







Subfeature Ensemble-Based Hyperspectral Anomaly Detection Algorithm

Shuo Wang , Graduate Student Member, IEEE, Wei Feng , Yinghui Quan , Member, IEEE, Wenxing Bao, Gabriel Dauphin , Lianru Gao , Senior Member, IEEE, Xian Zhong, and Mengdao Xing , Fellow, IEEE

Abstract—Hyperspectral images (HSIs) have always played an important role in remote sensing applications. Anomaly detection has become a hot spot in HSI processing in recent years. The popular detecting method is to accurately segment anomalies from the background. Informative bands are very important for the accuracy improvement of the detection technology. However, most of the abnormal targets segmentation methods focus on the usage of all the spectral features, thus are easily affected by redundant bands or feature noise. A hyperspectral anomaly detection algorithm based on subfeature ensemble is proposed in this article. The proposed method consists of the following steps. First, the bands of the original HSI are normalized and randomly divided into several subfeature sets according to different proportions. Second, six methods including the prior-based tensor approximation algorithm (PTA), Reed–Xiaoli method, a low-rank and sparse representation method, a low-rank and sparse matrix decomposition-based Mahalanobis distance method, the graph and total variation regularized low-rank representation-based method, and a method based on tensor principal component analysis are applied to detect anomalies on the original HSI, and the method with the best performance is used to obtain an enhanced feature set. Then, the enhanced features and the subfeatures are ensembled iteratively to construct a new dataset. Finally, the PTA method is operated on the dataset with ensemble features to get the final abnormal target results. Six hyperspectral datasets are used in the experiment. Seven methods are employed as comparisons. The results are analyzed from both qualitative and quantitative perspectives. Extensive experimental results illustrate that the proposed method performs best on all datasets.

Index Terms—Anomaly detection, hyperspectral images (HSIs), feature ensemble, unsupervised object detection, remote sensing.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) can provide rich spectral and spatial information [1] and are highly valued by remote sensing scholars and scientists around the world [2]. Compared with the traditional remote sensing data, HSIs have strong practicability and can be used in many aspects such as public safety [3], [4], environmental detection [5], [6], image classification [7], [8], city planning [9], food hygiene [10], and identification of geological rocks and mines [11]. Anomaly detection [12]–[14] in HSI is an essential technique to find out the anomalies in the region of interests (ROI), especially in the case that anomalies are weak [15].

Anomaly detection is considered as a binary classifier, which aims to label pixels as anomalous or background based on their spectral characteristics [16]. Generally speaking, target detection [17]–[19] can be divided into two categories, supervised and unsupervised, according to the existence of prior object information. Unsupervised target detection is also called anomaly detection and works without prior object information [20]. However, in the presence of prior object information, supervised target detection often achieves better performance than anomaly detection. But in practical situations, there is often not enough prior information, so anomaly detection has more research value than target detection. Over the past years, abundant anomaly detection algorithms have been proposed. One of the classical anomaly detection methods is Reed–Xiaoli (RX) [21]. The core idea of RX is to calculate the Mahalanobis distance of each pixel between the target and the background [22]. In order to obtain the distance, the covariance matrix and the mean vector of the background are inferred. Depending on the pixel range, the classic RX is divided into two categories, global-RX detection (GRXD) and local-RX detection (LRXD) [23]. GRXD estimates the background with all pixels in the HSI, while LRXD with neighbor pixels. However, in the real world, the image background usually contains complex objects. And RX is established on the basis of setting the background as a multivariate normal Gaussian distribution [24]. The Gaussian distribution cannot realistically simulate the complex features of the HSI [25]. In addition, some anomalies and noises cannot be eliminated, which may contaminate the background and reduce its detection performance [26]. Therefore, the false alarm rate of RX is relatively high.

Manuscript received 7 April 2022; revised 2 June 2022 and 8 July 2022; accepted 12 July 2022. Date of publication 20 July 2022; date of current version 1 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61772397 and Grant 12005159; in part by the Natural Science Basic Research Program of Shaanxi under Grant 2021JC-23; and in part by the Science Technology and Development Project of Yulin Science and Technology Bureau under Grant CXY-2020-094. (Shuo Wang and Wei Feng contribute equally to this work.) (Corresponding authors: Wei Feng; Yinghui Quan; Wenxing Bao.)

Shuo Wang, Wei Feng, Yinghui Quan, and Xian Zhong are with the Department of Remote Sensing Science and Technology, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: shuow@stu.xidian.edu.cn; wfeng@xidian.edu.cn; yhqun@mail.xidian.edu.cn; xzhong7@stu.xidian.edu.cn).

Wenxing Bao is with the School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China (e-mail: baowenxing@nun.edu.cn).

Gabriel Dauphin is with the Laboratory of Information Processing and Transmission, L2TI, Institut Galilée, University Paris XIII, 93430 Villetaneuse, France (e-mail: gabriel.dauphin@univ-paris13.fr).

Lianru Gao is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: gaolr@aircas.ac.cn).

Mengdao Xing is with the Academy of Advanced Interdisciplinary Research, Xidian University, Xi'an 710071, China (e-mail: xmd@xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3191725

To remedy this defect, some variants of the RXD have been proposed. For example, the weighted RXD method is proposed to reduce the effect of anomalies on the covariance matrix when estimating the background [27]. The kernel-RXD [28] aims to change the dimension of all pixels through mapping pixels into high-dimensional feature space based on the kernel theory. This has a positive effect on distinguishing anomalous and background pixels. In addition, an improved version of KRX with a higher computational efficiency is proposed by Khazai and Mojaradi [29]. Unlike the kernel-RXD, the clustering-based anomaly detection (CBAD) method [30] detects anomalies by segmenting the HSI into several clusters. In order to reduce the influence of background elements on extracting anomalies, Zhao and Zhang [31] proposed to use k-means clustering method to segment HSI. And the principal component analysis (PCA) method is adopted to preprocess the HSI to suppress the influence of noise. However, although these methods are proven to be effective in improving the RX performance, the core problem is not solved. The distribution model of their background is not changed. To avoid modeling, the distribution of the background, some anomaly detection methods for real-environment HSIs are proposed.

Because the background pixels are similar, the abnormal pixels are different, scholars assume that a few atoms in a sparse dictionary can represent background pixels well [16]. Thus, sparse representation (SR) [32] is proposed. The advantage of SR-based methods is that no assumptions need to be established for the statistical data distribution of the HSI dataset [33]. Inspired by SR, low-rank and sparse matrix decomposition (LSDM) [34] is proposed based on the concept that there are so few pixel categories in HSI scenes. In 2016, an anomaly detection method based on the low-rank sparse matrix factorization technique was proposed by Zhang *et al.* [26]. The theory decomposes the HSI matrix into three parts: low-rank matrix, sparse matrix, and noise matrix, which are processed separately to obtain detection results. In 2019, in order not to destroy the spatial structure of the HSI, Cheng and Wang proposed an anomaly detection method based on graph and total variation regularized low-rank representation (GTVLRR) [35]. HSI is a 3-D data cube, which can be regarded as a third-order tensor in essence: one spectral dimension and two spatial dimensions. Most anomaly detection methods ignore this point. Therefore, tensors are introduced to maintain all dimensional structures, which has been applied in many ways [36]–[38]. Tensor decomposition-based anomaly detection (TDAD) [39], [40] and tensor principal component analysis anomaly detection (TPCA) [41] become research hotspots.

In recent years, generative adversarial networks are applied to anomaly detection [42], [43], which has shown great advantages in unsupervised application scenarios with complex data distribution. Li *et al.* proposed an unsupervised generative adversarial network with background spatial feature enhancement and redundancy pooling to solve the problem of redundant information interference [44].

Although these methods have demonstrated their positive effect on anomaly detection in related studies, they focus on the usage of all the spectral features when spatial information is

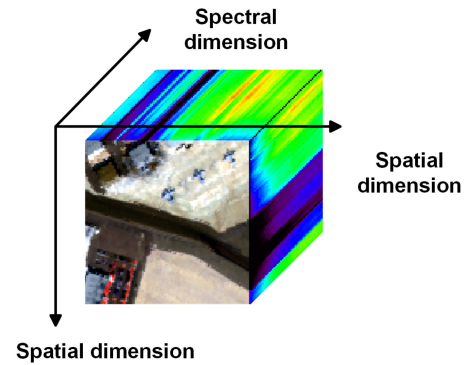


Fig. 1. Tensor explanation for HSI data.

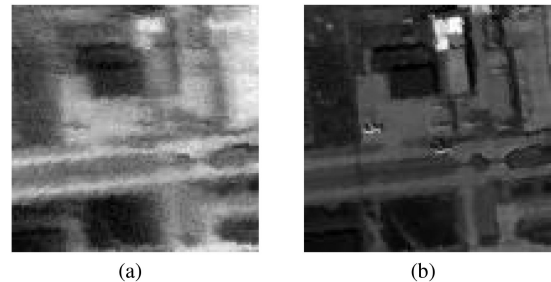


Fig. 2. Two band images of the Airport-1 dataset. (a) Noisy band. (b) Clean band.

not involved. This may affect the detection performance. First, as the number of spectra increases, too much information becomes redundant [45], and the time of processing the information increases. Second, the detection map is easily affected by feature noise [46]. In addition, the interference of noise will reduce the signal-to-noise ratio of the image [47]. Therefore, it is necessary to minimize the influence of noise and enhance the utilization of features. Wu and Liu [48] proposed an anomaly detection algorithm based on learning causal temporal relationships and feature discrimination. The causal temporal relationship (CTR) module is applied to capture local-scale temporal dependencies among features to enhance features. Zhao *et al.* [49] proposed an anomaly detection method based on stacked denoising autoencoders, and a spectral feature model was constructed to verify the effectiveness of the proposed algorithm.

Different from the aforementioned detectors, we proposed an anomaly detection algorithm based on subfeature ensemble (SED) that improves the detection accuracy by effectively using high-quality features, enhancing the difference between the abnormal part and the background. The contributions of the proposed algorithm are shown as follows.

- 1) The original spectral features of the HSI data are randomly selected to reduce the computation complexity of the proposed algorithm and reduce image noise.
- 2) Several diversity algorithms are employed to obtain the enhanced spectral features for the accuracy improvement of the detection algorithm.

This article is organized as follows. Section II introduces the PTA [50] algorithm and the symbols used in this article. Section III describes in detail the proposed methodology. Section IV

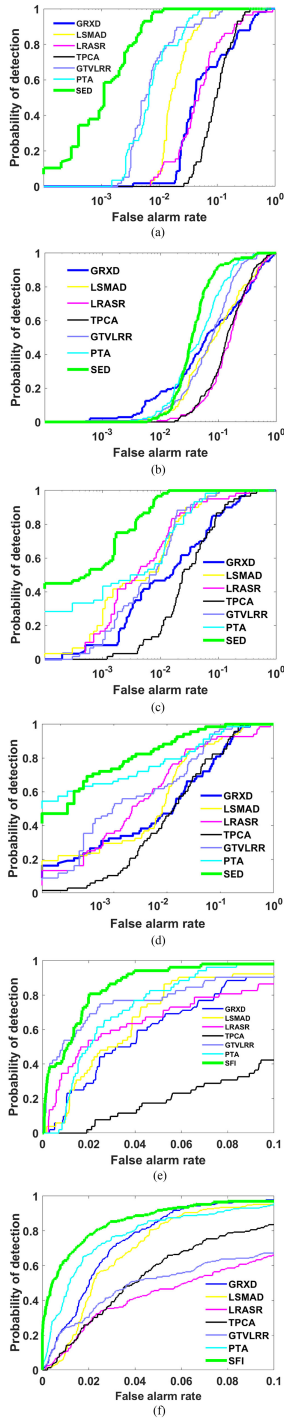


Fig. 3. ROC curves obtained by PTA, RX, LSMAD, LRASR, TPCA, GTVLRR, and the proposed method SED on six datasets. (a) San Diego. (b) Airport-1. (c) Airport-4. (d) Beach-4. (e) Urban-3. (f) Urban-5.

lists the results of the experiments and the discussion. Finally, Section V concludes this article.

II. RELATED WORK

PTA is a tensor-based anomaly detection method. Low-order, sparse, and segment-smooth prior information is added to its processing, which is helpful for the detection of objects of different sizes and has a wider application. In addition, the tensor

approximation allows both spectral and spatial dimensions to be considered, improving the detection accuracy. Assuming an original HSI Y 1 with H rows, W columns, and D spectral bands is denoted by $Y \in \mathbb{R}^{H \times W \times D}$.

Y can be divided into two parts: the background tensor X and the anomaly tensor S . The PTA method consists of the following steps. First, X is expanded along the spatial dimension to get X_1 and X_2 , respectively, then linear total variation norm regularization is applied to them. Second, X is expanded along the spectral dimension to form the matrix X_3 . Then, a novel truncated nuclear norm regularization is combined with X_3 . Third, the anomaly tensor is expanded as S_3 along its spectral dimension. An $l_{2,1}$ -norm regularization is applied on S_3 . Finally, PTA is calculated by the following equation:

$$\operatorname{argmin}_{\mathcal{X}, \mathcal{S}} \frac{1}{2} \left(\|\mathbf{D}_H \mathbf{X}_1\|_F^2 + \|\mathbf{D}_W \mathbf{X}_2\|_F^2 \right) + \alpha \|\mathbf{X}_3\|_r + \beta \|\mathbf{S}_3\|_{2,1} \quad (1)$$

$$\begin{cases} \mathcal{Y} = \mathcal{X} + \mathcal{S} \\ \mathcal{X}_1 = \operatorname{unfold}_1(\mathcal{X}) \\ \mathcal{X}_2 = \operatorname{unfold}_2(\mathcal{X}) \\ \mathcal{X}_3 = \operatorname{unfold}_3(\mathcal{X}) \\ \mathcal{S}_3 = \operatorname{unfold}_3(\mathcal{S}) \end{cases} \quad (2)$$

where $\mathbf{D}_H \in \mathbb{R}^{(H-1) \times H}$ and $\mathbf{D}_W \in \mathbb{R}^{(W-1) \times W}$ are defined as

$$\mathbf{D}_H = \begin{bmatrix} 1 & -1 & & & & \\ & 1 & -1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 1 & -1 \end{bmatrix} \quad (3)$$

$$\mathbf{D}_W = \begin{bmatrix} 1 & -1 & & & & \\ & 1 & -1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 1 & -1 \end{bmatrix}. \quad (4)$$

III. PROPOSED METHOD

SED is dedicated to enhancing the characteristics of abnormal parts of the HSI to improve the detection accuracy. The proposed method consists of three steps. First, the bands of the original HSI are randomly selected to obtain several subfeature sets. Second, six methods are applied to detect anomalies on the original HSI, and the best-performing algorithm is selected to obtain an enhanced feature set. Two feature sets are ensemble to construct the new dataset. Third, the PTA is applied with the new dataset to get the final abnormal target results. The algorithm flow is summarized in Algorithm 1.

A. Feature Selection

HSIs can provide rich spectral information for abnormal targets. However, feature redundancy and noise [51] of the original data (see Fig. 2) may lead to negative effects for the target

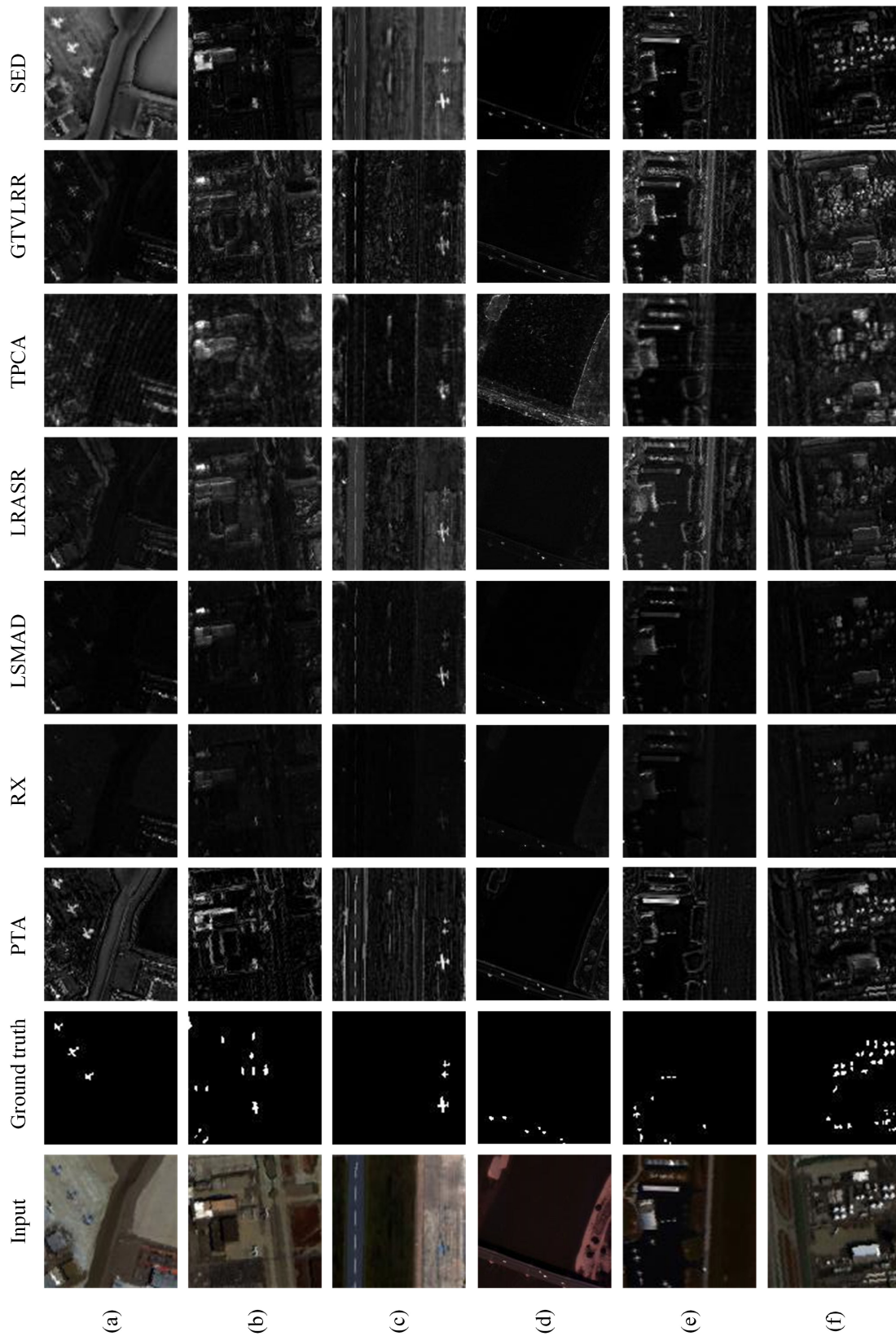


Fig. 4. Pseudocolor images (first column) and ground truth (second column). Detection results of PTA, RX, LSMAD, LRASR, TPCA, GTVLRR, and the proposed method SED, respectively on six datasets. (a) San Diego. (b) Airport-1. (c) Airport-4. (d) Beach-4. (e) Urban-3. (f) Urban-5.

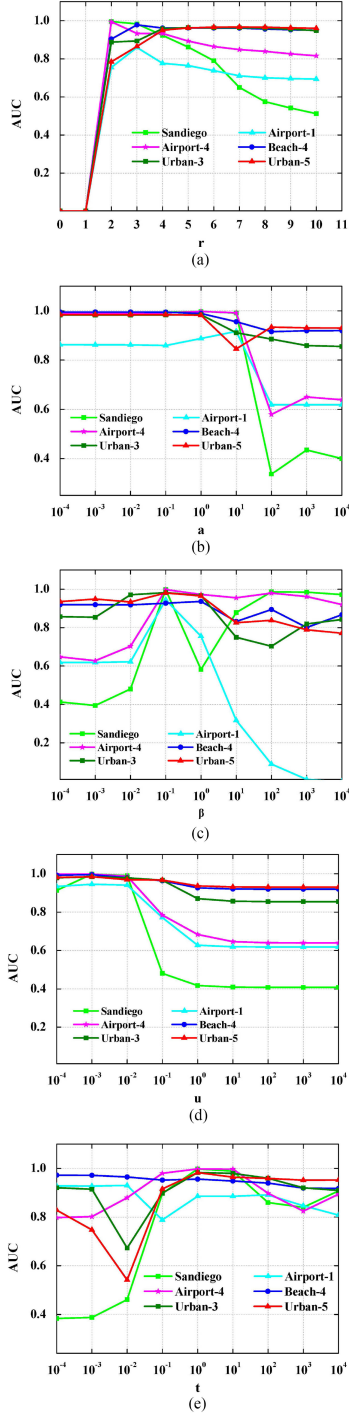


Fig. 5. Influence of parameters on the AUC scores on six datasets. The parameter γ is shown in (a), α is shown in (b), β is shown in (c), μ is shown in (d), and τ is shown in (e). (a) γ . (b) α . (c) β . (d) μ . (e) τ .

detection algorithm. In order to deal with the aforementioned problem, feature selection and enhancement are adopted to improve the detection accuracy.

Assuming that $\mathcal{Z} = \mathbb{R}^{H \times W \times D}$ is an HSI, it is stratified along the spectrum dimension as follows:

$$\mathcal{Z} = (B_1^{H \times W}, B_2^{H \times W}, \dots, B_D^{H \times W}) \quad (5)$$

where $B_i^{H \times W}$ represents the i th HSI layer with H rows and W columns.

Because there is often noise in collected hyperspectral data, if it is processed directly, the results will be affected.

In this article, some of the bands are selected to remain, and other bands that may contain noise are discarded. It has been illustrated that normalization is a necessary step to properly preprocess the dataset, which helps reduce unnecessary data bias and improves the accuracy of the analysis [52]. So, all D bands are normalized as follows:

$$B_i^{H \times W} = \frac{P - P_{\min}}{P_{\max} - P_{\min}} \quad (6)$$

where P represents the pixel value, P_{\max} represents the maximum pixel value, and P_{\min} represents the minimum pixel value.

In order to randomly divide the HSI into several subfeature sets according to different proportions, an array a containing random numbers is created. We set the length of the array a to be n , that is, a contains n random numbers. Because the numbers of bands included in various HSIs are different, and the proportions are also not the same, n changes accordingly. $a(k)$ represents the value of the k th element. In this article, the bands labeled $a(k)$ are selected as the subfeature set, denoted as \mathcal{Z}^O .

$$\mathcal{Z}^O = (B_{a(1)}^{H \times W}, B_{a(2)}^{H \times W}, \dots, B_{a(n)}^{H \times W}). \quad (7)$$

B. Feature Enhancement

To highlight the characteristics of the abnormal part, we especially add some results of anomaly detection. In this article, six algorithms including various detection principles are applied to perform anomaly detection on the original HSI. They are the classic method RX, a tensor representation method TPCA, and three LRASR methods (LRASR [25], LSMAD [26], and GTVLRR). Then, we can get six detection results. It is explained here that SED aims to propose a new framework to improve the detection accuracy of existing algorithms. And it can be extended by applying new algorithms, which can be done in future work. After analysis, the method with the best performance is retained. It is possible that two algorithms perform almost the same, then both of them will be selected. To accentuate the anomaly and reduce background influence, the number of the result layer is stacked to the tens to function.

The result layer is denoted as R_1 . The enhanced feature set is denoted as \mathcal{Z}^R , shown as follows:

$$\mathcal{Z}^R = (R_1^{H \times W}, \dots, R_1^{H \times W}). \quad (8)$$

Each \mathcal{Z}^R contains dozens of R_1 .

If we get two methods, the second result layer is denoted as R_2 . Then, \mathcal{Z}^R is set as follows:

$$\mathcal{Z}^R = (R_1^{H \times W}, \dots, R_1^{H \times W}, R_2^{H \times W}, \dots, R_2^{H \times W}). \quad (9)$$

C. Feature Ensemble

Ensemble learning can synthesize the advantages of each subset, and obtain results that are better than the individual [53].

Algorithm 1: Subfeature ensemble-based hyperspectral anomaly detection algorithm (SED).

Input: An HSI: \mathcal{Z} , the number of rows: H , the number of columns: W , and the number of spectral bands: D .

Process:

1: The HSI is normalized and stratified along the spectrum dimension: $B_i^{H \times W} = \frac{P - P_{\min}}{P_{\max} - P_{\min}}$,

$$\mathcal{Z} = (B_1^{H \times W}, B_2^{H \times W}, \dots, B_D^{H \times W}).$$

2: n bands are randomly selected as the sub-feature set $\mathcal{Z}^O = (B_{a(1)}^{H \times W}, B_{a(2)}^{H \times W}, \dots, B_{a(n)}^{H \times W})$.

3: Six algorithms including PTA, RX, LRASR, LSMAD, GTVLRR, and TPCA are applied to detect anomalies on the original HSI.

4: The method with the best performance is used to obtain an enhanced feature set $\mathcal{Z}^R = (R_1^{H \times W}, \dots, R_1^{H \times W})$.

5: Two kinds of features are ensembled as $\mathcal{Z}^* = \mathcal{Z}^O + \mathcal{Z}^R$.

6: PTA is applied to detect \mathcal{Z}^* to get the final abnormal map \mathcal{T} .

Output:

An anomaly detection map \mathcal{T} .

TABLE I
RATIO BETWEEN \mathcal{Z}^O AND \mathcal{Z}^R IS 3:1

Number of bands	20	40	60	80	100
San Diego	0.9980	0.9980	0.9978	0.9976	
Airport-1	0.9450	0.9442	0.9424	0.9442	0.9423
Airport-4	0.9976	0.9977	0.9978	0.9978	
Beach-4	0.9943	0.9949			
Urban-3	0.9801	0.9804	0.9834	0.9782	
Urban-5	0.9847	0.9810	0.9792	0.9817	0.9789

The best performers of all results are bolded.

The original HSI has the basic information of anomalies, and the enhanced feature set can highlight the anomalies and reduce the influence of the background. In order to obtain both effects at the same time, \mathcal{Z}^O and \mathcal{Z}^R are ensembled to obtain the new feature set, denoted as \mathcal{Z}^* .

$$\mathcal{Z}^* = \mathcal{Z}^O + \mathcal{Z}^R. \quad (10)$$

Because the PTA method has outstanding results in detection and the principle has been described previously. \mathcal{Z}^* is processed by the PTA, and the final result is obtained. In this article, PTA is applied twice. For the first time, the PTA is used to generate the result in the “feature enhancement” section like other algorithms. To clearly illustrate the algorithm flow, the overall description of the SED is shown in Algorithm 1.

IV. EXPERIMENTAL RESULTS

A. Datasets

Six sets of HSIs are applied to test the performance of the proposed algorithm.

The first HSI dataset is the San Diego dataset [25]. It was captured by an AVIRIS sensor in the San Diego Airport area in California, USA. The image consists of 100×100 pixels including 189 spectral channels. The spatial resolution is 3.5 m.

TABLE II
RATIO BETWEEN \mathcal{Z}^O AND \mathcal{Z}^R IS 3:2

Number of bands	20	40	60	80	100
San Diego	0.9973	0.9969	0.9974	0.9971	
Airport-1	0.9369	0.9408	0.9426	0.9412	0.9443
Airport-4	0.9980	0.9975	0.9977	0.9978	
Beach-4	0.9940	0.9948			
Urban-3	0.9827	0.9831	0.9826	0.9818	
Urban-5	0.9812	0.9798	0.9801	0.9817	0.9822

The best performers of all results are bolded.

TABLE III
RATIO BETWEEN \mathcal{Z}^O AND \mathcal{Z}^R IS 1:1

Number of bands	20	40	60	80	100
San Diego	0.9976	0.9970	0.9965	0.9968	
Airport-1	0.9423	0.9423	0.9430	0.9435	0.9447
Airport-4	0.9980	0.9977	0.9977	0.9976	
Beach-4	0.9949	0.9944			
Urban-3	0.9841	0.9841	0.9836	0.9833	
Urban-5	0.9823	0.9831	0.9785	0.9814	0.9812

The best performers of all results are bolded.

Three planes in the image are considered anomalous. The second and third datasets are from the airport-beach-urban (ABU) scenes, named Airport-1 and Airport-4 [54]. Airport datasets consist of 100×100 pixels. Airport-1 data contain 205 spectral channels and Airport-4 contains 191 bands, which were captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. Multiple aircraft in the dataset are considered anomalies. The fourth is beach data from the ABU scenes, named Beach-4. Beach-4 image consists of 150×150 pixels including 102 spectral channels, which were captured by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor. The anomalous targets are vehicles. The fifth and sixth datasets are Urban-3 and Urban-5 of the ABU scenes [54]. Urban datasets consist of 100×100 pixels. Urban-3 contains 191 spectral channels, and Urban-5 consists of 205 bands.

B. Evaluation Metrics

Two well-known methods, receiver operating characteristic (ROC) and area under curve value (AUC), are used to evaluate the performance of the anomaly detection algorithms. ROC [55] can reflect the relationship between detection rate (true positive rate, TPR) and false-positive rate (FPR), thus can comprehensively evaluate the effectiveness of the algorithm. The higher the curve, the better the algorithm performance. Moreover, the ROC curve could also analyze the performance of the algorithm from a qualitative point of view. AUC [56], the area under the ROC curve, can intuitively evaluate the pros and cons of the algorithm from a quantitative perspective. The larger the AUC value, the better, and the ideal value is 1.

C. Parameter Settings

First, about dictionary learning of LRASR and GTVLRR, the number of clusters K is 15 and the number of atoms in each cluster is 20. λ and β are set in range of 0.01–1. In GTVLRR,

TABLE IV
AUC VALUES AND CALCULATING TIME OF DIFFERENT METHODS ON SIX DATASETS

		PTA	RX	LSMAD	LRASR	TPCA	GTVLRR	SED
San Diego	AUC	0.9889	0.8886	0.9765	0.9144	0.8849	0.9881	0.9980
	Time/s	23.27	0.05	9.4	26.86	20.32	142.50	227.86
Airport-1	AUC	0.9243	0.8221	0.8277	0.7892	0.8023	0.8989	0.9450
	Time/s	44.10	0.06	21.40	30.10	9.07	144.11	257.15
Airport-4	AUC	0.9877	0.9526	0.9883	0.9789	0.9432	0.9864	0.9980
	Time/s	22.82	0.05	8.21	28.17	20.12	136.38	220.97
Beach-4	AUC	0.9831	0.9538	0.9681	0.9545	0.9583	0.9821	0.9949
	Time/s	40.67	0.04	10.13	81.03	18.42	321.41	486.37
Urban-3	AUC	0.9709	0.9513	0.9639	0.9493	0.8224	0.9714	0.9841
	Time/s	32.67	0.06	22.92	32.62	8.81	164.98	270.55
Urban-5	AUC	0.9741	0.9692	0.9598	0.9024	0.9370	0.9080	0.9847
	Time/s	33.57	0.06	22.81	30.47	9.08	155.25	258.35

The best performers of all results are bolded.

the hyperparameter $\lambda = 0.05$, $\beta = 0.2$, and $\gamma = 0.02$. In PTA, α and τ are set to 1 and $\beta = 0.01$. To simplify the algorithm, we set the ratio of the number of bands with the best results to the second-best algorithm to 2:1. The parameter settings of the six algorithms refer to [50].

Second, to improve the calculation efficiency, the number of \mathcal{Z}^O does not exceed half of the number of original HSI bands. In this article, the number of \mathcal{Z}^O is set to 20, 40, 60, 80, and 100. And three sets of ratio settings 3:1, 3:2, and 1:1 between \mathcal{Z}^O and \mathcal{Z}^R are applied. The AUC values of SED on all datasets are shown in the Tables I–III. It is confirmed that the best results can be achieved even when the number of \mathcal{Z}^R is the least. Therefore, through feature selection, the number of features is reduced, which has a great gain in computational efficiency.

D. Results and Analysis

The AUC scores of all the methods are shown in Table IV. Each AUC is the best result obtained after ten runs of the program. The best grades have been kept and highlighted in the Tables I–III. It can be seen that SED obtains the highest scores for all six datasets. We can see that as the number of bands increases, the value of AUC may become higher or lower. This is related to the feature importance of the input. In the case that the number of \mathcal{Z}^O is the same, if \mathcal{Z}^R is important, the AUC will become higher as the number of \mathcal{Z}^O increases. But if the added \mathcal{Z}^R is sufficient, increasing the number of \mathcal{Z}^R has no positive effect on the AUC. In the case that the number of \mathcal{Z}^R is the same, if the selected \mathcal{Z}^O is not enough, increasing its input will make the AUC higher. But if \mathcal{Z}^O is enough to identify anomalies, adding more original features will only increase the interference, affect the recognition results.

For the San Diego dataset, compared with PTA and GTVLRR, the AUC of SED is increased by 0.009 and 0.01, respectively. SED is ensembled by PTA and GTVLRR to produce higher accuracy, which proves the reliability and superiority of the SED algorithm. For Airport-1 dataset, only the PTA obtains a better score, so it is the only algorithm to be ensembled. The proposed method obtains 0.9450, compared with PTA, the AUC of SED is increased by 0.02. And it is much higher than other methods.

This set of data proves to a great extent that the algorithm can effectively improve the detection accuracy by ensembling features. For Airport-4 dataset, unlike the aforementioned datasets, the PTA and LSMAD obtain better scores, so they are ensembled.

Detection maps of different anomaly detection methods on six datasets are shown in Fig. 4. By comparing the detection maps of compared methods, we can see that the proposed method can effectively detect abnormalities. For the San Diego dataset, planes in the map of SED are clearer than in other detection images. Especially, when compared with the PTA, the map of SED protrudes abnormal objects and penalizes the background part. Without being confused with the background, we can know where the anomalous object is more clearly, especially on the Beach-4 dataset. Moreover, regardless of whether the number of abnormal targets is large or small, and the background is simple or complex, the detection map of SED is clearer than the maps of the comparison algorithms. And the ROC curves of all the methods are compared quantitatively for all datasets. As is shown in Fig. 3, the proposed method is superior to the comparison algorithms in most situations. Especially for the San Diego and Airport-4 datasets, the SED curve is always higher than other curves. For the other two datasets, when the false alarm rate is bigger than 0.01, the probability of detection of the proposed method is higher than other methods.

The computing time of all methods on six datasets is listed in Table IV. The experiments are carried out on the same computer. It can be seen that the computation time of the SED is the sum of the running time of all the algorithms plus the time of the PTA to process \mathcal{Z}^* . Additionally, the time at which the HSI is stratified is ignored. In fact, the time that \mathcal{Z}^* is detected by PTA is very short, only a few seconds. The total running time of SED can be reduced by integrating algorithms with good detection results and abandoning long-running algorithms.

To evaluate the influence of different parameter settings on PTA, we take experimental datasets as examples. We test the effect of γ on AUC in the range of 0–10. As is shown in Fig. 5(a), for all HSI datasets, the best values of γ are less than 4. However, the optimal value of γ still needs to be determined separately.

There are five hyperparameters in PTA. The parameters of the aforementioned six datasets are shown in Fig. 5. And we test the influence of one parameter on AUC under the condition of

fixing the other four parameters. The hyperparameters are set in the range of 10^{-4} , 10^{-3} , 10^{-2} , 10^{-1} , 1 , 10^1 , 10^2 , 10^3 , and 10^4 . Obviously, according to the performance of AUC, it is observed that the parameters β and τ are more sensitive than γ , α , and μ . In order to simplify the algorithm, the best parameters are set between 0.01 and 10.

V. CONCLUSION

In this article, a hyperspectral anomaly detection algorithm based on subfeature ensemble is proposed. In order to reduce the influence of noise and improve the detection efficiency, the band selection is performed on the HSI to obtain a subfeature set \mathcal{Z}^R . To highlight the abnormal features and reduce the interference of the background, six algorithms are applied to detect the abnormality of the original HSI, and the best algorithm is selected to obtain the enhanced feature set \mathcal{Z}^O . The two sets of features are ensemble to obtain the final detection results. The experimental results show that the method proposed in this article performs well on multiple evaluation indicators and is superior to the ensemble algorithms. The proposed algorithm can improve the probability of identifying anomalies and significantly reduce the influence of background. This fully demonstrates the effectiveness of ensemble subfeatures, which can be extended to other similar applications. In addition, although the SED operation time is long, the time for \mathcal{Z}^* to be processed by the PTA is short. We can optimize the efficiency of SED by adding feature importance selection, choosing ensemble algorithms with short running time, etc., in future work.

REFERENCES

- [1] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly detection in hyperspectral images based on an adaptive support vector method," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 646–650, Jul. 2011.
- [2] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Simultaneous joint sparsity model for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 676–680, Jul. 2011.
- [3] X. Liu, C. Wang, Q. Sun, and M. Fu, "Target detection of hyperspectral image based on convolutional neural networks," in *Proc. IEEE 37th Chin. Control Conf.*, 2018, pp. 9255–9260.
- [4] Q. Li, W. Feng, and Y.-H. Quan, "Trend and forecasting of the COVID-19 outbreak in China," *J. Infection*, vol. 80, no. 4, pp. 469–496, 2020.
- [5] S. Dong, Y. Quan, W. Feng, G. Dauphin, and M. Xing, "A pixel cluster CNN and spectral-spatial fusion algorithm for hyperspectral image classification with small-size training samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4101–4114, 2021.
- [6] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2020.
- [7] W. Feng *et al.*, "Semi-supervised rotation forest based on ensemble margin theory for the classification of hyperspectral image with limited training data," *Inf. Sci.*, vol. 575, pp. 611–638, 2021.
- [8] W. Feng *et al.*, "Dynamic synthetic minority over-sampling technique-based rotation forest for the classification of imbalanced hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2159–2169, Jul. 2019.
- [9] T. Guo, F. Luo, L. Fang, and B. Zhang, "Meta-pixel-driven embeddable discriminative target and background dictionary pair learning for hyperspectral target detection," *Remote Sens.*, vol. 14, no. 3, 2022, Art. no. 481.
- [10] Y.-Z. Feng and D.-W. Sun, "Application of hyperspectral imaging in food safety inspection and control: A review," *Crit. Rev. Food Sci. Nutr.*, vol. 52, no. 11, pp. 1039–1058, 2012.
- [11] T. Guo, F. Luo, L. Zhang, B. Zhang, X. Tan, and X. Zhou, "Learning structurally incoherent background and target dictionaries for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3521–3533, 2020.
- [12] R. Tao, X. Zhao, W. Li, H. C. Li, and Q. Du, "Hyperspectral anomaly detection by fractional Fourier entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.
- [13] W. Yao, L. Li, H. Ni, W. Li, and R. Tao, "Hyperspectral anomaly detection based on improved RPCA with non-convex regularization," *Remote Sens.*, vol. 14, no. 6, 2022, Art. no. 1343.
- [14] K. Li *et al.*, "Spectral-spatial deep support vector data description for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, Art. no. 5522316.
- [15] P. Lahaie, "Classification and anomaly detection algorithms for weak hyperspectral signal processing," in *Proc. 8th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2016, pp. 1–5.
- [16] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Jul. 2011.
- [17] X. Sun *et al.*, "Ensemble-based information retrieval with mass estimation for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–23, 2021, Art. no. 5508123.
- [18] J. Zhang, M. Xing, and Y. Xie, "FEC: A feature fusion framework for SAR target recognition based on electromagnetic scattering features and deep CNN features," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2174–2187, Mar. 2021.
- [19] J. Zhang, M. Xing, G.-C. Sun, and N. Li, "Oriented Gaussian function-based box boundary-aware vectors for oriented ship detection in multiresolution SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022, Art. no. 5211015.
- [20] C. I. Chang, "Target signature-constrained mixed pixel classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 5, pp. 1065–1081, May 2002.
- [21] I. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [22] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
- [23] S. Matteoli and G. Corsini, "Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images," *Opt. Eng.*, vol. 49, no. 4, pp. 258–258, 2010.
- [24] W. Zhang, X. Lu, and X. Li, "Similarity constrained convex nonnegative matrix factorization for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4810–4822, Jul. 2019.
- [25] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
- [26] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.
- [27] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.
- [28] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [29] S. Khazai and B. Mojaradi, "A modified kernel-RX algorithm for anomaly detection in hyperspectral images," *Arabian J. Geosci.*, vol. 8, no. 3, pp. 1487–1495, 2015.
- [30] M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 374–387, Feb. 2005.
- [31] C. Zhao and L. Zhang, "Spectral-spatial stacked autoencoders based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection," *Infrared Phys. Technol.*, vol. 92, pp. 166–176, 2018.
- [32] H. Ren and T. B. Moeslund, "Abnormal event detection using local sparse representation," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill.*, 2014, pp. 125–130.
- [33] R. Zhao, B. Du, and L. Zhang, "Hyperspectral anomaly detection via a sparsity score estimation framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3208–3222, Jun. 2017.

- [34] D. Lu, Z. Wu, X. Yang, L. Wei, and Z. Wei, "Kernel low-rank representation for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 477–480.
- [35] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jan. 2020.
- [36] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3071–3084, May 2020.
- [37] J. Xue, Y. Zhao, W. Liao, J. C.-W. Chan, and S. G. Kong, "Enhanced sparsity prior model for low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4567–4581, Nov. 2020.
- [38] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral images super-resolution via learning high-order coupled tensor ring representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4747–4760, Nov. 2020.
- [39] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Spatial-spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1030–1040, Apr. 2018.
- [40] X. Zhang, G. Wen, and W. Dai, "Anomaly detecting in hyperspectral imageries based on tensor decomposition with spectral and spatial partitioning," in *Proc. 8th Int. Congr. Image Signal Process.*, 2015, pp. 737–741.
- [41] Z. Chen, B. Yang, and B. Wang, "A preprocessing method for hyperspectral target detection based on tensor principal component analysis," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1033.
- [42] K. Jiang, W. Xie, Y. Li, J. Lei, G. He, and Q. Du, "Semisupervised spectral learning with generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 5224–5236, Jul. 2020.
- [43] T. Jiang, W. Xie, Y. Li, and Q. Du, "Discriminative semi-supervised generative adversarial network for hyperspectral anomaly detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2420–2423.
- [44] Z. Li, S. Shi, L. Wang, M. Xu, and L. Li, "Unsupervised generative adversarial network with background enhancement and irredundant pooling for hyperspectral anomaly detection," *Remote Sens.*, vol. 14, no. 5, 2022, Art. no. 1265.
- [45] M. Yousefan, H. E. Najafabadi, H. Amirkhani, H. Leung, and V. Hajishemi, "Deep anomaly detection in hyperspectral images based on membership maps and object area filtering," *Expert Syst. Appl.*, vol. 191, 2022, Art. no. 116200.
- [46] D. Esan, P. A. Owolawi, and C. Tu, "Anomalous detection in noisy image frames using cooperative median filtering and KNN," *IAENG Int. J. Comput. Sci.*, vol. 49, no. 1, pp. 1–13, 2022.
- [47] Y. Liu, W. Xie, Y. Li, Z. Li, and Q. Du, "Dual-frequency autoencoder for anomaly detection in transformed hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 5523613.
- [48] P. Wu and J. Liu, "Learning causal temporal relation and feature discrimination for anomaly detection," *IEEE Trans. Image Process.*, vol. 30, pp. 3513–3527, 2021.
- [49] C. Zhao, X. Li, and H. Zhu, "Hyperspectral anomaly detection based on stacked denoising autoencoders," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 042605.
- [50] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022.
- [51] W. Feng, Y. Quan, and G. Dauphin, "Label noise cleaning with an adaptive ensemble method based on noise detection metric," *Sensors*, vol. 20, no. 23, 2020, Art. no. 6718.
- [52] T. Vu, E. Riekeberg, Y. Qiu, and R. Powers, "Comparing normalization methods and the impact of noise," *Metabolomics*, vol. 14, no. 8, pp. 1–10, 2018.
- [53] W. Feng, G. Dauphin, W. Huang, Y. Quan, and W. Liao, "New margin-based subsampling iterative technique in modified random forests for classification," *Knowl.-Based Syst.*, vol. 182, 2019, Art. no. 104845.
- [54] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [55] P. Xiang, J. Song, H. Li, L. Gu, and H. Zhou, "Hyperspectral anomaly detection with harmonic analysis and low-rank decomposition," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 3028.
- [56] N. Ma, X. Yu, Y. Peng, and S. Wang, "A lightweight hyperspectral image anomaly detector for real-time mission," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1622.



Shuo Wang (Graduate Student Member) is currently working toward the master's degree in control science and engineering with the Department of Remote Sensing Science and Technology, School of Electronic Engineering, Xidian University, Xi'an, China. Her research interests include hyperspectral target extraction and remote sensing image classification.



Wei Feng received the B.S. degree in computer science and technology from Northeast Agricultural University, Harbin, China, in 2009, the M.Sc. degree in computer applications technology from North Minzu University, Yinchuan, China, in 2013, and the Ph.D. degree in information science and technology from the Université Michel de Montaigne-Bordeaux 3, Bordeaux, France, in 2017.

She worked as a Postdoctoral Researcher with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, from 2017 to 2019. She is currently an Associate Professor with the Department of Remote Sensing Science and Technology, School of Electronic Engineering, Xidian University, Xi'an, China. Her current research interests include remote sensing, machine learning, and image processing.



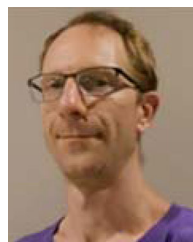
Yinghui Quan (Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 2004 and 2012, respectively.

He is currently a Full Professor with the Department of Remote Sensing Science and Technology, School of electronic engineering, Xidian University. His research interests include radar imaging, radar signal processing, and radar microsystem.



Wenxing Bao received the B.Eng. degree in industrial automation from Xidian University, Xi'an, China, in 1993, the M.Sc. degree in electrical engineering and the Ph.D. degree in electronic science and technology from Xi'an Jiaotong University, Xi'an, in 2001 and 2006, respectively.

He is currently a Professor and Vice President with the North Minzu University, Yinchuan, China. His current research interests include digital image processing and remote sensing image classification and fusion.



Gabriel Dauphin received the engineer's degree in signal and image processing from Mines ParisTech, Paris, France, in 1996, and the Ph.D. degree in signal and image processing from Télécom Paris Tech University, Paris, in 2001.

Since 2002, he has been an Associate Professor with the Laboratory of Information Processing and Transmission (L2TI), University Paris 13, Villetaneuse, France. His research interests include gate recognition, stereoscopic image compression, machine learning, and remote sensing.



Lianru Gao (Senior Member, IEEE) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2002, the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 2007.

He is currently a Professor with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, CAS. He also has been a Visiting Scholar with the University of Extremadura, Cáceres, Spain, in 2014, and with the Mississippi State University, Starkville, MS, USA, in 2016. He has authored and coauthored more than 160 peer-reviewed papers, and there are more than 80 journal papers included by Science Citation Index. His research interests include hyperspectral image processing and information extraction.



Mengdao Xing (Fellow, IEEE) was born in Zhejiang, China, in November 1975. He received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 1997 and 2002, respectively.

He is currently a Full Professor with the Academy of Advanced Interdisciplinary Research, Xidian University. He is also with the National Key Laboratory of Microwave Imaging Technology, Institute of Electronics, Chinese Academy of Sciences, Beijing, China. He is the author or co-author of two books and more than 200 papers. His research interests include synthetic aperture radar (SAR), inverse SAR, and sparse signal processing.



Xian Zhong received the graduate degree in information and communication engineering from the Department of Remote Sensing Science and Technology, School of Electronic Engineering, Xidian University, Xi'an, China, in 2022.

Her research interests include machine learning and deep learning.