Background Augmentation with Transformer-Based Autoencoder for Hyperspectral Anomaly Detection

Jianing Wang, Yichen Liu, Linhao Li

Background Augmentation With Transformer-based Autoencoder for Hyperspectral Anomaly Detection

Jianing Wang[1], Yichen Liu[2], and Linhao Li[3]

[1] Xidian University *jnwang@xidian.edu.cn*
[2] Xidian University *lycc*610@163.*com*
[3] Xidian University *lhli_*7@*stu.xidian.edu.cn*

Abstract. Aiming at handling the problem caused by the lack of prior spectral knowledge of anomalous pixels for hyperspectral anomaly detection (HAD). In this paper, we propose a background augmentation with transformer-based autoencoder for hyperspectral remote sensing image anomaly detection. The representative background pixels are selected based on sparse representation for obtaining typical background pixels as training samples of the transformer-based autoencoder. The selected typical background pixels can be used for training the transformer-based autoencoder to realize background pixel reconstruction. Thereafter, the pseudo background samples can be reconstructed from the transformer-based autoencoder, which is used to subtract the original image to obtain the residual image. Finally, Reed-Xiaoli (RX) is used to detect the anomalous pixels from residual image. Experiments results demonstrate that the proposed transformer-based autoencoder which can present competitive hyperspectral image anomaly detection results than other traditional algorithms.

Keywords: Hyperspectral Remote Sensing · Transformer-based Autoencoder · Anomaly Detection · Reed-Xiaoli

1 Introduction

Hyperspectral images collect rich information from the scene by hundreds of spectral bands. The rich spectral information offers a unique diagnostic identification ability for targets of interest. As a branch of hyperspectral target location, anomaly detection has been studied since the advent of hyperspectral technology[1]. A lot of hyperspectral anomaly detection methods has been proposed in recent years.

HAD based on statistical theory are the most classical methods. As an important milestone of anomaly detection, Reed–Xiaoli (RX) algorithm is proposed by Reed and Yu [2], which characterizes the background as a multivariate Gaussian distribution and assigns anomaly scores according to the Mahalanobis distance. Among that, the local RX (LRX) detector [3] is a typically evolved version of

RX, which estimates background (BKG) by utilizing local statistics. But Gaussian distribution assumption of BKG in these methods is not fully reasonable due to the complexity of hyperspectral images. To address this issue, a collaborative representation-based detector (CRD)[4] is proposed, which can approximately represent each pixel in BKG by its spatial neighborhoods. Subsequently, the priority-based tensor approximation (PTA)[5] is proposed by introducing tensors into low-rank and sparse priors for achieving better detection performance. Recently, deep learning (DL)-based algorithms have been widely utilized in HAD area. There are mainly supervised and unsupervised two mainstreams for DL-based HAD task. As for supervised DL-based HAD, a convolutional neural network (CNN) is proposed to fully exploit the spatial correlation[6]. In [7], a deep belief network (DBN) is exploited for the feature representation and reconstruction of background samples. As for unsupervised methods, autoencoder (AE) and its similar variant structures have been applied for HAD, the residuals between the reconstructed background image and the original image are utilized to detect anomalies [8]. Thereafter, the generative adversarial networks (GAN) have also been proposed and applied in HAD task. In [9], the generator-based and discriminator-based detectors are exploited for realizing better reconstruction performance of background samples in HAD task.

According to the capability of efficient local and global feature extraction, transformer [10] has been successfully applied in natural language processing (NLP). Recently, researchers have gradually applied transformer in computer vision (CV) area. Vision transformer (ViT) [11] has achieved the state-of-the-art performance on multiple image recognition benchmarks. Meanwhile, in [12], transformer encoder-decoder architecture is proposed and explored for object detection. Therefore, in this paper, in order to mitigate the problem caused by the lack of prior spectral knowledge of anomalous pixels for HAD, we proposed a background augmentation with transformer-based autoencoder structure, which can realize better reconstruction performance by exploiting local and global attention in the spectral perspective. The pseudo background samples can be reconstructed from the transformer-based autoencoder, then the highlight anomaly pixels can be easily detected by the RX algorithm. We implement our proposed algorithm on several HAD data sets, the experiment results demonstrate the competitive performance and results than other traditional algorithms.

The rest of the paper is organized as follows. Section 2 presents the detailed process and principles of the algorithm in this paper. Section 3 performs experiments on real hyperspectral data sets. Section 4 draws some conclusions and provides hints at plausible future research lines.

2 Methodology

The overall procedure of our detection process is shown in Figure 1. The main structure mainly composed of three main parts. The typical background samples selection, transformer-based autoencoder and RX detection. The hybrid pixel selection strategy based on sparse representation is exploited to select typical

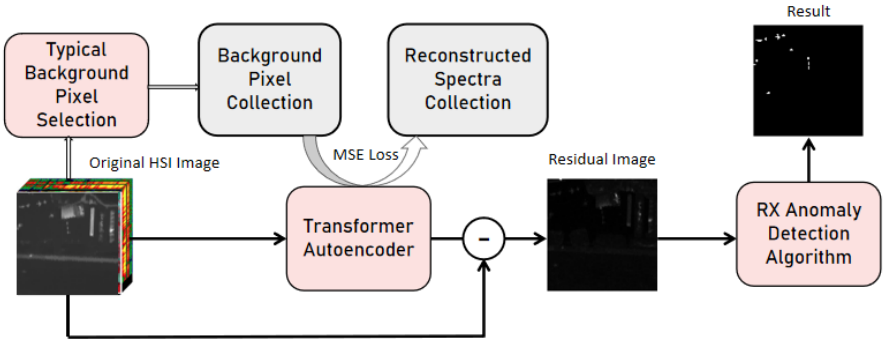


Fig. 1: Background augmentation with transformer-based autoencoder for hyperspectral anomaly detection process.

background samples, the selected samples would be fed into the transformer-based autoencoder for training and automatically reconstructing the pseudo background samples. Then, the residual image is obtained by subtracting the reconstructed hyperspectral image from the original hyperspectral image as

$$\mathbf{r} = \mathbf{s} - \mathbf{s}' \tag{1}$$

where $\mathbf{s}$ is the original hyperspectral image, $\mathbf{s}'$ is the reconstructed hyperspectral image and $\mathbf{r}$ is the residual image. Therefore, the backgrounds are augmented by suppressing the residual of background pixels in the residual image $\mathbf{r}$, since the RX algorithm can be utilized to realize more efficient anomaly detection performance.

2.1 Typical Background Samples Selection

The high-quality and representative background sample plays a key role for DL-based training procedure. Therefore, a hybrid pixel selection strategy based on sparse representation was presented. The given hyperspectral image $X \in \mathbb{R}^{H\times W\times C}$ is transformed into $X_{PCA} \in \mathbb{R}^{H\times W\times 3}$ by Principal Component Analysis (PCA) [13]. X_{PCA} aggregates into m clusters by the K-means algorithm, and each cluster consists of m_s pixels

$$\mathbf{E}^i = \left\{\mathbf{e}_1^i, \mathbf{e}_2^i, \ldots, \mathbf{e}_{m_s}^i\right\}, 1 \leqslant i \leqslant m \tag{2}$$

Inspired by sparse representation, each pixel $\mathbf{e}_s^i$, $(1 \leq s \leq m_s)$ in the ith cluster can be approximately represented as a linear combination of pixels $\bar{E}_i$ in that cluster

$$\begin{aligned} \mathbf{e}_s^i =& \overline{\mathbf{E}}_s^i \mathbf{a}_s^i, \quad 1 \leq i \leq m, 1 \leq s \leq m_s \\ & \text{s.t. } \overline{\mathbf{E}}_s^i = \left\{\mathbf{e}_1^i, \mathbf{e}_2^i, \ldots, \mathbf{e}_{s-1}^i, \mathbf{e}_{s+1}^i, \ldots, \mathbf{e}_{m_s}^i\right\} \end{aligned} \tag{3}$$

where $\mathbf{e}^i_{m_s}$ can be sparsely represented by the sparse vector $\mathbf{a}^i_s$ and the dictionary $\bar{E}^i_s$. Orthogonal Matching Pursuit (OMP) [14] is applied to access each $\mathbf{a}_i$. The sparsity of each pixel is indicated by the magnitude of $\mathbf{a}_i$ L1 normalization value. The greater the value and the less sparsity of the pixel. The less sparsity of a pixel can be more easily linearly represented by lots of other pixels so as to select lower sparse samples in each cluster as typical background pixels.

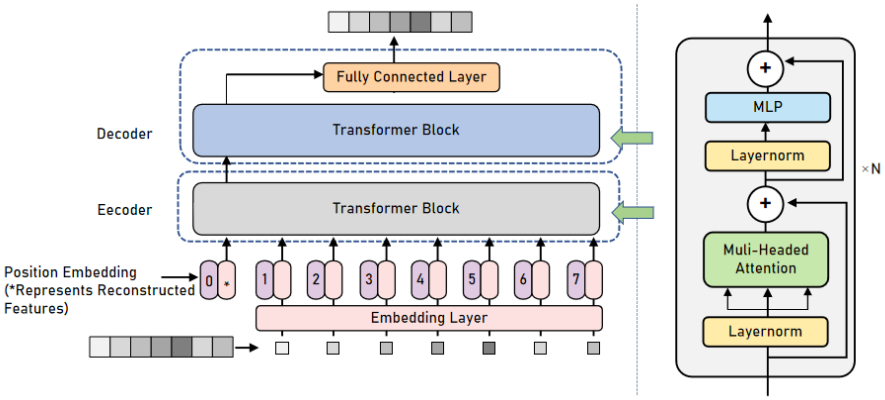


Fig. 2: Schematic diagram of transformer-based autoencoder.

2.2 Transformer-based Autoencoder

The overview model of our proposed transformer-based autoencoder as shown in Figure 2, which mainly consists of an encoder and a decoder with N transformer blocks and D dimensions latent vector in all layers. The spectral sequence of background pixel in a hyperspectral image with C bands can be represented as $\mathbf{x} = [x_1, x_2, \cdots, x_C] \in \mathbb{R}^C$, the input of multilayer perceptron $\mathbf{x}$ is mapped to D dimension, which can be concatenated with the position encoding information of each channel to form the embedded feature $\mathbf{z}_0$. Embedded features are sequentially fed into the encoder, decoder and fully connected layers to obtain the reconstructed spectral sequence $\mathbf{x}' \in \mathbb{R}^C$.

Similar to ViT [11], we add a learnable reconstruction feature $\mathbf{z}^0_0 = \mathbf{z}_{rebuild}$ at the head position of feature $\mathbf{z}_0$, which would fully interacts with features in other locations. Embedding vector generation and feature operations in transformer-based autoencoder is shown in the following formula as

$$
\begin{aligned}
&\mathbf{z}_0 = [\mathbf{z}_{rebuild}; x_1\mathbf{E}; x_2\mathbf{E}; \cdots; x_C\mathbf{E}] + \mathbf{E}_{pos}, \mathbf{E} \in \mathbb{R}^{1\times D}, \mathbf{E}_{pos} \in \mathbb{R}^{(C+1)\times D} \\
&\mathbf{z'}_n = \mathrm{MSA}(\mathrm{LN}(\mathbf{z}_{n-1})) + \mathbf{z}_{n-1}, n = 1, 2, \cdots, 2N \\
&\mathbf{z}_n = \mathrm{MLP}(\mathrm{LN}(\mathbf{z'}_n)) + \mathbf{z'}_{n-1}, n = 1, 2, \cdots, 2N \\
&\mathbf{x}' = \mathrm{MLP}(\mathrm{LN}(\mathbf{z}_{rebuild}))
\end{aligned}
\tag{4}
$$

where $\mathbf{E}$ represents the weight of the encoded fully connected layer, $\mathbf{E}_{pos}$ is the position encoding matrix, $\mathbf{z}_n$ represents the output of the nth transformer block, and $\mathbf{z'}_n$ is an intermediate variable. $\mathbf{x'}$ represents the reconstructed spectral sequence. MSA means multiple self-attention mechanism in transformer, the MLP and LN distributions represent multilayer perceptrons and layer normalization layers. The activation function used by each layer of MLP is a Gaussian Error Linear Unit (GELU), which can be formulated as

$$GELU(a) = a\Phi(a) \approx 0.5a(1 + \tanh[\sqrt{2/\pi}(a + 0.044715a^3)]) \tag{5}$$

where $\Phi(a)$ is the standard Gaussian distribution function.

Assuming that the set of background pixels $X = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M] \in \mathbb{R}^{M \times C}$ can be used as training samples, the reconstruction result of background pixel $\mathbf{x}_i$ by the transformer-based autoencoder is $\mathbf{x'}_i$, the MSE loss function can be expressed as

$$loss_{mse} = \frac{1}{M} \sum\nolimits_{i=1}^{M} (\mathbf{x}_i - \mathbf{x'}_i)^2 \tag{6}$$

2.3 RX Anomaly Detection Algorithm

The RX anomaly detection algorithm [2] performs anomaly detection by calculating the characteristics of the background and abnormal pixels, which assumes that the abnormal pixels and background pixels obey the same Gaussian distribution as

$$\begin{aligned} H_0 : \mathbf{x}_0 &\sim N(\mu, \sigma) \\ H_1 : \mathbf{x}_1 &\sim N(\mu, \sigma) \end{aligned} \tag{7}$$

where $\mathbf{x_0}$ is the background pixel, $\mathbf{x_1}$ is the abnormal pixel.

The hyperspectral image can be expressed as $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, H and W correspond to rows and columns of the hyperspectral image, the number C is the bands of hyperspectral image. While the three-dimensional matrix can be converted into a two-dimensional matrix of $\mathbb{R}^{(H \times W) \times C}$. Let $M = H \times W$, the mean and variance can be calculated as

$$\begin{aligned} \mu_b &= \frac{1}{M} \sum_{i=1}^{M} \mathbf{x}_i \\ C_b &= \frac{1}{M} \sum_{i=1}^{M} (\mathbf{x}_i - \mu_b)(\mathbf{x}_i - \mu_b)^T \end{aligned} \tag{8}$$

The value of $RX(\mathbf{x})$ can be calculated as

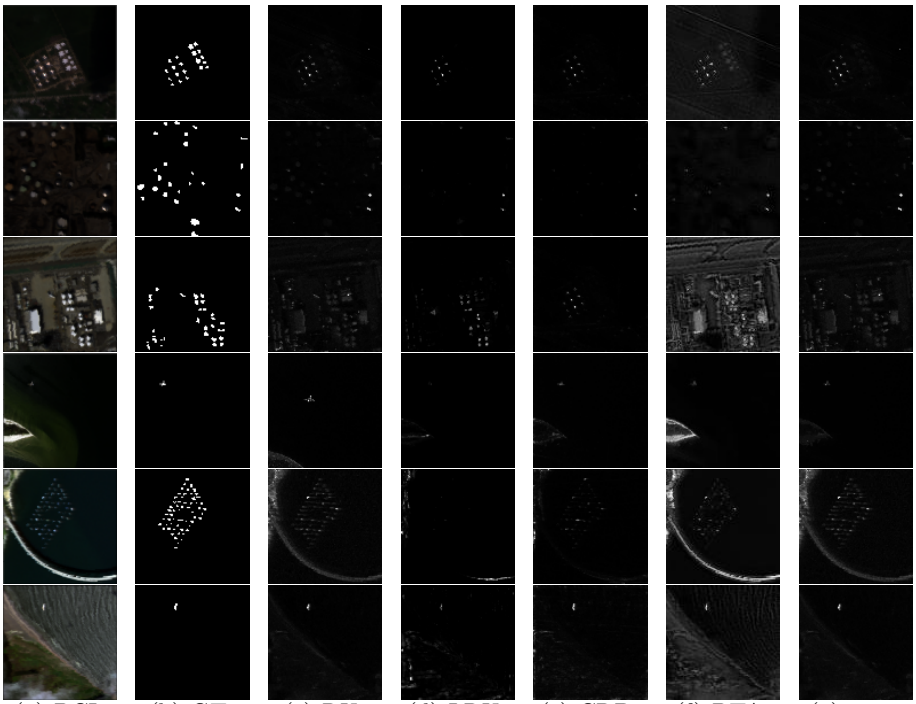
$$RX(\mathbf{x}) = (\mathbf{x} - \mu_b)^T {C_b}^{-1} (r - \mu_b) \tag{9}$$

The larger the value $RX(\mathbf{x})$ means the higher abnormal probability of the pixel.

3 Experiments and Results

3.1 Datasets and Implementation

Six different hyperspectral images obtained from the AVIRIS sensor are selected for anomaly detection test (Texas Coast, Los Angeles-1, Los Angeles-2, Cat Island, San Diego, Bay Champagne). We select 1000 background pixels as training and the number of transformer blocks is set to $N = 8$, while the dimension of transformer latent vector is set to $D = 64$. During the training process, the Adam optimization algorithm was selected as the optimizer, and the learning rate was set to 0.00001. In order to quantitatively evaluate the effect of object detection, AUC was used as evaluation metrics.



(a) PCI (b) GT (c) RX (d) LRX (e) CRD (f) PTA (g) ours

Fig. 3: Detection maps of each method. PCI indicates pseudo-color maps, GT means ground truth. The six datasets in order from top to bottom are: Texas Coast, Los Angeles-1, Los Angeles-2, Cat Island, San Diego, Bay Champagne.

3.2 Experimental Results

In this section, we mainly evaluate the proposed method with four related traditional anomaly detection algorithms: RX [2], LRX [3], CRD [4], PTA [5]. The AUC values of the detection results on the six datasets are shown in Table 1, respectively.

Table 1: AUC values of different methods on six HAD datasets.

HSIs	RX	LRX	CRD	PTA	ours
Texas Coast	0.9945	0.9370	0.8805	0.9769	**0.9949**
Los Angeles-1	0.9887	0.9489	0.9408	0.8257	**0.9893**
Los Angeles-2	0.9694	0.8883	0.9371	0.8257	**0.9696**
Cat Island	0.9660	0.9543	**0.9853**	0.9183	0.9815
San Diego	0.9103	0.8801	0.9040	0.8298	**0.9117**
Bay Champagne	0.9997	0.9394	0.9941	0.9202	**0.9998**

The experimental results demonstrate that our proposed method achieves the better performance, the transformer-based autoencoder proposed in this paper presents more stable effect in improving the anomaly detection, and strikes a satisfactory balance between high detection rate and low false positive rate. For all datasets, our method can detect most anomalies while preserving the overall object shape, the detection maps for each method are shown in Figure 3.

4 Conclusion

In this paper, a background augmentation with transformer-based autoencoder architecture is proposed to improve the effect of anomaly detection by efficiently enlarging the difference between background and abnormal pixels. Experiments demonstrate that our proposed method can be effectively applied in the field of hyperspectral remote sensing image interpretation and presents competitive anomaly detection results on several real hyperspectral datasets.

Acknowledgements Manuscript received May 10, 2022, accepted June 1, 2022. This work was supported in part by the National Natural Science Foundation of China(No.61801353), in part by The Project Supported by the China Postdoctoral Science Foundation funded project(No.2018M633474), in part by GHfund under grant number 202107020822 and 202202022633.

References

1. Su H, Wu Z, Zhang H, et al: Hyperspectral anomaly detection: A survey. IEEE Geoscience and Remote Sensing Magazine **10**(1), 64-90 (2021)

2. Reed I S, Yu X: 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 (1990)
3. Matteoli S, Veracini T, Diani M, et al: A locally adaptive background density estimator: An evolution for RX-based anomaly detectors. IEEE geoscience and remote sensing letters **11**(1), 323-327 (2013)
4. Li W, Du Q: Collaborative representation for hyperspectral anomaly detection. IEEE Transactions on geoscience and remote sensing **53**(3), 1463-1474 (2014)
5. Li L, Li W, Qu Y, et al: Prior-based tensor approximation for anomaly detection in hyperspectral imagery. IEEE Transactions on Neural Networks and Learning Systems (2020)
6. Fu X, Jia S, Zhuang L, et al: Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization. IEEE Transactions on Geoscience and Remote Sensing **59**(11), 9553-9568 (2021)
7. Ma N, Peng Y, Wang S, et al: An unsupervised deep hyperspectral anomaly detector. Sensors **18**(3), 693 (2018)
8. Lu X, Zhang W, Huang J: Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection. IEEE Transactions on Geoscience and Remote Sensing **58**(3), 1527-1537 (2019)
9. Jiang T, Li Y, Xie W, et al: Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection. IEEE Transactions on Geoscience and Remote Sensing **58**(7), 4666-4679 (2020)
10. Vaswani A, Shazeer N, Parmar N, et al: Attention is all you need. Advances in neural information processing systems **30** (2017)
11. Dosovitskiy A, Beyer L, Kolesnikov A, et al: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
12. Fang Y, Liao B, Wang X, et al: You only look at one sequence: Rethinking transformer in vision through object detection. Advances in Neural Information Processing Systems **34**, 26183-26197 (2021)
13. Abdi H, Williams L J. Principal component analysis. Wiley interdisciplinary reviews: computational statistics **2**(4), 433-459 (2010)
14. Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transactions on information theory **53**(12), 4655-4666 (2007)