

RETHINKING IDENTITY MAPPING IN SELF-SUPERVISED HYPERSPECTRAL ANOMALY DETECTION: A UNIFIED PERSPECTIVE ON NETWORK OPTIMIZATION

Yongchuan Cui^{1,2}, Jinhe Zhang³, Peng Liu^{1,2,✉}, Yan Ma^{1,2}, and Yi Zeng⁴

¹*Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China*

²*School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China*

³*School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan, China*

⁴*College of Information, Beijing Forestry University, Beijing, China*

✉Corresponding author, liupeng202303@aircas.ac.cn

Abstract—The surge of deep learning has catalyzed considerable progress in self-supervised Hyperspectral Anomaly Detection (HAD). The core premise for self-supervised HAD is that anomalous pixels are inherently more challenging to reconstruct, resulting in larger errors compared to the background. However, owing to the powerful nonlinear fitting capabilities of neural networks, self-supervised models often suffer from the Identity Mapping Problem (IMP). The IMP manifests as a tendency for the model to overfit to the entire image, particularly with increasing network complexity or prolonged training iterations. Consequently, the whole image can be precisely reconstructed, and even the anomalous pixels exhibit imperceptible errors, making them difficult to detect. Despite the proposal of several models aimed at addressing the IMP-related issues, a unified descriptive framework and validation of solutions for IMP remain lacking. In this paper, we conduct an in-depth exploration to IMP, and summarize a unified framework that describes IMP from the perspective of network optimization, which encompasses three aspects: perturbation, reconstruction, and regularization. Correspondingly, we introduce three solutions: superpixel pooling and upooling for perturbation, error-adaptive convolution for reconstruction, and online background pixel mining for regularization. With extensive experiments being conducted to validate the effectiveness, it is hoped that our work will provide valuable insights and inspire further research for self-supervised HAD. Code: <https://github.com/yc-cui/Super-AD>.

Index Terms—Hyperspectral anomaly detection, identity mapping, deep learning, self-supervised neural networks

I. INTRODUCTION

Hyperspectral anomaly detection (HAD) aims to identify pixels or regions in a hyperspectral image that exhibit spectral signatures significantly different from surroundings [1–3]. Traditional methods for HAD have relied heavily on statistical approaches, such as the Reed-Xiaoli (RX) [4, 5] detectors, collaborative representation (CR) [6, 7] and low-rank representation (LRR) [8, 9]. These methods, while effective in certain scenarios, often struggle with the complexity and variability of real-world hyperspectral data, leading to suboptimal detection performance [8]. Compared to traditional methods, the advent of parameterized neural networks for self-supervised learning [10–16], has emerged as a promising approach in HAD. The core premise is that the background, comprising the majority of the image, can be approximated well by the model, while

anomalies, being spectrally distinct, cannot be accurately represented by the learned background model [8, 17]. However, self-supervised models in HAD face a significant challenge known as the Identity Mapping Problem (IMP), which has been extensively mentioned in [15, 12, 11, 13, 17, 18]. The IMP arises from the powerful nonlinear fitting capabilities of deep neural networks, which can lead to overfitting to the entire image dataset. As the complexity of the network increases or the number of training iterations grows, these models tend to reconstruct both the background and anomalies with high fidelity, resulting in imperceptible errors for anomalous pixels [19]. Despite the introduction of various models attempting to tackle IMP-related issues, a comprehensive analytical framework and a unified validation of solutions for IMP in the context of self-supervised HAD are still missing, lacking a holistic view of the problem.

In this paper, we aim to fill this gap by conducting an in-depth exploration to the IMP. We propose a unified framework that describes the IMP from the perspective of network optimization, encompassing three key aspects: perturbation, reconstruction, and regularization. Each aspect corresponds to a specific solution that we introduce. Through extensive experiments on various hyperspectral datasets, we validate the effectiveness of our proposed solutions and demonstrate how they collectively contribute to overcoming the IMP.

II. METHODOLOGY

A. Unified Perspective for Self-supervised HAD

In the context of self-supervised HAD, we address the IMP by proposing a unified framework that encompasses perturbation, reconstruction, and regularization. Given a hyperspectral image $\mathbf{X} \in \mathbb{R}^{h \times w \times c}$ containing anomalous pixels, we utilize a neural network \mathcal{F} parameterized by θ to reconstruct the image. We summarize the optimization process of a theoretically well-performing neural network using the following unified formulation:

$$\hat{\theta} = \arg \min_{\theta} \mathcal{L} \left(\mathcal{F} \left(\mathcal{P}(\mathbf{X}), \hat{\mathbf{M}}; \theta \right), \mathbf{X} \right) + \lambda \mathcal{R}(\hat{\mathbf{M}}), \quad (1)$$

where $\mathcal{L}(\cdot, \cdot)$ denotes the reconstruction loss (commonly l_1 or l_2), and $\hat{\mathbf{M}}$ is the estimated anomaly map from the previous iteration. The critical components of gaining insights into IMP include the data perturbation operation $\mathcal{P}(\cdot)$, the guided

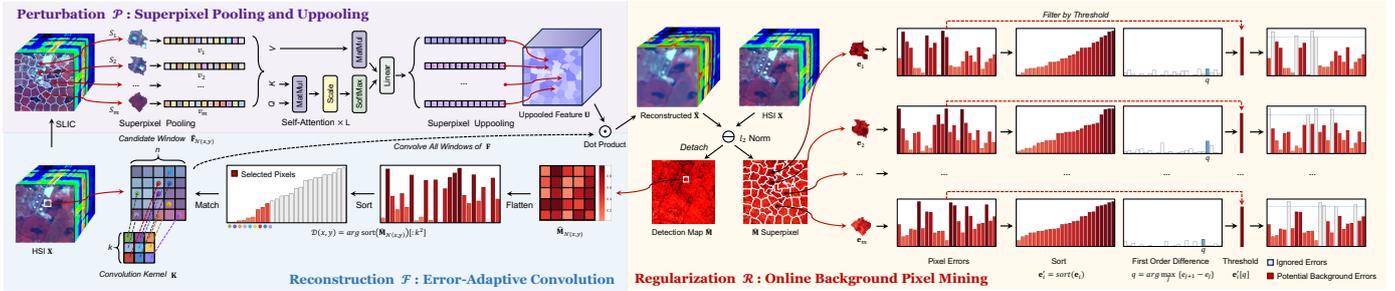


Fig. 1: Our proposed unified framework for self-supervised HAD (Zoom in for better view).

reconstruction function $\mathcal{F}(\cdot, \cdot; \theta)$, and the regularization term $\mathcal{R}(\cdot)$. λ balances the contribution of the regularization.

1) *Perturbation*: The data perturbation operation $\mathcal{P}(\cdot)$ is designed to perturb the spectral information. Applying perturbations to obscure the information of anomalous spectra before they can influence the network’s reconstruction process is a straightforward strategy to mitigate the IMP. The random masking strategy in SMCNet [14] and AETNet [20], the use of noise in AutoAD [10] and BSDM [21], *etc.*, are specific perturbation instances. The blind spot network series, BS3LNet [17], BockNet [13], PDBSNet [11], DirectNet [15], *etc.*, can actually be considered a form of perturbation, which can be viewed as applying a mask to the central pixel during convolution.

2) *Reconstruction*: The reconstruction function $\mathcal{F}(\cdot, \cdot; \theta)$ leverages the estimated anomaly map $\hat{\mathbf{M}} \in \mathbb{R}^{h \times w}$ from the previous iteration to guide the forward process of the network. Given the premise that anomalous spectra are difficult to reconstruct, the reconstruction error from the previous iteration can be treated as a confidence measure for the anomaly map. By designing a suitable weighting function for $\hat{\mathbf{M}}$, we can enhance the error associated with anomalous spectra while diminishing the error related to the background, leading to a more accurate anomaly map in subsequent iterations and thus mitigating the IMP. For instance, BiGS_eT [12] and MSNet [22] utilized the dot product to modify the reconstruction results. AutoAD [10], DeepLR [23], and S2DWMTrans [24] employed adaptive weights to alter the gradients during the backpropagation.

3) *Regularization*: This term imposes constraints on the estimated anomaly map to prevent the IMP. The weight coefficient λ balances the contributions of the reconstruction loss and the regularization term. In the optimization process of the neural network, this term is typically formulated as a loss function that imposes additional constraints on the anomaly map. For instance, BiGS_eT [12] and MSNet [22] applied the second-order Laplacian of Gaussian (LoG) operator to suppress anomalies. DeepLR [23] and RSAAE [25] applied a low-rank regularized loss to constrain the network to approximate the low-rank background. However, a common challenge in existing methods is the difficulty in determining the balance coefficient λ between reconstruction and regularization.

Although each part presents various methods, their limited consideration of the reconstruction process from a holistic perspective of network optimization results in constrained performance. In this paper, we meticulously designed these three key aspects, and experiments prove that our approach can achieve optimal results (see Fig. 1).

B. Design of the Perturbation Operation \mathcal{P}

Masking [14, 20, 17, 13, 11, 15] and noise [10, 21] cannot ensure the total elimination of anomalous spectra before sent into the network. To this end, we propose a new perturbation strategy, *i.e.*, superpixel pooling and unpooling (dubbed as SPP). Specifically, we first use Simple Linear Iterative Clustering (SLIC) [26] to segment the hyperspectral image into superpixels, and then apply average pooling to each region block to retain the average feature information. Since anomalies occupy a small proportion, they are easily wrapped in pixel blocks surrounded by the background. Due to the average pooling strategy, the block information will contain mostly background spectra while ignoring the anomalous spectra, which prevents the anomalous spectra from being reconstructed, thereby mitigating the IMP. Meanwhile, for the extracted all blocks, we use the self-attention mechanism [27, 28] to perform spectral reconstruction, learning the relationship between the blocks. Finally, all blocks will perform uppooling to revert to original size. Compared to masking [14, 20, 17, 13, 11, 15] and noise [10, 21] strategies, SPP effectively encapsulates anomalous pixels within background-dominated blocks, thereby preventing their influence on the reconstruction process.

Formally, given a hyperspectral image $\mathbf{X} \in \mathbb{R}^{h \times w \times c}$, SPP(\mathbf{X}) can be described as follows. Firstly, obtaining a series of superpixel blocks using the SLIC [26] algorithm,

$$\mathcal{S} = \text{SLIC}(\mathbf{X}), \quad (2)$$

where $\mathcal{S} = \{S_1, S_2, \dots, S_m\}$, S_i represents the i -th superpixel. Then, we apply average pooling to each superpixel to obtain the feature vectors $\mathcal{V} = \{v_1, v_2, \dots, v_m\}$. The pooling process can be expressed as,

$$v_i = \frac{1}{|S_i|} \sum_{p \in S_i} \mathbf{F}_p, \quad (3)$$

where \mathbf{F}_p denotes the feature vector of pixel p used in superpixel pooling and $|\cdot|$ is the cardinality of set (# of pixels). After forward the self-attention [27, 28], the feature vector will be restored to its original shape through uppooling,

$$\mathbf{U}(x, y) = \sum_{v_i \in \mathcal{V}} v_i \cdot 1_{p_{xy} \in S_i}, \quad (4)$$

where $\mathbf{U} \in \mathbb{R}^{h \times w \times c}$ is the upooled feature. $1_{p_{xy} \in S_i}$ is an indicator function that is 1 if p in (x, y) belongs to S_i , and 0 otherwise.

C. Design of the Reconstruction Function \mathcal{F}

Commonly, existing designs directly use the estimated anomaly map as a weight [10, 12, 22–24], which still allow anomalous pixels to affect the reconstruction process. In

contrast, we propose a novel guided reconstruction mechanism termed error-adaptive convolution (dubbed as AdaConv), which maximizes the non-utilization of anomalies. AdaConv performs dynamic convolution only on pixels that are most likely to be non-anomalous based on anomaly probability from the previous iteration.

Specifically, given a coordinate (x, y) , we get the indices of all elements of a candidate window of size $n \times n$,

$$\mathcal{N}(x, y) = \left\{ (i, j) \mid i \in \left[x - \frac{n-1}{2}, x + \frac{n-1}{2} \right], \right. \\ \left. j \in \left[y - \frac{n-1}{2}, y + \frac{n-1}{2} \right] \right\}. \quad (5)$$

For the estimation of the anomaly map obtained in the previous iteration, we sort the probabilities (or errors) ascendingly, and take the indices corresponding to the smallest top k^2 elements, where k^2 is the number of trainable parameters in the convolution kernel and $k \leq n$,

$$\mathcal{D}(x, y) = \text{argsort}(\hat{\mathbf{M}}_{\mathcal{N}(x, y)})[:k^2]. \quad (6)$$

Finally, convolve the feature map with elements taken from the corresponding indices in $\mathcal{D}(x, y)$,

$$\mathbf{F}'(x, y) = \mathbf{F}_{\mathcal{D}(x, y)} * \mathbf{K} \\ = \sum_{i=1}^k \sum_{j=1}^k \mathbf{F}(d_i, d_j) \cdot \mathbf{K}(i, j), \quad (7)$$

where \mathbf{F}' is the feature obtained by AdaConv, and \mathbf{K} is the trainable kernel with size of $k \times k$. The upooled features are reconstructed to the original image by performing a dot product with features extracted using AdaConv,

$$\hat{\mathbf{X}}^t = \mathbf{U} \odot \mathbf{F}'. \quad (8)$$

where t represents the current iteration. l_2 -norm is employed to calculate the anomaly score of pixel p ,

$$\hat{\mathbf{M}}_p^t = \left\| \hat{\mathbf{X}}_p^t - \mathbf{X}_p \right\|_2, \quad (9)$$

and the estimated detection map is used to guide the reconstruction process in the next iteration,

$$\hat{\mathbf{X}}^{t+1} = \mathcal{F}(\text{SPP}(\mathbf{X}), \hat{\mathbf{M}}^t; \theta). \quad (10)$$

D. Design of the Regularization Term \mathcal{R}

The regularization term imposes constraints on the anomaly map during the backward propagation of errors. However, determining the balance coefficient between the reconstruction and regularization terms remain challenging. We propose Online Background Pixel Mining (dubbed as OBPM) loss, which simultaneously achieves more efficient reconstruction and provides stronger constraints on anomalies. OBPM incorporates two key strategies: (1) For the reconstruction of the background, the more difficult the background is to reconstruct, the larger the gradient will be contributed. Gradient will be scaled exponentially with the reconstruction error. (2) For the regularization of anomalies, we enforce the disregard of gradients generated by potential anomalies. The two aspects ensure that the model focuses on reconstructing more complex background while avoiding the influence of anomalies that could distort the training process.

1) *Reconstructing Background*: Given the absolute background reconstruction error x , we desire that its backpropagation yields an exponentially scaled gradient,

$$g(x) = e^{\beta x} + \alpha, \quad (11)$$

here, the rate of exponential growth is determined by β , whereas α sets the minimum gradient. Thus, the reconstruction loss can be formulated as,

$$l(x) = e^{\beta x} / \beta + \alpha x. \quad (12)$$

2) *Regularizing Anomaly*: The ideal solution is to not allow anomalies to contribute any gradients, *i.e.*, discarding potential anomalies. Specifically, for superpixel S_i , the reconstruction error \mathbf{e}_i will be firstly sorted ascendingly,

$$\mathbf{e}_i' = \text{sort}(\mathbf{e}_i) = [e_1, e_2, \dots, e_{|S_i|}], \quad (13)$$

where $e_1 \leq e_2 \leq \dots \leq e_{|S_i|}$. Since the basic assumption is that the anomalies has significantly larger errors than the background, we set the index with the largest error change as the boundary,

$$q = \arg \max_j \{e_{j+1} - e_j\}, j = 1, 2, \dots, |S_i| - 1, \quad (14)$$

where $e_{j+1} - e_j$ represents the first order difference in sorted error, reflecting the magnitude of the error change. q is the index where the error changes the most. Any error greater than $\mathbf{e}_i'[q]$ will be ignored. Note that this will cause some background errors to be ignored, but since we provide exponential gradients, the remaining background can still provide enough gradients for network optimization.

Combining the reconstruction loss, the OBPM loss of an error x which belongs to S_i is expressed as follows,

$$\text{OBPM}(x \in S_i) = \begin{cases} e^{\beta x} / \beta + \alpha x, & \text{if } x \leq \mathbf{e}_i'[q] \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

III. EXPERIMENTAL RESULTS

A. Experimental Settings

We evaluate the performance of our methods using four widely recognized hyperspectral datasets: Coast, San Diego, HYDICE, and Pavia. Eight commonly-recognized models including traditional RXD [4] and CRD [6], and self-supervised methods with diverse architectures including GAED [29], MSNet [22], PDBSNet [11], PTA [30], AutoAD [10], and RGAE [31], were compared with the proposed methods. The network architecture was implemented using PyTorch. All experiments were conducted on an NVIDIA GeForce RTX 2080 Ti with 11 GB of memory. Access the source code: <https://github.com/yc-cui/Super-AD>.

B. Detection Performance Comparison

1) *Quantitative Comparison*: As shown in Table I, our model achieved state-of-the-art results in terms of Area Under the Receiver Operating Characteristic Curve (AUC) across the majority of the datasets, with a slight decrease than AutoAD [10] and PDBSNet [11] in performance on the HYDICE dataset. The comparative analysis of the Receiver Operating Characteristic (ROC) curve presented in Fig. 2 and the separability maps shown in Fig. 3 across all four datasets further validates the superior ability of the proposed model to distinguish anomalies from background compared to other models.

TABLE I: AUC values of the nine considered detectors on four datasets. The best performance is shown in **bold** and the second best is underlined.

| Model | Coast | San Diego | HYDICE | Pavia | Average |
|--------------|---------------|---------------|---------------|---------------|---------------|
| RXD [4] | 0.9906 | 0.9089 | 0.9933 | 0.9537 | 0.9616 |
| CRD [6] | 0.9910 | 0.8608 | 0.9975 | 0.9167 | 0.9415 |
| GAED [29] | 0.9779 | 0.9866 | 0.9845 | 0.9362 | 0.9713 |
| MSNet [22] | 0.9946 | <u>0.9907</u> | 0.9993 | 0.9889 | <u>0.9934</u> |
| PDBSNet [11] | <u>0.9950</u> | 0.9820 | <u>0.9996</u> | <u>0.9892</u> | 0.9914 |
| PTA [30] | 0.6992 | 0.9683 | 0.8659 | 0.9061 | 0.8599 |
| AutoAD [10] | 0.9938 | 0.9849 | 0.9998 | 0.9818 | 0.9901 |
| RGAE [31] | 0.9709 | 0.6991 | 0.7064 | 0.9053 | 0.8204 |
| Ours | 0.9982 | 0.9929 | 0.9993 | 0.9911 | 0.9954 |

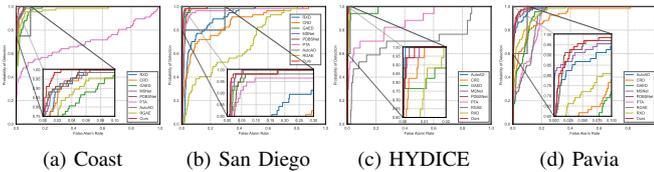


Fig. 2: ROC curves of nine considered detectors on four datasets (Zoom in for better view).

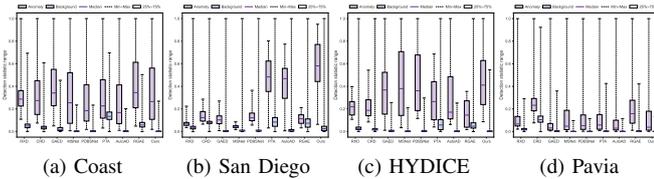


Fig. 3: Separability maps of nine considered detectors on four datasets (Zoom in for better view).

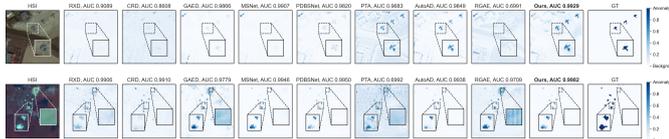


Fig. 4: Colored detection maps obtained by different methods on San Diego dataset (upper) and Coast dataset (lower). **Please zoom in for better view.**

2) *Visual Comparison*: Fig. 4 illustrates the anomaly detection maps for the San Diego and Coast datasets. Among all the evaluated models, only MSNet [22] and AutoAD [10] demonstrate competitive performance to our approach. While MSNet [22] yields impressive results on the Coast dataset, it struggles to identify anomalous pixels within the San Diego dataset. For AutoAD [10], although it also exhibited strong performance, our model assigns higher probabilities to anomalous points compared to AutoAD [10]. This clearly indicates that the proposed model effectively differentiates anomalies from the background, underscoring the efficacy of our method.

C. Ablation Study and Parameter Analysis

1) *Perturbation Operation SPP*: An ablation study was conducted to evaluate the contribution of the superpixel pooling and uppooling mechanism, as shown in Table II. The results clearly demonstrate the significant impact of SPP on the model's performance, with a noticeable increase in AUC scores

TABLE II: Ablation of perturbation operation SPP.

| | Coast | San Diego | HYDICE | Pavia | Average |
|---------|---------------|---------------|---------------|---------------|---------------|
| w/o SPP | 0.9938 | 0.9905 | 0.9960 | 0.9847 | 0.9913 |
| w/ SPP | 0.9982 | 0.9929 | 0.9993 | 0.9911 | 0.9954 |

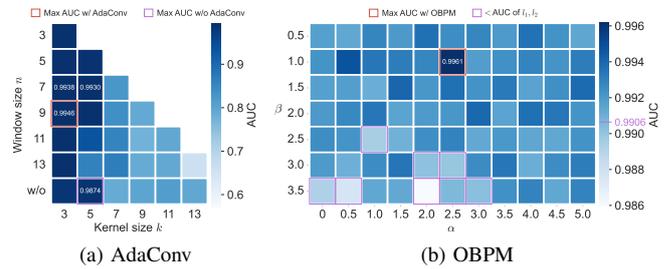


Fig. 5: Ablation studies and parameter analysis conducted on San Diego dataset for (a) AdaConv and (b) OBPM.

when SPP is incorporated. This indicates that SPP plays a crucial role in mitigating the IMP by encapsulating anomalous pixels within background-dominated blocks, thereby preventing their influence on the reconstruction process.

2) *Reconstruction Function AdaConv*: Fig. 5a present the results of ablation studies and parameter analysis on the reconstruction function AdaConv. Optimal performance is achieved with a window size of $n = 9$ and a kernel size of $k = 3$, yielding an AUC of 0.9946. We noticed AdaConv exhibits sensitivity to large kernels, such as $\{7, 9, 11, 13\}$, possibly due to the incorporation of irrelevant information by distant pixels, which diminishes the local correlation with the center pixel. As shown in the last row of Fig. 5a, without AdaConv, the optimal AUC is 0.9874. This indicates that AdaConv effectively targets non-anomalous pixels, enhancing the model's ability to reconstruct the background while disregarding anomalies, thereby preventing IMP and achieving superior results.

3) *Regularization Term OBPM*: Fig. 5b illustrates the effectiveness of the OBPM. Our method achieved an optimal AUC of 0.9961 compared to 0.9906 achieved by the commonly used l_1 or l_2 loss. The OBPM strategy is shown effective on reconstructing complex backgrounds while disregarding potential anomalies. The parameter sensitivity analysis reveals that the proposed OBPM performs consistently well within a range for $\beta \in [0.5, 2]$ and $\alpha \in [0, 5]$, indicating its robustness and stability across different parameter settings.

IV. CONCLUSION AND DISCUSSION

This paper presents a novel approach to address the identity mapping problem in self-supervised HAD, which is grounded in a unified framework that encompasses three critical aspects: perturbation, reconstruction, and regularization. Through extensive experiments on various hyperspectral datasets, we have demonstrated the effectiveness of our proposed solutions, including superpixel pooling and uppooling, error-adaptive convolution, and online background pixel mining. Our work presents a significant step forward in the field of self-supervised HAD, offering a robust and effective approach to tackle the challenges posed by the IMP. It is hoped that this paper will provide valuable insights and inspire further research for self-supervised HAD.

REFERENCES

- [1] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3d receiver operating characteristic curve analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 5131–5153, 2021.
- [2] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 1, pp. 64–90, 2022.
- [3] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.
- [4] I. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1760–1770, 1990.
- [5] H. Kwon and N. Nasrabadi, "Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, 2005.
- [6] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1463–1474, 2015.
- [7] H. Su, Z. Wu, Q. Du, and P. Du, "Hyperspectral anomaly detection using collaborative representation with outlier removal," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, pp. 5029–5038, 2018.
- [8] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 1990–2000, 2016.
- [9] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.
- [10] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "AutoAD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.
- [11] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. J. Plaza, "PDBSNet: Pixel-shuffle downsampling blind-spot reconstruction network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.
- [12] H. Liu, X. Su, X. Shen, L. Chen, and X. Zhou, "BiGSeT: Binary mask-guided separation training for dnn-based hyperspectral anomaly detection," 2023.
- [13] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. Plaza, "BockNet: Blind-block reconstruction network with a guard window for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.
- [14] Z. Wu, M. E. Paoletti, H. Su, P. Zheng, Z. Zhang, J. M. Haut, and A. Plaza, "SMCNet: Sparse-inspired masked convolutional network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–17, 2024.
- [15] D. Wang, L. Zhuang, L. Gao, X. Sun, X. Zhao, and A. Plaza, "Sliding dual-window-inspired reconstruction network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.
- [16] Z. Wu and B. Wang, "Transformer-based autoencoder framework for nonlinear hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.
- [17] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "BS3LNet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–18, 2023.
- [18] Z. Yang, R. Zhao, X. Meng, G. Yang, W. Sun, S. Zhang, and J. Li, "A multi-scale mask convolution-based blind-spot network for hyperspectral anomaly detection," *Remote Sensing*, vol. 16, no. 16, 2024.
- [19] Y. Liu, W. Xie, Y. Li, Z. Li, and Q. Du, "Dual-frequency autoencoder for anomaly detection in transformed hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.
- [20] Z. Li, Y. Wang, C. Xiao, Q. Ling, Z. Lin, and W. An, "You Only Train Once: Learning a general anomaly enhancement network with random masks for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–18, 2023.
- [21] J. Ma, W. Xie, Y. Li, and L. Fang, "BSDM: Background suppression diffusion model for hyperspectral anomaly detection," 2023.
- [22] H. Liu, X. Su, X. Shen, and X. Zhou, "MSNet: Self-supervised multiscale network with enhanced separation training for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.
- [23] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Deep low-rank prior for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.
- [24] S. Xiao, T. Zhang, Z. Xu, J. Qu, S. Hou, and W. Dong, "Anomaly detection of hyperspectral images based on transformer with spatial-spectral dual-window mask," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 1414–1426, 2023.
- [25] L. Wang, X. Wang, A. Vizziello, and P. Gamba, "RSAAE: Residual self-attention-based autoencoder for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.
- [26] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [27] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.
- [28] D. Alexey, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv: 2010.11929*, 2020.
- [29] P. Xiang, S. Ali, S. K. Jung, and H. Zhou, "Hyperspectral anomaly detection with guided autoencoder," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.
- [30] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 3, pp. 1037–1050, 2022.
- [31] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.