

# Hyperspectral Anomaly Detection With Robust Graph Autoencoders

Ganghui Fan<sup>ID</sup>, Yong Ma<sup>ID</sup>, Xiaoguang Mei<sup>ID</sup>, Member, IEEE , Fan Fan<sup>ID</sup>, Jun Huang<sup>ID</sup>, and Jiayi Ma<sup>ID</sup>, Senior Member, IEEE

**Abstract**—Anomaly detection of hyperspectral data has been gaining particular attention for its ability in detecting targets in an unsupervised manner. Autoencoder (AE), together with its variants can not only extract intrinsic features automatically but also detect anomalies that differ dramatically from others. Many AE-driven algorithms are, thus, proposed for anomaly detection in hyperspectral imagery (HSI), but they suffer from two problems: 1) when there exist anomalies in the training set, AE can generalize so well that it can also learn the abnormal patterns well, thereby reducing the ability to distinguish anomalies from the background and 2) geometric structure among samples are lost in latent space of AE, which is vital in hyperspectral anomaly detection. To tackle these problems, we propose a robust anomaly detector based on the AE framework, named robust graph AE (RGAE) detector, in this article. To be specific, we propose a robust AE framework with  $\ell_{2,1}$ -norm that is robust to noise and anomalies during training. Meanwhile, we embed a superpixel segmentation-based graph regularization term (SuperGraph) into AE. This strategy can preserve the geometric structure and the local spatial consistency of HSI simultaneously and also effectively reduce the searching space and execution time for each pixel. Extensive experiments are conducted on five datasets, and the results demonstrate that our method has a better detection performance, after comparing with other state-of-the-art hyperspectral anomaly detectors.

**Index Terms**—Autoencoders (AEs), graph Laplacian matrix, hyperspectral anomaly detection,  $\ell_{2,1}$ -norm, spatial information.

## I. INTRODUCTION

**H**YPERSPECTRAL imagery (HSI) is a 3-D cube stacked with hundreds of narrowband images over a wide range of wavelengths, providing more detailed spectral information than infrared and multispectral images [1], [2]. Materials contain different reflective (or emissive) properties in the visible-infrared region of the spectrum; HSI is, thus, feasible to distinguish ground materials from each other and

Manuscript received March 30, 2021; revised May 25, 2021 and June 14, 2021; accepted July 11, 2021. Date of publication July 26, 2021; date of current version January 17, 2022. This work was supported by the National Natural Science Foundation of China under Grant 61903279, Grant 62075169, Grant 62003247, and Grant 62061160370. (Corresponding author: Xiaoguang Mei.)

Ganghui Fan is with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: 2015301200292@whu.edu.cn).

Yong Ma, Xiaoguang Mei, Fan Fan, Jun Huang, and Jiayi Ma are with the Electronic Information School, Wuhan University, Wuhan 430072, China, and also with the Institute of Aerospace Science and Technology, Wuhan University, Wuhan 430079, China (e-mail: mayong@whu.edu.cn; meixiaoguang@gmail.com; fanfan@whu.edu.cn; junhwong@whu.edu.cn; jyma2010@gmail.com).

Digital Object Identifier 10.1109/TGRS.2021.3097097

shows its advantages in unmixing, classification, and target detection [3]–[6]. After decades of development, HSI has become an important part of the remote sensing domain and has been widely applied in many remote sensing tasks, such as environmental monitoring and intelligent agriculture [7]–[10].

As an essential part of HSI applications, anomaly detection aims at detecting potential targets that differ obviously from the dominant background in an unsupervised manner [11]. Anomaly detection in HSI usually possesses the following characteristics: 1) anomalies are small objects occupying only a relatively very small part of the entire image; 2) anomalies are significantly different from the background in terms of spectral signatures; 3) no prior spectral information about the background or anomalies are available; and 4) in some cases, anomalous targets are mixed with background and appear as mixed pixels or subpixels in real-world scenes. These properties make hyperspectral anomaly detection became a hot but difficult topic in the remote sensing domain [12]–[14].

As the *a priori* spectral information about background or targets is unknown in advance, it is still a challenge to separate these two components apart in hyperspectral anomaly detection. In recent years, a variety of algorithms are proposed to get rid of this dilemma, which can be divided into two categories.

- 1) *Distance-Based*: The background is modeled with Gaussian multivariate distribution assumption, and anomalies are detected based on the Euclidean distance or the Mahalanobis distance. Among them, Reed–Xiaoli (RX) [15] is now a benchmark in hyperspectral anomaly detection, which derives a set of advanced variants, such as local RX and cluster kernel RX [16]–[18].
- 2) *Representation-Based*: These algorithms reconstruct the background from a global perspective under the assumption that anomalies usually behave differently from the background [19], [20].

Abundance and dictionary-based low-rank decomposition (ADLR) [21], [22] exploits spectral unmixing to obtain distinctive abundance vectors, and the mean shift clustering is used to construct a more representative dictionary. The detection result shows its superiority in maintaining a low false alarm rate, whereas the abovementioned anomaly detectors only exploit low-level features, which limits the detection performance.

Recent years have witnessed the booming of deep learning in many fields, including, but not limited to, computer vision, natural language processing, and personalized

recommendation [23]–[26]. As one of the unsupervised neural networks, autoencoder (AE) captures abstract and hierarchical features in an automatic way, which has been successfully applied in many remote sensing tasks. Hyperspectral anomaly detection based on the AE framework is now getting more and more attention; a multitude of articles are proposed to promote detection performance. Chang *et al.* [27] trained two sparse AEs with pixels in the whole window and the outer window separately; anomalies can be selected by subtracting the reconstruction error of these two AEs. In order to facilitate the detection performance, Zhao and Zhang [28] decomposed the HSI into the low-rank part and sparse part, the low-dimensional latent features are learned by two stacked AEs separately, and the final detection map can be obtained with the local Mahalanobis distance. Ma *et al.* [29] proposed an online sparse AE-based detector to avoid the heavy computational cost. This detector can find the pixels containing significant features and then updates the mismatched model, so it is lightweight without degradation of detection accuracy.

Considering that many of the above AE-based detectors project each pixel into latent space separately and neglect the connections among them, graph regularization, one of the methods that can preserve the local geometric structure, has been widely exploited in the remote sensing domain. In [30], the AE framework is embedded with a manifold constraint to maintain locality characteristics; the detection results reveal superiority compared with the original version. Cheng and Wang [31] embedded the graph regularization and total variation regularization into the low-rank representation framework to preserve the local geometric structure. A vertex and edge-weighted graph together with a pixel selection process are utilized in [32] to locate the anomaly targets, which ensures a satisfying detection performance and robust to noise.

AE network performs well with the condition that it is properly trained. However, as there is no *a priori* spectral signature about the background or targets, the training set is prone to be corrupted by anomalous pixels. When there exist anomalies in the training set, AE “generalizes” so well that it learns abnormal patterns well and reconstructs anomalies with small reconstruction errors [33]. This does harm to the performance of anomaly detection in HSI. Worse still, the construction of graph Laplacian matrix is time-consuming because it has to go through the entire dataset for each sample. Moreover, spatial information is not taken into account, which is vital in hyperspectral image processing.

In order to handle the abovementioned drawbacks, a robust graph AE (RGAE) detector is proposed in this article. First, inspired by robust PCA [34] that adopts  $\ell_{2,1}$ -norm to avoid the effects of outliers, we introduce a robust AE (RAE) framework based on  $\ell_{2,1}$ -norm, as it normalizes the gradient for each sample. This simple step can alleviate the impact of anomalous part on parameter fine-tuning, which enables the AE network more robust to noise and anomalies than the Frobenius norm during training [35]. Moreover, in order to preserve the local geometric structure of the given high-dimensional data, a graph regularization term is embedded into the AE framework. To avoid the heavy computational cost in constructing the traditional graph and preserve the spatial information of

HSIs, a superpixel segmentation-based graph regularization term, named SuperGraph, is proposed and embedded into the RAE framework. SuperGraph can not only preserve the geometric structure and spatial consistency simultaneously but also effectively reduce the execution time.

The main contributions of the proposed RGAE for hyperspectral anomaly detection are listed as follows.

- 1) We propose an RAE framework to handle the conditions where serious noise and anomalies are present in the training set. The  $\ell_{2,1}$ -norm is introduced into the objective function of RAE because it is less sensitive to outliers than the Frobenius norm. The proposed RAE framework possesses the ability in selecting anomalies out with reconstruction residuals, especially when pure background training sets cannot or hard to obtain. To the best of our knowledge, it is the first time to introduce the  $\ell_{2,1}$ -norm into the AE framework to overcome the drawback of AE from being sensitive to outliers in the training set.
- 2) A graph constraint is embedded into the latent space of the proposed framework to maintain the geometric structure of the data. As the construction of graph Laplacian matrix costs a lot of time, we further adopt a superpixel segmentation strategy to effectively reduce the searching space and time. This graph regularization term named SuperGraph can preserve spatial consistency as well.

This article is an expanded version of our previous work [36]. The new contributions include the following three aspects. First, we describe the proposed RAE framework in more detail. Second, we provide a comprehensive evaluation of the proposed RGAE detector by conducting extensive experiments, such as execution time, sensitivity to parameters, and ablation study. Finally, we add more state-of-the-art methods on more challenging large-scale datasets to demonstrate the superiority of RGAE on hyperspectral anomaly detection.

The rest of this article is arranged as follows. Section II briefly introduces the framework of AE and graph regularization. In Section III, the proposed RGAE detector will be described in detail. Extensive experiments demonstrating the rationality and superiority of the proposed RGAE algorithm are conducted in Section IV. Section V draws the conclusions of our article.

## II. BACKGROUND

### A. AE Framework

As a typical unsupervised feature extractor, AE is a symmetric neural network that consists of an input layer, an output layer, and several hidden layers. Take the traditional AE networks with  $(M - 1)$  hidden layers as an example. For each sample  $\mathbf{x}_i \in \mathbb{R}^L$  with  $L$  features, the output  $\mathbf{h}_i^{(m)}$  for the  $m$ th layer of AE corresponding to  $\mathbf{x}_i$  is

$$\mathbf{h}_i^{(m)} = g\left(\mathbf{W}^{(m)}\left(\mathbf{h}_i^{(m-1)}\right)^T + \mathbf{b}^{(m)}\right)^T, \quad m = 1, \dots, M. \quad (1)$$

Note that  $\mathbf{h}_i^{(0)} = \mathbf{x}_i$ ,  $\mathbf{z}_i = \mathbf{h}_i^{(\frac{M}{2})}$ , and  $\tilde{\mathbf{x}}_i = \mathbf{h}_i^{(M)}$  are the input, the latent feature, and the output of AE, respectively.  $\mathbf{W}^{(m)}$  and

$\mathbf{b}^{(m)}$  represent the weights and biases between the  $(m - 1)$ th and  $m$ th layers. The sigmoid function  $g(\cdot) = 1/(1 + \exp(-(\cdot)))$  is chosen as the activation for each layer of AE in this article. AE learns the abstract and hierarchical characteristics automatically by reconstructing the original inputs as close as possible with encoder and decoder.

In fact, besides feature extraction, the AE network can also be applied for anomalies removal with the criterion of reconstruction residual. To be specific, the AE network is upgraded as an anomaly detector with a pure training set; anomalies are selected out with reconstruction residuals. However, as the labels are unknown in advance, to make full use of the data, AE-based anomaly detectors always exploit all of the samples for training. As anomalies usually occupy a relatively very small portion of the entire dataset, AE inclines to represent the background well while fail to learn the patterns of anomalies. The reconstruction residuals for the dominant background are small, while anomalies are just the opposite, and abnormal samples are, thus, selected out. However, the performance of these AE-based anomaly detectors [27], [30], [37] is sensitive to noise and anomalies; the existing outliers in the training set may play a major role in updating parameters of AE. This does harm to the discrimination between the background and anomalies, but it is hard to determine when to stop the iteration.

### B. Graph Regularization

A good performance always closely related to good semantic features, while combinations of features are usually more discriminative than individual ones, especially in detection tasks [31], [38]. To this end, Li *et al.* [39] introduced the pairwise graph constraint to ensure that the local intrinsic geometric structure of latent space is consistent with that in the original domain. The constraint is formulated as (2) under the local invariance assumption: if two adjacent samples are similar with respect to their features in the original space, their representations should be close to each other in the subspace as well

$$\min_{\mathbf{Z}} \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N w_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 = \min_{\mathbf{Z}} \frac{1}{N} \text{Tr}[\mathbf{Z}^T \mathbf{L} \mathbf{Z}] \quad (2)$$

where  $N$  is the number of samples and  $\mathbf{z}_i$  is the  $i$ th column of  $\mathbf{Z}$ , which denotes all the training samples in the latent space.  $w_{ij}$  is the  $(i, j)$ th entry of the adjacency matrix  $\mathbf{W}$  representing the similarity of  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Specifically,  $w_{ij}$  is defined as

$$w_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma^2}\right) & \mathbf{x}_i \in \mathcal{N}_k(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where  $\sigma$  is a scalar parameter and  $\mathcal{N}_k(\mathbf{x}_i)$  is the set of  $k$ -nearest neighbors of  $\mathbf{x}_i$ .  $\mathbf{L} = \mathbf{D} - \mathbf{W}$  is the graph Laplacian matrix, wherein the degree matrix  $\mathbf{D}$  is a diagonal matrix. The  $i$ th diagonal element of  $\mathbf{D}$  denotes the degree of the  $i$ th node (pixel) and is the summation of the  $i$ th row of adjacency

matrix  $\mathbf{W}$

$$d_{ii} = \sum_{j=1}^N w_{ij}. \quad (4)$$

In condition that sample  $\mathbf{x}_i$  is among the  $k$ -nearest neighbors of sample  $\mathbf{x}_j$  or  $\mathbf{x}_j$  is among the  $k$ -nearest neighbors of  $\mathbf{x}_i$ ,  $w_{ij}$  tends to be large. In order to minimize (2), the distance between the  $i$ th and  $j$ th samples in the latent space is enforced to be small, and the similarity between samples  $i$  and  $j$  is, thus, preserved. The graph regularization encourages samples within the same semantic region to share similar representation, and vice versa. Graph regularization is beneficial in distinguishing different clusters from each other and has been widely used in many fields [31], [40], [41].

## III. PROPOSED METHOD

### A. Robust Autoencoder Framework

We first explain that the traditional AE framework is sensitive to heavy noise and anomalies during training from a gradient perspective. Given a matrix  $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N \in \mathbb{R}^{N \times L}$  with  $N$  samples and its reconstructed version  $\tilde{\mathbf{X}}$ , the derivative of the objective function of AE with respect to  $\tilde{\mathbf{x}}_i$  is

$$\frac{\partial \|\tilde{\mathbf{X}} - \mathbf{X}\|_F^2}{\partial \tilde{\mathbf{x}}_i} = 2(\tilde{\mathbf{x}}_i - \mathbf{x}_i). \quad (5)$$

The gradient is related to the magnitude of the corresponding reconstruction error. Since the square of  $\ell_2$ -norm is adopted in each pixel under test (PUT), noisy or anomalous pixels tend to obtain greater reconstruction errors, which will play a major role in updating parameters of AE [42]. This corrupts the reconstruction performance of AE and, thus, fails to pick up some anomalous pixels.

To address this drawback, Zhou and Paffenroth [43] iteratively select the pure samples and train the AE network to alleviate the impact of anomalies in training set. However, the two steps are conducted separately, leading to unstable detection performance. The  $\ell_{2,1}$ -norm proposed in [34] encourages algorithms more robust to heavy noise and outliers and has been widely applied in many fields [42], [44]. Inspired by this, we propose an RAE framework by employing the  $\ell_{2,1}$ -norm as the objective function so that the network can better handle noise and outliers during training

$$\begin{aligned} \|\tilde{\mathbf{X}} - \mathbf{X}\|_{2,1} &= \sum_{i=1}^N \|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2 \\ &= \text{Tr}[(\tilde{\mathbf{X}} - \mathbf{X})^T \mathbf{P}(\tilde{\mathbf{X}} - \mathbf{X})] \end{aligned} \quad (6)$$

where  $\mathbf{P}$  is a diagonal matrix with  $P_{ii} = (1/\|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2 + \epsilon)$ , and  $\epsilon$  is a very small constant such that  $(1/\|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2 + \epsilon)$  approximates  $(1/\|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2)$ . The derivatives of the objective function in (6) with respect to  $\tilde{\mathbf{x}}_i$  are obtained as follows:

$$\frac{\partial \|\tilde{\mathbf{X}} - \mathbf{X}\|_{2,1}}{\partial \tilde{\mathbf{x}}_i} = 2 \frac{\tilde{\mathbf{x}}_i - \mathbf{x}_i}{\|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2 + \epsilon}. \quad (7)$$

It is not hard to find that the only difference between  $\ell_{2,1}$ -norm and the Frobenius norm with regard to the gradient is that  $\ell_{2,1}$ -norm employs a normalization step on the gradient.

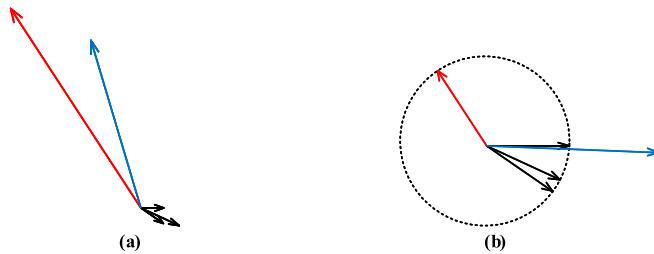


Fig. 1. Comparison of (a) traditional AE with (b) RAE in terms of gradients. Vectors in black, red, and blue denote the background, the outliers, and the corresponding resultants, respectively. The simple normalization step in RAE encourages AE to the majority part representing the background.

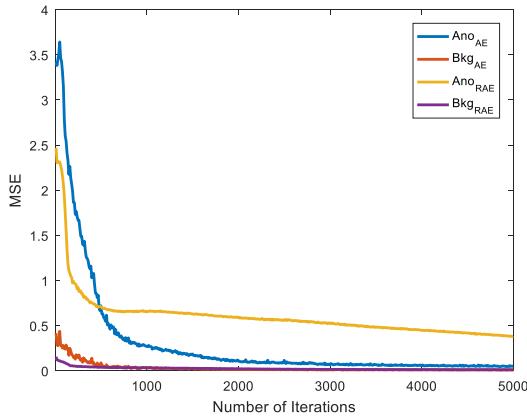


Fig. 2. RAE versus AE in reconstructing background and anomalies.

The simple normalization step can alleviate the impact of anomalous samples on parameter fine-tuning.

Fig. 1 gives a demonstration to better illustrate this. Anomalous samples tend to obtain greater reconstruction errors and, thus, have larger gradient magnitudes, while the background is just the opposite. As a consequence, the traditional AE tends to learn the patterns of anomalies, as shown in Fig. 1(a). This does harm to the discrimination between the background and anomalies, but it is hard to determine when to stop the iteration to get out of the dilemma. In Fig. 1(b), on the contrary, the proposed RAE framework normalizes the gradient for each sample, and the impact of anomalies on parameters fine-tuning is suppressed. The  $\ell_{2,1}$ -norm ensures the network to learn the patterns of the majority part representing the background, thus being more robust to noise and anomalies during training.

We conduct an extra experiment to validate the superiority of RAE. The Gulfport HSI dataset containing 0.6% anomalous samples is exploited here. The mean square error (mse) is used to evaluate the reconstruction performance of background and anomalies. Fig. 2 illustrates how mse changes with iterations. Both AE and RAE are capable to learn the background as the mse of background converges to zero during iterations. However, AE can also learn the characteristics of anomalies after sufficient training, which makes it hard to distinguish anomalies from the background. Nevertheless, RAE normalizes the reconstruction error for each sample before backpropagation; hence, it tends to learn more information about the background and suppress the influence of anomalies on network

fine-tuning. Thus, we can draw the conclusion that RAE with  $\ell_{2,1}$ -norm can better handle anomalies during training.

### B. SuperGraph

In Section II-B, we discuss the importance of graph regularization in obtaining a good representation for detection. It takes full advantage of the local intrinsic geometric structure to encourage similar samples to share similar representations, thus distinguishes anomalies from the background, whereas, there are two major disadvantages that need addressing.

- 1) In hyperspectral images, anomalies often appear as a group of pixels surrounded by the background, which is an additional *a priori* knowledge that can be exploited for anomaly detection. However, neither AE nor graph regularization term takes the spatial information into consideration, which is vital in hyperspectral anomaly detection.
- 2) Formation of the graph Laplacian matrix  $\mathbf{L}$  is time-consuming. For each PUT, it needs to go through the entire dataset to calculate the distances and sort them; only a small part is selected to formulate the similarity matrix, which introduces a lot of redundant operations.

To improve the speed of graph construction, one can first divide the dataset into several clusters with k-means or mean-shift. However, spatial information is not taken into consideration. Inspired by the LXR and CRD that adopt dual window strategy for each PUT, we hold that local geometric structure for each pixel only relies on its adjacent pixels, while those far away from the PUT have no contribution to the construction of the geometric structure. However, the performance of LXR and CRD is sensitive to the sizes of the dual windows. To adaptively select the most relevant samples for each pixel, we adopt a superpixel segmentation strategy to simultaneously preserve the geometric structure and local spatial consistency. To be specific, HSI is segmented into  $S$  nonoverlapping superpixels  $\mathcal{P} = \{P_s\}_{s=1}^S$  by applying simple linear iterative cluster (SLIC) [45] on the first principle component (PC) of HSI. Each superpixel in HSI is a homogeneous region that consists of a set of pixels with similar spatial and spectral characteristics [46]. Pixels within the same patches (superpixels) are active to construct the adjacency matrix  $\hat{\mathbf{W}}$ , while those in different patches are abandoned as they have no contribution to the construction of the geometric structure. The similarity  $\hat{w}_{ij}$  between  $\mathbf{x}_i \in P_u$  and  $\mathbf{x}_j \in P_v$  for  $1 \leq u, v \leq S$  in (3) is redefined as follows:

$$\hat{w}_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma^2}\right) & u = v \\ 0 & u \neq v. \end{cases} \quad (8)$$

Fig. 3 demonstrates the construction of the  $i$ th row of adjacency matrix  $\hat{\mathbf{W}}$ . For PUT  $\mathbf{x}_i$  (the red dot), we first find the pixels belonging to the same patches as  $\mathbf{x}_i$ ; then, the similarities between  $\mathbf{x}_i$  and these pixels sharing the homogeneous region are calculated by (8). The corresponding elements in the  $i$ th row of  $\hat{\mathbf{W}}$  are, therefore, replaced with these similarities, while others stay 0s as they are different from  $\mathbf{x}_i$  in spectral and spatial domains. Note that all of the pixels within the same

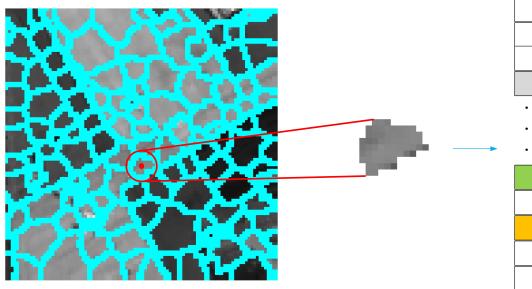


Fig. 3. Construction of one row of graph Laplacian matrix with SuperGraph.

superpixels are exploited to construct the adjacency matrix  $\hat{\mathbf{W}}$ , rather than selecting the first  $k$ -nearest neighbors, as in (3). The degree matrix  $\hat{\mathbf{D}}$  and the modified graph Laplacian matrix, termed SuperGraph matrix  $\hat{\mathbf{L}}$ , are constructed afterward. The local geometric structure is generated by preserving similarities within each superpixel while discarding those outside the region; thus, the local spatial information is integrated into the local geometric structure.

Besides preserving the spatial information, SuperGraph can reduce the execution time of formulating the SuperGraph matrix  $\hat{\mathbf{L}}$  as well because it can effectively reduce the searching space for each pixel, given a matrix  $\mathbf{X} \in \mathbb{R}^{N \times L}$ , where  $N$  and  $L$  are the total numbers of pixels and the number of bands. For each PUT, it takes  $O(NL)$  to obtain the similarities with other pixels and  $O(N \log N)$  to sort these distances and get the  $k$ -nearest neighbors. All of the  $N$  pixels should be calculated to form the graph Laplacian matrix  $\mathbf{L}$ ; hence, the total time cost is  $O(N^2L + N^2 \log N)$ . On the contrary, SuperGraph narrows the scope for searching similar samples. Suppose that the HSI is segmented into  $S$  nonoverlapping superpixels  $\mathcal{P} = \{P_s\}_{s=1}^S$  with  $P_s$  consisting of  $n_s$  pixels, and  $\sum_{s=1}^S n_s = N$ . It takes  $O(NL^2 + L^3)$  to implement this step. The computational complexity for any pixel  $\mathbf{x} \in P_s$  is  $O(n_s L)$  without sort operations. Thus, the total time cost for SuperGraph is  $O(NL^2 + L^3 + \sum_{s=1}^S n_s^2 L)$ , where  $L \ll N$  in HSIs, and  $\sum_{s=1}^S n_s^2 \ll N^2 = (\sum_{s=1}^S n_s)^2$  when  $S$  is large. This demonstrates that the formulation of SuperGraph matrix  $\hat{\mathbf{L}}$  is much faster than that of the matrix  $\mathbf{L}$ .

To prove the above computational complexity analysis, we implement an extra experiment to calculate the time cost of constructing  $\mathbf{L}$  and  $\hat{\mathbf{L}}$  corresponding to five hyperspectral image datasets. The size of these datasets is  $100 \times 100$  except for the last one with  $250 \times 300$ ; the numbers of bands are 205, 191, 205, 189, and 224, respectively. These five HSI datasets will also be used in the following experiments. The range of the number of superpixels  $S1-S5$  is set to  $\{50, 100, 150, 300, 500\}$  for all of the datasets except for the last one with  $\{300, 500, 1000, 1500, 2000\}$ , considering the varying scales. Fig. 4 diagrams the execution time of constructing the graph Laplacian matrix  $\mathbf{L}$  and the SuperGraph matrix  $\hat{\mathbf{L}}$  with different numbers of superpixels. Generally speaking, the execution time of constructing SuperGraph matrix  $\hat{\mathbf{L}}$  is less than 1% of that when constructing  $\mathbf{L}$ , especially in large datasets. We can find from Fig. 4 that it takes approximately only 3 s to construct  $\hat{\mathbf{L}}$  for small datasets and is not sensitive to

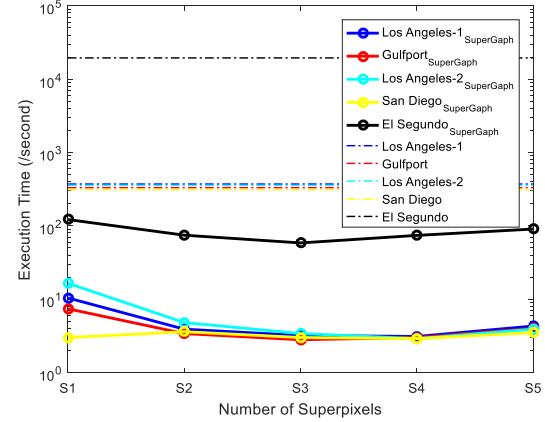


Fig. 4. Execution time of constructing matrices  $\mathbf{L}$  and  $\hat{\mathbf{L}}$  with different numbers of superpixels.

the number of superpixels. In terms of El Segundo, it takes more time to construct  $\hat{\mathbf{L}}$  because of the larger scale, whereas it can achieve up to 252 times faster than constructing  $\mathbf{L}$ , which is inspiring when exploiting geometric structure in HSI applications, such as classification and target detection. Thus, the execution time of forming the local geometric structure is another advantage of the proposed SuperGraph.

The procedure of constructing the SuperGraph matrix  $\hat{\mathbf{L}}$  is described in detail, as shown in Algorithm 1.

---

#### Algorithm 1 Construction of the SuperGraph Matrix $\hat{\mathbf{L}}$

---

##### **Input:**

- (1) the original HSI dataset  $\mathbf{X} \in \mathbb{R}^{m \times n \times L}$ ;
- (2) the number of superpixels  $S$ .

##### **Output:**

The SuperGraph matrix  $\hat{\mathbf{L}}$ .

##### **Procedure:**

- 1: Segment the first PC of the extended  $\mathbf{X}$  into  $S$  clusters with SLIC.
  - 2: Distribute all pixels into  $\mathcal{P} = \{P_s\}_{s=1}^S$  superpixels according to the segmentation result.
  - 3: For each superpixel  $P_s \subseteq \mathcal{P}$ :
  - 4:     For each pixel  $\mathbf{x}_i \in P_s$ :
  - 5:         Find all the pixels  $\mathbf{x}_j \in P_s$ .
  - 6:         Construct the  $i$ -th row of  $\hat{\mathbf{W}}$  by Eq. (8).
  - 7: Calculate the diagonal matrix  $\hat{\mathbf{D}}$  by Eq. (4).
  - 8: The SuperGraph matrix  $\hat{\mathbf{L}}$  is obtained by  $\hat{\mathbf{L}} = \hat{\mathbf{D}} - \hat{\mathbf{W}}$ .
- 

### C. RGAE Detector

In this article, we propose a modified AE by integrating the two objective functions of (6) and (2) together with a positive tradeoff parameter  $\lambda$ . The objective function of the proposed RGAE detector is given as follows:

$$\begin{aligned} J(\Theta) &= J_1(\Theta) + \lambda J_2(\Theta) \\ &= \min_{\Theta} \frac{1}{2N} \sum_{i=1}^N \|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2 + \frac{\lambda}{2N} \sum_{i=1}^N \sum_{j=1}^N \hat{w}_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2 \\ &= \min_{\Theta} \frac{1}{2N} \text{Tr}[(\tilde{\mathbf{X}} - \mathbf{X})^T \mathbf{P}(\tilde{\mathbf{X}} - \mathbf{X})] + \frac{\lambda}{N} \text{Tr}[\mathbf{Z}^T \hat{\mathbf{L}} \mathbf{Z}]. \quad (9) \end{aligned}$$

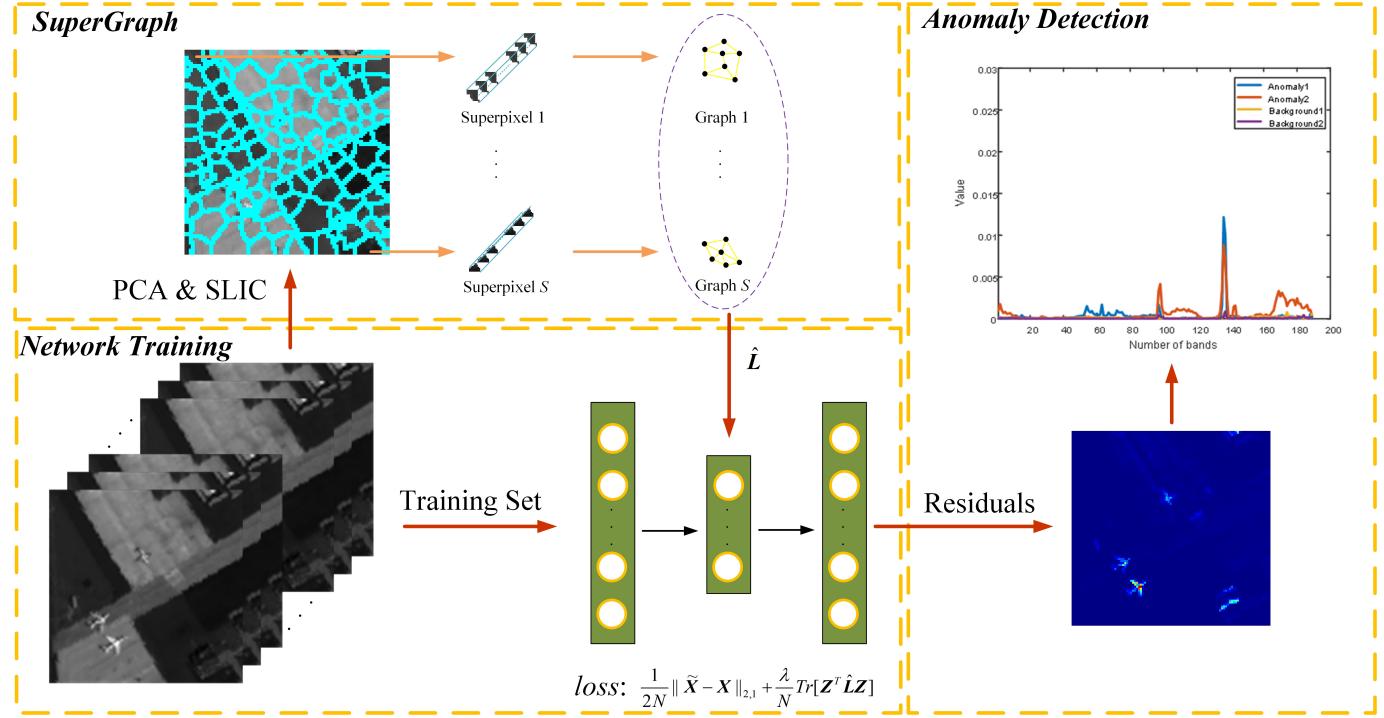


Fig. 5. Flowchart of the proposed RGAE algorithm for hyperspectral anomaly detection.

The first part representing the  $\ell_{2,1}$ -norm of reconstruction error induces RGAE to learn more background information and be more robust to severe noise and anomalies. The second term corresponding to SuperGraph regularization preserves the local geometric structure and spatial consistency simultaneously, which encourages samples within a local homogeneous region to share similar representations.

Fig. 5 describes the whole procedure of the proposed RGAE algorithm for hyperspectral anomaly detection. First, the SuperGraph matrix  $\hat{\mathbf{L}}$  is formulated; it can not only integrate the spatial consistency into the local geometric structure but also remarkably reduce the time cost. Then, an RAE-based hyperspectral anomaly detector is generated by optimizing (9). Note that a single-layer AE is utilized in our method because deeper architecture requires more training samples, which may lead to underfitting with a limited number of pixels. Experiments in Section IV indicate that vanilla AE can still obtain superior detection performance. The upgraded network is used to reconstruct the original HSI dataset, and the reconstruction residuals are used as the criterion for anomaly detection.

The traditional backpropagation algorithm is adopted to solve the optimization problem [see (9)]. Parameters  $\Theta = \{\mathbf{W}^{(m)}, \mathbf{b}^{(m)}\}_{m=1}^M$  in the  $t$ th iteration are updated as follows:

$$\begin{cases} \mathbf{W}_t^{(m)} = \mathbf{W}_{t-1}^{(m)} - \mu \frac{\partial J_t(\Theta)}{\partial \mathbf{W}_t^{(m)}} \\ \mathbf{b}_t^{(m)} = \mathbf{b}_{t-1}^{(m)} - \mu \frac{\partial J_t(\Theta)}{\partial \mathbf{b}_t^{(m)}} \end{cases}, m = 1, \dots, M. \quad (10)$$

The subgradients of  $J(\Theta)$  with respect to  $\mathbf{W}^{(m)}$  and  $\mathbf{b}^{(m)}$  ( $m = 1, \dots, M$ ) can be obtained by

$$\begin{cases} \frac{\partial J(\Theta)}{\partial \mathbf{W}^{(m)}} = \frac{\partial J_1(\Theta)}{\partial \mathbf{W}^{(m)}} + \lambda \frac{\partial J_2(\Theta)}{\partial \mathbf{W}^{(m)}} \\ = \frac{1}{N} \sum_{i=1}^N \left( \nabla_{1i}^{(m)} + \lambda \nabla_{2i}^{(m)} \right)^T \mathbf{h}_i^{(m-1)} \\ \frac{\partial J(\Theta)}{\partial \mathbf{b}^{(m)}} = \frac{\partial J_1(\Theta)}{\partial \mathbf{b}^{(m)}} + \lambda \frac{\partial J_2(\Theta)}{\partial \mathbf{b}^{(m)}} \\ = \frac{1}{N} \sum_{i=1}^N \left( \nabla_{1i}^{(m)} + \lambda \nabla_{2i}^{(m)} \right)^T \end{cases} \quad (11)$$

where  $\nabla_{1i}^{(m)}$  and  $\nabla_{2i}^{(m)}$  are given by

$$\nabla_{1i}^{(m)} = \begin{cases} \frac{\mathbf{h}_i^{(m)} - \mathbf{x}_i}{\|\mathbf{h}_i^{(m)} - \mathbf{x}_i\|_2 + \epsilon} \circ g'(\mathbf{h}_i^{(m)}) & m = M \\ (\nabla_{1i}^{(m+1)} \mathbf{W}^{(m+1)}) \circ g'(\mathbf{h}_i^{(m)}) & \text{otherwise} \end{cases} \quad (12)$$

$$\nabla_{2i}^{(m)} = \begin{cases} 0 & m > \frac{M}{2} \\ \mathbf{H}^{(m)} (\hat{\mathbf{L}} + \hat{\mathbf{L}}^T)_{\cdot i} \circ g'(\mathbf{h}_i^{(m)}) & m = \frac{M}{2} \\ (\nabla_{2i}^{(m+1)} \mathbf{W}^{(m+1)}) \circ g'(\mathbf{h}_i^{(m)}) & m = 1, \dots, \frac{M}{2} - 1 \end{cases} \quad (13)$$

where  $\circ$  is the Hadamard product.  $\mathbf{H}^{(m)}$  is the output of the  $m$ th layer of the entire dataset, and  $\mathbf{H}^{(0)} = \mathbf{X}$  is the input of AE. The derivative of the activation  $g(\mathbf{h}_i^{(m)})$  is  $g'(\mathbf{h}_i^{(m)})$ , and  $g'(\mathbf{h}_i^{(m)}) = g(\mathbf{h}_i^{(m)}) \circ (1 - g(\mathbf{h}_i^{(m)}))$  for sigmoid function.  $(\hat{\mathbf{L}} + \hat{\mathbf{L}}^T)_{\cdot i}$  is the  $i$ th column of the symmetric matrix  $(\hat{\mathbf{L}} + \hat{\mathbf{L}}^T)$ .

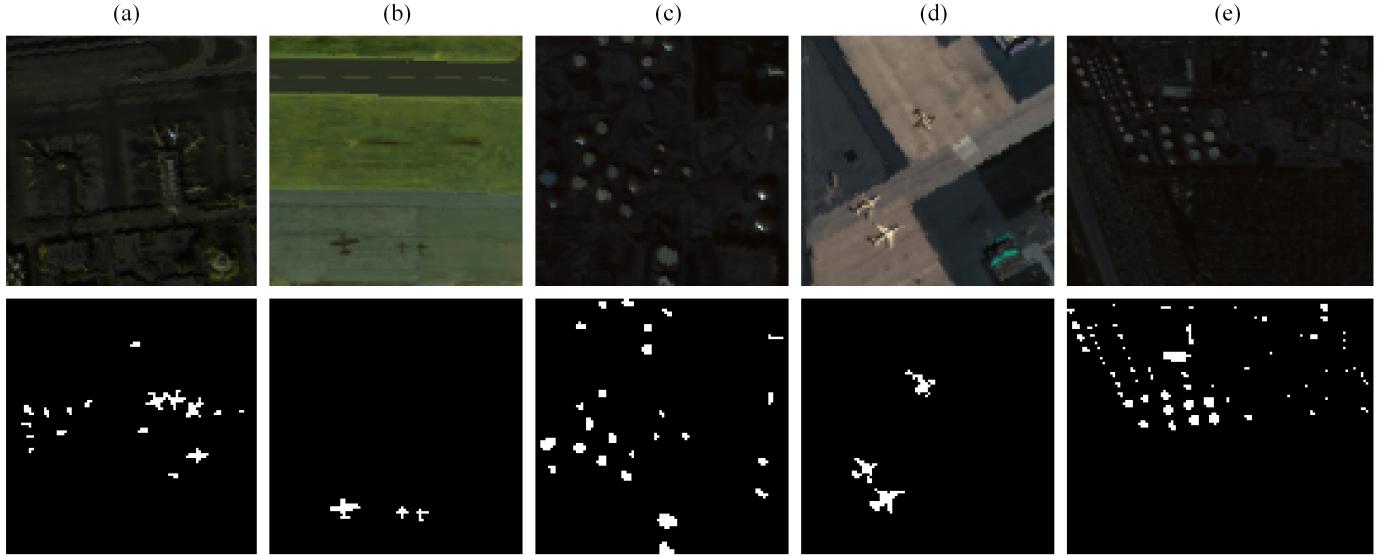


Fig. 6. Pseudocolor images and the corresponding ground truth of several pixels of the five HSI datasets. (a) Los Angeles-1. (b) Gulfport. (c) Los Angeles-2. (d) San Diego. (e) El Segundo.

When the mean of reconstruction residuals reduces to a predefined parameter  $\epsilon$  or the maximum number of iterations reached, the proposed AE framework is upgraded as a hyperspectral anomaly detector. For simplicity, we set a maximum number of iterations  $iter_{max}$  as the stop criterion in advance. The level of pixel  $\mathbf{x}_i$  being anomalous can be measured with

$$d_i = \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2. \quad (14)$$

The higher the value of  $d_i$  is, the more likely that the pixel  $\mathbf{x}_i$  is an anomaly.

Algorithm 2 summarizes the whole procedure of the proposed RGAE detector.

#### IV. EXPERIMENTS

In this section, we implement several experiments to verify the rationality and superiority of the proposed RGAE detector in detecting anomalies in hyperspectral images. All experiments are conducted in MATLAB on an Intel Core i5-6200U CPU with 4 GB of RAM.

##### A. Datasets

In this article, the five HSI datasets mentioned in Section III-B are employed to verify the effectiveness of the proposed algorithm. The five datasets are introduced here, and the pseudocolor images and the corresponding ground truth are illustrated in Fig. 6.

- 1) *Los Angeles-1*: This hyperspectral image was captured from the AVIRIS sensor with a 7.1-m spatial resolution. There are  $100 \times 100$  pixels with 205 bands after bad bands removed in wavelengths ranging from 400 to 2500 nm. There are 170 anomalous pixels representing aircraft.
- 2) *Gulfport*: This dataset was captured from the airborne visible/infrared imaging spectrometer (AVIRIS) sensor

---

##### Algorithm 2 Proposed RGAE Detector

---

**Input:**

- (1) the original HSI dataset  $\mathbf{X} \in \mathbb{R}^{m \times n \times L}$ ;
- (2) the positive tradeoff parameter  $\lambda$ ;
- (3) the number of superpixels  $S$ ;
- (4) the dimensionality of hidden layer  $d$ .

**Output:**

Hyperspectral anomaly detection map.

**Initialize:**

$$\Theta = \{\mathbf{W}^{(m)}, \mathbf{b}^{(m)}\}_{m=1}^M, \mathbf{H}^{(0)} = \mathbf{X}, iter_{max}, t = 1.$$

**Procedure:**

- 1: Obtain the SuperGraph matrix  $\hat{\mathbf{L}}$  by **Algorithm 1**.
  - 2: Repeat:
    - 3: Get  $\{\mathbf{H}_t^{(m)}\}_{m=1}^M$  by Eq. (1).
    - 4: Compute gradients  $\{\frac{\partial J_t(\Theta)}{\partial \mathbf{W}^{(m)}}\}_{m=1}^M$  and  $\{\frac{\partial J_t(\Theta)}{\partial \mathbf{b}^{(m)}}\}_{m=1}^M$  by Eq. (11)-(13).
    - 5: Update the parameters  $\Theta_t = \{\mathbf{W}_t^{(m)}, \mathbf{b}_t^{(m)}\}_{m=1}^M$  by Eq. (10).
    - 6:  $t := t + 1$ .
    - 7: Until:  $t \geq Iter_{max}$
    - 8: Get the anomalous probability for each PUT  $\mathbf{x}_i$  in  $\mathbf{X}$  by Eq. (14).
- 

with a 3.4-m spatial resolution. The size of the hyperspectral image is  $100 \times 100$ ; it has 191 spectral bands with the spectral wavelengths from 400 to 2500 nm after bad bands removed because of the water absorption regions and low signal-to-noise ratio (SNR). The types of anomalies in this HSI dataset are three airplanes at the bottom of the image, which occupies 60 pixels in total.

- 3) *Los Angeles-2*: This dataset was also derived from the AVIRIS sensor with a 7.1-m spatial resolution. This image scene covers an area of  $100 \times 100$  pixels,

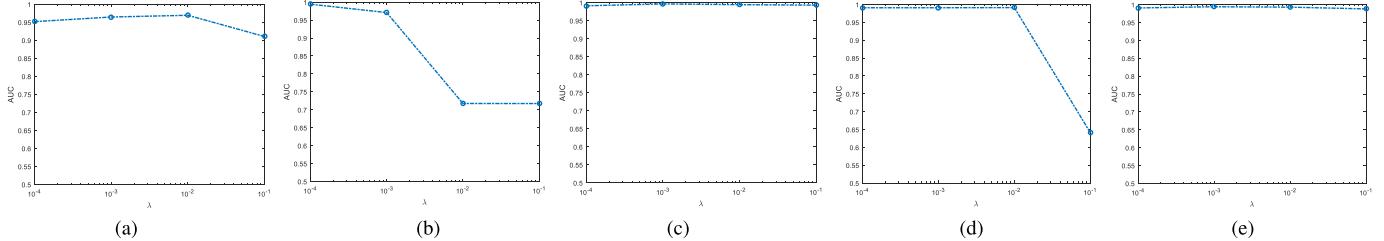


Fig. 7. Impact of the tradeoff parameter  $\lambda$  on the final detection result. (a) Los Angeles-1. (b) Gulfport. (c) Los Angeles-2. (d) San Diego. (e) El Segundo.

which has 205 spectral bands with the spectral wavelengths from 400 to 2500 nm after bad bands are removed; 272 pixels representing oil tanks are regarded as anomalies.

- 4) *San Diego*: The  $100 \times 100$  image was acquired by the AVIRIS sensor over the San Diego airport area. This dataset has a 20-m spatial resolution and a 10-nm spectral resolution with spectral wavelengths from 370 to 2510 nm. After removal of bad bands (1–6, 33–35, 97, 107–113, 153–166, and 221–224) corresponding to the water absorption regions, low SNR and 189 bands have remained for the following experiments. The three airplanes in the image that consists of 134 pixels are treated as anomalies.
- 5) *El Segundo*: This image was derived from the AVIRIS sensor over an area of El Segundo, which consists of 224 spectral channels with wavelengths ranging from 336 to 2496 nm. The  $250 \times 300$  image has a 7.1-m spatial resolution. The oil refinery consisting of storage tanks and towers is regarded as anomalies, which accounts for 2.73% of the entire image.

The receiver operating characteristic (ROC) is applied as the evaluation metric to quantify the performances of anomaly detectors. ROC represents the relationship between the true-positive rate (TPR) and false-positive rate (FPR) by setting different thresholds. The ROC curves of good anomaly detectors are usually near to the upper left than others. However, ROC curves describe the performance of anomaly detectors qualitatively; to further describe the detection performance of anomaly detectors from the quantitative perspective, the AUC value is added as another evaluation metric. AUC calculates the area under the ROC curve; a higher value of AUC demonstrates better detection performance.

### B. Parameter Settings

The proposed RGAE detector involves three parameters that need adjusting: the tradeoff parameter  $\lambda$ , the number of superpixels  $S$ , and the dimension of hidden layer  $n\_hid$ . We analyze the impact of these three parameters on the final detection performance here. All of the five datasets mentioned above are exploited, and the AUC values are utilized as the evaluation metric in this section.

1) *Tradeoff Parameter  $\lambda$* : The tradeoff parameter  $\lambda$  balances the contributions of the  $\ell_{2,1}$ -norm [i.e.,  $J_1(\Theta)$ ] and the SuperGraph regularization term [i.e.,  $J_2(\Theta)$ ] to the final detection performance. During training, the number of superpixels  $S$  is

set to be 200 for the first four datasets and 1000 for the last one when constructing the modified graph Laplacian matrix  $\hat{\mathbf{L}}$ , considering the different sizes of these datasets. The dimension of the hidden layer is fixed to be 100 for all of the datasets. The range of the tradeoff parameter  $\lambda$  is set to  $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ .

The AUC values with respect to  $\lambda$  are shown in Fig. 7. The tradeoff parameter  $\lambda$  has a great impact on the final detection performance, especially in Los Angeles-1 and San Diego datasets. Also, we can notice that the detection performance improved with the increase of  $\lambda$  until a certain value reached in the other three datasets. This reveals that both  $\ell_{2,1}$ -norm and the SuperGraph regularization term have significant influences on separating anomalies from the background. The optimal values of  $\lambda$  vary for these five datasets and are selected to be  $10^{-2}, 10^{-4}, 10^{-3}, 10^{-2}$ , and  $10^{-3}$ , respectively.

2) *Number of Superpixels  $S$* : The number of superpixels  $S$  determines the quality of the local geometric structure when constructing the SuperGraph matrix  $\hat{\mathbf{L}}$ . The tradeoff parameter  $\lambda$ , as analyzed above, is a very important parameter, which varies in different datasets and needs to be set in advance ( $10^{-2}, 10^{-4}, 10^{-3}, 10^{-2}$ , and  $10^{-3}$ , respectively). The dimension of hidden layer  $n\_hid$  is set to be 100 for all of the five datasets. The range of the number of superpixels  $S1-S5$  is set to  $\{50, 100, 150, 300, 500\}$  for the first four datasets and  $\{300, 500, 1000, 1500, 2000\}$  for the last one, in light of the varying scales in these datasets.

As demonstrated in Fig. 8, the proposed RGAE detector approaches the optimal detection performance gradually with the increase of the number of superpixels  $S$  until a certain value. The detector yields the worst detection performance in the Los Angeles-1 dataset, and we think that the complicated background degrades the distinguishing ability of superpixel segmentation strategies. Taking both the detection performance and the construction time of SuperGraph matrix  $\hat{\mathbf{L}}$  into consideration, the parameter  $S$  is set to be 150 for the first four datasets and 1000 for El Segundo.

3) *Dimension of Hidden Layer  $n\_hid$* : The dimension of hidden layer  $n\_hid$  influences the information of input layer preserved in latent space. In light of the analyses in Sections IV-B1 and IV-B2, the number of superpixels  $S$  is set to 150 for all of the datasets except for the last one with 1000, while the values of the tradeoff parameter  $\lambda$  are set to be  $10^{-2}, 10^{-4}, 10^{-3}, 10^{-2}$ , and  $10^{-3}$ , respectively. The range of the dimension of the hidden layer  $n\_hid$  is set to  $\{20, 40, 60, 80, 100, 120, 140, 160\}$ .

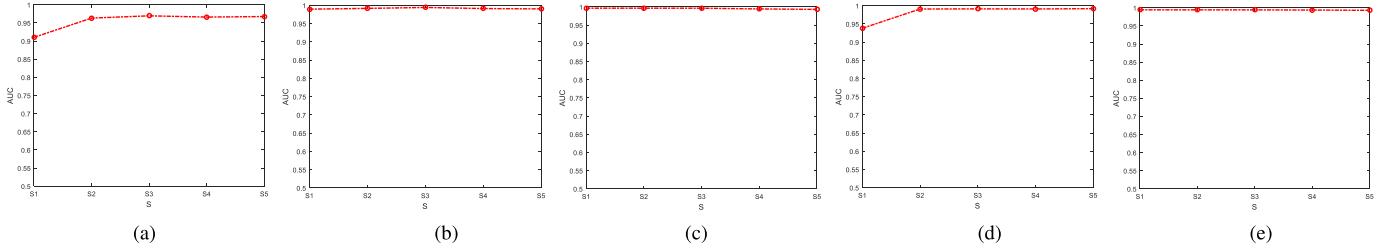


Fig. 8. Impact of the number of superpixel  $S$  on the final detection result. (a) Los Angeles-1. (b) Gulfport. (c) Los Angeles-2. (d) San Diego. (e) El Segundo.

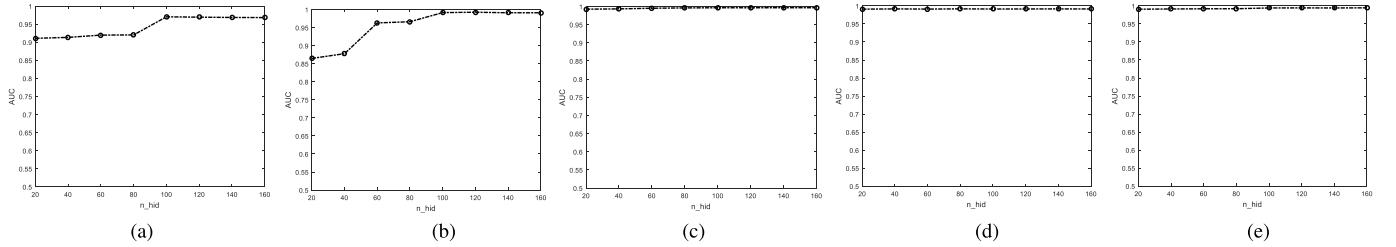


Fig. 9. Impact of the dimension of hidden layer  $n\_hid$  on the final detection result. (a) Los Angeles-1. (b) Gulfport. (c) Los Angeles-2. (d) San Diego. (e) El Segundo.

Fig. 9 shows the impact of  $n\_hid$  on the final detection performance. The AE network with small values of  $n\_hid$  loses some intrinsic information to help to separate anomalies from the background and, hence, yields unsatisfying detection results. With the increase of the value of  $n\_hid$ , AE can, thus, preserve the distinctive characteristics for anomaly detection. We can also find that the performance is less sensitive to  $n\_hid$  when  $n\_hid$  is greater than 100. For simplicity, the values of  $n\_hid$  are chosen to be 100 for all of the five datasets.

### C. Ablation Study

In this section, we evaluate the contributions of the modified objective functions of AE, i.e., the  $\ell_{2,1}$ -norm and the SuperGraph regularization term on the final detection performance. The details of the detectors based on the AE framework exploited in this experiment are listed as follows.

- 1) *AE*: The traditional AE-based anomaly detector with the square of the Frobenius norm of the reconstruction error as the objective function, which is rewritten as follows:

$$J_{\text{AE}} = \min_{\Theta} \frac{1}{2N} \|\tilde{\mathbf{X}} - \mathbf{X}\|_F^2. \quad (15)$$

- 2) *RAE*: Robust AE that exploits  $\ell_{2,1}$ -norm as the objective function, which is

$$J_{\text{RAE}} = \min_{\Theta} \frac{1}{2N} \|\tilde{\mathbf{X}} - \mathbf{X}\|_{2,1}. \quad (16)$$

- 3) *Robust Global Graph AE (RGGAE)*: It formulates the graph Laplacian matrix  $\mathbf{L}$  in a global point of view, as illustrated in Section II-B

$$J_{\text{RGGAE}} = \min_{\Theta} \frac{1}{2N} \|\tilde{\mathbf{X}} - \mathbf{X}\|_{2,1} + \frac{\lambda}{N} \text{Tr}[\mathbf{Z}^T \mathbf{L} \mathbf{Z}]. \quad (17)$$

- 4) *Proposed RGAE*: It formulates each row of the SuperGraph matrix  $\hat{\mathbf{L}}$  with a local sliding window

$$J_{\text{RGAE}} = \min_{\Theta} \frac{1}{2N} \|\tilde{\mathbf{X}} - \mathbf{X}\|_{2,1} + \frac{\lambda}{N} \text{Tr}[\mathbf{Z}^T \hat{\mathbf{L}} \mathbf{Z}]. \quad (18)$$

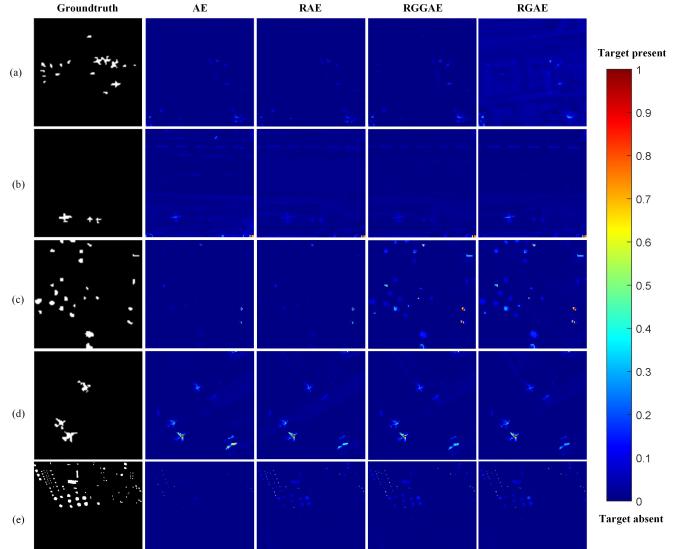


Fig. 10. Ablation study on five datasets. (a) Los Angeles-1. (b) Gulfport. (c) Los Angeles-2. (d) San Diego. (e) El Segundo.

Parameters of each anomaly detector are set to get the optimal detection performance, the detection maps and the corresponding AUC values for those detectors on five datasets are shown in Fig. 10 and Table I, and the best detection result for each dataset is highlighted in bold. Generally, RGAE is superior to other detectors in these five datasets.

AE with the Frobenius norm treats all the samples equally and is sensitive to heavy noise and outliers; it inclines to learn characteristics of these samples because of the large reconstruction residuals. As a consequence, AE fails to detect some small targets, as shown in Fig. 10, especially in Los Angeles-1 and El Segundo datasets. Worse still, AE yields

TABLE I  
COMPARISON OF RGAE AND ITS DERIVATES ON FIVE DATASETS

Datasets	Algorithms			
	AE	RAE	RGGAE	RGAE
Los Angeles-1	0.9381 (222.4 s)	0.9465 (244.1 s)	0.9658 (615.6 s)	<b>0.9712</b> (280.8 s)
	0.9665 (220.6 s)	0.9900 (230.5 s)	<b>0.9915</b> (562.8 s)	0.9912 (256.6 s)
Gulfport	0.9937 (229.0 s)	0.9948 (232.8 s)	0.9964 (589.6 s)	<b>0.9965</b> (300.0 s)
	0.9888 (216.4 s)	0.9900 (235.4 s)	0.9917 (546.6 s)	<b>0.9918</b> (304.5 s)
San Diego	0.9900 (1765 s)	0.9916 (1834 s)	0.9921 (21336 s)	<b>0.9942</b> (2298 s)

the lowest AUC values among these four AE-based anomaly detection frameworks, as revealed in Table I.

RAE is a robust version of AE with  $\ell_{2,1}$ -norm as the objective function, which avoids the impact of severe noise and outliers in the training set on parameters fine-tuning. As shown in Fig. 10, small targets are highlighted especially in the El Segundo dataset, and the detection performance is improved when comparing the AUC values with the traditional AE detector. However, features that RAE learns are not discriminative enough because samples are represented individually and the local geometric structure is discarded.

RGGAE introduces the graph regularization term into the RAE framework under the assumption that combinations of features are usually more discriminative than individual ones. Fig. 10 demonstrated that more anomalies are detected out in Los Angeles-2 compared with RAE. The graph regularization term encourages adjacent pixels in the spectral domain to share similar representations, and it is beneficial to better distinguish anomalies from the background, whereas RGGAE does not take spatial information into consideration, which is vital in hyperspectral anomaly detection. What is worse, the construction of graph Laplacian matrix  $\mathbf{L}$  is time-consuming.

RGAE takes advantage of the prior knowledge that anomalies usually appear as several groups of pixels surrounded by the background. It modifies the formulation of the graph Laplacian matrix  $\hat{\mathbf{L}}$  by introducing a SuperGraph regularization term. RGAE achieves superior detection performance in all of the five datasets except Gulfport, as shown in Table I. Though the detection performances of RGGAE and RGAE are similar in general, the time cost of the latter one is relatively smaller than that of RGGAE, as diagrammed in Table I.

In conclusion, the proposed RGAE detector outperforms other AE-based detectors, which demonstrates the rationality of our proposed algorithm.

#### D. Detection Performance

In this section, we compare our proposed RGAE detector with other state-of-the-art methods in hyperspectral anomaly detection, and then, we will analyze the results in detail. Before applying the experiment, we briefly introduce these methods as follows.

- 1) *GRX* [15]: The global RX detector is a benchmark in hyperspectral anomaly detection. Under the assumption that the background follows the multivariate normal distribution, GRX models the background with the inverse of the covariance matrix, and anomalies are selected out via the Mahalanobis distance.
- 2) *LRX* [16]: Local RX is one of the derivates of GRX that adopts a dual-window strategy to reconstruct the covariance matrix. Samples inside the outer window while outside the inner window centered at PUT are used to construct the inverse of the covariance matrix.
- 3) *LKRX* [17]: Local kernel RX is also one of the derivates of GRX. LKRX adopts kernel trick to project the data into a higher dimensional space to better represent the complex distribution of the HSI.
- 4) *CRD* [47]: The collaborative-representation-based detector is based on the assumption that background can be linearly represented by its surrounding pixels while anomalies cannot, and anomalies can be determined in the residual image, which is obtained by subtracting the predicted background from the original hyperspectral data.
- 5) *LSMAD* [48]: The LRaSMD-based Mahalanobis distance method sets the background apart from the anomalies via the low-rank and sparse matrix decomposition (LRaSMD) technique and explores the low-rank prior knowledge of the background. Anomalies are detected with the Mahalanobis distance.
- 6) *FrFE* [14]: The fractional Fourier entropy exploits the fractional Fourier transform (FrFT) to enhance the signal and suppress noise simultaneously. The optimal fractional transform order can be automatically determined with an FrFE-based step.
- 7) *KIFD* [49]: Kernel isolation forest is a kernel version of isolation forest (iForest) to better separate anomalies from the background. To take advantage of the local information, the initial anomaly detection map is refined with local iForest.
- 8) *MemAE* [33]: MemAE adds an extra memory module to extract the most relevant characteristics. Samples tend to learn the background because the memory module records the patterns of the background. Thus, the residuals of anomalies are larger than those of background samples.
- 9) *AE*: The traditional AE framework with the square of Frobenius norm of reconstruction error as the objective function. Anomalies are picked up based on the reconstruction error between the original image and the reconstructed image.

Parameters are set to get the optimal detection result for each hyperspectral anomaly detector. The detection results, i.e., the detection maps, ROC curves, and AUC values, are demonstrated in Figs. 11 and 12 and Table II, respectively. Basically, the proposed RGAE hyperspectral anomaly detector can better suppress the background and detects targets with a lower false alarm rate.

GRX is a simple and comprehensible hyperspectral anomaly detector. However, the Gaussian distribution assumption may

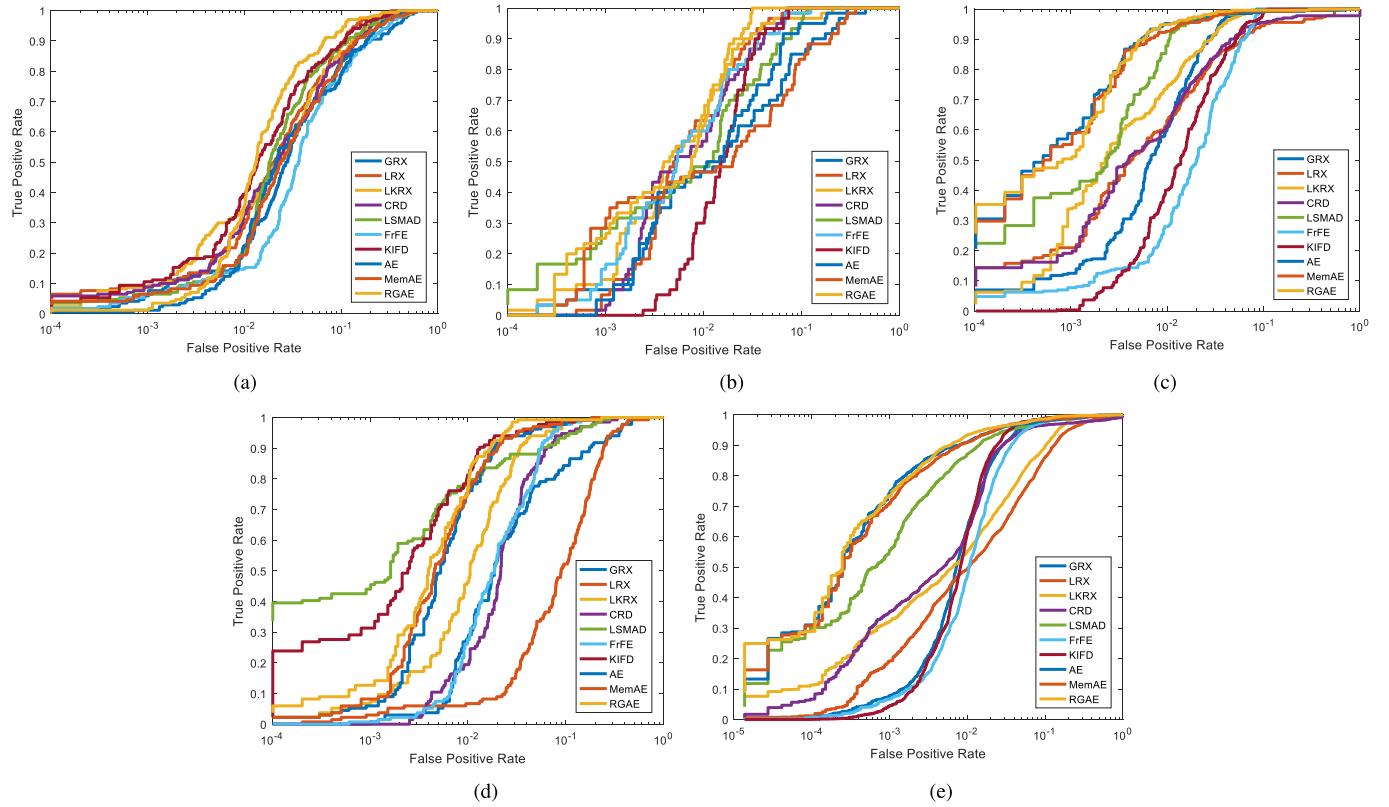


Fig. 11. ROC curves of  $(P_d, P_f)$  on five datasets. (a) Los Angeles-1. (b) Gulfport. (c) Los Angeles-2. (d) San Diego. (e) El Segundo.

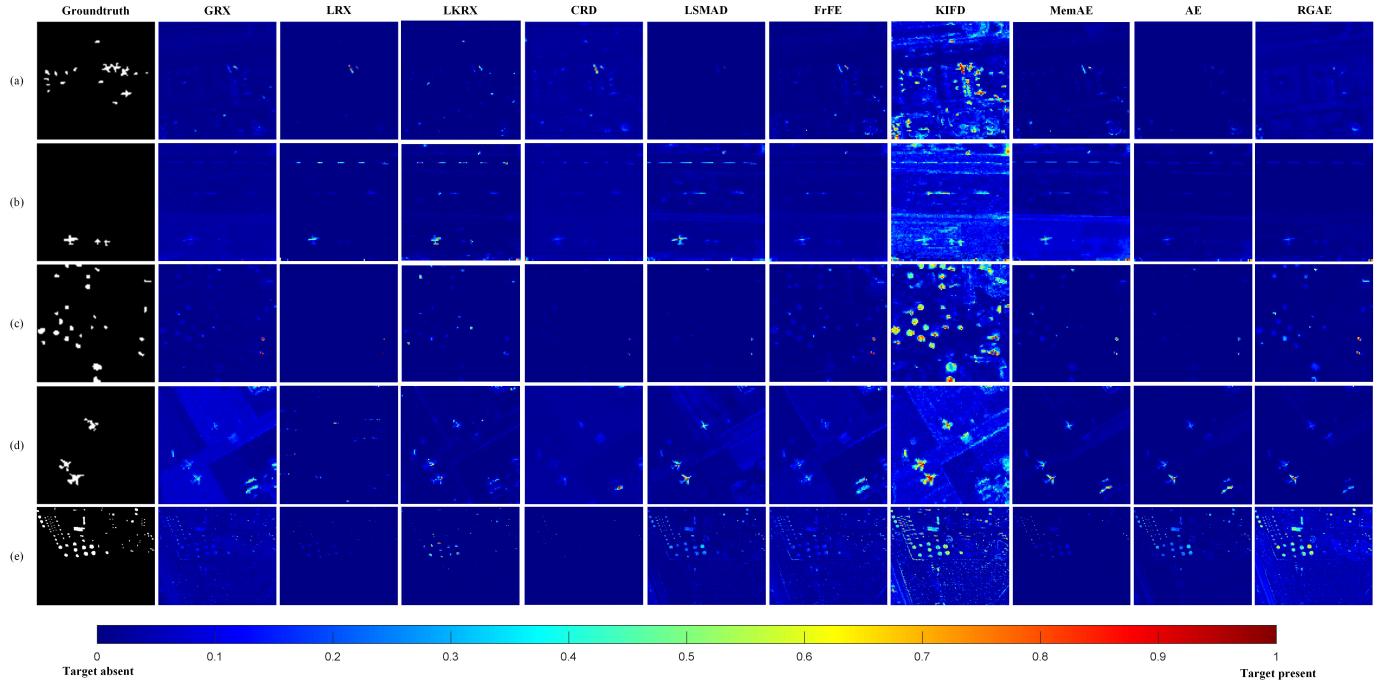


Fig. 12. Hyperspectral anomaly detection maps on five datasets. (a) Los Angeles-1. (b) Gulfport. (c) Los Angeles-2. (d) San Diego. (e) El Segundo.

be violated in some complicated scenes, which leads to degradation of detection performance. Anomalies are usually a group of pixels that differs dramatically from the surrounding background, which means that the distant pixels from PUT

make little contribution to the detection result. Under this assumption, LRX, LKRX, and CRD take spatial information into consideration by applying a sliding dual-window strategy. This improves the detection performance, but it also introduces

TABLE II  
AUC VALUES AND THE EXECUTION TIME ON FIVE DATASETS\*

Algorithms	AUC					Time Cost				
	HSI-1	HSI-2	HSI-3	HSI-4	HSI-5	HSI-1	HSI-2	HSI-3	HSI-4	HSI-5
<b>GRX</b>	0.9288	0.9526	0.9887	0.9403	0.9841	0.159	0.132	0.155	0.126	1.606
<b>LRX</b>	0.9593	0.9872	0.9680	0.8797	0.9483	81.1	63.1	63.7	59.6	796.4
<b>LKRX</b>	0.9486	0.9850	0.9900	0.9820	0.9679	7312.6	825.3	1369.7	6555.3	2025.6
<b>CRD</b>	0.9501	0.9858	0.9633	0.9664	0.9715	288.1	402.7	117.4	266.9	10358.3
<b>LSMAD</b>	0.9529	0.9780	0.9927	0.9821	0.9899	35.6	45.0	33.2	38.7	285.9
<b>FrFE</b>	0.9279	0.9853	0.9733	0.9724	0.9808	15.7	13.4	14.5	13.2	113.1
<b>KIFD</b>	0.9624	0.9803	0.9789	0.9917	0.9863	141.7	148.2	124.8	148.9	381.9
<b>MemAE</b>	0.9409	0.9450	0.9946	0.9898	0.9914	231.2	301.8	289.3	247.2	2236.1
<b>AE</b>	0.9381	0.9665	0.9937	0.9888	0.9900	222.4	220.6	229.0	216.4	1765.0
<b>RGAE</b>	<b>0.9712</b>	<b>0.9912</b>	<b>0.9965</b>	<b>0.9918</b>	<b>0.9942</b>	280.8	256.6	300.0	304.5	2298.0

\* HSI-1 to HSI-5 denote Los Angeles-1, Gulfport, Los Angeles-2, San Diego and El Segundo data set, respectively.

more additional parameters to adjust and more time cost to traverse the whole HSI.

LSMAD is known for its ability in separating the background from noise and anomalies, while FrFE has the advantage of simultaneously suppressing noise and enhancing the signal. Fig. 12 reveals that both LSMAD and FrFE can better detect anomalies and suppress the background at the same time. However, the same as LRx and CRD, the original raw data without extraction of latent and representative characteristics are exploited, which limits the detection performance.

The iForest is an ensemble method specially designed for outlier detection. KIFD exploits kernel tricks to better separate anomalies from the background, and a refinement operation can make full use of local information. Anomalies with KIFD are highlighted with strong amplitude, as shown in Fig. 12. However, the background is not suppressed well, especially in the El Segundo dataset, because the spatial information is not taken into consideration.

In order to extract more intrinsic features in an unsupervised manner, AE and MemAE act as an information bottleneck enforcing the network to extract the typical patterns of high-dimensional raw data. Nevertheless, they are sensitive to noise and outliers during training as it treats all the samples equally. The training set used in hyperspectral anomaly detection is contaminated with anomalies because there is no prior spectral information about background or anomalies. Samples with higher reconstruction errors tend to play a major role in parameter tuning, leading to the corruption of detection results.

RGAE can detect all the anomalies with a lower false alarm rate, as shown in Fig. 11. In addition, the AUC values also demonstrate that RGAE outperforms other state-of-the-art detectors. RGAE takes both robustness and geometric structure into consideration. On the one hand, it employs the  $\ell_{2,1}$ -norm instead of the traditional Frobenius norm to encourage pixels with smaller reconstruction errors to contribute more to the parameter tuning of the network and suppress the others corresponding to heavy noise and outliers. On the other hand, it creatively embeds the spatial information

into the geometric structure via SuperGraph, a superpixel segmentation-based graph construction strategy. Moreover, SuperGraph can also remarkably reduce the execution time of forming the graph Laplacian matrix  $\hat{\mathbf{L}}$ , which is another additional advantage of SuperGraph. In a nutshell, this subsection proves that the proposed RGAE detector is superior to other algorithms in hyperspectral anomaly detection.

## V. CONCLUSION

In this article, a modified AE framework-based algorithm named RGAE is proposed for hyperspectral anomaly detection. First, to alleviate the sensitivity of traditional AE to strong noise and outliers, we propose an RAE framework based on  $\ell_{2,1}$ -norm. The RAE framework is able to select anomalies out even with a contaminated training set. Then, the pairwise graph regularization is added into the objective function to ensure that the local geometric structure is consistent with that in the original domain. We adopt a superpixel segmentation-based graph construction strategy to embed the spatial consistency into the local geometric structure, and the time cost is reduced remarkably.

Experiments with five hyperspectral datasets verify the rationality and superiority of our RGAE algorithm. By comparing our algorithm with other versions of AE, the AUC values demonstrate that RGAE is more robust to anomalies during training, and it is feasible to integrate the local geometric structure with spatial consistency. The detection performance reveals the superiority of RGAE in detecting anomalies with a lower false alarm rate, after comparing our proposed algorithm with traditional and state-of-the-art detectors in five datasets. How to further reduce the execution time of constructing the SuperGraph matrix especially in large-scale datasets will be a focus in our future work.

## REFERENCES

- [1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

- [2] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.
- [3] J. Peng *et al.*, "Low-rank and sparse representation for hyperspectral image processing: A review," *IEEE Geosci. Remote Sens. Mag.*, early access, Jun. 10, 2021, doi: [10.1109/MGRS.2021.3075491](https://doi.org/10.1109/MGRS.2021.3075491).
- [4] X. Mei, Y. Ma, C. Li, F. Fan, J. Huang, and J. Ma, "Robust GBM hyperspectral image unmixing with superpixel segmentation based low rank and sparse representation," *Neurocomputing*, vol. 275, pp. 2783–2797, Jan. 2018.
- [5] X. Tian, Y. Chen, C. Yang, and J. Ma, "Variational pansharpening by exploiting cartoon-texture similarities," *IEEE Trans. Geosci. Remote Sens.*, early access, Jan. 15, 2021, doi: [10.1109/TGRS.2020.3048257](https://doi.org/10.1109/TGRS.2020.3048257).
- [6] X. Zhao *et al.*, "Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7355–7370, Oct. 2020.
- [7] M. Xu *et al.*, "Regionally and locally adaptive models for retrieving chlorophyll-A concentration in inland waters from remotely sensed multispectral and hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4758–4774, Jul. 2019.
- [8] T. A. Carrino, A. P. Crósta, C. L. B. Toledo, and A. M. Silva, "Hyperspectral remote sensing applied to mineral exploration in southern Peru: A multiple data integration approach in the Chapi Chiara gold prospect," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 64, pp. 287–300, Feb. 2018.
- [9] T. Adao *et al.*, "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sens.*, vol. 9, no. 11, p. 1100, 2017.
- [10] W. Sun, L. Tian, Y. Xu, B. Du, and Q. Du, "A randomized subspace learning based anomaly detector for hyperspectral imagery," *Remote Sens.*, vol. 10, no. 3, p. 417, 2018.
- [11] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.
- [12] B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, Nov. 2014.
- [13] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.
- [14] R. Tao, X. Zhao, W. Li, H. 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.
- [15] I. S. 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.
- [16] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.
- [17] 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.
- [18] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.
- [19] W. Wang, S. Li, H. Qi, B. Ayhan, C. Kwan, and S. Vance, "Identify anomaly componentbysparsity and low rank," in *Proc. IEEE 7th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Jun. 2015, pp. 1–4.
- [20] 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.
- [21] Y. Qu *et al.*, "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, Aug. 2018.
- [22] Y. Qu *et al.*, "Anomaly detection in hyperspectral images through spectral unmixing and low rank decomposition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 1855–1858.
- [23] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2019.
- [24] H. Xu, J. Ma, and X.-P. Zhang, "MEF-GAN: Multi-exposure image fusion via generative adversarial networks," *IEEE Trans. Image Process.*, vol. 29, pp. 7203–7216, Jun. 2020.
- [25] W. Sun, K. Ren, X. Meng, C. Xiao, G. Yang, and J. Peng, "A band divide-and-conquer multispectral and hyperspectral image fusion method," *IEEE Trans. Geosci. Remote Sens.*, early access, Feb. 25, 2021, doi: [10.1109/TGRS.2020.3046321](https://doi.org/10.1109/TGRS.2020.3046321).
- [26] X. Wang, Q. Tan, and L. Zhang, "A deep neural network of multi-form alliances for personalized recommendations," *Inf. Sci.*, vol. 531, pp. 68–86, Aug. 2020.
- [27] S. Chang, B. Du, and L. Zhang, "A sparse autoencoder based hyperspectral anomaly detection algorihtm using residual of reconstruction error," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 5488–5491.
- [28] C. Zhao and L. Zhang, "Spectral-spatial stacked autoencoders based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection," *Infr. Phys. Technol.*, vol. 92, pp. 166–176, Aug. 2018.
- [29] N. Ma, Y. Peng, and S. Wang, "On-line hyperspectral anomaly detection with hypothesis test based model learning," *Infr. Phys. Technol.*, vol. 97, pp. 15–24, Mar. 2019.
- [30] X. Lu, W. Zhang, and J. Huang, "Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1527–1537, Mar. 2020.
- [31] 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.
- [32] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 3123–3134, Dec. 2015.
- [33] D. Gong *et al.*, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1705–1714.
- [34] C. Ding, D. Zhou, X. He, and H. Zha, "R<sub>1</sub> PCA: Rotational invariant L<sub>1</sub> norm principal component analysis for robust subspace factorization," in *Proc. 23rd Int. Conf. Mach. Learn.*, Jun. 2006, pp. 281–288.
- [35] L. Zhen, D. Peng, W. Wang, and X. Yao, "Kernel truncated regression representation for robust subspace clustering," *Inf. Sci.*, vol. 524, pp. 59–76, Jul. 2020.
- [36] G. Fan, Y. Ma, J. Huang, X. Mei, and J. Ma, "Robust graph autoencoder for hyperspectral anomaly detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 1830–1834.
- [37] W. Xie, J. Lei, B. Liu, Y. Li, and X. Jia, "Spectral constraint adversarial autoencoders approach to feature representation in hyperspectral anomaly detection," *Neural Netw.*, vol. 119, pp. 222–234, Nov. 2019.
- [38] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 818–832, Apr. 2017.
- [39] Z. Li, J. Liu, J. Tang, and H. Lu, "Robust structured subspace learning for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 2085–2098, Oct. 2015.
- [40] Y. Liao, Y. Wang, and Y. Liu, "Graph regularized auto-encoders for image representation," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2839–2852, Jun. 2017.
- [41] S. Yang, L. Li, S. Wang, W. Zhang, and Q. Huang, "A graph regularized deep neural network for unsupervised image representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7053–7061.
- [42] Y. Ma, C. Li, X. Mei, C. Liu, and J. Ma, "Robust sparse hyperspectral unmixing with  $\ell_{2,1}$  norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1227–1239, Dec. 2017.
- [43] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 665–674.
- [44] L. Jia, Z. Zhang, L. Wang, W. Jiang, and M. Zhao, "Adaptive neighborhood propagation by joint L<sub>2,1</sub>-norm regularized sparse coding for representation and classification," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 201–210.
- [45] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [46] Z. Huang, L. Fang, and S. Li, "Subpixel-pixel-superpixel guided fusion for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 5998–6007, Sep. 2020.

- [47] L. Wei and D. Qian, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [48] 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.
- [49] S. Li, K. Zhang, P. Duan, and X. Kang, "Hyperspectral anomaly detection with kernel isolation forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 319–329, Jan. 2020.



**Ganghui Fan** received the B.S. degree from the Electronic Information School, Wuhan University, Wuhan, China, in 2019, where he is pursuing the master's degree with the Multi-Spectral Vision Processing Laboratory.

His research interests include hyperspectral imagery and machine learning.



**Yong Ma** graduated from the Department of Automatic Control, Beijing Institute of Technology, Beijing, China, in 1997. He received the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003.

From 2004 to 2006, he was a Lecturer with the University of the West of England, Bristol, U.K. From 2006 to 2014, he was with the Wuhan National Laboratory for Optoelectronics, HUST, where he was a Professor of electronics. He is a Professor with the Electronic Information School, Wuhan University, Wuhan. His general field of research is in signal and systems. His research projects include remote sensing of the LiDAR and infrared, as well as infrared image processing, pattern recognition, interface circuits to sensors, and actuators.



**Xiaoguang Mei** (Member, IEEE) received the B.S. degree in communication engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2007, the M.S. degree in communications and information systems from Central China Normal University, Wuhan, in 2011, and the Ph.D. degree in circuits and systems from HUST in 2016.

From 2010 to 2012, he was a Software Engineer with the 722 Research Institute, China Shipbuilding Industry Corporation, Wuhan. From 2016 to 2019, he was a Post-Doctoral Fellow with the Electronic Information School, Wuhan University (WHU), Wuhan, where he is an Associate Professor. His research interests include hyperspectral imagery, machine learning, and pattern recognition.



**Fan Fan** received the B.S. degree in communication engineering and the Ph.D. degree in electronic circuits and systems from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2015, respectively.

He is a Post-Doctoral Researcher with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan. His research interests include infrared thermal imaging, machine learning, and computer vision.



**Jun Huang** received the B.S. and Ph.D. degrees from the Department of Electronic and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

He is an Associate Professor with the Electronic Information School, Wuhan University, Wuhan. His main research interests are infrared image processing and infrared spectrum processing.



**Jiayi Ma** (Senior Member, IEEE) received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

He is a Professor with the Electronic Information School, Wuhan University, Wuhan. He has authored or coauthored more than 200 refereed journal articles and conference papers, including IEEE TPAMI/TIP, IJCV, CVPR, ICCV, and ECCV. His research interests include computer vision, machine learning, and

pattern recognition.

Dr. Ma has been identified in the 2020 and 2019 Highly Cited Researchers list from the Web of Science Group. He is also an Area Editor of *Information Fusion*, an Associate Editor of *Neurocomputing*, *Sensors*, and *Entropy*, and a Guest Editor of *Remote Sensing*.