mathematics

MDPI

Article

An Anomaly Detection Method for Multivariate Time Series Data Based on Variational Autoencoders and Association Discrepancy

Haodong Wang and Huaxiong Zhang *

School of Computer Science and Technology (School of Artificial Intelligence), Zhejiang Sci-Tech University, Hangzhou 310018, China; 202230603115@mails.zstu.edu.cn
* Correspondence: zhxhz@zstu.edu.cn

Abstract: Driven by rapid advancements in big data and Internet of Things (IoT) technologies, time series data are now extensively utilized across diverse industrial sectors. The precise identification of anomalies in time series data—especially within intricate and ever-changing environments—has emerged as a key focus in contemporary research. This paper proposes a multivariate anomaly detection framework that synergistically combines variational autoencoders with association discrepancy analysis. By incorporating prior knowledge of associations and sequence association mechanisms, the model can capture long-term dependencies in time series and effectively model the association discrepancy between different time points. Through reconstructing time series data, the model enhances the distinction between normal and anomalous points, learning the association discrepancy during reconstruction to strengthen its ability to identify anomalies. By combining reconstruction errors and association discrepancy, the model achieves more accurate anomaly detection. Extensive experimental validation demonstrates that the proposed methodological framework achieves statistically significant improvements over existing benchmarks, attaining superior F1 scores across diverse public datasets. Notably, it exhibits enhanced capability in modeling temporal dependencies and identifying nuanced anomaly patterns. This work establishes a novel paradigm for time series anomaly detection with profound theoretical implications and practical implementations.

Keywords: time series anomaly detection; deep learning; association discrepancy; variational autoencoder; multivariate time series

MSC: 37M10

check for updates

Academic Editor: Yanjie Fu

Received: 8 March 2025
Revised: 2 April 2025
Accepted: 5 April 2025
Published: 7 April 2025

Citation: Wang, H.; Zhang, H. An Anomaly Detection Method for Multivariate Time Series Data Based on Variational Autoencoders and Association Discrepancy. *Mathematics* **2025**, *13*, 1209. https://doi.org/10.3390/math13071209

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Driven by rapid advancements in industrial technology, time series anomaly detection has been extensively applied across diverse domains. Key applications include industrial equipment monitoring, spacecraft status monitoring, network security, and financial fraud detection [1]. The primary objective of this task is to identify anomalous instances that significantly differ from the majority of data points. This process is crucial for improving operational efficiency, reducing costs, and enhancing public safety. However, the imbalance between normal and anomalous samples makes it extremely challenging to obtain anomaly labels, leading most existing algorithms to rely on unsupervised learning methods for anomaly detection.

Although deep learning techniques have advanced considerably in time series anomaly detection over recent years, current methods continue to confront several challenges. For example, prediction-based anomaly detection methods, such as Long Short-Term Memory (LSTM) networks are prone to gradient explosion problems when handling long time series, and their sequential processing mechanism results in low detection efficiency. Additionally, reconstruction-based anomaly detection methods like variational autoencoders (VAEs) often neglect key features such as the sequential relationships and time intervals between data points during time series reconstruction. This impedes the models' ability to thoroughly capture the fundamental dynamic patterns within the sequences.

In order to overcome these obstacles, this study introduces a multivariate time series anomaly detection model based on variational autoencoders and association discrepancy —VAE-Anomaly. The model utilizes the self-attention mechanism to effectively capture temporal dependencies in multivariate time series, with a focus on modeling the prior associations and sequence associations of the data. Specifically, sequence association represents the distribution of relationships between time points in a time series, providing rich contextual information for capturing dynamic features (such as periodicity and trend). Prior association reflects the potential strong correlations between adjacent or nearby time points based on prior assumptions of their relative positions in the time series [2].

By incorporating prior associations, the model can predefine potential time-point dependencies without encountering actual data. This mechanism is particularly important for anomaly detection, as anomalous data, due to their sparsity, typically have weak correlations with other time points, and this weak correlation is primarily concentrated between adjacent time points. Based on this feature, this paper uses association discrepancy as an important metric for anomaly detection, which quantifies the difference between prior associations and sequence associations. Experimental results show that anomalous points usually have smaller association discrepancies, meaning their prior associations are closer to their sequence associations.

To enhance the model's anomaly detection capability, this paper combines variational autoencoders (VAEs) and utilizes the Evidence Lower Bound (ELBO) to estimate the probability distribution of input data. During reconstruction, the model improves its ability to identify anomalies by learning the association discrepancies of the reconstructed data. Since anomalous data typically generate larger reconstruction errors, reconstruction error is used as another critical criterion for anomaly detection. The latent variable prior distribution introduced by VAEs further enhances the model's robustness.

In terms of anomaly score thresholds, this paper adopts an Extreme Value Theory (EVT)-based Peak Over Threshold (POT) model [3] to automatically set detection thresholds. The experimental results demonstrate that the proposed model surpasses existing baseline methods on four public datasets and performs second-best on another dataset.

The main contributions of this article are as follows:

The first is the design of a triple-weighted loss function integrating reconstruction error, KL divergence, and association discrepancy loss to optimize model parameters. This unified framework combines association discrepancy with reconstruction error for anomaly score calculation, enabling adaptive threshold determination to detect temporal anomalies.

The second is enhanced anomaly discriminability through association discrepancy learning on reconstructed time series data. By reconstructing input sequences, the disparity between normal and anomalous points becomes amplified in the latent representation space, thereby strengthening the model's capacity to identify subtle yet critical deviations through association discrepancy analysis.

2. Related Work

Research on time series anomaly detection has attracted considerable attention from scholars, focusing on developing efficient and accurate detection methods to address challenges in various practical applications. Currently, the main research directions include anomaly detection methods based on statistical models, machine learning models, and deep learning models.

Traditional statistical methods are widely used in time series anomaly detection. For instance, the Autoregressive Integrated Moving Average (ARIMA) model [4] models the autocorrelation, trend, and seasonality in time series to predict future values and detect anomalies based on the predicted results. The ARIMA model is well suited for stationary and periodic data, effectively capturing linear relationships within the dataset. However, ARIMA needs high stationarity in the data and has limited adaptability to complex nonlinear relationships and abrupt changes, thus performing poorly when confronted with complex dynamic patterns. Additionally, methods such as exponential smoothing and the autoregressive moving average (ARMA) were proposed for anomaly detection [5,6].

With the advancement of machine learning, new approaches have provided more solutions for time series anomaly detection. A notable example is the Isolation Forest [7] proposed by Liu et al., which identifies anomalies by isolating data points based on the difficulty of their isolation.

With advancements in deep learning, neural-network-based approaches have achieved significant progress in time series anomaly detection. These include recurrent neural network (RNN) approaches exemplified by DeepAR [8] and LSTNet [9]; convolutional architectures including TCN [10] and SCINet [11]; Transformer-based innovations such as Informer [12] with ProbSparse attention and its variants Autoformer [13] leveraging seasonal-trend decomposition, Fedformer [14], and PatchTST [15]; MLP-driven frameworks featuring N-BEATS [16], which employ backward residual links for interpretable decomposition; and linear-based models (NLinear [17], DLinear [17]), which simplify temporal projection through channel-independent weighting [18–21]. Park et al. [22] proposed LSTM-VAE, where LSTM captures the spatiotemporal characteristics of the sequence, while VAE learns the latent representation of the data to reconstruct the original sequence. Anomalies are identified when the reconstruction error surpasses a predefined threshold. Su et al. [23] used SRNN, a variant of RNN, which introduces stochastic modeling, making the model more robust when handling multivariate time series data and aiding in capturing complex dynamic features. Zong et al. [24] pioneered a dual-pathway architecture integrating deep autoencoders with Gaussian mixture modeling (DAGMM) for unsupervised anomaly detection. The VRNN framework [25] implements a temporal variational inference mechanism that computes posterior approximations by contextually combining current observational data with preceding hidden states in recurrent computations. VRNN captures the temporal dependencies between random variables through iterative updates of the RNN hidden vector. The generative network uses the sampled value of the current random variable along with the previous RNN hidden state to generate the value distribution of time series data.

In summary, different time series anomaly detection methods have their strengths and limitations. Traditional statistical methods excel in handling linear and periodic data but perform poorly with nonlinear and complex dynamics. Machine learning methods can handle more complex data relationships but often face issues such as kernel function selection and overfitting. Deep learning methods have high computational complexity and a weak ability to extract local patterns.

3. Materials and Methods

3.1. Problem Definition

Time series data typically consist of values read from multiple sensors, the status of actuators, and control system commands, forming a multidimensional time series dataset $X_{1:T} = \{x_1, x_2, \cdots\cdots x_T\}$, where $x_T \in \mathbb{R}^D$ represents the D-dimensional measurements collected at time point T. For each individual data point in a dimension, it is denoted as $x_{i,t}$, where $x_{i,t} \in \mathbb{R}$, $i \in \{1, 2, \cdots\cdots, D\}$, $t \in \{1, 2, \cdots\cdots, T\}$ represents each time step in the time series.

In practical applications, anomaly detection often requires the introduction of label variables y_t, where $y_t \in \{0, 1\}$ represents the data state at time point t. If $y_t = 1$, it indicates that the measurement data at time point t are anomalous; if $y_t = 0$, it indicates that the data at that time point are normal.

Since this study assumes that the training dataset does not include anomaly labels, an unsupervised learning approach is used. The objective is to train an anomaly detection model by learning patterns from normal time series data. This model is then used to evaluate and determine whether the observed data in the test set exhibit anomalous behavior.

3.2. Network Architecture

3.2.1. VAE-Anomaly

The workflow of the model proposed in this paper is illustrated in Figure 1. In this model, the initial step involves capturing both long-term and short-term temporal dependencies in multivariate time series data through the association discrepancy layer, which models the series association and prior association. This allows the model to learn the association discrepancies of the raw time series data. Next, the feature vector output from the association discrepancy layer is mapped to the latent space via the reconstruction layer to capture its latent features. In the latent space, random sampling is performed using the reparameterization technique to reconstruct the time series data.

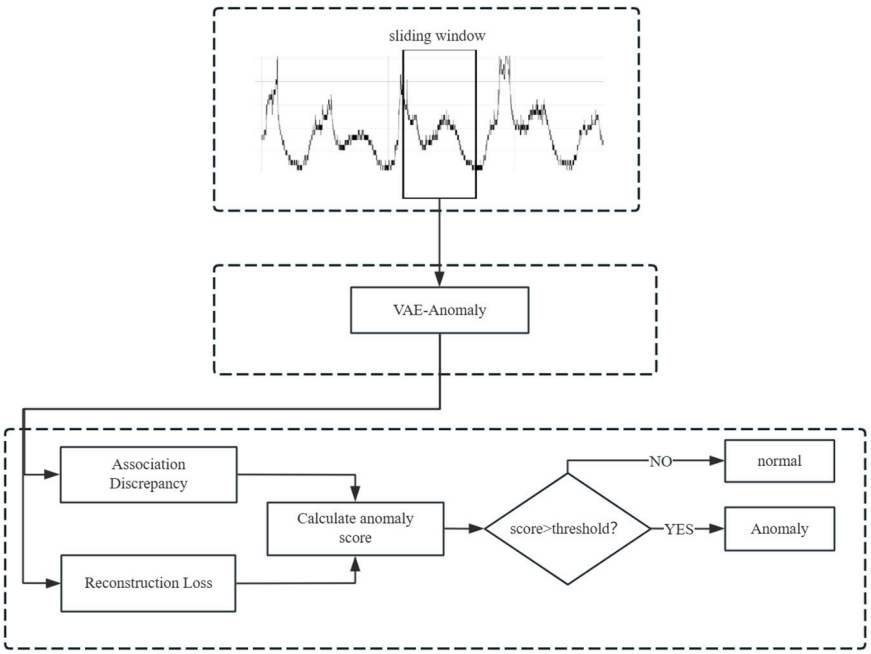


Figure 1. Workflow of VAE-Anomaly.

After reconstruction, the differences between anomaly and normal points in the time series data are further magnified, allowing the model to enhance its ability to distinguish anomalies by learning the association discrepancies from the reconstructed time series

data. Moreover, the randomness introduced during the VAE reconstruction process greatly improves the model's robustness and performance.

Subsequently, the model combines reconstruction loss and association discrepancy to identify anomalous points. The reconstruction loss reflects the model's effectiveness in reconstructing the time series, while the association discrepancy helps differentiate between normal and anomalous data points. The detection phase culminates in computing anomaly likelihood metrics for each temporal unit, followed by binary classification against statistically validated thresholds derived from training distributions. The structure of the VAE-Anomaly model is shown in Figure 2.

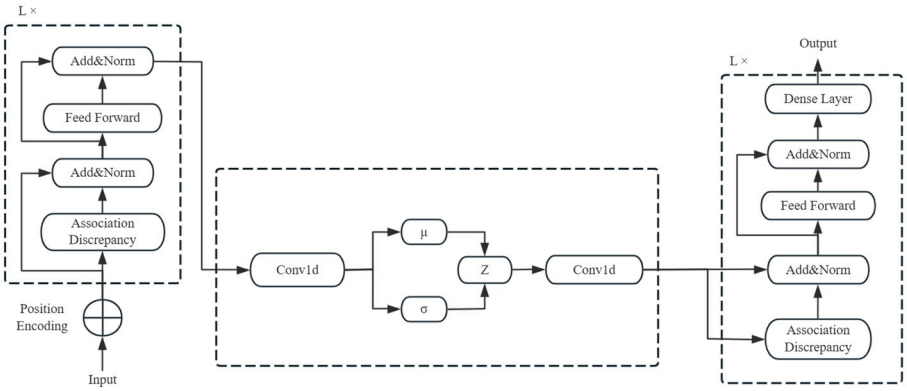


Figure 2. Model structure of VAE-Anomaly.

3.2.2. Association Discrepancy Layer

For multivariate time series input, the input vector is first element-wise added to a position encoding vector derived from sine and cosine functions, enabling the encoding of sequential information within the data. This method allows the model to capture the relative positions of elements within the sequence. The final input representation, denoted as X, incorporates both the temporal information of each time point and the encoded positional information of the sequence.

Assume that the association discrepancy layer contains L layers, and the input sequence is $X \in \mathbb{R}^{N \times D}$, where N is the length of the sequence and D is the feature dimension at each time point. The operation of the l-th layer can be expressed as

$$Z^l = \text{Norm}\left(\text{Association} - \text{Discrepancy}\left(X^{l-1}\right) + X^{l-1}\right), \tag{1}$$

$$X^l = \text{Norm}\left(\text{Feed} - \text{Forward}\left(Z^l\right) + Z^l\right) \tag{2}$$

where $X^l \in \mathbb{R}^{N \times d_{model}}$ and $l \in \{1, \cdots, L\}$ represent the output of the l-th layer, and d_{model} is the feature dimension of each layer. X^0 is the input X after position encoding, and $Z^l \in \mathbb{R}^{N \times d_{model}}$ is the intermediate state of the l-th layer. The association discrepancy module (ADM) is used to calculate the association discrepancy.

In the ADM, four important matrices are defined: the query vector matrix Q, the key vector matrix K, the value vector matrix V, and the scaling parameter matrix σ. The computation of these matrices is as follows:

$$Q = X^{l-1} W_Q^l, \tag{3}$$

$$K = X^{l-1} W_K^l, \tag{4}$$

$$V = X^{l-1} W_V^l, \tag{5}$$

$$\sigma = X^{l-1} W_\sigma^l \tag{6}$$

where $Q, K, V \in \mathbb{R}^{N \times d_{model}}$ and $\sigma \in \mathbb{R}^{N \times 1}$, W_Q^l, W_K^l, W_V^l are the parameter matrices of the l-th layer, belonging to $\mathbb{R}^{d_{model} \times d_{model}}$, while $W_\sigma^l \in \mathbb{R}^{d_{model} \times 1}$.

For the prior association matrix P^l of the l-th layer, the computation formula is

$$P^l = \text{Rescale}\left(\left[\frac{1}{\sqrt{2\pi}\sigma_i} exp\left(-\frac{|j-i|^2}{2\sigma_i^2}\right)\right]_{i,j\in\{1,\cdots,N\}}\right) \tag{7}$$

where the prior association matrix $P^l \in \mathbb{R}^{N \times N}$ is constructed using the learned scaling parameter $\sigma \in \mathbb{R}^{N \times 1}$, with the i-th element σ_i corresponding to the scaling value at the i-th time point. With the application of the Rescale(.) operation, the obtained association weights are divided by the sum of each row, converting them into a normalized discrete distribution P^l, such that the sum of the elements in each row equals 1.

At the same time, the sequence association matrix S^l is computed by calculating the inner product of the query vector Q and the key vector K and normalizing it using the Softmax function, as shown in the following formula:

$$S^l = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_{model}}}\right) \tag{8}$$

where $S^l \in \mathbb{R}^{N \times N}$ represents the sequence association matrix, and the Softmax operation is performed along the last dimension for normalization. Therefore, each row of S^l forms a discrete distribution, representing the association weights between each time point and the other time points.

By multiplying S^l with the value vector matrix V, the output feature vector e^l of the layer can be obtained:

$$e^l = S^l V \tag{9}$$

where $e^l \in \mathbb{R}^{N \times d_{model}}$ represents the feature representation of the input time series after deep encoding, capturing both contextual information and long-term dependencies.

By alternating the stacking of the ADM and feedforward layers, the encoder is able to learn underlying association patterns from deep and multi-level features. The association discrepancy (AssDis) represents the symmetric KL divergence between the prior association and sequence association. The overall association discrepancy is computed by averaging the discrepancies across multiple layers. Specifically, the formula for calculating the association discrepancy is

$$\text{AssDis}(P, S; X) = \left[\frac{1}{L}\sum_{l=1}^{L}\left(KL\left(P_{i,:}^l \| S_{i,:}^l\right) + KL\left(S_{i,:}^l \| P_{i,:}^l\right)\right)\right]_{i=1,\cdots,N} \tag{10}$$

where $KL(.)$ represents the Kullback–Leibler (KL) divergence. For each layer l, P^l and S^l are the prior association and sequence association matrices, and $P_{i,:}^l$ and $S_{i,:}^l$ represent the distributions of the i-th row, respectively. Finally, $\text{AssDis}(\text{P}, \text{S}; \text{X}) \in \mathbb{R}^{N \times 1}$ is a vector containing N elements, where each element represents the association discrepancy at the corresponding time point in the time series.

Based on the above formula, it can be inferred that for normal data points, the association discrepancy values are typically large, while for anomalous data points, the association discrepancy is usually smaller. This is because the association patterns between anomalies and normal points are typically quite different, leading to smaller differences between their prior association and sequence association distributions, which results in a lower association discrepancy.

3.2.3. Reconstruction Layer

The layer processes feature encodings e^l from the association discrepancy layer through convolutional operations to parameterize the variational distribution $q(Z|X)$, estimating posterior parameters (μ, σ^2) for latent variable Z while minimizing KL divergence between the variational approximation and target posterior $p(Z|X)$.

This process can be viewed as searching for a regular path in the high-dimensional data space, such that different feature vectors have a structured distribution in the latent space. Finally, the latent variable Z is sampled from $q(Z|X)$ using the reparameterization trick. Specifically, the reparameterization trick involves sampling a noise vector ϵ from a standard normal distribution and then combining it with the mean and standard deviation to generate the latent vector:

$$z = \mu + \epsilon \times \sigma \tag{11}$$

where $\epsilon \sim N(0, 1)$ represents the noise vector sampled from the standard normal distribution. This method introduces randomness while ensuring the gradient is computable, allowing the model to explore diverse feature combinations flexibly in the latent space.

After being processed by the reconstruction layer, the feature vectors of normal data form a relatively stable distribution pattern in the latent space. When the data point is anomalous, its feature vector will deviate from the normal distribution. By setting an appropriate threshold, the model can quickly and accurately identify anomalous data points.

3.2.4. Stochastic Association Discrepancy Layer

In the stochastic association discrepancy layer, the model learns from the time series data that have been reconstructed by the reconstruction layer, effectively extracting the association discrepancies between different time points in the series. Since the reconstruction process restores the latent features of the original data while introducing some randomness for each data point, the differences between anomalous and normal data points become more pronounced in the reconstructed data. By calculating these discrepancies, the association discrepancy mechanism can effectively distinguish between anomalous and normal states within a small error margin.

Moreover, the randomness in the reconstruction process not only enhances the model's generalization ability but also, by learning from diverse samples, allows the model to exhibit greater flexibility and robustness when handling different types of anomalies. This mechanism helps the model better capture the data distribution features in the time series, and especially in complex, dynamic environments, it enables the effective identification of potential anomalous patterns, thereby improving anomaly detection accuracy.

The output of the stochastic association discrepancy layer is the reconstructed time series data $\hat{X}$.

3.3. Optimization Objective

In this paper, reconstruction error serves as the objective for model training. The calculation of the reconstruction error combines mean squared error (MSE) and Kullback–Leibler divergence (KLD). The formula for Kullback–Leibler divergence (KLD) is

$$\mathcal{L}_{\mathrm{KLD}} = \frac{1}{2}\sum_{i=1}^{n}\left(\mu_i^2 + \sigma_i^2 - \log\left(\sigma_i^2\right) - 1\right) \tag{12}$$

where n represents the dimension of the latent variable, and μ_i and σ_i^2 represent the i-th component of the mean vector μ and variance vector σ^2 of the approximate posterior distribution. KL divergence (KLD) quantifies the difference between two probability

distributions. By minimizing the KLD, the model encourages the approximate posterior distribution $q(Z|X)$ to align with the true posterior distribution $p(Z|X)$.

The MSE is computed using the following formula:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{13}$$

where n denotes the dimension of the reconstructed time series, and x_i and $\hat{x}_i$ represent the i-th component of the original and reconstructed data. MSE primarily measures the model's reconstruction ability; i.e., it evaluates how well the model restores the input data. During model training, minimizing the MSE helps the decoder accurately reconstruct the original input data from the latent variables output by the encoder, thereby enabling the model to learn an effective feature representation of the data.

This paper applies an additional loss function to enhance the association discrepancy. Since the prior association usually exhibits unimodality, the discrepancy loss guides the sequence association to focus more on non-adjacent regions. For anomalous points, the related Gaussian distance tends to focus on neighboring points, and the attention mechanism reflects this trend, resulting in a smaller association discrepancy (AssDis) for anomalous points. By amplifying the value of AssDis, the model can direct the attention mechanism to focus on more distant points, thereby increasing the reconstruction difficulty for abnormal points and improving their detectability.

The loss function is expressed as

$$\mathcal{L}_{\text{Assdis}}\left(P, S, \lambda; X, \hat{X}\right) = \lambda \times \left(\left\|\text{AssDis}(P, S; X)\right\|_1 + \left\|\text{AssDis}\left(P, S; \hat{X}\right)\right\|_1\right) \tag{14}$$

where $\|.\|_k$ denotes the k-norm, λ is the parameter controlling the switch between minimization and maximization, $\text{AssDis}(P, S; X)$ is the association discrepancy of the original time series data, and $\text{AssDis}\left(P, S; \hat{X}\right)$ is the association discrepancy of the reconstructed time series data. When the model minimizes the loss, the goal is to maximize the association discrepancy. Directly maximizing the association discrepancy can lead to a rapid reduction in the Gaussian kernel's scale parameters, which may render the prior sequence ineffective. Therefore, to better control the association learning, this paper employs a min–max strategy.

Specifically, the optimization process is divided into two stages:

Minimization Stage: In this stage, the prior association P is optimized to make it as close as possible to the sequence association S learned from the original time series. This ensures that the prior association adapts to different time series patterns while preventing the scale parameters from becoming too small, thus maintaining the effectiveness of the prior sequence.

Maximization Stage: In this stage, the sequence association S is optimized to maximize the association discrepancy. This process encourages the sequence association to focus more on non-adjacent points, thereby increasing the difficulty of reconstructing abnormal points and enhancing their detectability.

Since the minimization and maximization stages optimize different objectives, different gradient propagation strategies are employed during the optimization process: the gradient in the minimization stage is propagated to the prior association P, while the gradient in the maximization stage is propagated to the sequence association S. By adjusting the sign of λ, the switch between the minimization and maximization stages is controlled.

The overall optimization objective function is

$$\mathcal{L}_{\text{total}} = (\mathcal{L}_{\text{KLD}} + \mathcal{L}_{\text{MSE}}) + k \times \left(\mathcal{L}_{\text{AssDis}}\left(P, S_{detach}, 1; X, \hat{X}\right) + \mathcal{L}_{\text{AssDis}}\left(P_{detach}, S, -1; X, \hat{X}\right)\right) \tag{15}$$

where k is the weight parameter, $\mathcal{L}_{\text{AssDis}}(P, S_{detach}, 1; X, \hat{X})$ is the loss function for the minimization phase, and $\mathcal{L}_{\text{AssDis}}(P_{detach}, S, -1; X, \hat{X})$ is the loss function for the maximization phase.

3.4. Anomaly Score

The final anomaly score calculation, incorporating the normalized association discrepancy into the reconstruction criterion, is as follows:

$$\text{AnomalyScore}(X, \hat{X}) = \text{Softmax}\left(-\left(\text{AssDis}(P, S; X) + \text{AssDis}(P, S; \hat{X})\right)\right) \bigodot \left(\text{MSE}(X, \hat{X}) + \text{KLD}\right) \tag{16}$$

where the association discrepancy of the original time series data is denoted as $\text{AssDis}(P, S; X)$, and the association discrepancy of the reconstructed time series data is denoted as $\text{AssDis}(P, S; \hat{X})$. The symbol $\odot$ represents element-wise multiplication, while MSE(.) and KLD represent the MSE and KLD at each point. The anomaly score synthesizes association discrepancy metrics with reconstruction errors. Anomalous instances typically exhibit reduced association consistency coupled with amplified reconstruction deviations, resulting in discriminative score elevation compared to normal patterns.

After the anomaly score for each time point is obtained, a threshold must be set to determine whether an anomaly occurs. An anomaly score exceeding this threshold will be classified as anomalous. Setting the threshold involves balancing two factors: suboptimal threshold calibration induces dual detection errors: under-thresholding elevates false positive rates through normal data misclassification, while over-thresholding suppresses anomaly sensitivity by increasing false negative rates.

To resolve this challenge, the proposed methodology leverages Extreme Value Theory (EVT)-based Peak Over Threshold (POT) modeling for automated threshold calibration. The model fits the extreme value probability distribution of the anomaly score sequence, derives the region beyond the threshold, and, following the Pickands theorem, approximates its tail distribution using the Generalized Pareto Distribution (GPD). The specific formula is

$$\overline{\text{F}}(s) = \text{P}(th - S > s | S < th) \sim \left(1 + \frac{\gamma s}{\beta}\right)^{-\frac{1}{\gamma}} \tag{17}$$

where th is the initial threshold for the anomaly score, γ and β parameterize the Generalized Pareto Distribution (GPD) shape and scale characteristics, and S represents any observable in the anomaly score sequence. The part below the threshold th is given by $th - S$ and is determined based on a lower quartile. The estimators $\hat{\gamma}$ and $\hat{\beta}$ are obtained through Maximum Likelihood Estimation (MLE) optimization.

The final threshold th_F is calculated as

$$th_F = th - \frac{\hat{\beta}}{\hat{\gamma}}\left(\left(\frac{qN}{N_{th}}\right)^{-\hat{\gamma}} - 1\right) \tag{18}$$

where q is the observed expected probability $S < th$, N is the sum total of observed data, and N_{th} is the number of time points below the initial threshold th.

After obtaining the final threshold th_F, this paper marks anomalies by checking whether the anomaly score surpasses this value. Specifically, the rule for determining anomalies and normal points is as follows:

$$y_i = \begin{cases} 1: & \text{Anomaly} \quad \text{AnomalyScore}(X_i) \geq th_F \\ 0: & \text{normal} \quad \text{AnomalyScore}(X_i) < th_F \end{cases} \tag{19}$$

4. Experiment

4.1. Datasets

To assess the effectiveness of the proposed anomaly detection model, five representative benchmark datasets from real-world application scenarios were selected for experimentation. The statistical characteristics of the datasets are summarized in Table 1, which include the training set length, testing set length, feature dimensions, and anomaly rates.

Table 1. Details of benchmark datasets.

Dataset	Training	Test (Labeled)	Dimension	Anomaly Ratio (%)
MSL	58,317	73,729	55	10.5
SMAP	135,183	427,617	25	12.8
PSM	132,481	87,841	25	27.8
SMD	708,405	708,420	38	4.2
SWaT	495,000	449,919	51	12.1

MSL (Mars Science Laboratory Dataset): Collected by NASA, it showcases sensor and actuator status data from the Mars rover [26].

SMAP (Soil Moisture Active Passive Dataset): Also collected by NASA, it provides soil sample and telemetry data used by the Mars rover. Compared to MSL, the SMAP dataset includes more point anomalies [26].

PSM (Pooled Server Metrics Dataset): A public dataset from eBay server machines, containing 25 dimensions [27].

SMD (Server Machine Dataset): A five-week dataset sourced from a computing cluster of an internet company, recording resource usage access traces of 28 machines [23].

SWaT (Secure Water Treatment Dataset): A dataset based on 51 sensor dimensions from a continuously operating critical infrastructure system [28].

4.2. Experimental Setup

This study adopts the standardized protocol established by Shen et al. [29], employing a non-overlapping sliding window mechanism with fixed window size (w = 100 across all datasets) for subsequence extraction. Dynamic threshold adaptation is implemented through Extreme Value Theory-based optimization.

This work implements the anomaly post-processing protocol by Xu et al. [30], where the detection of a single anomaly point within a continuous window triggers comprehensive labeling of all data points in that window. The rationale behind this strategy comes from practical observations: the detection of an anomaly time point usually triggers an alarm, which then leads to the focus on the entire anomaly segment.

In terms of model structure, the association discrepancy layer and the stochastic association discrepancy layer each consist of 3 layers. The hidden state dimension d_{model} is set to 512 and the number of attention heads h is fixed at 8. To balance the two components of the loss function, the hyperparameter k is consistently set to 3 across all datasets.

The training process uses the ADAM optimizer [31], with the initial learning rate set to 10^{-4}. Early stopping termination criteria are enforced, constraining maximum training iterations to 10 epochs with the batch size of 32.

4.3. Main Results

The proposed time series anomaly detection framework undergoes rigorous evaluation through empirical validation on five benchmark temporal datasets. The performance of the proposed model was compared with the following advanced baseline models: LSTM-

VAE [22], AnomalyTransformer [2], DCdetector [32], OmniAnomaly [23], InterFusion [33], and DAGMM [24]. The comparative results are systematically summarized in Table 2. The F1 score serves as the primary evaluation metric, with boldface denoting optimal performance and underlined entries indicating secondary superiority.

Table 2. Performance comparison of VAE-Anomaly and other methods. All results are in %, the best ones are in Bold, and the second ones are underlined.

Dataset	SMD			MSL			SMAP			SWaT			PSM		
Metric	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
DAGMM	67.30	49.89	57.30	89.60	63.93	74.62	86.45	56.73	68.51	89.92	57.84	70.40	93.49	70.03	80.08
LSTM-VAE	75.76	90.08	82.30	85.49	79.94	82.62	92.20	67.75	78.10	76.00	89.50	82.20	73.62	89.92	80.96
OmniAnomaly	83.68	86.82	85.22	89.02	86.37	87.67	92.49	81.99	86.92	81.42	84.30	82.83	88.39	74.46	80.83
InterFusion	87.02	85.43	86.22	81.28	92.70	86.62	89.77	88.52	89.14	80.59	85.58	83.01	83.61	83.45	83.52
AnomalyTrans	88.47	92.28	90.33	91.92	96.03	93.93	93.59	**99.41**	96.41	89.10	99.28	94.22	96.94	97.81	97.37
DCdetector	83.59	91.10	87.18	**93.69**	**99.69**	**96.60**	95.63	98.92	97.02	**93.11**	99.77	**96.33**	97.14	**98.74**	97.94
VAE-Anomaly	**93.55**	**93.02**	**93.28**	92.48	97.93	95.13	**95.98**	98.44	**97.19**	92.92	**100**	**96.33**	**98.89**	98.16	**98.52**

As evidenced in the comparative analysis, the proposed architecture demonstrates superior performance over five benchmark models across five standardized datasets (SMD, MSL, SMAP, SWaT, and PSM), achieving F1 scores of 93.28%, 95.13%, 97.19%, 96.33%, and 98.52%, respectively. Specifically, the F1 scores of the proposed model are improved by 10.98%, 12.51%, 19.09%, 14.13%, and 17.56% compared to LSTM-VAE on the SMD, MSL, SMAP, SWaT, and PSM datasets, respectively; improved by 8.06%, 7.46%, 10.27%, 13.50%, and 17.69% compared to OmniAnomaly; and improved by 7.06%, 8.51%, 8.05%, 13.32%, and 15.00% compared to InterFusion. These three baseline models are all sequence generation models based on VAE.

Compared to LSTM-VAE, the proposed model shows a significant improvement in F1 scores across the five datasets. This improvement is primarily due to the modeling of long-term dependencies in time series anomaly detection tasks. LSTM-VAE fails to effectively capture the temporal dependencies of the latent variables in the latent space, whereas the proposed model incorporates a multi-head self-attention mechanism that effectively models long-term dependencies.

OmniAnomaly uses GRU to capture temporal correlations, but it has limitations when handling long-term dependencies. Additionally, it constructs prior distributions based on a linear Gaussian state space model (LGSSM), which limits the model's ability to express complex nonlinear transformations. Moreover, OmniAnomaly does not fully integrate the hidden state information accumulated by GRU during the generation process, which hinders the model's ability to learn deep patterns.

InterFusion refines normal behavior patterns through hierarchical variational autoencoders, but it still relies on reconstruction probabilities as anomaly scores, rather than combining multiple features into a comprehensive score. As a result, its detection performance is not as effective as the proposed model.

The F1 scores of DAGMM are lower than the proposed model by 35.98%, 20.51%, 28.68%, 25.93%, and 18.44% on the SMD, MSL, SMAP, SWaT, and PSM datasets, respectively. DAGMM leverages deep autoencoders for dimensionality reduction and synergistically integrates Gaussian Mixture Models (GMMs) to perform probabilistic anomaly scoring. However, its autoencoders fail to effectively extract temporal features from multidimensional time series data, resulting in poor performance across these five datasets.

Additionally, the proposed model outperforms AnomalyTransformer by 2.95%, 1.2%, 0.78%, 2.11%, and 1.15% in terms of F1 score on the SMD, MSL, SMAP, SWaT, and PSM datasets, respectively. Although AnomalyTransformer also uses association differences to detect anomalies, it relies solely on the self-attention mechanism. Comparative analysis

reveals suboptimal efficacy in temporal pattern abstraction compared to our architecture, particularly in capturing latent structural dependencies within time series data. The proposed model, by learning association differences on reconstructed time series data, enhances the ability to distinguish anomalies based on association differences. Consequently, the proposed architecture demonstrates superior performance over AnomalyTransformer across all five benchmark datasets.

On the SMD, SMAP, and PSM datasets, the proposed model outperforms DCdetector by 6.1%, 0.17%, and 0.58%, respectively, while performing similarly on the SWaT dataset. On the MSL dataset, the proposed model's F1 score is 1.47% lower than DCdetector. Among all five datasets, the proposed model shows the most significant improvement on the SMD dataset, where it outperforms DCdetector by the largest margin. The SMD dataset contains many short-duration samples with subtle anomalous changes. DCdetector, which uses a dual-attention contrastive representation learning structure, struggles to effectively capture these small anomaly features.

Overall, the proposed model achieves an average F1 score that surpasses the baseline method, DCdetector, by 1.076% across the five datasets. These findings substantiate the model's superior generalization capabilities and operational efficacy in complex environments characterized by heterogeneous anomaly patterns.

4.4. Visualization Analysis

To better demonstrate the effectiveness of the association discrepancy and reconstruction error in time series modeling, this study selected a subset of test data from the SMD dataset for visualization analysis. Figure 3 delineates the fluctuation patterns across distinct sequences in the testing set, with tripartite representations comprising the association discrepancy vector, reconstructed sequence, and raw input. Anomalous regions are demarcated via red bounding boxes.

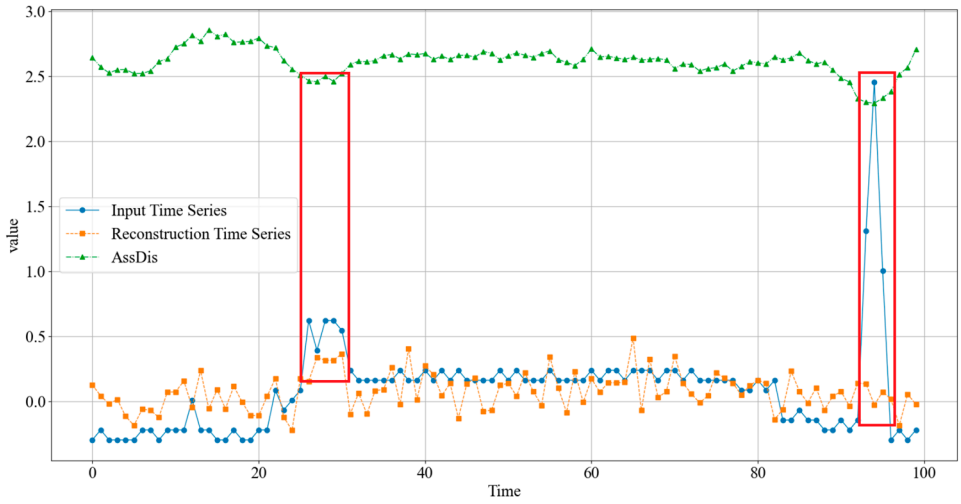


Figure 3. Visualization of test set data.

From the analysis of Figure 3, it is clear that the association discrepancy values corresponding to anomaly points in the original sequence are significantly lower than those of normal points, while the reconstruction error exhibits a notable increase near the anomaly points. This suggests that the association discrepancy mechanism effectively identifies variations in correlations between normal and abnormal states within the time series. The association discrepancy at anomaly points shows distinct characteristics compared to normal points. Additionally, the reconstruction error shows greater fluctuation at the anomaly

points, further proving that the reconstruction module can accurately reflect the anomalous features in the data.

By combining these two metrics, the model can more effectively differentiate between anomaly points and normal points, thereby improving anomaly detection accuracy.

At the bottom of Figure 3, the stability of the reconstructed sequence is shown. Despite small fluctuations in the original data, the reconstructed sequence remains relatively steady. This characteristic suggests that through the learning process of the reconstruction module, the model is able to capture the global trends in the time series and avoid misjudgments or overreactions caused by minor fluctuations, effectively reducing the false positive rate.

In summary, by combining the dual decision-making mechanisms of association discrepancy and reconstruction error, the model demonstrates higher sensitivity and robustness when faced with local fluctuations and sudden anomalies in time series data. Consequently, it demonstrates superior performance in detecting complex anomalies.

4.5. Parameter Experiment

To explore the model's sensitivity to hyperparameters, two of the most important hyperparameters were selected for experimentation: the sliding window length and the joint optimization weight.

To further investigate the influence of sliding window size on anomaly discernment accuracy, experiments were conducted on five different datasets, while keeping other parameters in the model unchanged and setting different sliding window sizes. Different sliding window sizes were tested, and Figure 4 delineates the resultant performance metrics. The figure indicates that the proposed model maintains a certain level of stability across varying sliding window sizes.

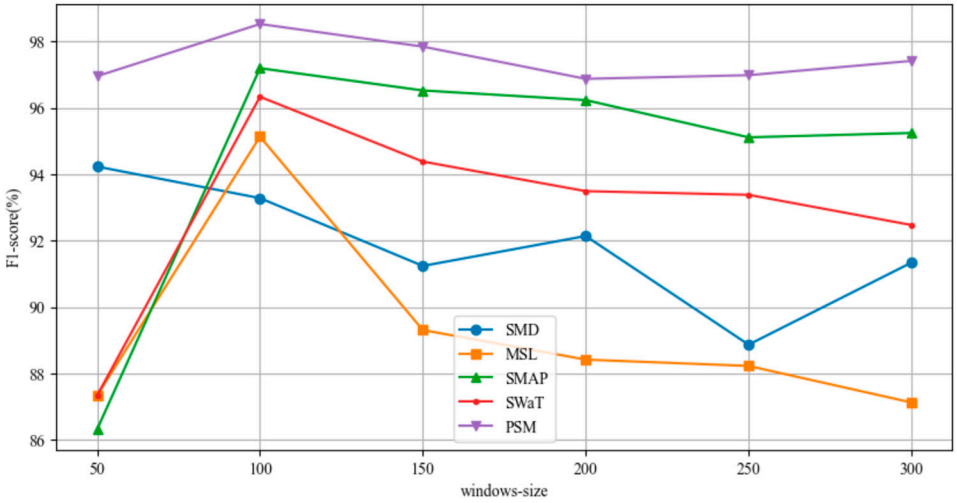


Figure 4. Influence of sliding window size on detection performance.

However, it is important to note that as the sliding window size grows, the model's memory consumption also increases, leading to a significant rise in both training and inference time costs. With smaller sliding windows, although each sliding step is shorter, allowing for more frequent updates to the model state, the limited information within the window often fails to capture sufficient contextual details from the sequence, leading to reduced anomaly detection accuracy.

To further examine the impact of the parameter k, which regulates the optimization weight between reconstruction error and association discrepancy, on model performance, extensive experiments were conducted on five datasets. Specifically, we explored the sensi-

tivity of the model's anomaly detection accuracy to different k values. The experimental results are presented in Figure 5.

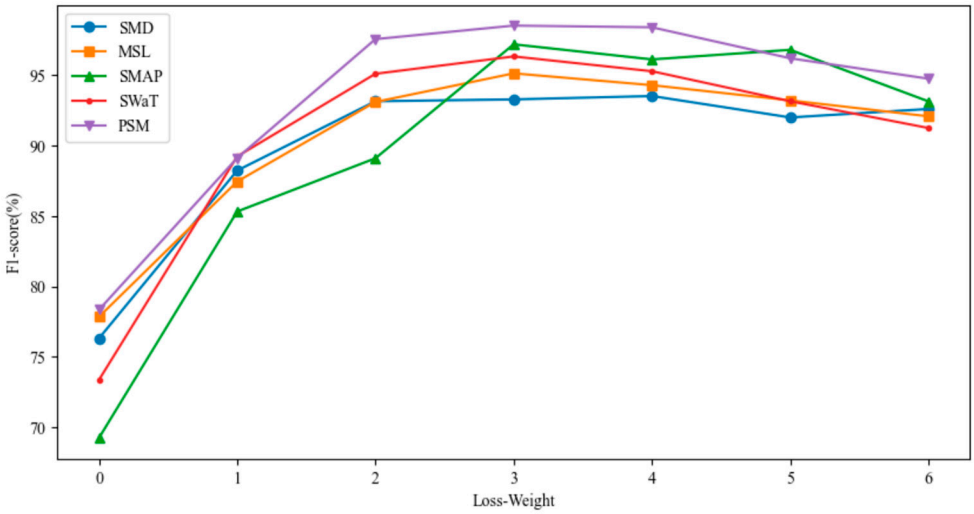


Figure 5. Influence of joint optimization weights on detection performance.

The results indicate that when $k = 3$, the model achieves the best F1 score across most datasets, suggesting that this value effectively balances reconstruction error and association discrepancy. Within the range of $k = 2$ to $k = 6$, the model performance remains stable. This demonstrates the model's broad adaptability in anomaly detection tasks.

From the experimental outcomes, the model's adjustment to the k value exhibits high adaptability, further confirming its ability to handle varying data characteristics and changes in real-world applications. This robustness ensures that the proposed anomaly detection method can effectively address various challenges in different datasets and complex environments.

As shown in Figure 6, the per-timestep processing time of our method under varying d_{model} dimensions on the SMD dataset demonstrates its real-time monitoring capability in industrial environments.

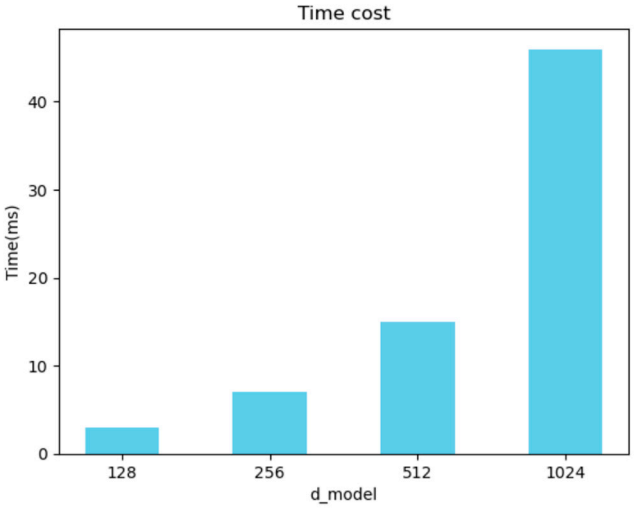


Figure 6. Single-time-step processing time with different d_{model} sizes.

4.6. Ablation Study

To validate the effectiveness of the design of each module in the proposed anomaly detection model, this study compares the model with three of its variants. The experimental results, in terms of F1 scores, are shown in Table 3.

Table 3. F1 scores in the ablation experiments.

	SMD	MSL	SMAP	SWaT	PSM
VAE-Anomaly-1	76.31	77.84	69.24	73.36	78.34
VAE-Anomaly-2	88.35	92.34	92.23	92.53	95.34
VAE-Anomaly-3	89.15	91.91	96.44	95.26	94.76
VAE-Anomaly	93.28	95.13	97.19	96.33	98.52

VAE-Anomaly-1 refers to using only reconstruction error as the criterion for anomaly detection, where reconstruction error includes MSE and KLD.

VAE-Anomaly-2 refers to using only association difference as the criterion for anomaly detection.

VAE-Anomaly-3 refers to using a combination of association difference and reconstruction error, but without using dynamic thresholds.

VAE-Anomaly refers to the anomaly detection model introduced in this paper.

The table shows that the proposed model outperforms the VAE-Anomaly-1 model in terms of F1 scores by 16.97%, 17.29%, 27.95%, 22.97%, and 20.18% for the SMD, MSL, SMAP, SWaT, and PSM datasets, respectively. It also outperforms the VAE-Anomaly-2 model by 4.93%, 2.79%, 4.86%, 3.80%, and 3.18% for the same datasets. This indicates that combining reconstruction error and association difference as the anomaly detection criteria leads to better identification of anomalies in time series data.

Reconstruction error provides an overall understanding of the underlying structural information in the time series data, ensuring that the learned latent features follow certain statistical patterns. The association difference captures the intrinsic relationships between time series and can effectively capture changes in these relationships under time, periodicity, seasonality, and anomaly conditions. The combination of both enables more comprehensive anomaly detection and feature learning, thereby helping to build a robust model.

The proposed model outperforms the VAE-Anomaly-3 model by 4.13%, 3.22%, 0.75%, 1.07%, and 3.76% for the SMD, MSL, SMAP, SWaT, and PSM datasets, respectively. This demonstrates that by dynamically setting the anomaly score threshold in an appropriate way, the model becomes more intelligent and flexible. It can continuously adjust the criteria for determining anomalies according to the actual data conditions, improving detection accuracy, and enhancing the model's robustness in dealing with complex situations such as data dynamics, local anomalies, and concept drift. This provides a more reliable solution for various time series anomaly detection tasks.

5. Conclusions

This paper proposes a novel multivariate time series anomaly detection model aimed at improving the performance of time series anomaly detection. By introducing association discrepancy, the model effectively captures the nature of anomalous changes in time series. The calculation of association discrepancy allows the model to quantify the changes in the degree of correlation between variables, thus enabling a more precise distinction between normal and abnormal states. Additionally, by using VAE to reconstruct the time series, the model identifies the factors that best represent the essential characteristics of the data. By further learning the association discrepancy of the reconstructed time series, the model's anomaly differentiation ability is enhanced and generalized.

Furthermore, the combination of mean squared error, KL divergence, and association discrepancy as anomaly detection criteria significantly enhances the model's capability to detect anomalies. The experimental results across five datasets demonstrate that the proposed model achieves an average F1 score 1.076% higher than the best baseline method, DCdetector, confirming the model's effectiveness and generalization capability.

In future work, we plan to further investigate how to reduce the computational complexity of the model, so that it can be better applied to real-world anomaly detection scenarios, enhancing its performance in real-time detection and large-scale datasets.

Author Contributions: Conceptualization, H.W. and H.Z.; methodology, H.W.; software, H.W.; validation, H.W. and H.Z.; formal analysis, H.W.; investigation, H.W.; resources, H.W.; data curation, H.W.; writing—original draft preparation, H.W.; writing—review and editing, H.W.; visualization, H.W.; supervision, H.W.; project administration, H.W.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province, grant numbers 2022C01220, 2024C01019.

Data Availability Statement: All data are publicly available. SMD: https://github.com/NetManAIOps/OmniAnomaly; MSL: https://github.com/khundman/telemanom; SMAP: https://github.com/khundman/telemanom; SWaT: https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/; PSM: https://github.com/eBay/RANSynCoders/tree/main/data.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, J.; Fang, J.; Lei, F. On-Line Detection Method for Abnormal Data of Power Quality. *Comput. Eng. Appl.* **2020**, *56*, 240–247.
2. Yao, Z.; Wang, R.; Chen, X.; Wang, P.; Guo, Y.; Yu, P.S. Anomaly Transformer: Time series anomaly detection with association discrepancy. *arXiv* **2021**, arXiv:2110.02642.
3. Siffer, A.; Fouque, P.-A.; Termier, A.; Largouet, C. Anomaly Detection in Streams with Extreme Value Theory. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1067–1075.
4. Yu, Q.; Jibin, L.; Jiang, L. An Improved ARIMA-Based Traffic Anomaly Detection Algorithm for Wireless Sensor Networks. *Int. J. Distrib. Sens. Netw.* **2016**, *12*, 9653230.
5. Pincombe, B. Anomaly Detection in Time Series of Graphs Using ARMA Processes. *Bull. Am. Soc. Overseas Res.* **2005**, *24*, 2.
6. Xu, H.; Sun, Z.; Cao, Y.; Bilal, H. A Data-Driven Approach for Intrusion and Anomaly Detection Using Automated Machine Learning for the Internet of Things. *Soft. Comput.* **2023**, *27*, 14469–14481. [CrossRef]
7. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 413–422.
8. Salinas, D.; Flunkert, V.; Gasthaus, J.; Januschowski, T. DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *Int. J. Forecast.* **2020**, *36*, 1181–1191. [CrossRef]
9. Lai, G.; Chang, W.-C.; Yang, Y.; Liu, H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 95–104.
10. Lea, C.; Flynn, M.D.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal Convolutional Networks for Action Segmentation and Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1003–1012.
11. Liu, M.; Zeng, A.; Chen, M.; Xu, Z.; Lai, Q.; Ma, L.; Xu, Q. SCINet: Time Series Modeling and Forecasting with Sample Convolution and Interaction. In Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022), New Orleans, LA, USA, 28 November–9 December 2022.
12. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, Virtual Conference, 2–9 February 2021; Volume 35, pp. 11106–11115.
13. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In Proceedings of the 35th Annual Conference on Neural Information Processing Systems (NeurIPS 2021), Virtual Conference, 6–14 December 2021.
14. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 27268–27286.
15. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A Time Series is Worth 64 Words: Long-Term Forecasting with Transformers. *arXiv* **2023**, arXiv:2211.14730.

16. Oreshkin, B.N.; Carpov, D.; Chapados, N.; Bengio, Y. N-Beats: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting. *arXiv* **2020**, arXiv:1905.1043.
17. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are Transformers Effective for Time Series Forecasting? *arXiv* **2022**, arXiv:2205.13504.
18. Jin, M.; Koh, H.Y.; Wen, Q.; Zambon, D.; Alippi, C.; Webb, G.I.; King, I.; Pan, S. A Survey on Graph Neural Networks for Time Series: Forecasting, Classification, Imputation, and Anomaly Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 10466–10485. [PubMed]
19. Iqbal, A.; Amin, R. Time Series Forecasting and Anomaly Detection Using Deep Learning. *Comput. Chem. Eng.* **2024**, *182*, 108560. [CrossRef]
20. Cui, Q.D.; Xu, C.; Xu, Y.; Ou, W.; Pang, Y.; Liu, Z.; Shen, J.; Baber, M.Z.; Maharajan, C.; Ghosh, U. Bifurcation and Controller Design of 5D BAM Neural Networks with Time Delay. *Int. J. Numer. Model.* **2024**, *37*, e3316. [CrossRef]
21. Maharajan, C.; Sowmiya, C.; Xu, C. Delay Dependent Complex-Valued Bidirectional Associative Memory Neural Networks with Stochastic and Impulsive Effects: An Exponential Stability Approach. *Kybernetika* **2024**, *60*, 317–356.
22. Park, D.; Hoshi, Y.; Kemp, C.C. A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder. *IEEE Robot. Autom. Lett.* **2018**, *3*, 1544–1551. [CrossRef]
23. Su, Y.; Liu, R.; Zhao, Y.; Sun, W.; Niu, C.; 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, Anchorage, AK, USA, 4–8 August 2019; pp. 2828–2837.
24. 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 Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018; IEEE Press: Piscataway, NJ, USA, 2018; pp. 1–19.
25. Chung, J.; Kastner, K.; Dinh, L.; Goel, K.; Courville, A.; Bengio, Y. A recurrent latent variable model for sequential data. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 2962–2970.
26. 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, London, UK, 19–23 August 2018; pp. 387–395.
27. Abdulaal, A.; Liu, Z.; Lancewicki, T. Practical Approach to Asynchronous Multivariate Time Series Anomaly Detection and Localization. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021; pp. 2485–2494.
28. Mathur, A.P.; Tippenhauer, N.O. SWaT: A water treatment testbed for research and training on ICS security. In Proceedings of the 2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater), Vienna, Austria, 11 April 2016; pp. 31–36.
29. Shen, L.; Li, Z.; Kwok, J.T. Timeseries anomaly detection using temporal hierarchical one-class network. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 13016–13026.
30. Xu, H.; Chen, W.; Zhao, N.; Li, Z.; Bu, J.; Li, Z.; Liu, Y.; Zhao, Y.; Pei, D.; Feng, Y.; et al. Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 187–196.
31. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
32. Yang, Y.; Zhang, C.; Zhou, T.; Wen, Q.; Sun, L. DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Long Beach, CA, USA, 6–10 August 2023; pp. 3033–3045.
33. Li, Z.; Zhao, Y.; Han, J.; Su, Y.; Jiao, R.; Wen, X.; Pei, D. Multivariate Time Series Anomaly Detection and Interpretation using Hierarchical Inter-Metric and Temporal Embedding. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021; pp. 3220–3230.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.