SGC-based Anomaly Detection for Multivariate Time Series

Kewei Hu, Qiang Tian, Biao Wang, Jiakun Wu, and He Li*

Xidian University, Xi'an 710000, China
heli@xidian.edu.cn

Abstract. In industrial facilities or IT systems, there are lots of multivariate time series generated from various metrics. Anomaly detection in multivariate time series is of great importance in applications such as fault diagnosis and root cause discovery. Recently, some unsupervised methods have made great progress in this task, especially the reconstruction architecture of autoencoders (AEs), learning normal distribution, and producing a significant error for anomalies. Although AEs can reconstruct subtle abnormal patterns well with the powerful generalization ability, it also leads to a high false negative. Moreover, these AE-based models ignore the dependence among variables at different time scales. In this paper, we propose an enhanced anomaly detection framework that builds upon the Multiscale Wavelet Graph Autoencoder (MEGA) by substituting the Graph Convolutional Network (GCN) with Simplified Graph Convolution (SGC) to augment the model's performance. The core idea is to leverage the spectral methods of SGC to process the multivariate time series data obtained by integrating Discrete Wavelet Transform (DWT) into the AE. Experiments have been conducted on three public multivariate time-series anomaly detection datasets. The results indicate that the improved model utilizing SGC performs comparably to MEGA, yet in certain scenarios, it may provide slightly better outcomes.

Keywords: Anomaly detection · discrete wavelet transform (DWT) · simple graph convolution(SGC) · multivariate time series.

1 Introduction

In time-series anomaly detection, identifying outliers from normal data distributions has gained increasing attention from academia and industry. Multivariate time series, which record multiple system indicators, are crucial for applications [24] like system monitoring and troubleshooting. In industrial environments [23], with numerous operational indicators generated continuously, manual monitoring is impractical, making automatic anomaly detection essential.

Traditional machine learning methods like KNN [2] and One-class SVM [13] have been proposed but struggle with high-dimensional and complex data. Recently, deep learning approaches, particularly deep autoencoders (AEs) [16], have been explored for unsupervised anomaly detection, which model temporal dependence and intervariable dependence to reconstruct time series for anomaly detection [18]. Although these reconstruction-based deep models have achieved good performance in detecting a distinct anomaly, that is, obviously deviated from normal patterns, they fail to detect subtle contextual anomalies that behave normally compared to their neighbors.

Moreover, system-level anomalies often involve inter-variable dependencies [26], which are not adequately captured by directly modeling the original multivariate time series. These time series consist of oscillations at multiple scales, leading to varying inter-variable dependencies.

This paper introduces an enhanced multivariate time-series anomaly detection framework using SGC [22]. SGC reduces training time, computational resources, and model complexity while better capturing long-term dependencies compared to traditional GCN [7]. Our contributions are:

1.Dynamic Graph Structures with SGC:After decomposing the time series into multifrequency components, we construct dynamic graph structures at each scale, using the frequency components as node features.We employ SGC to capture the complex inter-variable dependencies at different scales. SGC's spectral methods provide a more efficient and effective way to process these dependencies compared to GCN.

2.Extensive experiments on three public multivariate time-series anomaly detection datasets demonstrate that our model maintains good performance while significantly reducing computational complexity and training time.

2 Related Work

2.1 Multivariate Time-Series Anomaly Detection

Recent research on anomaly detection in multivariate time series has focused on modeling temporal and variable dependencies.Many approaches use AEs architectures for anomaly detection. OmniAnomaly [18] uses VAE-based networks to capture time dependencies and stochasticity in latent space. USAD [1] introduces adversarial AEs to model time representations and combines generator and discriminator scores for anomaly detection. RAMED [17] addresses temporal error accumulation using a multiresolution ensemble decoder, while MemAE [4] adds a memory module to enhance AE's generalization and detection capability. In modeling variable dependencies, MTAD-GAT [26] applies attention mechanisms for temporal and spatial anomaly detection, and GDN [3] uses graph networks to represent variable relationships and uses a prediction-based approach to accomplish anomaly detection. AddGraph [27] integrates dynamic graph structures with GRUs to capture both short-term and long-term patterns, while MSCRED [25] models correlations through system signature matrices. MTS-DCGAN [11] combines sliding windows and forgetting mechanisms in the anomaly detection phase to focus on the contribution of samples at different distances and achieve good results. CCG-EDGAN [10] combines cross-correlation graphs and encoder–decoder GAN to learn sequential correlation features among multiple time series and achieve good performance. MEGA [20] combines DWT, AE, and GCN to better detect subtle anomalies and capture multiscale dependencies.

2.2 DWT in Neural Networks

In signal processing, signals like temperature or KPIs are composed of different frequency components.Researchers have recently combined DWT with neural networks to capture multiscale frequency representations. For example, mWDN [21] uses multilevel DWT within deep neural networks for improved frequency learning in time-series analysis. DWT has also been applied successfully in computer vision tasks due to its noise robustness and efficiency. WaveCNet [9] integrates wavelets with CNNs for noise-resistant image classification, MW-GAN [19] enhances video quality using wavelets and GANs, WaveletMonoDepth [15] applies wavelet decomposition for depth prediction, and STMFANet [6] combines wavelets with spatial-temporal networks for video prediction.

2.3 GCNs

GCNs have become a powerful tool for learning on graph-structured data by aggregating information from neighboring nodes through graph convolutions. Traditional GCNs, introduced by Kipf and Welling [7], operate in the spectral domain and excel in tasks like semi-supervised node classification. However, issues like oversmoothing and high computational cost have led to various improvements. SGC [22] simplifies GCNs by removing nonlinearities and collapsing weight matrices, improving efficiency and reducing oversmoothing. GraphSAGE [8] introduces inductive learning to handle dynamic graphs. GCNs are now being applied to time-series anomaly detection. GDN [2] models variable relationships using graph networks, while MTAD-GAT [13] integrates attention mechanisms to capture both temporal and variable dependencies. Our work leverages SGC within an autoencoder framework to enhance detection of subtle dependencies and anomalies in multivariate time series.

3 Methodology

3.1 Overview

Our proposed methodology enhances the MEGA framework by integrating SGC to improve anomaly detection capabilities in multivariate time series. This approach capitalizes on the strengths of SGC to simplify the model architecture while retaining effectiveness.The specific structure is shown in Figure 1.

3.2 Multiscale Discrete Wavelet Decomposition

In the DWT segment of our methodology, we employ a multilevel DWT to decompose the original multivariate time series into a set of frequency components

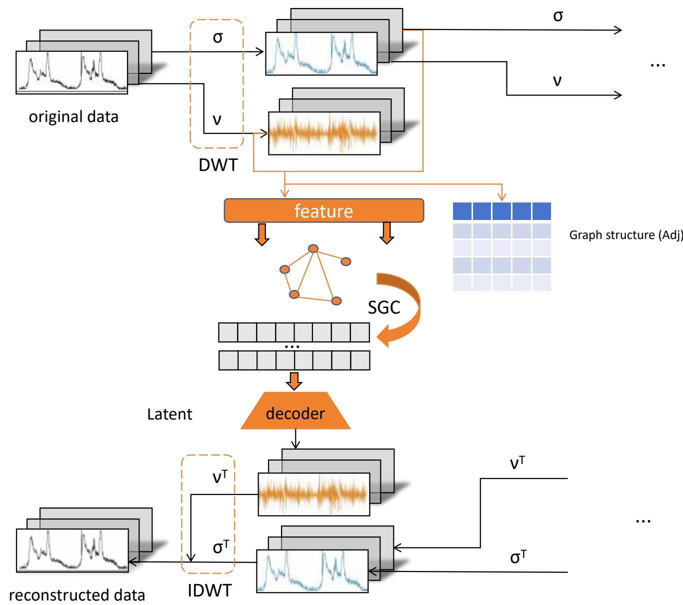

Fig. 1. This is a schematic of one layer of the model in this paper, mainly consisting of three parts, and the other two layers have the same structure as in the schematic.

that represent different scales within the data. This process is crucial for detecting anomalies that may manifest at various frequency bands.

The DWT uses a pair of filters—a low-pass filter (σ) and a high-pass filter (ν)—to decompose the time series into low-frequency components and high-frequency components. The decomposition is performed iteratively, with each iteration halving the length of the time series while doubling the number of frequency components.

The mathematical representation of the DWT process is as follows:

$$f_l^{i+1} = \sigma^i f_l^i \tag{1}$$

$$f_h^{i+1} = \nu^i f_h^i \tag{2}$$

where f_l^i represents the low-frequency components at the i-th level, f_h^i represents the high-frequency components at the i-th level.

We then feed the multiscale frequency components into the encoder to extract their latent representations. These representations are subsequently used by the decoder to regenerate the frequency components. Finally, we reconstruct the original time series using the IDWT.

3.3 Graph Structure Learning

After decomposing the multivariate time series into multiple frequency components using DWT, we construct a dynamic graph structure at each scale. The relationships between variables are modeled as edges, and the characteristics of each variable are represented as nodes. To efficiently capture these dependencies, we employ the SGC, which simplifies the traditional GCN by removing the nonlinearity and collapsing weight matrices.

The adjacency matrix A is defined as:

$$A = \text{LeakyReLU}\left(\tanh\left(\kappa\left(E_1 E_2^T - E_2 E_1^T\right)\right)\right) \tag{3}$$

where E_1 and E_2 are embeddings learned adaptively during training, and κ is a scaling factor. This formulation ensures the asymmetry of the adjacency matrix, reflecting the unidirectional impact of anomalies among variables.

Given the node features f_i from the DWT, the SGC operation can be expressed as:

$$f_i' = \tilde{A} f_i \tag{4}$$

where $\tilde{A} = \tilde{D}^{-1/2} A \tilde{D}^{-1/2}$ and $\tilde{D}_{ii} = 1 + \sum_j A_{ij}$ is the degree matrix. Here, f_i' represents the updated node features after the SGC layer, capturing the aggregated information from neighboring nodes.

We incorporate the frequency information of the time series into the process of SGC to obtain the latent representation. When anomalies occur, the frequency of the series changes, and the variable dependence of the graph convolution output is affected, which are helpful for anomaly detection. The entire encoding process can be formalized as follows:

$$z = E(f_l, f_h) \tag{5}$$

where E is the encoder function that maps the low-frequency components f_l and high-frequency components f_h to the latent representation z.

3.4 Frequency Generator and IDWT

Once we have the multiscale representation of the latent space, we utilize multiscale frequency generation and synthesis to reassemble the original time series.

Frequency Generator

The frequency generator uses the encoded latent representations z to generate the multiscale frequency components. The generation process can be expressed as:

$$f_l^{i+1} = \sigma^i z \tag{6}$$

$$f_h^{i+1} = \nu^i z \tag{7}$$

where f_l^{i+1} and f_h^{i+1} are the low-frequency and high-frequency components at scale, σ^i and ν^i are the learned low-pass and high-pass filters.

IDWT

The IDWT is used to reconstruct the original time series from the generated multiscale frequency components. The reconstruction process combines the low-frequency and high-frequency components at each scale and can be expressed as:

$$f_l^i = \mathrm{IDWT}(f_l^{i+1}, f_h^{i+1}) \tag{8}$$

where f_l^i is the reconstructed low-frequency component at scale i , IDWT is the inverse discrete wavelet transform operation that combines the low-frequency and high-frequency components to reconstruct the time series at the previous scale.

By integrating the frequency generator and IDWT, the MEGA framework effectively reconstructs the original time series, enabling the detection of anomalies through the comparison of the reconstructed and original time series.

3.5 Anomaly Detection

In the test phase, the loss obtained by feeding samples into the trained model is used as the score for anomaly determination. The loss function can be expressed as:

$$L = \alpha\|X-X'\|_2^2+\beta\|f_{1h}-f'_{1h}\|_2^2+\gamma\|f_{2h}-f'_{2h}\|_2^2+\delta\|f_{3l}-f'_{3l}\|_2^2+\lambda\|f_{3h}-f'_{3h}\|_2^2 \tag{9}$$

where $\alpha, \beta, \gamma, \delta$, and λ are the weighting coefficients that can be used to adjust the attention for different scale frequencies. X and X' are the original and reconstructed time series, respectively. $f_{1h}, f'_{1h}, f_{2h}, f'_{2h}, f_{3l}, f'_{3l}, f_{3h}$, and f'_{3h} are the high-frequency and low-frequency components at different scales.

If we want to place more emphasis on the reconstruction of frequency components at different scales in the anomaly scores, we can increase the corresponding loss weights during the training phase. In the final step, an anomaly threshold is set to classify samples with scores above the threshold as abnormal and those below the threshold as normal, thereby completing the anomaly detection.

4 Experiment

4.1 Datasets and Evaluation Metrics

In this article, we use three public datasets of multivariate time-series anomaly detection to test the effectiveness of our model, including two real-world datasets Mars Science Laboratory rover (MSL), Soil Moisture Active Passive satellite (SMAP) [5] collected from NASA, each of which has a training and testing datasets. Anomalies in both testing subsets have been labeled. And a five-week-long dataset was collected from a large Internet company, server machine dataset (SMD) [18] which is divided into two subsets of equal size: a training set and a testing set. MSL contains 27 entities, whose dimension is 55, while SMAP contains 55 entities, whose dimension is 25. SMD has 28 entities with 38 metrics of each. The detailed information on these datasets is shown in Table 1.

Table 1. Details of the Datasets

Dataset	Train	Test	Dimensions	Anomalies (%)
MSL	58317	73729	27×55	10.72
SMAP	135183	427617	55×25	13.13
SMD	708405	708420	28×38	4.16

In order to compare the performance between our model and other baselines, we use the same data preprocessing method and evaluation metrics as in previous works [18], [1] including

$$F1 = \frac{2 \times P \times R}{P + R} \tag{10}$$

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

We do not focus on the specific strategy for selecting the anomaly threshold. If a model does not provide a method to select thresholds, we use a brute force search to find the best F1 scores, ensuring a uniform threshold search strategy. To maintain consistency in the experimental settings, we apply the point-adjust strategy: for a segment anomaly, any subset that triggers an alert is considered acceptable. Therefore, if the model detects any observation as an anomaly within a ground-truth anomaly segment, we assume the entire segment is correctly detected. For general point anomalies, no adjustment is made.

4.2 Baseline Methods

We compare our model with five unsupervised anomaly detection methods to demonstrate the performance of our model including: IF [12] is a classical anomaly detection algorithm in machine learning that exploits the assumption of low data density at outlier points for data partitioning; AE is the most fundamental reconstruction-based deep model in the field of anomaly detection; DAGMM [28] combines AE and the Gaussian mixture model (GMM) to estimate the density of the representations in the latent space; LSTM-VAE [14] replaces the feed-forward neural network in VAE with an LSTM to capture the temporal dependence; MEGA [3] have been introduced in the related work.

4.3 Experimental Results

Table 2 shows the overall performance of our model and the other baselines. As shown in the table, MEGA and Ours, which use multiscale frequency modeling, achieve better results in accuracy, the metric evaluated for the anomaly detection task.

Of these models, IF had the worst results.IF attempts to distinguish abnormal data from normal data in the raw feature space, but it struggles with high-dimensional, complex data distributions. DAGMM also performs relatively poorly compared to other benchmarks due to the lack of time-dependent modeling. Since anomalies in time-series data are usually caused by temporal variations, multiscale frequency modeling can capture time dependence more effectively from multiple perspectives.

AE and LSTM-VAE are trained to utilize temporal information in a reconstructive manner. However, they have higher false negatives because their reconstructions are performed only in the original space without proper constraints in the latent space. These models do not specialize in subtle pattern anomalies and do not take into account anomalous relationships between variables, hence their limited performance.

In contrast, Our SGC-based models utilize both temporal and spatial relationships and are highly competitive in terms of performance; MEGA's GCN integrates multi-scale frequency information to capture spatial anomalies, while SGC in Ours simplifies the graph learning process by reducing computational overhead and removing non-critical operations. This allows Ours to improve its efficiency while maintaining strong performance, especially on the MSL dataset, where it even outperforms MEGA in terms of F1 score. although Ours is slightly inferior to MEGA in terms of overall F1 score, the trade-off in computational efficiency makes it a practical alternative.

The effective utilization of multi-scale frequency information and spatial anomaly detection using graph networks is responsible for MEGA's excellent performance. However, our model provides a more computationally efficient model that strikes a balance between performance and resource utilization, especially in high-dimensional time series anomaly detection tasks.

Table 2. Performance of different models

Model	**MSL**			**SMAP**			**SMD**		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
IF	0.5681	0.6740	0.5984	0.4423	0.5105	0.4671	0.5938	0.8532	0.5866
AE	0.8535	0.9748	0.8792	0.7216	0.9795	0.7776	0.8825	0.8037	0.8280
DAGMM	0.7562	0.9803	0.8112	0.6334	0.9984	0.7124	0.6730	0.8450	0.7231
LSTM-VAE	0.8599	0.9756	0.8537	0.7164	0.9875	0.7555	0.8698	0.7879	0.8083
MEGA	0.8561	0.8223	0.8388	0.9694	0.5564	0.7070	0.9835	0.9992	0.9913
Ours	0.8689	0.8118	0.8394	0.9300	0.5611	0.6999	0.8707	0.9974	0.9297

5 Conclusion

In this paper, we introduce an unsupervised multivariate time-series anomaly detection framework based on multiscale frequency decomposition and generation. We employ SGC instead of traditional Graph Convolutional Networks, which significantly reduces training time and computational resource requirements. Experiments on three representative datasets (MSL, SMD, and SMAP) demonstrate that our model can maintain or even improve accuracy while reducing model complexity.

References

1. Audibert, J., Michiardi, P., Guyard, F., Marti, S., Zuluaga, M.A.: Usad: Unsupervised anomaly detection on multivariate time series. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 3395–3404 (2020)
2. Chaovalitwongse, W.A., Fan, Y.J., Sachdeo, R.C.: On the time series k-nearest neighbor classification of abnormal brain activity. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **37**(6), 1005–1016 (2007)
3. Deng, A., Hooi, B.: Graph neural network-based anomaly detection in multivariate time series. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 4027–4035 (2021)
4. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., Hengel, A.v.d.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1705–1714 (2019)
5. Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T.: Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 387–395 (2018)
6. Jin, B., Hu, Y., Tang, Q., Niu, J., Shi, Z., Han, Y., Li, X.: Exploring spatial-temporal multi-frequency analysis for high-fidelity and temporal-consistency video prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4554–4563 (2020)
7. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
8. Li, C., Guo, L., Gao, H., Li, Y.: Similarity-measured isolation forest: Anomaly detection method for machine monitoring data. IEEE Transactions on Instrumentation and Measurement **70**, 1–12 (2021)
9. Li, Q., Shen, L., Guo, S., Lai, Z.: Wavelet integrated cnns for noise-robust image classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7245–7254 (2020)
10. Liang, H., Song, L., Du, J., Li, X., Guo, L.: Consistent anomaly detection and localization of multivariate time series via cross-correlation graph-based encoder–decoder gan. IEEE Transactions on Instrumentation and Measurement **71**, 1–10 (2021)
11. Liang, H., Song, L., Wang, J., Guo, L., Li, X., Liang, J.: Robust unsupervised anomaly detection via multi-time scale dcgans with forgetting mechanism for industrial multivariate time series. Neurocomputing **423**, 444–462 (2021)

12. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: 2008 eighth ieee international conference on data mining. pp. 413–422. IEEE (2008)
13. Ma, J., Perkins, S.: Time-series novelty detection using one-class support vector machines. In: Proceedings of the International Joint Conference on Neural Networks, 2003. vol. 3, pp. 1741–1745. IEEE (2003)
14. Park, D., Hoshi, Y., Kemp, C.C.: A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. IEEE Robotics and Automation Letters **3**(3), 1544–1551 (2018)
15. Ramamonjisoa, M., Firman, M., Watson, J., Lepetit, V., Turmukhambetov, D.: Single image depth prediction with wavelet decomposition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11089–11098 (2021)
16. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representations by error propagation, parallel distributed processing, explorations in the microstructure of cognition, ed. de rumelhart and j. mcclelland. vol. 1. 1986. Biometrika **71**(599-607), 6 (1986)
17. Shen, L., Yu, Z., Ma, Q., Kwok, J.T.: Time series anomaly detection with multiresolution ensemble decoding. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 9567–9575 (2021)
18. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D.: Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2828–2837 (2019)
19. Wang, J., Deng, X., Xu, M., Chen, C., Song, Y.: Multi-level wavelet-based generative adversarial network for perceptual quality enhancement of compressed video. In: European Conference on Computer Vision. pp. 405–421. Springer (2020)
20. Wang, J., Shao, S., Bai, Y., Deng, J., Lin, Y.: Multiscale wavelet graph autoencoder for multivariate time-series anomaly detection. IEEE Transactions on Instrumentation and Measurement **72**, 1–11 (2022)
21. Wang, J., Wang, Z., Li, J., Wu, J.: Multilevel wavelet decomposition network for interpretable time series analysis. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2437–2446 (2018)
22. Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., Weinberger, K.: Simplifying graph convolutional networks. In: International conference on machine learning. pp. 6861–6871. PMLR (2019)
23. Yu, J., Song, Y., Tang, D., Han, D., Dai, J.: Telemetry data-based spacecraft anomaly detection with spatial–temporal generative adversarial networks. IEEE Transactions on Instrumentation and Measurement **70**, 1–9 (2021)
24. Zeng, Z., Jin, G., Xu, C., Chen, S., Zeng, Z., Zhang, L.: Satellite telemetry data anomaly detection using causal network and feature-attention-based lstm. IEEE Transactions on Instrumentation and Measurement **71**, 1–21 (2022)
25. Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H., Chawla, N.V.: A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 1409–1416 (2019)
26. Zhao, H., Wang, Y., Duan, J., Huang, C., Cao, D., Tong, Y., Xu, B., Bai, J., Tong, J., Zhang, Q.: Multivariate time-series anomaly detection via graph attention network. In: 2020 IEEE international conference on data mining (ICDM). pp. 841–850. IEEE (2020)

27. Zheng, L., Li, Z., Li, J., Li, Z., Gao, J.: Addgraph: Anomaly detection in dynamic graph using attention-based temporal gcn. In: IJCAI. vol. 3, p. 7 (2019)
28. Zong, B., Song, Q., Min, M.R., Cheng, W., Lumezanu, C., Cho, D., Chen, H.: Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In: International conference on learning representations (2018)