Collaborative Representation Based Attention Network for Hyperspectral Anomaly Detection

Maryam Imani[1], *Senior Member, IEEE* and Daniele Cerra[2], *Member, IEEE*
[1] Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
[2]Remote Sensing Technology Institute, German Aerospace Center (DLR), Muenchener Strasse 20, Wessling, 82234, Germany

***Abstract*: The Collaborative Representation-based Detector (CRD) performs anomaly detection for hyperspectral data using a linear representation of local neighbors for background estimation, which may not fully capture the informational content and spectral variability in complex hyperspectral images with heterogenous background. To deal with this aspect, the Collaborative Representation-based Attention Network (CRAN) is introduced in this letter, providing a nonlinear representation of data samples for background estimation. Both local neighbors and global samples are used in parallel, and their outputs are fused through a cross-attention mechanism. Experimental results show a good performance of CRAN in comparison with several state-of-the-art anomaly detectors.**

***Index Terms*— collaborative representation, cross-attention, hyperspectral anomaly detection, convolutional neural network.**

I. Introduction

Hyperspectral (HS) sensors measure the reflected solar radiation in up to hundreds of contiguous and narrow spectral bands, allowing the estimation of physical properties of ground materials and their chemical composition, and the identification of specific targets [1]. Hyperspectral anomaly detection is an unsupervised task used to identify rare or unusual objects in a scene, without prior information about their spectral properties [2]. The target-to-anomaly conversion mechanism through a dummy variable trick idea is suggested in [3] to adapt the detection of a known target to the detection of an anomaly, regarded as an unknown target.

A comprehensive anomaly detection benchmark named ADBench is presented in [4]. Hyperspectral anomaly detection methods are generally divided into three groups [5]. The first one considers a statistical distribution for the background, and then performs a hypothesis test, with the Reed-Xiaoli (RX) being the most well-known method in this category [6]. RX is based on the generalized likelihood ratio test (GLRT), of which a two-step variant (2S-GLRT) is introduced in [7]. The covariance matrix is assumed to be known a priori in the first step of 2S-GLRT, while in the second step it is substituted by the maximum likelihood estimate of the covariance matrix. However, the considered statistical distribution for this category of anomaly detectors may not capture more complex distribution of the spectral information in hyperspectral data, degrading the performance. The kernel isolation forest-based anomaly detector (KIFD) assumes anomalies to be isolated from other pixels with respect to the background samples in a kernel space [8].

The second group of anomaly detectors is representation-based. These detectors do not assume any specific distribution of the background a priori, and therefore do not estimate data statistics. Detectors based on collaborative representation (CR) [9]-[10], sparse representation [11], and low-rank representation [12], belong to this group. The CR based detector (CRD) [13] assumes that a background pixel can be well approximated through a linear combination of its spatial neighbors, while this does not hold for the anomalies. While the CRD leads to a convex optimization problem, the sparse representation method adding a sparsity constraint to the coefficient vector leads to a more complex optimization problem.

Different CR-based detectors have been introduced so far. In [14], the Local Summation Unsupervised Nearest Regularized Subspace with an Outlier Removal Anomaly Detector (LSUNRSORAD) and the Local Summation Anomaly Detection based on CR and Inverse Distance Weight (LSAD-CR-IDW) are introduced. LSUNRSORAD removes outliers to mitigate the influence of anomalous targets. Moreover, it applies a local dual moving window to provide more accurate detection results through summation of the obtained residual maps. The LSAD-CR-IDW improves the CR by relying on the Euclidean distance of geometric coordinates in the regularization term. In the CRD with Background Purification and Saliency Weight (CRDBPSW) [15], the background is purified by applying an outlier removal strategy, and saliency weights are used to provide the spatial features within a local window.

A low-rank and sparse matrix decomposition-based dictionary reconstruction is introduced in [16], which uses the atom usage probability for the low-rank part, and the Euclidean distance for the sparse part. A relaxed collaborative representation method with low-rank and sparse matrix decomposition is also introduced in [17], which considers the similarity and differences between features. One effective low-rank and sparsity-matrix decomposition is Go Decomposition (GoDec), which iteratively discovers low-rank and sparse matrices according to predetermined matrix and sparsity cardinalities. An orthogonal subspace-projection version of GoDec is proposed in [18], and a tensor-based GoDec model is introduced in [19]. The effective anomaly space (EAS) concept is presented in [20] to improve the background suppression.

The mentioned CR-based detectors use a local dictionary to estimate the spectrum under test. In order to improve the detection accuracy, the global information guided CRD constructs a global dictionary, where each pixel in the local dictionary is reconstructed by using a linear combination of its entries [5]. A structure tensor selects beforehand the most suitable spectral bands for the spatial representation of the anomalous pixels.

The third group of anomaly detection methods is based on deep learning based, which allows high-level feature extraction

from both targets and background. Different networks have been proposed, such as autoencoders (AEs) [21] and generative adversarial networks [22] based on an unsupervised architecture. An autonomous anomaly detection (Auto-AD) network is introduced in [23], where the AE network reconstructs background, and anomalies are computed as the reconstruction errors. The robust graph AE (RGAE) embeds into AE a superpixel segmentation-based graph regularization term [24]. A diagonal subsampling reconstruction network is introduced in [25], performing a self-supervision training process and subsampling to provide a training data generative algorithm. Convolutional neural networks (CNNs) are utilized for anomaly detection in [26], trained using annotated instances. However, no labeled samples are available for fully unsupervised anomaly detection. To avoid using annotated samples, the CR-based Unsupervised CNN (CUCNN) utilizes the predetermined convolutional kernels selected from the CR for hierarchical feature extraction [27]. Subsequently, a residual term based on the difference of defined distances in inner and guard windows is introduced to compute the anomaly scores.

In this work, a CR-based deep neural network is proposed for hyperspectral anomaly detection, which mitigates the shortcomings of conventional CR-based detectors. In CRD, a dual window is centered around the given pixel, and the background estimation of the center is approximated through a linear representation of the local neighbors. In the proposed CR-based Attention Network (CRAN), a nonlinear representation of the background is estimated by applying convolutional layers and nonlinear activation functions. Moreover, in addition to the local neighbors, a global dictionary constructed from the clusters' centroids, obtained by applying the K-means clustering, is employed. Finally, the local and global feature maps are fused through a cross-attention mechanism to provide long-term dependencies and global correlations. The reported experiments show the superior performance of CRAN with respect to several state-of-the-art detectors on three real hyperspectral datasets.

II. Collaborative Representation-based Attention Network

A CR based detector is described in this section. As for CRD, a dual window is considered around each given pixel $\boldsymbol{y}_i \in \mathcal{R}^d; i = 1, \dots, N$ where d is the number of spectral bands and N the total number of pixels. The pixels inside the outer window having size $w_{out} \times w_{out}$ and outside the inner window having size of $w_{in} \times w_{in}$ compose a dictionary of local neighbors as $\boldsymbol{X}_i = \left[\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \dots, \boldsymbol{x}_{i,n}\right]^T \in \mathcal{R}^{n \times d}$ where $n = w_{out} \times w_{out} - w_{in} \times w_{in}$ is the number of local pixels in the local neighborhood. In a CR-based method, the aim is to estimate the central pixel $\boldsymbol{y}_i$ using its local neighbors $\boldsymbol{X}_i$. In CRD, $\boldsymbol{y}_i$ is estimated as a linear combination of local neighbors as $\widehat{\boldsymbol{y}}_i = \boldsymbol{X}_i \boldsymbol{\alpha}_i$ where the coefficient vector $\boldsymbol{\alpha}_i$ is estimated by solving a convex optimization problem. Nevertheless, in a hyperspectral image with complex and heterogenous background, the center $\boldsymbol{y}_i$ may not be appropriately estimated by using such linear combination of local neighbors.

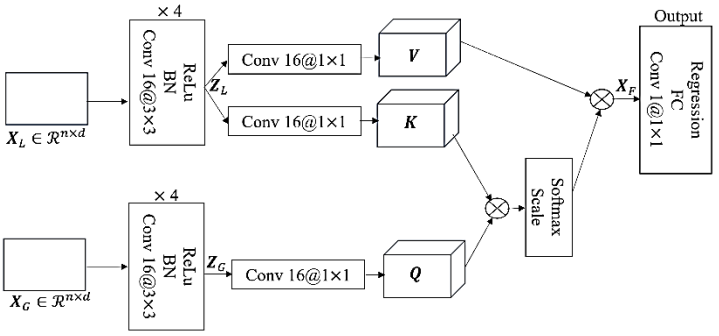

Fig. 1. Block diagram of the proposed CRAN detector.

To deal with this problem, the proposed CR-based Attention Network (CRAN) introduces two main solutions. First, in addition to local neighbors, the global information of the background is considered. Second, instead of a linear one, a nonlinear combination of dictionary spectra is provided by applying the convolutional blocks. The local and global features are finally fused by applying a cross- attention mechanism. Fig. 1 shows the block diagram of the proposed CRAN detector.

Associated with each pixel under test $\boldsymbol{y}_i \in \mathcal{R}^d$, two dictionaries are considered. The first is $\boldsymbol{X}_{L,i} = \left[\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \dots, \boldsymbol{x}_{i,n}\right]^T \in \mathcal{R}^{n \times d}$, which is composed by local neighbors in the dual window, and is the same as the one used in CRD. The second one is $\boldsymbol{X}_G = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m]^T \in \mathcal{R}^{m \times d}$, which is composed by m background pixels globally selected from the whole scene. To compose the global background dictionary, the K-means clustering is applied to the hyperspectral image and m cluster centers are selected to form $\boldsymbol{X}_G$ (as explained in eq. (4), $m = n$ is considered). The global dictionary $\boldsymbol{X}_G$ is therefore the same for all pixels. For simplicity, the subindex i is removed in Fig. 1 and following equations. Both local and global dictionaries are processed by applying 16 convolutional kernels with spatial size of 3×3 (conv16@3×3) and the "same padding" settings, followed by batch normalization (BN) and ReLu as nonlinear activation function. This convolutional block is applied sequentially four times. The outputs of the local and global branches are denoted by $\boldsymbol{Z}_L \in \mathcal{R}^{n \times d \times 16}$ and $\boldsymbol{Z}_G \in \mathcal{R}^{m \times d \times 16}$, respectively.

Subsequently, the processed local and global dictionaries are fused. To this end, a cross-attention mechanism is used to compute the long-term dependencies among local and global features. The value ($\boldsymbol{V} \in \mathcal{R}^{n \times d \times 16}$) and key ($\boldsymbol{K} \in \mathcal{R}^{n \times d \times 16}$) components are obtained from the local branch by applying 16 convolutional kernels with a size of 1×1. Similarly, the query ($\boldsymbol{Q} \in \mathcal{R}^{m \times d \times 16}$) component is obtained from the global branch as follows:

$$\boldsymbol{V} = \boldsymbol{Z}_L \boldsymbol{W}_V + \boldsymbol{b}_V \quad (1)$$

$$\boldsymbol{K} = \boldsymbol{Z}_L \boldsymbol{W}_K + \boldsymbol{b}_K \quad (2)$$

$$\boldsymbol{Q} = \boldsymbol{Z}_G \boldsymbol{W}_Q + \boldsymbol{b}_Q \quad (3)$$

where $\boldsymbol{W}_V, \boldsymbol{W}_K, \boldsymbol{W}_Q \in \mathcal{R}^{16 \times 16}$ and $\boldsymbol{b}_V, \boldsymbol{b}_K, \boldsymbol{b}_Q \in \mathcal{R}^{16}$ are the learnable weights and biases, respectively. Finally, the obtained components are fused together as follows:

Table I. The AUC values in the Ablation study.

	CRN	CRFN	CRAN
Airport	0.9853	0.8912	0.9962
Beach	0.9173	0.7127	0.9956
Urban	0.9715	0.9723	0.9835

$$\boldsymbol{X}_F = \boldsymbol{V} \otimes \text{softmax}\left(\frac{\boldsymbol{Q} \otimes \boldsymbol{K}}{\sqrt{p}}\right) \quad (4)$$

where $\otimes$ denotes elementwise multiplication, and p is the dimensionality of the feature maps $\boldsymbol{Q}$ and $\boldsymbol{K}$ (here $p = 16$). Note that to allow elementwise multiplication among $\boldsymbol{Q}$ and $\boldsymbol{K}$, their sizes must be the same, so we consider $m = n$. This means that the sizes of local and global dictionaries are considered to be the same, and that no additional parameter must be chosen to initialize the K-means algorithm. The multiplication of $\boldsymbol{Q}$ and $\boldsymbol{K}$ is scaled by $\sqrt{p}$, and its normalized value obtained by the softmax operator is multiplied in the value component to provide the fused feature maps denoted by $\boldsymbol{X}_F$.

Finally, a convolutional filter with size of 1×1 and fully connected (FC) layer is used for feature reduction, with a $d \times 1$ feature vector obtained as output of the regression layer estimates the central pixel $\hat{\boldsymbol{y}} \in \mathcal{R}^d$. In conventional CRD the residual term $r = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|$ is used in the detector's output, while the proposed CRAN obtains the detection output as:

$$r = 1 - \left|\frac{\boldsymbol{y} \cdot \hat{\boldsymbol{y}}}{\|\boldsymbol{y}\| \|\hat{\boldsymbol{y}}\|}\right| \quad (5)$$

where $\cdot$ is the inner product operator and r the anomaly score. The term subtracted from 1 in Eq. (5) is the normalized correlation among the original pixel $\boldsymbol{y}$ and its background estimation $\hat{\boldsymbol{y}}$. The probability of $\boldsymbol{y}$ to be a background pixel is considered proportional to its similarity to $\hat{\boldsymbol{y}}$.

III. Experiments

A) Datasets and parameter settings

Three real hyperspectral images acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) from the Airport-Beach-Urban (ABU) datasets are used in this section. The Airport image acquired over Gulfport has 191 spectral bands, spatial dimensions of 100×100, and a spatial resolution of 3.4m. The Beach image acquired over Cat Island contains 188 spectral bands, spatial dimensions of 150×150, and a spatial resolution of 17.2m. Finally, the Urban image acquired over Gainesville has 191 spectral bands, spatial dimensions of 100×100, and a spatial resolution of 3.5m.

The only free parameters, which should be set for the proposed CRAN method, are the size of the inner window w_{in} and the size of the outer window w_{out}. The appropriate window sizes are selected from the sets $w_{in} = \{3,5,7,9\}$ and $w_{out} = \{5,7,9,11\}$. Larger window sizes are not considered due to their impact on computational resources needed. All experiments are implemented on a laptop with Core i5 and 32G RAM using MATLAB R2022b.

For training, 25% of pixels of each dataset are randomly selected, while the remaining pixels are used for testing. As an unsupervised task, the class labels are not considered during training. For a given image element, the related local and global dictionaries are fed as input to the network, and the feature vector is considered as output. In the testing phase, the anomaly score is estimated according to (5). The mean-squared-error (MSE) loss function is used for the regression task. An Adam optimizer with a learning rate of 0.001, mini-batch size of 128, and 5 to 20 epochs is used for the network training.

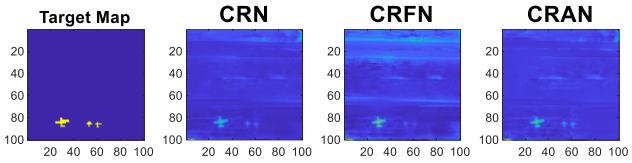

Fig. 2. Detection maps in the Ablation study for Airport dataset.

B) Ablation study

To assess performance of each of local and global dictionaries in the final results, and also to estimate the impact of the cross-attention for feature fusion, the three following cases are compared:

- CR-based Network (CRN): in this case, just the local dictionary is used as input of the network for the background estimation of the given pixel. The convolutional block consists here of conv16@3×3, BN and ReLu is applied to the input $\boldsymbol{X}_L$ four times. Finally, the extracted features $\boldsymbol{Z}_L$ are fed into the output block for the background estimation of the central pixel.
- CR-based fusion network (CRFN): in this case, both local and global dictionaries are used to estimate the background representation of the given pixel, but instead of the cross-attention mechanism, a simple elementwise addition operator is used to add $\boldsymbol{Z}_L$ and $\boldsymbol{Z}_G$ and provide the fused feature map: $\boldsymbol{X}_F = \boldsymbol{Z}_L \oplus \boldsymbol{Z}_G$.
- CR-based Attention Network (CRAN): the main proposed network.

The area under curve (AUC) of the receiver operating characteristic (ROC) curve obtained for all datasets in the ablation study is reported in Table I. For Airport and Beach datasets, the first to third rank belong to CRAN, CRN and CRFN, respectively. As expected CRAN provides the best detection results. In spite of CRN just using local information, while CRFN uses also global one, CRN is superior to CRFN with a significant difference. This suggests that using the addition operation for fusion of the processed local and global feature maps may degrade the discriminative information, and should not be used to integrate local and global feature maps. The detection maps obtained by the three configurations are shown for the Airport dataset in Fig. 2. Here, the number of false positives detected by CRFN is much higher than for CRAN and CRN.

C) Comparison with other detectors

The proposed algorithm CRAN is compared to the methods listed in Section I. The most appropriate parameters for each competitor are selected among the ones suggested by the authors or the best ones found during our experiments. The 3D ROC curves as a function of detection probability (P_D), false alarm probability (P_F), and threshold (τ) are shown in Fig. 3. The AUC values corresponding to 2D ROC curves, i.e., (P_D, P_F), (P_D, τ), and (P_F, τ), along with the running time

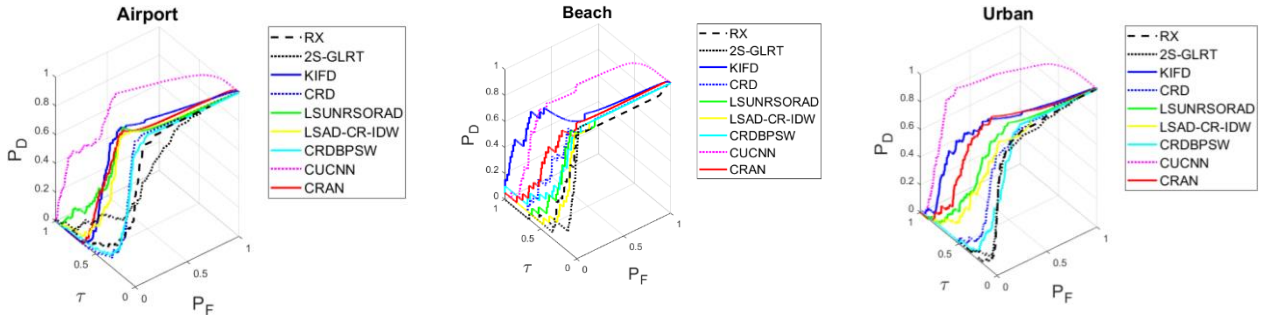

Fig. 3. 3D ROC curves of CRAN and the competitors.

Table II. The detection results for Airport dataset.

	RX	2S-GLRT	KIFD	CRD	LSUNRSORAD	LSAD-CR-IDW	CRDBPSW	CUCNN	CRAN
AUC(P_D, P_F)	0.9695	0.8898	0.9779	0.9821	0.9821	0.9927	0.9791	0.9343	**0.9962**
AUC(P_D, τ)	0.1338	0.2486	0.4210	0.1470	0.4767	0.3706	0.1540	**0.7848**	0.4080
AUC(P_F, τ)	**0.0065**	0.0189	0.1342	0.0336	0.0545	0.0534	0.0136	0.4934	0.0882
Time (seconds)	54.71	87.63	45.97	79.75	1046.51	1014.88	**21.95**	98.37	122.55

Table III. The detection results for Beach dataset.

	RX	2S-GLRT	KIFD	CRD	LSUNRSORAD	LSAD-CR-IDW	CRDBPSW	CUCNN	CRAN
AUC(P_D, P_F)	0.9514	0.9954	0.9897	0.9892	0.9880	0.9913	0.9936	0.9694	**0.9956**
AUC(P_D, τ)	0.2092	0.1056	**0.7687**	0.3905	0.2942	0.1972	0.3388	0.6875	0.4691
AUC(P_F, τ)	**0.0003**	0.0017	0.0949	0.0092	0.0042	0.0023	0.0037	0.4550	0.0464
Time (seconds)	39.35	56.64	37.27	46.97	452.65	454.90	**12.53**	316.35	116.89

Table IV. The detection results for Urban dataset.

	RX	2S-GLRT	KIFD	CRD	LSUNRSORAD	LSAD-CR-IDW	CRDBPSW	CUCNN	CRAN
AUC(P_D, P_F)	0.9515	0.9304	0.9825	0.9336	0.9615	0.9679	0.9267	0.9806	**0.9835**
AUC(P_D, τ)	0.0377	0.0549	0.6216	0.1740	0.3763	0.3027	0.1107	**0.7985**	0.5083
AUC(P_F, τ)	**0.0008**	0.0024	0.1260	0.0488	0.0578	0.0157	0.0147	0.5942	0.1123
Time (seconds)	18.11	18.11	48.34	18.88	12.07	11.59	**2.14**	39.92	109.07

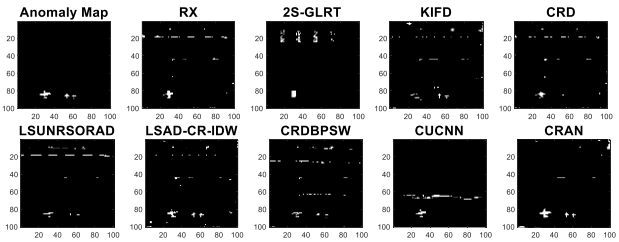

Fig. 4. Binarized detection maps with CL of 99% for Airport dataset.

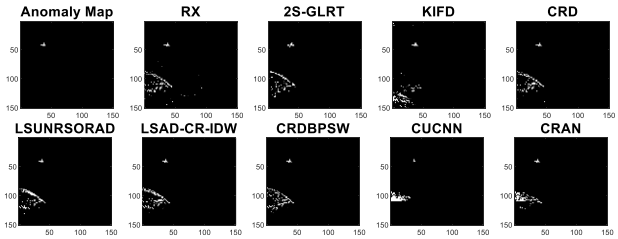

Fig. 5. Binarized detection maps with CL of 99% for Beach dataset.

required to predict the detection maps, are reported in Tables II to IV, for Airport, Beach and Urban datasets, respectively where the best values in each row are shown in bold. Larger values of AUC(P_D, P_F) and AUC(P_D, τ) denote better results, while the smaller the value of AUC(P_F, τ) the better.

In the Airport dataset, the highest AUC(P_D, P_F) values are obtained by CRAN and LSAD-CR-IDW, respectively,

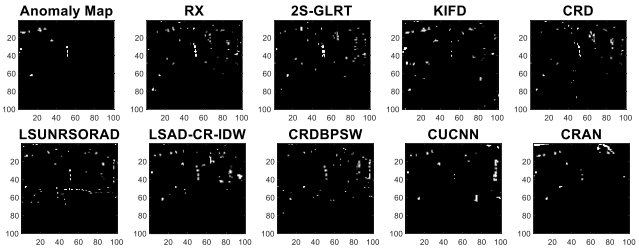

Fig. 6. Binarized detection maps with CL of 99% for Urban dataset.

followed by LSUNRSORAD and CRD. In terms of AUC(P_D, τ), CUCNN, LSUNRSORAD and CRAN rank first to third, respectively. The RX anomaly detector provides the lowest AUC (P_F, τ) value. In terms of running time, CRDBPSW is the fastest method. In contrast, LSUNRSORAD and LSAD-CR-IDW require the highest running time. The binarized detection maps having a confidence level (CL) of 99% are shown in Fig. 4 for the Airport dataset. It can be seen that CRAN results in the lowest false positive rate in the detection map.

In the Beach dataset, CRAN, 2S-GLRT and CRDBPSW are the best methods in terms of AUC(P_D, P_F). KIFD, CUCNN and CRAN rank first to third in terms of AUC(P_D, τ), respectively. Although CRAN is not among the best methods in terms of AUC (P_F, τ) , its performance is satisfactory in terms of

AUC(P_F, τ). As in the previous case, CRDBPSW is the fastest method, while LSAD-CR-IDW and LSUNRSORAD are the slowest ones. The detection maps for the Beach dataset are shown in Fig. 5.

For the Urban dataset, CRAN, KIFD and CUCNN achieve the highest AUC(P_D, P_F). In terms of AUC(P_D, τ), CUCNN, KIFD and CRAN rank first to third, respectively. Among the three mentioned methods, the lowest AUC (P_F, τ) value is obtained by CRAN, which outperforms its competitors, but also needs the highest running time. The detection maps for the Urban image are shown in Fig. 6.

Generally, in all datasets, CRAN ranks first in terms of AUC (P_D, P_F), is among three best methods in terms of AUC(P_D, τ), and provides a relatively low AUC(P_F, τ) value. The CRAN model has 1.4M learnable parameters and 266.47M FLOPs. Although CRAN needs higher computational resources with respect to some competitors such as CRD and CRDBPSW, these are still relatively low, making CRAN relatively lightweight in the frame of deep learning based anomaly detectors.

IV. Conclusion

A collaborative representation based deep neural network is proposed for hyperspectral anomaly detection in this work, providing a nonlinear representation of the background data captured in both local and global dictionaries. The obtained feature maps are then fused using a cross-attention mechanism. The ablation study shows that the addition of global information has a positive impact in detection accuracy. Moreover, the proposed network shows better detection performance compared to several benchmark methods. In the proposed method, the sizes of the global and local dictionaries are considered to be equal, thing which may fail at characterizing efficiently larger scenes with heterogeneous background samples. Future works will define a global dictionary with tunable size and design a suitable strategy for fusing local and global dictionaries having different size.

References

[1] H. Su, Z. Wu, H. Zhang and Q. Du, "Hyperspectral Anomaly Detection: A survey," in IEEE Geoscience and Remote Sensing Magazine, vol. 10, no. 1, pp. 64-90, March 2022.

[2] R. Wang and J. Hu, "Gaussian-Inspired Attention Mechanism for Hyperspectral Anomaly Detection," in IEEE Geoscience and Remote Sensing Letters, vol. 22, pp. 1-5, Art no. 5500705, 2025.

[3] C. -I. Chang, "Target-to-Anomaly Conversion for Hyperspectral Anomaly Detection," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-28, Art no. 5540428, 2022.

[4] S. Han, X. Hu, H. Huang, M. Jiang, Y. Zhao, "ADBench: Anomaly Detection Benchmark", https://arxiv.org/abs/2206.09426, 2022.

[5] M. Song, Z. Guo, L. Li, S. Liu, H. Bao and J. Li, "Global Information and Structure Tensor Guided Collaborative Representation for Anomaly Detection," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 3236-3252, 2025.

[6] M. Imani, "RX Anomaly Detector with Rectified Background," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 8, pp. 1313-1317, August 2017.

[7] J. Liu, Z. Hou, W. Li, R. Tao, D. Orlando and H. Li, "Multipixel Anomaly Detection With Unknown Patterns for Hyperspectral Imagery," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 10, pp. 5557-5567, 2022.

[8] S. Li, K. Zhang, P. Duan and X. Kang, "Hyperspectral Anomaly Detection With Kernel Isolation Forest," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 1, pp. 319-329, 2020.

[9] C. Zhao, C. Li and S. Feng, "A Spectral–Spatial Method Based on Fractional Fourier Transform and Collaborative Representation for Hyperspectral Anomaly Detection," in IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 7, pp. 1259-1263, July 2021.

[10] M. Imani, "Adaptive Window Based Collaborative Representation for Hyperspectral Anomaly Detection with Fusion of Local and Global Information," The Egyptian Journal of Remote Sensing and Space Sciences, vol. 26, no. 2, pp. 369–380, 2023.

[11] T. Guo, Y. Yang, L. He, C. Fu and F. Luo, "Anomaly Detection of Hyperspectral Image by Coarse-to-Fine Tensor Two-Level Decomposition," in IEEE Geoscience and Remote Sensing Letters, vol. 22, pp. 1-5, Art no. 5501105, 2025.

[12] Q. Yu, G. Yan, X. Li, J. Xu and X. Yang, "Graph Regularized Low-Rank Representation for Hyperspectral Anomaly Detection," 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS), Chengdu, China, 2023, pp. 1162-1165.

[13] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," IEEE Trans. Geosci. Remote Sens., vol. 53, no. 3, pp. 1463–1474, 2015.

[14] K. Tan, Z. Hou, F. Wu, Q. Du, Y. Chen, "Anomaly Detection for Hyperspectral Imagery Based on the Regularized Subspace Method and Collaborative Representation," Remote Sensing, vol. 11, no. 11, pp. 1318, 2019.

[15] Z. Hou, W. Li, R. Tao, et al., Collaborative representation with background purification and saliency weight for hyperspectral anomaly detection, Sci. China Inf. Sci., 65, 112305, 2022.

[16] Y. Xu, B. Du, L. Zhang and S. Chang, "A Low-Rank and Sparse Matrix Decomposition- Based Dictionary Reconstruction and Anomaly Extraction Framework for Hyperspectral Anomaly Detection," in *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 7, pp. 1248-1252, July 2020.

[17] H. Su, H. Zhang, Z. Wu and Q. Du, "Relaxed Collaborative Representation With Low-Rank and Sparse Matrix Decomposition for Hyperspectral Anomaly Detection," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 6826-6842, 2022.

[18] C. -I. Chang, H. Cao, S. Chen, X. Shang, C. Yu and M. Song, "Orthogonal Subspace Projection-Based Go-Decomposition Approach to Finding Low-Rank and Sparsity Matrices for Hyperspectral Anomaly Detection," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2403-2429, March 2021.

[19] M. Song, X. Zhang, L. Li, H. Cao and H. Bao, "A Tensor-Based Go Decomposition Method for Hyperspectral Anomaly Detection," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 4584-4600, 2025.

[20] C.-I. Chang, "Effective anomaly space for hyperspectral anomaly detection," IEEE Trans. Geosci. Remote Sens., vol. 60, Art. no. 5526624, Mar. 2022.

[21] A. R. Rezvanian, M. Imani and H. Ghassemian, "Patch-based Sparse and Convolutional Autoencoders for Anomaly Detection in Hyperspectral Images," 2020 28th Iranian Conference on Electrical Engineering (ICEE), Tabriz, Iran, pp. 1-5, 2020.

[22] A. Emoto and R. Matsuoka, "Unsupervised Anomaly Detection in Hyperspectral Imaging: Integrating Tensor Robust Principal Component Analysis With Autoencoding Adversarial Networks," in IEEE Access, vol. 13, pp. 21422-21433, 2025.

[23] S. Wang, X. Wang, L. Zhang and Y. Zhong, "Auto-AD: Autonomous Hyperspectral Anomaly Detection Network Based on Fully Convolutional Autoencoder," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-14, Art no. 5503314, 2022.

[24] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang and J. Ma, "Hyperspectral Anomaly Detection With Robust Graph Autoencoders," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-14, Art no. 5511314, 2022.

[25] H. Wen, X. Hu and P. Zhong, "DSRNet: Diagonal Subsampling Reconstruction Network for Hyperspectral Anomaly Detection," in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-16, 2024, Art no. 5628916.

[26] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," IEEE Geosci. Remote Sens. Lett., vol. 14, no. 5, pp. 597–601, May 2017.

[27] M. Imani, "Collaborative Representation Based Unsupervised CNN for Hyperspectral Anomaly Detection," Infrared Physics and Technology, vol. 141, 105498, Sept. 2024.