Hyperspectral Anomaly Detection using Kernel Auto Encoder

Katinka Müller, Vinay Chakravarthi Gogineni, Milica Orlandic, Stefan Werner
Dept. of Electronic Systems, Norwegian University of Science and Technology-NTNU, Norway
E-mails: katinkam@stud.ntnu.no, {vinay.gogineni, milica.orlandic, stefan.werner}@ntnu.no

***Abstract*—Anomaly detection in hyperspectral images has attracted colossal attention due to its ability to uncover small distinctive objects dispersed across large geographical areas in an unsupervised manner. Autoencoders (AEs) have recently been demonstrated to be effective tools for detecting hyperspectral anomalies. Using pre-processing techniques along with AEs improves accuracy by removing noise and irrelevant information from the data, and also improves computational efficiency by reducing the dimensionality of the data or transforming it into a more appropriate representation. Therefore, this paper proposes to utilize kernel principal component analysis (KPCA) based pre-processing methods in conjunction with the autonomous hyperspectral anomaly detection autoencoder (AUTO-AD). Further, we propose to use multiple kernels in KPCA-based preprocessing in order to better capture the complexity of the data. Although KPCA- and MKPCA-based pre-processing shows excellent results when combined with hyperspectral anomaly detection algorithms, it has a high computational cost, which becomes crucial in resource-constrained applications. As a solution, we use random Fourier features (RFF) to approximate KPCA-based pre-processing. We conduct a series of experiments on various datasets to demonstrate the performance of the proposed framework. The experiments reveal that utilizing KPCA as a pre-processing step lead to better results than state-of-the-art hyperspectral anomaly detection approaches.**

***Index Terms*—Hyperspectral imaging, anomaly detection, autoencoder, kernel principle component analysis, random Fourier features**

I. Introduction

Hyper Spectral Imaging (HSI) is a technique that uses a wide range of electromagnetic spectrum to analyze each pixel [1]. Thus, it is possible to distinguish objects by taking advantage of the different absorption wavelengths of light for different materials. Hyperspectral anomaly detection (HAD) refers to the identification of pixels or sub-pixels with significantly different spectral characteristics from their neighboring pixels [2, 3]. It is worth noting that a pixel in HSI is a vector whose dimension is equals to the spectral dimension of the HSI. Remote sensing applications such as search-and-rescue, mine detection, and environmental monitoring rely heavily on the HAD. Consequently, there has been an increased interest in HAD as a research field, leading to the development of several state-of-the-art models using both traditional and deep learning methods [4, 5].

In a few traditional methods, the background of the HSI is modeled as a multidimensional Gaussian distribution, and the deviation of a test vector from this distribution is calculated using the Mahalanobis distance [6, 7]. However, due to the complexity of HSI, the Gaussian distribution is not always sufficient to accurately model the background distribution. Representation-based methods, such as collaborative representation detection (CRD) [8] assumes that background pixels can be modeled as a linear combination of the surrounding pixels while anomalies cannot. Additionally, tensor decomposition-based methods, such as prior-based tensor approximation (PTA) [9], considers the 3D structure of the HSI by treating the 3D hypercube as a third-order tensor. It is, however, difficult to generalize CRD and PTA to new datasets as both require manual parameter setting. A common issue with traditional methods is achieving good feature extraction for different types of data.

HAD approaches based on deep learning are able to overcome this problem by learning automatically how to extract features based on a variety of datasets [10, 11]. Auto encoders (AEs), which reconstruct the original input and use the error as an anomaly score, have recently been explored in HAD [12, 13, 14]. A recently proposed autonomous hyperspectral anomaly detection network (AUTO-AD) utilizes the same concept of reconstructing the background for separating the anomalies [15]. Instead of using the hyperspectral image as input, AUTO-AD uses uniform noise to train the model, which has been shown to improve the performance of the model. AUTO-AD has demonstrated promising results in terms of computational cost and detection accuracy, and has outperformed several state-of-the-art HAD models. The AUTO-AD, however, does not employ any form of preprocessing, which would potentially enhance its performance.

In this paper, we propose the use of KPCA as a preprocessing method along with AE in the AUTO-AD model. Using KPCA, the underlying structure of the data can be efficiently captured, which aids in identifying anomalies that may not be apparent in the raw data, making it easier for an AE to identify them. Recent study has demonstrated that using multiple kernels can improve hyperspectral data representation, making background and anomalies more distinct [16]. Therefore, the paper proposes to use multiple kernels in KPCA-based preprocessing, which results in multi-KPCA (MKPCA) preprocessing. Through this approach, various combinations of kernels can be used to capture different aspects of the data, therefore providing the AUTO-AD model with useful representations that could potentially improve its performance. However, this advantage comes at the expense of

This work is partly supported by the Research Council of Norway.

high computational costs associated with KPCA when dealing with large datasets, which is even more pronounced in the case of MKPCA. To reduce the computational cost associated with KPCA and MKPCA, this paper adopts the principles of random Fourier features (RFF) [17].

II. Autonomous Hyperspectral Anomaly Detection (Auto-AD)

The AEs for HAD are based on the principle that anomaly pixels are less likely to be reconstructed than background pixels, i.e., the areas corresponding to anomalies results in high reconstruction error [15]. However, after a large number of iterations, research has shown that the AE learns how to reconstruct the anomalies. To address this issue, the AUTO-AD uses adaptive-weight loss function to suppress the anomalies further [15]. The Fig. 1 illustrates the architecture of the AE proposed in the AUTO-AD. The AE consists of a decoder and encoder which are built up using fully convolutional layers, together with skip connections to complement the encoder with spatial details.

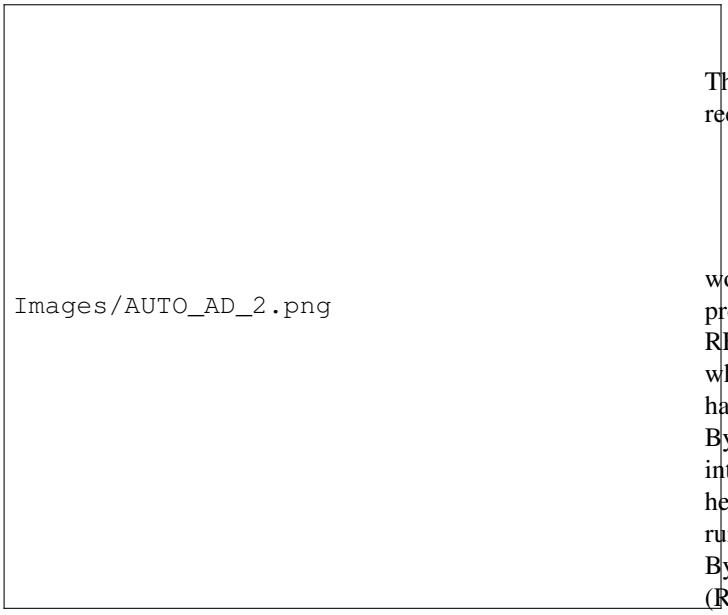

Fig. 1: Architecture of the Autonomous Hyperspectral Anomaly Detection (AUTO-AD).

In the AUTO-AD, the encoder consists of 15 convolutional layers. The input to the encoder is a hypercube, of the same dimension as the dataset, filled with uniform noise $\mathcal{U} \in [0, 0.1]$. The encoder performs spectral and spatial dimensionality reduction using convolutional filters, and utilizes skip connections. The decoder contains 11 convolutional layers together with up-sampling blocks that use nearest-neighbour interpolation to increase the spectral dimensionality. The input to to the decoder is a concatenation of the output of the encoder and outputs given from the skip connections. The output of the decoder contains the reconstructed background, which is sent as input to the adaptive weight-loss function.

To compute the loss-function, a weight map is constructed using the reconstruction error of each pixel $\mathbf{e}_{i,j}$, which is calculated using (1), where $\mathbf{x}_{i,j}$ is a pixel in the dataset $\mathbf{X} \in \mathbb{R}^{\mathbf{H}\times\mathbf{W}\times\mathbf{B}}$ with coordinates (i, j) and $\hat{\mathbf{x}}_{i,j}$ is a pixel in the output of the network. It is worth noting that each pixel $\mathbf{x}_{i,j}$ is a vector of same dimension as the spectral dimension B.

$$\mathbf{e}_{i,j} = ||\mathbf{x}_{i,j} - \hat{\mathbf{x}}_{i,j}||_2^2 \tag{1}$$

This gives the reconstruction error matrix $\mathbf{E} = \{e_{i,j}\}_{i=1,j=1}^{H,W}$, which is used to find the weighr map $\mathbf{W} = \{w_{i,j}\}_{i=1,j=1}^{H,W}$, where each weight $w_{i,j}$ is calculated using (2).

$$w_{i,j} = \max(\mathbf{E}) - \mathbf{e}_{i,j}, \tag{2}$$

The weight map is updated every 100 iterations. This gives the adaptive weight-loss function (3), where the loss is summed over the full dataset and used together with ADAM[18] to update the parameters of the network.

$$\mathcal{L} = \sum_{i=1}^{H}\sum_{i=j}^{W} ||(\mathbf{x}_{i,j} - \hat{\mathbf{x}}_{i,j})w_{i,j}||_2^2, \tag{3}$$

The result of the algorithm is a hyper-spectral image with the reconstructed background and a 2D detection map.

III. Proposed HAD

This paper proposes an implementation using the AE network proposed in the AUTO-AD paper, together with the pre-processing method MKPCA which has been optimized using RFF. The AE is chosen because it has shown great results when compared to other state-of-the-art models, additionally it has a low time-cost and requires no manual parameter setting. By utilizing KPCA as a pre-processing method the AE is introduced to the underlying structure of the data which can help to identify the anomalies. To reduce the time-cost of running the KPCA, RFF is used to compute the kernel matrix. By introducing MKPCA, with the random basial function (RBF) and Laplacian kernels, the model can be introduced to different representations or different features. A visualization of the proposed model is shown Figure 2.

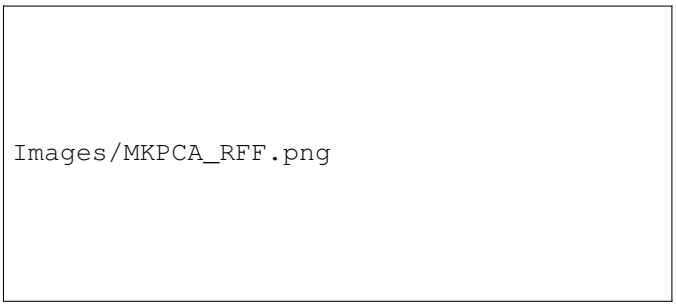

Fig. 2: Proposed implementation of AUTO-AD utilizing the MKPCA-based preprocessing.

A. Kernel Principle Component Analysis

PCA is a linear method and can only be used on linearly separable datasets. This is a common problem for complex data sets which are not linearly separable, like HSI. Kernel PCA is a solution for data sets that are not linearly separable since it projects the data onto a higher dimensional feature space where it is linearly separable [19]. KPCA uses the same principles as PCA but has an extra step at the beginning where the dataset $\mathbf{X}$ is projected onto a higher dimension using the kernel trick. The kernel trick is a mathematical technique that allows for the computation of the dot product between two vectors $\mathbf{x}$ and $\mathbf{x}'$ in a high-dimensional feature space without requiring the actual transformation of the data [20]. The kernel function is defined as

$$\kappa(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^{\mathbf{T}} \phi(\mathbf{x}') \tag{4}$$

where ϕ is a mapping that takes a vector from a dimension N to a new dimension M. The selection of kernel functions is not fixed but based on the application. Some commonly used kernel functions are presented by Patle and Chouhan [21]. The first step in KPCA is to express the covariance matrix $\hat{\mathbf{C}}$ as a dot product between two transformed vectors, as given in the function (4), which results in the equation below

$$\hat{\mathbf{C}} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{i=1}}^{\mathbf{N}} \phi(\mathbf{x})^{\mathbf{T}} \phi(\mathbf{x}') \tag{5}$$

Since $\hat{C}$ is expressed as a product of two vectors, it can be described as the kernel matrix $\mathbf{K}$ which is used as the covariance matrix in the eigendecomposition problem as given in (6), where $\mathbf{P}$ is the eigenvectors and $\mathbf{\Lambda}$ are the eigenvalues.

$$\mathbf{KP} = \mathbf{\Lambda P} \tag{6}$$

The eigenvectors represent $\mathbf{P}$ the principle components which are arranged based on the eigenvalues $\mathbf{\Lambda}$, representing the amount of variance in each component. For dimensionality reduction only ξ principle components are chosen. By applying KPCA, the underlying structure of the data can be captured, which helps identify the anomalies which might not be apparent in the raw data[19]. However, the computation of a kernel matrix of higher dimension comes with a computational cost [22].

B. Multi Kernel Learning

In recent years, researchers have introduced methods that use a combination of kernels, where each kernel corresponds to a different representation or features of the data. This approach is called multi kernel learning (MKL)[16]. In this paper MKL is utilized to obtain the detection map. The dataset is first pre-processed separately using Laplace and RBF kernels and sent into the AE in parallel to retrieve two detection maps,$\mathbf{D_1}$ and $\mathbf{D_2}$, as can be seen in Figure 2. These detection maps fused together to obtain a final detection map $\mathbf{D}$as described in (7).

$$\mathbf{D} = \alpha \mathbf{D}_1 + (1 - \alpha)\mathbf{D}_2 \tag{7}$$

C. Random Fourier Features

RFF can be used to accelerate the computation of the kernel matrix when the amount of data samples D, is large. This involves mapping the input data to a randomized low-dimensional feature space before applying a feature transformation[17]. The RFF algorithm approximates the feature transformation $\phi(\cdot) \approx z(\cdot)$ used to compute the kernel function $\kappa(\mathbf{x}, \mathbf{x}')$ in KPCA. This yields a new kernel function $\kappa(\mathbf{x}') = \phi(\mathbf{x})^{\mathbf{T}} \phi(\mathbf{x}) \approx \mathbf{z}(\mathbf{x})^{\mathbf{T}} \mathbf{z}(\mathbf{x}')$. The approximated transformation is expressed as

$$\begin{aligned} z(\mathbf{x}) = \sqrt{\frac{1}{M}} [&\cos(w_1 \mathbf{x} + b_1) ... \cos(w_M \mathbf{x} + b_M) \\ &\sin(w_1 \mathbf{x} + b_1) ... \sin(w_M \mathbf{x} + b_M)] \end{aligned} \tag{8}$$

where the weights w are drawn from a distribution $p(w)$ that is determined by the choice of kernel method being employed.The offset b is drawn from a uniform distribution $U \in [0, 2\pi]$.

IV. Experimental Results

The proposed model was tested on a public dataset and compared with state-of-the-art models[4][5], CRD, PTA and RX as well as the AUTO-AD model. The evaluation is based on the Area Under Curve (AUC) of the Region of Convergance (ROC) curve. To evaluate the effectiveness of this approach, an implementation of the default AUTO-AD model and three traditional models, RX, PTA and CRD will be used as a comparison in the results. **Add computer configurations if I want to add comptuational cost.**

A. Datasets

The dataset used for testing is the public Airport-Urban-Beach (ABU) and was made available at *Data sets* [23]. The dataset scenes cover an area between 100×100 and 150×150, with a range of $102 - 207$ spectral bands. The datasets have been manually extracted from images of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), except for the ABU-Beach-4 which is taken from the Reflective Optical System Imaging Spectrometer (ROSIS-03).

B. Parameter Settings

Parameters for PTA were set according to Li et al. [9]. The parameters for CRD were set according to the parameters that gave the best result for abu-airport-4, found in Song et al. [24]. The RX model did not need any parameter setting. The parameter configuration for the AUTO-AD model is set according to Wang et al. [15]. The MKPCA is configured using $\xi = 100$ principle components for both the Laplacian and RBF kernel functions, and the AE is configured in the same ways as the AUTO-AD.

C. Results

The results are given as a total average AUC score over the full ABU dataset. Table I show the AUC score together with the total average for each of the scenes, Airport, Beach and Urban. The results are shown for the AUTO-AD model

with and without the proposed pre-processing method, RFF-MKPCA. For comparison the results are also shown for KPCA with the RBF kernel and PCA bot with $\xi = 100$ principle components. Additionally the results from three traditional methods, RX, PTA and CRD are shown. To further evaluate the models, the computational time for three of the datasets: Airport 1, Airport 4 and Urban 2 are shown in Table II.

Utilizing KPCA obtains the best total average AUC score of $-$, outperforming the original AUTO-AD with $-$. Additionally, it outperforms all three state-of-the-art HAD models, PTA, CRD and RX, commonly used as reference models for HAD. This shows that utilizing pre-processing greatly improves the results.

Acknowledgment

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

V. Conclusions

Overall the proposed implementation of AUTO-AD using KPCA and min-max feature scaling proved to give the highest results of all the models, with an improvement of 10% over the original AUTO-AD model. Additionally, the AUTO-AD model outperformed three state-of-the-art HAD models, RX, PTA and CRD. Based on these results, it can be concluded that the proposed implementation of the AUTO-AD is an improvement and has been shown to give better results than several state-of-the-art HAD. For the future further testing and research using the proposed implementation of the AUTO-AD will be conducted. In general, the KPCA algorithm and the training of AE should be optimized, together with tests using different configurations of the AE network.

References

[1] Gamal ElMasry and Da-Wen Sun. "CHAPTER 1 - Principles of Hyperspectral Imaging Technology". In: *Hyperspectral Imaging for Food Quality Analysis and Control*. Ed. by Da-Wen Sun. San Diego: Academic Press, Jan. 1, 2010, pp. 3–43. ISBN: 978-0-12-374753-2. DOI: 10.1016/B978-0-12-374753-2.10001-2. URL: https://www.sciencedirect.com/science/article/pii/B9780123747532100012 (visited on Dec. 2, 2022).

[2] D.W.J. Stein et al. "Anomaly detection from hyperspectral imagery". In: *IEEE Signal Processing Magazine* 19.1 (Jan. 2002). Conference Name: IEEE Signal Processing Magazine, pp. 58–69. ISSN: 1558-0792. DOI: 10.1109/79.974730.

[3] Yuan Yuan, Dandan Ma, and Qi Wang. "Hyperspectral Anomaly Detection by Graph Pixel Selection". In: *IEEE Transactions on Cybernetics* 46.12 (Dec. 2016). Conference Name: IEEE Transactions on Cybernetics, pp. 3123–3134. ISSN: 2168-2275. DOI: 10.1109/TCYB.2015.2497711.

[4] Yichu Xu et al. "Hyperspectral Anomaly Detection Based on Machine Learning: An Overview". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022). Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 3351–3364. ISSN: 2151-1535. DOI: 10.1109/JSTARS.2022.3167830.

[5] Xing Hu et al. "Hyperspectral Anomaly Detection Using Deep Learning: A Review". In: *Remote Sensing* 14.9 (Apr. 20, 2022), p. 1973. ISSN: 2072-4292. DOI: 10.3390/rs14091973. URL: https://www.mdpi.com/2072-4292/14/9/1973 (visited on Dec. 4, 2022).

[6] I.S. Reed and X. Yu. "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 38.10 (Oct. 1990). Conference Name: IEEE Transactions on Acoustics, Speech, and Signal Processing, pp. 1760–1770. ISSN: 0096-3518. DOI: 10.1109/29.60107.

[7] Chunhui Zhao et al. "Global and Local Real-Time Anomaly Detectors for Hyperspectral Remote Sensing Imagery". In: *Remote Sensing* 7.4 (Apr. 2015). Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, pp. 3966–3985. ISSN: 2072-4292. DOI: 10.3390/rs70403966. URL: https://www.mdpi.com/2072-4292/7/4/3966 (visited on Feb. 28, 2023).

[8] Wei Li and Qian Du. "Collaborative Representation for Hyperspectral Anomaly Detection". In: *Geoscience and Remote Sensing, IEEE Transactions on* 53 (Mar. 1, 2015), pp. 1463–1474. DOI: 10.1109/TGRS.2014.2343955.

[9] Lu Li et al. "Prior-Based Tensor Approximation for Anomaly Detection in Hyperspectral Imagery". In: *IEEE Transactions on Neural Networks and Learning Systems* 33.3 (Mar. 2022). Conference Name: IEEE Transactions on Neural Networks and Learning Systems, pp. 1037–1050. ISSN: 2162-2388. DOI: 10.1109/TNNLS.2020.3038659.

[10] Xiyou Fu et al. "Hyperspectral Anomaly Detection via Deep Plug-and-Play Denoising CNN Regularization". In: *IEEE Transactions on Geoscience and Remote Sensing* 59.11 (Nov. 2021), pp. 9553–9568. ISSN: 0196-2892, 1558-0644. DOI: 10.1109/TGRS.2021.3049224. URL: https://ieeexplore.ieee.org/document/9329138/ (visited on Dec. 4, 2022).

[11] Wei Li, Guodong Wu, and Qian Du. "Transferred Deep Learning for Anomaly Detection in Hyperspectral Imagery". In: *IEEE Geoscience and Remote Sensing Letters* 14.5 (May 2017). Conference Name: IEEE Geoscience and Remote Sensing Letters, pp. 597–601. ISSN: 1558-0571. DOI: 10.1109/LGRS.2017.2657818.

[12] Chunhui Zhao and Lili Zhang. "Spectral-spatial stacked autoencoders based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection". In: *Infrared Physics & Technology* 92 (Aug. 1, 2018), pp. 166–176. ISSN: 1350-4495. DOI: 10.1016/j.infrared.

TABLE I: AUC score for CRD, PTA, RX, AUTO-AD and the proposed implementation of AUTO-AD with three different preprocessing methods: PCA, KPCA and RFF-MKPCA, utilizing RBF and RBF-Laplace respectively with $\xi = 100$ principle components.

ABU scene	RX	PTA	CRD	Original	PCA	KPCA	RFF-MKPCA
Airport 1	0.8221	0.7331	0.9246	0.6941	0.9394	0.9537	0.9333
Airport 2	0.8404	0.9096	0.8931	0.6764	0.9378	0.9789	0.9287
Airport 3	0.9288	0.5476	0.9456	0.921	0.9411	0.9367	0.9243
Airport 4	0.9526	0.9955	0.8664	0.5509	0.9879	0.9943	0.9899
Total Average	0.8860	0.7965	0.9074	0.7106	0.9515	**0.9659**	0.9441
Beach 1	0.9828	0.9638	0.9882	0.9605	0.9730	0.9708	0.9856
Beach 2	0.9106	0.8300	0.9257	0.9042	0.9702	0.9014	0.9649
Beach 3	0.9998	0.9203	0.9932	0.9276	0.9959	0.9996	**0.9998**
Beach 4	0.9887	0.9660	0.9417	0.9898	0.9704	0.9847	0.9494
Total Average	0.9705	0.9200	0.9622	0.9455	**0.9774**	0.9668	0.9749
Urban 1	0.9907	0.9055	0.9887	0.9833	0.8816	0.9886	0.9393
Urban 2	0.9946	0.9770	0.9294	0.9994	0.9962	0.9883	0.9106
Urban 3	0.9513	0.8346	0.9414	0.7663	0.9750	0.9790	0.9615
Urban 4	0.9887	0.8257	0.9549	0.9965	0.9866	0.9931	0.9951
Urban 5	0.9692	0.8258	0.9371	0.962	0.8834	0.9793	0.9168
Total Average	0.9789	0.8737	0.9503	0.9415	0.9446	**0.9857**	0.9447

TABLE II: Computational cost in seconds given for the different models, including the AUTO-AD with utilizing PCA, KPCA and RFF-MKPCA pre-processing, together with the results from RX, PTA and CRD.

ABU scene	RX	PTA	CRD	Original	PCA	KPCA	RFF-MKPCA
Airport 1	< 1	18	517	21	14	114	48
Airport 4	< 1	15	409	5	12	70	44
Urban 2	< 1	21	506	10	4	80	40

2018.06.001. URL: https://www.sciencedirect.com/science/article/pii/S1350449518301646 (visited on Feb. 28, 2023).

[13] Ganghui Fan et al. "Hyperspectral Anomaly Detection With Robust Graph Autoencoders". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022). Conference Name: IEEE Transactions on Geoscience and Remote Sensing, pp. 1–14. ISSN: 1558-0644. DOI: 10.1109/TGRS.2021.3097097.

[14] Shizhen Chang, Bo Du, and Liangpei Zhang. "A Sparse Autoencoder Based Hyperspectral Anomaly Detection Algorihtm Using Residual of Reconstruction Error". In: *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. ISSN: 2153-7003. July 2019, pp. 5488–5491. DOI: 10.1109/IGARSS.2019.8898697.

[15] Shaoyu Wang et al. "Auto-AD: Autonomous Hyperspectral Anomaly Detection Network Based on Fully Convolutional Autoencoder". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–14. ISSN: 0196-2892, 1558-0644. DOI: 10.1109/TGRS.2021.3057721. URL: https://ieeexplore.ieee.org/document/9382262/ (visited on Sept. 29, 2022).

[16] Yen-Yu Lin, Tyng-Luh Liu, and Chiou-Shann Fuh. "Multiple Kernel Learning for Dimensionality Reduction". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.6 (June 2011). Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1147–1160. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2010.183.

[17] Ali Rahimi and Benjamin Recht. "Random Features for Large-Scale Kernel Machines". In: *Advances in Neural Information Processing Systems*. Vol. 20. Curran Associates, Inc., 2007. URL: https://papers.nips.cc/paper/2007/hash/013a006f03dbc5392effeb8f18fda755-Abstract.html (visited on Jan. 31, 2023).

[18] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. Jan. 29, 2017. arXiv: 1412.6980[cs]. URL: http://arxiv.org/abs/1412.6980 (visited on Oct. 1, 2022).

[19] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. "Nonlinear Component Analysis as a Kernel Eigenvalue Problem". In: *Neural Computation* 10.5 (July 1, 1998). Publisher: MIT Press, pp. 1299–1319. ISSN: 0899-7667. DOI: 10.1162/089976698300017467. URL: https://direct.mit.edu/neco/article/10/5/1299/6193/Nonlinear-Component-Analysis-as-a-Kernel (visited on Oct. 3, 2022).

[20] K.-R. Muller et al. "An introduction to kernel-based learning algorithms". In: *IEEE Transactions on Neural Networks* 12.2 (Mar. 2001). Conference Name: IEEE Transactions on Neural Networks, pp. 181–201. ISSN: 1941-0093. DOI: 10.1109/72.914517.

[21] Arti Patle and Deepak Singh Chouhan. "SVM kernel functions for classification". In: *2013 International Conference on Advances in Technology and Engineering (ICATE)*. 2013 International Conference on Advances in Technology and Engineering (ICATE). Jan. 2013, pp. 1–9. DOI: 10.1109/ICAdTE.2013.6524743.

[22] L. J. Cao et al. "A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector

machine". In: *Neurocomputing*. Support Vector Machines 55.1 (Sept. 1, 2003), pp. 321–336. ISSN: 0925-2312. DOI: 10.1016/S0925-2312(03)00433-8. URL: https://www.sciencedirect.com/science/article/pii/S0925231203004338 (visited on Dec. 10, 2022).

[23] *Data sets*. Xudong Kang's Homepage. URL: http://xudongkang.weebly.com/data-sets.html (visited on Dec. 3, 2022).

[24] Xiangyu Song et al. "Spectral–Spatial Anomaly Detection of Hyperspectral Data Based on Improved Isolation Forest". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022). Conference Name: IEEE Transactions on Geoscience and Remote Sensing, pp. 1–16. ISSN: 1558-0644. DOI: 10.1109/TGRS.2021.3104998.