SPARSE CODING-INSPIRED GAN FOR WEAKLY SUPERVISED HYPERSPECTRAL ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Anomaly detection (AD) on hyperspectral images (HSIs) is of great importance in both space exploration and earth observations. However, the challenges caused by insufficient datasets, no labels, and noise corruption substantially downgrade the quality of detection. For solving these problems, this paper proposes a sparse coding-inspired generative adversarial network (GAN) for weakly supervised HAD, named sparseHAD. It can learn a discriminative latent reconstruction with small errors for background samples and large errors for anomaly samples. First, we design a novel background-category searching step to eliminate the difficulty of data annotation and prepare for weakly supervised learning . Then, a sparse coding-inspired regularized network is integrated into an end-to-end GAN to form a weakly supervised spectral mapping model consisting of two encoders, a decoder, and a discriminator. This model not only makes the network more robust and interpretable both experimentally and theoretically but also develops a new sparse coding-inspired path for HAD. Subsequently, the proposed sparseHAD detect anomalies in latent space rather than original space, which also contributes to the robustness of the network against noise. Quantitative assessments and experiments over real HSIs demonstrate the unique promise of such an approach.

1 INTRODUCTION

Anomaly detection (AD) tackles the problem of exploring unknown space or material species due to the limitation of prior knowledge and samples, which has been studied in many visual processing tasks Schlegl et al. (2019), Akcay et al. (2018), Jiang et al. (2020b). Due to the rare and unbounded nature of anomalies, such a problem stimulates the re-application of representation theory, among which the anomaly detection based on sparse coding (SC) Olshausen & Field (1996) Olshausen & Field (1997) is one. SC has successfully implanted its potential in AD Cong et al. (2011) Lu et al. (2013), including the interpretability of brain neuron activity, the prevention of overfitting, and enhancement of network robustness Arora et al. (2015). The common technique is to model the normal samples so that the normal samples can be reconstructed with small errors, while the anomaly samples have large reconstruction errors Luo et al. (2019).

However, above AD in traditional color images has its inherent limitations. Detectors tend to confuse in some challenging scenarios, e.g., noise interference, background clutter, object camouflage, similar appearance of different materials Xiong et al. (2018). In these cases, however, the intrinsic material information of the anomaly and its surrounding background is distinguishable, and the most common image capable of capturing the intrinsic material information is the hyperspectral images (HSIs) Goetz et al. (1985). Although HSIs are capable of material recognition, due to the inherent factors of spectral imaging and rarity of the anomaly, how to use this information effectively for AD still faces many challenges. Especially, the challenges caused by **insufficient datasets**, **no labels**, and **noise corruption** substantially downgrade the quality of detection Ghamisi et al. (2017). The above problems lead us to develop a weakly supervised perspective in the absence of prior knowledge and propose a novel anomaly detection framework termed **Sparse** Coding-inspired Generative Adversarial Network for **H**yperspectral **A**nomaly **D**etection **(SparseHAD)**.

In our work, we first tackle the problem of **insufficient datasets** in the AD method based on deep learning. Anomaly usually has the attribute of low occurrence probability. Furthermore, in HSIs, the attribute with a dramatic difference from the background in terms of the spectral vector is further

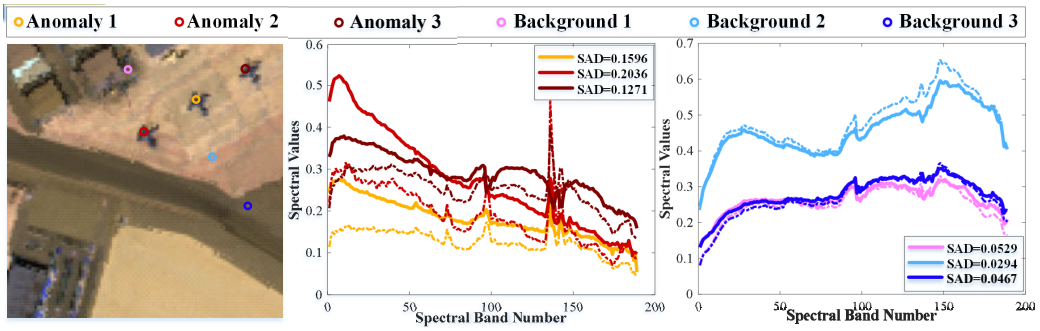

Figure 1: The spectral reconstruction results and their SAD values corresponding to different pixels obtained by a GAN model. *Left*: The locations of anomaly pixels and background pixels in the color composites. *Middle* and *Right*: The thick solid line and thin dotted line correspond to the spectral vector curves at the same location in the original HSI and the reconstructed HSI, respectively. The smaller the corresponding SAD values indicate that the spectral vectors are more similar, which also means that the reconstruction errors are smaller.

attached, making it reasonable to extract a spectral vector corresponding to a spatial pixel as a sample for AD. At this point, a $100 \times 100 \times 200$ HSI would have 100×100 samples with 200 dimensions.

Then we move on to the second problem, **no labeling**. Relying on the notion that the initially given labels of weakly supervised learning (WSL) are not always ground-true Zhou (2018), we introduce the WSL. Thus, a background-category searching step is designed to eliminate potential anomaly samples and form a coarse background sample set. The embedding of this step not only breaks the limitation of the sample label of unsupervised learning but also matches the efficient detection performance of supervised learning Ergen & Kozat (2019). Subsequently, the coarse background samples can be regarded as normal samples and then fed into the proposed weakly supervised spectral mapping model based on an end-to-end GAN. This model is expected to reconstruct the background spectral vectors as far as possible in the case that the anomalous spectral vectors cannot be well performed, as shown in Figure 1.

Noise corruption is the driving force for the proposal of the sparse coding-inspired regularized network (sparseNet), but the role of sparse coding in the proposed model has been far more than making the network more robust to noise. A deeper investigation shows that sparse coding can assist the above end-to-end GAN with autoencoder (AE) to avoid learning an approximation of the identity function Zhou et al. (2020). Furthermore, it has strong interpretability, which makes our network robust both in theory and practice Luo et al. (2017). To make optimization of our attached sparseNet tractable, we recast it into the above GAN, which is a data-driven sparse coding-inspired network without the need of precomputed sparse codes. Accordingly, we model a weakly supervised spectral mapping function and finally arrived at our sparseHAD by exploring different reconstruction error calculation strategies: original space reconstruction (*Strategy 1*) and latent space reconstruction (*Strategy 2*). We prefer the latent reconstruction, which can eliminate the influence of image noise caused by degradation mechanisms, as shown in Figure 2 (Note that the stripe noise can still be seen in *Middle*, while no trace of the stripe noise can be seen in *Right*.) This illustrates that latent reconstruction makes our model highly robust to noise because the latent space makes the network pay more attention to the difference of the intrinsic attribute of the spectrum Akcay et al. (2018).

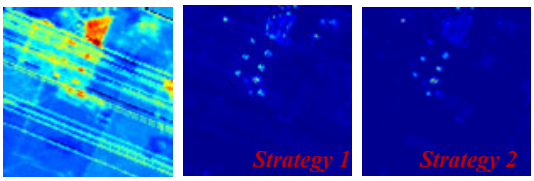

Figure 2: Taking strip noise in HSIs as an example. *Left*: The spatial map of the 190th band of the Texas-1 dataset, *Middle* and *Right*: detection maps on Texas-1.

The contributions of sparseHAD are three-fold: 1) We introduce the concept of WSL into HAD and design a novel background-category searching step to satisfy the need of WSL; 2) A novel interpretable and more robust sparesNet is developed, which can be efficiently implemented and seamlessly integrated into an end-to-end GAN. To the best of our knowledge, this is the first time

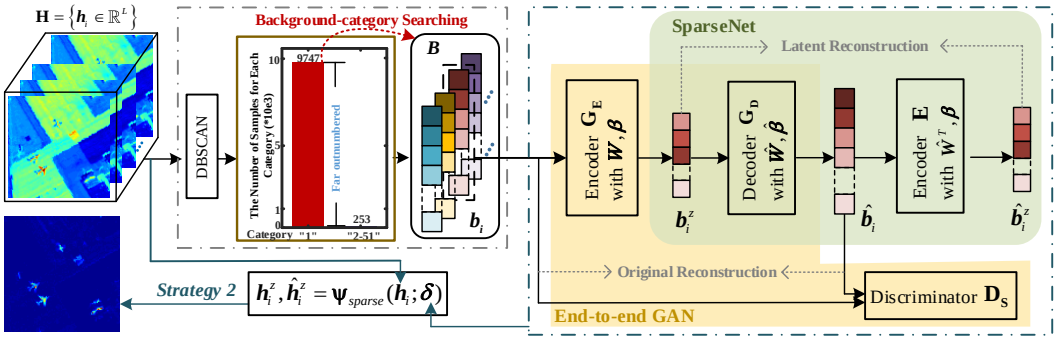

Figure 3: High-level overview of the proposed sparseHAD method in HSI.

sparse coding has been introduced into HAD; 3) The proposed framework predicts the anomalies in latent space rather than the original space. Thus, our solution may eliminate the difficulty in data annotation, make the network more robust, and avoid the dimension disaster.

2 Related Work

Considering the difficulty of obtaining labels for HAD data, most existing work on deep learning-based HAD can be roughly categorized into two-folds: semi-supervised and unsupervised. Some semi-supervised learning works, like Jiang et al. (2020a) Xie et al. (2020b), exposes a common weakness that they are unable to acquire the expected background labels that are always ground-truth. As for the unsupervised approach, stacked AE Hinton & Zemel (1994) and adversarial AE are used to extract in-depth features in the latent layer Zhao & Zhang (2018) Xie et al. (2019). Hosseiny & Shah-Hosseini (2020) realize nonlinear feature extraction by the convolution stacked AE network. MC-AEN Lu et al. (2020) combines global and local reconstruction errors of AE to achieve detection. HADGAN Jiang et al. (2020b) first employ GAN Goodfellow et al. (2014) to model background in HAD. Li et al. (2017) uses a transferred convolutional neural network (CNN) to capture the difference between pixel pairs, while NDDLR Song et al. (2019) employs a CNN to extract the abundance maps. The above methods can produce good detections, yet, the absence of labels will eventually limit the detection performance.

3 Proposed Method

Problem Definition: This work proposes a weakly-supervised HAD framework. It learns a discriminative latent reconstruction with small errors for background samples and large errors for anomaly samples, which means the latent differential image (i.e., latent reconstruction errors) can highlight the anomaly instances and restrain the background instances. Such a framework could be constructed by recasting sparse coding as a novel regularized AE unit to GAN, termed sparseHAD. As shown in Figure 3, it consists of three modules: **data preparation**, **adversarial learning**, and **inference**. Mathematically, we define and formulate our problem as follows:

An 3D cube HSI $\mathbf{H} = \left\{\boldsymbol{h_i} \in \mathbb{R}^L | i = (i_1, i_2) = N \cdot (i_1 - 1) + i_2\right\}_{i_1=1, i_2=1}^{i_1=M, i_2=N}$ with L dimensions can be reshaped into a 2D matrix $\boldsymbol{H} = \left\{\boldsymbol{h_i} \in \mathbb{R}^L\right\}_{i=1}^{M \times N}$, where (i_1, i_2) represents the position coordinates of the ith spectral vector in the spatial domain, M and N are the height and width in the spatial dimension of HSI, respectively. Besides, a coarse background samples set $\boldsymbol{B} = \left\{\boldsymbol{b_i} \in \mathbb{R}^L\right\}_{i=1}^{N_B}$ can be obtained by discarding the potential anomaly samples in set $\boldsymbol{H}$, where N_B is the number of coarse background samples. Consider the introduction of sparse coding, we need four symbols, $\boldsymbol{H}$, $\boldsymbol{H}^Z$, $\hat{\boldsymbol{H}}$, and $\hat{\boldsymbol{H}}^Z$, where $\boldsymbol{H}$ and $\hat{\boldsymbol{H}}$ belong to the original HSI space, and $\boldsymbol{H}^Z$ and $\hat{\boldsymbol{H}}^Z$ belong to the the latent HSI space. Meanwhile, $\hat{\boldsymbol{H}}$ is the reconstruction of $\boldsymbol{H}$, and $\boldsymbol{H}^Z$ and $\hat{\boldsymbol{H}}^Z$ are latent mappings of $\boldsymbol{H}$ and $\hat{\boldsymbol{H}}$, respectively. Based on these notations, we are to establish a nonlinear spectral mapping model $\Psi_{sparse}(\cdot)$ from the original HSI space to the latent HSI space as follows:

$$\boldsymbol{h}_i^z, \hat{\boldsymbol{h}}_i, \hat{\boldsymbol{h}}_i^z = \Psi_{sparse}\left(\boldsymbol{h}_i; \boldsymbol{B} | \boldsymbol{B} \subset \boldsymbol{H}\right) \tag{1}$$

where $\boldsymbol{h}_i^z \subset \boldsymbol{H}^Z \in \mathbb{R}^{(M\times N)\times L_Z}$, $\hat{\boldsymbol{h}}_i \subset \hat{\boldsymbol{H}} \in \mathbb{R}^{(M\times N)\times L}$, $\hat{\boldsymbol{h}}_i^z \subset \hat{\boldsymbol{H}}^Z \in \mathbb{R}^{(M\times N)\times L_Z}$, and $L_Z(\ll L)$ refers to the dimension of the latent space. The learning objective of the model $\Psi_{sparse}(\cdot)$ is to capture the discriminative feature between the anomaly and the background, resulting in latent differential image $\mathcal{A}^{\mathcal{Z}}$ between $\boldsymbol{H}^Z$ and $\hat{\boldsymbol{H}}^Z$ more discriminative and robust than the original differential image $\mathcal{A}$ between $\boldsymbol{H}$ and $\hat{\boldsymbol{H}}$.

3.1 Data Preparation for Weakly Supervised Learning

This step, also termed background-category searching, is to prepare for weakly supervised learning and tries to predict coarse labels of the background samples in the given input. According to the fact that there is no prior knowledge in HAD, we exploit an unsupervised clustering method, named DBSCAN Ester et al. (1996) (See Appendix A for details), to obtain a category probability map $\boldsymbol{C} = \{c_i | i = (i_1, i_2)\}_{i_1=1,i_2=1}^{i_1=M,i_2=N}$ on given input HSI $\boldsymbol{H}$. For different HSIs, the ranges of values of c_i (number of categories) are quite different and large, even up to fifty. Focusing on the theory that the probability of anomaly occurrence is far lower than the background, we design a novel probability-based background searching step which starts from calculating the sample number of each category in $\boldsymbol{C}$. Experimentally, we find that the number of samples of category "1" in $\boldsymbol{C}$ is much higher than that of other categories, such as 9747 to 253. Naturally, if the probability of occurrence of background samples is high, the number of background samples will be large. Thus, we search for samples of category "1" and classify them into the background samples set $\boldsymbol{B}$ with coarse label "1", while the remaining categories are discarded as potential anomaly samples as:

$$\boldsymbol{B} = \left\{ \boldsymbol{h}_i | i \leftarrow c_{(i_1,i_2)} = 1 \right\} \tag{2}$$

For coarse background samples, they are allowed to have some category-errors, i.e., the predicted coarse labels of $\boldsymbol{B}$ are not always ground-truth. Consequently, the model $\Psi_{sparse}(\cdot)$ is learned on $\boldsymbol{B}$ in the form of weakly supervised learning.

3.2 Sparse Coding-inspired Adversarial Learning

SparseHAD Pipeline: A novel sparse coding-inspired GAN network, consisting of two encoders, a decoder, and a discriminator, is constructed to form three subnetworks, as shown in Figure 3.

To meet the requirements of reconstruction and overcome the imbalance between GAN's generator and discriminator, we construct the **first subnetwork**, namely **an end-to-end GAN with AE**, as Perera et al. (2019). The AE, consists of an encoder ($\mathbf{G_E}$) and a decoder ($\mathbf{G_D}$), behaves as the generator ($\mathbf{G}$) in GAN, and plays a role of reconstructing. We learn model $\Psi_{sparse}(\cdot)$ on $\boldsymbol{B}$ by the mapping $\boldsymbol{B} \rightarrow \boldsymbol{B^Z} = \mathbf{G_E}(\boldsymbol{W}\boldsymbol{B} + \boldsymbol{\beta})$, and then by the mapping $\boldsymbol{B}^Z \rightarrow \hat{\boldsymbol{B}} = \mathbf{G_D}\left(\hat{\boldsymbol{W}}\boldsymbol{B}^Z + \hat{\boldsymbol{\beta}}\right)$. Especially, $\boldsymbol{B}^Z = \left\{\boldsymbol{b}_i^z \in \mathbb{R}^{L_Z}\right\}_{i=1}^{N_B}$ and $\hat{\boldsymbol{B}} = \left\{\hat{\boldsymbol{b}}_i \in \mathbb{R}^{L}\right\}_{i=1}^{N_B}$ correspond to the latent space and the original space, respectively. The spectral discriminator $\mathbf{D_S}$ not only see through the $\mathbf{G}$'s tricks but also avoid blurring of the reconstructed image $\hat{\boldsymbol{B}}$ caused by AE Pidhorskyi et al. (2018).

The reconstructed encoder network E is the **second subnetwork** and attempts to compress $\hat{\boldsymbol{B}}$ into $\hat{\boldsymbol{B}}_Z$ by $\hat{\boldsymbol{B}}^Z = \mathbf{E}\left(\hat{\boldsymbol{B}}\right)$, where $\hat{\boldsymbol{B}}^Z = \left\{\hat{\boldsymbol{b}}_i^z \in \mathbb{R}^{L_Z}\right\}_{i=1}^{N_B}$. Compared to the previous methods, in which latent vectors are obtained by minimizing the distance between encoder and decoder, this subnetwork acquires latent vectors via minimizing latent distance with its parametrization. One of the reasons this subnet works this way is to mitigate the effects of noise (e.g. stripe noise in HSIs) Akcay et al. (2018), and another has to do with the third subnetwork.

The third subnetwork is **a sparse coding-inspired regularized network (SparseNet)** whose objective is to build networks from a neuroscience perspective to make them more robust Arora et al. (2015) and conducive to anomaly detection Luo et al. (2017). Meanwhile, out of concern that the end-to-end generator $\mathbf{G}$ might learn an approximation to the identity function without additional network constraints, which would make it impossible for the generator to identify the anomaly from the background Zhou et al. (2020), we strongly recommend attaching a novel SparseNet.

Mathematically, we model SC as a generative model as neuroscientists Olshausen & Field (1997) do. Given a set of features corresponding to the normal samples as $\boldsymbol{Y} = \{\boldsymbol{y}_i \in \mathbb{R}^p\}_{i=1}^m$, our goal

is to linearly reconstruct $\boldsymbol{Y}$ by an unknown coding dictionary matrix $\boldsymbol{A} \in \mathbb{R}^{p \times q}$ with small error $\boldsymbol{\varepsilon}$, i.e., $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\alpha} \in \mathbb{R}^{q \times m}$ is the unknown sparse coefficients. The final nonconvex sparse coding objective function is as follows:

$$\sum_{i=1}^{i=m} \|\boldsymbol{y_i} - \boldsymbol{A} \cdot \boldsymbol{\alpha_i}\|_2^2 + \sum_{i=1}^{i=m} \mathbf{S}\left(\boldsymbol{\alpha_i}\right) \tag{3}$$

where $\mathbf{S}(\cdot)$ is a nonlinear sparsity penalty function. In the past practice, heuristics based on alternating minimization is usually used to solve the above problem.

The most important problem is how to apply SC to the GAN model built above. Based on the understanding of SC theory and the deconstruction of Formula 3, we employ the following principles to build a sparse regularized network: 1) Eliminating the impact of pre-computing and reducing computational overhead. Most SC methods are computationally expensive because of the need to pre-compute $\boldsymbol{\alpha}$, which largely affects the performance of those methods Joey et al. (2018). To solve this problem, we decided to adopt a neural network optimization structure (GAN) instead of the original heuristics based on the alternating minimization method. 2) Learning overcomplete representations. The advantages of overcomplete representation are as follows: stronger robustness in the case of unavoidable noise, more sparsity, greater flexibility in data matching structure, and compliance with some response characteristics of primary visual cortex neurons Lewicki & Sejnowski (2000). An overcomplete representation can be constructed when the number of basis vectors on an overcomplete basis is greater than the dimension of the input and the input representation is not unique Lewicki & Sejnowski (2000), that is, dimension q must be greater than dimension p in $\boldsymbol{A}$ for Formula 3. Considering the above two points and the GAN constructed above, the most appropriate means of applying SC to our model is to regard the latent layer output (i.e., $\boldsymbol{B}^Z$) of the end-to-end GAN model as the input of the sparseNet. Then combined with encoder network $\mathbf{E}$, the forward feedback of sparse network can be expressed as: $\hat{\boldsymbol{B}} = \mathbf{G_D}\left(\boldsymbol{B}^Z\right)$ and $\hat{\boldsymbol{B}}^Z = \mathbf{E}\left(\hat{\boldsymbol{B}}\right)$.

Learning Objective: Our purpose is to fully mine the representative background information from the coarse background samples set $\boldsymbol{B}$ during the learning process. Therefore, when the anomaly samples pass through the well-learned network, there are large errors between the reconstructed anomaly samples and the input anomaly samples because the network parameters are not suitable for the anomaly samples. Moreover, if the above original reconstructed output does not retain the anomaly information well, then the latent reconstructed output obtained from the original reconstruction encoded by the subnetwork $\mathbf{E}$ will correspondingly fail to retain the valid information of anomalies Akcay et al. (2018). Such a situation would result in a large error between the latent layer and the latent reconstructed layer. To verify the efficacy of our idea, we incorporate two losses (adversarial, sparsity) into our learning objectives, each of which optimizes the corresponding subnetworks and accounts for its contribution.

The Adversarial Loss of the end-to-end GAN is to reconstruct the background samples set $\boldsymbol{B} = \{\boldsymbol{b}_i\}_{i=1}^{N_B}$. Meanwhile, to capture the distribution of background samples more effectively, L1 normalization is applied to calculate the reconstruction error Akcay et al. (2018). The adversarial loss with the reconstruction loss is reformulated as:

$$L_A = \min_{\mathbf{G}}(\alpha_d \max_{\mathbf{D_S}}(L_d) + \alpha_r L_r) \tag{4}$$

$$L_d = \mathbb{E}_{\boldsymbol{b}_i \sim p(\boldsymbol{b}_i)} \|\mathbf{D_S}(\boldsymbol{b}_i) - \mathbb{E}_{\boldsymbol{b}_i \sim p(\boldsymbol{b}_i)} \mathbf{D_S}(\hat{\boldsymbol{b}}_i)\|_2 \tag{5}$$

$$L_r = \mathbb{E}_{\boldsymbol{b}_i \sim p(\boldsymbol{b}_i)} \|\boldsymbol{b}_i - \hat{\boldsymbol{b}}_i\|_1, \boldsymbol{b}_i = \mathbf{G}(\boldsymbol{b}_i) \tag{6}$$

The Sparsity Loss is obtained by integrating subnetworks SparseNet and $\mathbf{E}$. To avoid learning the parameters of another new encoder and to reduce the computational overhead, the decoder $\mathbf{G_D}$ parameters $\hat{\boldsymbol{W}}$ are reused in $\mathbf{E}$ as $\hat{\boldsymbol{W}}^T$, just like tied weighted autoencoder. Particularly, we regard $\hat{\boldsymbol{W}}$ as the dictionary $\boldsymbol{A}$ in the sparseNet. Meanwhile, the L1 normalization is chosen as the sparse penalty term. We specify $\boldsymbol{s} = \hat{\boldsymbol{b}}_i$ for sparse coding and recast the sparse loss formula as follows:

$$L_S = \mathbb{E}_{\boldsymbol{b}_i \sim p(\boldsymbol{b}_i)} \left(\|\boldsymbol{b}_i^z - \hat{\boldsymbol{W}}^T \boldsymbol{s}\|_F^2 + \|\boldsymbol{s}\|_1 \right), \boldsymbol{b}_i^z = \mathbf{G_E}(\boldsymbol{b}_i) \tag{7}$$

The Final Objective of the model $\Psi_{sparse}(\cdot)$ gives:

$$L_F = L_A + L_S \tag{8}$$

Table 1: Details of the datasets used in the experiment.

Datasets	Scene	Resolution	Sensor	Spatial size	Bands	Anomaly material	Anomaly proportion
G1	Airport	3.4m	AVIRIS	100×100	191	Three planes	0.6%
G2	Urban	3.5m	AVIRIS	100×100	191	unknown	0.52%
S1	Urban	3.5m	AVIRIS	100×100	189	Three planes	0.57%
S2	Airport	unknown	AVIRIS	100×100	189	Three planes	1.34%

*Note that AVIRIS means the Airborne Visible Infrared Imaging Spectrometer

Model $\Psi_{sparse}(\cdot)$ is optimized by minimizing the loss function. When the multi-networks are jointly learned to a certain epoch, the parameters $\boldsymbol{\delta} = (\boldsymbol{W}, \boldsymbol{\beta}, \hat{\boldsymbol{W}}, \hat{\boldsymbol{\beta}})$ of model $\Psi_{sparse}(\cdot)$ are frozen and used to inference for anomaly detection.

3.3 Inference for Anomaly Detection

During the **Inference process**, the given $\boldsymbol{H} = \{\boldsymbol{h}_i\}_{i=1}^{M \times N}$ is fed into the well-learned model as:

$$\boldsymbol{h}_i^z, \hat{\boldsymbol{h}}_i, \hat{\boldsymbol{h}}_i^z = \Psi_{sparse}\left(\boldsymbol{h}_i; \boldsymbol{\delta}\right) \tag{9}$$

By the above inference, the model $\Psi_{sparse}(\cdot)$ learns three outputs. Due to only focusing on the learning backgrounds characteristics, this model can reconstruct the background samples well but cannot reconstruct the anomalies well. This results in small errors between the reconstructed background samples and the input background samples, and large errors between the reconstructed anomaly samples and the input anomaly samples. Aiming to finding the anomalies in the HSI, we acquire two differential images from Formula 9, which are the original differential image $\mathcal{A}$ and the latent differential image $\mathcal{A}^{\mathcal{Z}}$, respectively, as follows:

$$Strategy1: \mathcal{A} = \|\hat{\boldsymbol{h}}_i - \boldsymbol{h}_i\|_1, Strategy2: \mathcal{A}^{\mathcal{Z}} = \|\hat{\boldsymbol{h}}_i^z - \boldsymbol{h}_i^z\|_1 \tag{10}$$

where $\mathcal{A} = \left\{\boldsymbol{a}_i \in \mathbb{R}^L\right\}_{i=1}^{M \times N}$ and $\mathcal{A}^{\mathcal{Z}} = \left\{\hat{\boldsymbol{a}}_i \in \mathbb{R}^{Lz}\right\}_{i=1}^{M \times N}$. Formula 10 describes two strategies for calculating the reconstruction error corresponding to Section 1.

During the **Detection process**, the Mahalanobis distance is adopted on $\mathcal{A}^{\mathcal{Z}}$/$\mathcal{A}$ to detect anomalies. Although both $\mathcal{A}$ and $\mathcal{A}^{\mathcal{Z}}$ can highlight the anomaly instances and restrain the background instances, $\mathcal{A}^{\mathcal{Z}}$ has better robustness than $\mathcal{A}$. Consistent with this, the Mahalanobis distances in $\mathcal{A}^{\mathcal{Z}}$ will get clearer anomalies and less background interference. In subsequent experiments, we will assess the effectiveness of the above statement.

4 Experiments

4.1 Experimental Setup

The Datasets that we employed in evaluating the sparseHAD method include Gulfport Kang et al. (2017), Gainesville Kang et al. (2017), San Diego-1 Zhang et al. (2015), and San Diego-2 Xu et al. (2016), the details of which are listed in Table 1. It is worth noting that the datasets we used contain both point and structural anomalies, and are in the form of different scales.

The Evaluation Metrics we used to report experimental results are the receiver operating characteristic (ROC) curve Bradley (1997) and the area under the ROC curve (AUC) Bradley (1997), where ROC can be plotted by the true positive rate (P_d) and the false positive rate (P_f) at various thresholds (τ). Furthermore, the closer the AUC of (P_d, P_f) value is to 1, the better the detection performance. Conversely, the closer the AUC of (P_f, τ) value is to 0, the lower the probability of false detection. In anomaly detection, the main goal is to find and locate anomalies. Therefore, the AUC of (P_d, P_f) value is taken as our primary consideration standard, followed by the AUC of (P_f, τ) value.

The Baseline Methods we picked are five frequently cited as well as state-of-the-art HAD methods, including a deep learning-based method SAFL Xie et al. (2020a), four traditional-based methods CRD Li & Du (2014), LSMAD Zhang et al. (2015), FrFE Tao et al. (2019), and LSDM_MoG Li et al. (2020). For fine comparison, we reproduced the SAFL network as recommended in the paper Xie et al. (2020a) and replaced the detector with the RX detector, termed SAFL_RX.

Implementation Details and **Parameter Analysis** see Appendix B.

Table 2: Evaluation AUC scores of the model's effectiveness on different datasets.

Datasets	The AUC scores of (P_d, P_f)				The AUC scores of (P_f, τ)			
	sparseHAD_L	_O	_EL	_EO	sparseHAD_L	_O	_EL	_EO
G1	**0.9886**	0.9690	0.9842	0.9700	0.0185	0.0203	**0.0088**	0.0204
G2	**0.9792**	0.9475	0.9678	0.9474	**0.0052**	0.0322	0.0074	0.0323
S1	**0.9846**	0.7598	0.9797	0.7598	**0.0101**	0.0273	0.0103	0.0273
S2	**0.9914**	0.9467	0.9895	0.9468	0.0140	0.0322	**0.0117**	0.0322
Average	**0.9860**	0.9060	0.9803	0.9060	0.0120	0.0280	**0.0096**	0.0280

Table 3: Evaluation AUC scores of the ablation study on different datasets.

Datasets	The AUC scores of (P_d, P_f)				The AUC scores of (P_f, τ)			
	#1	#2	#3	#4	#1	#2	#3	#4
G1	**0.9886**	0.9811	0.9715	0.9520	**0.0185**	0.0387	0.0243	0.0257
G2	**0.9792**	0.9658	0.9583	0.9511	**0.0052**	0.0305	0.0298	0.0360
S1	**0.9846**	0.9632	0.9578	0.9054	**0.0101**	0.0265	0.0301	0.0390
S2	**0.9914**	0.9865	0.9800	0.9515	**0.0140**	0.0323	0.0464	0.0713
Average	**0.9860**	0.9742	0.9669	0.9400	**0.0120**	0.0320	0.0327	0.0430

4.2 Experimental Setup

Effectiveness Assessment: According to Section 3.3 ,there are two strategies to calculate the reconstruction error. As shown in Figure 2, we have visually observed that *Strategy 2* is more robust to noise than *Strategy 1*. Furthermore, *Strategy 2* is more resource-efficient, because $\mathcal{A}^{\mathcal{Z}}$'s dimension is much smaller than $\mathcal{A}$. In this section, we will prove the superiority of *Strategy 2* quantitatively. First, we take sparseHAD_O to denote the detection method corresponding to *Strategy 1*, and use sparseHAD_L to represent the detection method on *Strategy 2*. Then, to strengthen the argument, we add a new set of comparison experiments, in which $\mathbf{E}$ with the updatable parameters is used instead of $\mathbf{E}$ with the tied weighted $\mathbf{G_D}$' s parameters, which form two new comparison methods, sparseHAD_EO, and sparseHAD_EL. The performance comparisons are reported in Table 2. One could find that sparseHAD_L/sparseHAD_EL outperforms sparseHAD_O/sparseHAD_EO consistently on all evaluation metrics, which provides another piece of evidence for the superiority of *Strategy 1*. We also find that the AUC scores of (P_d, P_f) for sparseHAD_L are always better than sparseHAD_EL, which explains the advantages and rationality of using weighted AE in our network. Considering the above analysis, we decide to adopt sparseHAD_L as our sparseHAD.

Ablation Study: To justify the benefits of the proposed sparseHAD method, we conduct the following ablation studies: #1 denotes the proposed method, #2 means the end-to-end GAN without sparse coding is adopted as the core network, #3 represents that we only use AE to obtain reconstruction error, and #4 means that HSIs are directly fed into the RX detector to calculate the Mahalanobis distances without prior access to any network. The AUC scores are tabulated in Table 3. We note that the average AUC scores of (P_d, P_f) obtained for #1 is already high at 0.94003. Therefore, from this point on, even the slightest improvement is important. When the deep neural network is introduced, the performance of the system improved by an average of 2.69%. That is to say, the deep neural network is beneficial. By replacing AE with an end-to-end GAN, the performance was further improved by an average of 0.725%, which indicates that the adversarial learning has a positive influence on HAD. Finally, by adding a sparse regularized network, the performance improves further by 1.18%, indicating that the sparse regularized network is effective. By observing the AUC scores of (P_f, τ), we can also get a conclusion consistent with the AUC scores of (P_d, P_f), which once again strengthens the effectiveness and superiority of the network we designed.

Performance Comparison:

The AUC results of the competing methods and the sparseHAD method are shown in Table 4, where sparseHAD by itself delivers the best performance in terms of AUC scores of (P_d, P_f). Specifically, relative to the AUC score of (P_d, P_f), it exceeds the second-best LSDM_MoG by an average of 0.975%. Although its AUC scores of (P_f, τ) are lower than SAFL_RX, it still shows a great competitive advantage over other comparison algorithms. The corresponding visual detection maps

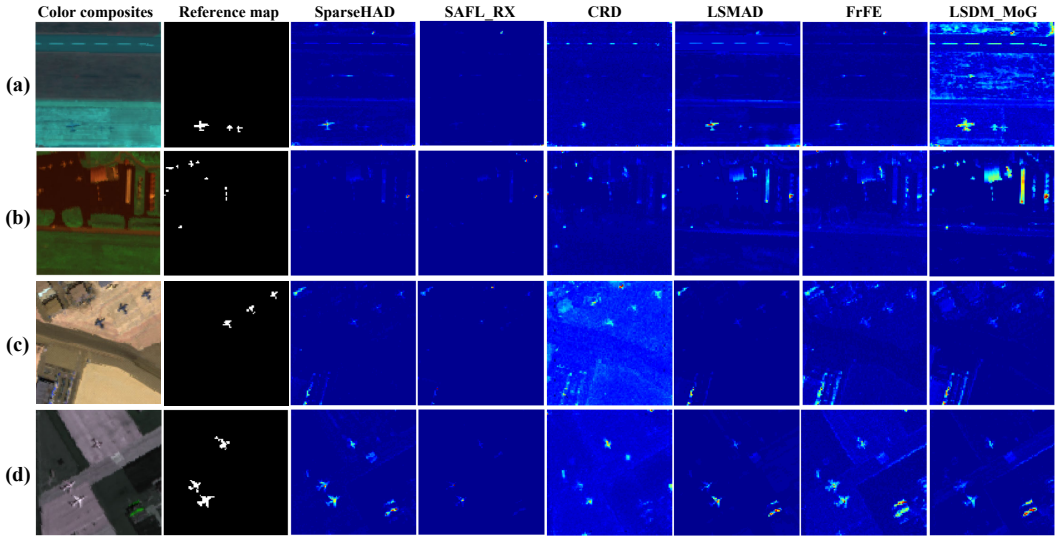

Figure 4: Pseudo-color image, binary reference maps and detection maps of the compared methods for (a) Gulfport, (b) Gainesville, (c) San Diego-1, and (d) San Diego-2.

Table 4: Evaluation AUC scores of the different network-based methods on different datasets.

Datasets	The AUC scores of $(P_d, P_f)/(P_f, \tau)$					
	sparseHAD	SFAL_RX	CRD	LSMAD	FrFE	LSDM_RMoG
G1	**0.9886**/0.0185	0.8983/0.0323	0.9239/0.0324	0.9762/0.0213	0.9828/0.1135	0.9521/0.0037
G2	**0.9792**/0.0052	0.9533/0.0216	0.9493/0.0254	0.9649/0.0356	0.9728/0.0272	0.9577/0.0039
S1	**0.9846**/0.0101	0.9655/0.1526	0.9816/0.0089	0.9534/0.0369	0.9750/0.0231	0.9789/0.0026
S2	**0.9914**/0.0140	0.9444/0.0742	0.9792/0.0124	0.9832/0.0461	0.9742/0.0177	0.9893/0.0016
Average	**0.9860**/0.0120	0.9403/0.0701	0.9585/0.0198	0.9694/0.0349	0.9762/0.0453	0.9695/0.0029

(as shown in Figure 4), supporting by the quantitative results of Table 4, reveal that the detection maps obtained by the proposed sparseHAD are closest to the ground truths. The corresponding ROC curve of (P_d, P_f) is shown in Appendix C.

Overall results verify that our method yields superior results than leading state-of-the-art methods.

5 Conclusion

In this paper, we study the AD task on HSIs from a novel weakly supervised perspective where a spectral mapping model is constructed to reconstruct background samples with small errors. The proposed sparseHAD examines the role of three subnetworks, especially the sparseNet, and reformulate sparse coding into the objective function of the other two GAN-based subnetworks. Furthermore, we dive deep into the mechanics of reconstruction error calculation in different spaces. The empirical findings in the assessment experiment provide an insight into the robustness of the proposed method. Further ablation study ensures the reliability and effectiveness of the proposed ideas and methods. Experiments of performance comparison show the sparseHAD method outperforms many state-of-the-art detection methods. The superior performance of our method exhibits the value of weakly supervised learning and sparse coding in hyperspectral tasks.

References

Samet Akcay, Amir Atapour-Abarghouei, and Toby P. Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Proc. Asian Conf. Comput. Vis.*, pp. 622–637, 2018.

Sanjeev Arora, Rong Ge, Tengyu Ma, and Ankur Moitra. Simple, efficient, and neural algorithms for sparse coding. In *Proc. Conf. Learn. Theory*, pp. 113–149, 2015.

Andrew P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recogn.*, 30(7):1145–1159, 1997.

Yang Cong, Junsong Yuan, and Ji Liu. Sparse reconstruction cost for abnormal event detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 3449–3456, 2011.

Tolga Ergen and Suleyman Serdar Kozat. Unsupervised anomaly detection with lstm neural networks. *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–15, 2019.

Martin Ester, Hans-Peter Kriegel, Jrg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. 1996 Int. Conf. Knowl. Disc. Data Min. (KDD '96)*, pp. 226–231, 1996.

Pedram Ghamisi, Naoto Yokoya, Jun Li, Wenzhi Liao, Sicong Liu, Javier Plaza, Behnood Rasti, and Antonio Plaza. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geosci. Remote Sens. Mag.*, 5(4):37–78, 2017.

Alexander F.H. Goetz, Gregg Vane, Jerry E. Solomon, and Barrett N. Rock. Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–1153, 1985.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proc. Adv. Neural Inf. Process. Syst.*, pp. 2672–2680, 2014.

Geoffrey E Hinton and Richard S Zemel. Autoencoders, minimum description length and helmholtz free energy. In *Proc. Adv. Neural Inf. Process. Syst.*, pp. 3–10, 1994.

Benyamin Hosseiny and Reza Shah-Hosseini. A hyperspectral anomaly detection framework based on segmentation and convolutional neural network algorithms. *Int. J. Remote Sens.*, 41(18):6946–6975, 2020.

Kai Jiang, Weiying Xie, Yunsong Li, Jie Lei, Gang He, and Qian Du. Semisupervised spectral learning with generative adversarial network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.*, pp. 1–13, 2020a.

Tao Jiang, Yunsong Li, Weiying Xie, and Qian Du. Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.*, pp. 1–14, 2020b.

Joey, Jiawei Du, Kai Di, Xi Peng, Hao Yang, Sinno Jialin Pan, Ivor Tsang, Yong Liu, Zheng Qin, and Rick Siow Mong Goh. Sc2net: Sparse lstms for sparse coding. In *Proc. AAAI Conf. on Artif. Intel.*, pp. 4588–4595, 2018.

Xudong Kang, Xiangping Zhang, Shutao Li, Kenli Li, Jun Li, and Jón Atli Benediktsson. Hyperspectral anomaly detection with attribute and edge-preserving filters. *IEEE Trans. Geosci. Remote Sens.*, 55(10):5600–5611, 2017.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Michael S. Lewicki and Terrence J. Sejnowski. Learning overcomplete representations. *Neural Comput.*, 12(2):337–365, 2000.

Lu Li, Wei Li, Qian Du, and Ran Tao. Low-rank and sparse decomposition with mixture of gaussian for hyperspectral anomaly detection. *IEEE Trans. Syst. Man Cybern. Syst.*, pp. 1–10, 2020.

Wei Li and Qian Du. Collaborative representation for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.*, 53(3):1463–1474, 2014.

Wei Li, Guodong Wu, and Qian Du. Transferred deep learning for anomaly detection in hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.*, 14(5):597–601, 2017.

Cewu Lu, Jianping Shi, and Jiaya Jia. Abnormal event detection at 150 fps in matlab. In *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2720–2727, 2013.

Xiaoqiang Lu, Wuxia Zhang, and Ju Huang. Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.*, 58(3):1527–1537, 2020.

Weixin Luo, Wen Liu, and Shenghua Gao. A revisit of sparse coding based anomaly detection in stacked rnn framework. In *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 341–349, 2017.

Weixin Luo, Wen Liu, Dongze Lian, Jinhui Tang, Lixin Duan, Xi Peng, and Shenghua Gao. Video anomaly detection with sparse coding inspired deep neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, (99):1–1, 2019.

Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

Bruno A. Olshausen and David J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1 ? *Vision Res.*, 37(23):3311–3325, 1997.

Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. Ocgan: One-class novelty detection using gans with constrained latent representations. In *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, pp. 2898–2906, 2019.

Stanislav Pidhorskyi, Ranya Almohsen, and Gianfranco Doretto. Generative probabilistic novelty detection with adversarial autoencoders. In *Proc. Adv. Neural Inf. Process. Syst.*, pp. 6823–6834, 2018.

Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Med. Image Anal.*, 54:30–44, 2019.

Shangzhen Song, Huixin Zhou, Yixin Yang, and Jiangluqi Song. Hyperspectral anomaly detection via convolutional neural network and low rank with density-based clustering. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 12(9):3637–3649, 2019.

Ran Tao, Xudong Zhao, Wei Li, Heng-Chao Li, and Qian Du. Hyperspectral anomaly detection by fractional fourier entropy. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 12(12): 4920–4929, 2019.

Weiying Xie, Jie Lei, Baozhu Liu, Yunsong Li, and Xiuping Jia. Spectral constraint adversarial autoencoders approach to feature representation in hyperspectral anomaly detection. *Neural Netw.*, 119:222–234, 2019.

Weiying Xie, Baozhu Liu, Yunsong Li, Jie Lei, Chein-I Chang, and Gang He. Spectral adversarial feature learning for anomaly detection in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.*, 58(4):2352–2365, 2020a.

Weiying Xie, Baozhu Liu, Yunsong Li, Jie Lei, and Qian Du. Autoencoder and adversarial-learning-based semisupervised background estimation for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.*, pp. 1–12, 2020b.

Fengchao Xiong, Jun Zhou, Xi Li, Kun Qian, and Yuntao Qian. Material based object tracking in hyperspectral videos. *IEEE Trans. Image Process.*, 2018.

Yang Xu, Zebin Wu, Jun Li, Antonio Plaza, and Zhihui Wei. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Trans. Geosci. Remote Sens.*, 54(4): 1990–2000, 2016.

Yuxiang Zhang, Bo Du, Liangpei Zhang, and Shugen Wang. A low-rank and sparse matrix decomposition-based mahalanobis distance method for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.*, 54(3):1376–1389, 2015.

Chunhui Zhao and Lili Zhang. Spectral-spatial stacked autoencoders based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection. *Infrared Phys. Technol.*, 92:166–176, 2018.

Table 5: Parameters of all methods over the four datasets

Datasets	SparseHAD	SAFL_RX	FrFE	LSDM_MoG	CRD	LSMAD
	Fixed parameters					
All	MinPts=1	As Xie et al. (2020a)	–	k=4	λ=10^{-6}	r=3,k=0.003
–	Settable parameters					
G1	ϵ=0.13	–	–	r=6	w_{out}=17,w_{in}=13	–
G2	ϵ=0.08	–	–	r=2	w_{out}=9,w_{in}=7	–
S1	ϵ=0.19	–	–	r=4	w_{out}=13,w_{in}=7	–
S2	ϵ=0.22	–	–	r=4	w_{out}=17,w_{in}=13	–

Kang Zhou, Shenghua Gao, Jun Cheng, Zaiwang Gu, Huazhu Fu, Zhi Tu, Jianlong Yang, Yitian Zhao, and Jiang Liu. Sparse-gan: Sparsity-constrained generative adversarial network for anomaly detection in retinal oct image. In *Proc. IEEE Int. Symp. Biomed. Imaging*, pp. 1227–1231, 2020.

Zhihua Zhou. A brief introduction to weakly supervised learning. *Natl. Sci. Rev.*, 5(1):44–53, 2018.

A Appendix: Preliminaries

DBSCAN Ester et al. (1996) is an unsupervised clustering method with two parameters (ϵ: the minimum distance between adjacent points, MinPts: the minimum number of points), named density-based spatial clustering of applications with noise. Given input HSI $\boldsymbol{H}$, the DBSCAN scan every pixels neighbors within ϵ distance. Once the number of neighbors of the core pixel exceeds MinPts, a cluster is formed. Subsequently, the neighbors within ϵ distance of the core pixel are assembled iteratively to obtain the category probability map $\boldsymbol{C} = \{c_i | i = (i_1, i_2)\}_{i_1=1,i_2=1}^{i_1=M,i_2=N}$.

B Appendix: Implementation Details and Parameter Analysis

Implementation Details contain learning details and parametric analysis. First, we implement our model in TensorFlow (v1.10.0, Python 3.6.0, and CUDA 10.0) by optimizing the networks using Adam Kingma & Ba (2014) with an initial learning rate 0.0001, the number of hidden nodes 20. The batch size is the same as the number of input spatial pixels, and the epoch is 3000. The weighting parameters in Formula 4 is chosen as $\boldsymbol{\alpha_d} = 0.5$ and $\boldsymbol{\alpha_r} = 50$ empirically. Both $\mathbf{G_E}$ and $\mathbf{G_D}$ dopt two fully connected network layers, which are then activated by leaky ReLU function. Empirically, we set the number of latent layer units to 20. The $\mathbf{D_S}$ has the same structure as the $\mathbf{G_D}$. $\mathbf{E}$ and $\mathbf{G_E}$ share the same structural details but different parameters

Parameter Analysis: In the data preparation stage, there are two parameters: ϵ and MinPts. Considering the pixel-level processing characteristics of the proposed method and its insensitivity to parameter ϵ, we fixed MinPts as 1. Therefore, here we focus on the parameter ϵ, which mainly controls the outputs of the category probability map. With Eps varying from 0.01 to 0.25, the optimal ϵ values for different datasets are shown in Figure 5. The AUC scores achieve the bests when ϵ is 0.13 for Gulfport, 0.08 for Gainesville, 0.19 for San Diego-1, and 0.22 for San Diego-2, respectively. For the baseline methods, we find the optimal parameters for them to achieve comparable performance. Their exact configurations are listed in Table 5.

C Appendix: ROC Curves of (P_d, P_f) for the Methods

This visualization reveals that the proposed sparseHAD usually demonstrates higher P_d as P_f varying from 10e-4 to 1 as shown in Figure 6. The conclusions are consistent with Table 3.

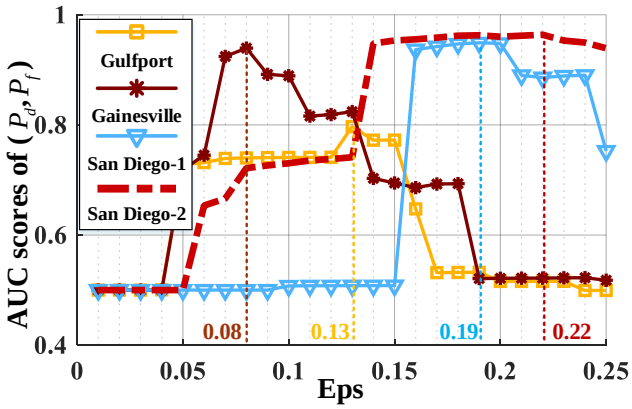

Figure 5: Effects of the parameter Eps (ϵ) over the AUC scores of (P_d, P_f) on each dataset.

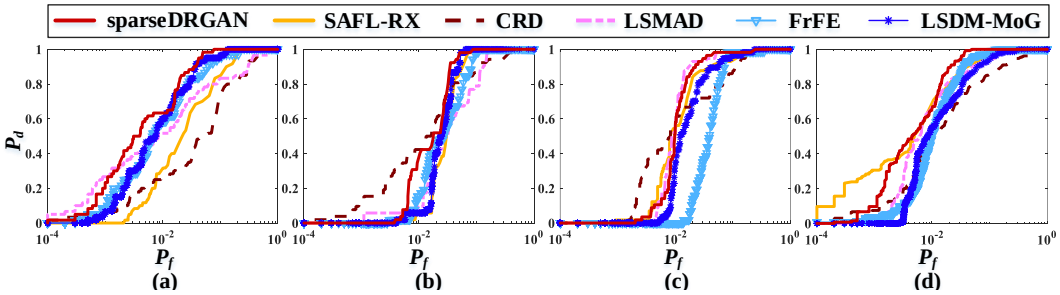

Figure 6: ROC curves of (P_d, P_f) for the methods on (a) Gulfport, (b) Gainesville, (c) San Diego-1, and (d) San Diego-2.