

RESEARCH ARTICLE

CNTS: Cooperative Network for Time Series

JINSHENG YANG¹, YUANHAI SHAO¹, AND CHUN-NA LI¹

Management School, Hainan University, Haikou 570228, China

Corresponding author: Yuanhai Shao (shaoyuanhai21@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 12271131, Grant 61966024, Grant 62106112, Grant 61866010, and Grant 11871183; and in part by the Natural Science Foundation of Hainan Province under Grant 120RC449 and Grant 620QN234.

ABSTRACT The use of deep learning techniques in detecting anomalies in time series data has been an active area of research with a long history of development and a variety of approaches. In particular, reconstruction-based unsupervised anomaly detection methods have gained popularity due to their intuitive assumptions and low computational requirements. However, these methods are often susceptible to outliers and do not effectively model anomalies, leading to suboptimal results. This paper presents a novel approach for unsupervised anomaly detection, called the Cooperative Network Time Series (CNTS) approach. The CNTS system consists of two components: a detector and a reconstructor. The detector is responsible for directly detecting anomalies, while the reconstructor provides reconstruction information to the detector and updates its learning based on anomalous information received from the detector. The central aspect of CNTS is a multi-objective optimization problem, which is solved through a cooperative solution strategy. Experiments on three real-world datasets demonstrate the state-of-the-art performance of CNTS and confirm the effectiveness of the detector and reconstructor. The source code for this study is publicly available on GitHub (<https://github.com/BomBooooo/CNTS/tree/main>).

INDEX TERMS Deep learning, time series, anomaly detection, reconstruction, cooperative network.

I. INTRODUCTION

The field of anomaly detection has a rich history dating back to the 1960s [1], with an increasing number of applications in areas such as fraud detection [2] and intrusion detection [3]. The proliferation of data has further expanded the scope of anomaly detection. The presence of anomalies in Internet of Things (IoT) data poses significant security risks, hindering the development of the IoT [4]. Deep Learning, a potent machine learning technique, has recently garnered significant attention from researchers and demonstrated its superiority in anomaly detection [5].

The problem of anomalies is very common in real life, but it is difficult to obtain high-quality labeled data [6]. Therefore, there is an urgent need for unsupervised anomaly detection methods. Deep learning approaches for anomaly detection can broadly be classified into three categories [5]: (i) Feature extraction: the objective is to use deep learning to

reduce high-dimensional and/or non-linearly separable data into low-dimensional feature representations for downstream anomaly detection [7]. (ii) Learning feature representations of normality: these methods integrate feature learning with anomaly assessment, unlike the previous category, which separates the two processes. Examples include autoencoders [8], GANs [9], predictability modeling [10], and self-supervised classification [11]. (iii) End-to-end anomaly score learning: the goal is to learn scalar anomaly scores in a direct manner without relying on existing anomaly metrics. The neural network in this approach learns to identify outliers directly [12]. The first two methods of anomaly detection are predominantly utilized in unsupervised tasks, whereas the third method is more commonly employed in supervised tasks, as it necessitates the availability of anomalous labels.

The first category of methods involves a separation of tasks and models, while the third category requires labeled abnormal data. As a result, the most commonly used method for unsupervised anomaly detection currently belongs to the second category, i.e., reconstruction-based methods [13]. These

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy¹.

methods operate by inputting time series data and calculating its reconstruction error, which is then used as an anomaly score to determine the abnormality of the data. Once the anomaly score has been obtained, a suitable threshold is selected to identify anomalies [14]. The underlying assumption of this method is that deep networks can effectively learn low-dimensional features of the data and normalize outliers by reconstruction.

Although these methods have demonstrated good performance, they suffer from two key limitations. Firstly, they do not provide any special treatment for outliers, which can negatively impact the results of the model since they are present throughout the training process. Secondly, the goal of the network is reconstructing the data, rather than detecting anomalies, resulting in suboptimal performance [5].

In this paper, we present a cooperative network for time series (CNTS) to address the aforementioned limitations of existing reconstruction-based methods. The CNTS consists of two deep learning networks, a reconstructor (**R**) and a detector (**D**). **R** is responsible for reconstructing the original data and providing abnormal labels by calculating reconstruction error, while **D** trains based on these labels and returns the results to **R** for weight adjustment. Through this cooperative training process, **D** and **R** work together to improve learning performance. The CNTS reduces the impact of outliers by screening and selecting appropriate data during training, and enhances anomaly detection performance by separating the tasks of data reconstruction and anomaly detection into separate networks. The CNTS solves a multiobjective programming problem and is based on the FEDFormer network [15]. The main contributions of this work can be summarized as follows:

- **Reconstruction:** a novel reconstruction network, based on anomaly weights, is proposed in this work. This reconstruction network has the ability to adapt its own training loss to the abnormal conditions, thus enhancing its ability to distinguish between abnormal and normal samples after reconstruction.
- **Detection:** we present a direct training method for unsupervised anomaly detection, which involves separating the tasks of reconstruction and anomaly detection. This approach prevents suboptimal results that may arise from poor feature extraction.
- **Cooperation Network:** A new cooperative training mode for the network is proposed in this work. The network consists of two parts: the reconstructor (**R**) that learns the characteristics of normal samples and the detector (**D**) that detects anomalies. **R** and **D** work together and enhance each other's performance. Both are part of each other's loss function in this cooperative training mode.
- **Experiment:** Experiments were conducted on three databases, totaling 128 real-world data, and the results showed that the detection performance of **D** surpasses that of nine baseline methods in terms of accuracy, F1 value, and AUC. Additionally, the ability to distinguish

between normal and abnormal samples was further improved by the reconstructed data of **R**.

II. PROBLEM FORMULATION

Given a period of univariate time series data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, with $x_t \in \mathbb{R}^1$, where N is the length of data. The abnormal labels are denoted as $\mathbf{Y} = \{y_1, y_2, \dots, y_N\}$, where

$$y_t = \begin{cases} 1 & \text{if } \mathbf{x}_t \text{ is abnormal,} \\ 0 & \text{else.} \end{cases} \quad (1)$$

Therefore, the problem of anomaly detection can be described as follows: given a time series data $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, the aim is to train the model to minimize the following objective function:

$$\min_{\theta_{\mathcal{M}}} \mathcal{L}(\mathcal{M}(\mathbf{X}, \theta_{\mathcal{M}}), \mathbf{Y}). \quad (2)$$

where \mathcal{L} represents the loss function and $\mathcal{M}(\cdot, \theta_{\mathcal{M}})$ represents the model \mathcal{M} with parameter $\theta_{\mathcal{M}}$. Since this paper mainly discusses the problem of unsupervised anomaly detection, \mathbf{Y} does not exist in the training set and only exists in the testing set, which is the same as the usual unsupervised anomaly detection settings.

III. RELATED WORKS

The objective of reconstruction-based anomaly detection models is to learn an optimized model that accurately reconstructs normal data instances while failing to reconstruct anomalies effectively [13]. The use of reconstruction models for anomaly detection is based on the assumption that the reconstructed data can differentiate normal data from abnormal data. The objective function of these models can be formulated as follows:

$$\min_{\theta_{\mathcal{M}}} \sum_{i=1}^n [(1 - y_i) \mathcal{L}(\mathbf{x}_i, \mathcal{M}(\mathbf{x}_i, \theta_{\mathcal{M}})) - y_i \mathcal{L}(\mathbf{x}_i, \mathcal{M}(\mathbf{x}_i, \theta_{\mathcal{M}}))] \quad (3)$$

The effectiveness of the reconstruction model in detecting anomalies depends on its ability to produce a small reconstruction error for normal values and a large reconstruction error for abnormal values. This will result in a relatively low value for formula (3). On the other hand, formula (2) has an advantage over formula (3) as it models anomalies directly, whereas formula (3) relies on the hope that the reconstruction error can differentiate between normal and abnormal values.

In practical modeling, the labels for anomalies are always unknown, and the objective function is formulated as Equation (4). Equation (4) simplifies the second term in Equation (3). Despite the fact that outliers represent only a small fraction of the original data, Equation (4) remains effective.

$$\min_{\theta_{\mathcal{M}}} \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i - \mathcal{M}(\mathbf{x}_i, \theta_{\mathcal{M}})\|^2 \quad (4)$$

There are already several models that are based on these concepts. For instance, the Deep Autoencoding Gaussian

Mixture Model (DAGMM) [16] is an unsupervised anomaly detection method that combines an autoencoder with a Gaussian mixture model. The autoencoder is used to compress and reconstruct the data, and the reconstruction error is fed into the Gaussian mixture model. The entire model is trained end-to-end. MAD-GAN [17] is a model that employs a Generative Adversarial Network (GAN) architecture for training, and uses both the reconstruction loss and the discriminative loss as anomaly scores in the testing phase. The Unsupervised Anomaly Detection (USAD) [18] method introduces another decoder based on the traditional autoencoder, and trains the model unsupervised through adversarial learning. The reconstruction errors from the two decoders play a crucial role in the training process. MTAD-GAT [19] optimizes both the forecast-based model and the reconstruction-based model and achieves a better representation of the time series by combining single-timestamp forecasting and the reconstruction of the entire time series. CAE-M [20] simultaneously optimizes two networks, the characterization network, and the memory network, where the characterization network is based on the reconstruction loss.

Methods such as these can partially alleviate the impact of outliers because normal values typically constitute the majority of the data. However, it is not feasible to use the same weights for all data, as outliers can affect the training process. Additionally, the lack of direct training for anomalies in the objective function can result in suboptimal outcomes.

The inspiration for the cooperative approach in this paper mainly comes from the adversarial process in GANs [21]. As is well-known, the original GAN was primarily used for training generative models, and various modifications have since been developed. In GANs, the generator and discriminator engage in a game and eventually reach a Nash equilibrium. The patterns generated by the generator in the equilibrium state are sufficiently similar to real patterns to confuse the discriminator. This paper introduces a Cooperative Network for Time Series (CNTS), which also contains two networks: a reconstructor (**R**) and a detector (**D**). However, the relationship between the two networks is not adversarial, but cooperative. **R** and **D** work together towards similar objectives, ultimately achieving their individual goals. Specifically, **R**'s goal is to reconstruct normal samples and distinguish them from abnormal samples, while **D**'s goal is to detect abnormalities. There is a degree of overlap between the two goals, allowing for mutual reinforcement.

IV. COOPERATIVE NETWORK FOR TIME SERIES

An anomaly is a singular or a series of observations that significantly deviate from the general distribution of the data. The set of anomalies constitutes a small portion of the data set [22]. Therefore, anomaly detection can be viewed as the identification of values that do not conform to the known data distribution. To address this challenge, we propose a deep learning network that integrates reconstruction and detection. Reconstruction aims to reconstruct values that closely match the known data distribution, while detection focuses on

identifying values that do not align with the fitted data distribution. These two tasks are both distinct and interrelated.

A. DIFFERENCE

There are two main differences between the objectives of the reconstructor (**R**) and the detector (**D**). Firstly, **R** aims to fit the known data distribution, whereas **D** aims to identify values that deviate from the normal data distribution. Secondly, the output information differs: **R** provides reconstructions, while **D** provides anomaly scores.

B. SIMILARITY

Despite the differences in their objectives, **R** and **D** share similarities in their input forms, and the output information can be used to some extent to judge the degree of abnormality.

The mutual utilization of information between two entities, **D** and **R**, has several advantages. Firstly, the reconstruction error of **R** can be used as a labeling mechanism for **D** during its training process. Secondly, **D** can provide information to **R** regarding anomalies, thereby reducing the influence of potential outliers on **R**'s training. This separation of the task of anomaly detection from data reconstruction allows CNTS to be modeled as a multi-objective programming problem. The objective function can be formulated as follows:

$$\begin{aligned} & \left[\begin{array}{l} \min_{\theta_D, \theta_R} \sum_{i=1}^n \mathcal{L}_D(\mathbf{D}(\mathbf{x}_i, \theta_D), \mathcal{L}_R(\mathbf{x}_i, \mathbf{R}(\mathbf{x}_i, \theta_R))) \\ \min_{\theta_D, \theta_R} \sum_{i=1}^n (1 - \hat{y}_i(\mathbf{x}_i, \theta_D)) \mathcal{L}_R(\mathbf{x}_i, \mathbf{R}(\mathbf{x}_i, \theta_R)) \end{array} \right] \\ s.t. \hat{y}_i(\mathbf{x}_i, \theta_D) = & \begin{cases} 1 & \text{if } \mathbf{D}(\mathbf{x}_i, \theta_D) > Top_{k\%}(\mathbf{D}(\mathbf{X}, \theta_D)), \\ 0 & \text{else.} \end{cases} \end{aligned} \quad (5)$$

where $\mathbf{D}(\cdot, \theta_D)$ is the **D** with parameters θ_D , $\mathbf{R}(\cdot, \theta_R)$ is the **R** with parameters θ_R , \mathcal{L}_D and \mathcal{L}_R denote the loss of **D** and **R** respectively, and $\hat{y}_i(\cdot, \theta_D)$ is a category label, if the value of $\mathbf{D}(\mathbf{x}_i, \theta_D)$ accounts for the top $k\%$ of all $\mathbf{D}(\mathbf{X}, \theta_D)$, then $\hat{y}_i(\mathbf{x}_i, \theta_D)$ equals to 1, otherwise it equals to 0. Through this label, the data with a large abnormal score is screened out during the training process, thereby reducing the impact of outliers on the reconstruction model.

The overall structure of CNTS is shown in Figure 1. It consists of two parts, the reconstructor and the detector. After the data **W** is input, two outputs can be obtained, namely the reconstructed data \mathbf{W}_r and the abnormal score *Score*. In order to improve the ability of CNTS on reconstruction and detection tasks, it is necessary to make the **R** and **D** cooperate. During the training process, while the numerical value of the abnormal reconstruction error may not necessarily increase, it should increase relative to the reconstruction error of the normal points. At the same time, **D** is continually improving. The improvement of **D** is achieved by using the reconstruction error provided by **R** as a criterion for detecting anomalies. As **D** becomes more robust, the anomaly detection is passed

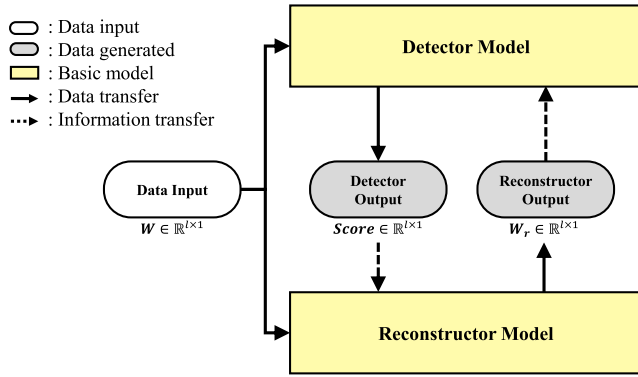


FIGURE 1. Structure of CNTS. The black dotted line indicates the cooperative information transfer during training.

back to \mathbf{R} , which results in a clearer distinction between normal and anomalous reconstructions. The following sections describe a highly effective cooperative learning strategy for optimizing both objectives simultaneously.

C. BASIC MODEL

In this study, both the anomaly detector and the reconstructor are based on the FEDFormer [15] network structure. The FEDFormer architecture integrates the Transformer model with the seasonal-trend decomposition method, leading to improved performance for long-term time series forecasting. This approach leverages the fact that many time series data can be represented sparsely in well-known bases and augments the Transformer with a frequency-based approach. The FEDFormer method has been shown to produce a reduction in prediction error of 14.8% for multivariate and 22.6% for univariate time series, compared to state-of-the-art methods, while being more computationally efficient than standard Transformers. As such, the FEDFormer structure serves as the foundation of the CNTS model in this paper.

D. RECONSTRUCTOR

Figure 2 depicts the training process of \mathbf{R} . It is assumed that the detector, trained by the method outlined in the previous section, has the capability to detect outliers. The input data pass through \mathbf{D} to obtain the abnormal scores, and then are processed by \mathbf{R} to generate the reconstructed result. Prior to computing the loss between the input data and the reconstructed data, it is necessary to exclude from the data the values with high abnormal scores, as identified by the detector. The loss is then calculated by combining the input data and reconstructed data, and this guides the training of \mathbf{R} . During the training of \mathbf{R} , the parameters of \mathbf{D} are held constant.

To mitigate the impact of outliers on the reconstruction process, \mathbf{R} employs a strategy of excluding from consideration the top 10% of values with the highest abnormal scores, as determined by \mathbf{D} . This reduces the effect of outliers on the calculation of the loss, which is used to guide the training of \mathbf{R} . The objective function of \mathbf{R} is to minimize this calculated

loss between the input and reconstructed data.

$$\mathbf{W}_r(\mathbf{W}, \theta_R) = \mathbf{R}(\mathbf{W}, \theta_R) \tag{6}$$

$$\theta_R^* = \arg \min_{\theta_R} \mathcal{L}_M(\mathbf{W}_s, \mathbf{W}_{r_s}(\mathbf{W}, \theta_R)) \tag{7}$$

where $\mathbf{R}(\cdot, \theta_R)$ is the \mathbf{R} with parameters θ_R , \mathbf{W}_r represents reconstructed data, $\mathcal{L}_M(\cdot, \cdot)$ is the mean squared error (MSE) loss function, θ_R is the parameter of \mathbf{R} , \mathbf{W}_s and \mathbf{W}_{r_s} represent the selected \mathbf{W} and \mathbf{W}_r .

E. DETECTOR

The specific training process of \mathbf{D} is shown in Figure 3. In the training of deep learning, before the data is input into the model, it needs to be split into subsequences using a sliding window of length l , denoted as \mathbf{W} , where $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$, and $\mathbf{w}_i = \{x_i, x_{i+1}, \dots, x_{i+l-1}\}$.

In the training of \mathbf{D} , it is assumed that data with larger reconstruction errors are more likely to be anomalies. The data entered into the model follows two paths. The first path involves inputting the data into \mathbf{R} to obtain the reconstructed window \mathbf{W}_r and calculating the reconstruction error \mathbf{e}_r . The second path involves inputting the data into \mathbf{D} to obtain the abnormal score $Score$. Before the two paths converge to calculate the loss, the outputs of both parts are compared, and the top k windows with the highest reconstruction errors are selected as anomalies for training the model. During the training of \mathbf{D} , the parameters of \mathbf{R} remain fixed.

The \mathbf{e}_r can be calculated as

$$\mathbf{e}_r(\mathbf{W}, \theta_R) = \mathcal{L}_M(\mathbf{W}_r(\mathbf{W}, \theta_R), \mathbf{W}) \tag{8}$$

The abnormal score $Score$ is calculated by

$$Score(\mathbf{W}, \theta_D) = \mathbf{D}(\mathbf{W}, \theta_D). \tag{9}$$

To avoid the training collapse that can result from class imbalance, only the $Score$ values corresponding to the largest 20% of points are selected as training points. The objective function can be expressed as follows:

$$\theta_D^* = \arg \min_{\theta_D} \mathcal{L}_C(\text{Softmax}(\mathbf{e}_{r_s}(\mathbf{W}, \theta_R)), \text{Softmax}(Score_s(\mathbf{W}, \theta_D))) \tag{10}$$

where \mathcal{L}_C is CrossEntropy loss, and \mathbf{e}_{r_s} and $Score_s$ represent selected \mathbf{e}_r and $Score$, respectively.

F. ALGORITHM

This subsection provides the pseudocode for the complete CNTS algorithm, referred to as Algorithm 1. In the algorithm, where $MSE(\mathbf{X}_1, \mathbf{X}_2) = (\mathbf{X}_1 - \mathbf{X}_2)^2$, the term ‘‘epoch’’ represents the number of iterations of the entire algorithm. ‘‘R_epochs’’ and ‘‘D_epochs’’ denote the number of iterations of \mathbf{R} and \mathbf{D} , respectively, in each loop of the algorithm. During the training of one model, the parameters of the other model remain fixed.

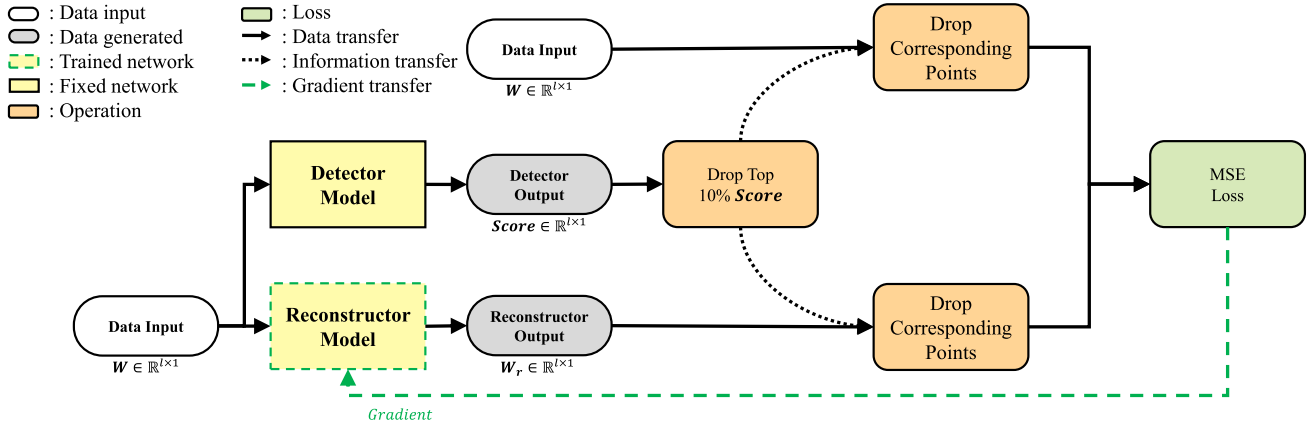


FIGURE 2. Train phase of R. During the training of R, the parameters of D are fixed.

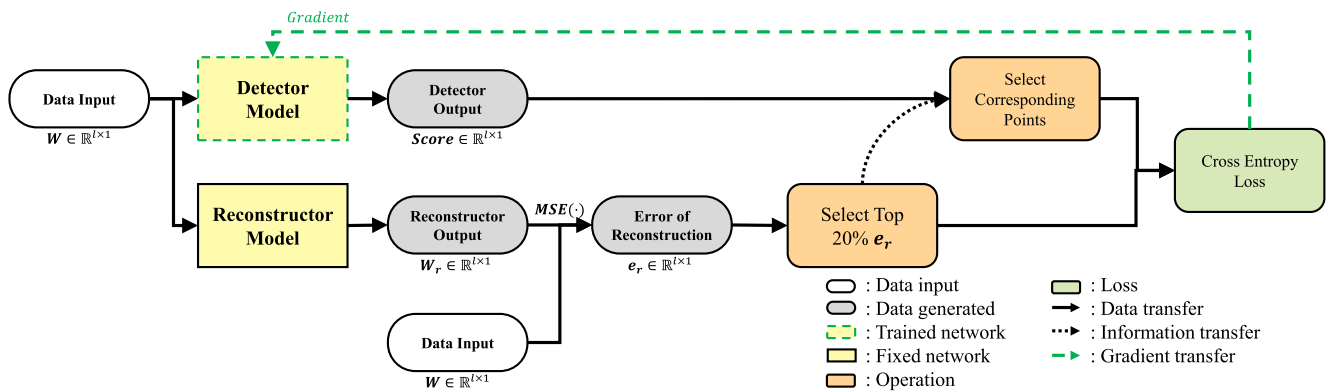


FIGURE 3. Train phase of D. During the training of D, the parameters of R are fixed.

V. EXPERIMENTS

A. MODELS

As the outstanding performance of FEDformer [15] in time series forecasting tasks demonstrates its suitability for time series data, this paper selects it as the network architecture for D and R. This study compares nine deep learning methods and presents them as follows:

- MERLIN [23]: an algorithm that can efficiently and exactly find discords of all lengths in massive time series archives.
- LSTM-NDT [24]: a long short-term memory network that effectively solves the problem of unsupervised anomaly detection.
- OmniAnomaly [25]: a stochastic recurrent neural network for multivariate time series anomaly detection that works well robustly for various devices.
- MSCRED [26]: a model that formulates the anomaly detection and diagnosis problem as three underlying tasks, i.e., anomaly detection, root cause identification, and anomaly severity (duration) interpretation and addresses these issues jointly

- MAD-GAN [17]: a framework considers the entire variable set concurrently to capture the latent interactions amongst the variables.
- USAD [18]: a multivariate time series anomaly detection model based on an autoencoder architecture whose learning is inspired by GANs.
- MTAD-GAT [19]: a self-supervised framework for multivariate time-series anomaly detection
- CAE-M [20]: a unsupervised deep learning based anomaly detection approach for multi-sensor time series data with two main subnetworks: characterization network and memory network.
- GDN [27]: a structure learning approach with graph neural networks, additionally using attention weights to provide explainability for the detected anomalies.

B. DATASETS

The final model comparison in this study utilizes the following two datasets:

- NAB [28]: This dataset consists of labeled real-world and artificial time series, including metrics from AWS servers metrics, online ad click-through rates, real-time

Algorithm 1 Algorithm of CNTS

Require: The input data \mathbf{W} ; Initialize the model \mathbf{R} with parameters θ_R ; Initialize the model \mathbf{D} with parameters θ_D ; Train epoch of CNTS epochs; Train epoch of \mathbf{R} R_epochs ; Train epoch of \mathbf{D} D_epochs ;

Ensure: The trained models \mathbf{R} and \mathbf{D} with parameters θ_R^* and θ_D^* , respectively.;

```

1: for e in epochs:
2:   # Train phase of  $\mathbf{R}$ .
3:   for r_e in R_epochs:
4:      $\mathbf{W}_r = R(\mathbf{W}, \theta_R)$ ;
5:      $\mathbf{W}_{r_s}$  and  $\mathbf{W}_s$  are selected by  $\mathbf{W}_r$  and  $\mathbf{W}$ ,
       respectively;
6:      $\mathcal{L}_R = MSE(\mathbf{W}_{r_s}, \mathbf{W}_s)$ ;
7:     Update parameters of  $\mathbf{R}$  to minimize the loss  $\mathcal{L}_R$ ;
8:   end for
9:   # Train phase of  $\mathbf{D}$ .
10:  for d_e in D_epochs:
11:     $\mathbf{e}_r = MSE(\mathbf{W}, R(\mathbf{W}, \theta_R))$ ;
12:     $Score = \mathbf{D}(\mathbf{W}, \theta_D)$ ;
13:     $Score_s$  and  $\mathbf{e}_{r_s}$  are selected by  $Score$  and  $\mathbf{e}_r$ ,
       respectively;
14:     $\mathcal{L}_D = CrossEntropy(Softmax(Score_s),$ 
                           $Softmax(\mathbf{e}_{r_s}))$ ;
15:    Update parameters of  $\mathbf{D}$  to minimize the loss  $\mathcal{L}_D$ ;
16:  end for
17: end for

```

traffic data, and a collection of Twitter mentions from large public companies.

- NASA-SMAP and NASA-MSL [29]: These datasets consist of real spacecraft telemetry data from the Soil Moisture Active Passive (SMAP) satellite and the Curiosity rover (MSL), respectively.

For the purposes of this study, only the first dimension representing continuous data is retained, while the remaining dimensions for binary data are omitted. In the experiments, the data is divided into a training set and a test set. The training set is used for model training and does not include abnormal labels, while the test set contains abnormal labels for comparison of results. Data without anomalies in the test set is removed. Finally, the effective data information is obtained, as shown in Table 1. The header of Table 1 indicates the name of the datasets, the number of points in the training set, the number of points in the test set, the number of abnormal points in the test set, and the abnormal proportion.

TABLE 1. Information of datasets.

Name	Nums	Train Nums	Test Nums	Abnormal Nums	Abnormal Rate (%)
NAB	47	164096	164130	18456	11.24
NASA-MSL	27	58317	73729	7766	10.53
NASA-SMAP	54	138004	435826	55817	12.81
Total	128	360417	673685	82039	11.53

C. DETECTION

In the experiments, anomaly detection is performed on all data separately, and the average value of each dataset is finally calculated. During the test phase, the data is fed into \mathbf{D} to obtain the anomaly scores and the anomaly labels are used to calculate performance metrics. However, due to the over-estimation of the anomaly detection model by the anomaly adjustment strategy [30], the detection results reported in many previous studies are no longer credible [31]. Hence, this paper eliminates the anomaly adjustment strategy and utilizes a unified detection metric to compare different methods. This metric calculates the result by combining all the test data, instead of adjusting for anomalies. A point with an anomaly score greater than a selected threshold δ is considered an anomaly. The output anomaly result $\mathbf{S}^\delta = \{s_1^\delta, s_2^\delta, \dots, s_N^\delta\}$ is the following formula

$$s_i^\delta = \begin{cases} 1 & \text{if } Score(\mathbf{x}_i) > \delta, \\ 0 & \text{else.} \end{cases} \quad (11)$$

where $Score(\mathbf{x}_i)$ is the anomaly score of \mathbf{x}_i . After labeling, the accuracy, precision, recall, and F1 score for the evaluation are computed as follows:

$$acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

$$precision = \frac{TP}{TP + FP} \quad (13)$$

$$recall = \frac{TP}{TP + FN} \quad (14)$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (15)$$

where TN, TP, FP, and FN denote the number of true negatives, true positives, false positives and false negatives, respectively.

There are numerous ways to choose the threshold [32], [33]. In this paper, the focus is on evaluating the model's ability to distinguish between abnormal and normal data, therefore, the threshold that maximizes the F1 score of the model results is selected.

$$\delta^* = \underset{\delta}{argmax} F1(\mathbf{Y}, \mathbf{S}^\delta) \quad (16)$$

where $F1(\mathbf{Y}, \mathbf{S}^\delta)$ is F1 score between \mathbf{Y} and \mathbf{S}^δ . The accuracy (ACC), F1 value and area under the curve (AUC) of receiver operating characteristic (ROC) of the detection results are shown in the table 2.

The results of the comparison between different detection methods are depicted in Table 2, with the optimal result highlighted in bold. It is evident that the CNTS outperforms the other methods on all three datasets after removing the influence of the abnormal adjustment. This confirms the previous finding that the abnormal adjustment tends to overestimate the performance of the model [31], [34]. The table shows that the CNTS has greatly improved, resulting in the best results in terms of ACC, F1, and AUC. Compared to the baseline method, there has been an improvement of 15%, 17%, and

TABLE 2. Comparison of detection results.

	NAB			NASA-MSL			NASA-SMAP		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
TranAD	0.5180	0.2989	0.5818	0.7820	0.4587	0.6765	0.7775	0.4526	0.7133
GDN	0.5665	0.2909	0.5785	0.8111	0.4578	0.6871	0.8050	0.4207	0.7022
MAD_GAN	0.5161	0.2944	0.5781	0.8099	0.4503	0.6816	0.8043	0.4251	0.7019
MTAD_GAT	0.5611	0.2978	0.5873	0.7730	0.4411	0.6780	0.8146	0.4030	0.6951
MSCRED	0.4940	0.2845	0.5727	0.7709	0.4407	0.6791	0.7613	0.3982	0.6788
USAD	0.5383	0.2952	0.5823	0.7954	0.4489	0.6825	0.7954	0.4161	0.6989
OmniAnomaly	0.5348	0.2926	0.5768	0.8126	0.4305	0.6704	0.7479	0.4048	0.6847
LSTM_AD	0.4887	0.2873	0.5746	0.6916	0.4065	0.6603	0.6297	0.3751	0.6460
CAE_M	0.5504	0.2941	0.5828	0.7605	0.4400	0.6720	0.8033	0.4139	0.7003
Detection*	0.5937	0.3177	0.6071	0.7555	0.4715	0.7109	0.7499	0.4852	0.6682
CNTS	0.7010	0.3717	0.6350	0.8277	0.5459	0.7320	0.8733	0.5685	0.7396

* This method is a network structure that takes out the detector separately in CNTS and uses the reconstruction loss as the training target.

6% respectively in these indicators. Additionally, it has been demonstrated that the use of FEDformer alone, as an unsupervised anomaly detection model based on reconstruction, can still produce good results for some indicators. This indicates a degree of correlation across different time series data tasks.

D. RECONSTRUCTION

The obtained reconstructed results form **R** and the original data are used to calculate the metrics. The reconstruction error $e_r = \{e_1, e_2, \dots, e_N\}$. The reconstruction error of normal and abnormal values will be used to measure the quality of the reconstructed model. It is calculated as follows:

$$\begin{aligned}
 MSE_n &= \frac{\sum_{i=0}^N e_i * (1 - y_i)}{\sum_{i=0}^N (1 - y_i)} \\
 MSE_a &= \frac{\sum_{i=0}^N e_i * y_i}{\sum_{i=0}^N y_i}
 \end{aligned}
 \tag{17}$$

where MSE_n and MSE_a represent the MSE errors of normal and abnormal, respectively.

Figure 4 presents a comparison of the reconstruction effectiveness between a single reconstruction network (**R**) and CNTS. The length of each bar in the histogram represents the proportionality of the models, not the actual values, with **R** being equal to 1. The numerical value for each model's index is indicated on the top of each bar. The histogram further quantifies the capability of the models in distinguishing between normal and anomalous data points. This value can be calculated using the formula provided.

$$Dis = \frac{MSE_a - MSE_n}{MSE_n}
 \tag{18}$$

The histogram in Figure 4 clearly illustrates the superiority of CNTS over the baseline model in terms of discriminative power. As can be observed from the figure, the *Dis* value of CNTS is higher than that of the baseline model across all datasets, with a margin of approximately three times higher on the NASA-MSL data. This result further emphasizes the detrimental impact of outliers on the training process of reconstruction models, reducing their sensitivity to anomalies.

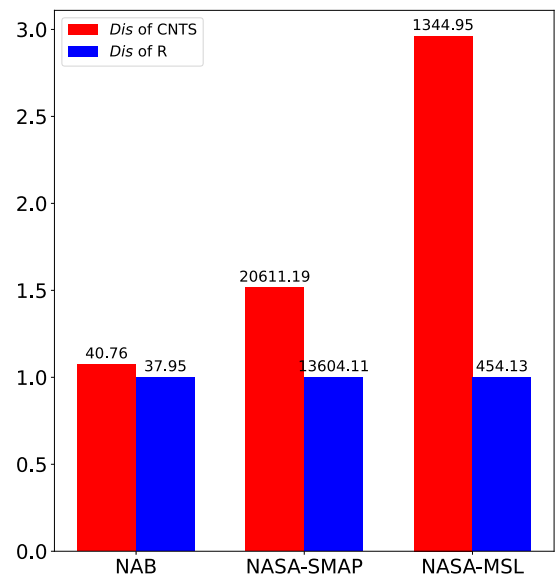


FIGURE 4. Comparisons of reconstruction effects.

E. COOPERATION

Figure 5 displays the training process for six data points from three different databases. The x-axis in Figure 5 represents the number of alternating training rounds, while the left and right y-axes represent the reconstruction loss for normal data points and the F1 value of the detector, respectively.

As the training progresses, the reconstruction loss of CNTS for normal values continually decreases, thereby exposing outliers to an increasingly higher reconstruction error. As a result, the detection component of CNTS continuously learns from this reconstruction error to enhance its ability to detect anomalies.

By combining these results, it can be observed that CNTS outperforms the baseline method after one stage of training in some data sets, while in others it may require several stages of training to gradually surpass the baseline method. Regardless of the starting point, as the training phase advances, the performance of CNTS's detector and reconstructor continues to improve. This demonstrates that there is a cooperative

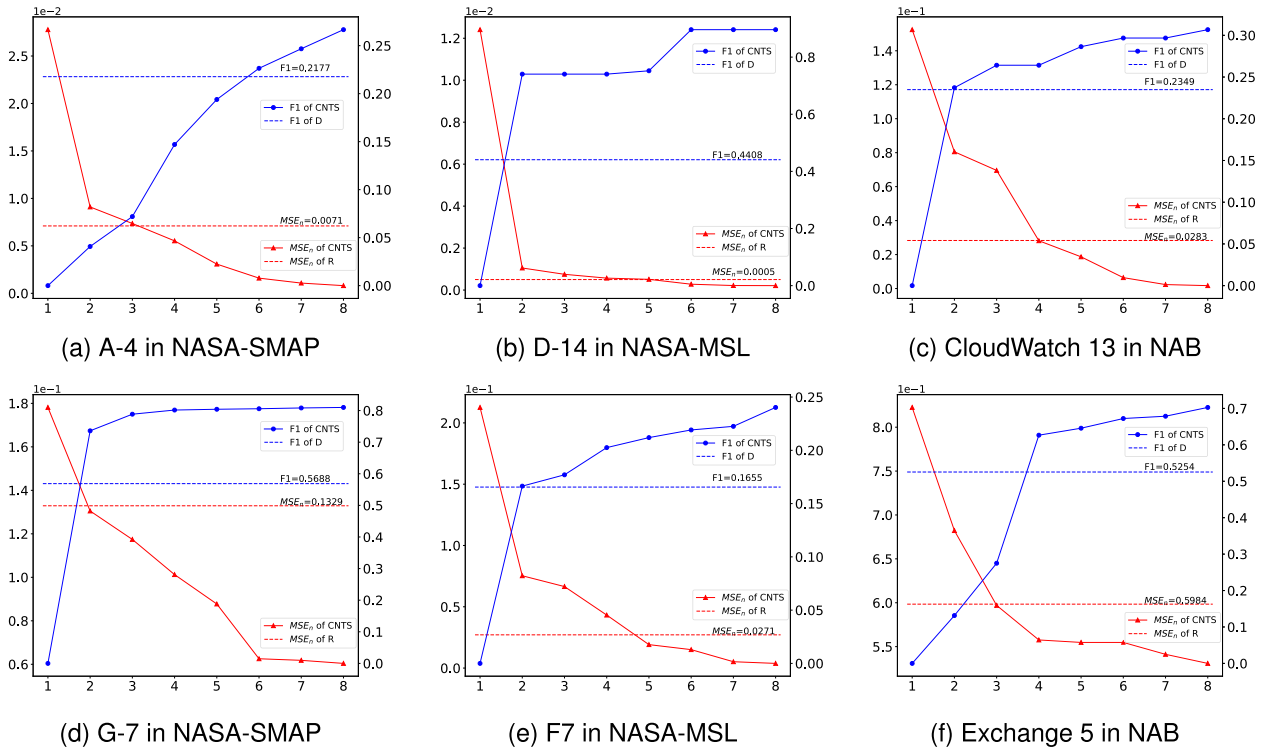


FIGURE 5. Metric changes of cooperative training. The solid lines represent the results of CNTS, and the dashed lines represent the results of single D or R.

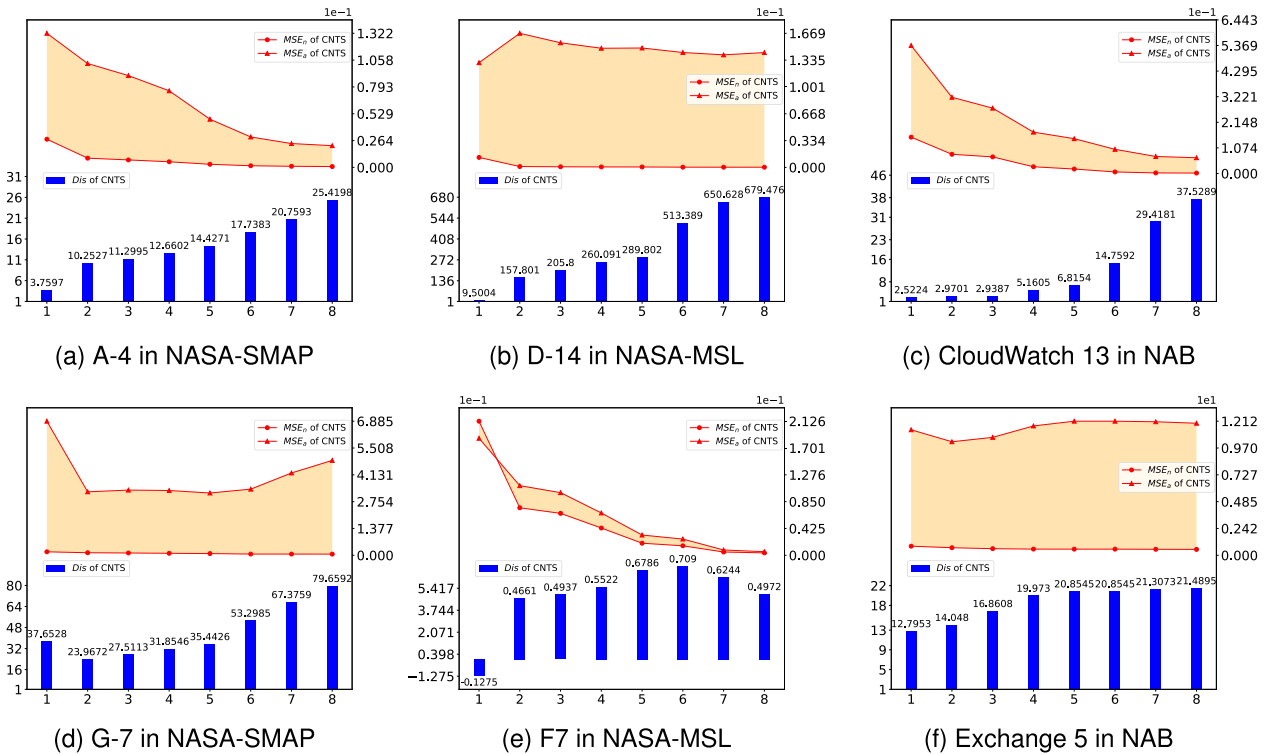


FIGURE 6. Metric changes of cooperative training. The red line represents the reconstruction error, with triangles and dots representing the reconstruction errors of outliers and normal values, respectively. The histograms depict the Dis of the model at different stages of training.

relationship between the detector and reconstructor during the training process, resulting in superior performance compared to training them individually.

Figure 6 illustrates the variations in various indicators during the training process of CNTS's R as the stage advances. The red line represents the reconstruction error, with triangles

and dots representing the reconstruction errors of outliers and normal values, respectively. The histograms depict the Dis of the model at different stages of training.

The line graph reveals that the reconstruction errors of normal values consistently decrease, demonstrating that the model is learning effectively. However, the reconstruction errors of outliers are more fluctuating. The histograms provide another perspective, showing that \mathbf{R} has the ability to differentiate between outliers and normal values. Although there may have been some fluctuations in earlier stages, the overall trend indicates that the Dis is increasing, which implies that the discriminative ability of \mathbf{R} improves as the number of model training increases.

VI. CONCLUSION

In this paper, we propose a novel cooperation-based unsupervised anomaly detection method for time series data, referred to as CNTS. This method is based on reconstruction loss and aims to address the limitations of existing reconstruction-based unsupervised anomaly detection methods by mitigating the impact of outliers on reconstruction performance. CNTS integrates a separate network dedicated to learning the anomaly detection task. The experimental results demonstrate that CNTS effectively helps the reconstruction model to distinguish between normal and abnormal time series data.

There are three aspects that can be further studied in the CNTS. One area of focus could be the selection of the basic model, such as FEDformer. Currently, it is not specifically designed for anomaly detection or reconstruction, and it may be beneficial to investigate the use of a dedicated network for these tasks. Additionally, in order to reduce the time complexity of CNTS, which results from its dual-network structure, one could alleviate it by incorporating a more efficient approach, such as missing value imputation, prediction, or classification, into a unified framework for processing time series data. Finally, further experimentation is useful to determine the optimal proportion of data to be selected or removed during the data selection process performed by detectors and reconstructors.

REFERENCES

- [1] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, Feb. 1969.
- [2] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," 2010, *arXiv:1009.6119*.
- [3] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 1, pp. 41–50, Feb. 2018.
- [4] Y. Liu, S. Garg, J. Nie, Y. Zhang, Z. Xiong, J. Kang, and M. S. Hossain, "Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6348–6358, Apr. 2020.
- [5] G. Pang, C. Shen, L. Cao, and A. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, 2020.
- [6] A. A. Cook, G. Misirlili, and Z. Fan, "Anomaly detection for IoT time-series data: A survey," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6481–6494, Jul. 2020.
- [7] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe, "Learning deep representations of appearance and motion for anomalous event detection," 2015, *arXiv:1510.01553*.
- [8] T. Kieu, B. Yang, C. Guo, and C. S. Jensen, "Outlier detection for time series with recurrent autoencoder ensembles," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 2725–2732.
- [9] Y. Li, X. Peng, J. Zhang, Z. Li, and M. Wen, "DCT-GAN: Dilated convolutional transformer-based GAN for time series anomaly detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3632–3644, Apr. 2021.
- [10] Z. Chen, D. Chen, X. Zhang, Z. Yuan, and X. Cheng, "Learning graph structures with transformer for multivariate time-series anomaly detection in IoT," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9179–9189, Jun. 2022.
- [11] S. Wang, Y. Zeng, X. Liu, E. Zhu, J. Yin, C. Xu, and M. Kloft, "Effective end-to-end unsupervised outlier detection via inlier priority of discriminative network," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.
- [12] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, "Adversarially learned one-class classifier for novelty detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3379–3388.
- [13] L. Ruff, R. K. Jacob, A. V. Robert, G. Montavon, W. Samek, M. Kloft, G. D. Thomas, and K. R. Müller, "A unifying review of deep and shallow anomaly detection," in *Proc. IEEE*, vol. 109, no. 5, pp. 756–795, Feb. 2021.
- [14] C. Yin, S. Zhang, J. Wang, and N. N. Xiong, "Anomaly detection based on convolutional recurrent autoencoder for IoT time series," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 1, pp. 112–122, Feb. 2022.
- [15] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 27268–27286.
- [16] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–19.
- [17] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S.-K. Ng, "MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2019, pp. 703–716.
- [18] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "USAD: UnSupervised anomaly detection on multivariate time series," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 3395–3404.
- [19] H. Zhao, Y. Wang, J. Duan, C. Huang, D. Cao, Y. Tong, B. Xu, J. Bai, J. Tong, and Q. Zhang, "Multivariate time-series anomaly detection via graph attention network," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 841–850.
- [20] Y. Zhang, Y. Chen, J. Wang, and Z. Pan, "Unsupervised deep anomaly detection for multi-sensor time-series signals," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 2118–2132, Feb. 2021.
- [21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [22] M. Braei and S. Wagner, "Anomaly detection in univariate time-series: A survey on the state-of-the-art," 2020, *arXiv:2004.00433*.
- [23] T. Nakamura, M. Imamura, R. Mercer, and E. Keogh, "MERLIN: Parameter-free discovery of arbitrary length anomalies in massive time series archives," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 1190–1195.
- [24] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 387–395.
- [25] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2828–2837.
- [26] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, and N. V. Chawla, "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 1409–1416.
- [27] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4027–4035.
- [28] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp. 134–147, Nov. 2017.

- [29] P. Benecki, S. Piechaczek, D. Kostrzewa, and J. Nalepa, "Detecting anomalies in spacecraft telemetry using evolutionary thresholding and LSTMs," in *Proc. Genetic Evol. Comput. Conf. Companion*, Jul. 2021, pp. 143–144.
- [30] H. Xu, Y. Feng, J. Chen, Z. Wang, H. Qiao, W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y. Liu, Y. Zhao, and D. Pei, "Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications," in *Proc. World Wide Web Conf.*, 2018, pp. 187–196.
- [31] S. Kim, K. Choi, H.-S. Choi, B. Lee, and S. Yoon, "Towards a rigorous evaluation of time-series anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 7, 2022, pp. 7194–7201.
- [32] A. Siffer, P.-A. Fouque, A. Termier, and C. Largouet, "Anomaly detection in streams with extreme value theory," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 1067–1075.
- [33] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *J. Roy. Statist. Soc. B Statist. Methodol.*, vol. 63, no. 2, pp. 411–423, 2001.
- [34] W.-S. Hwang, J.-H. Yun, J. Kim, and B. G. Min, "Do you know existing accuracy metrics overrate time-series anomaly detections?" in *Proc. 37th ACM/SIGAPP Symp. Appl. Comput.*, Apr. 2022, pp. 403–412.



YUANHAI SHAO received the B.S. degree in information and computing science from the College of Mathematics, Jilin University, in 2006, and the master's degree in applied mathematics and the Ph.D. degree in operations research and management from the College of Science, China Agricultural University, China, in 2008 and 2011, respectively. He is currently a Full Professor with the School of Management, Hainan University, Haikou, China. He has published over 100 refereed articles in his research areas, including *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, *IEEE TRANSACTIONS ON COMPUTERS*, *PR*, and *NV*. His research interests include support vector machines, optimization methods, machine learning, and data mining.



JINSHENG YANG received the bachelor's degree from Sichuan University, in 2016. He is currently pursuing the master's degree with Hainan University. His main research interest includes time series data analysis.



CHUN-NA LI received the master's and Ph.D. degrees from the Department of Mathematics, Harbin Institute of Technology, China, in 2009 and 2012, respectively. She is currently a Professor with the Management School, Hainan University. Her research interests include optimization methods, machine learning, and data mining.

...