

Technical paper

Multivariate time series anomaly detection: Missing data handling and feature collaborative analysis in robot joint data



Bo Yang ^{a,*¹}, Weishan Long ^{a,2}, Yucheng Zhang ^{b,3}, Zerui Xi ^{a,2}, Jian Jiao ^{a,1}, Yufeng Li ^{a,1}

^a State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, China

^b Seres Group Co. Ltd., Chongqing 401335, China

ARTICLE INFO

Keywords:

Missing data
Anomaly detection
Comprehensive imputation strategy
Collaborative graph attention mechanism
Strong feature correlation
Collaborative learning models

ABSTRACT

The efficient operation of industrial robots relies on reliable anomaly detection systems, but the problem of missing data caused by sensor failures, data transmission errors, and system maintenance or updates in the industrial field poses a serious challenge to their accuracy and may lead to delays in recognizing critical faults, which in turn threatens production safety and causes economic losses. To address this problem, this paper constructs an Anomaly Detection Method with Combined Data Imputation and Feature Collaboration (AD-CIFC), which firstly employs a comprehensive imputation strategy to impute the missing data in order to better adapt to the natural flow of time series and complex patterns; then two graph attention mechanisms are proposed to be used in parallel to capture the relationship between nodes more comprehensively by introducing synergies, thus effectively utilizing the correlation between features to make the model more sensitive; finally, two collaborative learning models are proposed to be used in parallel to more comprehensively take into account the strong correlation between features in the data, and a collaborative loss function computation method is proposed, thus improving the overall detection accuracy and discrimination ability. Through anomaly detection experiments on industrial robots with different levels of missing joint data, it is demonstrated that the AD-CIFC model demonstrates a performance that matches or even outperforms other mainstream anomaly detection models in complete fault data tests, while the model performance using the comprehensive imputation strategy to impute in the joints is about 25 % higher on average compared to the model performance of traditional linear interpolation methods, and meanwhile, in the case of missing data at the minor to severe levels, the model still demonstrates robustness beyond traditional methods, which is very suitable for industrial robot production lines.

1. Introduction

With the rapid development of automation technology, industrial robots have become a key force in driving the acceleration of production processes and ensuring high standards of manufacturing quality, especially in these areas of precision industry and automotive manufacturing [1]. These robots are responsible for a wide range of tasks in highly cooperative multi-jointed motions, from welding to precision assembly, where every link is critical [2,3]. Real-time monitoring of joint motion parameters such as angles, velocities, and moments is fundamental to ensuring robot performance and timely troubleshooting. However, with

the complexity of these robotic systems, even minor malfunctions can lead to the disruption of the entire production line, which in turn can cause significant economic losses [4]. Therefore, the development of accurate anomaly detection techniques is essential to maintain the stability and reliability of robotic systems. Particularly within the robot joints, there are complex interactions between multiple features, and the effective analysis and pattern recognition of the time series data generated by these features poses a major challenge [5–7]. In this context, deep learning techniques have become a research hotspot due to their potential in processing multivariate time-series data, and these techniques are particularly suitable for analyzing the synergistic effects

* Correspondence to: State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, 174 Shazheng Street, Shapingba District, Chongqing 400044, China.

E-mail address: yangbo61@cqu.edu.cn (B. Yang).

¹ Research area: intelligent manufacturing system

² Research area: advanced manufacturing technology

³ Research area: automotive engineering

of features within robot joints and allowing a more precise understanding of the complex interactions and temporal dependencies between the joints, which in turn provide new opportunities for designing and implementing more efficient anomaly detection algorithms [8–10].

At the same time, missing data is a common problem in industrial automation systems due to sensor failures (e.g., physical damage or calibration issues, which are particularly common in harsh operating environments), data transmission errors (including network failures and software bugs), and temporary interruptions during system maintenance or software updates [11,12]. In addition, the data acquisition and conversion process itself is a significant source of missing data, especially in configurations where multiple sensors share the same acquisition card, a setup that can lead to problems in data processing [13]. Also, the fact that robots operate in dynamic and complex environments increases the risk of missing data [14]. Missing data not only affects the real-time monitoring capability of the robot but also may interfere with the accuracy of the anomaly detection system, especially when dealing with multivariate time-series data, missing data points are particularly misleading and may lead to misinterpretation of the actual performance and state of the robot joints [15]. To cope with this problem, traditional interpolation methods have evolved to employ advanced machine learning-based techniques, such as deep learning, to process and reconstruct missing data more efficiently [16,17]. However, while these techniques perform well in data recovery, they are usually not optimized for highly dynamic and feature-rich industrial robot environments and are particularly deficient in integrating multiple sources of data and their complex interdependencies. Furthermore, while some anomaly detection systems integrate built-in data interpolation capabilities, these methods typically fail to consider the complex relationships between interdependent data features, leading them to cause erroneous fault diagnosis and delayed maintenance response when confronted with dynamic changes and frequent data missing that characterize industrial robotic systems, further exacerbating production losses [18,19]. Therefore, the development of a model that can adapt to data missing while maintaining efficient anomaly detection performance is of great significance in enhancing the stability and reliability of industrial robot systems, imputing an obvious gap in the existing technology for industrial applications.

In this paper, for the anomaly detection challenges of industrial robot multivariate time series data when facing various missing data scenarios, we propose an Anomaly Detection Method with Combined Data Imputation and Feature Collaboration (AD-CIFC). It employs an innovative comprehensive imputation strategy to impute the missing data regions, which takes into account the time-dependent and nonlinear characteristics of the data, and focuses on capturing complex feature relationships in multiple joints of industrial robots while maintaining their performance in the presence of missing data by collaborating feature and temporal graph-attention networks; followed by collaborative prediction and reconstruction modeling, which more comprehensively takes into account the strong correlation of features between data features' strong correlations, which makes the model more adaptable when facing different types of anomalies, AD-CIFC shows good results in solving the problem of anomaly detection in multivariate time series under different missing data situations. The main contributions of the methodological part of this study are as follows:

- 1) An innovative data imputation method is proposed: a comprehensive strategy of Adjacent Data imputation, Cosine Similarity imputation, and GraphSAGE imputation is used to consider the missing data regions comprehensively, and the average of Cosine Similarity results and GraphSAGE results are first computed, and then weighted and summed up with the results of Adjacent Data imputation. This strategy, which integrates the neighboring data information and advanced imputation techniques, takes into account the time-dependent and nonlinear properties of the data while maintaining

data integrity, and effectively enhances the stability and robustness of the model in the face of incomplete data.

- 2) Two graph attention mechanisms, Collaborative Graph Attention Mechanism (CGA) and Temporal Graph Attention Mechanism (TGA), are proposed and used CGA achieves comprehensive updating of graph structural information by means of a combination of collaborative features, while TGA focuses on dealing with the change of node relationships in the graph over time. By introducing synergy, the relationships between nodes are captured in a more comprehensive way, thus effectively utilizing the correlation between features to make the model more sensitive.
- 3) Two collaborative learning models, The Collaborative Forecasting-Based (CFBM) and Collaborative Reconstruction-Based Models (CRBM), are proposed and used jointly. Through this joint approach, we more comprehensively consider the strong correlation of features between the data and propose a new loss function calculation method for the case of missing data by utilizing the feature synergy, which makes the model more adaptive when facing different types of anomalies, and thus improves the overall detection accuracy and discriminative ability.

2. Related works

This section reviews the progress of related research in two ways, first summarizing the research on multivariate timing anomaly detection. It then briefly reviews recent advances in missing data imputation strategies, a section that currently has only a small number of research efforts.

2.1. Multivariate time series anomaly detection

In the field of anomaly detection, traditional machine learning methods like Isolation Forest (IF) [20] and Random Forest [21] have made significant contributions to identifying anomalies in high-dimensional data. These methods show high efficiency on high-dimensional datasets by effectively isolating anomalies, especially isolated forests isolate observations by constructing a special tree structure of isolated trees. However, these methods show some limitations when confronted with multivariate time series data characterized by complex dependencies and dynamic changes. With the evolution of deep learning technologies, new strategies have been proposed to address these shortcomings. Models such as Recurrent Neural Networks (RNN) [22], Long Short-Term Memory networks (LSTM) [23], and AutoEncoder (AE) [24] have excelled in processing long-term dependencies and complex dynamic relationships in time series data, especially LSTM is able to capture long term dependencies through its unique memory unit to effectively deal with time series data, and also solves the problem of gradient disappearance of RNN when handling long sequences, and AE effectively recognizes anomalous patterns by learning a low-dimensional representation of the data. Nevertheless, these methods still have challenges in highly nonlinear and large-scale multivariate data processing.

To address these challenges, researchers have explored innovative methods that incorporate modern techniques. For example, the SLMR [25] method efficiently extracts short-term local dependency patterns and long-term global trend patterns by combining masked self-supervised representation learning, multi-scale residual expansion convolution, and Gated Recurrent Units (GRUs). The innovation of the method lies in its spatiotemporal masked self-supervised representation learning and sequence segmentation strategy, which enables a deep understanding of temporal context and feature correlation and is suitable for the dynamic prediction of complex systems. Generative Adversarial Networks (GANs) such as GAN-Li [26] and MAD-GAN [27] can reveal complexity in data by incorporating Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN) into the GAN framework, which acts as both a generator and discriminator base model to capture

temporal correlation of time series and generate new data instances by modeling complex data distributions that are capable of revealing complex patterns and potential interactions between variables, providing more accurate inputs for anomaly detection. In particular, MAD-GAN also introduces the concept of a multi-agent discriminator, which enhances the understanding of complex interactions among variables in multivariate time series and further optimizes the performance of anomaly detection. MTAD-GAT [28] treats each univariate time series as a separate feature and contains two graph attention layers oriented to features and time in parallel to dynamically learn the complex interactions among different time series as well as the complex dependencies within the time series; meanwhile, the framework jointly employs prediction-based and reconstruction-based models, and this joint optimization strategy not only improves the performance of anomaly detection but also enhances the model's representation of Data Augmentation for time series data. Meanwhile, LSTM-VAE [29] utilizes the LSTM-VAE model to fuse and reconstruct multimodal sensor signals for state-based thresholding and anomaly detection at each time step. Meanwhile, the model innovatively introduces a prior based on task progression changes and is experimentally validated in a robot-assisted feeding task, where it demonstrates excellent performance in a detection task with 12 representative anomaly types. LSTM-NDT [30] uses LSTM-RNN as a predictor to model normal behavior and subsequently uses the cumulative sum method to identify anomalous behaviors. The innovation of this method is to provide a hybrid model that excels in handling abnormal patterns in complex multivariate time series data through a combination of deep learning and statistical methods. Notably, recent research advances have also provided cutting-edge techniques for multivariate time-series anomaly detection [31,32]. For example, Wang et al. [33] proposed a technique combining Adaptive Extended Kalman Filtering (AEKF) and One-Class Support Vector Machine (OCSVM), which optimizes sensor readings from Connected Automated Vehicles (CAVs) through a nonlinear vehicle following model to effectively cope with sensor anomalies and communication delays. In addition, Talagala et al. [34] developed a new approach for non-stationary environments to effectively identify anomalies in continuous data streams by reducing the time series to a two-dimensional feature space and using a bivariate two-sample nonparametric test. Munir et al. [35] proposed the FuseAD model that incorporates statistical ARIMA and deep learning CNN techniques to enhance the accuracy of anomaly detection. This fusion strategy not only improves the performance of the model in complex time series data but also demonstrates better AUC values than traditional methods in Yahoo Webscope benchmarks. However, despite the outstanding capabilities of these methods in time series analysis, their performance is still limited in handling large-scale datasets, meeting real-time performance demands, and coping with incomplete data. Particularly in environments with high levels of missing data, the performance of these models may significantly degrade as they rely heavily on complete data to learn and recognize anomalous patterns. Therefore, while existing techniques perform well on standard datasets, they may face challenges of practicality and effectiveness in specific real-world application scenarios, such as complex industrial environments.

In the application of anomaly detection in industrial robots, although related literature is relatively scarce, some innovative techniques have shown potential. For example, Zhong et al. [36] proposed a Sliding Window One-Dimensional Convolution AutoEncoder (SW1DCAE) method for unsupervised anomaly detection in industrial robots, directly applicable to raw vibration signals, effectively improving detection accuracy and validated with vibration signals collected from industrial robot test beds. Azzalini et al. [37] introduced a novel Variational AutoEncoder architecture, utilizing a new incremental training method for modeling very long multivariate sensor logs, inducing a progress-based latent space suitable for runtime and offline anomaly detection. Castellini et al. [38] defined a method based on Hidden Markov Models (HMM) for generating adversarial examples for anomaly

detectors, then proposed data augmentation and retraining techniques to enhance anomaly detection performance. He et al. [39] proposed a method for generalized fault diagnosis in industrial robot joints using a Multi-Joint Attention Residual Network (MJAR) model, incorporating a Multi-Space Reconstruction (MSR) module based on sparse sampling for multi-scale feature extraction adaptable to real industrial signals. Cui et al. [40] developed a multi-layer Recurrent Neural Network (RNN) approach for capturing and recognizing patterns in complex data collection scenarios, with experiments on industrial robots also validating the effectiveness of their proposed method. While these studies have made progress in their respective technical paths, they still face limitations in fully understanding the complex dynamic interactions of multivariate and multi-featured industrial robotic systems as well as data incompleteness.

Taken together, these advanced techniques have achieved remarkable results in the field of multivariate time series anomaly detection, especially in processing high-dimensional data and recognizing complex patterns, but there are certain limitations when these techniques are applied to industrial robots. The operation of an industrial robot system relies on the collaborative work of various components and their interactions under complex operating conditions, and the characteristics of its different joints (e.g., rotational speed, current, and rotation angle) do not exist independently, but rather interact with each other and jointly determine the overall performance and stability of the robot. However, most of the existing anomaly detection models focus on analyzing a single feature and ignore the interactions between these features, thus failing to comprehensively capture and understand the overall state of the system, resulting in less accurate or timely identification of the robot's anomalous state. It is worth noting that there are almost no studies in the current literature focusing on the collaborative analysis of industrial robot features, suggesting that this is an important area to be further explored and developed. In addition, missing data is a common problem in real-world applications due to sensor failures or communication errors. Most of the current anomaly detection techniques suffer a significant performance degradation when data integrity is compromised because they fail to process or impute the missing data efficiently. Therefore, an in-depth study of feature co-analysis and missing data handling not only provides a more comprehensive view of anomaly detection, especially in understanding dynamic interactions in complex systems, but also enhances the understanding of how industrial robotic systems operate as a whole, thereby optimizing performance and reliability.

2.2. Data imputation strategies

In industrial automation, ensuring the data integrity of a robotic system is essential to enable effective real-time monitoring and anomaly detection. The problem of missing data, whether caused by sensor failures or transmission errors, may seriously affect the analysis of multivariate time series and lead to a decrease in the accuracy of anomaly detection. Facing this problem, researchers have developed various data imputing techniques. First, simple interpolation methods, such as linear interpolation and median interpolation [16], were initially favored for their computational efficiency, but these techniques often failed to provide sufficient accuracy when dealing with time series data with complex relationships and dynamic patterns. With the advancement of time series analysis, deep learning techniques have been gradually utilized to estimate missing values. Gao et al. [41] first applied the "3-sigma" rule to eliminate outliers and erroneous data, introducing the Gray Relation Analysis-based K-Nearest Neighbors Imputation (GRAKNNI) algorithm, which completes anomalies and missing values based on spatiotemporal features. Then the convolutional bi-directional GRU network (SA-CNN-BiGRU) incorporating the self-attention mechanism achieves efficient and interpretable anomaly identification. Although this approach performs well in interpolating local data points, its performance may be limited when dealing with large-scale and

high-dimensional industrial data, and it does not adequately capture the nonlinear complex relationships in the data. Feng et al. [42] proposed a unified framework composed of multiple deep AutoEncoders (AEs) with skip connections, generating missing data from available sources in the latent space. Yang et al. [43] designed a deep probabilistic graphical model, SCNF, approximating the true complex posterior distribution of missing values through variational inference and deep neural networks, but in practice, it is computationally costly and extremely sensitive to the choice of parameters and susceptible to overfitting. Zhang et al. [44] developed an anomaly detection method based on DVAE, serving both as a probabilistic NBM for anomaly detection and a generative model for missing data recovery. Chen et al. [45] proposed a generative adversarial network (GAN) model based on a federated learning framework and a long and short-term memory network, known as the FedGAN model, in which a discriminator is used for anomaly detection and a generator is used for anomaly data repair in the GAN structure. However, its federated learning framework requires distributed training of the model on multiple nodes, which requires a long training time. Ma et al. [46] developed an anomaly detection algorithm for long-term structural health monitoring (SHM) systems based on probabilistic principal component analysis (PPCA) to evaluate the anomaly indices and their probability distributions, to process incomplete data and to recover lost data, but its static modeling framework is difficult to adapt in industrial environments where dynamics are changing rapidly. In addition, recent research advances have emphasized the importance of data integrity in improving equipment safety. Cai et al. [47] proposed a hybrid physical model and data-driven remaining useful life (RUL) estimation method based on dynamic Bayesian network (DBN) for submarine pipelines. The method copes with data insufficiency by collecting data from a physical model, which not only improves prediction accuracy but also highlights the critical role of data integrity in ensuring equipment safety. Cai et al. [48] used a Wiener process-based approach to re-predict the remaining useful life (RUL) of a subsea Christmas tree system and applied an expectation maximization (EM) algorithm to a dynamic Bayesian network (DBN) model to deal with the uncertainty caused by missing data. This further confirms the importance of maintaining data integrity in RUL estimation for complex systems. It is worth noting that the above methods, although effective in specific scenarios, generally suffer from deficiencies in dealing with complex and dynamically changing industrial data, especially in integrating multiple data sources and dealing with complex dependencies among data.

In the field of industrial robot anomaly detection in the face of missing data, although research is still in its early stages, there have already been studies discussing the feasibility of addressing this issue. Basurto et al. [49] applied traditional (such as Linear Regression) and innovative (Decision Trees and Neural Networks) models for comparative analysis, concluding through experimental analysis that the N-LR method yielded the best results in terms of MSE, while the RBFN method achieved the shortest time across all attributes. In the research on data imputation techniques and their application in robotic anomaly detection, while current advanced data imputation techniques play a significant role in addressing sensor data incompleteness, they often fail to fully consider the temporal dependencies and non-linear characteristics of data. These systems typically rely on continuous sensor readings to respond in real-time to environmental changes, making accurate prediction and imputation of these gaps crucial for maintaining stable system operation and timely identification of potential faults.

In our study, we combine recent developments in two fields, data impute techniques and robotic anomaly monitoring methods, to address the problem of missing data common to industrial robotic systems in dynamic environments and to significantly improve the accuracy of anomaly detection. Although many data impute methods have been extensively studied, they often exist independently of anomaly detection systems and fail to fully utilize the specific environment and requirements of robotic systems. Our approach not only integrates these techniques to optimize data integrity, but also enhances the system's

sensitivity and responsiveness to anomalous states by synergizing various data sources and features. The development of this approach imputes a gap in the application of existing techniques to industrial robot anomaly monitoring, especially in dealing with missing data and maintaining system stability.

3. Method

The aim of this paper is to solve the problem of anomaly detection in the case of industrial robots with some missing data for one or more key parameter features, and to explore how to make the model have a high anomaly detection recognition rate. To address this problem, we propose an Anomaly Detection Method with Combined Data Imputation and Feature Collaboration (AD-CIFC), and mainly divided into two parts: (a) Data reorganization model and (b) Subject anomaly detection model, as shown in Fig. 1.

When there is some missing data for one or several key parameter features of industrial robots, we first constructed a data reorganization model, preprocessed the data to obtain normal data, copied the normal data and processed it with random Mask, and then proposed a comprehensive strategy of Adjacent Data imputation, Cosine Similarity imputation, and GraphSAGE imputation to impute in the missing data regions thus form variant data, and splice with normal data to form a combined dataset containing variants, which lays the foundation for subsequent subject model training and realizes the purpose of data augmentation. Next, we constructed the subject anomaly detection model, analyzing the intra-group and global relationships of key features by Collaborative Graph Attention Mechanism (CGA), so as to effectively utilize the correlation between features to make the model more sensitive; meanwhile, we adopted Temporal Graph Attention Mechanism (TGA) is used in parallel in order to learn the curve features before and after the masked region, so that the potential anomalies in the partially missing data in the test set can be detected and recognized more effectively. Subsequently, through the joint use of The Collaborative Forecasting-Based and Collaborative Reconstruction-Based Models, we are able to effectively mine the interconnections between the features of the joint data, thus fully utilizing the advantages of the two models to improve the robustness and generalization ability of the overall model. Finally, the data reconstruction model and the subject anomaly detection model are combined to solve the problem of anomaly detection in industrial robots with some missing features of one or more key parameters.

3.1. Data reorganization model

(1) Data preprocessing: In order to ensure the stability of model training, eliminate the scale differences between different features, and also eliminate noise and outliers to ensure the reliability of model training, data normalization, and cleaning are required. We adopt the Min-Max normalization method to scale the time series data of the training and test sets to a range of 0 to 1, and perform data cleaning operations only on the training set.

(2) Data reorganization: First, we preprocess a piece of raw industrial robot data to obtain normal training data $x \in \mathbb{R}^{L \times n}$, where L is the total length of the time series, n is the total number of key parameter features of the robot, and at the same time, each joint of the robot has m features that are strongly correlated with each other. Next, a base normal data is replicated and the replicated dataset is subjected to dynamic random Mask processing: In each training iteration of the subject anomaly detection model, we randomly select different positions of 1 to $m - 1$ features in each joint of the robot to be Masked, and adopt a unified combination of missing rate v and missing length s to simplify the data processing process and to ensure consistency and manageability of the training, and to record the specific location of each Mask region. However, the missing data largely affects the accuracy of subsequent anomaly detection, so the reasonable design of the data-imputation

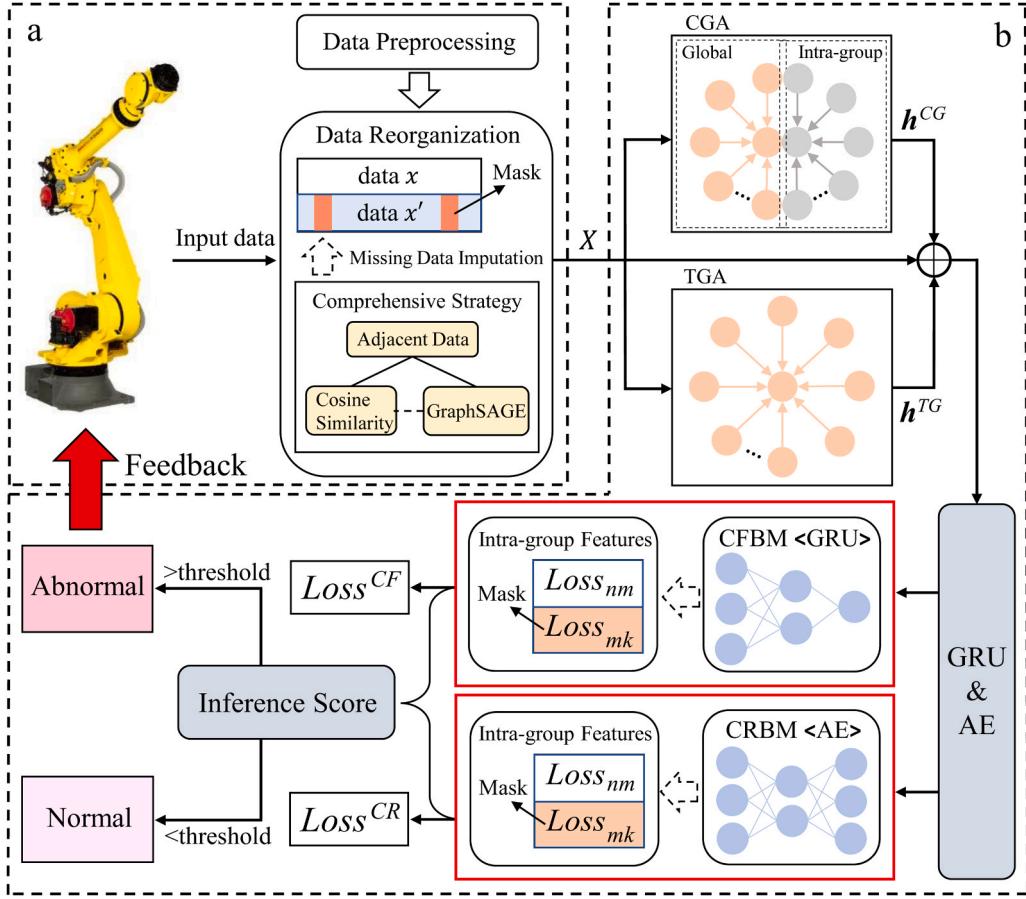


Fig. 1. An overview of our method for anomaly detection. ((a) Data reorganization model. (b) Subject anomaly detection model).

strategy is crucial to solving the problem of anomaly detection in missing data scenarios.

In order to impute in the missing data more accurately and reliably, we adopt the comprehensive strategy of Adjacent Data imputation, Cosine Similarity imputation, and GraphSAGE imputation to consider the missing data regions comprehensively. Among them, Adjacent Data imputation basically imputes based on physical or temporal proximity to ensure data continuity; Cosine Similarity imputation predicts missing values by comparing the similarity between time series, which is suitable for periodic or pattern-similar datasets; and GraphSAGE imputation makes use of the graph structural information and predicts the characteristics of the missing nodes by learning the node's neighborhood characteristics, which is an effective way to cope with the structural missing. Specifically, we realize the combined consideration of global feature similarity and local graph structure information for the data by calculating the average of the two methods, Cosine Similarity and GraphSAGE imputation. At the same time, hyperparameters are introduced to regulate the balance between this comprehensive result and Adjacent Data imputation, so as to achieve comprehensive consideration of short and long-term series and optimize the overall imputation effect. This approach not only imparts the accuracy of Data Augmentation but also enhances the adaptability of the model to complex data structures.

1. Adjacent Data imputation: For the time point t_{mk} of the Mask missing data region, we extracted the data segment of length $s/2$ before and after the Mask region respectively, without any computation, which is directly used as the imputing value of the missing region.

$$x_{t_{mk}}^{AD} = \begin{cases} x_{t_{mk}-(s/2)}, & t_{mk} < t_c \\ x_{t_{mk}+(s/2)}, & t_{mk} \geq t_c \end{cases} \quad (1)$$

where $x_{t_{mk}}^{AD}$ is the imputation value at time point t_{mk} using the Adjacent

Data imputation strategy; and t_c is the centrally located time point in the missing region.

2. Cosine Similarity imputation: We use the sliding window function to generate a fixed length of l as a reference window, the Mask region of the missing data points within the window is assigned the value of 0, and input the feature vector of the window, calculate the window with the highest Cosine Similarity and impute the corresponding missing region of the value. For the time point t_{mk} of the Mask missing data region, calculate the estimated impute value for that time point.

$$x_{t_{mk}}^{CS} = \left(\underset{A \in D}{\operatorname{argmax}} \left(\frac{A \cdot B}{\|A\| \|B\|} \right) \right)_{t_{mk}} \quad (2)$$

where $x_{t_{mk}}^{CS}$ is the imputation value at time point t_{mk} using the Cosine Similarity imputation strategy; A is the data window vector corresponding to the target time point t_{mk} ; D is the set of all data windows in the entire time series except for the window vector A ; B is one of the data window vectors in the set D ; argmax is the function used to find a window vector B to maximize the Cosine Similarity; $A \cdot B$ is the dot product between the vectors A and B ; $\|A\|$ and $\|B\|$ are the modulus lengths of the vectors A and B .

3. GraphSAGE imputation: We use the sliding window function to generate a fixed length of l as a reference window, impute the Mask region within the window with adjacent data, and use the GraphSAGE model to learn the embedding representation of each node within the window, while the edges between nodes are specifically defined as a single propagation in order to better capture dependencies in the time series data and a network containing layer k , and finally the missing point values are estimated by averaging the feature embeddings of the neighboring nodes. In each layer i ($1 < i < k$) of the GraphSAGE model, the features for any time point t_v within the window are updated by

$$f_{t_v}^{(i)} = \text{ReLU}\left(W^{(i)} \cdot \text{MEAN}\left(\left\{f_{t_u}^{(i-1)} : t_u \in P_l(t_v)\right\}\right) + B^{(i)}\right) \quad (3)$$

where $f_{t_v}^{(i)}$ is the feature representation of time point t_v in layer i ; $W^{(i)}$ is the weight matrix in layer i ; MEAN is the mean aggregation function, which calculates the average value of the features of the neighboring nodes; $f_{t_u}^{(i-1)}$ is the feature representation of time point t_u in layer $i-1$; $P_l(t)$ is the set of all the time points that precede the time point t , in a time window of length l ; and $B^{(i)}$ is the bias term in layer i . Once the model has been trained, the last layer of feature representations that have been learned can be used to predict the impute values for the time point t_{mk} in the missing region of the Mask.

$$x_{t_{mk}}^{\text{GS}} = \text{Pred}\left(f_{t_{mk}}^{(k)}\right) \quad (4)$$

where $x_{t_{mk}}^{\text{GS}}$ is the imputation value at time point t_{mk} using the GraphSAGE imputation strategy; and Pred is the prediction function used to convert the feature representation into an imputed value.

4. Combined result calculation: For the time point t_{mk} of the Mask missing data region, the average of the Cosine Similarity result and the GraphSAGE result is first calculated, and then a hyper-parameter δ , between 0 and 1, is used to adjust the weight between this average and the adjacent data imputation result. The combination of local continuity (Adjacent Data imputation), global similarity patterns (Cosine Similarity imputation), and local structural information (GraphSAGE imputation) provide a comprehensive view of data imputing.

$$x_{t_{mk}}^{\text{FN}} = \delta \left(\frac{x_{t_{mk}}^{\text{CS}} + x_{t_{mk}}^{\text{GS}}}{2} \right) + (1 - \delta)x_{t_{mk}}^{\text{AD}} \quad (5)$$

where $x_{t_{mk}}^{\text{FN}}$ is the combined result of using the comprehensive imputation strategy at time point t_{mk} .

Through this data comprehensive imputation approach, we finally obtain the variant data x' , that has been processed by random Mask, and the variant data has the same dimensions as the original normal data x . Subsequently, the original normal data is fused with the variant data to form a combined dataset containing variants, thus increasing the diversity of the data. The dataset for multivariate time series anomaly detection is denoted by $X = [x, x'] \in \mathbb{R}^{2L \times n}$, as shown in Fig. 2. Ultimately, the spliced dataset is fed into the subject model for training, which effectively improves the model's adaptability to various possible data variants. In addition, the dynamic randomization processing strategy adopted further enhances the performance of the subject model in generalization and real application scenarios.

3.2. Subject anomaly detection model

3.2.1. Graph attention mechanism

After the data processing work, we enter the subject anomaly detection model design phase. Here, we propose two graph attention mechanisms, Collaborative Graph Attention Mechanism (CGA) and Temporal Graph Attention Mechanism (TGA) and use them jointly to enhance the performance and adaptability of the model. CGA works to improve the model performance by introducing the synergies to capture the relationships between nodes in a more comprehensive way. It allows nodes to jointly focus on a set of related nodes, and achieves a comprehensive update of graph structure information through the combination of synergistic features in a way that enhances generalization, while providing support for a new loss function computation method proposed for the problem of partially missing data. TGA, on the other hand, focuses on dealing with node relationships in the graph over time. By introducing time dimension, temporal attention mechanism and dynamic graph update, TGA can effectively capture the time-varying relationships among nodes while integrating temporal features to enable the model to understand the evolution of node behaviors and relationships more comprehensively.

(1) Collaborative Graph Attention Mechanism (CGA): For the long sequence problem, we use the sliding window function to generate a fixed length of l , the number of key parameter features of n as the data input, and through the CGA method to study the relationship between the key parameter features, as shown in Fig. 3. The number of key parameter features n is used as nodes, and the length l intercepted by the sliding window function is used as the number of features in each node, i.e., the input set of node features is $h_i = \{h_1, h_2, \dots, h_n\}, h_i \in \mathbb{R}^l$. At the same time in accordance with the first node in order to select m nodes as a group, the next group will start from $m+1$ nodes and continue to select m nodes as a group, assuming that n can be evenly divided by m , there are a total of $g \in [1, n/m]$ groups. Consider intra-group interactions for each group and also global interactions for all n nodes, and finally a combination of intra-group and global influences for each feature.

1. Attention weight calculation (intra-group influence): For m key parameter features intra-group, calculate the attention weight.

$$e_{ij}^{\text{IN}} = \text{LeakyReLU}\left(\alpha_1^T W_g^{\text{IN}} h_i + \alpha_2^T W_g^{\text{IN}} h_j\right) \quad (6)$$

where e_{ij}^{IN} is the attention coefficient between nodes i and j in the intra-group; $\alpha = [\alpha_1 || \alpha_2] \in \mathbb{R}^{2l}$ is the parameter of intra-group learning, and $\alpha_1, \alpha_2 \in \mathbb{R}^l$, l is the transformed feature dimensions; $W_g^{\text{IN}} \in \mathbb{R}^{l \times l}$ is the weight matrix of group g where nodes i and j are located.

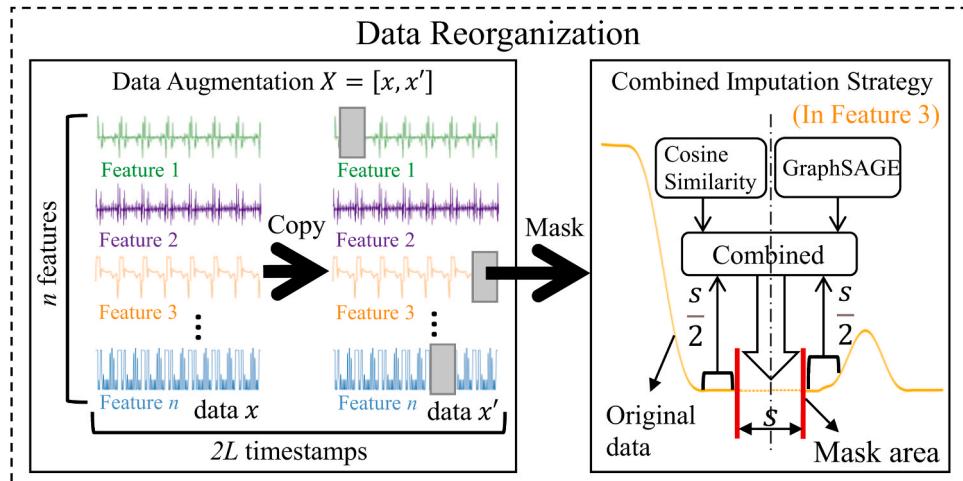


Fig. 2. Schematic diagram of data reorganization.

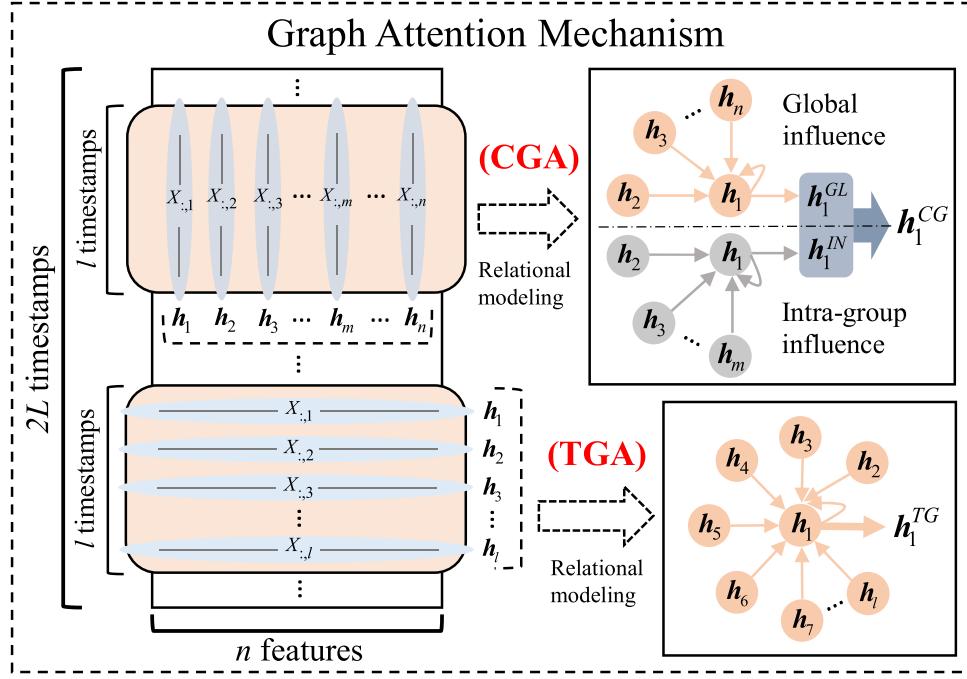


Fig. 3. Schematic diagram of CGA and TGA.

2. Attention weight normalization (intra-group influence): Use the softmax function to normalize the attention weight of intra-group features.

$$\text{attn}_{ij}^{IN} = \text{softmax}_j(e_{ij}^{IN}) = \frac{\exp(e_{ij}^{IN})}{\sum_{k \in N_g(i) \cup \{i\}} \exp(e_{ik}^{IN})} \quad (7)$$

where attn_{ij}^{IN} is the normalized attentional weight between node i and its neighboring node j in the group; $N_g(i) \cup \{i\}$ is the set of other neighboring nodes including node i in group g .

3. Updating the feature representation (intra-group influence): The feature representation is updated by weighted averaging of the features using the attention weights.

$$h_i^{IN} = \text{Sigmoid} \left(\sum_{j \in N_g(i) \cup \{i\}} \text{attn}_{ij}^{IN} \cdot W_g^{IN} h_j \right) \quad (8)$$

where h_i^{IN} is the localized effect of other nodes in the group on node i .

4. Attention weight calculation (global influence): For all n features, calculate the attention weight. We use the same method of intra-group influence for attention weight computation, adjusting the learning parameters to $\beta = [\beta_1 || \beta_2] \in \mathbf{R}^{2l^p}$, and $\beta_1, \beta_2 \in \mathbf{R}^{l^p}$, and using the global weight matrix $W^{GL} \in \mathbf{R}^{l^p \times l}$. Then the weight normalization is computed by the softmax function and the node representation is updated using the attentional weights:

$$h_i^{GL} = \text{Sigmoid} \left(\sum_{j=1}^n \text{attn}_{ij}^{GL} \cdot W^{GL} h_j \right) \quad (9)$$

where h_i^{GL} is the overall impact of other nodes globally on node i .

5. Combination of intergroup and individual direct impacts: weighted sum of intra-group and global impacts for each feature.

$$h_i^{CG} = \lambda \cdot h_i^{IN} + (1 - \lambda) \cdot h_i^{GL} \quad (10)$$

where h_i^{CG} is the final feature representation after fusing local and global influences, and λ is a weight parameter between 0 and 1 to balance the

importance of intra-group and global influences.

(2). Temporal Graph Attention Mechanism (TGA): In order to better learn the potential characteristics of the Masked region, we introduce TGA to capture the temporal dependence in the time series and enhance the correlation learning of the time series before and after the Masked region. At the same time, constructing multiple variant datasets that have been Masked, ultimately enables the model to learn and evaluate the characteristics of the missing part of the data more comprehensively and accurately. We take the length l of each sliding window function intercept as a node, and the number of key parameter features n as the number of features in each node, i.e., the input set of node features is $h_i^t = \{h_1^t, h_2^t, \dots, h_l^t\}, h_i^t \in \mathbf{R}^n$, and for each temporal node i we compute its weight with respect to the neighboring node j , as shown in Fig. 3.

For the sliding window function to generate a fixed length of length l , we use the same method of CGA for the attention weight computation, adjusting the learning parameters to $\gamma = [\gamma_1 || \gamma_2] \in \mathbf{R}^{2n'}$, and $\gamma_1, \gamma_2 \in \mathbf{R}^{n'}$; the weight matrix is adjusted to $W^{TG} \in \mathbf{R}^{n' \times n}$, and n' is the transformed feature dimension. Then the calculation of weight normalization is performed by the softmax function and the node representation is updated using the attention weights:

$$h_i^{TG} = \text{Sigmoid} \left(\sum_{j=1}^l \text{attn}_{ij}^{TG} \cdot W^{TG} h_j^t \right) \quad (11)$$

where h_i^{TG} is the overall effect of the other nodes in the time series on node i .

Finally, we combine the outputs of parallel co-features and temporal graph attention networks (CGA and TGA) and fuse them with the data-processed robot joint data. This approach not only improves the sensitivity of the model to learning different feature dimensions but also enhances the ability to recognize potential anomalies in partially missing data. Further, these fused data are fed into an integrated neural network module that combines GRU and AE to optimize the capability of analyzing long-time series trends. GRU focuses on dynamic time series forecasting, effectively supporting the Collaborative Forecasting-Based Model (CFBM); while AE is responsible for data reconstruction and anomaly identification, which is the core of the Collaborative Reconstruction-Based Model (CRBM). This design framework not only

clarifies the functional positioning of each module but also provides the theoretical foundation and data support for an in-depth discussion of CFBM and CRBM models.

3.2.2. Collaborative optimization

Here, we propose The Collaborative Forecasting-Based (CFBM) and Collaborative Reconstruction-Based Models (CRBM) and use them jointly to address the strong correlation of features among joints of welding industrial robots, as shown in Fig. 4. Among them, the CFBM part captures the underlying patterns and trends of the data, while the CRBM can effectively deal with local anomalies and missing data. Through this joint approach, we more comprehensively consider the strong correlation of features among the data, which makes the model more adaptable in the face of different types of anomalies, and thus improves the overall detection accuracy and discriminative ability. Because for a time series in which one or more features are partially missing, the other features in their corresponding groups are strongly correlated with the missing data features, we propose a new loss function calculation method for the missing data case by utilizing feature synergy. If feature i of a time series is in the unmasked region at a certain time, the loss function is calculated as normal, and if it is in the masked region, the loss function is calculated by defining a hyperparameter ϵ between 0 and 1 and weighting the average of the other features in the group corresponding to the feature, to sum up, the loss function. For a segment of missing data, we can judge it by the joint information of other features in the group, so as to improve the anomaly detection ability for part of the missing data.

(1) The Collaborative Forecasting-Based Model: In CFBM, we combine a gated recurrent unit (GRU) and a fully connected layer focusing on the prediction of future values of time series data. The GRU optimizes the information flow through its gating mechanism and effectively handles time dependency, especially when the data are incomplete; the fully-connected layer makes predictions based on the output of the GRU and optimizes the performance of the model by using the mean-square error (MSE) as a loss function. $\hat{x}_{t,i}^{CF}$ denotes the predicted value of feature i modeled at time t and $x_{t,i}$ denotes the true value of feature i at time t . For the i th feature at time t the underlying MSE is

computed as:

$$\text{Loss}_{t,i}^{CF} = (\hat{x}_{t,i}^{CF} - x_{t,i})^2 \quad (12)$$

Next, depending on whether the feature belongs to the Mask region or not, we adjust the MSE calculation. For features in non-Mask regions use the standard MSE calculation, for features in Mask regions, find the group g in which the feature is located, calculate the MSE mean value of non-Mask features within the group, and then we use the hyperparameter ϵ to combine the MSE mean value with the feature's own MSE for joint consideration. Assuming that $S_g(i)$ is the set of all non-Mask features within group g , and $G_g(i)$ is the number of non-mask features within the group, the adjusted MSE for features within the Mask region is:

$$\text{Loss}_{t,i(adjusted)}^{CF} = \epsilon \text{Loss}_{t,i}^{CF} + \left(\frac{1-\epsilon}{G_g(i)} \right) \cdot \sum_{j \in S_g(i)} \text{Loss}_{t,j}^{CF} \quad (13)$$

Thus, the total loss of CFBM at time t is the summed value of the MSEs (adjusted and unadjusted) for all features:

$$\text{Loss}_t^{CF} = \sum_{i=1}^n \text{Loss}_{t,i(adjusted)}^{CF} \quad (14)$$

(2) The Collaborative Reconstruction-Based Model: In CRBM, we use AE to enhance the ability to reconstruct the original data, and evaluate the reconstruction quality through the MSE to detect data anomalies. AE deeply learns the intrinsic characteristics of the data through its encoding and decoding mechanism and optimizes the identification of missing or abnormal parts, thereby improving the reproduction of data details and the accuracy of anomaly detection. Its calculation method is the same as that of CFBM, replacing $\hat{x}_{t,i}^{CF}$ in its basic MSE formula with $\hat{x}_{t,i}^{CR}$, which indicates the reconstructed value of feature i in the model at the moment of t . At the same time, we determine whether the feature is currently in the Mask region, and then we use the same coefficient ϵ as CFBM to jointly consider the MSE mean value of non-Mask features in the group g where the feature is located and the MSE of the feature itself. Ultimately, the total loss of CRBM at time t is the summed value of the

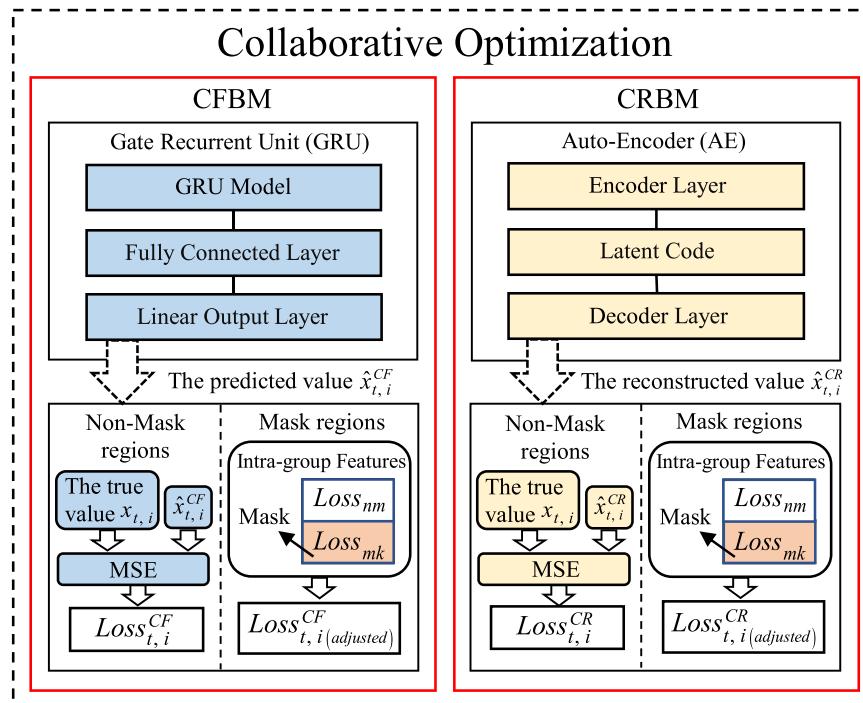


Fig. 4. Joint optimization framework using CFBM and CRBM.

MSEs (adjusted and unadjusted) of all features:

$$\text{Loss}_t^{\text{CR}} = \sum_{i=1}^n \text{Loss}_{t,i\text{ori}(\text{adjusted})}^{\text{CR}} \quad (15)$$

3.2.3. Anomaly detection

For joint optimization, we have two inference results for each timestamp. One is obtained based on the co-prediction model and the other is obtained based on the co-reconstruction model. The final inference score balances their benefits to maximize the overall effectiveness of anomaly detection. By setting a hyperparameter μ between 0 and 1, we combine the prediction error with the reconstruction probability, compute the inference score $s_{t,i}$ for each feature at moment t , and use the sum of all features as the final inference score. Specifically, the inference score can be calculated by the following method:

$$\text{Score} = \sum_{i=1}^n s_{t,i} = \sum_{i=1}^n \frac{\text{Loss}_{t,i\text{ori}(\text{adjusted})}^{\text{CF}} + \mu \text{Loss}_{t,i\text{ori}(\text{adjusted})}^{\text{CR}}}{1 + \mu} \quad (16)$$

For the selection of global thresholds, we adopt an approach similar to best-F1 in previous work. By adjusting the threshold, we evaluate the balance between precision and recall of the model and select the optimal global threshold. In the task of multivariate time series anomaly detection, when the inference score exceeds this threshold, we mark the data as anomalous and finally output a vector $y_t \in \{0, 1\}$ indicating whether the t th timestamp is anomalous or not, thus realizing the goal of anomaly detection. For hyperparameters δ , λ , ε , and μ are selected by grid search on the validation set.

4. Experimental results

In order to evaluate the performance of the proposed methods and models and their advantages, we used the operational data of a six-axis welding industrial robot in an automotive production line for test and analysis.

4.1. Dataset

The operating signals of the welding robot are collected in real-time by the built-in sensors of the individual joints and are received by the control software and stored in the network storage facility of the production line. All data collected in this paper is derived from actual field data from a New Energy Vehicle (NEV) production line (see Fig. 5) in Chongqing, China, from the welding robot's control system (BOS6000) and the associated production line database.

On a body-in-white welding line, the welding robot operates according to a preset program, sequentially moving the end welding actuator to the specified position and performing the welding task.



Fig. 5. Automated welding robot in car manufacturing process.

These operations are reflected in the robot's operating signal data, which exhibit phase and cycle characteristics. The key signals collected cover the rotational Speed (SPD), Current (CUR), and Rotation Angle (RA) of the six joints, which effectively reflect the real-time operating status of the industrial robot, i.e., the joints are characterized by $m = 3$ strong correlation with each other. It should be noted that due to the extreme lack of on-line abnormal cycle signal data of welding robots in the industrial field, we collected and analyzed part of the operation signal data of joint 2 and joint 4 during the maintenance and debugging period of the robot, as shown in Fig. 6, with the red areas being the abnormal signal locations. We observed strong correlations between the joint features during the actual overhaul period, but due to the low percentage of anomalous signals in the dataset, these data are not suitable for use as a standard test set, nor can they be adequately used in a missing data (Mask) test environment. In order to more effectively simulate the real-time operating state of the robot and enhance the recognition ability of the model, we adopt the strategy of injecting random Gaussian noise on top of the existing normal state data. Specifically, we considered injecting a fault rate of 12.826 % for single-joint 2 % and 13.259 % for double-joints 2 and 4, while accurately recording the location of each noise injection during the injection process and labeling these locations as abnormal, thus artificially generating more distinct fault features. The use of this ratio avoids the risks associated with too low anomaly rates: If the anomaly rate is set too low, there is a risk that missing data may not be able to cover the anomalous locations, thus limiting a full assessment of the actual anomaly detection capability of the model.

For the training and test sets, we have rationally designed them to accommodate the different experimental requirements of the training and test phases, as shown in Table 1. The dataset in the training phase is the normal signal after preprocessing of the original data, while the dataset in the testing phase employs faulty data and contains different degrees of missing data to simulate the diverse data integrity challenges that may be encountered in real applications. Specifically, the test set contains three combinations of faulty data with different levels of missingness: The first one is 20 data points in length with a missing rate of 10 % to simulate a mild missing data scenario; the second one is 30 data points in length and a missing rate of 15 %, with the same setup as the training set, for evaluating the performance of the model under a moderate level of missing data; and the third one is a more severe scenario of length of 40 data points and a missing rate of 20 %. In addition, the test set also includes the complete faulty data without any missing processing, which is used for comparative analysis with the missing-processed data, so as to comprehensively evaluate the model's ability to handle data with varying degrees of faults.

4.2. Experimental procedure

In the process of measuring the model performance, we used three key metrics: precision (P), recall (R), and F1 score (F1). In the training phase of the data reorganization model, we selected the normal signal dataset after data preprocessing for training with a uniform missing rate of 15 % and missing length of 30 data points combination strategy, while the GraphSAGE model was trained with the Adam optimizer with 100 epochs, the initial learning rate was set to 0.001, and a 2-layer GraphSAGE model to capture the relationship between neighbors, and the root mean square error (RMSE) was used as the loss function for model training. In addition, in the training phase of the subject anomaly detection model, we selected a combined signal dataset containing variants for training and trained 100 epochs using the Adam optimizer with an initial learning rate set to 0.001 and a batch size of 256. All experiments were conducted under the deep learning framework Pytorch, while 20 % of the training data was set aside as the validation set, and a sliding window $l = 100$ of the same size with a sliding step of 1 was used to ensure the uniformity of the experimental conditions and the reliability of the results. By performing a grid search on the vali-

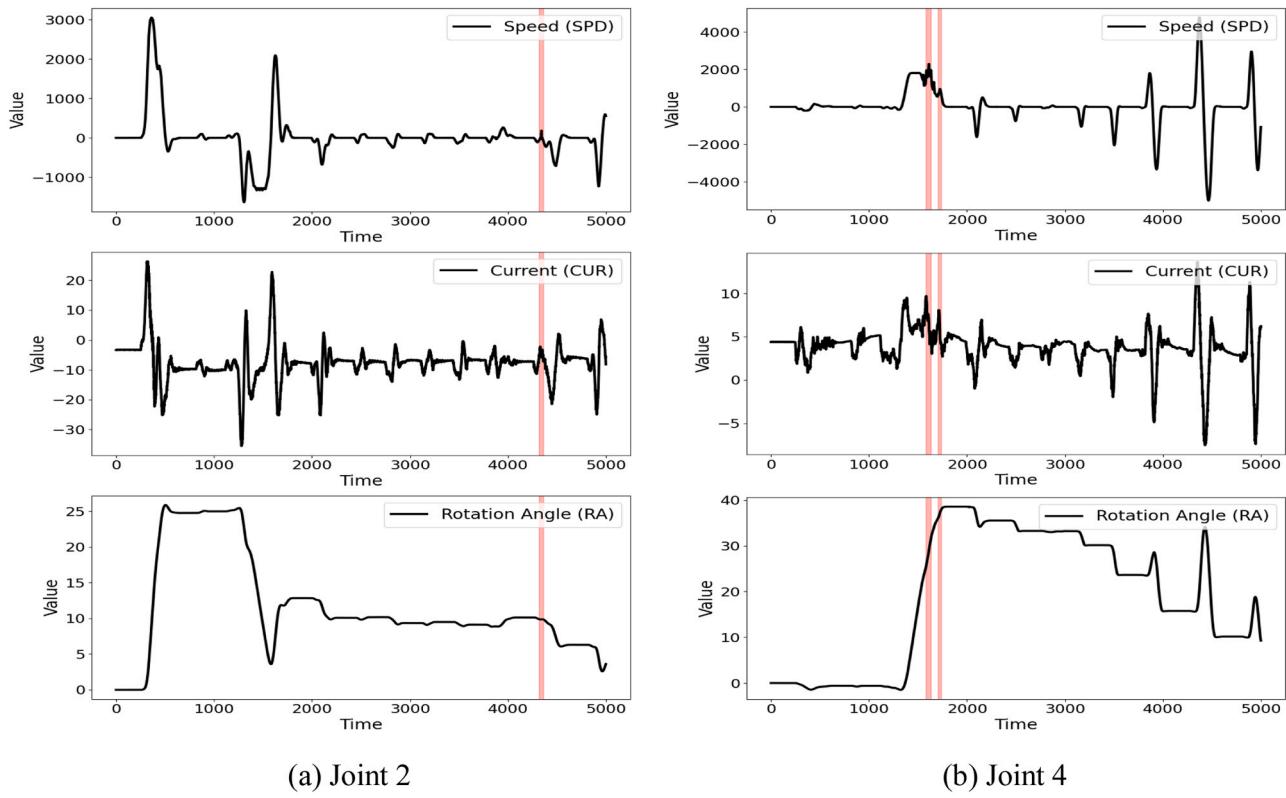


Fig. 6. Signal data and anomaly markings for two joints in maintenance state.

Table 1
Dataset configurations for training and testing.

Dataset	Description	Signal Variables	Data Volume	Anomaly Rate (%)	Missing Rate	Missing Length
Training Set	Normal Signals	cur , spd , ra	65,000	0	0	0
Test Set	Joint 4 Fault	cur , spd , ra	26,000	12.826	0	0
	Joint 4 Fault with Missing-1	cur , spd , ra	26,000	12.826	10 %	20
	Joint 4 Fault with Missing-2	cur , spd , ra	26,000	12.826	15 %	30
	Joint 4 Fault with Missing-3	cur , spd , ra	26,000	12.826	20 %	40
	Joints 2 and 4 Fault	cur , spd , ra	26,000	13.259	0	0
	Joints 2 and 4 Fault with Missing-1	cur , spd , ra	26,000	13.259	10 %	20
	Joints 2 and 4 Fault with Missing-2	cur , spd , ra	26,000	13.259	15 %	30
	Joints 2 and 4 Fault with Missing-3	cur , spd , ra	26,000	13.259	20 %	40

dation set, the hyperparameters δ , λ , ε , and μ are set to 0.4, 0.7, 0.3, and 0.8, respectively.

During the experimental testing phase, we performed a detailed randomized Mask treatment on three key performance metrics (SPD, CUR, and RA) of the robot joints in the test set. This processing consisted of six Mask operations on these metrics individually and in two-by-two combinations. After each operation, we summed up the results and computed the average to serve as a basis for subsequent analysis, and uniformly used a comprehensive imputation strategy to impute the missing data regions of the test set. The experiments first focused on a benchmark missing combination with the same settings as the training set, i.e., a missing length of 30 data points and a missing rate of 15 %. The purpose of the in-depth analysis of this combination is to evaluate the performance of the AD-CIFC model by comparing it to the current mainstream multivariate time series anomaly detection models, which include IF [20], AE [24], SLMR [25], GAN-Li [26], MAD-GAN [27], MTAD-GAT [28], LSTM-VAE [29], and LSTM-NDT [30]. In order to comprehensively evaluate the performance of AD-CIFC in data integrity and missing scenarios, we specifically introduce the DVAE [44] and FedLGAN [45] models designed for the missing data problem for comparative analysis. Meanwhile, in order to assess the performance of

the model under the complete data conditions, we also used the unmasked dataset to compare with the processed data. In addition, we conducted ablation experiments on the AD-CIFC model to explore the specific effects of different imputing strategies as well as key components of the model on the model performance. Subsequently, the experiments turned to evaluate two other different combinations of missingness: A length of 20 data points with a 10 % missingness rate, and a length of 40 data points with a 20 % missingness rate. This step allowed us to comprehensively assess and compare the model's adaptability and generalization ability in the face of different missing data conditions. Finally, in order to verify the validity and applicability of the model in real-world applications, we conducted a model validation phase using some of the operational signal data collected for robot overhaul and commissioning.

It is worth noting that any Mask labeled regions encountered by the model during training and testing are not automatically determined to be anomalous, but rather by whether the Score exceeds the threshold. In addition, we indicate by red color the abnormal regions correctly identified by the model (True Positives); the green color indicates the normal regions correctly identified by the model (True Negatives); and the yellow color indicates the misjudged regions of the model, including

False Positives (i.e., the model incorrectly labels the normal regions as abnormal) and False Negatives (i.e., actual anomalies that the model failed to detect).

4.3. Hyperparameter sensitivity analysis

In the hyperparameter sensitivity analysis portion of this study, we employed a controlled variable approach by designing a series of experiments to evaluate the performance of the model under different scenarios. Specifically, the experiments were divided into two main scenarios: one with only faults and no missing data, and another more complex scenario with both faults and concomitant missing data, and focused on four key hyperparameters: the weights for missing data handling (δ), the graph attention network weights (λ), the co-optimization weights (ε), and the scoring threshold adjustment weights (μ). These hyperparameters play a key role in handling missing data and anomaly detection, and we use sensitivity experiments to gain a deeper understanding of the specific impact of each hyperparameter on the performance of the model, especially on the F1 score, as shown in Fig. 7. The test results for each hyperparameter are averaged over the results of the corresponding test set, and such an approach not only ensures the reliability of the experimental results but also facilitates the evaluation of the model's generalization ability over multiple test sets.

The experimental results show that for the case of faults only without missing data, as shown in Figs. 7(a) and 7(c), we find that the role of δ and ε is relatively limited in this scenario, which may be because missing data does not occur, thus reducing the reliance on data impute strategies. Meanwhile, as shown in Figs. 7(b) and 7(d), we notice that the fine-tuning of hyperparameters λ and μ fails to significantly improve the model performance, although it has an impact on the model performance. This may imply that the model is less sensitive to these hyperparameters in fault detection scenarios without missing data or indicate that the optimization potential of these parameters is limited. In contrast, in the scenario with both faults and missing data, the fine-

tuning of the hyperparameters significantly affected the F1 scores of the model. In particular, as shown in Fig. 7(a) and (d), the tuning of hyperparameters δ and μ improved model performance particularly, highlighting the important role of these two hyperparameters in optimizing the data impute strategy and scoring threshold setting. This suggests that the model's ability to handle missing Data Augmentation and recognize anomalies can be significantly enhanced by fine-tuning these key parameters. In comparison, the tuning of hyperparameters λ and ε , as shown in Fig. 7(b) and (c), contribute to the model but its impact is relatively small. These two hyperparameters focus on the information weighting of the graph attention network and the joint optimization mechanism of the model, respectively, demonstrating the potential of hyperparameter sensitivity tests in enhancing the comprehensive identification ability of the model.

Considering the results of the sensitivity experiments, we finally determined the optimal values of hyperparameters δ , λ , ε , and μ as 0.4, 0.7, 0.3, and 0.8, respectively. This setting not only achieves the optimal performance in complex scenarios with faults and missing data but also maintains the stable performance of the model in the case of faults only. In addition, our study reveals the importance of hyperparameter tuning during data processing and model optimization, and also provides effective methodological support for further improving model performance. These findings are expected to provide more accurate and reliable technical support for anomaly detection and missing data processing in practical applications.

4.4. Model performance analysis

4.4.1. Performance comparison

In the performance comparison section of this study, we thoroughly compare the AD-CIFC model with several current mainstream anomaly detection techniques, including IF, AE, SLMR, GAN-Li, MAD-GAN, MTAD-GAT, LSTM-VAE, and LSTM-NDT, and introduce the DVAE and FedLGAN models designed for the missing data problem for a

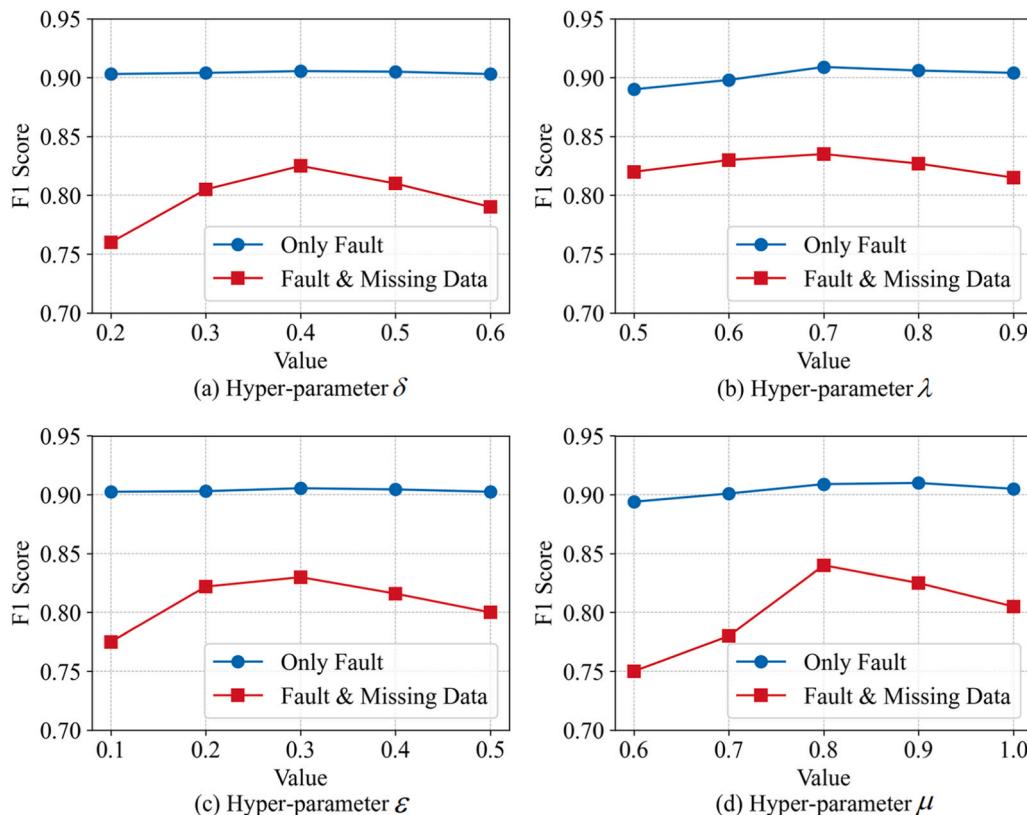


Fig. 7. Sensitivity of hyperparameters across fault and data missing scenarios.

comprehensive comparative analysis, as shown in [Table 2](#). Firstly, performance comparison under data-complete conditions: Our AD-CIFC model shows similar or even better performance than other mainstream models, including DVAE and FedGAN, in the complete fault data test, which highlights its high efficiency in recognizing standard anomaly patterns. Secondly, performance comparison under partially missing data conditions: In the missing data scenario with a missing length of 30 data points and a missing rate of 15 %, we are able to find a significant performance degradation of the other models, which indicates that they are unable to accurately recognize the key information in the missing data, and exposes their limitations in dealing with incomplete data. On the contrary, our model still achieves high accuracy, recall, and F1 value even under the condition of partially missing data, which proves its significant robustness in dealing with missing data. Specifically, IF and AE incurred significant performance loss in the presence of missing data, indicating that these models are highly sensitive to data completeness; whereas LSTM-VAE and LSTM-NDT, despite their good performance on intact data, appeared to be insufficiently adaptive to missing data; furthermore, the deep learning-based GAN-Li, MAD-GAN, and MTAD-GAT showed significant robustness to missing data in the GAN-Li, MAD-GAN, and MTAD-GAT based on deep learning show some adaptability in the face of missing data, but still fail to reach the level of our model. As for SLMR, it employs random Mask portions of data in the data preprocessing stage to infer representations in incomplete time series, and exhibits some robustness to anomaly detection in partially missing data, but it still lags behind our method in the comprehensive performance evaluation. Meanwhile, the DVAE model utilizes its denoising variational self-encoder design to specifically optimize its ability to handle missing data, while the FedGAN model utilizes federated learning and generative adversarial networks to exhibit high recall in handling complex missing data and anomaly detection scenarios. Although DVAE and FedGAN are designed to have their advantages in specific missing data scenarios, the AD-CIFC model provides a more comprehensive and stable performance in a wide range of anomaly detection application scenarios.

4.4.2. Demonstration of model performance with missing data

In the Model Performance Demonstration with Missing Data section of this study, we implement anomaly detection across different regions under fault and data impute (Masked) conditions for joints 2 and 4 in the double joint with an anomaly rate of 13.259 %, a Mask length of $s = 30$, and a missing data rate of $v = 15\%$, and show all the scenarios for a single versus a combination of the two and the two, as shown in [Figs. 8 and 9](#) show. First of all, we analyze Joint 2 in-depth, in [Fig. 8\(a\)](#) we can see that the SPD feature has carried out the Masked operation, and its real data curve is a straight line after data impute, at the same time, the CGA and the TGA in our model realize the comprehensive update of the graph structure information through the combination of synergistic features, and with the joint optimization of the CFBM and the CRBM, the model generates the prediction and reconstruction curves fluctuate

synchronously with other features in the missing regions. Eventually, the model correctly identifies this region as an anomalous region Area1 through the synergistic loss function calculation method. intuitively, we can infer that this region may have originally been an anomalous protrusion, and this finding confirms the ability of our model to detect anomalies under the condition of missing data. Similarly, it is a similar situation in [Fig. 8\(b\)](#), where our model is still able to correctly determine the normal region Area1 and the abnormal region Area2 after the SPD and RA features have gone through the Masked operation, a process that demonstrates the model's continued stability in anomaly detection. However, in [Fig. 8\(c\)](#) the model has misjudged the situation. After the RA features have gone through the Masked operation, probably due to the larger fluctuation of the SPD and CUR features in the data itself, their thresholds are also larger in comparison to the RA features as well, although the anomaly scores of these features do not exceed the respective thresholds of the SPD and the CUR, through the calculation of the synergistic loss function, the loss function of the RA features has also disguisedly increased and exceeded its own threshold, and it also incorrectly Area1 in the RA feature is judged as an anomalous area. This phenomenon reveals a challenge for our model and also indicates an important direction for future model optimization. In addition, the significant difference between the predicted and reconstructed values observed in [Fig. 8\(c\)](#) also suggests that we need to adjust the sensitivity of the model's response to different anomalous features.

Then we explore and analyze joint 4. In [Fig. 9\(a\)](#), we can see that after the Masked operation of CUR and RA features, the model is still able to identify the normal area Area1 and abnormal area Area2 normally, and the prediction and reconstruction curves generated by the model continue to fluctuate in the missing area in synchronization with the other features after the Masked operation on RA. This further demonstrates the sophistication of our model under missing data conditions. However, in [Fig. 9\(b\)](#), we can see that after the Masked operation on the SPD and CUR features, even though the SPD and RA features fluctuate to a certain extent in Area1, the model exceeds the thresholds set by the model and correctly determines it as an abnormal area. However, it may be because the faults of the key features in the area are not obvious, coupled with the fact that Area1 in the CUR features after passing the Masked process imputes in a certain amount of normal data for the missing area, after the comprehensive imputation strategy, which leads to its calculation by the synergistic loss function does not exceed the set threshold and incorrectly determines Area1 in the CUR features as a normal area. This may reflect that the fault features in this area are not obvious enough, revealing certain limitations of our model in recognizing milder fault conditions and providing important insights for future improvement. As for Area 2, the model is still able to recognize the normal area normally. Finally, in [Fig. 9\(c\)](#), we can see that after the Masked operation on the CUR features, our model successfully generates prediction and reconstruction curves similar to the real data curves and is able to correctly determine Area1 as an abnormal area, which also shows that our model maintains a high degree of sensitivity to the data

Table 2
Performance comparison of anomaly detection models on differing data conditions.

方法	关节4故障			关节4故障缺失-2			关节2, 4故障			关节2, 4故障缺失-2		
	P	R	F1									
IF	0.910	0.768	0.833	0.388	0.448	0.416	0.864	0.679	0.760	0.372	0.456	0.410
AE	0.937	0.739	0.826	0.453	0.632	0.528	0.735	0.681	0.707	0.407	0.480	0.440
SLMR	0.889	0.917	0.903	0.663	0.755	0.706	0.895	0.830	0.861	0.647	0.680	0.663
GAN-Li	0.686	0.928	0.789	0.625	0.760	0.686	0.682	0.836	0.751	0.559	0.709	0.625
MAD-GAN	0.891	0.783	0.834	0.666	0.729	0.696	0.820	0.866	0.842	0.593	0.742	0.659
MTAD-GAT	0.854	0.862	0.858	0.671	0.741	0.704	0.879	0.918	0.898	0.625	0.693	0.657
LSTM-VAE	0.805	0.936	0.866	0.633	0.485	0.549	0.771	0.956	0.854	0.545	0.625	0.582
LSTM-NDT	0.935	0.742	0.827	0.521	0.530	0.525	0.893	0.755	0.818	0.481	0.771	0.592
DVAE	0.901	0.870	0.885	0.766	0.810	0.787	0.875	0.864	0.869	0.715	0.788	0.750
FedGAN	0.909	0.892	0.900	0.721	0.879	0.792	0.892	0.907	0.899	0.673	0.816	0.738
AD-CIFC	0.895	0.919	0.907	0.792	0.875	0.831	0.889	0.915	0.902	0.743	0.857	0.796

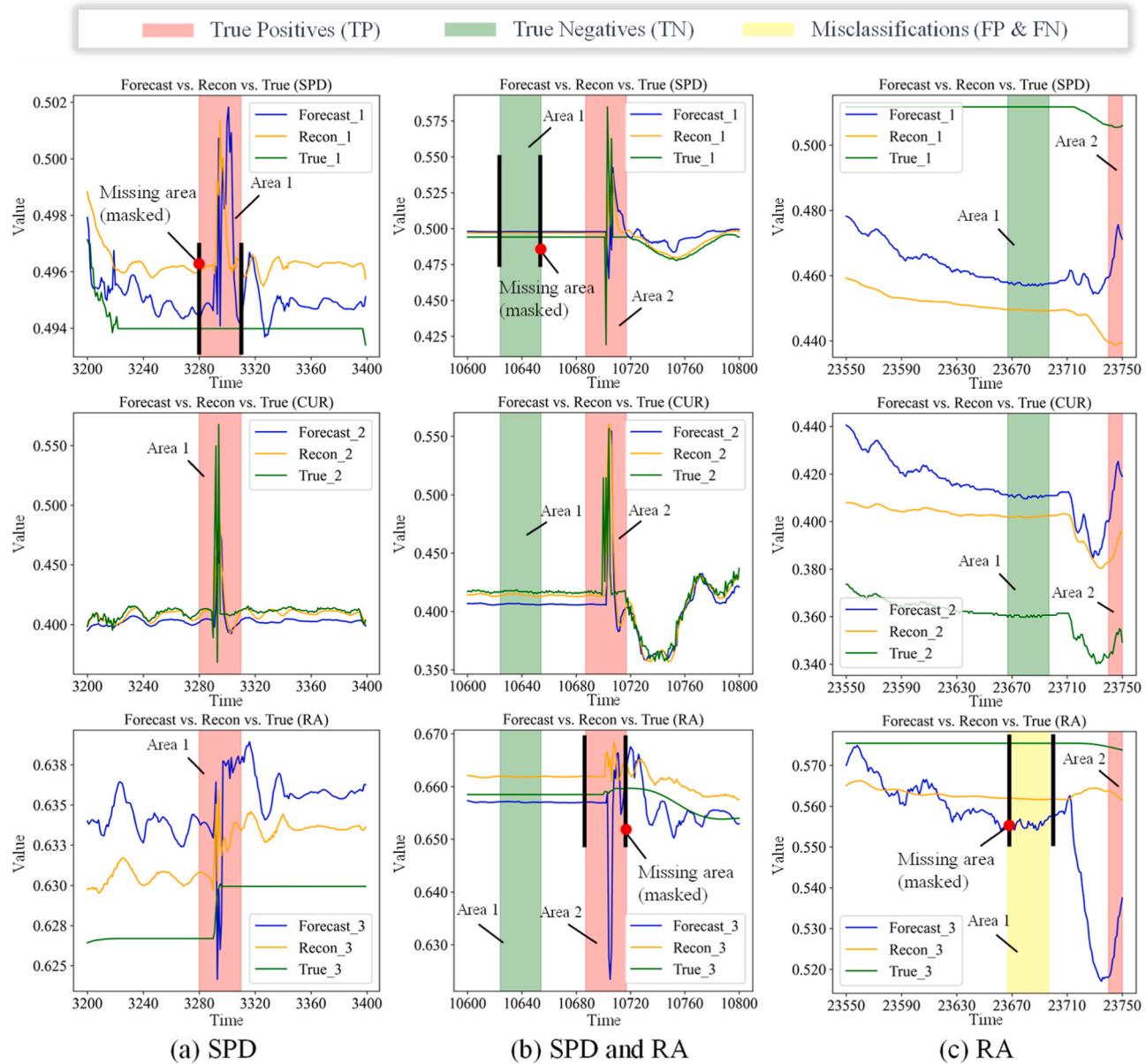


Fig. 8. Cross-region anomaly detection under fault and masking conditions of Joint 2.

even in the case of missing data and is able to accurately perform anomaly detection. In particular, while the anomaly patterns observed in the injected fault data provide valuable insights, they may differ from those in real industrial fault scenarios. Therefore, although the model performs well on the current dataset, its application to real industrial environments requires further validation and adaptation.

4.4.3. Ablation experiments

In the ablation experiments section of this study, we conducted ablation experiments on the AD-CIFC model in single-joint2 and double-joint2,4 and delved into the specific effects of different impute strategies as well as key components of the model on the model performance, as shown in Tables 3 and 4. Particular attention is paid to the fact that if, during the ablation experiments, we encountered the absence of some of these modules that prevented us from making specialized hyperparameter adjustments, we would use the hyperparameter values already determined in the full model setup as a substitute. This approach ensures the consistency of the experimental conditions, even in scenarios where some features are missing.

The ablation experiments with different imputing strategies apply different combinations of Adjacent Data imputation (AD), Cosine Similarity imputation (CS), and GraphSAGE imputing (GS) strategies, respectively, and introduce the traditional Linear Interpolation (LI) method for comparative analysis, and we observe the differences in the performance of the different strategies under the missing data condition, as shown in Table 3. In the ablation experiments using different combinations of data impute strategies, we found that for single and double joints the performance in the missing condition was different. Specifically, in the single-joint missing condition, each impute strategy combination shows a higher F1 score compared to the double-joint missing condition, with AD (Adjacent Data imputation) + GS (GraphSAGE imputation) showing a relatively outstanding performance. This suggests that information can be effectively recovered by this strategy when the missing data is relatively centralized. However, in the more complex bi-articulated missing cases, we observe that the CS (Cosine Similarity imputation) and GS (GraphSAGE imputation) approach demonstrates greater adaptability, which may be because the combination is better able to capture and utilize the complex relationships between the data,

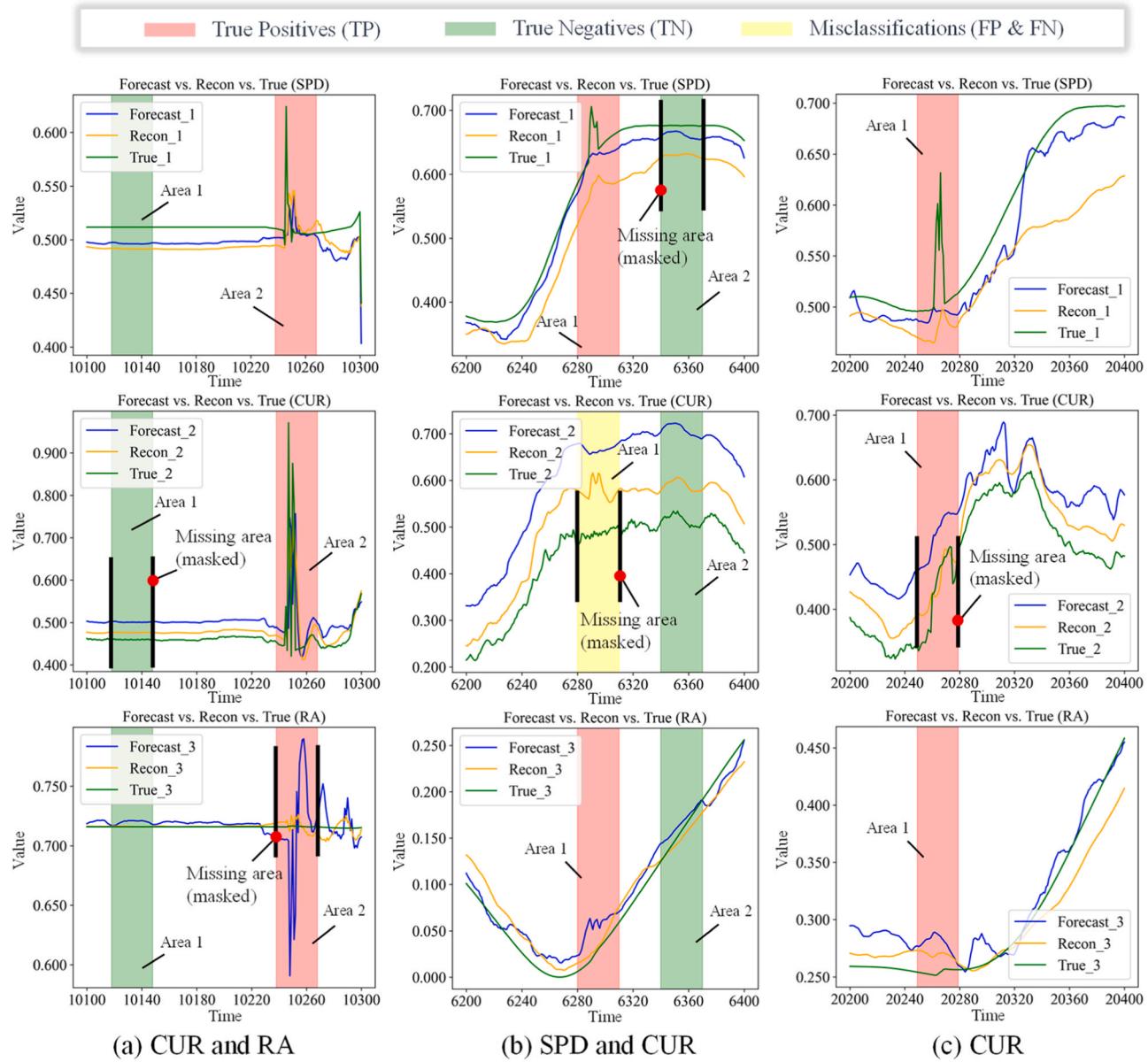


Fig. 9. Cross-region anomaly detection under fault and masking conditions of Joint 4.

Table 3

Performance of AD-CIFC with various imputation strategies (F1 Score).

Different combinations	Joint 4 Fault with Missing-2	Joints 2 and 4 Fault with Missing-2
AD (Adjacent Data)	0.733	0.680
CS (Cosine Similarity)	0.753	0.728
GS (GraphSAGE)	0.782	0.757
AD + CS	0.763	0.709
AD + GS	0.821	0.783
CS + GS	0.802	0.786
LI (Linear Interpolation)	0.665	0.631
AD + CS + GS (Ours)	0.831	0.796

thus providing more accurate imputations in the case of multi-articulated missing data. It is worth noting that when the three imputation methods, AD, CS, and GS, are employed in combination, they demonstrate superior model performance relative to the other combinations in both the single-articulated and double-articulated missing

Table 4

Efficacy analysis of our method and variants in ablation study (F1 Score).

Method	Joint 4 Fault	Joint 4 Fault with Missing-2	Joints 2 and 4 Fault	Joints 2 and 4 Fault with Missing-2
AD-CIFC	0.907	0.831	0.902	0.796
w/o Data Augmentation	0.897	0.703	0.891	0.642
w/o Collaborative Graph	0.819	0.725	0.809	0.666
w/o Temporal Graph	0.865	0.757	0.857	0.708
w/o Collaborative Forecasting	0.878	0.724	0.859	0.671
w/o Collaborative Reconstruction	0.850	0.712	0.825	0.663

cases, which suggests that the comprehensive imputation strategy can effectively leverage the advantages of each to cope with more complex missing data scenarios. Whereas for the LI (linear interpolation)

approach the performance is the lowest among all the tested scenarios, the simultaneous use of AD, CS, and GS combined imputation strategy demonstrates 24.96 % and 26.15 % higher model performance relative to LI in both single and double joint missing scenarios, which further emphasizes the importance of adopting a more advanced and comprehensive imputation strategy for data imputing.

The ablation experiments for the key components of the model focus on four main parts: Data Augmentation, Collaborative Graph, Temporal Graph, Collaborative Forecasting, and Collaborative Reconstruction, as Table 4 shows. For complete data, the removal of these components, while affecting the overall performance of the model, has a relatively limited impact. However, the same ablation experiments show more significant performance degradation with a missing length of 30 data points and a missing rate of 15 %. For example, Data Augmentation's removal in the missing data condition made the model less adaptable to various possible data variants, and in particular, the model's performance showed a significant degradation when unexpected missing data were encountered. Meanwhile, the removal of the Collaborative Graph in the missing data condition results in the inability to exploit the strong correlation synergy between features, which directly makes the model less sensitive to anomaly identification, reflecting its central role in maintaining data integrity and Data Augmentation performance. Similarly, the removal of the Temporal Graph in a missing data environment exacerbates the model's performance loss in terms of temporal dependency capture. The ablation of the Collaborative Forecasting and Collaborative Reconstruction components in the presence of missing data also reveals their importance in integrating inter-feature interactions within a group information, which is crucial for judgment and prediction, and their absence leads to significant model degradation in terms of accuracy and data recovery capability. Meanwhile, the performance of the dual-jointed model with missing data decreases more significantly compared to the single-jointed model, also indicating that the absence of key model components has a more significant negative impact on performance when faced with more complex missing data scenarios. These comparative results not only reveal the unique contributions of each component to handle different data integrity conditions but also provide valuable guidance for future model optimization, especially in improving the robustness of the model in the face of missing data.

4.4.4. Missing data test

In the missing data testing part of this study, we evaluated the performance of three different missing combinations in single-jointed 2 and double-jointed 2, 4. The main purpose of this is to investigate the performance of the model under different missing combinations, as well as

the performance of the other two different missing combinations compared to the performance with a length of 30 data points and a missing rate of 15 %, as shown in Fig. 10. For single-jointed 2 and double-jointed 2, 4, there is a similarity in the performance of the model under different missing combinations. At a length of 20 data points and a missing rate of 10 %, the model shows excellent performance under single and double joints and maintains high precision, recall, and F1 values, a result that demonstrates the robustness of the model to minor missing data. However, as the missing length and missing rate increased to 40 % and 20 %, respectively, we noticed a significant drop in the overall performance of the model, and this was especially noticeable when dealing with double-jointed data. While the model still maintains the ability to recognize anomalous patterns, its key performance metrics decline in this higher missingness condition compared to the low missingness case. This observation reveals the limitations of the model when dealing with higher levels of missing data and provides direction for future optimization and improvement. Notably, even under mild to severe levels of missing data, our model still demonstrates robustness beyond traditional methods, highlighting its potential for multivariate time series anomaly detection and providing key insights into performance enhancement for dealing with more extreme conditions of missing data.

4.5. Overhaul data validation

In the overhaul data validation part of this study, we implement the data impute (Masked) process and anomaly detection at the location of the fault point in the overhaul data of Joint 2 and Joint 4, respectively, where the data with a Mask length of is missing, as shown in Fig. 11. These data are mainly collected during routine maintenance rather than directly originated from typical faults. We found that the model fails to accurately identify all potential anomalies in a given situation due to the insignificance of the fault signals and the mixing with normal operation data. Specifically, when analyzing Area1 in Joints 2 and 4, the fault signals in the data are relatively significant, and the model is able to correctly identify some anomaly patterns under the condition of missing data, which suggests that the model has good sensitivity to the more significant fault signals and shows its effectiveness in handling data with clear anomaly characteristics. However, when dealing with Area2 in the CUR feature of Joint 4, the anomalous information of the other two features is relatively subtle, which leads to the fact that it does not exceed the set threshold after being calculated by the synergistic loss function, and incorrectly judges Area2 in the CUR feature as a normal area. This observation not only reveals the challenges of the model in dealing with atypical fault data with obscure features but also points to

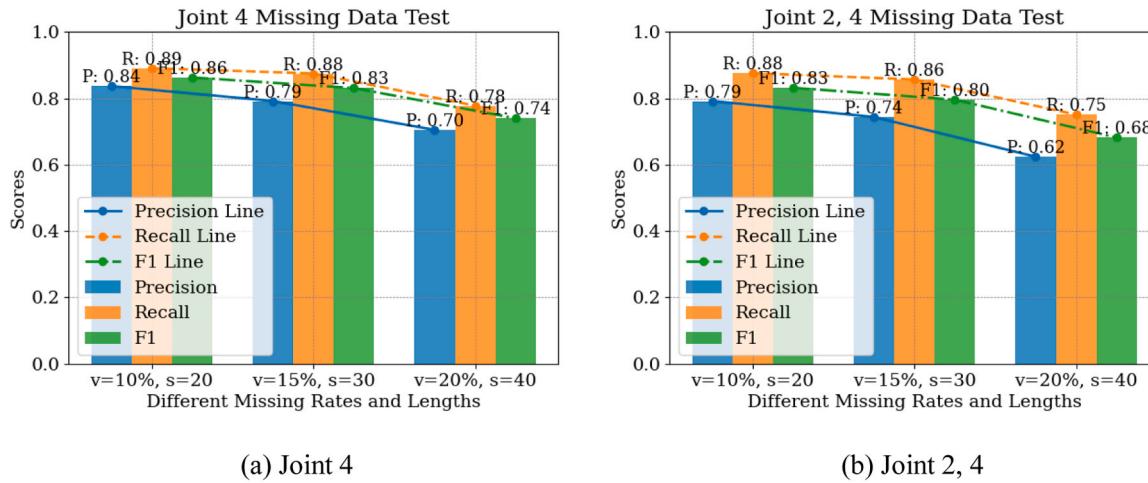


Fig. 10. Performance analysis of joint missing data tests.

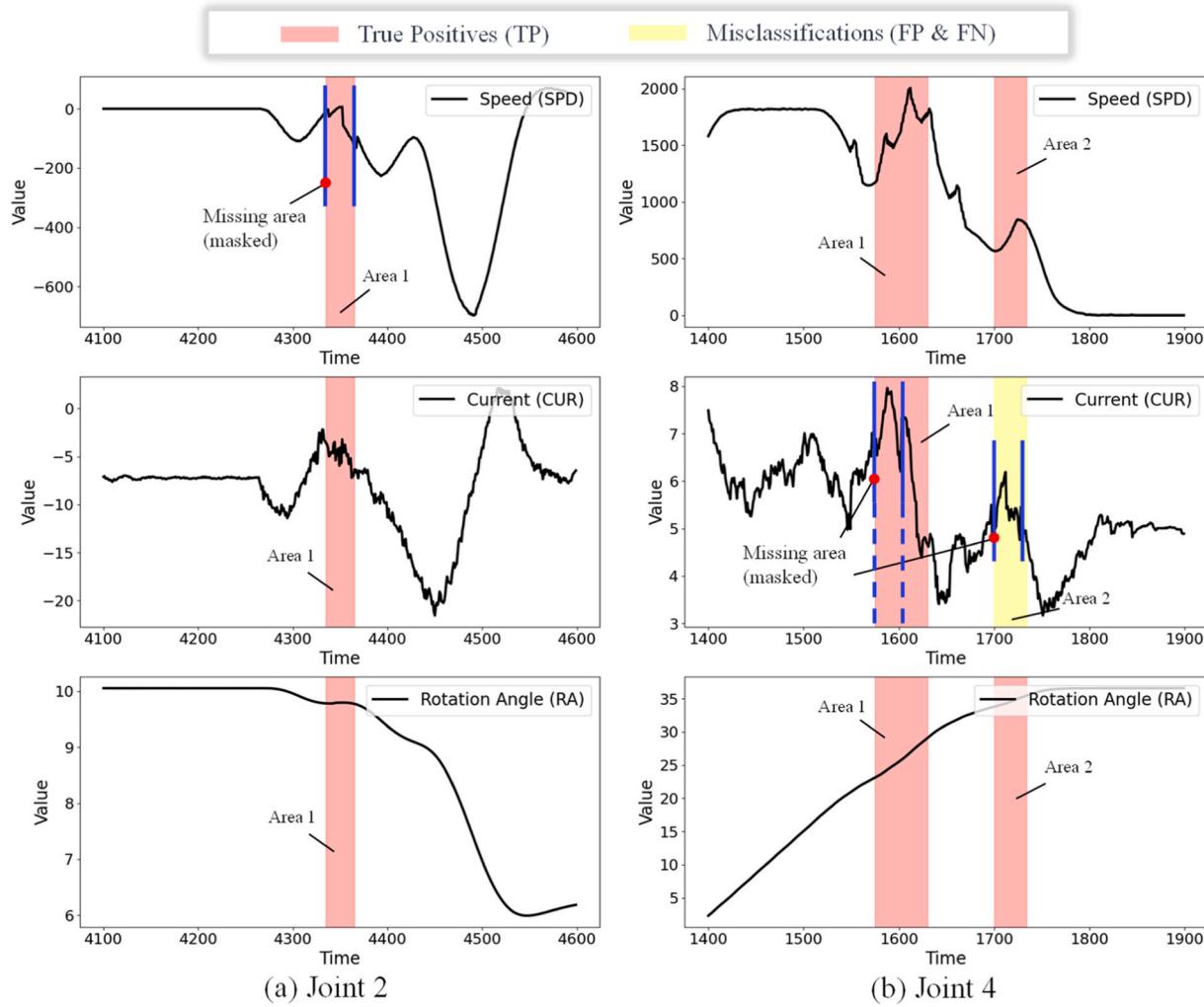


Fig. 11. The performance of the model in the maintenance state of two joints.

the direction of future model improvement, especially in terms of Augmentation of the model's sensitivity and adaptability to situations where the fault features are obscure, as well as potential avenues for optimization for specific joint characteristics.

5. Conclusion

This study focuses on the challenge of anomaly detection in industrial robot multivariate time series data when faced with various data-missing scenarios. To this end, we constructed AD-CIFC for identifying anomalous behaviors in variable environments with missing data. Firstly, an innovative data imputation method is proposed, i.e., a comprehensive imputation strategy of Adjacent Data imputation, Cosine Similarity imputation, and GraphSAGE imputation is used to comprehensively consider the missing data regions to better adapt to the natural flow and complex patterns of the time series, and the time-dependent and nonlinear characteristics of the data are especially considered; CGA and TGA are two graph noticing mechanisms, and CGA is used to recognize the missing data regions through the combination of synergistic features to achieve a comprehensive update of the graph structure information, and TGA focuses on dealing with the node relationships in the graph over time, through the joint use of the two in order to enhance the performance and adaptability of the model. Further, combining the outputs of CGA and TGA fusing them with the data-processed robot data, and inputting them to GRU deepens the understanding of long-term time series trends. Finally, by CFBM and CRBM and using them jointly, the

strong correlation of features between the data is considered more comprehensively, and a new loss function calculation method for the case of missing data is proposed by utilizing the feature synergy, which makes the model have a stronger adaptive ability in the face of different types of anomalies. The effectiveness of the proposed method and the superiority of AD-CIFC are verified by the anomaly detection experiments on industrial robots with different levels of missing joint data. The experimental results show that the AD-CIFC model demonstrates a comparable or even better performance than the other mainstream anomaly detection models in the complete fault data test, and the performance of the other models decreases significantly under the condition of partially missing data, whereas our model still maintains a high level of accuracy; in the case of slight to severe missing data, the AD-CIFC model still demonstrates robustness beyond the traditional methods; at the same time, the performance of the model using the comprehensive imputation strategy is about 25 % higher than that of the traditional linear interpolation method; and finally, in the validation of overhauling data, it is concluded that the model is able to recognize the majority of the anomaly patterns correctly under the condition of missing data. Finally, the model can correctly identify most of the anomaly patterns under the condition of missing data under the validation of maintenance data. Therefore, the model has the potential to be applied in corporate production lines, and its application is expected to guarantee the continuity of the production process and the consistency of product quality, especially when faced with the challenge of incomplete data.

Future work will focus on the following four areas: (1) Challenges when dealing with atypical fault data with obscure characteristics, especially in terms of Data Augmentation sensitivity and adaptability to situations where fault characteristics are obscure, and potential ways to optimize for specific joint characteristics; (2) Performance limitations when dealing with a higher percentage of missing data in order to enhance its application to extreme situations; and (3) Obtaining real robot failure data for further testing and optimization of the model, and exploring the model's performance in the face of more complex and less obvious anomaly patterns; (4) It will focus on the application of the model in actual industrial environments, especially the robot anomaly monitoring software developed in cooperation with a car company in Chongqing, which is expected to greatly improve the fault prediction capabilities of the production line.

CRediT authorship contribution statement

Jian Jiao: Formal analysis, Resources, Validation. **Zerui Xi:** Project administration, Resources. **Yufeng Li:** Data curation, Visualization. **Weishan Long:** Investigation, Methodology, Writing – original draft, Writing – review & editing. **Bo Yang:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **Yucheng Zhang:** Funding acquisition, Resources, Supervision, Validation.

Declaration of Competing Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service or company that could be construed as influencing the position presented in the manuscript entitled.

Acknowledgments

The presented work was supported by the National Key Research and Development Project of China (no. 2023YFB3306800), the Regional Innovation Cooperation Project of Sichuan Province (2023YFQ0019), the Natural Science Foundation of Chongqing (no. cstc2021jcyj-msxmX0732) and the Key Project of Technological Innovation and Application Development Plan of Chongqing (no. cstc2019jscx-mbdX0056).

References

- [1] Zhang R, Lv JH, Li J, Bao JS, Zheng P, Peng T. A graph-based reinforcement learning-enabled approach for adaptive human-robot collaborative assembly operations. *J Manuf Syst* 2022;63:491–503.
- [2] Li SF, Zheng P, Fan JM, Wang LH. Toward proactive human-robot collaborative assembly: a multimodal transfer-learning-enabled action prediction approach. *IEEE Trans Ind Electron* 2022;69:8579–88.
- [3] Yang B, Wang SL, Li S, Bi FY. Digital thread-driven proactive and reactive service composition for cloud manufacturing. *IEEE Trans Ind Inf* 2023;19:2952–62.
- [4] Rahman M, Liu HC, Masri M, Durazo-Cardenas I, Starr A. A railway track reconstruction method using robotic vision on a mobile manipulator: a proposed strategy. *Comput Ind* 2023;148:103900.
- [5] Chen C, Liu C, Wang T, Zhang A, Wu WH, Cheng LL. Compound fault diagnosis for industrial robots based on dual-transformer networks. *J Manuf Syst* 2023;66: 163–78.
- [6] Yang B, Huang HP, Bi FY, Yin LQ, Yang Q, Shen H. Analysis of cure kinetics of CFRP composites molding process using incremental thermochemical information aggregation networks. *Compos Struct* 2024;331:117904.
- [7] Zhou X, Zeng HL, Chen C, Xiao H, Xiang ZL. An attention-enhanced multi-modal deep learning algorithm for robotic compound fault diagnosis. *Meas Sci Technol* 2023;34:014007.
- [8] Wang JK, Liang WX, Yang JA, Wang SZ, Yang ZX. An adaptive image enhancement approach for safety monitoring robot under insufficient illumination condition. *Comput Ind* 2023;147:103862.
- [9] He YM, Chen JH, Zhou X, Huang SF. In-situ fault diagnosis for the harmonic reducer of industrial robots via multi-scale mixed convolutional neural networks. *J Manuf Syst* 2023;66:233–47.
- [10] Yang B, Xu WL, Bi FY, Zhang Y, Kang L, Yi LL. Multi-scale neighborhood query graph convolutional network for multi-defect location in CFRP laminates. *Comput Ind* 2023;153:104015.
- [11] Khalid A, Kirisci P, Khan ZH, Ghairai Z, Thoben KD, Pannek J. Security framework for industrial collaborative robotic cyber-physical systems. *Comput Ind* 2018;97: 132–45.
- [12] Vangipuram R, Gunupudi RK, Puligadda VK, Vinjamuri J. A machine learning approach for imputation and anomaly detection in IoT environment. *Expert Syst* 2020;37:e12556.
- [13] Li X, Xu YX, Li NP, Yang B, Lei YG. Remaining useful life prediction with partial sensor malfunctions using deep adversarial networks. *IEEE-CAA J Autom Sin* 2023; 10:121–34.
- [14] Basurto N, Arroyo A, Cambra C, Herrero A. A hybrid machine learning system to impute and classify a component-based robot. *Log J IGPL* 2023;31:338–51.
- [15] Basurto N, Cambra C, Herrero A. Improving the detection of robot anomalies by handling data irregularities. *Neurocomputing* 2021;459:419–31.
- [16] Tong CF, Chen HL, Xuan Q, Yang XH. A framework for bus trajectory extraction and missing data recovery for data sampled from the internet. *Sensors* 2017;17: 342.
- [17] Chen H, Lin MW, Liu JQ, Yang HS, Zhang C, Xu ZS. NT-DPTC: a non-negative temporal dimension preserved tensor completion model for missing traffic data imputation. *Inf Sci* 2024;653:119797.
- [18] Kim D, Choi JY. Efficient imputation of missing data using the information of local space defined by the geometric one-class classifier. *Expert Syst Appl* 2024;242: 122775.
- [19] Memarian A, Varanasi SK, Huang B, Slot G. Smart optimization with PPCR modeling in the presence of missing data, time delay and model-plant mismatch. *Chemom Intell Lab Syst* 2023;237:104812.
- [20] Togbe MU, Chabchoub Y, Boly A, Barry M, Chiky R, Bahri M. Anomalies detection using isolation in concept-drifting data streams. *Computers* 2021;10:13.
- [21] Elbasiny RM, Sallam EA, Eltobely TE, Fahmy MM. A hybrid network intrusion detection framework based on random forests and weighted k-means. *Ain Shams Eng J* 2013;4:753–62.
- [22] Xie ZT, Jin L, Luo X, Sun ZB, Liu M. RNN for repetitive motion generation of redundant robot manipulators: an orthogonal projection-based scheme. *IEEE Trans Neural Netw Learn Syst* 2022;33:615–28.
- [23] Zhu H, Yi C, Rho S, Liu S, Jiang F. An interpretable multivariate time-series anomaly detection method in cyber-physical systems based on adaptive mask. *IEEE Internet Things* 2024;11:2728–40.
- [24] Hinton G.E., Salakhutdinov R.R. Reducing the dimensionality of data with neural networks. *Science* 2006;313:504–507.
- [25] Miao QC, Xu CF, Zhan J, Zhu D, Wu CK. An unsupervised short- and long-term mask representation for multivariate time series anomaly detection. *Neural Information Processing*. Singapore: Springer Nature Singapore; 2023.
- [26] Li D, Chen D.C., Goh J, Ng S.K. Anomaly detection with generative adversarial networks for multivariate time series. 2018.
- [27] Li D, Chen DC, Jin BH, Shi L, Goh J, Ng SK. MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks. in *Artificial Neural Networks and Machine Learning-ICANN 2019: Text and Time Series*. Cham: Springer International Publishing.; 2019.
- [28] Zhao H, Wang Y, Duan J, Huang C, Cao D, Tong Y, Xu B, Bai J, Tong J, Zhang Q. Multivariate time-series anomaly detection via graph attention network. 2020 IEEE Int Conf Data Min (ICDM) 2020.
- [29] Park D, Hoshi Y, Kemp CC. A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. *IEEE Robot Autom Lett* 2018; 3:1544–51.
- [30] Goh J, Adepu S, Tan M, Lee ZS. Anomaly detection in cyber physical systems using recurrent neural networks. 2017 IEEE 18th Int Symp High Assur Syst Eng (HASE) 2017.
- [31] Choi K, Yi J, Park C, Yoon S. Deep learning for anomaly detection in time-series data: review, analysis, and guidelines. *IEEE Access* 2021;9:120043–65.
- [32] Erhan L, Ndubuaku M, Di Mauro M, Song W, Chen M, Fortino G, Bagdasar O, Liotta A. Smart anomaly detection in sensor systems: a multi-perspective review. *Inf Fusion* 2021;67:64–79.
- [33] Talagala PD, Hyndman RJ, Smith-Miles K, Kandanaarachchi S, Muñoz MA. Anomaly detection in streaming nonstationary temporal data. *J Comput Graph Stat* 2020;29:13–27.
- [34] Wang YY, Masoud N, Khojandi A. Real-time sensor anomaly detection and recovery in connected automated vehicle sensors. *IEEE T Intell Transp* 2021;22:1411–21.
- [35] Munir M, Siddiqui SA, Chattha MA, Dengel A, Ahmed S. FuseAD: unsupervised anomaly detection in streaming sensors data by fusing statistical and deep learning models. *Sensors* 2019;19:2451.
- [36] Zhong ZD, Zhao Y, Yang AY, Zhang HB, Qiao DH, Zhang ZH. Industrial robot vibration anomaly detection based on sliding window one-dimensional convolution autoencoder. *Shock Vib* 2022;1179192.
- [37] Azzalini D, Bonali L, Amigoni F. A minimally supervised approach based on variational autoencoders for anomaly detection in autonomous robots. *IEEE Robot Auton Lett* 2021;6:2985–92.
- [38] Castellini A, Masillo F, Azzalini D, Amigoni F, Farinelli A. Adversarial data augmentation for Hmm-based anomaly detection. *IEEE T Pattern Anal* 2023;45: 14131–43.
- [39] He YM, Zhao C, Zhou X, Shen WM. MJAR: A novel joint generalization-based diagnosis method for industrial robots with compound faults. *Robot Comput-Integr Manuf* 2024;86:102668.

- [40] Cui T, Song R, Li FM, Fu TY, Wang CQ, Li YB. Fast recognition of snap-fit for industrial robot using a recurrent neural network. *IEEE Robot Autom Lett* 2023;8:1635–42.
- [41] Gao BX, Kong XY, Li SZ, Chen Y, Zhang XY, Liu ZY, Lv WJ. Enhancing anomaly detection accuracy and interpretability in low-quality and class imbalanced data: a comprehensive approach. *Appl Energy* 2024;353:122157.
- [42] Feng Y, Liu ZJ, Chen JL, Lv HX, Wang J, Zhang XW. Unsupervised multimodal anomaly detection with missing sources for liquid rocket engine. *IEEE Trans Neural Netw Learn Syst* 2023;34:9966–80.
- [43] Yang JY, Yue ZG, Yuan Y. Deep probabilistic graphical modeling for robust multivariate time series anomaly detection with missing data. *Reliab Eng Syst Saf* 2023;238:109410.
- [44] Zhang YC, Su XJ, Meng K, Dong ZY. Robust fault detection approach for wind farms considering missing data tolerance and recovery. *IET Renew Power Gener* 2020;14:4150–8.
- [45] Chen ZL, Ni XH, Li H, Kong XJ. FedGAN: a method for anomaly detection and repair of hydrological telemetry data based on federated learning. *PeerJ Comput Sci* 2023;9:e1664.
- [46] Ma Z, Yun CB, Wan HP, Shen YB, Yu F, Luo YZ. Probabilistic principal component analysis-based anomaly detection for structures with missing data. *Struct Control Health Monit* 2021;28:e2698.
- [47] Cai BP, Shao XY, Liu YH, Kong XD, Wang HF, Xu HQ, Ge WF. Remaining useful life estimation of structure systems under the influence of multiple causes: subsea pipelines as a case study. *IEEE T Ind Electron* 2020;67:5737–47.
- [48] Cai BP, Fan HY, Shao XY, Liu YH, Liu GJ, Liu ZK, Ji RJ. Remaining useful life prediction methodology based on Wiener process: Subsea Christmas tree system as a case study. *Comput Ind Eng* 2021;151:106983.
- [49] Basurto N, Arroyo A, Cambra C, Herrero A. Imputation of missing values affecting the software performance of component-based robots. *Comput Electr Eng* 2020;87:106766.