Skewed t-distribution for Hyperspectral Anomaly Detection based on Autoencoder

Koray Kayabol, *Senior Member IEEE*, Ensar Burak Aytekin, Sertac Arisoy and Ercan Engin Kuruoglu, *Senior Member, IEEE,*

***Abstract*—We propose multivariate skewed t-distribution for hyperspectral anomaly detection. The proposed distribution model is able to increase the detection performance of autoencoder-based anomaly detectors. In the proposed method, the reconstruction error of a deep autoencoder is modeled with a skewed t-distribution. The deep autoencoder network is trained based on adversarial learning strategy by feeding its input with the hyperspectral data cubes. The parameters of the t-distribution model are estimated using variational Bayesian approach. We define a multivariate skewed t-distribution based detection rule for pixel-wise anomaly detection. We compare our proposed method with those based on the multivariate normal distribution and the robust multivariate normal variance-mean mixture distributions on real hyperspectral data sets. The experimental results show that the proposed approach outperforms other detectors in the benchmark.**

***Index Terms*—multivariate skewed t-distribution, anomaly detection, hyperspectral image, autoencoder, variational Bayes.**

I. Introduction

ANOMALY detection (AD) is an unsupervised inference method that takes place in various domains. In this study, we focus on hyperspectral anomaly detection problem. Anomalous pixels deviate from the normal behaviour of hyperspectral data. Therefore, the main idea in hyperspectral AD is to find the pixels which exhibit different behaviour compared to the rest of the image. The RX AD [1], that is based on multivariate normal (MVN) distribution, is a fundamental AD algorithm for hyperspectral images.

Local AD methods need more computation times than the global methods because the local background statistics are calculated for each test pixel. Generally, the mixture models are preferred for global background estimation. The global mixture-model-based ADs are able to find the anomalies without knowing their sizes. Gaussian mixture models and Student's t mixture models are used in [2] and [3], respectively. Low-rank and sparse representation (LRASR) is also used for anomaly detection in hyperspectral images in [4].

Recently, deep learning methods take place in hyperspectral anomaly detection. Among the deep learning methods, autoencoders (AEs) and generative adversarial networks (GANs) [5] exhibit very good performances in anomaly/novelty detection. The main idea in these methods is to train the network to learn the background and, then, to use the reconstruction error to detect the anomalies. A deep autoencoder AD (DAEAD) method is proposed in [6]. The Wasserstein GAN (WGAN) model is applied to hyperspectral AD in [7]. In [8], an adversarial autoencoder (AAE) model is proposed for feature representation of HSI and anomaly detection. In [9] and [10], a GAN model is used to estimate the background for anomaly detection where the score is calculated by the combination of the spectral and spatial information. In [11], autoencoding adversarial network (AEAN) was proposed for unsupervised pixel-wise anomaly detection using spectral and spatial information together. In this study, we use the AEAN model proposed in [11] to find the reconstruction error and, then, propose to use a multivariate skewed t-distribution (MVSkt) to model the data based on reconstruction error. We also propose a new anomaly score function based on the proposed distribution.

In the most of the aforementioned deep-learning-based AD studies, the reconstruction error is assumed to be in the form of ℓ_1 or ℓ_2 norm. The use of ℓ_1 norm is more realistic because the reconstruction error is sparse in its nature. The Laplace distribution is the probabilistic counterpart of the ℓ_1 norm. Multivariate Laplace distribution or any other multivariate sparse distribution can be used to model the reconstruction error. The normal variance-mean mixture formulation provides a general distribution model that includes a large family of distributions as special cases [12], [13], e.g. multivariate Laplace (MVL), MacKay's Bessel (MVMB), Student's t (MVSt), Cauchy (MVC), Jeffrey's (MVJ). MVSt distribution has been already used for robust hyperspectral background modeling [14] and anomaly detection [3] for a long time. In this study, we use these robust distributions to model the reconstruction error of an autoencoder and for anomaly detection. However, our main contribution is the use of MVSkt that is also formulated as a special case of the normal variance-mean mixture. According to our experiment, the MVSkt model has a better detection performance compared to other sparse multivariate distribution models. To the best of our knowledge although some robust detectors are proposed for hyperspectral AD [3], [15], the MVSkt distribution has not been used.

Since the parameter estimation in the MVSkt distribution cannot be performed by simple estimation methods, we need to use a sophisticated numerical method. A numerical method is proposed in [16] for maximum likelihood estimation. While an expectation-maximization (EM) algorithm is used for parame-

Manuscript received April 14, 2021; revised August 18, 2021. This work is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Project No. 118E295.

K. Kayabol and S. Arisoy are with the Electronics Engineering Department, Gebze Technical University, Gebze, Kocaeli, Turkey, e-mail: koray.kayabol@gtu.edu.tr, sarisoy@gtu.edu.tr.

E. B. Aytekin is with the TUBITAK BILGEM Information Technologies Institute, Kocaeli, Turkey, e-mail: burak.aytekin@tubitak.gov.tr.

E. E. Kuruoglu is with Tsinghua-Berkeley Shenzhen Institute China, on leave from ISTI, CNR, Pisa, Italy, e-mail: ercan.kuruoglu@isti.cnr.it.

ter estimation in [17], a Markov chain Monte Carlo (MCMC) method is proposed in [18]. If we use the normal variance-mean mixture formulation, the parameter estimation of the distribution can be performed by variational Bayesian (VB) approach. In this study, we propose a VB inference method to estimate the parameters of the MVSkt distribution.

The proposed method has three consecutive tasks which are background learning with an autoencoder, estimation of the parameters of MVSkt distribution with VB inference and anomaly detection using the detection rule obtained from MVSkt distribution. The organization of the paper is as follows. Section II introduces the AEN-based background learning. Section II presents the MVSkt distribution model for reconstruction error and its VB-based parameter estimation method. The simulation results are reported in Section V. Section VI summarizes the conclusion and future work.

II. Deep Autoencoding Adversarial Network for Background Learning

We use the autoencoding adversarial network (AEAN) method proposed in [11] to model the background with an autoencoder. In the AEAN, the autoencoder sub-network is defined as a generator and adversarial learning is performed between the autoencoder and the discriminator sub-networks. In this paper, we use the 3D-AEAN model for background learning. We train the model with small image cubes extracted from the HSI and thus the autoencoder learns to reconstruct small image cubes. The autoencoder sub-network consists of encoder and decoder networks which have convolutional and deconvolutional layers. The detailed explanation of the network architectures can be found in [11].

We denote HSI with $M_1 \times M_2$ pixels and L spectral bands as $\mathbf{F} \in \mathbb{R}^{M_1 \times M_2 \times L}$ and ith pixel in HSI as $\mathbf{f}_i \in \mathbb{R}^L$, where $i = 1, \ldots, N$ is the lexicographically ordered pixel index. In order to obtain the training set for the AEAN model, the background purification, [19], is applied to the image and a binary map is obtained for background and anomaly pixels. Then, the training set $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ is formed by extracting $\mathbf{s} \in \mathbb{R}^{m \times m \times L}$ image cubes from the background regions in the binary map obtained after background purification. After training, we use the autoencoder to generate synthesized HSI $\hat{\mathbf{F}} \in \mathbb{R}^{M_1 \times M_2 \times L}$ in which ith pixel is shown as $\hat{\mathbf{f}}_i$.

III. Autoencoder Reconstruction Error

After the reconstruction of the hyperspectral image with autoencoder, we calculate the difference between the reconstructed and the original spectral vectors for each pixel as $\mathbf{d}_i = \mathbf{f}_i - \hat{\mathbf{f}}_i$. Then, we apply a local standard deviation filter to each difference spectral image constructed by $\mathbf{d}_i$. If the filter's kernel is located at an anomaly pixel, the local standard deviation that is the output of the filter becomes higher than that of the background. Otherwise, the local standard deviation becomes similar to that of background. Thus, after filtering, we obtain the anomalies-highlighted reconstruction error $\mathbf{r}_i \in \mathbb{R}_+^L$. Since the standard deviation is always positive, we model $\mathbf{r}_i$'s with a multivariate skewed t-distribution instead of symmetric t-distribution. In this model $\mathbf{r}_i$ is given by

$$\mathbf{r}_i = \mathbf{m} + z\mathbf{b} + \sqrt{z}\mathbf{T}^{-1/2}\mathbf{w}_i. \tag{1}$$

where $\mathbf{w}_i \sim \mathcal{N}(0, \mathbf{I})$ is white Gaussian random vector, $z \sim \mathcal{GIG}(\alpha, \beta, \gamma)$ is a generalized inverse Gaussian random variable and $\mathbf{b}$ is the skewness vector. In this study, we choose as $\mathbf{b} = [b, b, \ldots, b]^T$ where $b \geq 2$. We define a normal-Wishart joint prior for mean vector $\mathbf{m}$ and the precision (inverse covariance) matrix $\mathbf{T}$ as $\{\mathbf{m}, \mathbf{T}\} \sim \mathcal{N}\left(\mathbf{m}_0, (\lambda\mathbf{T})^{-1}\right) \mathcal{W}^{-1}(((\nu_0 - L - 1)\Psi_0)^{-1}, \nu_0))$ where $\mathbf{m}_0$, λ, ν_0 and Ψ_0 are the hyperparameters. By integrating (1) over the latent variable z, we have the following hyperbolic distribution for $\mathbf{r}_i$ [12]:

$$f(\mathbf{r}_i|\mathbf{m}, \mathbf{T}; \alpha, \beta, \gamma) = \frac{\gamma^{\frac{\alpha}{2}} |\mathbf{T}|^{\frac{1}{2}} \exp\left\{(\mathbf{r}_i - \mathbf{m})^T \mathbf{T}\mathbf{b}\right\}}{\beta^{\frac{\alpha}{2}} K_\alpha(\sqrt{\beta\gamma})(2\pi)^{\frac{L}{2}}} \times \left(\frac{\beta + R_i}{\gamma + \mathbf{b}^T\mathbf{T}\mathbf{b}}\right)^{\frac{\alpha}{2} - \frac{L}{4}} K_{\alpha - \frac{1}{2}}\left(\sqrt{(\beta + R_i)(\gamma + \mathbf{b}^T\mathbf{T}\mathbf{b})}\right) \tag{2}$$

where $R_i = (\mathbf{r}_i - \mathbf{m})^T \mathbf{T} (\mathbf{r}_i - \mathbf{m})$ and $K_\alpha(\cdot)$ is the modified Bessel function of second kind.

The distribution $f(z)$ determines the form of the marginal distribution $f(\mathbf{r}_i|\mathbf{m}, \mathbf{T}; \alpha, \beta, \gamma)$ as given in Table I. By changing the parameters of $f(z)$, we can obtain different distributions for $\mathbf{r}_i$. For $\alpha < 0$, as γ goes to infinity, (2) reduces to the MVSkt distribution given in Table I. The other multivariate distributions arisen from (2) are also shown in Table I.

After the parameter estimation, the anomaly detection score can be found by using the probability distribution function (pdf) given in Table I. Rather than directly using the pdfs, we use the negative logarithm of them. For instance, by taking the negative logarithm of pdf in the last row of Table I, the multivariate skewed t-distribution-based anomaly detection score is found by

$$\mathrm{AD}(\mathbf{r}_i) = -\frac{2\alpha + L}{2} \log\left(1 + \frac{R_i}{\beta}\right) - (\mathbf{r}_i - \mathbf{m})^T \mathbf{T}\mathbf{b}. \tag{3}$$

IV. Variational Bayesian Inference for Multivariate Skewed t-distribution Model

The posterior inference of the latent (unobserved) variables, $\Phi = \{\mathbf{m}, \mathbf{T}, z\}$, and the parameters Θ from observations $\mathbf{r}_{1:N}$ can be made by some approximation methods such as MCMC or VB [20]. In MCMC methods, the variables are sampled in iterative way that after a sufficient number of iterations, the true posterior is approximated by the generated samples. Since MCMC methods need long times for convergence, we resort to VB methods. Let us consider the following maximum marginal likelihood estimation problem:

$$\hat{\Theta} = \max_{\Theta} \log f(\mathbf{r}_{1:N}|\Theta) \tag{4}$$

where Θ is the set of parameters of the MVSkt distribution model given in Table I and the latent variables, Φ, are integrated out. Using an approximate posterior distribution $q(\Phi)$, the marginal log-likelihood in (4) can be expressed as follows [20]:

$$\log f(\mathbf{r}_{1:N}|\Theta) = F(q, \Theta) + D_{KL}(f(\Phi|\mathbf{r}_{1:N}, \Theta)||q) \tag{5}$$

where $F(q, \Theta) = \int \log\left(\frac{f(\mathbf{r}_{1:N}, \Phi|\Theta)}{q(\Phi)}\right) q(\Phi) d\Phi$ is the free energy and $D_{KL}(f||q) = -\int \log\left(\frac{f(\Phi|\mathbf{r}_{1:N}, \Theta)}{q(\Phi)}\right) q(\Phi) d\Phi$ is

TABLE I: Normal variance-mean mixture distributions.

Distribution	Parameters	pdf
Student's t	$\mathbf{b}=0$, $\alpha<0$, $\gamma\longrightarrow\infty$	$f(\mathbf{r}_i)=\frac{\lvert\mathbf{T}\rvert^{\frac{1}{2}}\Gamma(-\alpha+L/2)}{\beta^{\alpha}\Gamma(-\alpha)(\pi)^{\frac{L}{2}}}(\beta+R_i)^{\alpha-\frac{L}{2}}$
Cauchy	$\mathbf{b}=0$, $\alpha=-1/2$, $\beta=1$, $\gamma\longrightarrow\infty$	$f(\mathbf{r}_i)=\frac{\lvert\mathbf{T}\rvert^{\frac{1}{2}}\Gamma((L+1)/2)}{\beta^{\alpha}\Gamma(1/2)(\pi)^{\frac{L}{2}}}(1+R_i)^{-\frac{(L+1)}{2}}$
Laplace	$\mathbf{b}=0$, $\alpha=(L+1)/2$, $\beta\longrightarrow\infty$	$f(\mathbf{r}_i)\propto\gamma^{\frac{L}{2}}\exp\{-\sqrt{\gamma R_i}\}$
Jeffry's	$\mathbf{b}=0$, $\alpha\longrightarrow 0$, $\beta\longrightarrow 0$, $\gamma\longrightarrow 0$	$f(\mathbf{r}_i)\propto R_i^{-L}$
skewed t	$\mathbf{b}>1$, $\alpha<0$, $\gamma\longrightarrow\infty$	$f(\mathbf{r}_i)=\frac{\lvert\mathbf{T}\rvert^{\frac{1}{2}}\Gamma(-\alpha+L/2)\exp\{(\mathbf{r}_i-\mathbf{m})^T\mathbf{T}\mathbf{b}\}}{\beta^{\alpha}\Gamma(-\alpha)(\pi)^{\frac{L}{2}}}(\beta+R_i)^{\alpha-\frac{L}{2}}$

the Kullback-Leibler (KL) divergence between $f(\Phi|\mathbf{r}_{1:N},\Theta)$ and $q(\Phi)$. In the VB inference, the approximate density $q(\Phi)$ is found by maximizing $F(q,\Theta)$ or minimizing $D_{KL}(f||q)$. Unlike the EM algorithm, in VB, we use the factorized approximation of the exact posterior such that $f(\Phi|\mathbf{r}_{1:N},\Theta)\approx q(\Phi)=q(\mathbf{m})q(\mathbf{T})q(z)$. This approximation makes the expectation integrals be computable. In VB method, we first find $q(\mathbf{m})$, $q(\mathbf{T})$ and $q(z)$, and then update the parameter set Θ in an iterative way. In Section IV-A, we give the details of determination of $q(\mathbf{m})$, $q(\mathbf{T})$ and $q(z)$. The update equations for parameters are presented in Section IV-C.

A. Determination of $q(\mathbf{m})$, $q(\mathbf{T})$ and $q(z)$

The optimum distribution $q(\mathbf{m})$ must satisfy the following equation:

$$\begin{aligned}\log q(\mathbf{m}) =&\mathrm{E}\left[-\frac{1}{2z}\sum_{i=1}^{N}(\mathbf{r}_i-\mathbf{m}-z\mathbf{b})^T\mathbf{T}(\mathbf{r}_i-\mathbf{m}-z\mathbf{b})\right.\\&\left.-\frac{\lambda}{2}(\mathbf{m}-\mathbf{m}_0)^T\mathbf{T}(\mathbf{m}-\mathbf{m}_0)\right]+\text{const.}\end{aligned}\tag{6}$$

The resulting density should be an MVN as $q(\mathbf{m})=\mathcal{N}(\boldsymbol{\mu}_m,\mathbf{C}_m)$ with

$$\boldsymbol{\mu}_m=\frac{\mathrm{E}[z^{-1}]N\bar{\mathbf{r}}-N\mathbf{b}+\lambda\mathbf{m}_0}{\mathrm{E}[z^{-1}]N+\lambda}\tag{7}$$

and $\mathbf{C}_m=((\mathrm{E}[z^{-1}]N+\lambda)\mathrm{E}[\mathbf{T}])^{-1}$ where $\bar{\mathbf{r}}=\frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i$.

The optimum distribution $q(\mathbf{T})$ must satisfy the following equation:

$$\begin{aligned}\log q(\mathbf{T}) =&-\frac{1}{2}\mathrm{tr}\Bigg(\mathbf{T}\Bigg(\mathrm{E}[z^{-1}]\sum_{i=1}^{N}(\mathbf{r}_i-\mathrm{E}[\mathbf{m}])(\mathbf{r}_i-\mathrm{E}[\mathbf{m}])^T\\&-2\sum_{i=1}^{N}(\mathbf{r}_i-\mathrm{E}[\mathbf{m}]-\frac{\mathrm{E}[z]}{2}\mathbf{b})\mathbf{b}^T\\&+\lambda(\mathrm{E}[\mathbf{m}]-\mathbf{m}_0)(\mathrm{E}[\mathbf{m}]-\mathbf{m}_0)^T\\&+\mathbf{C}_m(\mathrm{E}[z^{-1}]N+\lambda)+(\tau-L-1)\Psi_0\Bigg)\Bigg)\\&+\frac{1}{2}(\tau+L+N+2)\log|\mathbf{T}|+\text{const.}\end{aligned}\tag{8}$$

The resulting density should be a Wishart as $q(\mathbf{T})=\mathcal{W}(\Psi_N^{-1},\nu_N)$ with

$$\begin{aligned}\Psi_N =&\mathrm{E}[z^{-1}]\sum_{i=1}^{N}(\mathbf{r}_i-\mathrm{E}[\mathbf{m}])(\mathbf{r}_i-\mathrm{E}[\mathbf{m}])^T\\&-2\sum_{i=1}^{N}(\mathbf{r}_i-\mathrm{E}[\mathbf{m}]-\frac{\mathrm{E}[z]}{2}\mathbf{b})\mathbf{b}^T\\&+\lambda(\mathrm{E}[\mathbf{m}]-\mathbf{m}_0)(\mathrm{E}[\mathbf{m}]-\mathbf{m}_0)^T\\&+\mathbf{C}_m(\mathrm{E}[z^{-1}]N+\lambda)+(\tau-L-1)\Psi_0\end{aligned}\tag{9}$$

and $\nu_N=(\tau+N+1)$.

The optimum distribution $q(z)$ must satisfy the following equation:

$$\begin{aligned}\log q(z) =&-\frac{1}{2z}\mathrm{tr}\Bigg(\mathrm{E}[\mathbf{T}]\Bigg(\sum_{i=1}^{N}(\mathbf{r}_i-\mathrm{E}[\mathbf{m}])(\mathbf{r}_i-\mathrm{E}[\mathbf{m}])^T\Bigg)\Bigg)\\&+\frac{N}{2z}\mathrm{tr}(\mathrm{E}[\mathbf{T}]\mathbf{C}_m)-\frac{Nz}{2}\mathrm{tr}(\mathrm{E}[\mathbf{T}]\mathbf{b}\mathbf{b}^T)\\&+(\alpha-\frac{NL}{2}-1)\log z-\frac{1}{2z}\beta+\text{const.}\end{aligned}\tag{10}$$

The resulting density should be a generalized inverse Gaussian as $q(\mathbf{z})=\mathcal{GIG}(\alpha-\frac{NL}{2},c,d)$ with $c=N\mathrm{tr}(\mathrm{E}[\mathbf{T}]\mathbf{b}\mathbf{b}^T)$ and

$$d=\mathrm{tr}\Bigg(\mathrm{E}[\mathbf{T}]\Bigg(\sum_{i=1}^{N}(\mathbf{r}_i-\mathrm{E}[\mathbf{m}])(\mathbf{r}_i-\mathrm{E}[\mathbf{m}])^T+N\mathbf{C}_m\Bigg)\Bigg)+\beta.\tag{11}$$

B. Expectations

The expectations used to calculate (6)-(11) are given by $\mathrm{E}[\mathbf{m}]=\boldsymbol{\mu}_m$, $\mathrm{E}[\mathbf{T}]=\nu_N\Psi_N^{-1}$,

$$\mathrm{E}[z^{-1}]=\frac{K_{\alpha-\frac{NL}{2}-1}(\sqrt{cd})}{K_{\alpha-\frac{NL}{2}}(\sqrt{cd})}\left(\frac{d}{c}\right)^{-\frac{1}{2}},\tag{12}$$

$$\mathrm{E}[z]=\frac{K_{\alpha-\frac{NL}{2}+1}(\sqrt{cd})}{K_{\alpha-\frac{NL}{2}}(\sqrt{cd})}\left(\frac{d}{c}\right)^{\frac{1}{2}}.\tag{13}$$

C. Update of parameters

As defined in (4), the parameters can be found by maximizing the marginal log-likelihood in (5) that is equivalent to maximizing the lower bound $F(q, \Theta)$. We estimate all the parameters except τ and α .

The lower bound for $\mathbf{m}_0$, λ, Ψ_0 and β are given by

$$F(q, \mathbf{m}_0) = -\mathrm{E}\left[\frac{\lambda}{2}(\mathbf{m}-\mathbf{m}_0)^T\mathbf{T}(\mathbf{m}-\mathbf{m}_0)\right] \quad (14)$$

$$F(q, \Psi_0) = \frac{\tau}{2}\log|\Psi_0| - \mathrm{tr}\left((\tau-L-1)\Psi_0\mathrm{E}[\mathbf{T}]\right) \quad (15)$$

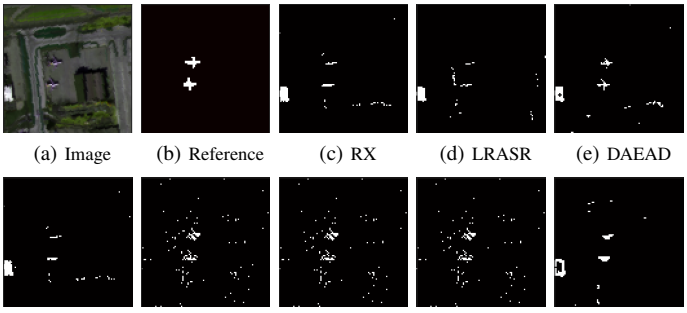
$$F(q, \beta) = -\alpha\log\beta - \frac{\beta\mathrm{E}[z^{-1}]}{2} \quad (16)$$

$$F(q, \lambda) = \frac{L}{2}\log\lambda - \frac{\lambda}{2}\mathrm{E}\left[(\mathbf{m}-\mathbf{m}_0)^T\mathbf{T}(\mathbf{m}-\mathbf{m}_0)\right] \quad (17)$$

The estimators which maximize (14)-(16) are given respectively by $\mathbf{m}_0 = \mathrm{E}[\mathbf{m}]$, $\Psi_0 = \frac{\tau}{\tau-L-1}(\mathrm{E}[\mathbf{T}])^{-1}$, $\beta = -\frac{2\alpha}{\mathrm{E}[z^{-1}]}$,

$$\lambda = \frac{L}{\mathrm{tr}\left(\mathrm{E}[\mathbf{T}](\mathrm{E}[\mathbf{m}]-\mathbf{m}_0)(\mathrm{E}[\mathbf{m}]-\mathbf{m}_0)^T + \mathbf{C}_m\right)} \quad (18)$$

We determine τ over new parameter p as $\tau = \frac{(N-L-1)p+L+1}{1-p}$ where $p = \frac{\tau-L-1}{N+\tau-L-1}$ and satisfies that $0 \leq p \leq 1$. If p is close to 0, the contribution of samples to calculate $\mathrm{E}[\mathbf{T}]$ is increased. Otherwise, the contribution of Ψ_0 is increased. We set α to $L/2$.



(a) Image (b) Reference (c) RX (d) LRASR (e) DAEAD

(f) MVN (g) MVL (h) MVSt (i) MVJ (j) MVSkt

Fig. 1: Detection maps for ABU Airport2 at false alarm rate 0.01.

V. Experimental Results

In this section, we evaluate the detection performances of the proposed MVSkt model with those of other multivariate distributions including MVN, MVC, MVL, MVSt and MVJ. Since these are symmetric distributions, except MVSkt, generally used in sparsity modeling, we fit these distributions directly to the reconstruction error using variational inference method. We also use RX, LRASR [4] and DAEAD [6] methods for comparison. The detection performances are measured with the area under the receiver operating characteristic (ROC) curve (AUC). We also use FAR@100 metric that gives the false alarm rate (FAR) corresponding to the maximum threshold value required to detect all anomaly pixels.

In the experiments, we use the ABU data set provided by [21]. In the ABU data set, the HSIs were captured by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor, except the ABU Beach4 image, which was captured by the reflective optic system imaging spectrometer (ROSIS-03) sensor. The sizes of images are between 100×100 and 150×150 and the numbers of spectral bands are between 102 and 205. The different kinds of man-made objects are assumed as anomalies. While in the urban area, the isolated small buildings are called as anomalies, in the airport scene, air-crafts are assumed to be anomalies. In the beach areas, the boats are considered as anomalies. We use 3D-AEAN and its parameter setting proposed in [11] to obtain reconstruction error.

The AUC and FAR@100 results are given in Table II and III respectively. According to the AUC metric, the MVSkt model gives the best results on average. If we look at the FAR@100 results given in Table III, the MVSkt model detects all anomaly pixels with the lowest false alarm rate compared to the other methods. If we consider the AUC and FAR@100 metrics together, the MVSkt model shows the best detection performance. Fig. 1 shows the detection maps for the ABU Airport2 image at false alarm rate 0.01. As seen from Fig. 1, the detection maps obtained by the MVSkt model is closer to the reference map.

In the last row of Table II, we give the average computation times of the methods. Since the training times of the deep networks are naturally higher than those of classical methods, we give the computation time of detection. As seen from the table, MVSkt model performs better than the other models at the cost of increased computation time.

VI. Conclusion

In this paper, we propose to use multivariate skewed t-distribution to model the reconstruction error of an autoencoder and to detect anomaly pixels in hyperspectral images. The proposed distribution is from the normal variance-mean mixture distribution family and allows us a single global probabilistic model and pixel-wise anomaly detection. By using multivariate skewed t-distribution model anomaly detection performance of the autoencoder is increased. In this study, we use a GAN-based autoencoder model but the proposed anomaly detection method is not restricted with the GAN-based autoencoders. It can be applied to the reconstruction error of any other autoencoder to improve its performance.

References

[1] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust. Speech Sig. Process.*, vol. 38, no. 10, pp. 1160–1170, 1990.

[2] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, 2002.

[3] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, 2014.

[4] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, 2016.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

TABLE II: AUC results of the methods for each image, average AUC and computation time (in second).

Images	RX	LRASR	DAEAD	MVN	MVC	MVL	MVSt	MVJ	MVSkt
Airport1	0.8821	0.6940	0.8994	0.7476	0.8790	0.8825	0.8790	0.8786	**0.9556**
Airport2	0.8404	0.6759	0.9166	0.9107	0.9348	0.9385	0.9348	0.9348	**0.9947**
Airport3	0.9288	0.9210	**0.9322**	0.9071	0.8709	0.8781	0.8709	0.8709	0.9207
Airport4	0.9526	0.5509	0.9493	0.9193	0.9645	0.9663	0.9645	0.9639	**0.9859**
Urban1	0.9907	0.9834	0.9868	0.9555	0.9767	0.9772	0.9767	0.9767	**0.9948**
Urban2	0.9946	**0.9994**	0.9907	0.9776	0.9968	0.9968	0.9968	0.9968	0.9836
Urban3	0.9513	0.7663	0.9524	0.8448	**0.9739**	0.9737	**0.9739**	**0.9739**	0.9658
Urban4	**0.9887**	0.9781	0.9551	0.9679	0.9684	0.9697	0.9684	0.9673	0.9314
Urban5	**0.9692**	0.9618	0.9592	0.9506	0.9351	0.9382	0.9351	0.9351	0.9620
Beach1	0.9807	0.9433	0.9747	0.9773	0.9887	0.9875	0.9886	0.9891	**0.9955**
Beach2	0.9106	0.9040	0.8936	0.8502	0.9310	**0.9427**	0.9309	0.9372	0.9041
Beach3	**0.9998**	0.9272	0.9921	0.9964	0.9966	0.9901	0.9966	0.9949	0.9997
Beach4	0.9538	0.9059	0.9583	0.9862	0.9914	0.9933	0.9914	0.9927	**0.9952**
Average	0.9495	0.8624	0.9508	0.9224	0.9544	0.9565	0.9544	0.9548	**0.9684**
Time	0.312	1.894	0.025	0.093	8.763	4.340	4.533	1.916	7.181

TABLE III: FAR@100 results of the methods for each image and average FAR@100.

Images	RX	LRASR	DAEAD	MVN	MVC	MVL	MVSt	MVJ	MVSkt
Airport1	0.9265	0.9762	0.6121	0.6769	0.8337	0.8129	0.8336	0.8334	**0.3614**
Airport2	0.9336	1.0000	0.5736	0.5777	0.6405	0.5945	0.6405	0.6399	**0.0961**
Airport3	0.6024	0.7098	**0.3804**	0.5865	0.9177	0.9059	0.9177	0.9176	0.7336
Airport4	0.2910	0.8282	0.1180	0.3659	0.3416	0.3483	0.3415	0.3462	**0.0460**
Urban1	0.0685	0.1011	0.0627	0.1352	0.3934	0.3812	0.3934	0.3931	**0.0302**
Urban2	0.0364	**0.0086**	0.1262	0.2068	0.0810	0.1058	0.0808	0.1061	0.2617
Urban3	0.2398	0.6577	0.1328	0.2831	**0.1249**	0.1258	**0.1249**	**0.1249**	0.1362
Urban4	**0.0845**	1.0000	0.3406	0.2911	0.7311	0.7882	0.7307	0.7636	0.6457
Urban5	0.4896	1.0000	0.3333	0.4148	0.7477	0.7876	0.7471	0.7806	**0.2023**
Beach1	0.2800	1.0000	0.1199	0.0641	0.0453	0.0675	0.0453	0.0448	**0.0141**
Beach2	**0.2146**	0.8289	0.7097	0.3332	0.4725	0.5334	0.4723	0.4740	0.2225
Beach3	0.0015	0.2907	0.0126	0.0167	0.0279	0.0926	0.0277	0.0440	**0.0011**
Beach4	0.3641	0.2347	0.2514	0.1454	0.1077	0.0889	0.1079	**0.0829**	0.1893
Average	0.3487	0.6643	0.2903	0.3152	0.4204	0.4333	0.4203	0.4278	**0.2262**

[6] A. Taghipour and H. Ghassemian, "Unsupervised hyperspectral target detection using spectral residual of deep autoencoder networks," in *Proc. 4th Int. Conf. Pattern Recognit. Image Anal. (IPRIA)*, 2019, pp. 52–57.

[7] A. W. W. Eide, "Applying generative adversarial networks for anomaly detection in hyperspectral remote sensing imagery," Master's thesis, NTNU, 2018.

[8] W. Xie, J. Lei, B. Liu, Y. Li, and X. Jia, "Spectral constraint adversarial autoencoders approach to feature representation in hyperspectral anomaly detection," *Neural Netw.*, vol. 119, pp. 222–234, 2019.

[9] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4666–4679, 2020.

[10] K. Jiang, W. Xie, Y. Li, J. Lei, G. He, and Q. Du, "Semisupervised spectral learning with generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 5224–5236, 2020.

[11] S. Arisoy, N. M. Nasrabadi, and K. Kayabol, "Unsupervised pixel-wise hyperspectral anomalydetection via autoencoding adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, 2021.

[12] O. Barndorff-Nielsen, J. Kent, and M. Sorensen, "Normal variance-mean mixtures and z distributions," *Statistical Review / Revue Internationale de Statistique*, vol. 50, no. 2, pp. 145–159, 1982.

[13] S. D. Babacan, S. Nakajima, and M. N. Do, "Bayesian group-sparse modeling and variational inference," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2906–2921, 2014.

[14] D. B. Marden and D. G. Manolakis, "Using elliptically contoured distributions to model hyperspectral imaging data and generate statistically similar synthetic data," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, S. S. Shen and P. E. Lewis, Eds., vol. 5425, International Society for Optics and Photonics. SPIE, 2004, pp. 558 – 572.

[15] J. Frontera-Pons, M. A. Veganzones, F. Pascal, and J.-P. Ovarlez, "Hyperspectral anomaly detectors using robust estimators," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 720–731, 2016.

[16] A. Azzalini and A. Capitanio, "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 65, no. 2, p. 367–389, 2003.

[17] K. Wang, S. K. Ng, and G. J. McLachlan, "Multivariate skew t mixture models: Applications to fluorescence-activated cell sorting data," in *2009 Digital Image Computing: Techniques and Applications*. IEEE, 2009, pp. 526–531.

[18] S. Fruhwirth-Schnatter and S. Pyne, "Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions," *Biostatistics*, vol. 11, no. 2, pp. 317–336, 2010.

[19] C. Zhao, X. Yao, and Y. Yan, "Modified kernel RX algorithm based on background purification and inverse-of-matrix-free calculation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 544–548, 2017.

[20] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, 2008.

[21] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, 2017.