A Novel Approach to Lighten the Onboard Hyperspectral Anomaly Detector

Ning Ma, Yu Peng, Shaojun Wang(✉), and Jingyi Dong

Harbin Institute of Technology, Harbin 150080, China
{maning,pengyu,wangsj}@hit.edu.cn, jennied_jiyue@163.com

Abstract. Hyperspectral image (HSI) anomaly targets detection is always applied for timeliness and onboard mission. For high detection accuracy, deep learning based HSI anomaly detectors (ADs) are widely employed in recent researches. However, their huge network scale for high-level representation ability leads to great computation burden for the onboard computation system. To decrease the computation complexity of the detector, a lightweight network is expected for the HSI AD. In this paper, by creating a multiobjective optimization with nondominated sorting genetic algorithm II (NSGA-II), an automatic evolution based deep learning network HSI AD (Auto-EDL-AD) is proposed to explore a lightweight network. The experimental results on an HSI dataset show that the proposed Auto-EDL-AD can generate an optimal network for the HSI anomaly detection which reaches up to 170% speedup without any detection accuracy loss.

Keywords: Hyperspectral image · Deep learning · Real-time processing · Multiobjective optimization

1 Introduction

As an important remote sensing technology, hyperspectral image (HSI) can reflect the material of the ground objects with spectral information from the hundreds of narrow contiguous bands and the spatial information. It is widely used in precision agriculture, environmental protection, military mission and city plan by classification, anomaly target detection (AD) and targets recognization. In those applications, anomaly detection is widely required to found the interest targets along with the image capture.

In general, an anomaly target is defined as an object or pixel which is different from its surrounding background. The anomaly targets present in a lower probability than the background, such as a fire point in a forest. To provide an early alarm for disaster monitor, the traditional processing procedure which contains image collection, data compression, downlink transmission, data decompression,

M. Jia et al. (Eds.): WiSATS 2019, LNICST 281, pp. 432–445, 2019.
https://doi.org/10.1007/978-3-030-19156-6_41

and targets detection is too long to fit the response requirement. A simple procedure that executed the detection onboard besides with the remote image collecting become a promising approach. However, the restriction on weight, power, size and the hash space environment seriously limit the onboard computing performance. Furthermore, the sharply increased spectral number and the spatial precision brings more burdens for the detector. Therefore, it is significant to lessen an onboard HSI AD for a real-world detection mission.

With the effort from the researchers all over the world, a various of HSI ADs have been proposed, such as the Reed-Xiaoli detector (RXD) [13], low probability detection [1] and random-selection-based anomaly detector (RSAD) [5], which is based on the whole HSI scene, namely global based detector. While to reduce the processing delay for a quick response, locally based detectors are studied which employ a small window slid on the HSI to detect the targets after the data collection, such as local based RXD (LRXD). For local based detector, the output delay is approximately equal to the time of capturing a certain number lines of pixels (depend on the height of the local window). To reduce such delay for real-time applications, a series of various RXD, such as progressive band processing of anomaly detection (PBP-AD) [2], real-time causal RXD (RT-CK-RXD) [3]. are proposed. By using a half window to replace the local window, the processing delay can be decreased to several milliseconds. However, those detectors have been based an assumption that the HSI background follows the Gaussian distribution while this assumption cannot be always right for real-world HSI and leads to false alarm.

To get a higher detection accuracy, the sparse representation theory based detectors are proposed, such as collaborative representation based anomaly detector (CRD) [8], sparse representation-based detector (SRD) [9]. In CRD, a dictionary is first set up by the pixels in a local window to represent the under test pixels which is in the central of the local window. A normal pixel can be represented with a sparse vector in a small residual, but an anomaly pixel cannot be. So, the anomaly pixel can be identified.

With the development of machine learning, especially the high-level representation ability of the deep learning, more deep learning based HSI ADs are proposed for high detection accuracy, such as stacked denoising autoencoder (SDAE) anomaly detection [16], Deep Belief Network (DBN) HSI AD [11], Weight based SAE HSI AD [12] and Transferred Deep Learning based HSI AD [10]. According to the fact that the anomaly targets are in low appearance probability, in the training stage, the anomaly pixels contribute less to optimize the parameters of the network than the background pixels. Therefore, during the detection, the residual of the anomaly pixel is greater than the residual of the background pixel. With the deep network structure and the abundant non-linear function of the neurons, the features can be learned well and a better detection accuracy can be reached. However, the large scale of the network leads to a huge computation which sets an enormous challenge for real-time onboard detection in a satellite.

To balance the conflicts between the detection accuracy and the computation amount, a minimum scale network which performs high detection accuracy

is expected for the onboard HSI AD. In this paper, a multiobjective optimization is proposed to find the probable solution for the above problem with the nondominated sorting genetic algorithm II. For deep learning based online HSI AD, an automatic deep learning network constructing method (Auto-EDL-AD) is proposed to generate the hyperparameters and model structure by a nondominated sorting genetic algorithm II. With the proposed method, a small scale of deep learning model is built to accelerate the detection without accuracy loss for onboard HSI AD.

The remainder of the paper is arranged as follow. A common deep learning based HSI AD is introduced briefly in Sect. 2. In Sect. 3, the details of the proposed approach are stated. A real HSI data set is employed to validate the proposed approach in Sect. 4.

2 Deep Learning Based Online HSI AD

Recently, deep learning based methods have performed as a promising technology for hyperspectral image anomaly detection. A widely studied network for HSI AD is the multiplayer stacked auto-encoder which learns the high-level features of the HSI by its abundant non-linear functions. After the model training, the features of an HSI pixel can be extracted and represented by the middle layer (code layer) of the network. Due to the anomaly pixel performs less contribution to build the network, the residual of the anomaly pixel is higher than that of the background pixel. Such residual can be employed to mitigate the local contamination. By calculating the distance between the under test pixels and its surrounding pixels, the anomaly score can be figured out to determine the anomaly pixels. In general, a deep learning based HSI AD can be described as in Fig. 1.

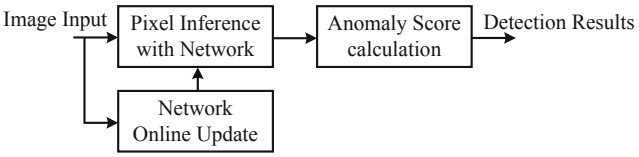


Fig. 1. The structure of a common deep learning based HSI AD

An autoencoder network is stacked by several layers as in Fig. 2 in a symmetrical structure. The layers are connected together by the connection weights. Through mapping the output as equal as possible with the input data during the training, the network can be learned the features of HSI in an unsupervised way.

As shown in Fig. 2, the layers are constructed by different number of neurons which deal the inputs with a non-linear activation function. In general, the sigmoid function, Rectified Linear Unit (ReLu) function or Leaky Relu function

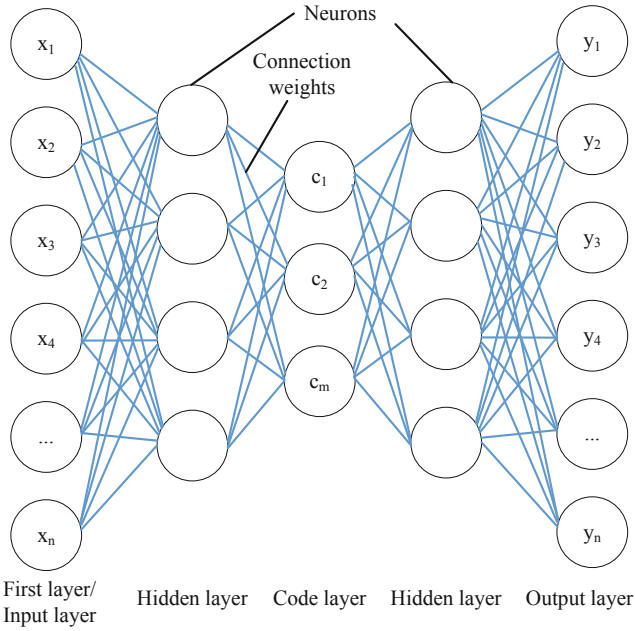


Fig. 2. The structure of a SAE network

are selected as the activation function. For onboard HSI AD mission, considering the computer resources limitation, the Leaky ReLu function is selected which is described as in formula 1. Its derivative is described as in formula 2.

$$f(z) = \begin{cases} z & when\ z > 0; \\ \alpha \cdot z & when\ z \leq 0; \end{cases} \tag{1}$$

$$f'(z) = \begin{cases} 1 & when\ z > 0; \\ \alpha & when\ z \leq 0; \end{cases} \tag{2}$$

where z is the input of each neuron. α is the leaky value.

For HSI AD application, the neurons number of the first layer and the output layer are equal to the spectral band number of the HSI. The output of neurons in the $(l+1)$-th layer is represented as $a^{(l+1)}$ or $h_{W,b}(x)$ in formulas 3a and 3b.

$$z^{(l+1)} = W^{(l)} a^{(l)} + b^{(l)} \tag{3a}$$

$$h_{W,b}(x) = a^{(l+1)} = f(z^{(l+1)}) \tag{3b}$$

where $W^{(l)}$ and $b^{(l)}$ denote the connection weight and the bias of the l-th layer respectively. Both of them are updated during the training.

To get the parameters W and b, a model training is executed by minimizing the loss function as formula 4 with gradient descent [7].

$$J(W,b) = \frac{1}{2m}\sum_{i=1}^{m}\left\|h_{W,b}(x^{(i)}) - y^{(i)}\right\|^2 + \frac{\lambda}{2}\sum_{l=1}^{n_{l-1}}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{l}\right)^2 \tag{4}$$

where λ is weight decay parameter, m is the number of training samples, y denotes the decoding output of the network, n_l is the total layer numbers of the network, s_l is the neuron number of l-th layer.

To determine the anomaly pixels, the distance between the under test pixel (PUT) and its surrounding pixels are calculated with the output of the neurons in the middle layer. The location relationship of the under test pixel and its surrounding pixels are shown in Fig. 3.

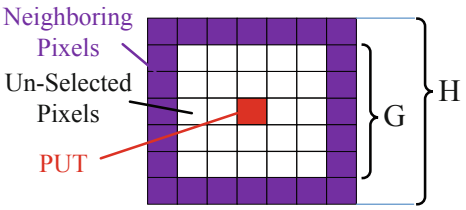


Fig. 3. The location relationship between PUT and surrounding pixels.

To reduce the anomaly pixels selected as background pixels in surrounding pixels, the Un-Selected pixels in Fig. 3 are employed. If some anomaly pixels located in outside as background pixels, it is called local anomaly contamination. The contamination may lead to a false alarm. To mitigate such contamination, the reconstruction errors from the network are employed as weights to adjust the contribution of each neighboring pixels in the anomaly score calculation as in formula 5. A PUT with a greater anomaly score means that it is more likely to be an anomaly targets.

$$\delta_d = 1/K\sum_{j=1}^{K} 1/t_j\left(\sum_{i=1}^{D}|x_{ji} - y_i|^2\right)^{1/2}, \tag{5}$$

where K is the total number of local neighboring pixels, t is the weight of local pixel which is reciprocal of its reconstruction error and is described as formula 6.

$$t = \sum_{i=1}^{B}|x_i - \hat{x}_i|^2, \tag{6}$$

where B is spectral bands number of the input HSI. $\hat{x}$ is the output of the network.

For the onboard mission, the low delay response and high processing throughput contradict the limitation of onboard computing resources. To reduce such gap, a light-weight network which can keep the detection accuracy with fewer neurons in each layer should be designed by exploring the network design space.

3 The Proposed Light-Weight Network Exploring Approach

To build up a light-weight network for the onboard HSI AD mission, a deep learning network structure exploring approach is proposed. The computation amount and the detection accuracy are employed as two objects for optimization. Both of them are related to the factors including neurons number, local window size, leaky value α of the network in formula 1. However, the computation amount and the detection accuracy conflict with each other with a latent relation. It is a constrained multiobjective optimization problem. In fact, this is a multi-objective programming (MOP) problem which is an NP-hard problem. In this paper, a genetic algorithm based optimization approach named nondominated sorting genetic algorithm II (NSGA-II) [14] is employed to build up an automatic evolution deep learning based HSI AD (Auto-EDL-AD). The basic flow of the NSGA-II is shown in Fig. 4.

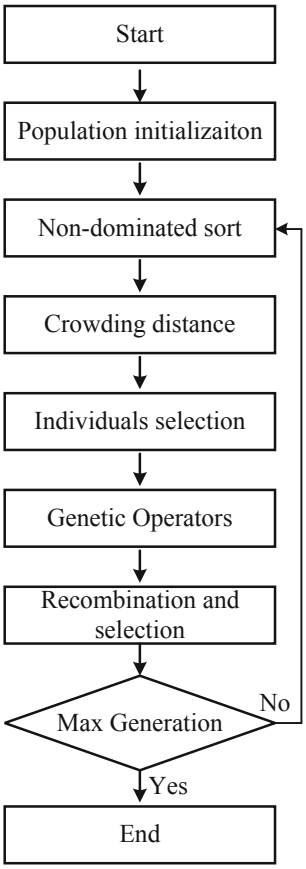


Fig. 4. The basic flow of the NSGA-II.

As shown in Fig. 4, the population is initialized within a specified range. The chromosome vector is generated which not only contains the basic genes but also contains the results of the fitness function and the crowding distance information.

The goal of the optimization is to maximize the detection accuracy and minimize the computation amount, the optimization model can be described as formula 7. Both of the objects are calculated depending on basic genes in the fitness function which is as shown in Fig. 5. Based on the effects for both of the objects, the neurons number, local window size, leaky value α of the network are employed as basic genes for the NSGA-II. The range of the basic genes are listed in Table 1. The objects are calculated as the formula 8 and the formula 11.

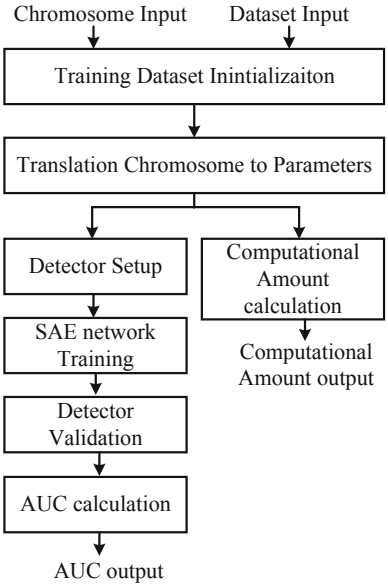


Fig. 5. The basic structure of the fitness function.

$$\begin{cases} min\ C = (G-H)\cdot(G+H)\cdot\sum_{l=1}^{n_{l-1}} s_l \cdot s_{l+1} \\ max\ AUC = \int_{\infty}^{-\infty} TPR(T)\cdot FPR(T)dT \\ s.t.\ \ G < H \\ \quad s_1 = s_{n_l} = B \\ \quad s_l > s_{l+1}, l < n_l/2 \\ \quad s_l < s_{l+1}, l > n_l/2 \end{cases} \tag{7}$$

where B is spectral bands number of the input HSI. H and G are the local window height and the guard window height of the local window. s_l is the neurons number of a layer. C is the computation amount as in formula 8. AUC is the area under the curve which is defined in formula 11.

The first object is to minimize the computation amount C which is approximatively quantized in terms of the number of multiply operations by the number of the neurons and the window size as formula 8.

$$C = (G - H) \cdot (G + H) \cdot \sum_{l=1}^{n_l - 1} s_l \cdot s_{l+1} \tag{8}$$

where H and G are the local window height and the guard window height of the local window respectively as shown in Fig. 3. From the previous researches [12], the n_l is set as 5. Due to the symmetrical structure of the autoencoder network, the neurons number of the first and the last layer are equal to the spectral band number B. Therefore, only the neuron number of the second (n_2) and the middle layer (n_m) are employed as the basic genes. The range of the n_2 and the n_m are specified as the Table 1.

Table 1. The range of the basic genes list.

Gene name	Leaky value α	n_2	n_m	H	G
Data type	Real	Integer	Integer	Integer	Integer
Maximum value	1	100	80	30	29
Minimum value	0.0001	20	1	8	6

The second object is to maximize the area under the curve (AUC) value which is used to indicate the detection accuracy. AUC is a common criterion to evaluate a detector in terms of accuracy. The curve is the receiver operating characteristic curve (ROC) which is created by drawing the true positive rate (TPR) against the false positive (FPR) when changing the threshold value for the anomaly score. TPR and FPR are defined as formulas 9 and 10 respectively. The anomaly score is the result of the detector which has been introduced in formula 5.

$$TPR = \frac{TP}{TP + FN} \tag{9}$$

where TP represents the true positive, which is the number of the correctly identified anomaly pixels. FN represents the false negative, which is the number of the incorrectly identified background pixels.

$$FPR = \frac{FP}{FP + TN} \tag{10}$$

where FP represents the false positive, which is the number of the incorrectly identified anomaly pixels. TN represents the true negative, which is the number of the correctly identified background pixels.

A greater AUC value means a high detection accuracy performance of a detector. The AUC value varies between 0 and 1. AUC is given by formula 11.

$$AUC = \int_{\infty}^{-\infty} TPR(T) \cdot FPR(T) dT \tag{11}$$

where T is the threshold value.

As shown in Fig. 4, after population initialization, a non-dominated sort and the crowding distance computing are executed for the chromosomes to gain its rank and the crowding distance to decide the position in the front. During the evolution process, the individuals with better fitness results are selected as a parent in the individual selection stage of the Fig. 4. To get the offspring chromosome, the crossover and mutation operator is performed in the genetic operator stage, the fitness results are calculated in this stage as well. The best solution depends on the rank and the crowding distance are selected for offspring generation. More details about NSGA-II can be found in Ref. [14].

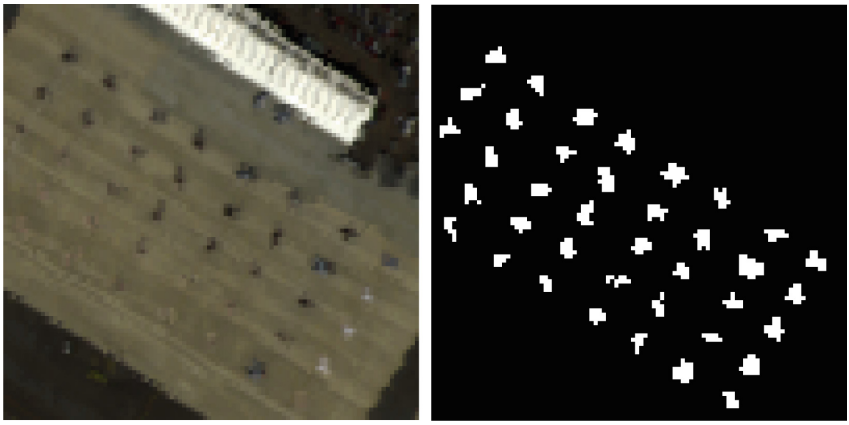
After exploration, Pareto optimal solutions can be generated which contains a series of basic genes. Then, a balanced decision between computation amount and the detection accuracy can be made based on those solutions.

4 Experiments and Results

To evaluate the proposed Auto-EDL-AD, a real HSI dataset is employed in the experiment. The dataset is captured over the Sandiego airport by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). The wavelength is from 0.4 μm–2.4 μm. The color image of the dataset is shown in Fig. 6. A part of the image with the size of 100×100 as shown in Fig. 7(a) is employed for the experiment. 166 spectral bands are selected from the 224 bands after removing the water absorption and low signal-to-noise ration bands with the central frequency of 0.37 μm–0.38 μm, 0.90 μm–0.97 μm, 1.11 μm–1.16 μm, 1.33 μm–1.50 μm and 1.78 μm–1.98 μm [4,6,15]. 38 planes are regarded as the anomaly objects for the HSI AD. The ground truth which indicates the location of the anomalies is shown in Fig. 7(b).

The classic HSI anomaly detector Reed-Xiaoli detector(RXD) which always used as the baseline algorithm for the HSI anomaly detection research, a state-of-the-art HSI AD, collaborative representation HSI anomaly detector (CRD) [8] and an SAE based AD (SAE-AD) [12] are implemented for comparison. The ROC and the AUC are employed as the criteria to evaluate the accuracy which has been introduced in Sect. 3. The throughputs of the above detectors are given as well to illustrate the computational efficiency. All the detectors are run by Matlab 2017a on a Dell T7910 workstation which contains 2 Intel Xeon CPU with the type of E5-2630 at 2.4 GHz and 32 Gigabyte Double-Data-Rate Three Synchronous Dynamic Random Access Memory (DDR3 SDRAM).

Fig. 6. The color image of the Sandiego airport HSI dataset.

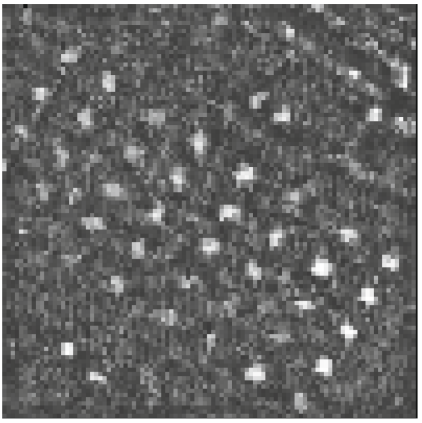


(a) The colour image and the anomaly targets of the Sandiego data set.

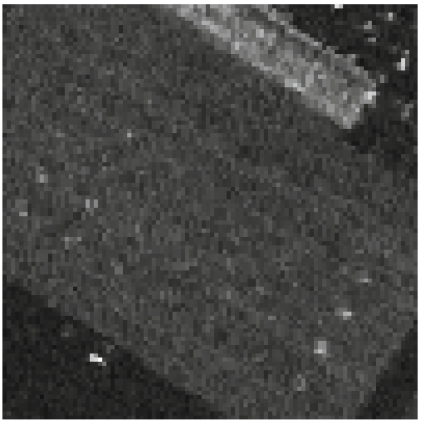
(b) The ground truth image of the Sandiego data set.

Fig. 7. The colour image and the ground truth image of Sandiego data set

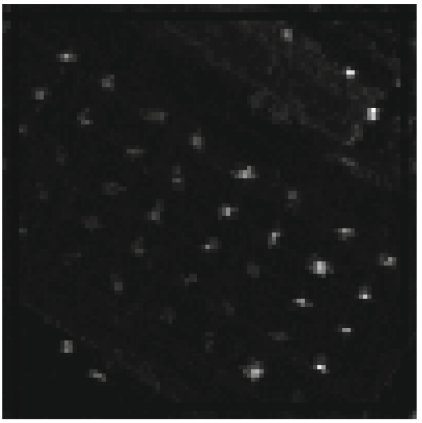
The detection results from the proposed Auto-EDL-AD, RXD, CRD and SAE-AD are given in term of anomaly score in Fig. 8.



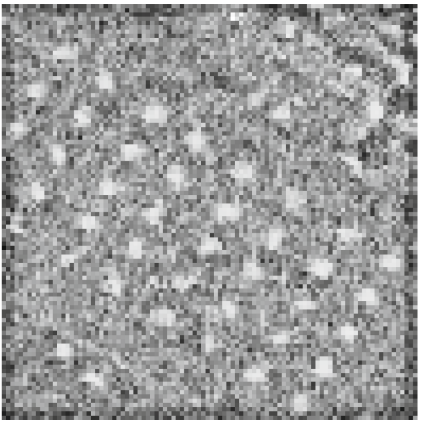
(a) The anomaly score image of the Sandiego by proposed Auto-EDL-AD.



(b) The anomaly score image of the Sandiego by LRXD.



(c) The anomaly score image of the Sandiego by CRD.



(d) The anomaly score image of the Sandiego by SAE-AD.

Fig. 8. The detection results for the Sandiego dataset.

The ROC curve of those detectors for the Auto-EDL-AD, RXD, CRD, and SAE-AD are given in Fig. 9. The AUC value and the processing time are listed in Table 2.

The anomaly score is converted to gray image in Fig. 8. The whiter pixel has a higher probability to be an anomaly pixel than the background pixel. From the Fig. 8(a), the proposed Auto-EDL-AD can indicate most of the anomaly objects clearly. From detection results by LRXD in Fig. 8(b), the anomaly targets show almost the same brightness with the background for the Sandiego dataset. In the results by CRD as shown in Fig. 8(c), the background are suppressed at a low level, while some of the anomaly pixels are suppressed to be the same level with background pixel. Figure 8(d) shows that the result difference between

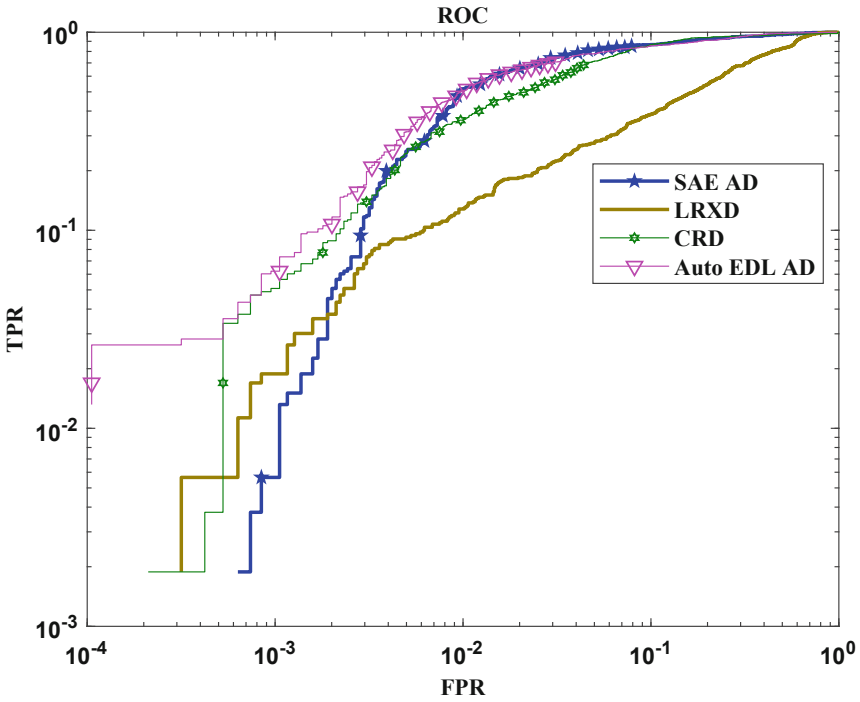


Fig. 9. The ROC curve for sandiego dataset.

anomaly pixels and the background pixels is not very great and more false alarm is generated.

From the definition of the ROC, a detector whose curve is located higher than others performs better. In Fig. 9, the ROCs are drawn for proposed Auto-EDL-AD, SAE-AD, LRXD and the CRD. The proposed Auto-EDL-AD outperforms other detectors in low FPR (FPR ≤ 0.01). When FPR ≤ 0.01, The ROC curve of the proposed Auto-EDL-AD overlap the ROC curve of CRD.

Table 2. The AUC value and the time consumption.

Detector name	Auto-EDL-AD	LRXD	CRD	SAE-AD
AUC value	0.922	0.776	0.917	0.914
Detection time (s)	9.41	80.01	34.38	16.23

From the Table 2, the AUC value of the proposed Auto-EDL-AD is almost the same with the state-of-the-art CRD and the SAE-AD. Due to comparison detectors are run with the same platform, less computation time means low computational complexity for a detector. From the Table 2, in term of the process

time, the proposed Auto-EDL-AD perform over 360% speedup with CRD and 172% speedup with SAE-AD. It is because that after the optimization by NSGA-II with the proposed approach, a light-weight network with the structure of $[166, 21, 5, 21, 166]$ is built up. The time in the Table 2 is the detection time which does not include the training time for exploring the network. The network exploring by the proposed approach required over 20 h which is executed during the satellite design and have no impacts for the detection stage.

5 Conclusion

To lessen the computation complexity of an HSI AD for an onboard mission, in this paper an automatic evolution SAE network is proposed to build up a light-weight HSI AD. The computation amount and the detection accuracy are employed as the optimization goals for multiobject programming by the NSGA-II. As a result, a fitness structure of the SAE-network are created for HSI AD. From the experimental results on a real HSI dataset, the HSI AD created by the proposed approach can reach up to over 360% and 170% respectively with CRD and SAE-AD. Therefore, the proposed approach can be used for onboard HSI AD design to fit the computation resources limitation and the fast response requirement for the online mission.

References

1. Chang, C.I., Chiang, S.S.: Anomaly detection and classification for hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. **40**(6), 1314–1325 (2002). https://doi.org/10.1109/tgrs.2002.800280
2. Chang, C.-I., Li, Y., Hobbs, M.C., Schultz, R.C., Liu, W.-M.: Progressive band processing of anomaly detection in hyperspectral imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **8**(7), 3558–3571 (2015). https://doi.org/10.1109/jstars.2015.2415782
3. Chen, S.-Y., Wang, Y., Wu, C.-C., Liu, C., Chang, C.-I.: Real-time causal processing of anomaly detection for hyperspectral imagery. IEEE Trans. Aerosp. Electron. Syst. **50**(2), 1510–1533 (2014). https://doi.org/10.1109/taes.2014.130065
4. Curran, P.J., Dungan, J.L.: Estimation of signal-to-noise: a new procedure applied to AVIRIS data. IEEE Trans. Geosci. Remote Sens. **27**(5), 620–628 (1989). https://doi.org/10.1109/tgrs.1989.35945
5. Du, B., Zhang, L.: Random-selection-based anomaly detector for hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. **49**(5), 1578–1589 (2011). https://doi.org/10.1109/TGRS.2010.2081677
6. Gao, B.C., Heidebrecht, K.B., Goetz, A.F.H.: Derivation of scaled surface reflectances from AVIRIS data. Remote Sensi. Environ. **44**(2–3), 165–178 (1993). https://doi.org/10.1016/0034-4257(93)90014-o
7. Hinton, G., Osindero, S., Teh, Y.: A fast learning algorithm for deep belief nets. Neural Comput. **7**(18), 1527–1554 (2006). https://doi.org/10.1162/neco.2006.18.7.1527
8. Li, W., Du, Q.: Collaborative representation for hyperspectral anomaly detection. IEEE Trans. Geosci. Remote Sens. **53**(3), 1463–1474 (2015A). https://doi.org/10.1109/tgrs.2014.2343955

9. Li, J., Zhang, H., Zhang, L., Ma, L.: Hyperspectral anomaly detection by the use of background joint sparse representation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **8**(6), 2523–2533 (2015). https://doi.org/10.1109/jstars.2015.2437073
10. Li, W., Wu, G., Du, Q.: Transferred deep learning for anomaly detection in hyperspectral imagery. IEEE Geosci. Remote Sens. Lett. **5**(14), 597–601 (2017). https://doi.org/10.1109/LGRS.2017.2657818
11. Ma, N., Wang, S., Yu, J., Peng, Y.: A DBN based anomaly targets detector for HSI. In: Asundi, A.K., Zhao, H., Osten, W. (eds.) AOPC 2017: 3D Measurement Technology for Intelligent Manufacturing, vol. 10458, pp. 1–6 (2017). https://doi.org/10.1117/12.2285766
12. Ma, N., Peng, Y., Wang, S., Gao, W.: A weight SAE based hyperspectral image anomaly targets detection. In: Proceedings of the International Conference on Electronic Measurement & Instruments (ICEMI), Yangzhou, China, October 2017, pp. 511–515. IEEE (2017). https://doi.org/10.1109/ICEMI.2017.8265874
13. Reed, I.S., Yu, X.L.: Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. IEEE Trans. Acoust. Speech Sig. Process. **38**(10), 1760–1770 (1990). https://doi.org/10.1109/29.60107
14. Ramesh, S., Kannan, S., Baskar, S.: Application of a fast and elitist multi-objective genetic algorithm to reactive power dispatch. Serb. J. Electr. Eng. **6**(6), 119–133 (2009). https://doi.org/10.1109/4235.996017
15. Rodger, A., Lynch, J.M.: Determining atmospheric column water vapour in the 0.4–2.5 μm spectral region. In: Proceedings of the AVIRIS Workshop Pasadena, California, USA, 27 February–02 March 2001. Jet Propulsion Laboratory (JPL) publication (2001)
16. Zhao, C., Li, X., Zhu, H.: Hyperspectral anomaly detection based on stacked denoising autoencoders. J. Appl. Remote Sens. **11**, 1–19 (2017). https://doi.org/10.1117/1.jrs.11.042605