

# Rank-Aware Generative Adversarial Network for Hyperspectral Band Selection

Xin Zhang<sup>✉</sup>, Weiyng Xie<sup>ID</sup>, Member, IEEE, Yunsong Li<sup>ID</sup>, Member, IEEE, Jie Lei<sup>ID</sup>, Member, IEEE,  
Qian Du<sup>ID</sup>, Fellow, IEEE, and Geng Yang<sup>ID</sup>

**Abstract**—Traditional clustering-based band selection (BS) methods treat each band as individuals, and selection is conducted by enlarging the difference between clusters, which leads to the loss of band interaction and information saliency evaluation. In this article, we propose a BS method named rank-aware generative adversarial network (R-GAN) to address these problems. First, centralized reference feature extraction (FE) with GAN aids R-GAN to combine interpretability and inter-band relevance. Then, the reference feature is refined with the saliency estimation provided by the rank-aware strategy. According to data characteristics, there are two versions of rank computation including tensor and matrix. Finally, the structural similarity index measurement (SSIM) maps the saliency to the original data space to obtain the final BS result. Extensive comparison experiments with popular existing BS approaches on five hyperspectral images (HSIs) datasets show that the proposed R-GAN can address spectral saliency effectively and select more informative band subsets, which outperforms other competitors for both detection and classification tasks. For example, on the SD-1 dataset, the ten bands selected by R-GAN achieve  $0.982 \pm 0.003$  with an improvement of 13.7% in the area under the curve (AUC) value of anomaly detection performance. The peaked accuracy surpasses the baseline by 0.46% for the classification on the PaviaU dataset.

**Index Terms**—Band selection (BS), hyperspectral images (HSIs), rank-aware generative adversarial network (R-GAN), spectral saliency.

Manuscript received June 6, 2021; revised September 2, 2021, December 6, 2021, and January 6, 2022; accepted January 7, 2022. Date of publication January 11, 2022; date of current version March 11, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62121001, Grant 62071360, Grant 61801359, Grant 61571345, Grant 91538101, Grant 61501346, Grant 61502367, and Grant 61701360; in part by the Young Talent Fund of University Association for Science and Technology in Shaanxi of China under Grant 20190103; in part by the Special Financial Grant from the China Postdoctoral Science Foundation under Grant 2019T120878; in part by the 111 Project under Grant B08038; in part by the Fundamental Research Funds for the Central Universities under Grant XJS200103; in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2019JQ153; in part by the General Financial Grant from the China Postdoctoral Science Foundation under Grant 2017M620440; in part by the Yangtze River Scholar Bonus Schemes under Grant CJT160102; in part by the Ten Thousand Talent Program; in part by the Fundamental Research Funds for the Central Universities; and in part by the Innovation Fund of Xidian University under Grant 5001-20109215456. (*Corresponding author: Weiyng Xie*)

Xin Zhang, Weiyng Xie, Yunsong Li, Jie Lei, and Geng Yang are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: xinzhang\_xd@163.com; wyxie@xidian.edu.cn; ysli@mail.xidian.edu.cn).

Qian Du is with the Department of Electronic and Computer Engineering, Mississippi State University, Starkville, MS 39759 USA (e-mail: du@ece.msstate.edu).

Digital Object Identifier 10.1109/TGRS.2022.3142173

## I. INTRODUCTION

THE wide deployment of cost-sensitive services and low-powered edge applications highlights the importance of computing load and memory footprint reduction [1]–[3]. Simplifying network design [4], [5] and reducing data dimension [6], [7] are two prevalent approaches for removing redundancy. Hyperspectral images (HSIs) [8]–[10] describe various surface features by using solar reflectance in hundreds of spectral bands. The resulted spectral redundancy is not conducive to downstream tasks, such as anomaly/target detection [11], [12] and classification [13]. Fig. 1 shows the similarity between the 205 bands and the reference band. More than 60 bands achieve similarity greater than 0.9, and there are more than four-fifths with similarity over 0.5, which indicates the necessity of removing spectral redundancy through dimensionality reduction (DR).

Band selection (BS) [14]–[17] is an effective DR technique, solving the redundancy reduction problem in an interpretable way called “Take the essence, discard the dregs.” BS explores a salient and representative subset from the original hyperspectral data without a detrimental impact on its performance on follow-up tasks. Instead of projecting data to a low-dimensional space, such as feature extraction (FE) [18]–[20], BS preserves the information of the original data in the physical sense. Roughly, there are two BS schemes: supervised and unsupervised. Due to more robust performance and higher application prospects, the latter one is preferred. Recently, there has been a surge of interest in clustering-based unsupervised or semisupervised BS methods [21]–[26]. Exemplified with optimal clustering framework (OCF) [27] and enhanced fast density-peak-based clustering (EFDPC) [28], clustering-based methods treat each band as individuals and cluster under some criteria. The most relevant band with other bands in each cluster is selected as a representative. However, this strategy fails to consider the global data properties and the relevance between bands. Besides, the information saliency is not evaluated, generating a BS result with significant individual differences but less effective content.

The first dilemma can be solved by deep FE to a certain degree. Being virtual of the superiority of solving nonlinear problems, deep FE shows its superiority in summarizing high-dimensional data. Autoencoder (AE) and its variants are basic deep FE approaches and rely on a small hidden layer to generate deep features. Stacked autoencoder (SAE),

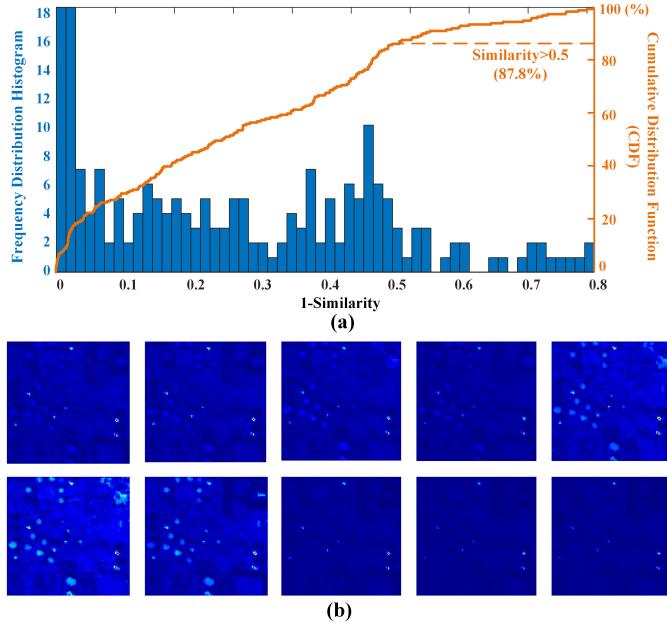


Fig. 1. Similarity between the reference band (130) and other bands. As shown in (a), there are 180 bands with similarities over 0.5, accounting for over four-fifths (87.8%). The reference band and the bands with a similarity varying from 1 to 0.2 (interval = 0.1) are exhibited in (b) from top left to the bottom right, and only the similarity lower than 0.5 results in a perceptible visual difference.

e.g., compact and discriminative stacked autoencoder (CDSAE) [29] and stacked sparse autoencoder (SSAE) [30], and variational autoencoder (VAE), e.g., spatial revising variational autoencoder (SRVAE) [31], have succeeded in low-dimensional feature mapping. Besides, a convolutional neural network (CNN) is also used for spectral FE [32], [33]. The adversarial game introduced by the generative adversarial network (GAN) [34]–[36] makes it stand out of the deep networks for FE. For example, [37] proposed a GAN-based FE network for HSI where the original Jensen–Shannon divergence is replaced by the Wasserstein distance. The spectral–spatial generative adversarial network and conditional random field (SS-GAN-CRF)-based framework [36] integrates a semisupervised deep learning and a probabilistic graphical model for the HSI classification task. As for the second problem, rank is an appropriate measurement for data information. For example, [38] uses ranks of feature maps for filter pruning. Considering the high-dimensional situation, rank has been extended to tensor.

Hence, we propose a novel BS method, i.e., rank-aware GAN (R-GAN), to address the spectral saliency problem. First, spectral vectors are inputted into GAN for deep latent FE, which completes the global information and the inherent nonlinear relationship between bands lost in traditional BS methods. Then, the latent feature is refined with a rank-aware strategy. Specifically, the tensor rank is used to determine the scale of latent features with a criterion of being small but informative. The saliency of bands is evaluated with rank. Finally, the most salient bands are mapped to the original data space to obtain the BS result. The main contributions of this article are given as follows.

- 1) We propose a deep-learning-based BS method, which combines the global modeling capability of deep FE and the interpretability of BS. The interbands' relevance provided by deep FE helps to construct a complete reference feature.
- 2) According to the data characteristics of different stages, a rank-aware strategy containing tensor rank and matrix rank is used for deep feature refinement to keep guiding reference informative.
- 3) The structural similarity index measurement (SSIM) bridges the latent space and the original data space, and derives the final informative BS result.

The rest of this article is structured as follows. We first introduce the proposed R-GAN framework in Section II. Next, in Section III, we conduct the experiments to investigate the performance of the proposed method and comparison BS methods. Finally, we conclude with a summary and final remarks in Section IV.

## II. METHODOLOGY

Let an HSI cube with  $B$  spectral bands and  $N \times M$  pixels be defined as a three-way tensor  $\mathcal{H} \in \mathbb{R}^{N \times M \times B}$ . Let the  $(i, j, b)$ th entry be denoted as  $\mathcal{H}_{ijb}$ , and  $\mathcal{H}(i, :, :)$ ,  $\mathcal{H}(:, j, :)$ , and  $\mathcal{H}(:, :, b)$  represent the  $i$ th horizontal,  $j$ th lateral, and  $b$ th frontal slice, respectively. Spatially,  $\mathcal{H}$  consists of  $N \times M$  spectral vectors  $\mathcal{H} = \{\mathbf{h}_i\}_{i=1}^{N \times M}$ . Along the band direction, the HSI can be regarded as a collection of  $B$  frontal slices, namely, band images  $\mathcal{H} = \{\mathcal{H}(:, :, b)\}_{b=1}^B$ . For convenience, it is abbreviated to  $\mathcal{H} = \{\mathcal{H}_b\}_{b=1}^B$ . In BS,  $\mathcal{H}$  can be split into two categories, i.e., a remaining subset crammed with salient bands  $\mathcal{I} = \{\mathcal{H}_{\mathcal{I}_1}, \mathcal{H}_{\mathcal{I}_2}, \dots, \mathcal{H}_{\mathcal{I}_P}\} = \{\mathcal{I}_p\}_{p=1}^P$  and a subset, with redundancy and less saliency, to be discarded  $\mathcal{U} = \{\mathcal{H}_{\mathcal{U}_1}, \mathcal{H}_{\mathcal{U}_2}, \dots, \mathcal{H}_{\mathcal{U}_Q}\} = \{\mathcal{U}_q\}_{q=1}^Q$ . Here,  $\mathcal{I}_p$  and  $\mathcal{U}_q$  are the indices of the  $p$ th selected and  $q$ th unselected band image, and  $\mathcal{I} \cup \mathcal{U} = \mathcal{H}$ ,  $\mathcal{I} \cap \mathcal{U} = \emptyset$ , and  $P + Q = B$ .

In essence, BS involves two core issues: how many bands and which bands are preferred, i.e., to determine  $\mathcal{I}_p$  and  $P$ . To solve these problems, BS is formulated as an optimization problem aiming to maximize the saliency of the reserved bands  $\mathcal{I}$

$$\begin{aligned} & \max_{\delta_b} \sum_{b=1}^B \delta_b \mathcal{R}(\mathcal{H}_b) \\ & \text{s.t. } \sum_{b=1}^B \delta_b = P \end{aligned} \quad (1)$$

where  $\delta_b$  is a pointer indicating if  $\mathcal{H}_b$  is included in  $\mathcal{I}$  ( $\delta_b = 1$ ) or in  $\mathcal{U}$  ( $\delta_b = 0$ ). The saliency of each band is measured by  $\mathcal{R}(\cdot)$ . Solving the optimization problem in (1) is equivalent to removing  $Q$  redundant and less salient bands. Most of prior works calculate the related indicators and determine the saliency of band images using clustering, leading to the loss of global data properties and relevance between bands. In this article, being virtue of the powerful ability of GAN in learning hierarchical and nonlinear features, a reasonable  $\mathcal{R}(\cdot)$  is constructed. First, the original HSI  $\mathcal{H}$  is fed into the network to generate the information-centralized latent

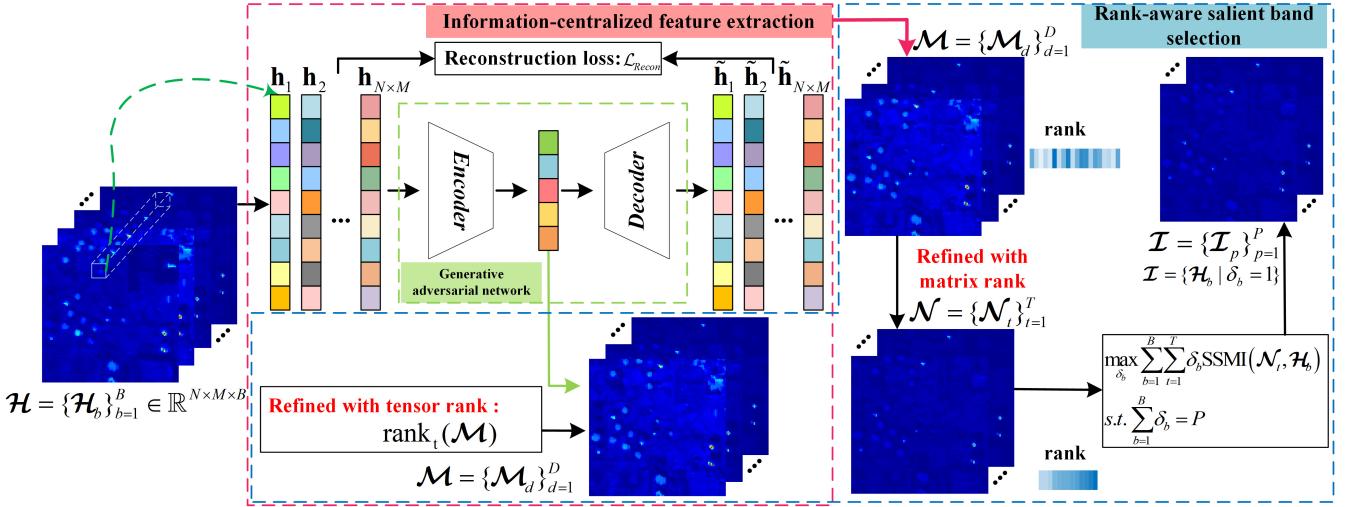


Fig. 2. Schematic of the proposed R-GAN for BS in HSIs.

feature  $\mathcal{M} = \{\mathcal{M}_d\}_{d=1}^D$ . Then,  $\mathcal{M}$  is further refined in a rank-aware way consisting of two rank operations specific to current data. Finally, BS is induced by the refined feature  $\mathcal{N} = \{\mathcal{N}_t\}_{t=1}^T$  through SSIM. SSIM captures the structural information which plays a pivotal role in the human visual system (HSV). The larger SSIM is, the more similar  $\mathcal{H}_b$  and informative  $\mathcal{N}_t$  are, and the more salient  $\mathcal{H}_b$  is. Then, (1) can be reformulated as

$$\begin{aligned} & \max_{\delta_b} \sum_{b=1}^B \sum_{t=1}^T \delta_b \text{SSIM}(\mathcal{N}_t, \mathcal{H}_b) \\ & \text{s.t. } \sum_{b=1}^B \delta_b = P \end{aligned} \quad (2)$$

and the optimal band subset  $\mathcal{I}$  can be obtained by

$$\mathcal{I} = \{\mathcal{H}_b | \delta_b = 1\}. \quad (3)$$

The schematic of the proposed R-GAN is shown in Fig. 2.

#### A. Information-Centralized FE

The sample adequacy of HSI makes it possible to model distribution with data-driven deep learning tools (an HSI of size  $100 \times 100$  provides 10 000 training samples). Here, GAN is chosen to extract referential latent features from  $B$  band images. Typically, GAN consists of a bowtie-shaped generator  $G$  and a discriminator  $D$  in the latent space. However, such a structure often comes with a nonnegligible reconstruction performance penalty [39]. Thus, two discriminators  $D_s$  and  $D_z$  are added to enforce the alignment in the spectral and latent spaces, respectively. The GAN is trained by solving the following optimization problem:

$$\arg \min_{\Theta_G} \max_{\Theta_{D_s}, \Theta_{D_z}} \mathcal{L}_{\text{Adv}}(\Theta_G, \Theta_{D_s}, \Theta_{D_z}) + \mathcal{L}_{\text{Recon}}(\Theta_G) \quad (4)$$

where  $\Theta_G$ ,  $\Theta_{D_s}$ , and  $\Theta_{D_z}$  denote the parameters of generator  $G$ , spectral discriminator  $D_s$ , and latent discriminator  $D_z$ ,

respectively. The first item is the adversarial loss and can be factorized into two parts as follows:

$$\mathcal{L}_{\text{Adv}}(\Theta_G, \Theta_{D_s}, \Theta_{D_z}) = \mathcal{L}_s(\Theta_G, \Theta_{D_s}) + \mathcal{L}_z(\Theta_G, \Theta_{D_z}) \quad (5)$$

where  $\mathcal{L}_s(\Theta_G, \Theta_{D_s})$  is the spectral adversarial loss matching the distribution of the decoded spectral  $p(\tilde{\mathbf{h}}_i)$  and the known input data distribution  $p(\mathbf{h}_i)$

$$\begin{aligned} \mathcal{L}_s(\Theta_G, \Theta_{D_s}) &= \mathbb{E}_{\mathbf{h}_i \sim p(\mathbf{h}_i)} [\log(D_s(\mathbf{h}_i, \Theta_{D_s}))] \\ &\quad + \mathbb{E}_{\tilde{\mathbf{h}}_i \sim p(\tilde{\mathbf{h}}_i)} [\log(1 - D_s(\tilde{\mathbf{h}}_i, \Theta_{D_s}))]. \end{aligned} \quad (6)$$

$\mathcal{L}_z(\Theta_G, \Theta_{D_z})$  is formulated in the similar fashion as (6) aiming to match distribution of the latent feature  $p(\mathbf{z})$  with the prior distribution  $p(\tilde{\mathbf{z}})$

$$\begin{aligned} \mathcal{L}_z(\Theta_G, \Theta_{D_z}) &= \mathbb{E}_{\tilde{\mathbf{z}} \sim p(\tilde{\mathbf{z}})} [\log(D_z(\tilde{\mathbf{z}}, \Theta_{D_z}))] \\ &\quad + \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} [\log(1 - D_z(\mathbf{z}, \Theta_{D_z}))]. \end{aligned} \quad (7)$$

Here,  $p(\tilde{\mathbf{z}})$  is recommended to be  $N(0, I)$  [40].

In addition,  $\mathcal{L}_{\text{Recon}}(\Theta_G)$  is the reconstruction loss used to align the real input with the reconstructed output. Therefore, the reconstruction loss can be expressed by the MSE loss

$$\mathcal{L}_{\text{Recon}}(\Theta_G) = \frac{1}{N \times M} \sum_{i=1}^{N \times M} \|\mathbf{h}_i - \tilde{\mathbf{h}}_i\|_2^2 \quad (8)$$

where  $\tilde{\mathbf{h}}_i = G(\mathbf{h}_i, \Theta_G)$ .

By optimizing (4) with stochastic gradient descent (SGD), we obtain the trained GAN  $\Phi(\Theta_G, \Theta_{D_s}, \Theta_{D_z})$ , where  $\Theta_G = [\Theta_{\text{En}}, \Theta_{\text{De}}]$  and  $\Theta_{\text{En}}$  and  $\Theta_{\text{De}}$  are the parameters of encoder and decoder, respectively. The encoder is dedicated to map the high-dimensional HSI into a low-dimensional latent space, revealing the essential characteristics, i.e., the latent features

$$\mathcal{M} = \text{Encoder}(\mathcal{H}, \Theta_{\text{En}}). \quad (9)$$

In Section III-E, FE, we design a BS method based on the rank of slices dubbed as R-BS to show how the information-centralized feature aids the proposed method in the subsequent tasks.

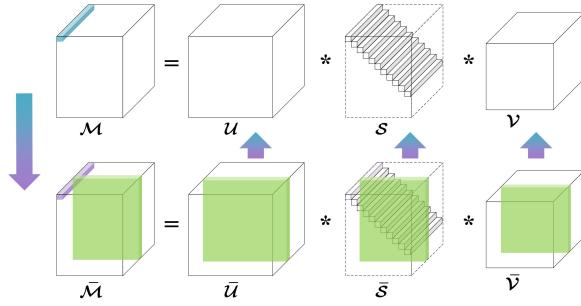


Fig. 3. Illustration of the t-SVD of tensor  $\mathcal{M}$ . First, perform discrete Fourier transformation (DFT) on every vector of  $\mathcal{M}$  to obtain  $\bar{\mathcal{M}}$ . Then, compute each front slice of  $\bar{\mathcal{U}}$ ,  $\bar{\mathcal{S}}$ , and  $\bar{\mathcal{V}}$  from  $\bar{\mathcal{M}}$  by SVD. Finally,  $\mathcal{U}$ ,  $\mathcal{S}$ , and  $\mathcal{V}$  are obtained via inverse DFT.

### B. Rank-Aware Salient BS

For the latent features, it is crucial to keep the sufficiency and conciseness of the contained information. Information insufficiency hinders the superiority of the data-driven approach. On the contrary, information redundancy conflicts over the intention of BS. A linear algebraic measure, rank, turns to be an appropriate measurement for spectral saliency. Here, different data characteristics result in two rank calculation methods. First, the high-dimensional nature of  $\mathcal{M}$  prefers global tensor tubal rank instead of calculating the rank of the 2-D matrix after being reshaped. For tensor  $\mathcal{M} \in \mathbb{R}^{N \times M \times D}$ , the tensor tubal rank is defined as the number of nonzero singular tube of  $\mathcal{S}$ , where  $\mathcal{S}$  is from the tensor singular value decomposition (t-SVD). t-SVD shown in Fig. 3 is solved with the aid of time–frequency domain conversion and factorizes a 3-D tensor into three components, including two orthogonal tensors,  $\mathcal{U}$  and  $\mathcal{V}$ , and a f-diagonal tensor,  $\mathcal{S}$

$$\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*. \quad (10)$$

The tensor tubal rank  $\text{rank}_t(\mathcal{M})$  can be defined as

$$\text{rank}_t(\mathcal{M}) = \text{counts}\{i, \mathcal{S}(i, i, :) \neq 0\}. \quad (11)$$

Considering the property of inverse DFT

$$\mathcal{S}(i, i, 1) = \frac{1}{n_3} \sum_{j=1}^{n_3} \bar{\mathcal{S}}(i, i, j). \quad (12)$$

Here,  $n_3$  is the number of slice in  $\mathcal{S}$ .  $\mathcal{S}(i, i, 1) = 0$  derives  $\bar{\mathcal{S}}(i, i, j) = 0$  and then  $\mathcal{S}(i, i, j) = 0$ , which indicates a zero singular tube. Only  $\mathcal{S}(i, i, 1) \neq 0$  contributes to the rank calculation. That is, the tensor tubal rank depends on the first slice  $\mathcal{S}(i, i, 1)$  of  $\mathcal{S}$

$$\text{rank}_t(\mathcal{M}) = \text{counts}\{i, \mathcal{S}(i, i, 1) \neq 0\}. \quad (13)$$

In accordance with rank, we can keep  $\mathcal{M}$  informative and restrict its dimension to reduce the computation load.

With  $\mathcal{M}$  of an appropriate dimension, the local matrix rank is applied for further evaluation on each slice to find the most salient bands. Specifically,  $\mathcal{M}$  is sorted by the rank of slice  $\mathcal{M}_d$ , and  $\mathcal{N}$  is a collection of slices with top- $T$  rank

$$\begin{aligned} \mathcal{M} &= \text{sort}(\mathcal{M}, \text{key} = \text{rank}(\mathcal{M}_d)) \\ \mathcal{N} &= \mathcal{M}[:, :, D - T : D]. \end{aligned} \quad (14)$$

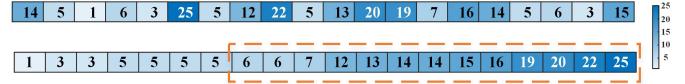


Fig. 4. Illustration of salient slices selection. The intensity of the color depends on the rank whose value is tagged in the corresponding position.

The rationale lies in that a larger rank suggests a larger saliency. Fig. 4 gives an intuitive illustration. Here, to evaluate the contribution of the rank-aware strategy, we also design a rank-unaware strategy BS method (RU-GAN), that is, the bands are selected by mapping latent features into the original data space directly. See Section III-E, rank-aware strategy, for detailed analysis.

### C. Mapping Saliency to Original Data Space

To obtain the final BS result, the data-driven saliency is mapped to the original data space with SSIM as (2). In other words,  $\mathcal{H}_b$  having the maximal SSIM with  $\mathcal{N}_t$  is regarded as an alternative in the original data space.  $\mathcal{H}_b$  is prioritized by the noise level before matching [41]. Here, SSIM [42] is adopted to evaluate the brightness, contrast, and structure to characterize the data for favorable human vision perception, which is defined as

$$\begin{aligned} \text{SSIM} &= A * C * S \\ A &= \frac{2u_1 u_2 + C_1}{u_1^2 + u_2^2 + C_1} \\ C &= \frac{2\sigma_1 \sigma_2 + C_2}{\sigma_1^2 + \sigma_2^2 + C_2} \\ S &= \frac{\sigma_{12} + C_3}{\sigma_1 \sigma_2 + C_3} \end{aligned} \quad (15)$$

where  $u_1$  and  $u_2$  are the means of  $\mathcal{N}_t$  and  $\mathcal{H}_b$ , respectively,  $\sigma_1$  and  $\sigma_2$  are the corresponding variances,  $\sigma_{12}$  is covariance, and  $C_1$ ,  $C_2$ , and  $C_3$  denote constants guaranteeing the validity of fraction. The larger SSIM between  $\mathcal{N}_t$  and  $\mathcal{H}_b$  is, the more salient  $\mathcal{H}_b$  is, and the more likely  $\mathcal{H}_b$  plays a critical role in subsequent tasks. Then, the selected subset can be determined by (3). Other mapping methods, such as the mean square error (MSE) and the Euclidean distance (ED), are also studied. See Section III-E, mapping saliency method, for detailed analysis.

## III. EXPERIMENTS

### A. Experimental Setup

**Dataset:** Five datasets (six HSIs) are used in our experiments including four for detection and one for classification. All the datasets are introduced as follows.

1) *San Diego*: The San Diego dataset covers two scenes of the San Diego airport area, CA, USA, and is captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. Both scenes consist of  $100 \times 100$  pixels, and each pixel is imaged in wavelengths ranging from 370 to 2510 nm with 189 bands. We name the scene with three airplanes in the upper left corner as SD-1 and the other one as SD-2. The pseudo-RGB images and the ground truths are shown in Fig. 5(a) and (b).

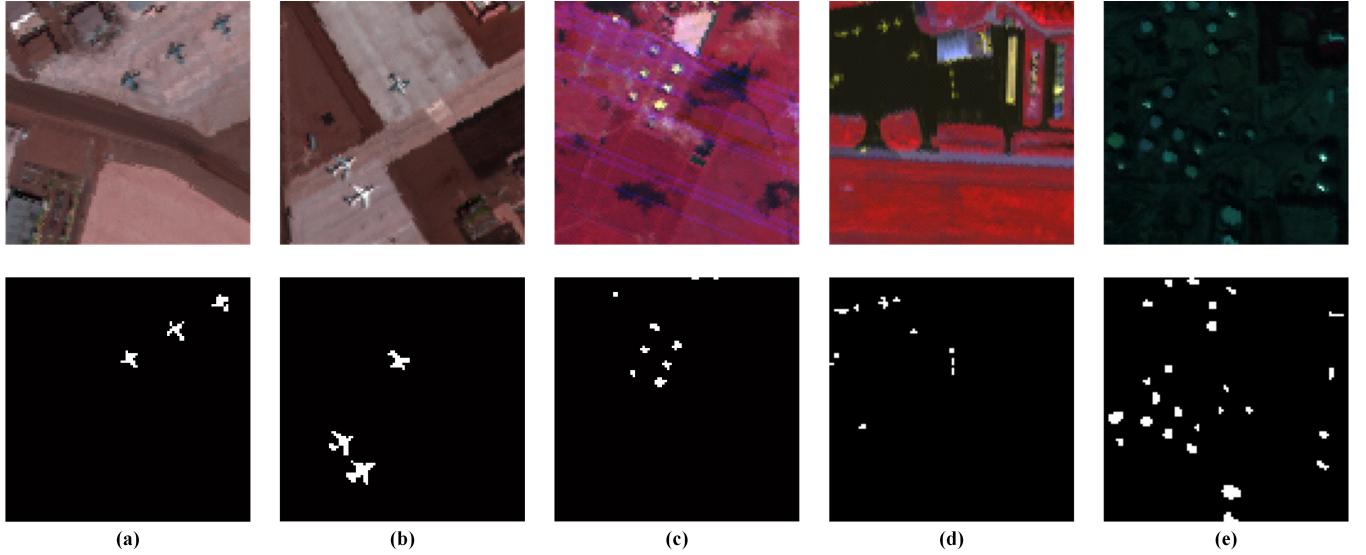


Fig. 5. Pseudo-RGB images and ground truths of experimental datasets (a) SD-1, (b) SD-2, (c) TC, (d) GA, and (e) LA for the detection tasks.

2) *Texas Coast*: This dataset abbreviated as TC is taken by AVIRIS sensor in 2010. In the experiment,  $100 \times 100$  pixels are used, which cover the Texas coast area with a resolution of 17.2 m and encompass a wavelength range of 450–1350 nm with 204 bands. Eleven most disturbed bands are removed in our experiments. Fig. 5(c) shows the pseudo-RGB image and the ground truth.

3) *Gainesville*: This dataset is taken in Gainesville, FL, USA, in 2010 by AVIRIS sensor and is named GA in experiments. The spatial size is  $100 \times 100$ , and the resolution is 3.5 m. The expansion of spectral direction covers wavelengths from 450 to 1350 nm. There are some ships considered as outliers. The pseudo-RGB image and the ground truth are shown in Fig. 5(d).

4) *Los Angeles*: This dataset composed of  $100 \times 100$  pixels describes the Los Angeles city area in 2011 with a resolution of 7.1 m. This dataset shares the same imaging technique as above and extends each pixel to 205 bands with a range of 430–860 nm. Conveniently, it is shortened to LA. Fig. 5(e) shows the pseudo-RGB image and the ground truth.

5) *Pavia University*: The Pavia University dataset (abbreviated as PaviaU) is acquired by the reflective optics system imaging spectrometer (ROSIS) sensor. It covers Pavia University in the Northern Italy with  $610 \times 340$  pixels. Each pixel has 103 spectrum bands ranging from 430 to 860 nm. Fig. 6 shows the pseudo-RGB image and the ground truth.

In this article, we resort AED [43] to generate the ground truth for each dataset, which is still a rough guide. Since the targets are fully resolved spatially, the readers can obtain a more convincing ground truth through high-resolution panchromatic data (pansharpening) or super-resolution methodology.

**Compared Methods:** The effectiveness of R-GAN is evaluated by comparison with eight existing BS methods, e.g., dominant-set-extraction-based selector (DSEBS) [44], EFPDC [28], OCF [27], saliency bands and scale selection

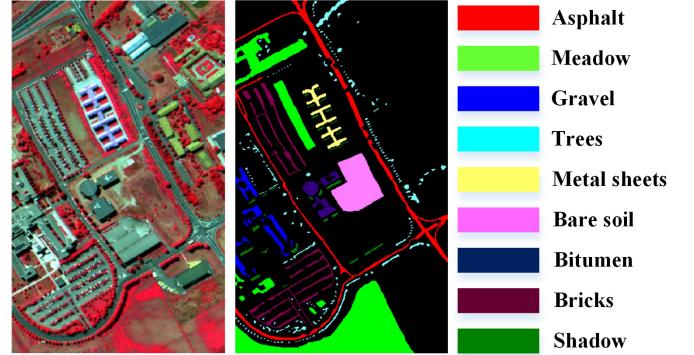


Fig. 6. Pseudo-RGB image and ground truth of experimental dataset PaviaU for the classification task.

(SBSS) [7], adaptive subspace partition strategy (ASPS) [41], scalable one-pass self-representation learning (SOPSLR) [15], fast neighborhood grouping method (FNGBS) [45], and BS network (BS-Net) [33]. The proposed and the competitor algorithms use different features for selection. R-GAN selects bands according to the spatial rank with the guidance of spectral information centralized feature. DSEBS exploits structure information by means of local spatial-spectral consistency analysis, which calculates the gradients in three directions. EFPDC draws support from the spectral feature by calculating the similarity matrix between bands. OCF chooses the bands by using EFPDC to assess whether the band is a suitable cluster center and only considering the spectral feature. SBSS combines spectral and spatial information and treats the spatial coordinates and spectral coordinates as the same. Both ASPS and FNGBS are spectral-spatial methods, the difference is the spatial feature used to select representative band. ASPS considers the band noise level, while FNGBS uses band information evaluated with entropy. SOPSLR only utilizes the spectral feature, and the lost spatial information leads to a suboptimal selection. The fully connected BS-Net used for

TABLE I  
PARAMETERS OF COMPARED METHODS

Methods	Parameters	Value
R-GAN	depth $d$	2
	learning rate $lr$	$10e^{-5}$
DSEBS	standard deviation of Gaussian kernel $\sigma$	0.5
	decay of the exponential function $\lambda$	0.5
SOPSR	regularization parameters $\tau_1$ , $\tau_2$ , and $\tau_3$	$10e^{-4}$ , $10e^{-5}$ , $10e^5$
	number of neighboring pixels $k$	10
	weight $r$	2
FNGBS	data scale for entropy operation $Z$	0.1
BS-Net	epoch	100
	learning rate $lr$	$10e^{-4}$

comparison is trained with spectral vectors, which means that it is also a spectral method. Note that EFPDC, OCF, SBSS, and ASPS are parameter-free, and the parameters impacting on R-GAN, DSEBS, SOPSR, FNGBS, and BS-Net are listed in Table I. The proposed R-GAN is trained for 500 epochs for all the experimental datasets, and it only takes nearly 13 min for the largest dataset PaviaU ( $610 \times 340 \times 103$ ). Considering the very different characteristics of each HSI dataset, the training procedure is dataset-customized.

*Evaluation Metric:* Two typical detectors, including a statistical modeling-based technique, Reed–Xiaoli (RX) algorithm [46], and a well-known subpixel target detector, constrained energy minimization (CEM) algorithm [47], are used to offer the objective detection evaluation for all the BS methods. Note that the dimension reduction necessitates the recalculation of prior target information “ $d$ ” functioning in CEM. We generate “ $d$ ” for the results of R-GAN and the above competitors by averaging all the target spectral vectors. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve is calculated to provide the quantitative analysis. ROC curve describes the varying relationship between the probability of detection (PD) and probability of false alarm (PFA) based on the ground truth. In practice, the PD at low PFA is the most important to the applications, particularly for the military/defense operations. Hence, the AUC and ROC curves integrated with the low PFA range are particularly studied. Here, the ROC curves and AUC values with a low PFA range ( $1e-4$  and  $1e-2$ ) are considered.

Besides, k-nearest neighborhood (KNN) is employed to assess the effectiveness of the classification task. 10% samples from each class based on selected bands are randomly chosen as the training set, and the remaining 90% samples are used for the test. The classification results are evaluated by the overall accuracy, which is defined as the number of correctly classified pixels out of the total number of all test pixels in percentage.

Because the desired number of bands is applicant-dependent, we implement experiments in the range of five to 30 bands to offer a comprehensive assessment. Considering the bell-shaped accuracy tendency caused by the “Hughes” curse of dimensionality phenomenon, the peaked accuracy is studied for each method.

#### B. Performance Comparison on Detection Tasks

Table II reports the anomaly detection performance. For the SD-1 dataset, R-GAN shows overwhelming dominance in

all the studied band sizes except for 5. The subsets selected by R-GAN outperform the baseline for at least 12.6% and 2.4% on SD-1 and SD-2, respectively. R-GAN only has a gap of 0.1% on SD-1 and SD-2 with the best-compared methods (BS-Net and EFPDC) with respect to the peaked accuracy. On the TC dataset, the good performance of all the compared methods shows that minimizing accuracy loss is a reasonable criterion to decide the best BS method. When the band size is 5, R-GAN achieves the second-highest performance after SBSS. As the number of bands is increasing, R-GAN shows its superiority over other methods. As for peaked AUC, R-GAN exceeds the runner-up (BS-Net) by 0.1%. For the GA dataset, R-GAN performs well on all the studied band numbers, which leads to a higher global and peaked accuracy. The peaked performance of R-GAN outperforms FNGBS by 0.73%. On the LA dataset, R-GAN maintains the highest accuracy for all the studied dataset scales. The peaked accuracy of R-GAN exceeds the best comparison methods (DSEBS and SBSS) by 0.2%.

As for target detection performance reported in Table III, R-GAN can still outperform. For the SD-1 dataset, R-GAN achieves the highest accuracy values at all subset scales. Even the worst performance of R-GAN on SD-1 is over 0.988, which is nonexistent in other compared methods. On the TC dataset, R-GAN leads at different band scales except for 5. However, the integral superiorities make it more competitive.

In some military/defense application scenarios, the desired detector should be able to keep high PD when PFA is low, which is displayed as the ROC curve close to the upper left corner. As reported in Table II, the AUC values with a low PFA range ( $1e-4$  and  $1e-2$ ) outperform other methods and the baseline for all the experimental datasets. The corresponding ROC curves are illustrated in Figs. 7 and 8. The proposed R-GAN envelopes other methods on SD and GA datasets and is a little lower than ASPS or DSEBS on TC and LA datasets, which indicates its competitiveness and effectiveness. The results of the target detection task are shown in Table III and Fig. 9. The numerical superiority on the SD-1 dataset is embodied in the ideal ROC curves in Fig. 9(a), which is above the comparison methods in the whole studied PFA range. For the TC dataset, the AUC value of R-GAN is lower than the baseline, which is consistent with the ROC curves in Fig. 9(b). Except for the baseline, the ROC curve of R-GAN is only lower than FNGBS. In general, the proposed R-GAN performs well in the low PFA range on all the experimental datasets, which indicates it is a suitable BS method for the detection tasks in military/defense operations.

#### C. Performance Comparison on Classification Task

We conduct the proposed R-GAN on the classification task to further evaluate the generalization for various downstream tasks. As listed in Table IV, R-GAN can distinguish the nine class objects with less error than the comparison methods when the band number is set more than 15. With more bands, the performance of R-GAN even surpasses the baseline. The peaked OA of R-GAN outperforms OCF and BS-Net by 2.4%

TABLE II  
AREA UNDER THE CURVE (AUC) VALUES OF DIFFERENT METHODS ON EXPERIMENTAL DATASETS FOR ANOMALY DETECTION

Dataset	Num of Bands	Compared Methods								Baseline	
		R-GAN	DSEBS	EFPDC	OCF	SBSS	ASPS	SOPSR	FNGBS		
SD-1	5	<u>0.989±0.002</u>	0.967	0.955	0.961	<u>0.984</u>	<u>0.975</u>	<u>0.973</u>	<u>0.972</u>	<b>0.990</b>	0.864
	10	<b>0.982±0.003</b>	<u>0.977</u>	<u>0.968</u>	0.966	0.978	0.953	<u>0.973</u>	0.968	0.980	
	15	<b>0.978±0.003</b>	0.946	0.963	0.964	0.864	0.954	0.960	0.955	0.977	
	20	<b>0.974±0.003</b>	0.973	0.965	0.963	0.969	0.951	0.966	0.960	<b>0.974</b>	
	25	<b>0.976±0.002</b>	0.973	0.962	0.963	0.967	0.958	0.962	0.964	<b>0.976</b>	
	30	<b>0.973±0.001</b>	0.972	0.955	<u>0.968</u>	0.962	0.959	0.958	0.953	<b>0.974</b>	
	Peaked AUC	0.989±0.002	0.977	0.968	0.968	0.984	0.975	0.973	0.972	<b>0.990</b>	
SD-2	Low PFA Range	<b>2.7e-4±0.0001</b>	0	4.5e-5	2.9e-5	1.3e-5	0	6.0e-5	1.1e-5	7.1e-5	9.1e-5
SD-2	5	<u>0.985±0.002</u>	0.967	<b>0.986</b>	<u>0.985</u>	0.968	<u>0.977</u>	0.969	<u>0.984</u>	0.972	0.954
	10	<b>0.981±0.001</b>	<u>0.977</u>	0.978	0.977	<u>0.975</u>	0.976	<u>0.975</u>	0.977	<u>0.978</u>	
	15	<b>0.979±0.001</b>	0.971	0.975	0.974	0.973	0.976	0.974	0.974	<b>0.979</b>	
	20	<b>0.978±0.001</b>	0.973	0.975	0.974	0.968	0.975	0.970	0.977	0.977	
	25	<b>0.978±0.001</b>	0.973	0.974	0.975	0.966	0.969	0.973	0.975	0.977	
	30	<b>0.977±0.000</b>	0.972	<b>0.977</b>	0.974	0.968	<u>0.977</u>	0.967	0.973	<b>0.977</b>	
TC	Peaked AUC	0.985±0.002	0.977	<b>0.986</b>	0.985	0.975	0.977	0.975	0.984	0.978	0.992
TC	Low PFA Range	<b>1.3e-3±0.0001</b>	8.8e-4	4.8e-4	6.2e-4	3.2e-4	6.1e-4	4.0e-4	5.5e-4	5.6e-4	
TC	5	0.992±0.002	0.988	0.981	0.985	<u>0.993</u>	0.990	0.985	0.986	<u>0.994</u>	
	10	<b>0.995±0.001</b>	0.992	0.987	0.990	<u>0.993</u>	<u>0.994</u>	<u>0.993</u>	0.982	0.992	
	15	<b>0.995±0.001</b>	<u>0.993</u>	0.988	<u>0.993</u>	0.992	0.989	0.992	0.991	0.992	
	20	<b>0.995±0.001</b>	0.992	0.989	<u>0.993</u>	0.991	0.990	0.992	0.990	0.992	
	25	<b>0.995±0.000</b>	0.990	0.990	0.993	0.990	0.991	0.992	0.991	0.992	
GA	30	<b>0.995±0.000</b>	0.991	<u>0.991</u>	<u>0.993</u>	0.991	0.991	<u>0.993</u>	<u>0.993</u>	0.991	0.951
GA	Peaked AUC	<b>0.995±0.000</b>	0.993	0.991	0.993	0.993	0.994	0.993	0.993	0.994	
	Low PFA Range	<b>7.2e-3±0.0001</b>	0.006	0.005	0.006	0.005	0.007	0.006	0.005	0.006	
GA	5	<b>0.968±0.000</b>	0.899	0.958	0.947	0.968	0.957	0.958	0.952	0.842	0.951
	10	<b>0.970±0.002</b>	0.962	0.938	0.946	<u>0.970</u>	0.957	0.942	0.941	<u>0.969</u>	
	15	<b>0.971±0.002</b>	0.960	0.957	0.953	0.965	0.955	0.958	0.945	0.963	
	20	<b>0.971±0.003</b>	0.963	0.967	<u>0.968</u>	0.967	0.951	0.956	0.959	0.963	
	25	<b>0.971±0.003</b>	0.958	0.966	<u>0.968</u>	0.967	<u>0.966</u>	<u>0.969</u>	<u>0.964</u>	0.965	
	30	<b>0.970±0.002</b>	<u>0.965</u>	<u>0.968</u>	0.965	0.965	0.964	0.965	0.959	0.965	
	Peaked AUC	<b>0.971±0.002</b>	0.965	0.968	0.968	0.970	0.966	0.969	0.964	0.969	
LA	Low PFA Range	<b>1.2e-3±0.0001</b>	7.9e-4	2.8e-4	5.0e-4	0.001	3.1e-4	8.3e-4	2.2e-4	4.7e-4	3.4e-4
LA	5	<u>0.992±0.000</u>	<u>0.990</u>	0.975	0.976	<u>0.990</u>	0.978	0.979	0.977	0.969	0.989
	10	<b>0.988±0.001</b>	0.986	0.981	0.984	0.987	0.982	0.985	<u>0.987</u>	0.982	
	15	<b>0.990±0.001</b>	0.986	<u>0.983</u>	<u>0.985</u>	0.981	0.982	0.986	<u>0.987</u>	0.979	
	20	<b>0.990±0.001</b>	0.986	0.982	0.984	0.980	0.981	0.985	0.984	<u>0.989</u>	
	25	<b>0.990±0.000</b>	0.986	0.981	0.982	0.979	<u>0.984</u>	<u>0.987</u>	0.982	0.982	
	30	<b>0.991±0.000</b>	0.985	0.982	0.982	0.979	0.983	<u>0.987</u>	0.981	0.984	
	Peaked AUC	<b>0.992±0.000</b>	0.990	0.983	0.985	0.990	0.984	0.987	0.987	0.989	0.003

\* The best-performing method at each data scale is in **bold**. The peaked accuracy of each method is underlined

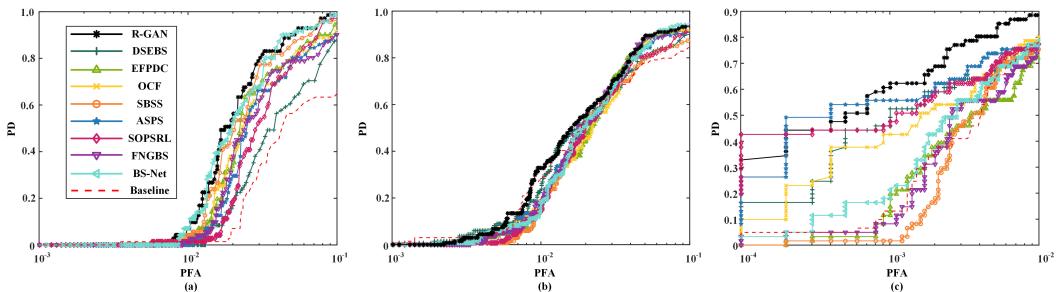


Fig. 7. ROC curves (in low PFA range) of different methods on (a) SD-1, (b) SD-2, and (c) TC datasets for anomaly detection.

and 6.6%. In Fig. 10, our work is extended into more bands toward all bands, and the overall superiority and stability are clear.

#### D. Network Architecture Analysis

GAN exhibits three game schemes according to the diversity of discriminator in quantity and position, as illustrated

TABLE III  
AUC VALUES OF DIFFERENT METHODS ON EXPERIMENTAL DATASETS FOR TARGET DETECTION

Dataset	Num of Bands	Compared Methods								Baseline	
		R-GAN	DSEBS	EFPDC	OCF	SBSS	ASPS	SOPSR	FNGBS		
SD-1	5	<b>0.998±0.000</b>	0.974	<b>0.998</b>	0.990	0.997	0.990	0.991	0.989	0.994	1
	10	<b>0.999±0.000</b>	0.984	<b>0.997</b>	0.997	<b>0.999</b>	0.993	<b>0.998</b>	0.997	0.994	
	15	<b>0.999±0.000</b>	0.987	<b>0.997</b>	0.998	<b>0.997</b>	0.996	<b>0.988</b>	0.997	0.996	
	20	<b>0.999±0.000</b>	<b>0.998</b>	<b>0.998</b>	0.998	0.998	0.996	<b>0.998</b>	0.997	0.998	
	25	<b>0.999±0.000</b>	<b>0.998</b>	<b>0.998</b>	0.998	0.998	0.998	<b>0.998</b>	<b>0.998</b>	<b>0.999</b>	
	30	<b>0.999±0.000</b>	<b>0.998</b>	<b>0.998</b>	<b>0.999</b>	0.998	<b>0.999</b>	<b>0.998</b>	<b>0.998</b>	<b>0.999</b>	
	Peaked AUC	<b>0.999±0.000</b>	0.998	0.998	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	0.998	0.998	<b>0.999</b>	
TC	Low PFA Range	<b>9.1e-3±7e-5</b>	0.003	0.008	0.005	0.008	0.004	0.006	0.004	0.008	0.009
	5	0.997±0.001	0.994	0.970	0.985	<b>0.998</b>	0.995	0.988	0.992	0.995	1
	10	<b>0.999±0.000</b>	0.994	0.984	0.989	0.998	0.996	0.991	0.977	0.995	
	15	<b>0.999±0.000</b>	<b>0.996</b>	0.988	<b>0.999</b>	0.998	<b>0.999</b>	0.997	0.995	0.989	
	20	<b>0.999±0.000</b>	0.995	0.998	<b>0.999</b>	<b>0.999</b>	0.984	0.998	0.998	0.990	
	25	<b>0.999±0.000</b>	0.993	0.998	<b>0.999</b>	0.998	<b>0.999</b>	0.995	0.998	0.996	
	30	<b>0.999±0.000</b>	<b>0.996</b>	<b>0.999</b>	<b>0.999</b>	0.995	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	0.998	
TC	Peaked AUC	<b>0.999±0.000</b>	0.996	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	0.998	
	Low PFA Range	<b>9.0e-3±6e-5</b>	0.008	0.007	0.008	0.008	0.008	0.008	0.008	0.008	0.010

\* The best-performing method at each data scale is in **bold**. The peaked accuracy of each method is underlined

TABLE IV  
OA VALUES OF DIFFERENT METHODS ON THE PAVIAU DATASET FOR CLASSIFICATION WITH KNN

Dataset	Num of Bands	Compared Methods								Baseline	
		R-GAN	DSEBS	EFPDC	OCF	SBSS	ASPS	SOPSR	FNGBS		
PaviaU	5	0.837±0.010	0.828	0.840	<b>0.843</b>	0.624	0.833	0.832	0.828	0.784	0.867
	10	0.857±0.007	0.832	<b>0.858</b>	0.843	0.759	0.850	0.830	0.840	<u>0.817</u>	
	15	<b>0.867±0.009</b>	0.847	0.863	0.850	0.798	0.858	0.861	0.853	0.796	
	20	<b>0.868±0.005</b>	0.850	0.864	0.844	0.808	0.854	0.867	0.850	0.799	
	25	<b>0.870±0.005</b>	0.856	<u>0.868</u>	0.846	0.853	0.864	0.868	<u>0.869</u>	0.802	
	30	<b>0.871±0.004</b>	<u>0.860</u>	0.867	<u>0.851</u>	<u>0.866</u>	<u>0.870</u>	<u>0.869</u>	0.866	0.807	
	Peaked OA	<b>0.871±0.004</b>	0.860	0.868	0.851	0.866	0.870	0.869	0.869	0.817	

\* 10% samples from each class based on selected bands are randomly chosen as the training set and the remaining 90% samples are used for test

\* The best-performing method at each data scale is in **bold**. The peaked accuracy of each method is underlined

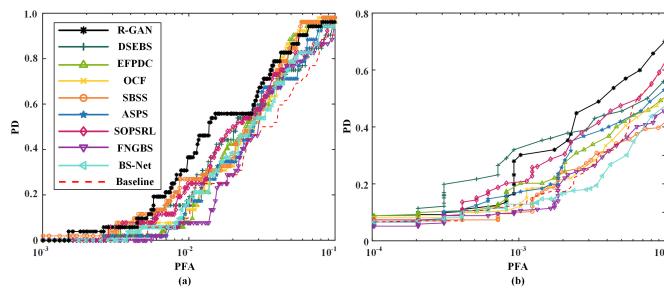


Fig. 8. ROC curves (in low PFA range) of different methods on (a) GA and (b) LA datasets for anomaly detection.

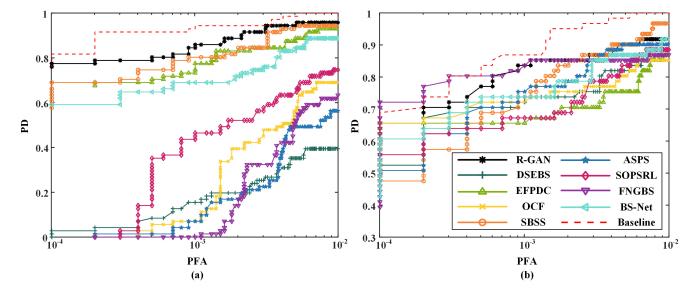


Fig. 9. ROC curves (in low PFA range) of different methods on (a) SD-1 and (b) TC datasets for target detection.

in Fig. 11. The discriminator in Fig. 11(a) establishes adversarial learning between the input and reconstructed spectral vectors, forcing the imitation to be similar. Fig. 11(b) imposes a prior distribution in the latent space and regresses the reconstructed spectral vector to the input only with the reconstruction loss. R-GAN is based on the network architecture shown in Fig. 11(c) realizing the alignments in both spectral and latent spaces. We report the anomaly detection performance with two band scales (10 and 20) in Table V, and the three

schemes in Fig. 11 are represented by  $s$ ,  $z$ , and  $s + z$ , respectively. The spectral-latent adversarial scheme achieves a dominant position in all the experimental datasets. For example, on the SD-1 dataset, latent, spectral, and spectral-latent present an increased whole range AUC value on the studied data scales. When detect with 20 bands, the added spectral discriminator blooms the latent pattern by 0.5% (28.6% in low PFA range AUC value). Overall, the spectral-latent scheme benefiting from the two alignments in both spectral and latent

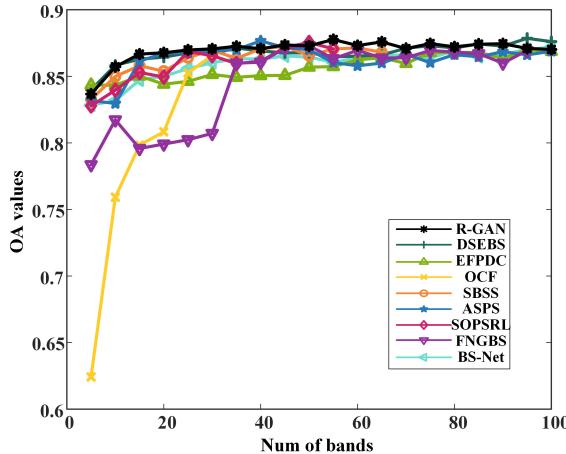


Fig. 10. Comparison of OA values using KNN under different numbers of bands selected for the PaviaU dataset.

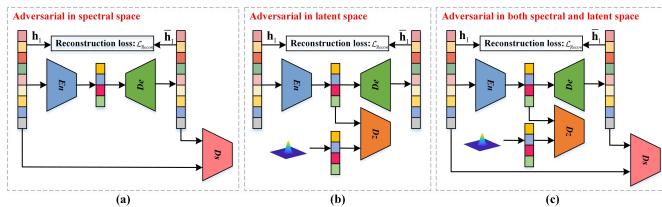


Fig. 11. Illustration of three adversarial patterns. (a) Adversarial in spectral space. (b) Adversarial in latent space. (c) Adversarial in spectral and latent spaces.

TABLE V

AUC VALUES OF DIFFERENT NETWORK ARCHITECTURES ON EXPERIMENTAL DATASETS FOR ANOMALY DETECTION

Data	Network	Anomaly Detection Performance (RX)		
		10 bands	20 bands	20 bands (Low PFA Range)
SD-1	s	0.977±0.004	0.970±0.005	6.5e-5±5e-5
	z	0.977±0.005	0.969±0.006	2.1e-4±0.0003
	s+z	<b>0.982±0.003</b>	<b>0.974±0.003</b>	<b>2.7e-4±0.0001</b>
SD-2	s	0.976±0.004	0.973±0.004	2.1e-4±0.0003
	z	0.980±0.005	0.972±0.005	2.2e-4±0.0003
	s+z	<b>0.981±0.001</b>	<b>0.978±0.001</b>	<b>1.3e-3±0.0001</b>
TC	s	0.992±0.001	0.993±0.001	7.0e-3±0.0003
	z	0.992±0.001	0.994±0.001	7.1e-3±0.0004
	s+z	<b>0.995±0.001</b>	<b>0.995±0.001</b>	<b>7.2e-3±0.0001</b>
GA	s	0.958±0.014	0.966±0.005	9.5e-4±0.0004
	z	0.967±0.008	0.966 ± 0.004	8.4e-04±0.0003
	s+z	<b>0.970±0.002</b>	<b>0.971±0.003</b>	<b>1.2e-3±0.0001</b>
LA	s	0.985±0.003	0.990±0.001	0.0041±0.0005
	z	0.987±0.001	0.990±0.001	0.0042±0.0004
	s+z	<b>0.988±0.001</b>	<b>0.990±0.001</b>	<b>4.5e-3±0.0003</b>

spaces achieves the highest average AUC values no matter whole or low PFA range and relatively low standard deviation, which indicates better FE ability and stability.

### E. Component Analysis

As stated in Section II, our method is implemented with three important steps. Here, we discuss their contribution to the final performance.

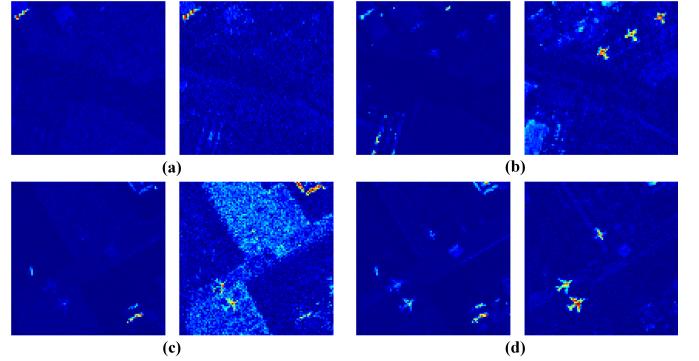


Fig. 12. FE of component analysis: (a) anomaly and target detection maps of R-BS on the SD-1 dataset, (b) anomaly and target detection maps of R-GAN on the SD-1 dataset, (c) anomaly and target detection maps of R-BS on the SD-2 dataset, and (d) anomaly and target detection maps of R-GAN on the SD-2 dataset.

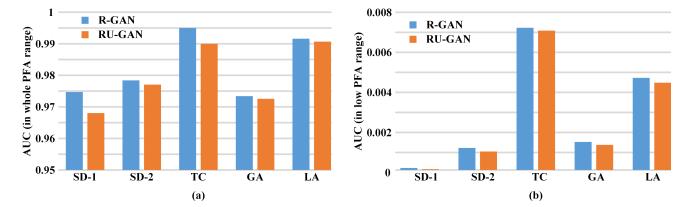


Fig. 13. Rank-aware strategy of component analysis: (a) anomaly detection performance of RU-GAN and R-GAN on experimental datasets with data scale of 20 evaluated with AUC in the whole PFA range and (b) anomaly detection performance of RU-GAN and R-GAN on experimental datasets with data scale of 20 evaluated with AUC in the low PFA range.

1) *Feature Extraction*: The first step provides the information-centralized referential latent feature for follow-up actions. A BS method based on the rank of slices dubbed as R-BS is compared with R-GAN to evaluate the necessity of referential information. Taking the SD dataset and data scale of 20 for example, the detection maps are exhibited in Fig. 12. It is obvious that R-GAN leads to a more prominent target and anomaly, and less noisy background. The performance gap is caused by the information-centralized guiding feature and indicates that the bands selected with the guiding feature contain more useful information for the detection task.

2) *Rank-Aware Strategy*: In R-GAN, a rank-aware strategy is conducted to keep the referential latent feature informative while reducing dimension. Here, we design a rank-unaware strategy BS method (RU-GAN), that is, the bands are selected by mapping latent features into the original data space with SSIM directly. The bar chart in Fig. 13 shows the performance of RU-GAN and R-GAN on experimental datasets for anomaly detection tasks evaluated with two criteria. No matter which kind of AUC is used for assessment, R-GAN is higher than RU-GAN on all the datasets, which indicates that the rank-aware strategy containing two rank operations can select salient features without key information loss.

3) *Mapping Saliency Method*: SSIM considering brightness, contrast, and structure is used for mapping salient features into the original data space. Another two popular distance metrics, i.e., MSE and ED, are discussed. Note that, the larger SSIM and smaller MSE and ED, the more likely

TABLE VI  
AUC VALUES OF DIFFERENT MAPPING SALIENCY METHODS ON EXPERIMENTAL DATASETS FOR ANOMALY DETECTION

Data	Mapping Methods					
	MSE		ED		SSIM	
	20 bands	20 bands (Low PFA Range)	20 bands	20 bands (Low PFA Range)	20 bands	20 bands (Low PFA Range)
SD-1	0.972±0.003	1.1e-4±0.0001	0.968±0.007	9.7e-5±7.4e-5	<b>0.974±0.003</b>	<b>2.7e-4±0.0001</b>
SD-2	0.976±0.003	<b>1.5e-3±0.0003</b>	0.977±0.003	1.1e-3±0.0003	<b>0.978±0.001</b>	1.3e-3±0.0001
TC	0.992±0.002	6.9e-3±0.0006	0.995±0.001	7.0e-3±0.0002	<b>0.995±0.001</b>	<b>7.2e-3±0.0001</b>
GA	0.966±0.006	9.3e-4±0.0003	0.969±0.004	1.0e-3±0.0003	<b>0.971±0.003</b>	<b>1.2e-3±0.0001</b>
LA	0.989±0.002	3.8e-3±0.0006	0.990±0.001	3.7e-3±0.0005	<b>0.990±0.001</b>	<b>4.5e-3±0.0003</b>

TABLE VII  
TIME COMPLEXITY FOR THE COMPARED METHODS

Methods	Time complexity	Parameters
R-GAN	$O(B \times \text{Num} \times h + D \times \text{Num}^{3/2} + B \times T \times \text{Num})$	$B$ : original band number $\text{Num}$ : pixel number
DSEBS	$O(B \times \text{Num} + B^2 \times \text{Num} + B^3)$	$h$ : max hidden nodes number
EFPDC	$O(2B^2 \times \text{Num} + B^2)$	$D$ : latent feature size
OCF	$O(B^2 \times \text{Num} + B^3 + B \times P)$	$P$ : selected band number
SBSS	$O(B \times \text{Num})$	$n$ : selected block number
ASPS	$O(B^2 \times \text{Num} + B \times n \times b^2)$	$b$ : block size
SOPSRL	$O(\sum_{m=1}^M T_{1t} \times T_{2t} \times b_t \times \max(b_t, n_t)^2)$	$m$ : selected sample number $b_t, n_t$ : sample matrix size
FNGBS	$O(B^2 \times \text{Num} + 2B + B \times \text{Num})$	$T_{1/2t}$ : iteration number of Algorithm 1/2
BS-Net	$O(B \times \text{Num} \times h)$	

the corresponding band is to be selected. Table VI reports the AUC values integrated with the whole and low PFA ranges of different mapping saliency methods on experimental datasets for anomaly detection. SSIM shows its superiority in mapping saliency for all the datasets except the low PFA range AUC on the SD-2 dataset. Exampled with the GA dataset, SSIM outperforms MSE and ED by 0.5% and 0.2% in the whole PFA range and 29.03% and 20% in the low PFA range.

#### F. Computational Complexity Analysis

The conduction of the proposed algorithm can be decomposed into three steps. In the following, we will discuss the theoretical computational complexity for each step.

1) *Extraction Step*: As stated in Section II-A, a GAN containing a generator (an encoder and a decoder) and two discriminators (latent discriminator  $D_z$  and spectral discriminator  $D_s$ ) is constructed for information-centralized FE. Only the forward propagation in (9) is considered for computational complexity computation, and it takes  $O(B \times \text{Num} \times h)$ . Here,  $\text{Num} = N \times M$  denotes the number of pixels in an his, and  $h$  is the max hidden nodes' number in the encoder.

2) *Selection Step*: In this step, the complexity comes from rank computation including both tensor and matrix version. The tensor rank implemented by t-SVD in Fig. 3 generates  $O(D \times \text{Num}^{3/2})$  complexity, which is the maximum between DFT/IDFT (implemented by FFT/IFFT) [ $O(D \times \log D \times \text{Num})$ ] and slicewise SVD [ $O(D \times \text{Num}^{3/2})$ ]. The matrix rank takes  $O(D \times \text{Num}^{3/2})$ . Here,  $D$  is the size of information-centralized latent feature outputted from the encoder.

3) *Mapping Step*: In this step, the similarity between refined feature  $\mathcal{N} = \{\mathcal{N}_t\}_{t=1}^T$  and HSI  $\mathcal{H} = \{\mathcal{H}_b\}_{b=1}^B$  is evaluated by SSIM and takes  $O(B \times T \times \text{Num})$  complexity.

In a sum, R-GAN takes  $O(B \times \text{Num} \times h + D \times \text{Num}^{3/2} + B \times T \times \text{Num})$ . EFPDC, OCF, SBSS, ASPS, and SOPSRL have reported complexity in respective articles, and the complexity of FNGBS can be derived according to ASPS. Although the core code of DSEBS is not open source, we can estimate the computation is more than  $O(B \times \text{Num} + B^2 \times \text{Num} + B^3)$  from the algorithm. As listed in Table VII, our better-performing R-GAN shares the same order of complexity with the most of methods except for SBSS (linear complexity).

## IV. CONCLUSION

In this article, we propose a novel BS method named R-GAN to address the spectral saliency in HSIs. The main idea behind R-GAN is to select the most salient subset containing as much as possible information and minimize the performance degradation in the follow-up tasks. R-GAN consists of three components, i.e., information-centralized FE, rank-aware salient BS, and mapping saliency to the original data space. The contributions of components are evaluated, and the effectiveness and superiority of R-GAN are demonstrated by comparing with other BS methods on detection and classification tasks. For example, the subset (ten bands) selected by R-GAN from the SD-1 dataset boosts the baseline detection performance by 13.7%. The classification performance on the PaviaU dataset with 30 bands is 0.46% better than the baseline. Note that our experiments are based on the ground truths generated by AED, which is still a rough guide. The readers can obtain a more convincing ground truth through high-resolution panchromatic data (pansharpening) or super-resolution methodology.

## REFERENCES

- [1] Z. Yang *et al.*, “CARS: Continuous evolution for efficient neural architecture search,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1826–1835.
- [2] S. Lin *et al.*, “Towards optimal structured CNN pruning via generative adversarial learning,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2785–2794.
- [3] H. You *et al.*, “ShiftAddNet: A hardware-inspired deep network,” in *Proc. NeurIPS*, 2020, pp. 1–14.
- [4] Y. Tang *et al.*, “Manifold regularized dynamic network pruning,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5018–5028.
- [5] Y. Tang *et al.*, “SCOP: Scientific control for reliable neural network pruning,” 2020, *arXiv:2010.10732*.

- [6] R. R. Lam, O. Zahm, Y. M. Marzouk, and K. E. Willcox, "Multifidelity dimension reduction via active subspaces," *SIAM J. Sci. Comput.*, vol. 42, no. 2, pp. A929–A956, Jan. 2020.
- [7] P. Su, D. Liu, X. Li, and Z. Liu, "A saliency-based band selection approach for hyperspectral imagery inspired by scale selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 572–576, Apr. 2018.
- [8] Z. Zhu, J. Hou, J. Chen, H. Zeng, and J. Zhou, "Hyperspectral image super-resolution via deep progressive zero-centric residual learning," *IEEE Trans. Image Process.*, vol. 30, pp. 1423–1438, 2021.
- [9] R. Zhao, Q. Wang, J. Fu, and L. Ren, "Exploiting block-sparsity for hyperspectral Kronecker compressive sensing: A tensor-based Bayesian method," *IEEE Trans. Image Process.*, vol. 29, pp. 1654–1668, 2020.
- [10] S. Song, H. Zhou, L. Gu, Y. Yang, and Y. Yang, "Hyperspectral anomaly detection via tensor-based endmember extraction and low-rank decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1772–1776, Oct. 2020.
- [11] Z. Li and Y. Zhang, "Hyperspectral anomaly detection via image super-resolution processing and spatial correlation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2307–2320, Mar. 2021.
- [12] X. Song, L. Zou, and L. Wu, "Detection of subpixel targets on hyperspectral remote sensing imagery based on background endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2365–2377, Mar. 2021.
- [13] R. Hang, Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya, "Hyperspectral image classification with attention-aided CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2281–2293, Mar. 2021.
- [14] Z. Dou, K. Gao, X. Zhang, H. Wang, and L. Han, "Band selection of hyperspectral images using attention-based autoencoders," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 147–151, Jan. 2021.
- [15] X. Wei, W. Zhu, B. Liao, and L. Cai, "Scalable one-pass self-representation learning for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4360–4374, Jul. 2019.
- [16] C. Sui, C. Li, J. Feng, and X. Mei, "Unsupervised manifold-preserving and weakly redundant band selection method for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1156–1170, Feb. 2020.
- [17] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, Jan. 2011.
- [18] S. Jia *et al.*, "3-D Gaussian–Gabor feature extraction and selection for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8813–8826, Nov. 2019.
- [19] W. Li, Z. Wang, L. Li, and Q. Du, "Feature extraction for hyperspectral images using local contain profile," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5035–5046, Dec. 2019.
- [20] H. Gao, Z. Chen, and C. Li, "Sandwich convolutional neural network for hyperspectral image classification using spectral feature enhancement," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3006–3015, 2021.
- [21] H. Su, H. Yang, Q. Du, and Y. Sheng, "Semisupervised band clustering for dimensionality reduction of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1135–1139, Nov. 2011.
- [22] H. Su, P. Du, and Q. Du, "Semi-supervised dimensionality reduction using orthogonal projection divergence-based clustering for hyperspectral imagery," *Proc. SPIE Opt. Eng.*, vol. 51, no. 11, 2012, Art. no. 111715.
- [23] H. Su and Q. Du, "Hyperspectral band clustering and band selection for urban land cover classification," *Geocarto Int.*, vol. 27, no. 5, pp. 395–411, 2012.
- [24] W. Sun, J. Peng, G. Yang, and Q. Du, "Correntropy-based sparse spectral clustering for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 484–488, Mar. 2020.
- [25] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 773–777, May 2017.
- [26] P. Hu, X. Liu, Y. Cai, and Z. Cai, "Band selection of hyperspectral images using multiobjective optimization-based sparse self-representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 452–456, Mar. 2019.
- [27] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.
- [28] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, Jan. 2016.
- [29] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4823–4833, Jul. 2019.
- [30] X. Kang, C. Li, S. Li, and H. Lin, "Classification of hyperspectral images by Gabor filtering based deep network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1166–1178, Apr. 2018.
- [31] W. Yu, M. Zhang, and Y. Shen, "Spatial revising variational autoencoder-based feature extraction method for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1410–1423, Feb. 2021.
- [32] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Jul. 2016.
- [33] Y. Cai, X. Liu, and Z. Cai, "Bs-Nets: An end-to-end framework for band selection of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, pp. 1969–1984, 2020.
- [34] I. J. Goodfellow *et al.*, "Generative adversarial networks," 2014, *arXiv:1406.2061*.
- [35] 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.
- [36] Z. Zhong, J. Li, D. A. Clausi, and A. Wong, "Generative adversarial networks and conditional random fields for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3318–3329, Jul. 2020.
- [37] M. Zhang, M. Gong, Y. Mao, J. Li, and Y. Wu, "Unsupervised feature extraction in hyperspectral images based on Wasserstein generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2669–2688, May 2019.
- [38] M. Lin *et al.*, "HRank: Filter pruning using high-rank feature map," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1529–1538.
- [39] S. Pidhorskyi, R. Almohsen, D. A. Adjeroh, and G. Doretto, "Generative probabilistic novelty detection with adversarial autoencoders," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6823–6834.
- [40] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4666–4679, Jul. 2020.
- [41] Q. Wang and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, Dec. 2019.
- [42] De Rosal Ignatius Moses Setiadi, "PSNR vs SSIM: Imperceptibility quality assessment for image steganography," *Multimedia Tools Appl.*, vol. 80, pp. 8423–8444, Mar. 2020.
- [43] 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.
- [44] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 227–239, Jan. 2016.
- [45] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5028–5039, Jun. 2021.
- [46] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, early access, Sep. 9, 2021, doi: [10.1109/MGRS.2021.3105440](https://doi.org/10.1109/MGRS.2021.3105440).
- [47] Z. Zou and Z. Shi, "Hierarchical suppression method for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 330–342, Jan. 2016.



**Xin Zhang** received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2019, where she is currently pursuing the Ph.D. degree with the State Key Laboratory of Integrated Services Networks, Image Coding and Processing Center.

Her research interests include remote sensing image processing and efficient deep learning.



**Weiyong Xie** (Member, IEEE) received the B.S. degree in electronic information science and technology from the University of Jinan, Jinan, China, in 2011, the M.S. degree in communication and information systems from Lanzhou University, Lanzhou, China, in 2014, and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2017.

She is currently an Associate Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. Her research interests include neural networks, machine learning, hyperspectral image processing, and high-performance computing.



**Yunsong Li** (Member, IEEE) received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 1999 and 2002, respectively.

He joined the School of Telecommunications Engineering, Xidian University, in 1999, where he is currently a Professor. He is also the Director of the State Key Laboratory of Integrated Services Networks, Image Coding and Processing Center. His research interests include image and video processing, and high-performance computing.



**Jie Lei** (Member, IEEE) received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2006 and 2010, respectively.

He had been a Visiting Scholar with the Department of Computer Science, University of California at Los Angeles, Los Angeles, CA, USA, from 2014 to 2015. He is currently an Associate Professor with the School of Telecommunications Engineering, Xidian University, where he is also a member with the State Key Laboratory of Integrated Services Networks, Image Coding and Processing Center. His research interests include image and video processing, computer vision, and customized computing for big-data applications.



**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. She is also an Adjunct Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is also a fellow of the SPIE-International Society for Optics and Photonics. She received the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was the Co-Chair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013, and the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was the General Chair of the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held in Shanghai in 2012. She has served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the *Journal of Applied Remote Sensing*, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.



**Geng Yang** received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2019, where she is currently pursuing the Ph.D. degree with the State Key Laboratory of Integrated Services Networks, Image Coding and Processing Center.

Her research interests include remote sensing image processing, computer vision, and efficient deep learning.