LREN: Low-Rank Embedded Network for Sample-Free Hyperspectral Anomaly Detection

Kai Jiang*, Weiying Xie*, Jie Lei, Tao Jiang, Yunsong Li

State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China
xdjiangkai@foxmail.com, wyxie@xidian.edu.cn, {jielei, ysli}@mail.xidian.edu.cn, tjiang_2@stu.xidian.edu.cn

Abstract

Hyperspectral anomaly detection (HAD) is a challenging task because it explores the intrinsic structure of complex high-dimensional signals without any samples at training time. Deep neural networks (DNNs) can dig out the underlying distribution of hyperspectral data but are limited by the labeling of large-scale hyperspectral datasets, especially the low spatial resolution of hyperspectral data, which makes labeling more difficult. To tackle this problem while ensuring the detection performance, we present an unsupervised low-rank embedded network (LREN) in this paper. LREN is a joint learning network in which the latent representation is specifically designed for HAD, rather than merely as a feature input for the detector. And it searches the lowest rank representation based on a representative and discriminative dictionary in the deep latent space to estimate the residual efficiently. Considering the physically mixing properties in hyperspectral imaging, we develop a trainable density estimation module based on Gaussian mixture model (GMM) in the deep latent space to construct a dictionary that can better characterize the complex hyperspectral images (HSIs). The closed-form solution of the proposed low-rank learner surpasses existing approaches on four real hyperspectral datasets with different anomalies. We argue that this unified framework paves a novel way to combine feature extraction and anomaly estimation-based methods for HAD, which intends to learn the underlying representation tailored for HAD without the prerequisite of manually labeled data. Code available at https://github.com/xdjiangkai/LREN.

Introduction

Recently, great strides have been made towards more precise target detection in natural images thanks in part to the large scale of labeled datasets, such as the COCO dataset (Lin et al. 2014), ImageNet (Deng et al. 2009) and DOTA dataset (Xia et al. 2018). In hyperspectral anomaly detection (HAD), however, it is difficult to manually and accurately label because of unavailable prior knowledge, complicated imaging environment, and low spatial resolution (Zare, Jiao, and Glenn 2018). Due to these real-world resource constraints, no labeled samples are captured for HAD, forming a typical sample-free detection task. Under these circumstances, it is highly essential to design a network specifically for the task of HAD.

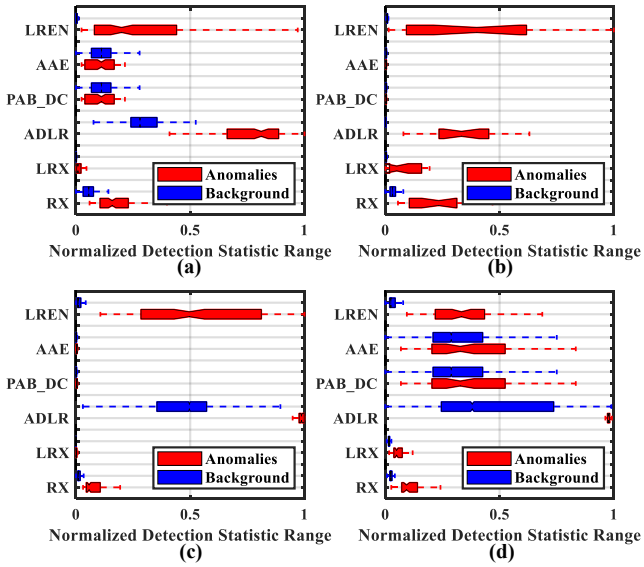

Figure 1: Background-anomaly separability results on (a) SanDiego, (b) Hydice, (c) Coast, and (d) Pavia. Each detector row corresponds to two boxes, where the red box indicates the distribution interval of the anomaly and the blue box indicates the distribution interval of the background. The distance between the upper boundary of the blue box and the lower boundary of the red box indicates the separability of the background-anomaly. The existing HAD methods cannot suppress the background and enhance the separation of the anomaly box from the background box.

A series of efforts dedicated to determining hyperspectral anomalies by DNNs without any prior knowledge have been developed. Most of which employ generative models including AE (Dong et al. 2018), DBN (Hinton and Salakhutdinov 2006) and GAN (Goodfellow et al. 2014) to extract features in the deep latent space or the reconstruction space by minimizing the error between the original and reconstructed spectra. Subsequently, separate detectors such as RX (Reed and Yu 1990), LRX (Matteoli et al. 2013), CRD (Li and Du

*These authors contribute equally to this work.

2014), AED (Kang et al. 2017), and LSDM-MoG (Li et al. 2020) are performed on these features to achieve final detection. Typically, these methods are optimized in two separate steps: extracting features of hyperspectral images (HSIs) with DNNs, and then performing detection on the extracted features. Although these methods tend to mitigate the problem of dimensional disaster, their objective functions are not customized for anomaly detection. This would lead to a suboptimal solution because the two steps have not been learned together to maximize their respective strengths through cooperation. Moreover, due to remote non-contact observations and numerous imaging scenes, real HSIs often contain rich and complex structures. In addition, noise corruption further poses a challenge to HAD. Consequently, when applied to real scenes, most existing methods fail to effectively distinguish the anomaly from the background, as shown in Figure 1.

A natural question is: Particularly for this sample-free task, is it possible to establish a unified network with both dimension reduction and anomaly detection to ensure detection performance? In order to address this issue, we systematically discuss various structures and propose our network structure based on the following three properties: 1) the background still preserves a low-rank property lying in a low-dimensional manifold; 2) the presence probability of the anomaly is much lower than that of the background; and 3) the latent representation serves the anomaly estimation, which optimally updates the parameters of the deep latent space. In this way, we propose a low-rank embedded network (LREN) for HAD to adapt the learned latent representation to a specific density estimation. To the best of our knowledge, this is the first method tackling this challenging problem. In contrast to existing approaches, an integrated loss, which includes a reconstruction error and a density estimation loss, is defined to optimize the LREN jointly. We formulate the unsupervised learning-based HAD problem as a low-rank optimization with a closed-form solution in a low-dimensional manifold. This method constructs a representative and discriminative dictionary that best characterizes the underlying structure of complex HSIs, breaking the dilemma of limited detection performance for sample-free detection. Notably, Gaussian mixture model, a weighted finite sum of Gaussian components, is embedded into the LREN to estimate the complex HSIs. Extensive experiments demonstrate the efficacy of LREN in generalizing to unknown targets and different datasets. Our main contributions are summarized as follows.

- We propose an unsupervised low-rank learner, called LREN, which is a unified framework for formulating the sample-free HAD problem as a low-rank optimization in a low-dimensional manifold. This design can not only effectively alleviate the misalignment between feature extraction and anomaly estimation, but also avoid the difficulty of decomposing high-dimensional data.
- A novel learnable density estimation module is developed, which models the physically mixing properties in hyperspectral imaging into the LREN to construct a discriminative dictionary while retaining the latent representations that are customized end-to-end for dictionary construction. This universal module can be embedded into any other HAD networks.
- To find the optimum of the low-rank learner, LREN, more efficiently, we shrink the solution space to the latent space of a low-dimensional closed-form solution. We show that LREN achieves state-of-the-art results on four real hyperspectral datasets.

Related Work

Broadly speaking, anomaly detection is a significant task in artificial intelligence to find anomalous instances that substantially deviate from a normal distribution. The DNN-based anomaly detection technology has been applied to many fields (Liu et al. 2013; Zhao et al. 2017; Abati et al. 2019; Markovitz et al. 2020; Pang et al. 2020; Park, Noh, and Ham 2020). They can be grouped into two typical categories including reconstruction-based and estimation-based methods to learn models describing normal samples. Most of these methods are trained on normal samples and even on samples that are both normal and anomalous. Therefore, they have limited applications in HAD in the absence of any background or anomaly samples at training time.

With respect to HAD, the existing DNN-based works mainly focus on feature extraction or dimension reduction through unsupervised learning, because it is a sample-free task. For example, Lei *et al*. (Lei et al. 2019) adopt DBN to learn the feature matrix and input it into the subsequent detector. Xie *et al*. (Xie et al. 2020) present a spectral adversarial feature learning (SAFL) architecture for HAD to extract underlying spectral features in deep latent space. Very recently, Jiang *et al*. (Jiang et al. 2020) propose a spectral constrained GAN to reduce the data dimensionality. These approaches can effectively capture the general structure, however, lead to a suboptimal solution, because the feature extraction and anomaly detection have not been learned together to maximize their respective strengths through cooperation. Moreover, the physically mixing properties of complex HSIs in real scenarios are not yet plugged into the network. In contrast, the proposed LREN can successfully address these issues.

Low-Rank Embedded Network

Problem Formulation

In significant contrast to the background, anomalies in HSIs are rarely present with sparse property, making it difficult to describe anomalies directly using data-driven models in the absence of samples. Besides, the presence of noise poses a challenge to HAD. LREN infers the anomaly based on the motivations that in the low-dimensional manifold, the background still preserves a low-rank property, while the anomaly presents a sparse property. Let us denote by $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_i, \cdots, \mathbf{h}_N] \in \mathbb{R}^{L \times N}$ an input HSI, where $\mathbf{h}_i, (i = 1, ..., N)$ is the ith spectral vector with L dimensions in the original data space. Denoting the dictionary as $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_i, \cdots, \mathbf{d}_{2k}] \in \mathbb{R}^{l \times 2k}$, we focus on minimizing the low-rank optimization with a closed-form solution in

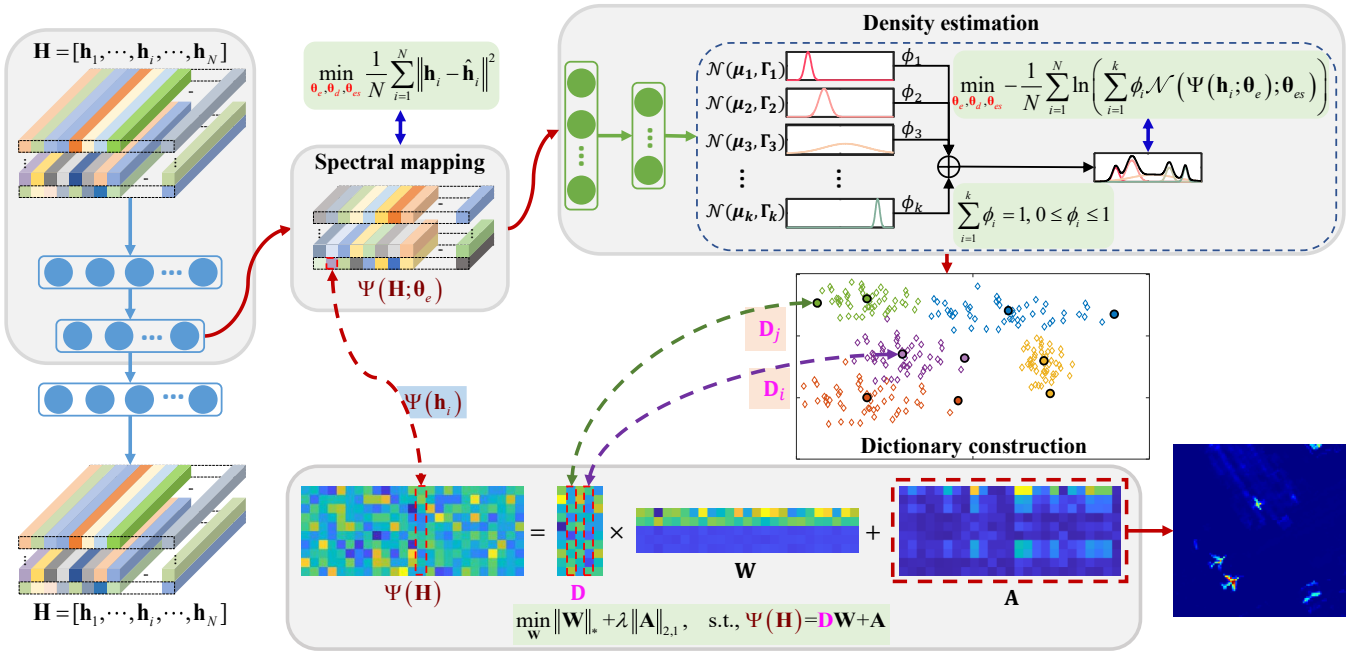

Figure 2: Overview of our LREN-based HAD with k Gaussian mixture-components. Note that the spectral mapping parameterized by $\boldsymbol{\theta}_e$ and $\boldsymbol{\theta}_d$, and the density estimation parameterized by $\boldsymbol{\theta}_{es}$ are jointly learned.

a low-dimensional manifold as follows:

$$\begin{aligned} \min_{\mathbf{W}} \quad & ||\mathbf{W}||_* + \lambda ||\mathbf{A}||_{2,1} \\ \text{s.t.}, \quad & \Psi(\mathbf{H}) = \mathbf{DW} + \mathbf{A}, \end{aligned} \tag{1}$$

where $||\mathbf{W}||_*$ represents the nuclear norm of the weight matrix $\mathbf{W} \in \mathbb{R}^{2k \times N}$ and $\|\mathbf{A}\|_{2,1} = \sum_{i=1}^{L} \sqrt{\sum_{j=1}^{N} (a_{i,j})^2}$ denotes the $\ell_{2,1}$-norm of the anomaly matrix $\mathbf{A}$. $\Psi(\cdot)$ represents the spectral mapping function from the high-dimensional data space to the low-dimensional representation space. The weight matrix $\mathbf{W}$ is the coefficient for $\Psi(\mathbf{H})$ with respect to the dictionary $\mathbf{D}$. The latent representation matrix $\Psi(\mathbf{H}) = [\Psi(\mathbf{h}_1), \cdots, \Psi(\mathbf{h}_i), \cdots, \Psi(\mathbf{h}_N)] \in \mathbb{R}^{l \times N}, (l \ll L)$ containing most of the critical information in the original data space, where $\Psi(\mathbf{h}_i)$ is the ith spectral representation with l dimensions.

To make Eq. (1) convex, we should achieve spectral mapping and dictionary construction. It is worth noting that LREN innovatively takes into account both the spectral mapping function and the dictionary as shown in Figure 2, that is, the features extracted in the deep latent space are customized for dictionary construction in an end-to-end manner. We next discuss in detail the optimized solution to Eq. (1) via the unified network.

Network Architecture

Instead of separately modeling the spectral mapping and dictionary as in Eq. (1), we present a deep network, i.e., LREN, to directly model the latent representation and its dictionary, which learns the intrinsic characterization that is customized for HAD implicitly. In this way, both the spectral mapping module and the dictionary construction module can be jointly learned in an unsupervised manner as

$$\begin{aligned} \min_{\boldsymbol{\theta}_e, \boldsymbol{\theta}_d, \boldsymbol{\theta}_{es}} & \frac{1}{N} \sum_{i=1}^{N} \left[\left\| \mathbf{h}_i - \hat{\mathbf{h}}_i \right\|^2 + \Phi\left(\Psi\left(\mathbf{h}_i; \boldsymbol{\theta}_e\right); \boldsymbol{\theta}_{es}\right) \right] \\ \text{s.t.}, \quad & \hat{\mathbf{h}}_i = \Pi\left(\Psi\left(\mathbf{h}_i; \boldsymbol{\theta}_e\right); \boldsymbol{\theta}_d\right), \end{aligned} \tag{2}$$

where $\Psi\left(\cdot; \boldsymbol{\theta}_e\right)$ denotes the proposed spectral encoder network that takes the observed spectral vector $\mathbf{h}_i$ as input and outputs the latent spectral representation $\mathbf{z}_i = \Psi\left(\mathbf{h}_i; \boldsymbol{\theta}_e\right)$ that is input into the proposed spectral decoder $\Pi\left(\cdot; \boldsymbol{\theta}_d\right)$ to generate the reconstructed version of $\mathbf{h}_i$. The second term is the proposed objective function for density estimation $\Phi\left(\cdot; \boldsymbol{\theta}_{es}\right)$ with the goal of constructing a strongly representative and discriminating dictionary.

Spectral mapping module. We describe the spectral mapping module in detail in Figure 2. The spectral mapping autoencoder takes the original spectral vector $\mathbf{h}_i$ as the input, whereas the reconstructed spectral vector $\hat{\mathbf{h}}_i$ is the output as follows:

$$\mathbf{z}_i = \Psi\left(\mathbf{h}_i; \boldsymbol{\theta}_e\right), \tag{3}$$

$$\hat{\mathbf{h}}_i = \Pi\left(\mathbf{z}_i; \boldsymbol{\theta}_d\right). \tag{4}$$

Note that both the mapping function $\Psi\left(\cdot; \boldsymbol{\theta}_e\right)$ from the input to the latent representation parameterized by $\boldsymbol{\theta}_e$ and the mapping function $\Pi\left(\cdot; \boldsymbol{\theta}_d\right)$ from the latent representation to the reconstructed version of the input parameterized by $\boldsymbol{\theta}_d$ contain multi-layer fully connected layers followed by the tanh activation. We solve the following problem to learn the

spectral mapping module.

$$\min_{\boldsymbol{\theta}_e,\boldsymbol{\theta}_d,\boldsymbol{\theta}_{es}} \frac{1}{N}\sum_{i=1}^{N}\left\|\mathbf{h}_i-\hat{\mathbf{h}}_i\right\|^2. \tag{5}$$

This objective function makes the reconstructed spectral vector $\hat{\mathbf{h}}_i$ approximate to the input spectral vector as much as possible. The features of the latent space can be constructed entirely with spectral vectors that are consistent with the input. Consequently, the objective function, in essence, aims to dig out the representative and discriminative information of the original high-dimensional spectral vector, which is defined as $\Psi(\mathbf{H}) = [\Psi(\mathbf{h}_1), \cdots, \Psi(\mathbf{h}_i), \cdots, \Psi(\mathbf{h}_N)] \in \mathbb{R}^{l\times N}$ in Eq. (1).

Dictionary construction module. The density estimation $\Phi(\cdot;\boldsymbol{\theta}_{es})$ parameterized by $\boldsymbol{\theta}_{es}$ in Eq. (2) is the key problem in the dictionary construction module. The proposed estimation discovers a set of clusters from the latent representation, denoted as k, and assigns the centroid of each cluster as the basis to form a dictionary atom. To further reflect the size of each cluster, we select the sample farthest from the centroid and let it act as a dictionary atom along with the centroid in $\mathbf{D}$. Notably, a critical issue is the establishment of $\Phi(\cdot;\boldsymbol{\theta}_{es})$ on the latent representation for the high-efficiency purpose of density estimation. We define the objective function for density estimation parameterized by $\boldsymbol{\theta}_{es}$ as

$$\min_{\boldsymbol{\theta}_e,\boldsymbol{\theta}_d,\boldsymbol{\theta}_{es}} \frac{1}{N}\sum_{i=1}^{N}\Phi\left(\Psi(\mathbf{h}_i;\boldsymbol{\theta}_e);\boldsymbol{\theta}_{es}\right). \tag{6}$$

In the proposed LREN, the output $\Phi(\Psi(\mathbf{h}_i;\boldsymbol{\theta}_e);\boldsymbol{\theta}_{es})$ is input to the softmax activation to obtain a k-dimensional vector for the mixture-component prediction, i.e.,

$$\beta_i = \text{softmax}\left(\Phi(\mathbf{h}_i;\boldsymbol{\theta}_e);\boldsymbol{\theta}_{es}\right). \tag{7}$$

Considering the complexity of HSIs, we further formulate our proposed density estimation as

$$\begin{aligned}\min_{\boldsymbol{\theta}_e,\boldsymbol{\theta}_d,\boldsymbol{\theta}_{es}} & -\frac{1}{N}\sum_{i=1}^{N}\ln\left(\sum_{i=1}^{k}\phi_i\mathcal{N}\left(\Psi(\mathbf{h}_i;\boldsymbol{\theta}_e);\boldsymbol{\theta}_{es}\right)\right)\\ \text{s.t.,}\quad & \sum_{i=1}^{k}\phi_i=1,\ 0\leqslant\phi_i\leqslant 1,\end{aligned} \tag{8}$$

following the framework of GMM because of its strong representational capability of complex data. Here, $\boldsymbol{\theta}_{es}$ contains the mean vector μ_i and the covariance matrix $\mathbf{\Gamma}_i$ for component i. ϕ_i is the nonnegative weight that sums to 1. All these parameters are trained and updated efficiently by the proposed LREN given a batch of Q spectral samples as follows:

$$\phi_k = \sum_{i=1}^{Q}\frac{\beta_{ik}}{Q}, \tag{9}$$

$$\mu_k = \frac{\sum_{i=1}^{Q}\beta_{ik}\Psi(\mathbf{h}_i;\boldsymbol{\theta}_e)}{\sum_{i=1}^{Q}\beta_{ik}}, \tag{10}$$

$$\mathbf{\Gamma}_k = \frac{\sum_{i=1}^{Q}\beta_{ik}\left(\Psi(\mathbf{h}_i;\boldsymbol{\theta}_e)-\mu_k\right)\left(\Psi(\mathbf{h}_i;\boldsymbol{\theta}_e)-\mu_k\right)^T}{\sum_{i=1}^{Q}\beta_{ik}}. \tag{11}$$

The parameters of the density estimation network are learned together with the parameters of the previous latent representation network in our LREN. Once the parameters are updated to the optimum, the dictionary $\mathbf{D}\in\mathbb{R}^{l\times 2k}$ is constructed by the centroid of each cluster (i.e., β_i) along with the sample farthest from the centroid.

Solving the Optimized Model

Given the latent representation $\Psi(\mathbf{H})$ and the dictionary $\mathbf{D}$, we can solve the convex function Eq. (1) with a closed-form solution by the alternating direction method in (Liu, Lin, and Yu 2010). An auxiliary variable $\mathbf{J}$ which is used as a split variable to replace $\mathbf{W}$, is introduced to make the objective function separable. Thus, the first term $||\mathbf{W}||_*$ in the objective function can be replaced with $||\mathbf{J}||_*$. And the original problem (1) can be reformulated as the following problem:

$$\begin{aligned}\min_{\mathbf{W}}\quad & ||\mathbf{W}||_* + \lambda||\mathbf{A}||_{2,1}\\ \text{s.t.,}\quad & \begin{cases}\Psi(\mathbf{H}) = \mathbf{DW} + \mathbf{A}\\ \mathbf{W}=\mathbf{J}\end{cases}.\end{aligned} \tag{12}$$

The augmented Lagrangian function of (12) is

$$\begin{aligned}L =& ||\mathbf{J}||_* + \lambda||\mathbf{A}||_{2,1} + tr[\mathbf{Y}_1^T(\Psi(\mathbf{H})-\mathbf{DW}-\mathbf{A})]\\ &+ tr[\mathbf{Y}_2^T(\mathbf{W}-\mathbf{J})]\\ &+ \frac{\mu}{2}(||\Psi(\mathbf{H})-\mathbf{DW}-\mathbf{A}||_F^2 + ||\mathbf{W}-\mathbf{J}||_F^2)\\ =& ||\mathbf{J}||_* + \lambda||\mathbf{A}||_{2,1} + \frac{\mu}{2}(||\mathbf{J}-(\mathbf{W}+\mathbf{Y}_2/\mu)||_F^2\\ &+ ||\mathbf{A}-(\Psi(\mathbf{H})-\mathbf{DW}+\mathbf{Y}_1/\mu)||_F^2)\\ &- \frac{1}{2\mu}(||\mathbf{Y}_1||_F^2 + ||\mathbf{Y}_2||_F^2),\end{aligned} \tag{13}$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are Lagrange multipliers, $\mu>0$ denotes the penalty parameter, $tr(\cdot)$ refers to the trace of a matrix, and $||\cdot||_F$ refers to the Frobenius norm. Therefore, the problem becomes:

$$\begin{aligned}\min_{\mathbf{W},\mathbf{J},\mathbf{A},\mathbf{Y}_1^T,\mathbf{Y}_2^T} & ||\mathbf{J}||_* + \lambda||\mathbf{A}||_{2,1}\\ &+ \frac{\mu}{2}(||\mathbf{J}-(\mathbf{W}+\mathbf{Y}_2/\mu)||_F^2\\ &+ ||\mathbf{A}-(\Psi(\mathbf{H})-\mathbf{DW}+\mathbf{Y}_1/\mu)||_F^2)\\ &- \frac{1}{2\mu}(||\mathbf{Y}_1||_F^2 + ||\mathbf{Y}_2||_F^2).\end{aligned} \tag{14}$$

Problem (14) can be solved with the alternating direction method by alternatively updating only one variable a time while fixing the others. During the kth iteration, the problem can be divided into the following steps:

1) $\mathbf{J}$ step:

$$\min_{\mathbf{J}} ||\mathbf{J}||_* + \frac{\mu}{2}(||\mathbf{J}-(\mathbf{W}+\mathbf{Y}_2/\mu)||_F^2. \tag{15}$$

2) $\mathbf{W}$ step:

$$\min_{\mathbf{W}} \frac{\mu}{2}(||\mathbf{J}-(\mathbf{W}+\mathbf{Y}_2/\mu)||_F^2 + ||\mathbf{A}-(\Psi(\mathbf{H})-\mathbf{D}\mathbf{W}+\mathbf{Y}_1/\mu)||_F^2) \quad (16)$$

$$\mathbf{W} = (\mathbf{D}^\mathbf{T}\mathbf{D}+\mathbf{I})^{-1}[\frac{1}{\mu}(\mathbf{D}^\mathbf{T}\mathbf{Y_1}-\mathbf{Y_2}) + \mathbf{D}^\mathbf{T}\Psi(\mathbf{H})-\mathbf{D}^\mathbf{T}\mathbf{A}+\mathbf{J}]. \quad (17)$$

3) $\mathbf{A}$ step:

$$\min_{\mathbf{A}} \lambda||\mathbf{A}||_{2,1}+||\mathbf{A}-(\Psi(\mathbf{H})-\mathbf{D}\mathbf{W}+\mathbf{Y}_1/\mu)||_F^2). \quad (18)$$

4) $\mathbf{Y}_1, \mathbf{Y}_2$ step:

$$\mathbf{Y}_1 = \mathbf{Y}_1 + \mu(\Psi(\mathbf{H})-\mathbf{D}\mathbf{W}-\mathbf{A}), \quad (19)$$

$$\mathbf{Y}_2 = \mathbf{Y}_2 + \mu(\mathbf{W}-\mathbf{J}). \quad (20)$$

Once the optimization is finished, the anomaly matrix $\mathbf{A}$ can be extracted, and the anomaly score for the ith spectrum $\mathbf{x}_i$ can be determined as $||\mathbf{A}_i||_1$, where $\mathbf{A}_i$ is the ith column of anomaly matrix $\mathbf{A}$.

Experiments

Experimental Settings

Datasets We evaluate our LREN on four benchmark hyperspectral datasets, including SanDiego (Xu et al. 2016), Hydice (Li and Du 2014), Coast (Kang et al. 2017), and Pavia (Kang et al. 2017). SanDiego dataset collected by the Airborne Visible/ Infrared Imaging Spectrometer (AVIRIS) sensor consists of 100×100 pixels and 189 spectral bands covering the wavelength range from 400 to 2500nm. Three aircrafts with structural information are considered as anomalous targets of interest. Hydice dataset was collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) airborne sensor over an urban area, CA, USA. There are 80×100 pixels and 162 spectral bands in a wavelength range of 400 to 2500nm. The vehicles and roofs are considered as anomalies and their pixels account for about 0.24% of the entire scene. Coast dataset was also acquired by the AVIRIS sensor, which contains 10000 spectral vectors, each with 207 channels, covering a range of 450 to 1350nm. Buildings of different scales are regarded as anomalies. Pavia dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor over Pavia city, which consists of 100×100 pixels and 102 spectral bands ranging from 430 to 860nm. The vehicles are considered as anomalies containing 63 pixels and account for 0.63% of the entire scene. It is worth noting that the datasets we used contain both point and structural anomalies, and are in the form of different scales.

Evaluation metrics To quantitatively evaluate the performance of our LREN and the state-of-the-art approaches, the receiver operating characteristic curve (ROC) together with the area under it are used. The area under the ROC curve of (P_d, P_f) demonstrates the detection accuracy, and the closer this value is to 1, the better the detection capability. Conversely, the closer the area value under the (P_f, τ) is to 0, the better the false alarm rate. Here, P_d, P_f and τ represent the true positive rate, false positive rate and threshold, respectively. Besides, Box-Whisker Plot (Manolakis and Shaw 2002) is also used to indicate the degree of background suppression and separation from the anomaly.

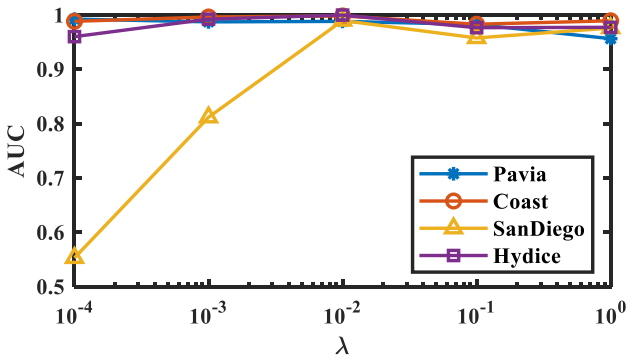

Figure 3: Effects of the parameter λ for SanDiego, Hydice, Coast, and Pavia.

Implementation details The proposed LREN is composed of encoder, decoder and estimator, and each part is composed of a two-layer fully connected network. The number of hidden nodes in the deep latent space is set to 9. We train LREN with SGD in an end-to-end fashion setting the learning rate to 10^{-4} and the batch size to the number of input spatial pixels. Intuitively, it is reasonable to estimate the distribution of complex HSIs with the larger parameter k, however, it takes too much time to search the dictionary atom. In LREN, we empirically set the number of Gaussian mixture-components k to 7 considering the tradeoff between detection performance and searching time. We terminate the learning process in 1000 epochs. In order to get a better understanding of the anomaly in HSIs, we evaluate the effect of different values of λ in problem (1) on detection. With λ varying from 10^{-4} to 1, the optimal λ values for different datasets are shown in Figure 3. We can see that the AUC scores achieve the bests when λ is 10^{-2} for SanDiego, Hydice, Coast datasets, and 10^{-3} for Pavia dataset, respectively. We implement our method by TensorFlow on one NVIDIA 2080 Ti GPUs with 8 GB memory.

Ablation Study

In this section, we conduct an ablation study to evaluate the contribution of different components in the proposed LREN to HAD. In detail, the following five scenarios are designed: (A) Our proposed LREN eliminates $\ell_{1,2}$-norm of the anomaly. (B) The proposed dictionary construction module is replaced with Kmeans algorithm. (C) We search the lowest rank representation in the original space. (D) The spectral mapping module and the dictionary construction module are separately learned. (E) We consider the full proposed model, LREN. The results are summarized in Table 1.

As displayed in Table 1, the LREN achieves a 0.9951 AUC score of (P_d, P_f) on average. By removing $\ell_{1,2}$ -norm of the anomaly, we can see that the result is decreased to

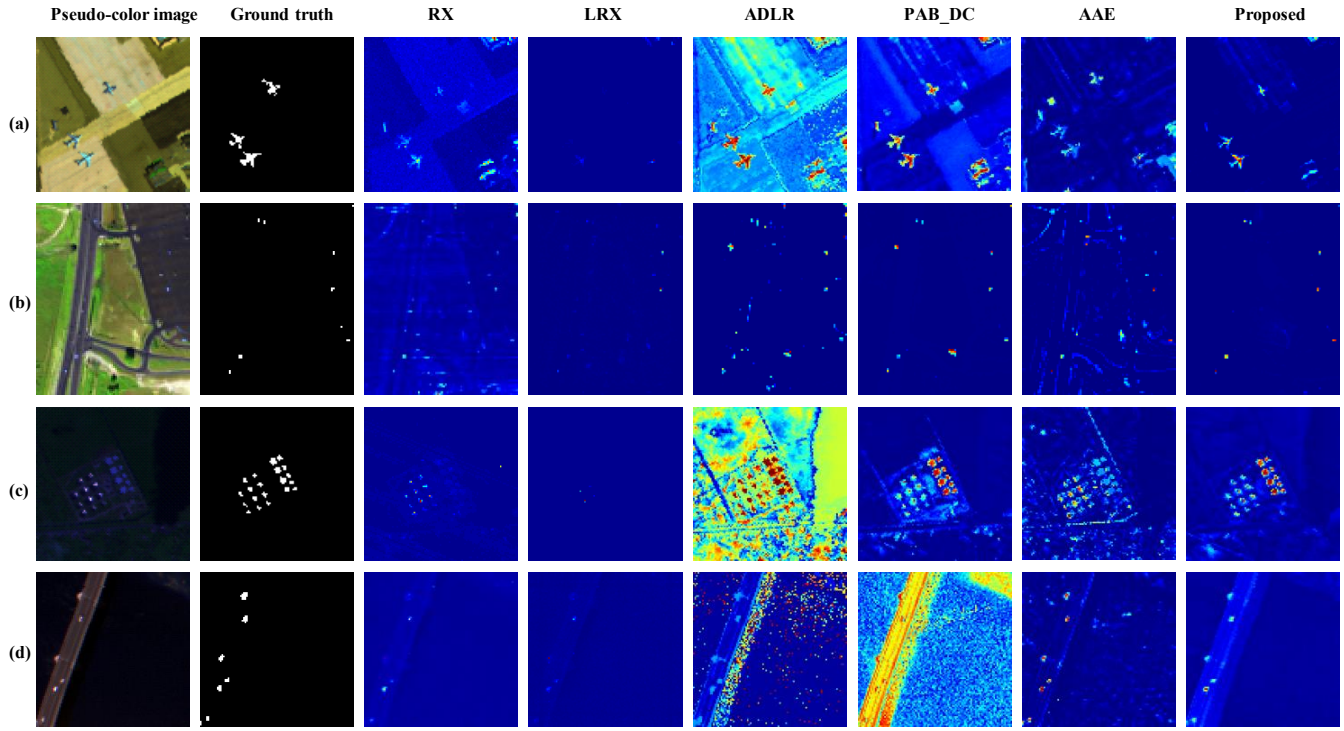

Figure 4: Visual quality comparison on four typical scenes in HSI datasets. The pseudocolor image, ground truth, and detection map of the compared methods for (a) SanDiego, (b) Hydice, (c) Coast, and (d) Pavia.

Description	(P_d, P_f)	(P_f, τ)
(A)	0.8692	0.1418
(B)	0.7677	0.2180
(C)	0.9843	0.0329
(D)	0.9428	0.1554
(E)	**0.9951**	**0.0236**

Table 1: Average results of ablation study for LREN performed on four datasets. The best result is highlighted.

Method	RX	LRX	ADLR	PAB_DC	AAE	Proposed
SanDiego	0.9403	0.9687	<u>0.9894</u>	0.9861	0.9595	**0.9897**
Hydice	0.9763	0.9327	0.9962	<u>0.9983</u>	0.9218	**0.9998**
Coast	<u>0.9946</u>	0.9926	0.9937	0.9870	0.9681	**0.9982**
Pavia	<u>0.9887</u>	0.9840	0.9507	0.9000	0.9822	**0.9925**
Average	0.9750	0.9695	<u>0.9825</u>	0.9679	0.9579	**0.9951**

Table 2: AUC scores of (P_d, P_f) for different algorithms on four datasets. The best and second best results are highlighted and underlined.

0.8692, delivering a decreasement of 0.1259, which indicates that the $\ell_{1,2}$ -norm of the anomaly is effective in HAD. In the (B) row, by simply replacing the density estimation network with Kmeans, the AUC score of (P_d, P_f) is significantly reduced to 0.7677 and the AUC score of (P_d, P_f) is increased to 0.2180, which illustrates the density estimation network has a positive influence on HAD. The (C) row validates that the low-rank optimization in the latent space is also effective in our LREN. Besides, decomposing the matrix on the low-dimensional latent space reduces the complexity of decomposition on the higher-dimensional original space. When jointly learning the spectral mapping and dictionary construction modules, we find that it raises the AUC score of (P_d, P_f) from 0.9428 to 0.9951 and decreases the AUC score of (P_f, τ) from o.1554 to 0.0236 as listed in (D) and (E) rows.

Evaluation of HAD Algorithm

For comparison, five commonly used methods, including RX (Reed and Yu 1990), LRX (Matteoli et al. 2013), ADLR (Qu et al. 2018), PAB_DC (Huyan et al. 2019), and AAE (Xie et al. 2020), are selected. All these methods are reimplemented according to their papers and open-source codes. The quantized AUC scores of (P_d, P_f) and (P_f, τ) are listed in Table 2 and Table 3, respectively. From Table 2 and Table 3, we can observe that the proposed model outperforms the competing methods on these benchmark datasets.

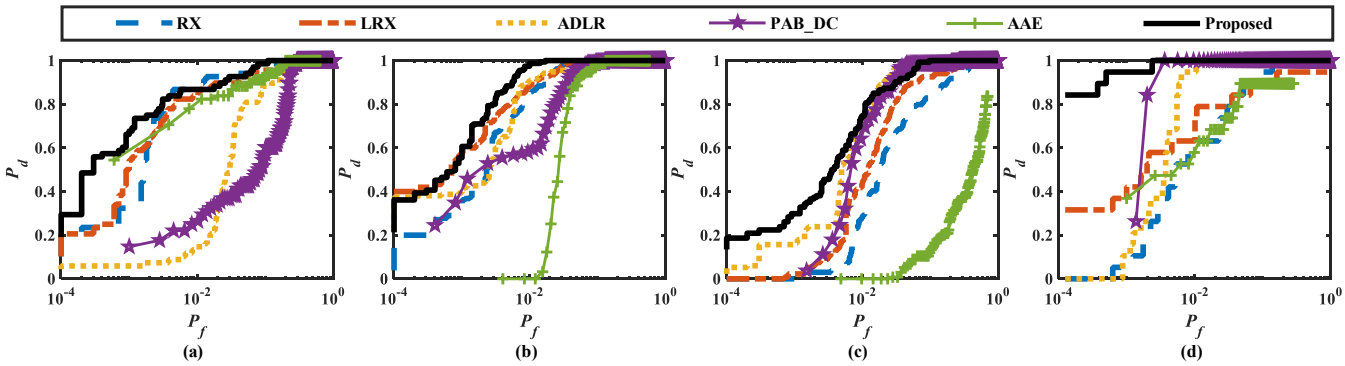

Figure 5: ROC curves of (P_d, P_f) for different algorithms on (a) SanDiego, (b) Hydice, (c) Coast, and (d) Pavia.

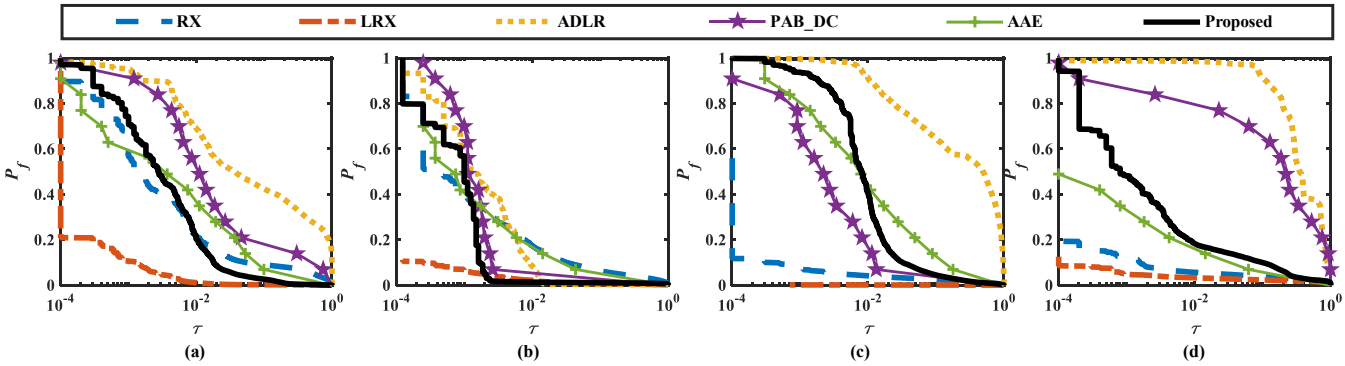

Figure 6: ROC curves of (P_f, τ) for different algorithms on (a) SanDiego, (b) Hydice, (c) Coast, and (d) Pavia.

Method	RX	LRX	ADLR	PAB_DC	AAE	Proposed
SanDiego	0.0599	**0.0025**	0.3049	0.1193	0.0324	<u>0.0134</u>
Hydice	0.0390	**0.0050**	<u>0.0060</u>	<u>0.0060</u>	0.0106	0.0102
Coast	0.0145	**0.0015**	0.4658	<u>0.0067</u>	0.0433	0.0276
Pavia	0.0268	**0.0178**	0.4615	0.3425	<u>0.0189</u>	0.0433
Average	0.0351	**0.0067**	0.3096	0.1186	0.0263	<u>0.0236</u>

Table 3: AUC scores of (P_f, τ) for different algorithms on four datasets. The best and second best results are highlighted and underlined.

Specifically, relative to the AUC score of (P_d, P_f), it exceeds the second-best ADLR by an average of 1.26%. More notably, its AUC score of (P_f, τ) is an order of magnitude less than that of ADLR on average. The corresponding visual detection maps, ROC curves of (P_d, P_f), and ROC curves of (P_d, τ) over four datasets are shown in Figure 4, Figure 5, and Figure 6, respectively. It is easy to observe that the detection maps obtained by the proposed LREN are closest to the ground truths. Meanwhile, the proposed LREN usually demonstrates higher P_d as P_f varying from 10^{-4} to 1 as shown in Figure 5.

Conclusion

One key issue in HAD is how to develop a unified HAD model for performing a joint feature extraction and density estimation without labeled samples, which has not been adequately resolved under the previous two-step framework. We propose an unsupervised low-rank learner called LREN to solve the HAD task as a low-rank optimization in a low-dimensional manifold. In consideration of the physically mixing properties in HS imaging, we design a trainable density estimation module in the deep latent space to construct a dictionary that can better characterize the complex HSIs. Our approach shrinks the solution space to the latent space of a low-dimensional closed-form solution which can not only maximize the respective strengths of feature extraction and density estimation through joint learning but also avoid the difficulty of decomposing high-dimensional data. Extensive experiments are conducted to validate the LREN has higher performance than those state-of-the-art approaches for different anomalies.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 62071360, Grant 61801359, in part by the Fundamental Research Funds for the Central Universities and the Innovation Fund of Xidian University under Grant 20109205456.

References

Abati, D.; Porrello, A.; Calderara, S.; and Cucchiara, R. 2019. Latent space autoregression for novelty detection. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 481–490.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Li, F.-F. 2009. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Dong, G.; Liao, G.; Liu, H.; and Kuang, G. 2018. A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images. *IEEE Geoscience and Remote Sensing Magazine* 6(3): 44–68.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *arXiv preprint arXiv:1406.2661* 187–208.

Hinton, G. E.; and Salakhutdinov, R. R. 2006. Reducing the dimensionality of data with neural networks. *Science* 313(5786): 504–507.

Huyan, N.; Zhang, X.; Zhou, H.; and Jiao, L. 2019. Hyperspectral anomaly detection via background and potential anomaly dictionaries construction. *IEEE Transactions on Geoscience and Remote Sensing* 57(4): 2263–2276.

Jiang, T.; Li, Y.; Xie, W.; and Du, Q. 2020. Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection. *IEEE Transactions on Geoscience Remote Sensing* 58(7): 4666–4679.

Kang, X.; Zhang, X.; Li, S.; Li, K.; Li, J.; and Benediktsson, J. A. 2017. Hyperspectral anomaly detection with attribute and edge-preserving filters. *IEEE Transactions on Geoscience Remote Sensing* 55(10): 5600–5611.

Lei, J.; Xie, W.; Yang, J.; Li, Y.; and Chang, C.-I. 2019. Spectral-spatial feature extraction for hyperspectral anomaly detection. *IEEE Transactions on Geoscience Remote Sensing* 57(10): 8131–8143.

Li, L.; Li, W.; Du, Q.; and Tao, R. 2020. Low-rank and sparse decomposition with mixture of gaussian for hyperspectral anomaly detection. *IEEE Transactions on Cybernetics,* doi: 10.1109/TCYB.2020.2968750.

Li, W.; and Du, Q. 2014. Collaborative representation for hyperspectral anomaly detection. *IEEE Transactions on Geoscience Remote Sensing* 53(3): 1463–1474.

Lin, T.-Y.; Maire, M.; Belongie, S. J.; Hays, J.; Perona, P.; Ramanan, D.; Dollr, P.; and Zitnick, C. L. 2014. Microsoft COCO: Common objects in context. In *Proceedings of the European Conference on Computer Vision*, 740–755.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(1): 171–184.

Liu, G.; Lin, Z.; and Yu, Y. 2010. Robust subspace segmentation by low-rank representation. In *Proceedings of the IEEE International Conference on Machine Learning*, 663–670.

Manolakis, D.; and Shaw, G. 2002. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine* 19(1): 29–43.

Markovitz, A.; Sharir, G.; Friedman, I.; Zelnik-Manor, L.; and Avidan, S. 2020. Graph embedded pose clustering for anomaly detection. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 10539–10547.

Matteoli, S.; Veracini, T.; Diani, M.; and Corsini, G. 2013. A locally adaptive background density estimator: An evolution for RX-based anomaly detectors. *IEEE Geoscience and Remote Sensing Letters* 11(1): 323–327.

Pang, G.; Yan, C.; Shen, C.; van den Hengel, A.; and Bai, X. 2020. Self-trained deep ordinal regression for end-to-end video anomaly detection. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 12173–12182.

Park, H.; Noh, J.; and Ham, B. 2020. Learning memory-guided normality for anomaly detection. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 14372–14381.

Qu, Y.; Wang, W.; Guo, R.; Ayhan, B.; Kwan, C.; Vance, S.; and Qi, H. 2018. Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition. *IEEE Transactions on Geoscience and Remote Sensing* 56(8): 4391–4405.

Reed, I. S.; and Yu, X. 1990. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Transactions on Acoustics Speech and Signal Processing* 38(10): 1760–1770.

Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; and Zhang, L. 2018. DOTA: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3974–3983.

Xie, W.; Liu, B.; Li, Y.; Lei, J.; Chang, C.-I.; and He, G. 2020. Spectral adversarial feature learning for anomaly detection in hyperspectral imagery. *IEEE Transactions on Geoscience Remote Sensing* 58(4): 2352–2365.

Xu, Y.; Wu, Z.; Li, J.; Plaza, A.; and Wei, Z. 2016. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Transactions on Geoscience and Remote Sensing* 54(4): 1990–2000.

Zare, A.; Jiao, C.; and Glenn, T. 2018. Discriminative multiple instance hyperspectral target characterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40(10): 2342–2354.

Zhao, Y.; Deng, B.; Shen, C.; Liu, Y.-W.; Lu, H.; and Hua, X.-S. 2017. Spatio-temporal autoencoder for video anomaly detection. In *Proceedings of the ACM Multimedia*, 1933–1941.