the mean vector term to improve the flexibility of the model under small samples.
2) A sample augmentation method based on the principle of minimizing the residual of representation is proposed. Through the two steps of MWCRC's initial classification and subspace representation to find the residuals, the augmented samples with the highest confidence level are determined, which realizes the doubling of the training samples.

The rest of the article is organized as follows. Section II describes the basic principles of representation learning and EMP. The detailed description of the proposed spatial-spectral joint-based MWCRC (JMWCRC) principle as well as the sample augmentation method is given in Section III. Experiments and results analysis are given in Section IV. Section V discusses

the relevant parameters. Finally, conclusions are drawn in Section VI.

II. Related Work

A. SRC and CRC

In the representation learning-based classification model, the spectral features of a testing pixel can be approximated by a dictionary composed of labeled pixels. Assuming that the number of feature classes is C, $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N] = [\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_C] \in R^{B\times N}$ denotes a dictionary consisting of N training samples, where each dictionary atom is a B-dimensional vector. Also, the whole dictionary can be divided into subdictionaries of C classes, and each subdictionary can be denoted as $\boldsymbol{D}_i = [\boldsymbol{d}_1^i, \boldsymbol{d}_2^i, \ldots, \boldsymbol{d}_{N_i}^i]$, where N_i denotes the number of dictionary atoms corresponding to the ith class and $\sum_{i=1}^{C} N_i = N$. $\boldsymbol{Y} = \{\boldsymbol{y}_i\}_{i=1}^{n}$ is defined as the testing set consisting of the testing samples $\boldsymbol{y}$. The common objective function for SRC and CRC models is as follows:

$$\arg\min_{\boldsymbol{\alpha}} \left(\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_{l_p} \right). \tag{1}$$

When $p = 0$ or 1, the model is SRC [49], where λ is the balance parameter. SRC mainly imposes the sparsest possible constraints on the representation coefficients to determine the optimal solution, but it is clear that (1) at $p = 0$ is an NP-hard combinatorial search problem, and at $p = 1$, it needs to be solved by a greedy tracking algorithm. In short, the SRC is more complicated to solve.

When $p = 2$, the model becomes CRC [50]. Studies have shown that the main contribution of the representation model comes from synergy without overemphasizing the sparsity of the coefficients, so the use of l_2 norm instead of l_0 or l_1 norm makes the model easier to solve while achieving better classification results. The solution to the model is as follows:

$$\boldsymbol{\alpha} = \left(\boldsymbol{D}^T\boldsymbol{D} + \lambda \boldsymbol{I}\right)^{-1}\boldsymbol{D}^T\boldsymbol{y} \tag{2}$$

where $\boldsymbol{I}$ is an identity matrix. After obtaining the representation coefficients a, an approximate representation of the testing sample is obtained

$$\boldsymbol{y} \approx \boldsymbol{D}\boldsymbol{\alpha} = \boldsymbol{D}_1\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{D}_i\boldsymbol{\alpha}_i + \cdots + \boldsymbol{D}_C\boldsymbol{\alpha}_C. \tag{3}$$

Finally, the residuals are derived using the resulting approximation with the real sample, and the predictive labels of $\boldsymbol{y}$ are determined by a certain discriminant criterion. Based on the fact that the probability of $\boldsymbol{y}$ belonging to each class is proportional to the contribution of each class in the representation coefficient, the discriminant criterion can be defined as

$$\text{class}\,(\boldsymbol{y}) = \arg\min_{i=1,2,\ldots,C} \left(\frac{\|\boldsymbol{y} - \boldsymbol{D}_i\boldsymbol{\alpha}_i\|_2^2}{\|\boldsymbol{\alpha}_i\|_2^2} \right). \tag{4}$$

B. Extended Morphology Profile

Extended morphological profile (EMP) [51] is a method for extracting multiscale spatial information from each single-band grayscale image one by one using grayscale morphological operations. First, expansion and erosion are two basic morphological operations on which other morphological composite operations are based. By defining the structural element (SE) of the morphological operation, the morphological opening is the process of first performing the corrosion operation on the image through the SE and then performing the expansion operation; conversely, the morphological closing is the process of first performing the expansion operation on the image through the SE and then performing the corrosion operation. The opening suppresses bright details smaller than the SE, while the closing suppresses dark details, so they are often combined for image smoothing and denoising.

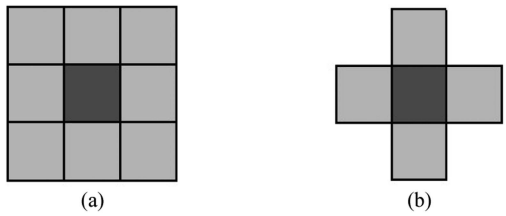

Fig. 1. SEs in 3 × 3 size. (a) Square. (b) Discoidal.

Morphological profile (MP) is the process of obtaining multiple spatial feature images at different scales by performing multiple opening and closing operations simultaneously on a single grayscale image. At this point, the difference between different scales of opening and closing is the size of the sliding subwindow, i.e., the size of the SE. Common types of SEs are square SEs and disk SEs, as shown in Fig. 1 for two types of SEs of 3 × 3 size. The connectivity components that are smaller than the size of the SE and brighter than the neighboring regions can be removed by the opening operation; conversely, the dark connectivity components can be removed by the closing operation. The specific implementation of MP is as follows:

$$\Pi_\gamma(z) = \left\{ \Pi_{\gamma\lambda} : \Pi_{\gamma\lambda} = \gamma_\lambda^*(z), \quad \forall\lambda \in [0, x] \right\} \tag{5}$$

$$\Pi_\varphi(z) = \left\{ \Pi_{\varphi\lambda} : \Pi_{\varphi\lambda} = \varphi_\lambda^*(z), \quad \forall\lambda \in [0, x] \right\} \tag{6}$$

where Π_γ and Π_φ denote the opening profile and closing profile, respectively, z is a pixel of the single-band image $\boldsymbol{I}$, λ denotes the defined SE size, and x denotes the size of the largest SE. $\gamma_\lambda^*(z)/\varphi_\lambda^*(z)$ denotes the execution of the opening/closing operation on the pixel z by utilizing the defined SE. MP is the series connection of Π_γ and Π_φ. The corresponding equation can be expressed as

$$\text{MP}_x(\boldsymbol{I}) = \{\varphi_1^*(\boldsymbol{I}), \ldots, \varphi_x^*(\boldsymbol{I}), \boldsymbol{I}, \gamma_1^*(\boldsymbol{I}), \ldots, \gamma_x^*(\boldsymbol{I})\}. \tag{7}$$

According to (7), it can be seen that if there are m SEs between [0, x], the MP operation is to utilize the m SEs to perform opening/closing profiles on the single-band image to obtain $(2m+1)$ spatial images of the bands, and then connect them in series.

The EMP operation, on the other hand, is to first perform an MP operation on multiple single-band images of the same scene, and then concatenate the obtained results. The specific operation formula is as follows:

$$\text{EMP} = \{\text{MP}_x(\boldsymbol{I}_1), \text{MP}_x(\boldsymbol{I}_2), \ldots, \text{MP}_x(\boldsymbol{I}_n)\}. \tag{8}$$

Each MP can extract $2m+1$ spatial feature images, so $n(2m+1)$ stacks of data containing multiple spatial features can be finally obtained by EMP operation.

III. Proposed Methods

A. MWCRC Model

Due to the influence of imaging conditions, hyperspectral remote sensing images are highly susceptible to noise, especially when the number of training samples is small. The traditional classification models based on representation learning have poor stability. Considering that fine classification under small samples cannot be achieved by only relying on the two conditions of CR among training samples and sparsity of representation coefficients, this article adopts the mean vector term and the representation weighting term to constrain the representation coefficients on the basis of CRC, thus constructing the MWCRC model. The main idea of the model is to enhance stability to the model while forcing the main contributing values of the representation coefficients to be more concentrated in the true class in the case of small samples. The objective function of the model is as follows:

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\alpha}} \left(\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 + \lambda \sum_{i=1}^{C} \|\boldsymbol{v}_i - \boldsymbol{D}_i\boldsymbol{\alpha}_i\|_2^2 + \gamma \sum_{i=1}^{C} \omega_i \|\boldsymbol{\alpha}_i\|_2^2 \right) \tag{9}$$

where the representation coefficient α can be denoted as $\boldsymbol{\alpha} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N]^T = [\boldsymbol{\alpha}_1; \boldsymbol{\alpha}_2; \ldots; \boldsymbol{\alpha}_C] \in R^{N\times 1}$, λ and γ are both balance parameters, $\sum_{i=1}^{C} \|\boldsymbol{v}_i - \boldsymbol{D}_i\boldsymbol{\alpha}_i\|_2^2$ is called the mean vector term, and $\sum_{i=1}^{C} \omega_i \|\boldsymbol{\alpha}_i\|_2^2$ is called the representation weighting term.

In the mean vector term, $\boldsymbol{v}_i$ denotes the vector mean of all samples from the ith class, estimated as

$$\boldsymbol{v}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \boldsymbol{d}_j^i. \tag{10}$$

First, the use of the mean vector in this item reduces the noise influence in the case of a small number of training samples; second, the use of the difference between the mean vector of each class and the corresponding reconstructed value of the subdictionary to sum up, and make the sum approach the minimum as the final constraint, which excludes the possibility of the main values of the representation coefficients being clustered in an erroneous class that is dissimilar to the real class, and thus forces the main values of the representation coefficients to be clustered towards the real class.

The representation weighting term uses the residuals from the best reconstruction of the testing sample in each subspace as the weights in the coefficients. The constrained weights for the coefficients in class i are as

$$\omega_i = \|\boldsymbol{y} - \boldsymbol{D}_i\widehat{\boldsymbol{\alpha}_i}\|_2^2, \quad i = 1, 2, \ldots, C \tag{11}$$

where $\widehat{\boldsymbol{\alpha}_i}$ is obtained when $\boldsymbol{y}$ is optimally reconstructed via a subdictionary from class i, i.e., the

$$\widehat{\boldsymbol{\alpha}_i} = \min_{\boldsymbol{\alpha}_i} \left(\|\boldsymbol{y} - \boldsymbol{D}_i\widehat{\boldsymbol{\alpha}_i}\|_2^2 \right). \tag{12}$$

Representation weighting is distinguished from the common distance-based weighting, both of which exploit the degree of similarity between training and testing samples, but the distance-based weighting is susceptible to noise in small-sample conditions, such as the CRT model utilizing Tikhonov regularization [52]. As can be seen from (11) and (12), the smaller the representation residuals (weights), the higher the degree of similarity between the samples, and the larger the value obtained by the representation coefficients of the corresponding class, thus increasing the strength of the constraints on the representation coefficients clustered in a particular class, making the classification model more discriminative.

MWCRC, like CRC, has a corresponding closed-form solution. For ease of solving, (9) is reorganized as follows:

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\alpha}} \left(\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 + \lambda \sum_{i=1}^{C} \left\|\boldsymbol{v}_i - \widehat{\boldsymbol{D}_i}\boldsymbol{\alpha}\right\|_2^2 + \gamma \sum_{i=1}^{C} \|\boldsymbol{W}\boldsymbol{\alpha}\|_2^2 \right) \tag{13}$$

where $\widehat{\boldsymbol{D}_i}$ and $\boldsymbol{W}$ are defined as

$$\widehat{\boldsymbol{D}_i} = [\boldsymbol{0}, \boldsymbol{0}, \ldots, \boldsymbol{D}_i, \ldots, \boldsymbol{0}, \boldsymbol{0}] \in \mathbb{R}^{B\times N} \tag{14}$$

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_1 & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{W}_C \end{bmatrix} \quad \boldsymbol{W}_i = \begin{bmatrix} \sqrt{\omega_i} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\omega_i} \end{bmatrix}. \tag{15}$$

The closed-form solution to the model is obtained by taking a partial derivative of α as

$$\boldsymbol{\alpha} = \left(\boldsymbol{D}^T\boldsymbol{D} + \lambda\boldsymbol{M} + \gamma\boldsymbol{W}^T\boldsymbol{W}\right)^{-1}\left(\boldsymbol{D}^T\boldsymbol{y} + \lambda\boldsymbol{V}\right) \tag{16}$$

where $\boldsymbol{M}$ and $\boldsymbol{V}$ are defined as

$$\boldsymbol{M} = \sum_{i=1}^{C} \widehat{\boldsymbol{D}_i}^T \widehat{\boldsymbol{D}_i} \tag{17}$$

$$\boldsymbol{V} = \sum_{i=1}^{C} \widehat{\boldsymbol{D}_i}^T \boldsymbol{v}_i. \tag{18}$$

In order to fully utilize the discriminative information, the classification guidelines used in this article are as follows:

$$\text{class}(\boldsymbol{y}) = \arg\min_{i=1,2,\cdots,C} \left(\frac{\|\boldsymbol{y} - \boldsymbol{D}_i\boldsymbol{\alpha}_i\|_2^2}{\|\boldsymbol{\alpha}_i\|_2^2} \right). \tag{19}$$

Algorithm 1 gives the detailed MWCRC pseudocode.

B. Spatial-Spectral Joint-Based MWCRC

Due to the imaging conditions, the spectra of the same kind of features often show certain differences [53]; meanwhile, the spectra between different classes of features can have strong similarities in complex scenes. These effects cause the intraclass

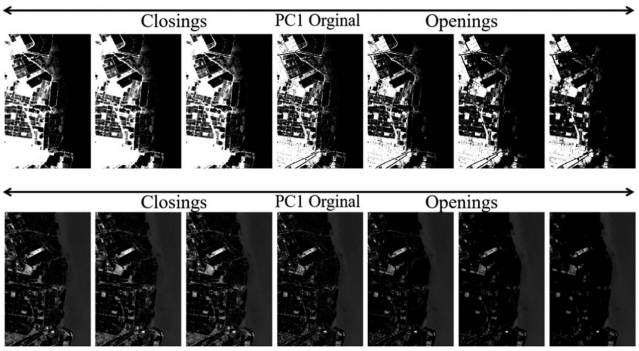

Fig. 2. Results of EMP operation based on square SEs. The SE sizes are 3, 5, and 7.

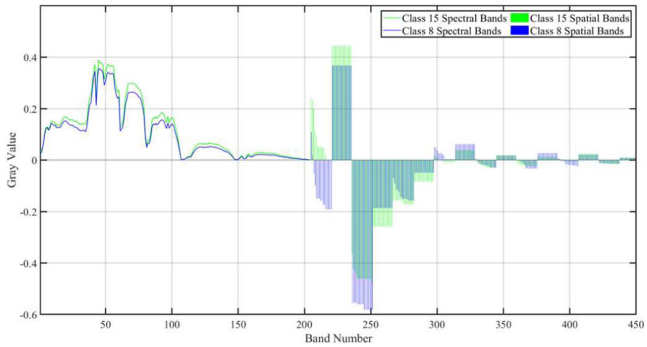

Fig. 3. Spatial-spectral features of similar features in the Salinas.

Algorithm 1: MWCRC.

Input: HSI data: $\boldsymbol{Y}$
Training samples set: $\boldsymbol{D}$
Balance parameters: λ, γ
For each testing sample $\boldsymbol{y}$ in $\boldsymbol{Y}$
1. Calculate the weighting matrix $\boldsymbol{W}$ of the coefficients by (11), (12), and (15)
2. Calculate the block diagonal matrix $\boldsymbol{M}$ by (14) and (17)
3. Calculate the column vector $\boldsymbol{V}$ by (10), (14), and (18)
4. Calculate the coefficient of representation α by (16)
5. Classify the testing sample $\boldsymbol{y}$ by (19) and predict its label $class(\boldsymbol{y})$

End
Output: Classification result of HSI: $class(\boldsymbol{Y})$

distance of feature spectra to increase and the interclass distance to decrease, making it difficult to improve the classification accuracy. Meanwhile, CR-based classification methods are all forced constraints [54], i.e., all bands of the training samples share the same coefficients, so the classification accuracy is affected. In this article, in order to fully utilize the advantages of the MWCRC model, the data preprocessing method is used to extract the spatial information in the original data, and the JMWCRC is constructed to solve the above-mentioned problems.

Considering that EMP can extract the spatial features of many images from multiple scales, the original hyperspectral data are first subjected to principal component analysis (PCA), and the first K principal components (PCs) associated with the K largest eigenvalues of the data covariance matrix are selected according to the principle of PCA, so that multiple single-band images of the same scene are obtained. They are subjected to EMP operation to obtain spatial cube data with multiscale spatial information.

According to the operation principle of EMP, using such features can increase the interclass distance and reduce the intraclass distance. Specifically, in hyperspectral imagery, there are many anomalous pixels interspersed in the connectivity region of a certain class of features, and this manifests itself in a certain single-band image as light/dark details (noise) within the region, which can be smoothed well after MP operation. Fig. 2 shows the effect of MP operation on the first two PC, with the original PC image in the center and the results of closing/opening at different scales on the left and right, respectively. It can be seen that, after the opening/closing operation, the light/dark details of the discrete or small connected areas within the connected areas of the same class of features are well eliminated.

According to Fig. 2, the resultant maps obtained from the opening/closing operations performed with different sizes of SEs are also significantly different from each other, mainly due to the interaction between the size of the features in the image and the size of the SEs. Therefore, a single size of SEs for information extraction is very likely to lead to the problem of insufficient information extraction or excessive smoothing. Meanwhile, when classifying HSI images, it is often unclear about the regional size of the feature, so EMP can effectively solve the above problems by stacking the results of the operation after opening/closing the SEs of different scales. When the feature size is small, the single-channel image obtained by the small-scale SE can constrain the large-scale image, thus preventing the missing boundary of the classification result caused by the transitional smoothing, whereas when the feature size is large, the large-scale SE can directly carry out the extraction of spatial features. Thus, the method can have the effect of reducing the parameter adjustment.

In order to fully utilize the spectral information, this article adopts the direct stacking method to combine the original spectral data with the spatial data extracted by EMP to construct the spatial-spectral joint cube data for JMWCRC classification. The construction formula of the spatial-spectral joint data is

$$\mathrm{EMP} = \left\{\mathrm{MP}_x\left(\mathrm{PC}_1\right), \mathrm{MP}_x\left(\mathrm{PC}_2\right), \ldots, \mathrm{MP}_x\left(\mathrm{PC}_K\right)\right\} \tag{20}$$

$$\mathbf{HSE} = \{\mathrm{HSI}; \mathrm{EMP}\}. \tag{21}$$

The method maintains the advantages of the original hyperspectral data while increasing the interclass distance of the features and decreasing the intraclass distance of the features through spatial data. As shown in Fig. 3 , the spectral-spatial features of two similar classes of features in the Salinas HSI are visualized, and it can be seen that the difference in the original spectral information of the two classes of features is very small, but the difference in spatial information is obvious in the interval

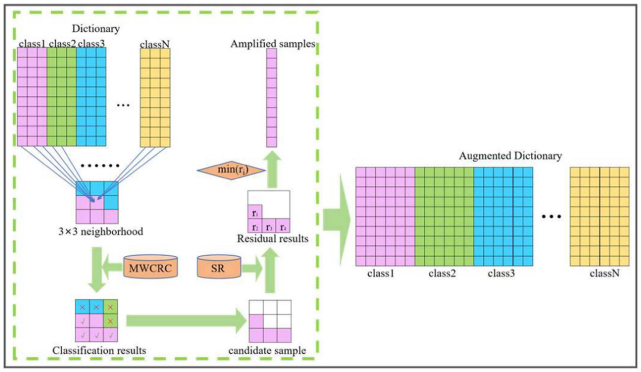

Fig. 4. Sample amplification process based on residual minimization.

Algorithm 2: JMWCRC.

Input: HSI data: $\boldsymbol{Y}$
Balance parameters: λ, γ
Data preprocessing:
1. Extraction of K PCs of $\boldsymbol{Y}$ by PCA
2. Extraction of multiscale spatial information for K PCs by (7)
3. Construct spatial cube data EMP by (20)
4. Construct spatial-spectral joint cube data **HSE** by (21)

Initiation Classification:
input: Training samples set $\boldsymbol{D}$ from **HSE**
for each testing sample $\boldsymbol{y}$ in **HSE**
Perform the MWCRC operation on the sample $\boldsymbol{y}$ by Algorithm 1
end
Output: Classification result: *class*($\boldsymbol{Y}$)

around 204–330 from stacked spatial features. Algorithm 2 gives the detailed JMWCRC pseudocode.

C. Sample Augmentation-Based JMWCRC

Considering that the improvement of classification accuracy is affected by the small number of training samples, in order to further improve the classification accuracy and to consolidate the stability of the model, this article proposes a sample augmentation method using the principle of minimizing the representational residuals (JMWCRC-SA) on the basis of JMWCRC. Fig. 4 shows the exact process. The method takes into account that if the augmented samples are misclassified, the accuracy improvement will be hindered, so the highest confidence samples need to be selected as augmented samples. The method first preliminarily classifies the 3×3 domain pixels of the initial training samples in the HSE based on the principle of feature continuity, thus excluding the interfering pixels. Then the initial classification results are represented in subspace, and the sample with the smallest representation residual is used as the augmentation sample. The specific realization process is as follows.

First, the first step in Fig. 4 is to select the initial training set $\boldsymbol{D}$ in the HSE and then find the 3×3 neighborhood pixels where

Algorithm 3: JMWCRC-SA.

Input: HSI data: $\boldsymbol{Y}$
Balance parameters: λ, γ
Data preprocessing:
HSE data and training sample set $\boldsymbol{D}$ are obtained by Algorithm 2
Sample Augmentation:
for each training sample $\boldsymbol{d}$ in $\boldsymbol{D}$
1. Find the 3×3 neighborhood pixels of $\boldsymbol{d}$ in **HSE**
2. Neighborhood pixels are classified by MWCRC and pixels of the same class as $\boldsymbol{d}$ are selected.
3. Calculate the representation residual r of the selected pixel by (22) and (23)
4. Determine the augmented sample $\boldsymbol{y}_j$ by r-minimizing and $\boldsymbol{D}_i = [\boldsymbol{D}_i, \boldsymbol{y}_j]$

end
output: Augmented training sample set $\boldsymbol{D}'$
Initiation Classification:
for each testing sample $\boldsymbol{y}$ in **HSE**
Perform the MWCRC operation on the sample $\boldsymbol{y}$ by Algorithm 1
end
Output: Classification result of HSI: *class*($\boldsymbol{Y}$)

the training samples are located. The purpose of choosing the 3×3 neighborhood here is that, considering the continuity of features, in the worst case, there is only one pixel of the same class in the connected domain of the training sample, so choosing the 3×3 neighborhood can get the augmented sample with the highest confidence.

Second, the pixels in the neighborhood are initially classified using MWCRC, and pixels of the same class as the center pixel are selected from the classification results. The main consideration of this process is to exclude the interfering pixels of other classes as much as possible.

Finally, the selected pixels are represented subspace-optimally one by one and the representation residuals are obtained r. The formula is given by

$$r_j = \left\| \boldsymbol{y}_j - \boldsymbol{D}_i \widehat{\boldsymbol{\alpha}_j} \right\|_2^2 \tag{22}$$

$$\widehat{\boldsymbol{\alpha}_j} = \arg\min_{\boldsymbol{\alpha}} \left(\left\| \boldsymbol{y}_j - \boldsymbol{D}_i \boldsymbol{\alpha} \right\|_2^2 \right) \quad i.e. \quad \widehat{\boldsymbol{\alpha}_j} = \left(\boldsymbol{D}_i^T \boldsymbol{D}_i \right)^{-1} \boldsymbol{D}_i^T \boldsymbol{y}_j \tag{23}$$

where $\boldsymbol{y}_j$ denotes the jth pixel selected in the second step and $\boldsymbol{D}_i$ denotes the subdictionary where the center pixel is located. The pixel with the smallest residual is considered to be most similar to the center pixel and is considered to have the highest confidence level. Therefore the pixel with the smallest residual r is selected as the augmented sample.

The above-mentioned process is performed on the training samples in $\boldsymbol{D}$ one by one, thus doubling the number of training samples i.e.,$\boldsymbol{D}' \in R^{B \times 2N}$. After obtaining the augmented dictionary, JMWCRC is then utilized to complete the classification.

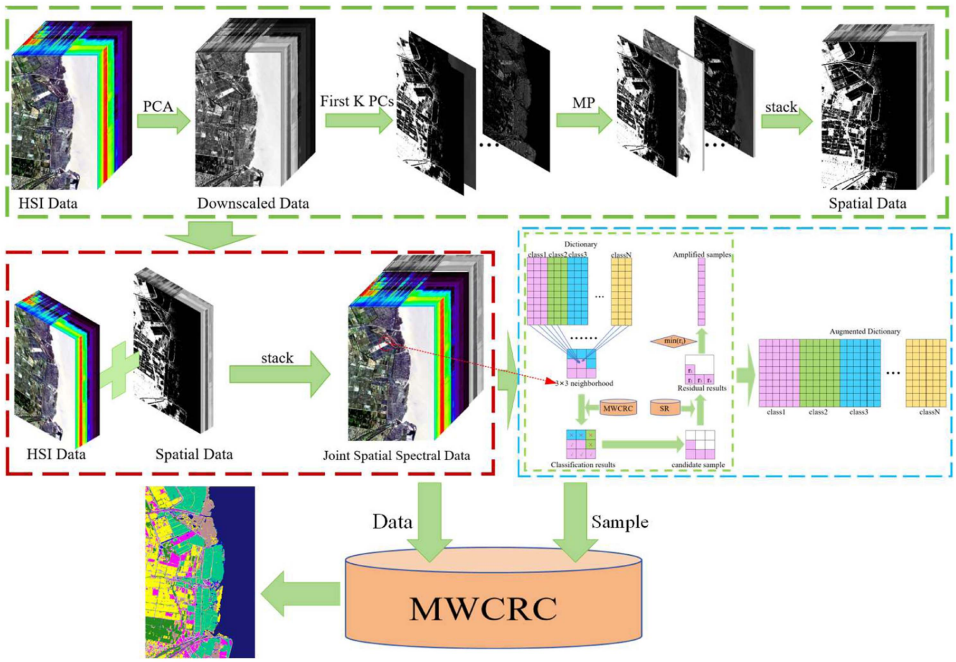

Fig. 5. Data flow of the proposed methods.

Algorithm 3 gives the detailed JMWCRC-SA pseudocode. Fig. 5 gives the detailed flow of the proposed method in this article.

IV. Experiments and Results

In this section, three experimental datasets are introduced, namely the Salinas dataset, which has more complex feature classes, and two wetland datasets, i.e., Yangtze River Delta and Dafeng Natural Reserve, which have more complex feature scenarios. A series of experiments are conducted to evaluate the performance of the proposed method. Salinas is a public dataset.[1] Yangtze River Delta and Dafeng Natural Reserve from [11].

A. Dataset

1) Salinas: The first set of data was acquired from the Airborne Visible Infrared Imaging Spectrometer over the Salinas Valley, California, United States of America, and the image consists of 512 × 217 pixels covering 204 spectral bands except for the water vapor absorption band. The region contains 16 feature classes, and the classes are complex, with a variety of very similar classes. The sample information of the data is shown in Table I. The false-color composite image and the distribution of ground labels are shown in Fig. 6.

2) Yangtze River Delta: The second set of data was taken by the AHSI on-board hyperspectral sensor. The sensor was mounted on the GF-5 satellite for ground-based observation. The area selected for data acquisition is located in the Yangtze River Delta in the Nantong region of eastern Jiangsu Province, China. The image size is 1160 × 1290, covering 268 spectral

1.[Online]. Available at: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

TABLE I
Sixteen Ground-Truth Classes of the Salinas Dataset

Number	Class	Number of samples
1	Brocoli_green_weeds_1	2009
2	Brocoli_green_weeds_2	3726
3	Fallow	1976
4	Fallow_rough_plow	1394
5	Fallow_smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes_untrained	11271
9	Soil_vinyard_develop	6203
10	Corn_senesced_green_weeds	3278
11	Lettuce_romaine_4wk	1068
12	Lettuce_romaine_5wk	1927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1070
15	Vinyard_untrained	7268
16	Vinyard_vertical_trellis	1807

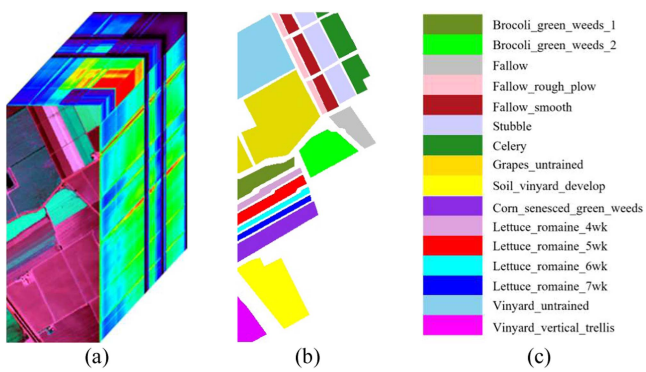

Fig. 6. Salinas dataset. (a) False color image. (b) Ground truth. (c) Legend.

TABLE II
NINE GROUND-TRUTH CLASSES OF THE YANGTZE RIVER DELTA DATASET

Number	Class	Number of samples
1	River	1273
2	Trees	575
3	Bare Land	200
4	Farmland	189
5	Lake	137
6	Mud Flat	476
7	Flood Plain	435
8	Impervious surface	1046
9	Plant	455

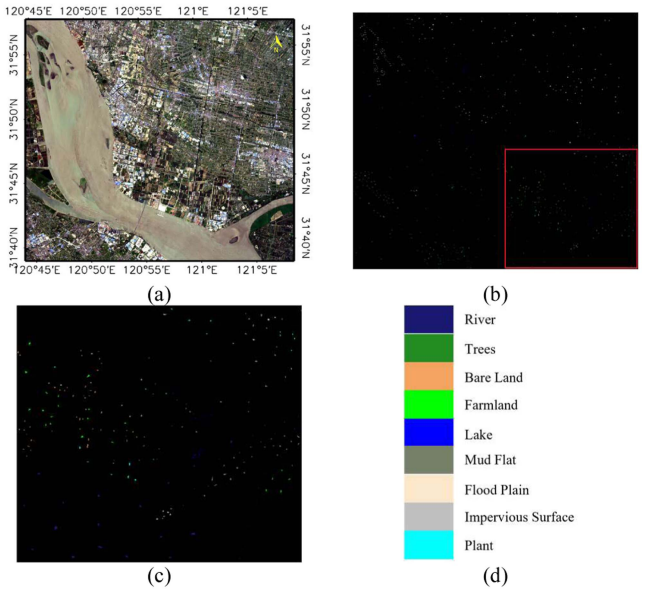

Fig. 7. Yangtze River Delta dataset. (a) False color image. (b) Ground truth. (c) Zoom of part of (b) to highlight the distribution. (d) Legend.

bands except 62 noisy bands. The region contains nine classes of features. The sample information of the data is shown in Table II, and the false-color composite image and the distribution of ground labels are shown in Fig. 7.

3) Dafeng Natural Reserve: The imaging equipment for the third set of data is the same as that for the second set. The area where the data were taken is located in the Dafeng Nature Reserve in the coastal wetland of Yancheng City, Jiangsu Province, China. The image size is 986 × 632, covering 256 spectral bands except 74 noise bands. The area contains nine classes of features, and the scenes in the data are more complex. The sample information of the data is shown in Table III, and the false-color composite image and the distribution of ground labels are shown in Fig. 8.

B. Experimental Setup

In order to fully verify the performance advantages of the proposed method, the following representative classification methods are selected for comparison. Among them, extreme learning machine (ELM) [55], three-dimensional (3-D)-CNN [25] are the more classical classification methods; CRC is the classical representation model classifier; JCRC [56] is the

TABLE III
NINE GROUND-TRUTH CLASSES OF THE DAFENG NATURAL RESERVE DATASET

Number	Class	Number of samples
1	Sea	1020
2	Mud Flat	528
3	Road	150
4	Buildings	89
5	Farmland	918
6	River	276
7	Trees	724
8	Aquiculture	746
9	Lake	79

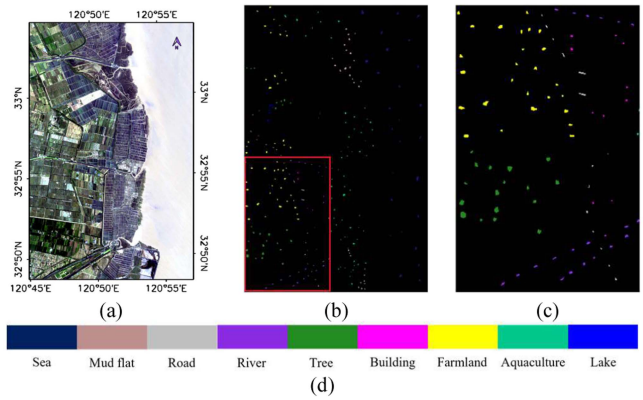

Fig. 8. Dafeng Natural Reserve dataset. (a) False color image. (b) Ground truth. (c) Zoom of part of (b) to highlight the distribution. (d) Legend.

TABLE IV
OPTIMAL PARAMETER SETTING

Methods	Parameter	Salinas	Yangtze River Delta	Dafeng Natural Reserve
MWCRC	λ	1e-6	1e-6	1e-4
	γ	1	1	1e1
JMWCRC	λ	1e-7	1e-6	1e-3
	γ	1e-3	1e-1	1e2
	x	31	9	7
	K/B_s	35/1085	35/315	30/210
JMWCRC -SA	λ	1e-7	1e-5	1e-2
	γ	1e-2	1e-1	1e2

CR model by combining spatial information; TCRC [57] and KProCRC [58] are the more popular CR models at present; KProCRC-LNSAE [47] is the latest classification method that combines spatial information to augment the dictionary. MWCRC, JMWCRC, and JMWCRC-SA are the three classification methods proposed in this article. In order to simulate the classification performance under small samples, the training sets in the experiments of this article are all set to ten training samples per class. The optimal parameters of the three methods proposed in this article are shown in Table IV, where x denotes the maximum SE size and square SEs are chosen for all SEs. K denotes the number of PCs selected after PCA processing, and B_s denotes the dimension of the spatial data. The classification results of the other methods are obtained under the optimal parameters.

TABLE V
CLASSIFICATION ACCURACY FOR THE SALINAS DATASET

Class	Train/Test	ELM	3D-CNN	CRC	JCRC	TCRC	KProCRC	KProCRC -LNSAE	MWCRC	JMWCRC	JMWCRC-SA
1	10/1999	99.57	93.82	99.47	98.65	97.69	99.98	100	99.58	95.42	94.85
2	10/3716	98.51	97.80	98.76	99.75	99.56	98.95	98.57	99.38	99.21	99.35
3	10/1966	90.40	81.27	91.11	99.55	98.44	96.19	93.03	94.49	95.13	96.22
4	10/1384	99.15	98.29	90.96	98.82	98.68	98.33	97.80	98.64	91.16	93.59
5	10/2668	92.09	96.13	89.32	90.43	94.28	89.59	93.90	91.41	98.71	98.96
6	10/3949	100.00	99.99	99.84	96.59	99.84	99.99	100	99.74	99.16	99.16
7	10/3569	98.81	90.66	99.48	99.69	99.84	99.38	98.28	99.41	96.88	96.69
8	10/11261	71.24	66.29	68.96	76.33	70.91	72.89	74.12	76.86	93.39	93.81
9	10/6193	98.49	94.77	91.74	96.31	93.57	99.19	99.48	99.08	98.67	99.31
10	10/3268	88.67	84.99	78.77	87.92	88.43	86.57	86.54	85.16	94.39	93.58
11	10/1058	90.90	62.38	90.29	90.93	93.01	89.77	90.04	94.56	90.45	94.10
12	10/1917	87.17	66.21	92.20	99.61	98.08	92.95	95.28	95.07	97.24	98.97
13	10/906	76.72	60.20	60.65	77.68	64.71	87.14	92.21	97.75	98.18	98.92
14	10/1060	95.61	85.28	74.02	83.46	76.39	96.26	93.66	94.53	84.30	86.93
15	10/7258	48.09	51.76	52.60	70.77	65.07	48.03	51.42	57.78	85.97	89.02
16	10/1797	99.71	96.64	98.94	99.37	99.99	98.97	99.43	98.72	94.12	92.97
OA(%)		82.81	79.12	81.76	86.66	84.95	83.38	84.37	86.72	94.26	95.07
AA(%)		89.70	82.90	86.07	91.62	89.91	90.89	91.48	92.63	94.52	95.40
Kappa(%)		80.97	76.71	79.72	85.16	83.26	81.61	82.71	85.26	93.62	94.51
Time(s)		26.59	46.58	22.56	29.25	40.63	63.09	208.20	95.78	194.02	435.99

TABLE VI
CLASSIFICATION ACCURACY FOR THE YANGTZE RIVER DELTA DATASET

Class	Train/Test	ELM	3D-CNN	CRC	JCRC	TCRC	KProCRC	KProCRC -LNSAE	MWCRC	JMWCRC	JMWCRC-SA
1	10/1263	99.60	99.80	98.44	98.91	97.35	99.75	99.82	98.66	99.32	99.39
2	10/565	72.72	82.67	73.21	79.02	72.00	86.70	90.30	79.13	83.23	85.68
3	10/190	77.30	81.07	69.42	75.95	67.96	87.91	90.86	89.60	84.99	92.47
4	10/179	77.18	91.33	88.82	91.99	91.35	88.28	90.76	95.36	96.87	99.30
5	10/127	49.57	53.05	68.05	72.64	74.12	66.01	67.62	75.17	70.72	70.53
6	10/466	59.27	68.85	63.62	66.95	63.24	69.31	72.03	71.68	78.17	82.12
7	10/425	79.12	91.87	93.20	95.37	96.42	91.09	91.51	94.53	99.63	99.69
8	10/1036	83.94	93.37	90.91	89.77	89.44	94.48	96.71	92.88	97.52	97.26
9	10/445	100.00	99.92	99.85	99.91	99.93	99.98	99.95	99.91	99.48	98.96
OA(%)		81.50	88.82	86.31	88.05	85.86	89.78	91.84	90.45	92.74	93.92
AA(%)		77.63	84.66	82.84	85.61	83.54	86.39	88.84	88.55	89.99	91.71
Kappa(%)		77.97	86.66	83.71	85.77	83.14	87.83	90.26	88.59	91.35	92.74
Time(s)		19.45	53.09	15.69	17.27	20.20	50.15	128.33	22.01	34.40	47.72

For quantitative assessment, overall accuracy (OA), average accuracy (AA), Kappa coefficient (Kappa), and class accuracy are adopted. In order to consider the computational efficiency, the running time to complete the classification of the testing set is used as a criterion. To avoid bias, all experimental results shown are the average of the results of ten runs of experiments. All experiments were conducted in a MATLAB environment on a 3.2-GHz CPU.

C. Classification Performance

The quantitative values of the classification results for the three datasets are given in Tables V–VII. The corresponding visualized classification result graphs are demonstrated in Figs. 9–11.

For Salinas data, JMWCRC-SA has the highest classification accuracy, and the OA value can reach 95.07%. Compared with MWCRC, it is improved by 8.35%, which shows that the incorporation of spatial information effectively improves the classification accuracy. Compared to other methods, MWCRC still has the highest classification accuracy, especially the OA is improved by 2.35% compared to the new KProCRC-LNSAE. According to Table V, it can be seen that the improvement in the accuracy of the proposed method mainly comes from class 8 and class 15 (vineyards in different states). Combined with Fig. 9, it can be seen that all the comparison methods have more severe misclassification between these two features, such as 3-D-CNN. While JCRC mitigates the problem by combining spatial information, it causes missing boundaries between other features. KProCRC-LNSAE can well mitigate the above-mentioned problem by augmenting the dictionary with spatial information, but the classification effect is weaker than that of the method presented in this article. From Fig. 9(j), it can be seen that JMWCRC extracts spatial information from multiple scales, which can distinguish similar features well without causing the problem of missing boundaries among other features.

For Yangtze River Delta data, JMWCRC-SA has the highest classification accuracy, and the OA value can reach 93.92%, which is improved by 3.47% compared with MWCRC. As KProCRC-LNSA combines spatial information, it has a better classification effect for the topographically complex data, but MWCRC is only considered from the spectral point of view, and

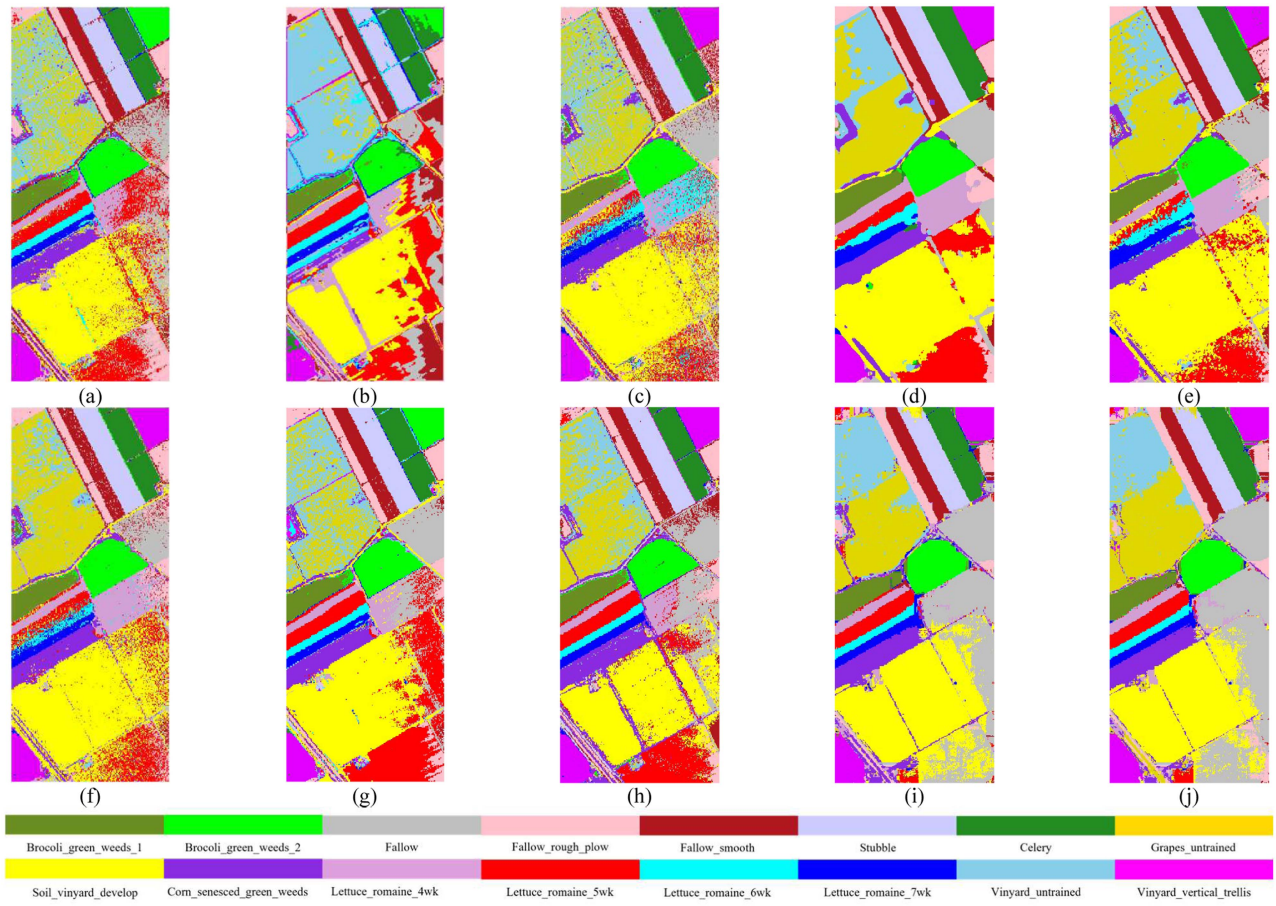

Fig. 9. Classification maps of Salinas. (a) ELM. (b) 3-D-CNN. (c) CRC. (d) JCRC. (e) TCRC. (f) KProCRC. (g) KProCRC-LNSAE. (h) MWCRC. (i) JMWCRC. (j) JMWCRC-SA.

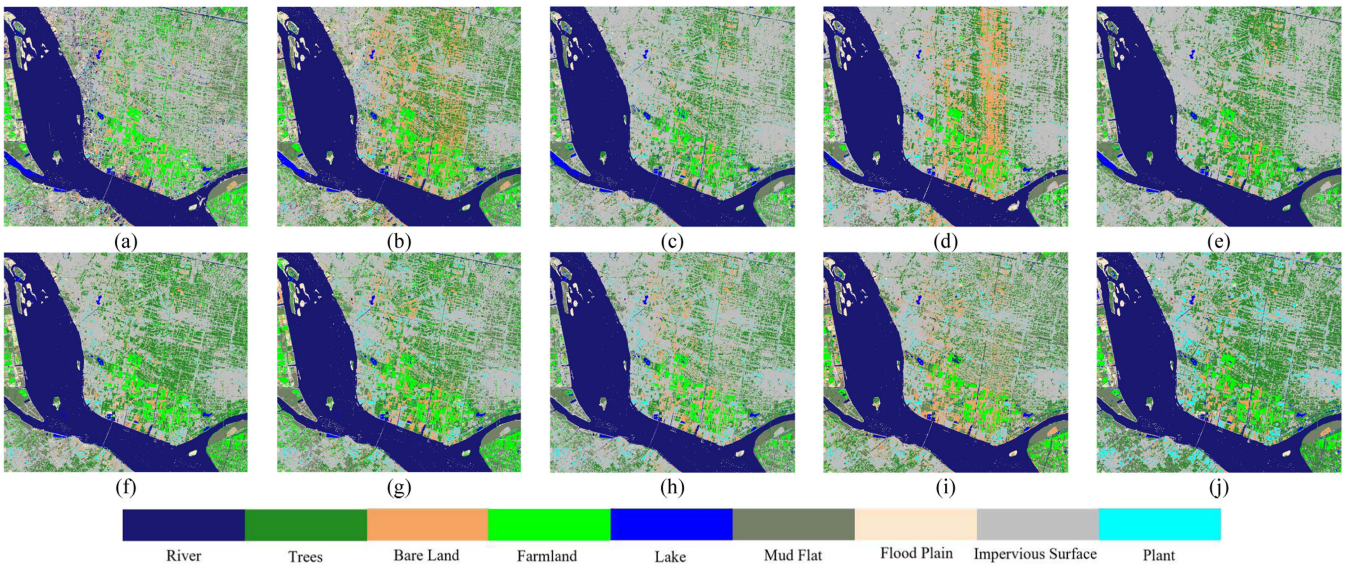

Fig. 10. Classification maps of Yangtze River Delta. (a) ELM. (b) 3-D-CNN. (c) CRC. (d) JCRC. (e) TCRC. (f) KProCRC. (g) KProCRC-LNSAE. (h) MWCRC. (i) JMWCRC. (j) JMWCRC-SA.

TABLE VII
CLASSIFICATION ACCURACY FOR THE DAFENG NATURAL RESERVE DATASET

Class	Train/Test	ELM	3D-CNN	CRC	JCRC	TCRC	KProCRC	KProCRC -LNSAE	MWCRC	JMWCRC	JMWCRC-SA
1	10/1010	99.97	100.00	99.99	99.94	99.81	99.99	100.00	99.87	100.00	100.00
2	10/518	89.66	94.11	92.64	90.24	90.86	93.22	98.97	94.21	95.86	98.36
3	10/140	41.79	71.22	45.09	37.52	42.60	61.90	72.09	57.48	78.74	83.18
4	10/79	99.85	100.00	94.52	98.24	99.12	99.75	97.95	96.15	97.46	93.81
5	10/908	95.73	96.09	91.36	90.08	89.47	96.37	97.60	94.45	96.70	98.81
6	10/266	76.79	78.32	74.28	77.16	76.24	82.61	90.43	84.33	97.19	97.53
7	10/714	95.94	94.13	90.63	89.59	87.09	94.60	90.96	94.88	93.61	90.36
8	10/736	81.80	95.48	93.02	95.84	96.41	97.14	98.06	94.52	97.85	96.87
9	10/69	46.06	69.73	51.18	78.83	73.50	82.32	92.30	72.59	85.70	95.96
OA(%)		89.06	93.84	88.65	88.82	88.88	93.68	95.46	92.44	95.93	96.30
AA(%)		81.95	88.79	81.41	84.16	83.90	89.77	93.15	87.61	93.68	94.99
Kappa(%)		87.01	92.62	86.48	86.64	86.70	92.46	94.57	91.00	95.14	95.57
Time(s)		14.83	25.27	10.75	14.30	14.76	30.42	88.22	12.76	25.31	33.31

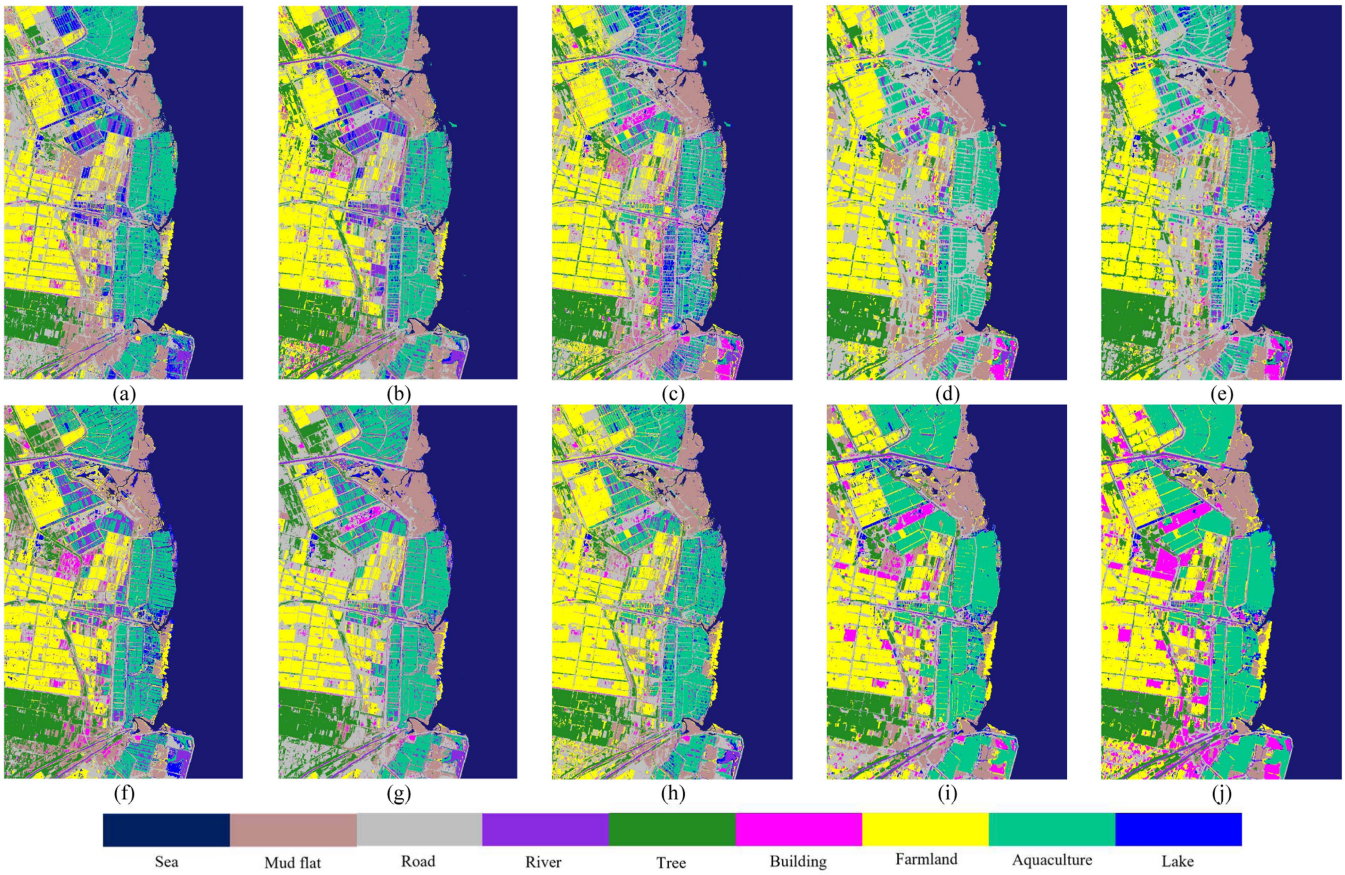

Fig. 11. Classification maps of Dafeng Natural Reserve. (a) ELM. (b) 3-D-CNN. (c) CRC. (d) JCRC. (e) TCRC. (f) KProCRC. (g) KProCRC-LNSAE. (h) MWCRC. (i) JMWCRC. (j) JMWCRC-SA.

the classification accuracy is only second to KProCRC-LNSA, which also shows that the model has a better differentiation for spectra. Combined with Fig. 10, it can be seen that trees and impervious surfaces are highly susceptible to misclassification with bare land when classified using the classifiers ELM, 3-D-CNN, CRC, JCRC, and TCRC. This is due to the complexity of the features resulting in the extreme similarity between the spectra. According to the classification graph of JMWCRC, it can be seen that trees, impervious surface and bare land can be well distinguished from each other.

For Dafeng Natural Reserve data, the overall classification accuracy of MWCRC was 92.44%. The highest classification accuracy was obtained by JMWCRC-SA with an OA value of 96.30%. Compared with MWCRC, the classification accuracy is improved by 3.86%. MWCRC classification accuracy is slightly lower than KProCRC and KProCRC-LNSAE, but compared

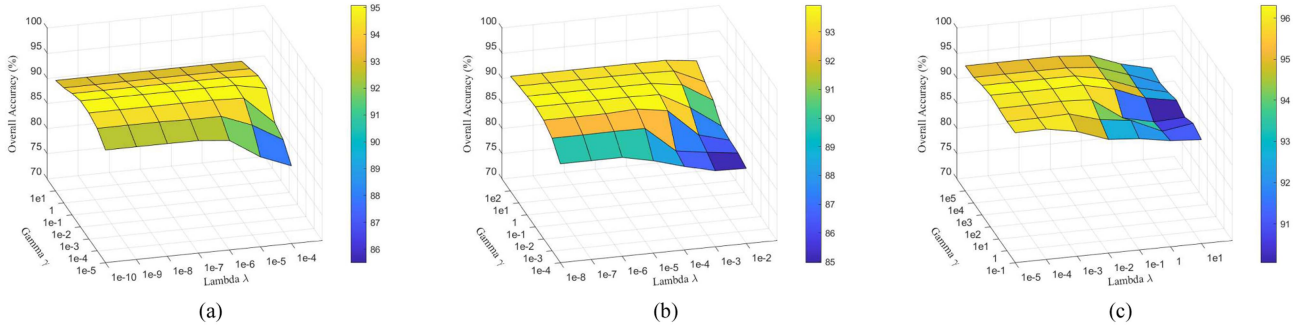

Fig. 12. Classification OA versus λ and γ. (a) Salinas. (b) Yangtze River Delta. (c) Dafeng Natural Reserve.

with other classifiers such as ELM, 3-D-CNN, and TCRC, there is a significant improvement in accuracy. This may be due to the high spectral similarity between features, and it is more difficult to rely only on spectral features for class separation. Combined with Fig. 11(a)–(e), misclassification between buildings and road and between aquaculture and lake are the main difficulties. As can be seen in Fig. 11(j), the incorporation of multiscale spatial information well increases the distinguishability between similar features.

In this article, the running time required to classify the testing samples has been calculated. Comprehensive Tables V–VII show that the running time may be increased compared to more traditional classifiers. However, compared to the more popular classifiers, such as KProCRC, it has more desirable classification results at almost the same running time. This also corresponds to the advantage of the CR model, where each pixel can be solved directly by a closed-form solution. KProCRC-LNSAE, on the other hand, mainly adopts the idea of ensemble learning, and thus its execution is more time-consuming.

V. Discussion

A. Analysis of Balance Parameters λ and γ

λ and γ are two hyperparameters in the MWCRC model, which are used to balance the residual and coefficient terms. The classification accuracy is optimized by manually tuning their values. λ is used to regulate the mean vector term and γ is used to regulate the weighted sparse term. These two terms are based on the case where the residuals are minimized, forcing the coefficients to have their main contribution come from the true class while increasing the stability of the model. In this article, the OA value of JMWCRC-SA is utilized as the performance discriminant criterion. Based on the optimal empirical parameters of the model known from Table IV, the experimental results of three groups of experimental data are generated using the grid search method, as shown in Fig. 12. Each set of data is subjected to 7×7 experiments.

First, after the OA value reaches the highest, it shows smooth fluctuation with the change of parameters in a certain region. This means that the model has better stability and generalization. Second, the model is slightly more sensitive to the parameters in the Yangtze River Delta dataset compared to the other two datasets. This may be related to the higher complexity of the terrain in the data, when the optimal parameters are kept at λ less than $1e$-3 and γ greater than $1e$-2. Finally, in all three data sets, the classification accuracy first increases and then smoothes as λ decreases, which may be due to the fact that the mean vector term can be very sensitive to the parameters when the strength of the sparse constraint is small. The classification accuracy first increases and then smooths as γ increases. This may be due to the fact that when the sparse constraints on the coefficients are fixed, increasing with γ will rely on the mean vector to attenuate the noise effect and further increase the sparse strength. Based on the range of parameter adjustment of λ and γ, it can be seen that the contribution of weighted sparse constraints in the model is slightly higher than that of mean vector constraints.

B. SE Size Range

Fig. 13 shows the results of the variation of the overall classification accuracy of JMWCRC-SA with the range of SE sizes, where x denotes the maximum SE size when the EMP operation is performed. Thus, the horizontal coordinates in the figure indicate the size range of the SEs. For Salinas data, the accuracy tends to grow steadily as the size range of the SEs increases. When the size is larger than 23 (pixels), the accuracy growth becomes slow. This is because the data contains a variety of large-sized features, and as the SE size increases, it keeps approaching the size of the features, which leads to increasing accuracy. For the other two data sets, the significant increase in accuracy is reflected in the SE size range in the interval of 3–7. As the size range continues to increase, the accuracy stays within a certain level without significant fluctuations. This is due to the small size of the features in these two data sets, with matching SEs around 7 pixels. When the size of the SE continues to increase, there is no significant decrease in accuracy. This is due to the fact that the proposed method adopts multiscale extraction of spatial information, and the multichannel mutual constraints prevent the problem of missing boundaries caused by too large SEs.

C. Performance on Small Size Samples Situation

In order to verify the classification performance of the proposed method with small samples, a comparison test with

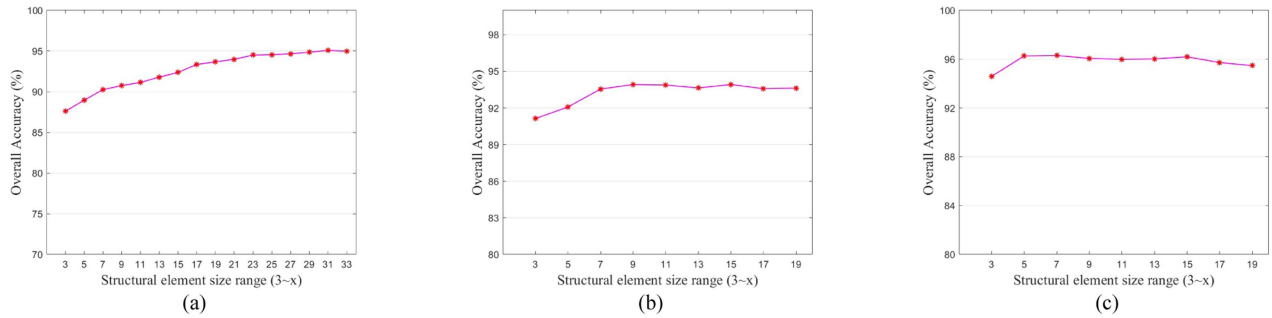

Fig. 13. Classification OA versus x. (a) Salinas. (b) Yangtze River Delta. (c) Dafeng Natural Reserve.

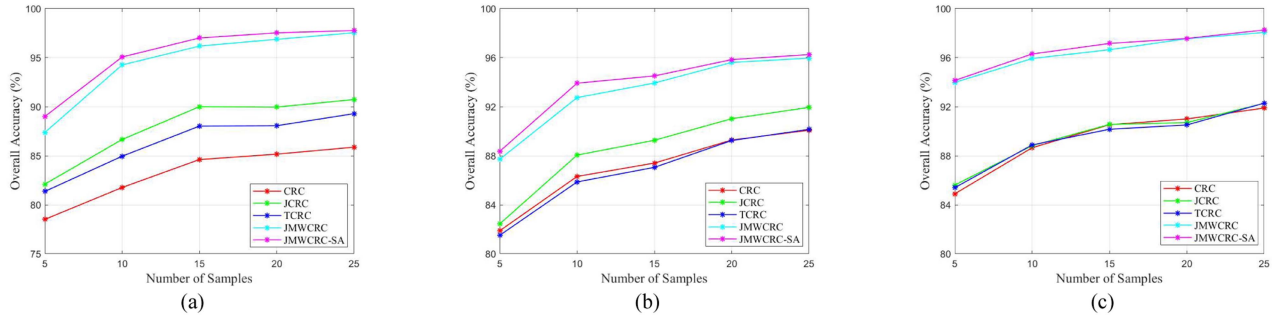

Fig. 14. Classification OA versus number of samples per class. (a) Salinas. (b) Yangtze River Delta. (c) Dafeng Natural Reserve.

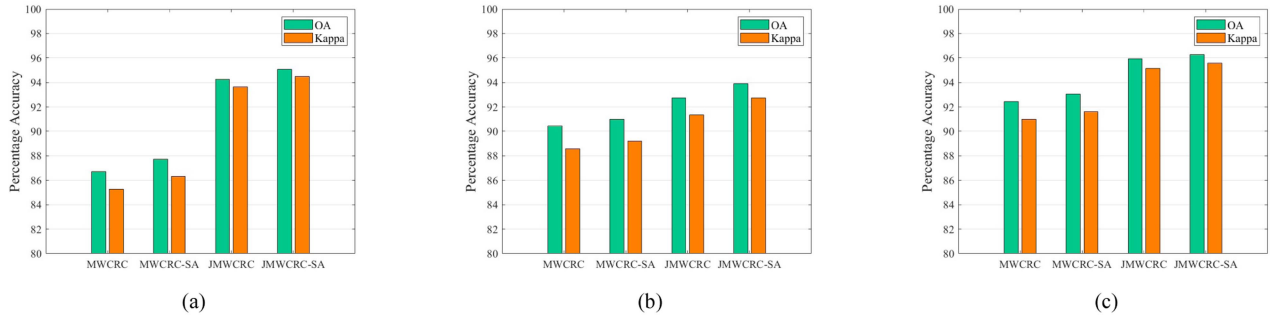

Fig. 15. Classification Accuracy versus model per class. (a) Salinas. (b) Yangtze River Delta. (c) Dafeng Natural Reserve.

the number of samples is designed. The OA value is used as the performance discrimination criterion, and the number of samples in each class is set as {5,10,15,20,25}. The experimental results are shown in Fig. 14. Combining the results of the three sets of experiments, with the change in the number of samples, the classification accuracies of JMWCRC and JMWCRC-SA are higher than the other classifiers, which also proves that the models have better stability and generalization. Compared to CRC, TCRC, and JCRC, JCRC produces the highest classification accuracy in Salinas data and Yangtze River Delta data, but there is no performance significant difference in Dafeng Natural Reserve data. Different spectral complexity and spatial complexity in these data may result in unstable classification results. Comparing the classification accuracies of JMWCRC and JMWCRC-SA shows that sample augmentation consolidates the stability of the model and further improves the classification accuracy.

D. Ablation Study

The proposed JMWCRC-SA framework consists of three main components: classification model, spatial information extraction, and sample augmentation. To evaluate the contribution of each part, this section removes each component of the JMWCRC-SA framework one by one for ablation experiments. A total of four models were generated by ablation.

1) MWCRC: contains only the classification model.
2) MWCRC-SA: spatial information is removed.
3) JMWCRC: sample augmentation is removed.
4) JMWCRC-SA.

The experimental results for the ablation models are given in Fig. 15, each using the average of the 10 sets of OA and Kappa as the performance discriminant criterion. As can be seen from the figure, the classification performance of JMWCRC-SA is significantly better than other variants. The joint use

of spatial and spectral data has the highest contribution in all three data sets, and the improvement in accuracy is particularly evident in the Salinas data, which has a large feature size. By comparing (1) with (2) and (3) with (4), it can be found that the sample augmentation method improves both OA and Kappa, indicating that the stability of the model has been consolidated.

VI. Conclusion

This article proposed MWCRC and JMWCRC-SA methods for HSI classification. MWCRC constrains the CRC by mean vector and representation weighting to increase the generalization of the model. To alleviate the small sample problem, the EMP method is introduced into MWCRC, and a sample augmentation method based on residual minimization is designed. The experimental results show that the classification performance of JMWCRC-SA under small sample conditions has desirable superiority.

Since the distribution of augmented samples is limited by the training samples, the improvement in classification accuracy after sample augmentation is not significant. Therefore, both the highest confidence level and the spatially uniform distribution of samples will be considered in future studies to obtain a more effective sample augmentation effect.

References

[1] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022, doi: 10.1109/MGRS.2021.3105440.

[2] B. Tu, X. Yang, C. Zhou, D. He, and A. Plaza, "Hyperspectral anomaly detection using dual window density," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8503–8517, Dec. 2020, doi: 10.1109/TGRS.2020.2988385.

[3] Y. Yang, H. Su, Z. Wu, and Q. Du, "Saliency-guided collaborative-competitive representation for hyperspectral anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6843–6859, 2023, doi: 10.1109/JSTARS.2023.3296876.

[4] Y. Peng, M. Fan, J. Song, T. Cui, and R. Li, "Assessment of plant species diversity based on hyperspectral indices at a fine scale," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018, doi: 10.1038/s41598-018-23136-5.

[5] L. Sun, S. Khan, and P. Shabestari, "Integrated hyperspectral and geochemical study of sediment-hosted disseminated gold at the Goldstrike district, Utah," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 1987, doi: 10.3390/rs11171987.

[6] X. Wang and W. Yang, "Water quality monitoring and evaluation using remote-sensing techniques in China: A systematic review," *Ecosyst. Health Sustainability*, vol. 5, no. 1, pp. 47–56, 2019, doi: 10.1080/20964129.2019.1571443.

[7] B. Du et al., "Mapping wetland plant communities using unmanned aerial vehicle hyperspectral imagery by comparing object/pixel-based classifications combining multiple machine-learning algorithms," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8249–8258, 2021, doi: 10.1109/JSTARS.2021.3100923.

[8] Y. Liang, H. Zheng, G. Yang, Q. Du, and H. Su, "Superpixel-based weighted sparse regression and spectral similarity constrained for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6825–6842, 2023, doi: 10.1109/JSTARS.2023.3298491.

[9] B. Tu, X. Zhang, X. Kang, G. Zhang, and S. Li, "Density peak-based noisy label detection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1573–1584, Mar. 2019, doi: 10.1109/TGRS.2018.2867444.

[10] J. Yang, B. Du, and L. Zhang, "From center to surrounding: An interactive learning framework for hyperspectral image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 197, no. 9, pp. 145–166, 2023, doi: 10.1016/j.isprsjprs.2023.01.024.

[11] H. Su, W. Yao, Z. Wu, P. Zheng, and Q. Du, "Kernel low-rank representation with elastic net for China coastal wetland land cover classification using GF-5 hyperspectral imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 171, no. 11, pp. 238–252, 2021, doi: 10.1016/j.isprsjprs.2020.11.018.

[12] H. Su, Y. Yu, Z. Wu, and Q. Du, "Random subspace-based k-nearest class collaborative representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6840–6853, Aug. 2021, doi: 10.1109/TGRS.2020.3029578.

[13] H. Su, Y. Yu, Q. Du, and P. Du, "Ensemble learning for hyperspectral image classification using tangent collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3778–3790, Jun. 2020, doi: 10.1109/TGRS.2019.2957135.

[14] K. Abend and T. J. Harley, "Comments 'on the mean accuracy of statistical pattern recognizers'," *IEEE Trans. Inf. Theory*, vol. IT- vol. 15, no. 3, pp. 420–423, May 1969, doi: 10.1109/TIT.1969.1054314.

[15] Y. Wang, D. Li, H. Wu, X. Li, F. Kong, and Q. Wang, "Multiple spectral-spatial representation based on tensor decomposition for HSI anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3539–3551, 2022, doi: 10.1109/JSTARS.2022.3170057.

[16] A. C. Karaca, "Spatial aware probabilistic multi-kernel collaborative representation for hyperspectral image classification using few labelled samples," *Int. J. Remote Sens.*, vol. 42, no. 3, pp. 839–864, 2021, doi: 10.1080/01431161.2020.1823516.

[17] C. Zheng, N. Wang, and J. Cui, "Hyperspectral image classification with small training sample size using superpixel-guided training sample enlargement," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7307–7316, Oct. 2019, doi: 10.1109/TGRS.2019.2912330.

[18] F. Fang et al., "Spatial context-aware method for urban land use classification using street view images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 192, no. 12, pp. 1–12, 2022, doi: 10.1016/j.isprsjprs.2022.07.020.

[19] H. Liu, W. Li, X. G. Xia, M. Zhang, C. Z. Gao, and R. Tao, "Central attention network for hyperspectral imagery classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8989–9003, Nov. 2022, doi: 10.1109/TNNLS.2022.3155114.

[20] B. Tu, C. Zhou, X. Liao, Q. Li, and Y. Peng, "Feature extraction via 3-D block characteristics sharing for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10503–10518, Dec. 2021, doi: 10.1109/TGRS.2020.3042274.

[21] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019, doi: 10.1109/TGRS.2018.2865102.

[22] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019, doi: 10.1109/TNNLS.2018.2874432.

[23] W. Li and Q. Du, "A survey on representation-based classification and detection in hyperspectral remote sensing imagery," *Pattern Recognit. Lett.*, vol. 83, pp. 115–123, 2016, doi: 10.1016/j.patrec.2015.09.010.

[24] 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, doi: 10.1109/TGRS.2004.831865.

[25] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019, doi: 10.1109/TGRS.2019.2907932.

[26] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Feb. 2016, doi: 10.1109/MGRS.2016.2540798.

[27] C. Wang and F. Wang, "GIS-automated delineation of hospital service areas in Florida: From Dartmouth method to network community detection methods," *Ann. GIS*, vol. 28, no. 2, pp. 93–109, 2022, doi: 10.1080/19475683.2022.2026470.

[28] Z. Liu, Y. Lin, J. Hoover, D. Beene, P. H. Charley, and N. Singer, "Individual level spatial-temporal modelling of exposure potential of livestock in the Cove Wash watershed, Arizona," *Ann. GIS*, vol. 29, no. 1, pp. 87–107, 2023, doi: 10.1080/19475683.2022.2075935.

[29] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011, doi: 10.1109/TGRS.2011.2129595.

[30] B. Xie, S. Mei, G. Zhang, Y. Zhang, Y. Feng, and Q. Du, "Extended collaborative representation-based hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6007905, doi: 10.1109/LGRS.2022.3159280.

[31] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009, doi: 10.1109/TPAMI.2008.79.
[32] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging : Sensing, reconstruction, and target detection," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 116–126, Jan. 2014, doi: 10.1109/MSP.2013.2279507.
[33] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014, doi: 10.1109/JSTARS.2013.2264720.
[34] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 471–478, doi: 10.1109/ICCV.2011.6126277.
[35] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 477–489, Jan. 2014, doi: 10.1109/TGRS.2013.2241773.
[36] J. Li, H. Zhang, Y. Huang, and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707–3719, Jun. 2014, doi: 10.1109/TGRS.2013.2274875.
[37] W. Li, Q. Du, and M. Xiong, "Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 48–52, Jan. 2015, doi: 10.1109/LGRS.2014.2325978.
[38] H. Su, Y. Gao, and Q. Du, "Superpixel-based relaxed collaborative representation with band weighting for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5525416, doi: 10.1109/TGRS.2022.3161139.
[39] H. Chi, H. Xia, L. Zhang, C. Zhang, and X. Tang, "Competitive and collaborative representation for classification," *Pattern Recognit. Lett.*, vol. 132, pp. 46–55, Apr. 2020, doi: 10.1016/j.patrec.2018.06.019.
[40] Y. Xu, Q. Du, W. Li, and N. H. Younan, "Efficient probabilistic collaborative representation-based classifier for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1746–1750, 2019, doi: 10.1109/LGRS.2019.2906839.
[41] J. Gou, X. He, J. Lu, H. Ma, W. Ou, and Y. Yuan, "A class-specific mean vector-based weighted competitive and collaborative representation method for classification," *Neural Netw.*, vol. 150, pp. 12–27, Jun. 2022, doi: 10.1016/j.neunet.2022.02.021.
[42] Z. Huang, H. Tang, Y. Li, and W. Xie, "HFC-SST: Improved spatial-spectral transformer for hyperspectral few-shot classification," *J. Appl. Remote Sens.*, vol. 17, no. 02, pp. 1–22, 2023, doi: 10.1117/1.jrs.17.026509.
[43] H. Pan, M. Liu, H. Ge, and S. Chen, "Semi-supervised spatial–spectral classification for hyperspectral image based on three-dimensional Gabor and co-selection self-training," *J. Appl. Remote Sens.*, vol. 16, no. 02, pp. 1–19, 2022, doi: 10.1117/1.jrs.16.028501.
[44] M. A. Moharram and D. M. Sundaram, "Spatial–spectral hyperspectral images classification based on krill herd band selection and edge-preserving transform domain recursive filter," *J. Appl. Remote Sens.*, vol. 16, no. 04, pp. 1–29, 2022, doi: 10.1117/1.jrs.16.044508.
[45] H. Su, B. Zhao, Q. Du, P. Du, and Z. Xue, "Multifeature dictionary learning for collaborative representation classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2467–2484, Apr. 2018, doi: 10.1109/TGRS.2017.2781805.
[46] Y. Gao, H. Su, H. Lu, and Q. Du, "Self-balancing dictionary learning for relaxed collaborative representation of hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539918, doi: 10.1109/TGRS.2022.3211209.
[47] H. Su, F. Shao, Y. Gao, H. Zhang, W. Sun, and Q. Du, "Probabilistic collaborative representation based ensemble learning for classification of wetland hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509517, doi: 10.1109/TGRS.2023.3267638.
[48] J. Hu, X. Shen, H. Yu, X. Shang, Q. Guo, and B. Zhang, "Extended subspace projection upon sample augmentation based on global spatial and local spectral similarity for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8653–8664, 2021, doi: 10.1109/JSTARS.2021.3107105.
[49] C. Li, Y. Ma, X. Mei, C. Liu, and J. Ma, "Hyperspectral image classification with robust sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 641–645, May 2016, doi: 10.1109/LGRS.2016.2532380.
[50] X. Chen, S. Li, and J. Peng, "Hyperspectral imagery classification with multiple regularized collaborative representations," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1121–1125, Jul. 2017, doi: 10.1109/LGRS.2017.2699667.
[51] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005, doi: 10.1109/TGRS.2004.842478.
[52] Y. Ma, C. Li, H. Li, X. Mei, and J. Ma, "Hyperspectral image classification with discriminative kernel collaborative representation and Tikhonov regularization," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 587–591, Apr. 2018, doi: 10.1109/LGRS.2018.2800080.
[53] B. Xie, Y. Zhang, S. Mei, G. Zhang, Y. Feng, and Q. Du, "Spectral variation augmented representation for hyperspectral imagery classification with few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5543212, doi: 10.1109/TGRS.2022.3220579.
[54] M. Yang, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2224–2231. doi: 10.1109/CVPR.2012.6247931.
[55] H. Su, Y. Cai, and Q. Du, "Firefly-algorithm-inspired framework with band selection and extreme learning machine for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 309–320, 2017, doi: 10.1109/JSTARS.2016.2591004.
[56] J. Yang and J. Qian, "Hyperspectral image classification via multiscale joint collaborative representation with locally adaptive dictionary," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 112–116, Jan. 2018, doi: 10.1109/LGRS.2017.2776113.
[57] H. Su, B. Zhao, Q. Du, and Y. Sheng, "Tangent distance-based collaborative representation for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1236–1240, Sep. 2016, doi: 10.1109/LGRS.2016.2578038.
[58] J. Liu, Z. Wu, J. Li, A. Plaza, and Y. Yuan, "Probabilistic-kernel collaborative representation for spatial-spectral hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2371–2384, Apr. 2016, doi: 10.1109/TGRS.2015.2500680.

Hongjun Su (Senior Member, IEEE) received the Ph.D. degree in cartography and geography information system from the Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, Nanjing, China, in 2011.

He is currently a Full Professor with the School of Earth Sciences and Engineering, Hohai University, Nanjing, China. His main research interests include hyperspectral remote sensing dimensionality reduction, classification, and spectral unmixing.

Dr. Su is an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Dezhong Shi received the B.E. degree in surveying and mapping engineering from the Shandong University of Science and Technology, Taian, China, in 2022. He is currently working toward the M.E. degree in surveying and mapping with the School of Earth Sciences and Engineering, Hohai University, Nanjing, China.

His research interests include hyperspectral remote sensing image processing, collaborative representation, and hyperspectral machine learning classification.

Zhaohui Xue (Member, IEEE) received the Ph.D. degree in cartography and geographic information system from Nanjing University, Nanjing, China, in 2015.

He is currently a Professor with the School of Earth Sciences and Engineering, Hohai University, Nanjing, China. He has authored more than 50 scientific papers. His research interests include hyperspectral image classification, time-series image analysis, pattern recognition, and machine learning.

Dr. Xue was the recipient of the Best Reviewer for the IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY. He is an Editorial Board Member with National Remote Sensing Bulletin (2020–2024).

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of the SPIE-International Society for Optics and Photonics. From 2016–2020, she was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She currently is a member of IEEE Periodicals Review and Advisory Committee.