

Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises

Shen Yan¹, Haidong Shao^{1,2*}, Yiming Xiao¹, Bin Liu³, Jiafu Wan⁴

1 College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China

2 Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, Luleå 97187, Sweden

3 Department of Management Science, University of Strathclyde, Glasgow G1 1XQ, UK

4 Provincial Key Laboratory of Technique and Equipment for Macromolecular Advanced Manufacturing, South China University of Technology, Guangzhou 510641, China

Corresponding author: Haidong Shao (hdshao@hnu.edu.cn)

Abstract: Anomaly detection of machine tools plays a vital role in the machinery industry to sustain efficient operation and avoid catastrophic failures. Compared to traditional machine learning and signal processing methods, deep learning has greater adaptive capability and end-to-end convenience. However, challenges still exist in recent research in anomaly detection of machine tools based on deep learning despite the marvelous endeavors so far, such as the necessity of labeled data for model training and insufficient consideration of noise effects. During machine operation, labeled data is often difficult to obtain; the collected data contains varying degrees of noise disturbances. To address the above challenges, this paper develops a hybrid robust convolutional autoencoder (HRCAE) for unsupervised anomaly detection of machine tools under noises. A parallel convolutional distribution fitting (PCDF) module is constructed, which can effectively fuse multi-sensor information and enhance network robustness by training in parallel to better fit the data distribution with unsupervised learning. A fused directional distance (FDD) loss function is designed to comprehensively consider the distance and angle differences among the data, which can effectively suppress the influence of noises and further improve the model robustness. The proposed method is validated by real computer numerical control (CNC) machine tool data, obtaining better performance of unsupervised anomaly detection under different noises compared to other popular unsupervised improved autoencoder methods.

Keywords: Machine tools; Deep learning; Unsupervised anomaly detection; Hybrid robust convolutional autoencoder; noises

1. Introduction

Machine tools are important and fundamental constitutes for manufacturing industry, which are widely used in various processes such as turning, milling and drilling [1-2]. In the era of Industry 4.0, machine tools are developing towards intelligence, automation, high precision, multi-function and high reliability, etc [3-5]. However, machine tools may exhibit unpredictable anomalies which can cause production downtime, serious losses or even casualties if not detected in time. Therefore, real-time anomaly detection of machine tools is important for economical and reliable production.

Data-driven methods have been proven to be effective for anomaly detection of machine tools, which can be divided into three procedures: signal collection, feature extraction and state detection. The collected signals can be vibration signals [6], current signals [7], acoustic signals [8] and temperature signals [9]. In general, the effectiveness of feature extraction directly affects the accuracy of state detection. Signal processing methods are widely used in feature extraction, such as empirical mode decomposition (EMD) [10], wavelet packet decomposition (WPD) [11], variational mode decomposition (VMD) [12], and local mean decomposition (LMD) [13]. The extracted features need to be subjected to machine learning models for state detection, such as hidden Markov models (HMM) [14], support vector machines (SVM) [15], and random forests (SF) [16]. Traditional machine learning models can only adapt to small-scale data [17], and feature extraction based on signal processing

methods changes with different work tasks and relies heavily on manual feature selection and prior knowledge, limiting the generalizability of the methods [18].

Unlike the above methods, deep learning-based intelligent methods have been widely studied in recent years for anomaly detection and condition monitoring of machine tools due to their adaptive feature mining capability and end-to-end convenience. Liang *et al.* [19] introduced a novel data-driven diagnosis system for supporting online anomaly detection in CNC machining. Huang *et al.* [20] constructed a deep CNN-based multi-domain feature fusion method to effectively characterize tool wear. Sun *et al.* [21] proposed a sparse autoencoder-based deep transfer learning network that can detect the health status of different machine tools. Cheng *et al.* [22] combined CNN and LSTM for tool wear value monitoring under variable working conditions. Deebak *et al.* [23] presented a digital-twin assisted method for analyzing the operating state of machine tools. Ou *et al.* [24] proposed a new nuclear autoencoder feature learning method for intelligent detection of tool wear states. Shi *et al.* [25] introduced multiple sparse autoencoders fusing features from multiple domains to accomplish a progressive tool wear monitoring task. Wang *et al.* [26] suggested a model with attention and monotonicity loss for simultaneous detection of tool wear and prediction. Sun *et al.* [27] designed a LSTM network to identify multiple rear tool wear values based on historical data. Yang *et al.* [28] presented an anomaly detection model based on residuals, which can effectively detect tool wear.

The above deep learning models have shown decent results for dealing with different challenges in anomaly detection and condition monitoring of machine tools. However, they all rely on sufficient labeled data, which limits their application in industrial scenarios. In the laboratory, it is easier to obtain data with accurate labels, but this is often difficult in the engineering practice [29]. Most of the machine tools are in normal status during actual operation, and labeled data is often very scarce or even unavailable. Therefore, unsupervised anomaly detection solely based on normal operating data provides more flexibility and extends the capability of the abovementioned deep learning methods. Autoencoder is a classical data reduction and reconstruction method, which has been explored and improved by many scholars in recent years to excel its performance in unsupervised anomaly detection of different fields. In 2018, Zong *et al.* [30] proposed a deep autoencoding Gaussian mixture model (DAGMM) by fusing autoencoder and gaussian mixture model that uses the joint optimization form to solve the problem of decoupling model learning and achieves good unsupervised anomaly detection results on public benchmark datasets. In 2019, Gong *et al.* [31] developed a memory-augmented autoencoder (MAE) by adding a memory module to increase the reconstruction error for anomaly data by enhancing the training memory for normal data. In 2021, Thill *et al.* [32] presented a temporal convolutional network autoencoder based on dilated convolution for unsupervised anomaly detection in electrocardiogram of patients with cardiac arrhythmia. In 2021, Li *et al.* [33] constructed an unsupervised fake news detection method based on autoencoder for detecting anomalies on social networks. All of the above improved autoencoder models achieve good unsupervised anomaly detection results in their respective fields. However, it is astonishing to find out few studies exist on unsupervised anomaly detection of machine tools that consider the improvement of autoencoder. In addition, monitoring data of machine tool collected by sensors in industrial scenarios inevitably contains various noise disturbances [34]. Therefore, to fill the research gap, this paper aims to study the improved autoencoder for unsupervised anomaly detection of machine tools under noises.

In this paper, a new improved autoencoder method called HRCAE is developed for unsupervised anomaly detection of machine tools under noises, which includes the construction of a PCDF module and the design of a FDD loss function. The PCDF module is constructed to effectively fuse multi-sensor information and to train symmetric convolutional networks in parallel to fit the distribution of normal data with unsupervised learning. The FDD loss function is designed to fully consider the distance and angle differences between the reconstructed data and the original data, thus effectively suppressing the noise effect on the data. The effectiveness of the proposed method is validated by real

CNC machine tool data. The results indicate that the developed HRCAE can effectively learn the distribution characteristics of normal data under different noises, and can obtain better unsupervised anomaly detection performance compared with other unsupervised improved autoencoder methods.

The sections are as follows: Section 2 introduces the related theory of basic CAE. Section 3 describes the developed HRCAE in detail, including construction of the PCDF module, design of the FDD loss function and general steps of the HRCAE. Section 4 shows the experimental verification of CNC machine tool data. Section 5 presents the conclusions and future research directions.

2. Basic CAE

CAE, as a classical unsupervised learning method, has been widely used in image denoising, data reconstruction and other relevant fields [35-37]. It can combine the feature extraction function of CNN with the unsupervised feature reconstruction function of autoencoder [38]. In general, the CAE consists of multiple convolution encoders and deconvolution decoders, which can be sketched in Fig. 1.

A convolution encoder usually consists of a convolution layer and a pooling layer to extract data feature, which can be expressed as:

$$H = pool(\sigma(\sum(X \odot w^i + b^i))) \quad (1)$$

where X is the input data; H is the feature obtained after encoding; w^i is the i th convolution kernel in the convolution layer, and b^i is the i th bias; \odot indicates the convolution operation; σ is the activation function; $pool(\cdot)$ indicates the pooling operation.

A convolution decoder generally includes a deconvolution layer and an up-sampling layer, which can be expressed as:

$$\hat{X} = ups(\sigma(\sum(H \otimes \hat{w}^i + \hat{b}^i))) \quad (2)$$

where \hat{X} is the data returned after decoding; \hat{w}^i is the i th deconvolution kernel in the deconvolution layer, and \hat{b}^i is the i th bias; \otimes indicates the deconvolution operation; $ups(\cdot)$ indicates the up-sampling operation.

The reconstruction error, which is used to optimize the network's parameters, is denoted as the Euclidean distance between X and \hat{X} as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{X}_i - X_i)^2 \quad (3)$$

where n is the data number; \hat{X}_i is the i th reconstructed data; X_i is the i th input data; MSE is the reconstruction error, and the network aims to minimize it.

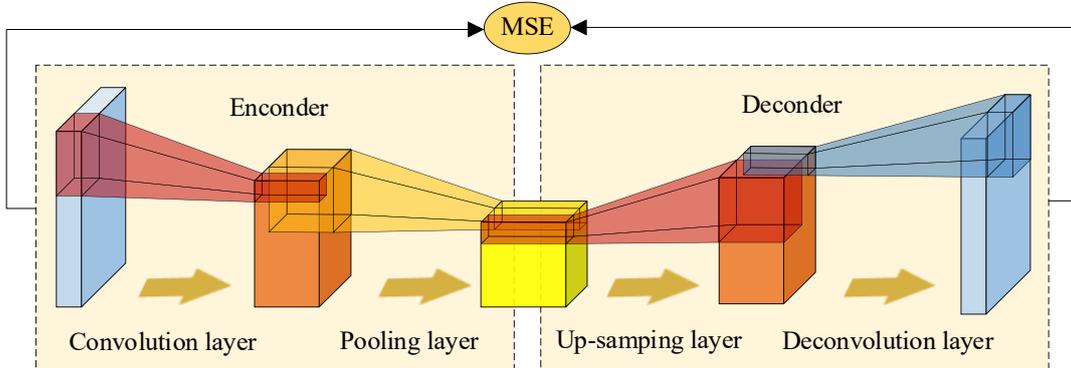


Fig.1. The structure of CAE.

3. The proposed method

We develop the HRCAE driven by multi-sensor information for unsupervised anomaly detection of machine tools under noises. The proposed method consists of two steps: constructing a PCDF module, and designing a FDD loss function.

3.1. Construction of the PCDF module

The model training in existing studies to anomaly detection of machine tools relies on labeled data, yet this is often very scarce or even unavailable in engineering practice. We construct the PCDF module that can cope with the above shortcomings, as described below.

The vibration data of machine tools with noises collected by multiple sensors is acquired and inputs to the convolutional network in multiple channels for the fusion to obtain richer health information. Inspired by the Inception network [39] with parallel convolutional blocks, we form a PCDF module by paralleling two convolutional networks to learn richer characteristics. The network width is enlarged by parallel training to improve network robustness, which can better fit the distribution of normal data for unsupervised anomaly detection. The structure of the PCDF module is shown in Fig. 2, where the convolution and deconvolution layers can be expressed by the following equations.

$$H = \sigma(\sum(X \odot w^i + b^i)) \quad (4)$$

$$\hat{X} = \sigma(\sum(H \otimes \hat{w}^i + \hat{b}^i)) \quad (5)$$

where Eq.(4) is the convolution layer; w^i , b^i are the i th one-dimensional convolution kernel and the corresponding bias here. Eq.(5) is the deconvolution layer; \hat{w}^i , \hat{b}^i are the i th one-dimensional deconvolution kernel and the corresponding bias here. Compared to Eqs.(1-2), we discard the pooling layer and the up-sampling layer because they are accompanied by a certain degree of data feature loss [40-41], which is not conducive to the distribution fitting of the normal data. The specific parameters of the PCDF module are shown in Table 1.

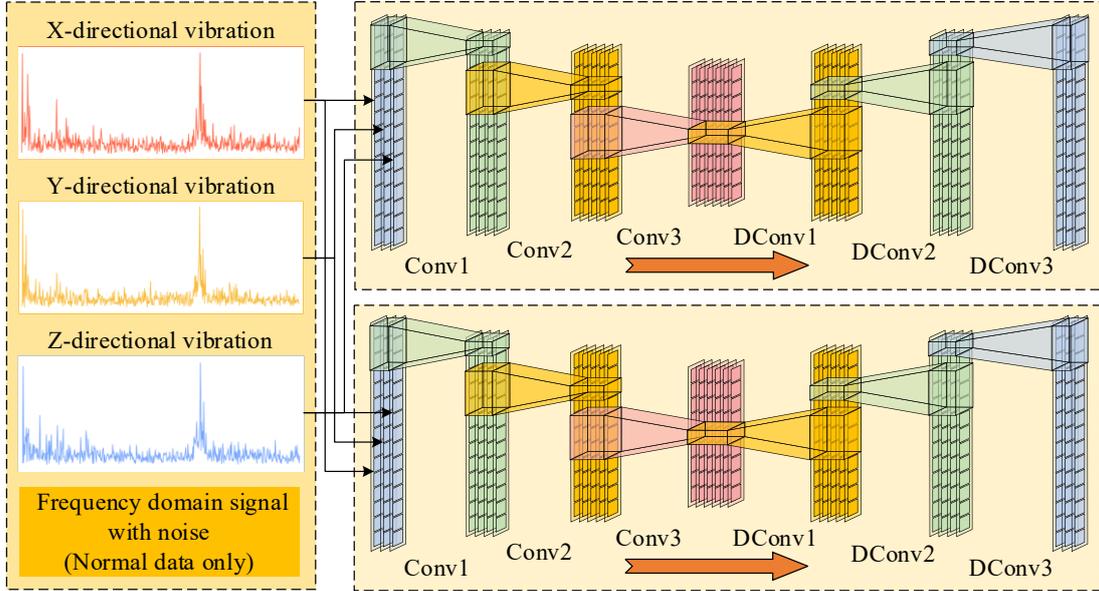


Fig.2. The structure of PCDF module.

Table 1

The structural parameters of PCDF module.

No.	Layer	Channel size	Kernel size/stride	Batch normalization	Activation function
1	Conv1	3×16	3/1	Yes	LeakyReLU
2	Conv2	16×32	3/1	Yes	LeakyReLU
3	Conv3	32×64	3/1	Yes	LeakyReLU
4	Dconv1	64×32	3/1	Yes	LeakyReLU
5	Dconv2	32×16	3/1	Yes	LeakyReLU
6	Dconv3	16×3	3/1	No	No

3.2. Design of the FDD loss function

The current prevalent CAE mainly chooses Eq. (3) as the reconstruction error. However, Euclidean distance is difficult to faithfully describe the similarity of complex feature spaces [42]. We design a new FDD loss function, which can measure the distribution similarity between data from the perspectives of angle and distance to effectively suppress the effect of noises, thus further improving the model robustness for anomaly detection of machine tools.

Several studies have demonstrated that effective combination of Euclidean distance and Cosine similarity can achieve good results in the fields of image recognition [43], language conversion [44-45], music classification [46], etc. The Euclidean distance is a spatial distance between data, focusing on the difference in values for each dimension. The Cosine similarity measures the geometric angle between data, and is insensitive to the numerical difference in each dimension, enabling better capture of feature difference in direction. A visual representation of the difference between Cosine similarity and Euclidean distance is shown in Fig. 3. The formula for calculating the Cosine similarity is expressed as follows.

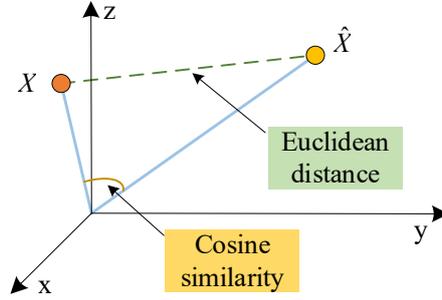


Fig.3. The difference between Euclidean distance and Cosine similarity.

$$CS = \frac{\sum_{i=1}^n (\hat{X}_i \times X_i)}{\sqrt{\sum_{i=1}^n (\hat{X}_i)^2} \times \sqrt{\sum_{i=1}^n (X_i)^2}} \quad (6)$$

where $CS \in [-1, 1]$ is the Cosine similarity between data, and the smaller value indicates the larger directional difference. CS can complement MSE well, and the distance and angle differences between data are fully considered. Thus, the designed FDD loss function can be obtained as follows:

$$FDD = MSE + \lambda \frac{(1 - CS)}{2} \quad (7)$$

where λ is an adjustable hyperparameter, which is used to balance the weight of distance and angle difference between data. The loss value is fed into Adam optimizer [47] to automatically optimize the network's learning parameters, which can be expressed as follows.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} J(\theta_{t-1}) \quad (8)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla_{\theta} J(\theta_{t-1}))^2 \quad (9)$$

$$\theta_t = \theta_{t-1} - \alpha \frac{m_t / (1 - \beta_1^t)}{(\sqrt{v_t / (1 - \beta_2^t)} + \varepsilon)} \quad (10)$$

where θ_t is the network's learning parameters at the t th iteration; α is the learning rate; $\nabla_{\theta} J(\theta_{t-1})$ is the computed gradient of the t th iteration; m_t is the exponential moving average of the calculated gradient for the t th iteration; v_t is the exponential moving average of the squared calculated gradient for the t th iteration; in general, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$.

3.3. General steps of the HRCAE

The general steps of the HRCAE are shown in **Fig. 4** and explained in detail as below.

Data Preparation: Acquire CNC machine tool vibration data collected by multiple sensors, adding Gaussian random noises, performing normalization operation and FFT transform. The data with noises is then divided into training set and validation set containing only normal data, and a test set containing both normal and anomaly data.

Model initialization: Set the model hyperparameters such as weight coefficient λ , learning rate α , iteration number T , and PCDF module structure. The learning parameters θ of the network are randomly initialized.

Model off-line training: The training set with only normal data is fed into the constructed PCDF module in multiple channels, and two convolutional distribution fitting networks are trained in parallel. The reconstructed data \hat{X} is obtained based on **Eqs. (4-5)**. The FDD loss values of the two networks are calculated and added based on **Eqs. (3), (6-7)**, and the learning parameters θ of the network are iteratively updated based on **Eqs. (8-10)**.

Model online validation: The validation set with only normal data is fed into the trained PCDF module, and after obtaining the reconstructed data, the sum of the FDD loss value for each sample in the two networks is calculated. Because there may be very few unfitted noise samples and considering the model's robustness, the loss value containing 95% of normal data is chosen as the threshold value for anomaly data.

Model online testing: The test set containing normal and anomaly data is fed into the trained PCDF module, and the sum of FDD loss value is calculated for each test sample to determine the normal or anomaly based on the threshold value. Finally, evaluation indexes such as accuracy, recall, precision and F-score are displayed.

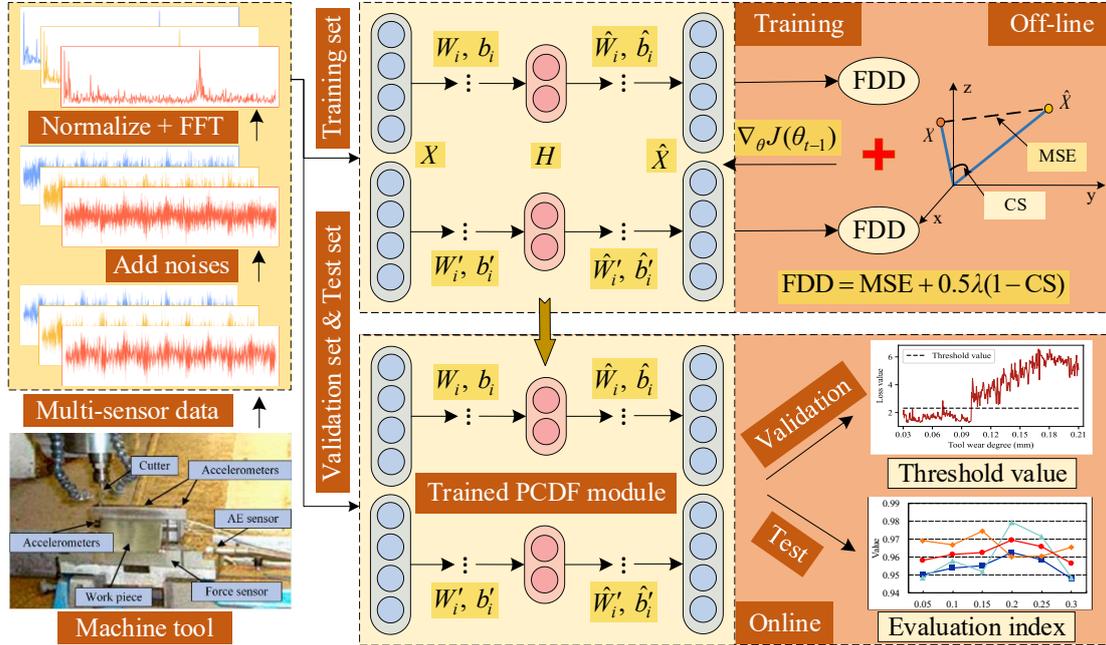


Fig.4. General steps of the HRCAE.

4. Experimental verifications

We test the effectiveness of the proposed method on a high-speed CNC machine. Section 4.1 describes the CNC machine tool datasets and division rules. Section 4.2 verifies the validity of the proposed method for unsupervised tool anomaly detection under different noises. Section 4.3 validates the benefit of fusing multi-sensor information and discusses the hyperparameter λ in the FDD loss function.

4.1. Description of datasets

Datasets are derived from a high-speed CNC milling machine operating under dry milling operation [48], and the data acquisition system is shown in Fig. 5. The specific operating parameters are set as follows: spindle speed is 10400 rpm; X-direction feed rate is 1555 mm/min; Y-direction cutting depth is 0.125 mm; Z-direction cutting depth is 0.2 mm. Tool vibration in the X, Y and Z directions is measured by piezoelectric accelerometers on the workpiece with sampling frequency of 50 kHz [49].

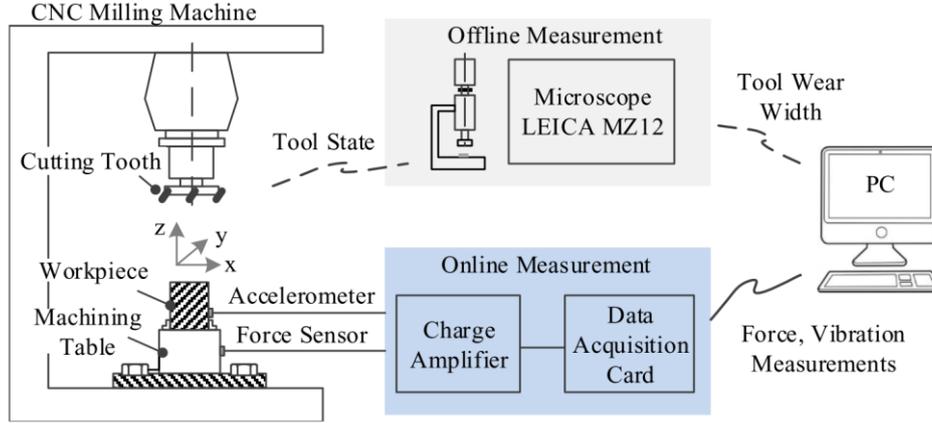


Fig.5. Data collection system [50].

A total of three datasets are used to record the wear degree of three tools, which are C1, C4 and C6, respectively. Each dataset has 315 sampling processes that record the variation of tool wear degree, with each sampling process recording tool wear degree in three directions. We average the wear degree in these three directions, and the wear degree above 0.10 mm is considered as anomaly data. Since there are few normal data in C6 dataset, we conduct unsupervised anomaly detection on the C1 and C4 datasets, and the dataset information is shown in Table 2. The C1 dataset contains data for tool wear degree from 0.04 mm to 0.16 mm. In order to simulate the full degradation process, for the training and validation sets, tool wear degree varies from 0.04 mm to 0.09 mm with normal data only, 100 samples randomly selected at 0.01 mm intervals, 600 samples in total; the test set of tool wear degree contains the full range of 0.04 mm to 0.16 mm, with 100 samples randomly selected at 0.01 mm intervals, for a total of 1300 samples. The length of each sample is 1024 sampling points; the training set, validation set and test set are from different sampling processes. The C4 dataset contains data with tool wear degree from 0.03 mm to 0.20 mm. The training set, validation set and test set are divided by the same rules as the C1 dataset in Table 2. It is important to emphasize that we only divide the dataset by tool wear degree and do not need to use specific wear degree labels in the training of the model, so the model training is unsupervised.

Table 2
Introduction to datasets.

Datasets	Training set (Only normal data)	Validation set (Only normal data)	Test set (Normal and anomaly data)
C1	600 (6×100) 0.04 mm-0.09 mm	600 (6×100) 0.04 mm-0.09 mm	1300 (13×100) 0.04 mm-0.16 mm
C4	700 (7×100) 0.03 mm-0.09 mm	700 (7×100) 0.03 mm-0.09 mm	1800 (18×100) 0.03 mm-0.20 mm

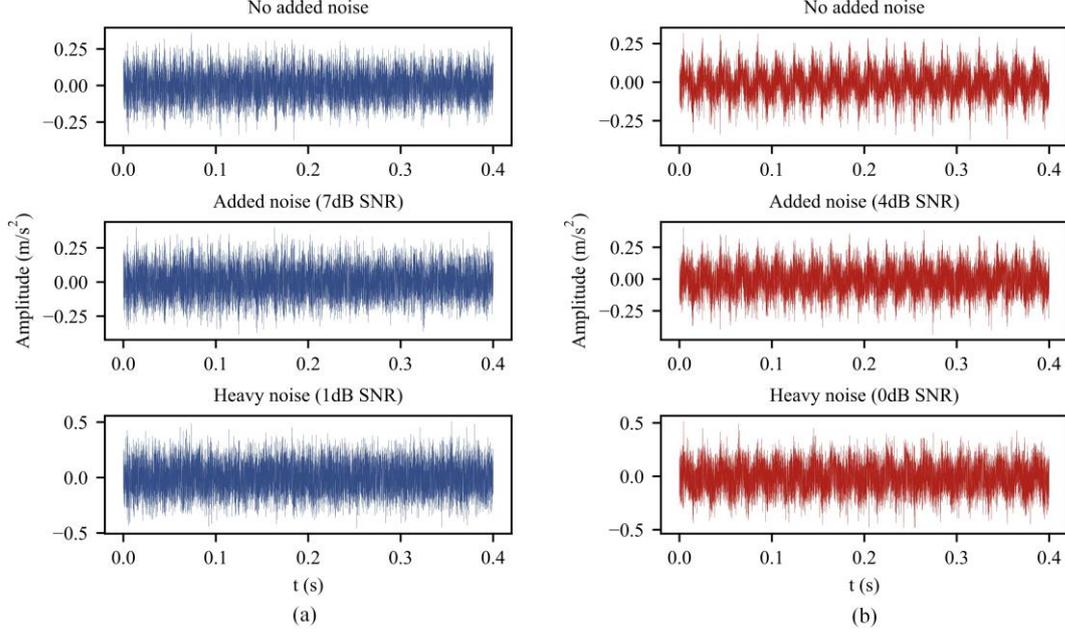


Fig. 6. Vibration signals of datasets under different noises: (a) C1 dataset; (b) C4 dataset.

Fig. 6 shows the vibration signals of the two datasets under different noises. As noises increase, the destructive effect on the vibration signals becomes more apparent. Under the heavy noise, the amplitudes of the vibration signals in both datasets are almost doubled compared to the signals without the addition of noise, which puts high demands on the robustness of the model. The value of signal-to-noise (SNR) is calculated as below:

$$\text{SNR} = 10 \log_{10} \left(\frac{P_s}{P_n} \right) \quad (11)$$

where P_s is the data power; P_n is the noise power; SNR is the signal-to-noise ratio in dB, and its smaller value means more heavy noise.

4.2. The validity of the HRCAE

In this paper, to reduce the random effect for experiments, all the methods are repeated five times. The algorithms are run in Pytorch 1.7.1 of Windows 10, with GPU GTX1650 4G. Based on the greedy algorithm idea [51], we search for a relatively better combination of hyperparameters by controlling the variables. The main hyperparameters are set as follows: the number of iterations is set to 300; the learning rate is set to 0.01, which decays 0.1 at 20, 200 iterations respectively; the hyperparameter λ in FDD loss function is set to 0.2. In the future, we will explore simpler and more general hyperparameter setting methods based on existing study [52].

First, we examine the unsupervised anomaly detection capability of the developed HRCAE under heavy noise and compare it with the commonly used unsupervised methods based on autoencoder, including CAE, MAE [31], DAGMM [30], sparse autoencoder (SAE) [53] and denoising autoencoder (DAE) [54], in which the superiority of the PCDF module and the FDD loss function are also verified. The evaluation indexes of anomaly detection on the C1 and C4 datasets for each method are calculated as follows:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (12)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (13)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (14)$$

$$\text{F-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (15)$$

where TP indicates correctly detected as anomaly data; FP indicates error detection as anomaly data; TN means correctly detected as normal data; FN means error detection as normal data. Accuracy can reflect the quality of global classification; F-score combines two indexes of precision and recall, which is an important index of how well anomaly data is detected.

Table 3 shows the evaluation indexes of each method for anomaly detection on the C1 dataset. The accuracy and F-score of the proposed method reach 95.95% and 96.21% under heavy noise with SNR=1 dB ($P_n / P_s \approx 80\%$), which are the optimal and the most stable among all methods. Although the recall of DAGMM reaches 98.43% and the precision of DAE-FDD reaches 97.81%, the other three evaluation indexes of them are unsatisfactory. The PCDF-MSE shows an improvement of 5.95%, 3.24%, 7.89% and 5.86% in each evaluation index compared to the CAE-MSE; the HRCAE shows an improvement of 1.27%, 0.80%, 1.60% and 1.32% in each evaluation index compared to the CAE-FDD. These improvements indicate that the constructed PCDF module has stronger robustness than the single convolutional distribution fitting network. After replacing MSE with FDD loss function, four evaluation indexes of MAE are improved by 8.21%, 5.28%, 10.34% and 8.27% respectively; SAE's evaluation indexes are improved by 6.90%, 5.79%, 7.63% and 7.22% respectively; DAE's evaluation indexes are improved by 6.08%, 2.79%, 9.58% and 7.80% respectively; CAE's evaluation indexes are improved by 5.78%, 3.14%, 7.69% and 5.60% respectively; PCDF module's evaluation indexes are improved by 1.10%, 0.70%, 1.40% and 1.06% respectively, and the stability of all methods is improved. The stability of anomaly detection results for above methods is also improved. It shows that the designed FDD loss function has universal applicability to various unsupervised improved autoencoder methods and can effectively measure the reconstruction errors of data from distance and angle directions to suppress the impact of heavy noise. The running times of all methods are shown in **Table 3**. The off-line training time of DAGMM is the longest, and the online validation test times of all methods are within 3s except DAGMM.

Table 3

Evaluation indexes and times of each method on the C1 dataset with SNR=1 dB.

Unsupervised methods	Evaluation indexes (%)				Times (s)	
	Accuracy	Precision	Recall	F-score	Off-line	Online
HRCAE	95.95±0.75	96.85±0.25	95.60±1.50	96.21±0.74	36	2.5
PCDF-MSE	94.85±1.61	96.15±0.98	94.20±2.75	95.15±1.58	36	2.5
CAE-FDD	94.68±1.65	96.05±0.75	94.00±3.59	94.89±1.64	19	1.9
CAE-MSE	88.90±4.46	92.91±5.75	86.31±6.56	89.29±4.29	19	2
MAE-FDD	94.00±1.75	96.02±0.84	92.71±3.63	94.30±1.78	78	2.4
MAE-MSE	85.79±3.96	90.74±4.67	82.37±9.15	86.03±4.37	78	2.5
DAGMM	84.51±1.56	78.35±1.17	98.43±2.51	87.24±1.38	355	3.1
SAE-FDD	90.62±6.23	96.76±1.16	85.37±11.35	90.41±6.84	20	2.0
SAE-MSE	83.72±10.05	90.97±7.97	77.74±15.67	83.19±11.13	20	2.0
DAE-FDD	88.11±8.62	97.81±0.61	79.69±16.25	87.03±10.98	19	1.9
DAE-MSE	82.03±11.67	95.02±2.20	70.11±22.15	79.23±14.96	19	2.0

Fig. 7 shows the visual test results of methods on the C1 dataset with SNR=1 dB. The horizontal coordinates indicate the tool wear degree of the test set, which ranges from 0.04 mm to 0.16 mm, and the vertical coordinates are the loss function value of each sample. Tool wear degrees less than 0.10 mm are normal samples; those above 0.10 mm are anomaly samples, and the dotted line is the threshold value for detecting anomaly samples. To express the aesthetics, the average loss value of five adjacent samples is taken as a point for plotting the curve. Because of the different model structures and loss functions of each method, as well as the different optimal spaces of each stochastic training, there are differences in the loss values of each method. It can be observed that the visual

test results of each method can correspond to the evaluation results in **Table 3**. The proposed method HRCAE and the PCDF-MSE differentiations between normal and anomaly samples are relatively more obvious compared to other methods. This indicates that the constructed PCDF module can better fit the distribution of normal samples to distinguish the anomaly samples. After replacing MSE with FDD loss function, the discriminative degree of each method for normal and anomaly samples is improved to some extent, which again demonstrates the robustness of FDD loss function under heavy noise.

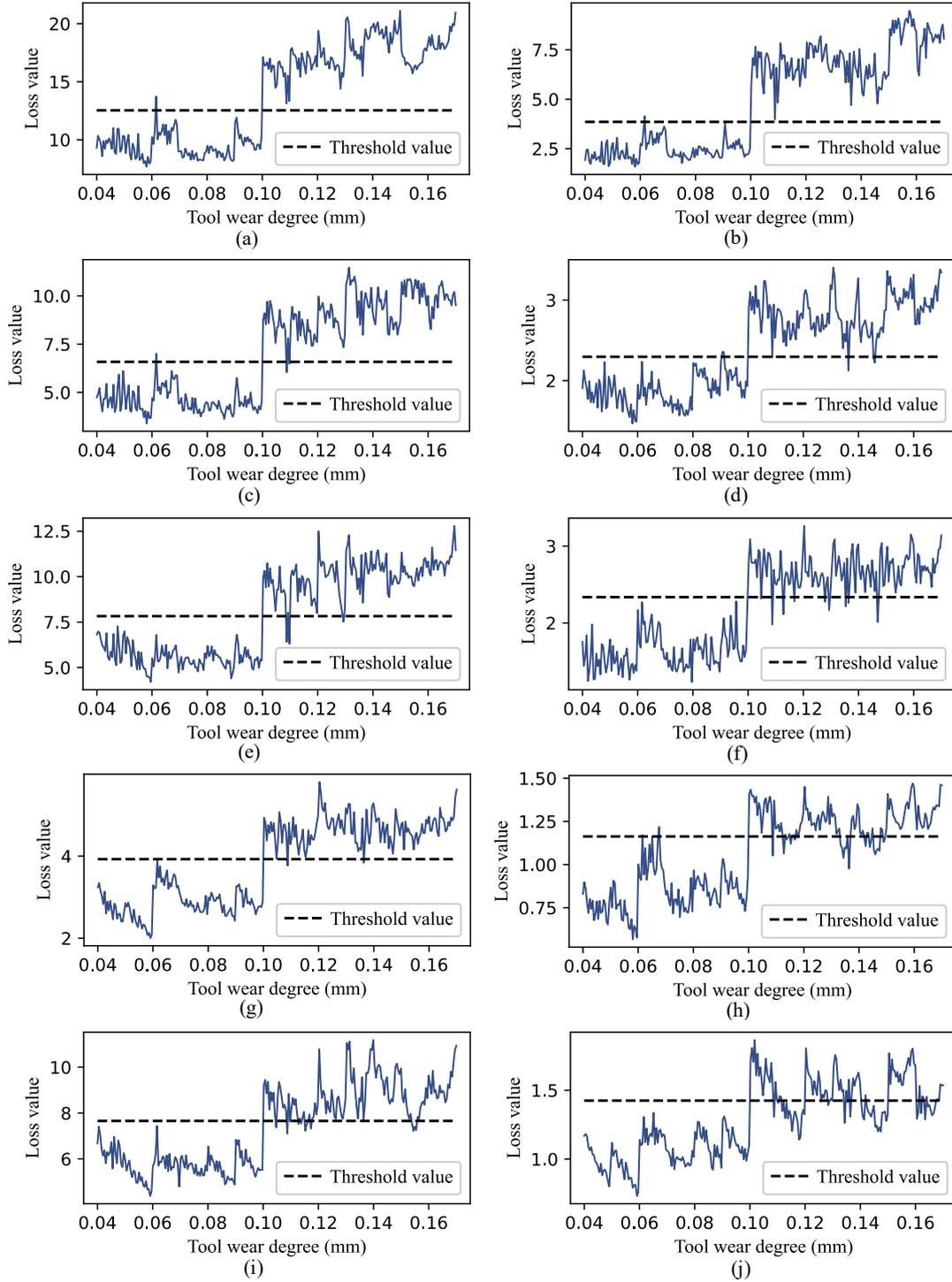
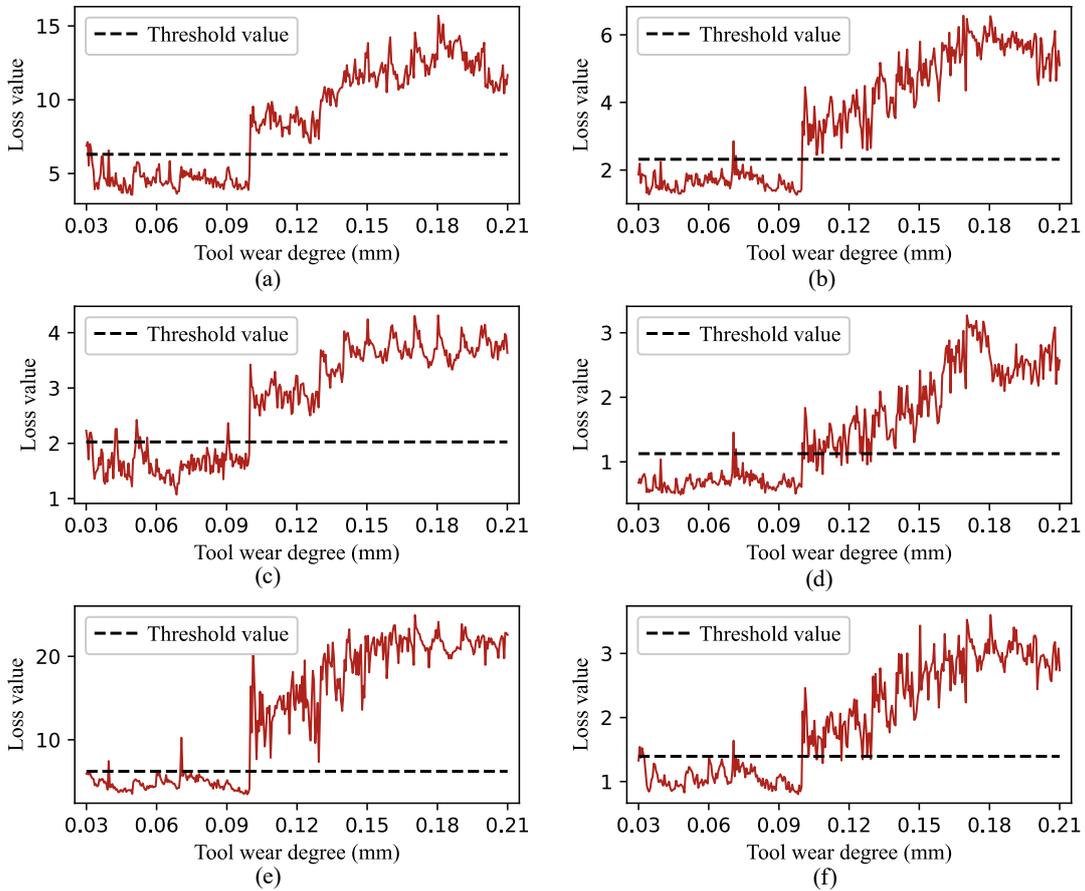


Fig.7. The visual test results of each method on the C1 dataset with SNR=1 dB: (a) HRCAE; (b) PCDF-MSE; (c) CAE-FDD; (d) CAE-MSE; (e) MAE-FDD; (f) MAE-MSE; (g) SAE-FDD; (h) SAE-MSE; (i) DAE-FDD; (j) DAE-MSE.

Table 4

Evaluation indexes and times of each method on the C4 dataset with SNR=0 dB.

Unsupervised methods	Evaluation indexes (%)				Time (s)	
	Accuracy	Precision	Recall	F-score	Off-line	Online
HRCAE	96.25±1.17	96.01±1.09	97.95±2.32	96.95±0.99	43	2.9
PCDF-MSE	93.91±2.42	96.72±0.48	93.20±4.15	94.89±2.18	42	2.9
CAE-FDD	93.00±2.79	95.71±2.52	92.89±6.85	94.10±2.58	23	2.3
CAE-MSE	88.31±5.48	96.11±2.95	84.26±7.84	89.67±5.03	23	2.3
MAE-FDD	94.43±1.83	96.34±0.89	94.49±3.31	95.38±1.59	94	2.9
MAE-MSE	91.14±4.32	95.78±1.65	89.56±8.49	92.35±4.08	94	2.9
DAGMM	53.22±2.63	97.81±0.32	23.98±4.32	38.37±5.39	424	3.9
SAE-FDD	92.21±1.65	96.25±1.08	90.80±2.55	93.43±1.47	23	2.3
SAE-MSE	78.95±8.05	95.69±1.19	68.53±12.91	79.37±9.20	23	2.3
DAE-FDD	88.66±6.68	95.21±3.63	85.60±8.54	89.90±6.12	23	2.3
DAE-MSE	77.57±8.47	94.66±2.98	66.89±13.39	77.87±10.09	23	2.3

**Fig.8.** The visual test results of each method on the C4 dataset with SNR=0 dB: (a) HRCAE; (b) PCDF-MSE; (c) CAE-FDD; (d) CAE-MSE; (e) MAE-FDD; (f) MAE-MSE.

On the C4 dataset, four evaluation indexes of all methods for anomaly detection are shown in **Table 4**. Under the heavy noise with SNR=0 dB ($P_n/P_s = 100\%$), the proposed method HRCAE achieves 96.25%, 96.01%, 97.95%, and 96.95% for the four evaluation indexes, respectively, showing superior and stable anomaly detection results over other methods. Although the DAGMM has the highest precision, its other three evaluation indexes are very bad, and the volatility is very high compared to the C1 dataset. After replacing the MSE with the FDD loss function, all methods show the

improvement in four evaluation indexes and have better stability, which are consistent with the performance on the C1 dataset. On the C4 dataset, the PCDF module also shows advantages compared to other methods. The running time regularity of all methods on the C4 dataset is consistent with that on the C1 dataset. **Fig. 8** shows the corresponding visual test results of six methods on the C4 dataset. Tool wear degree for the test set ranges from 0.03 mm to 0.20 mm, and wear degree above 0.10 mm are anomaly samples. Other plotting principles are the same as **Fig. 7**. On the C4 dataset, the PCDF module can better discriminate between normal and anomaly samples than other models, and the FDD loss function can further enhance the model's discrimination ability, which are consistent with **Fig. 7**.

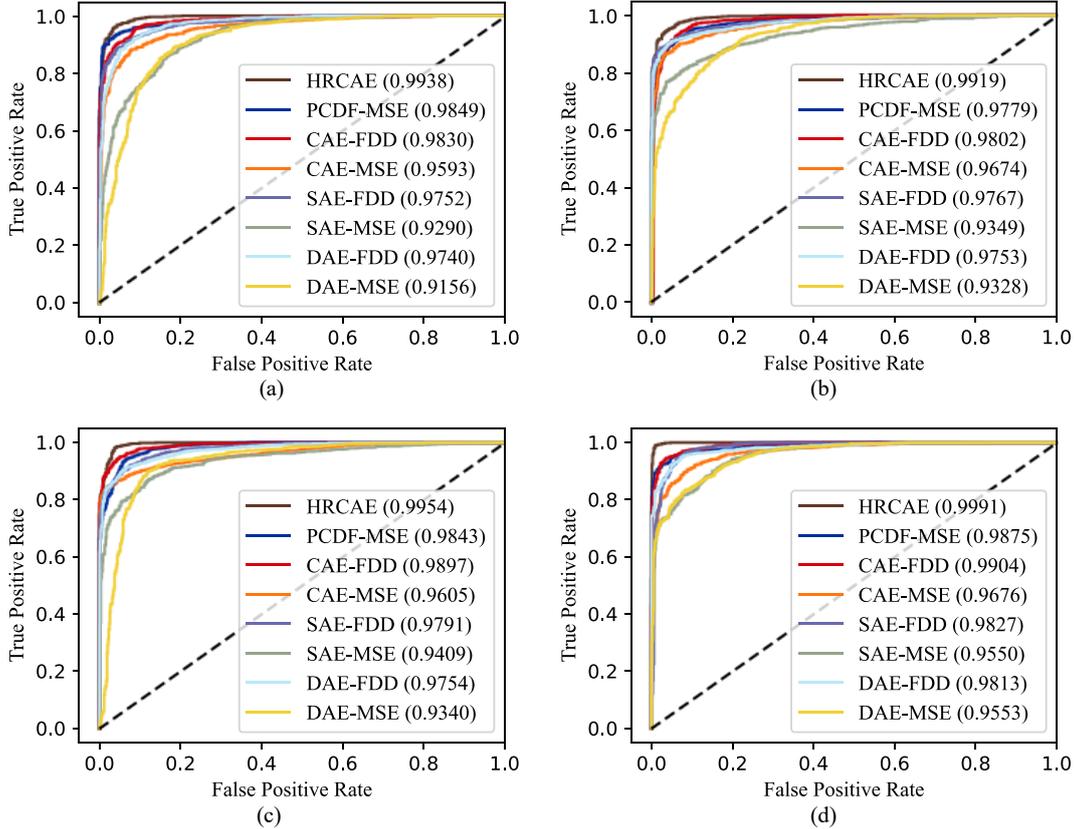


Fig.9. The ROC curves of methods under different noises on the C4 dataset: (a) SNR=0 dB; (b) SNR=1 dB; (c) SNR=2 dB; (d) SNR=4 dB.

Subsequently, we plot ROC curves and confusion matrixes to show the anomaly detection results of methods under different noise effects. **Fig. 9** shows the ROC curves for different methods on the C4 dataset with SNR of 0 dB, 1 dB, 2 dB, and 4 dB (P_n / P_s is approximately 100%, 80%, 60%, and 40%). The area under the curve (AUC) is used to measure the anomaly detection performance of each method and ranges from 0.5 to 1 [55]. The higher AUC value represents the better performance of the method, and the AUC values for methods are listed in parentheses of right corner. At SNR values of 0 dB, 1 dB, 2 dB, and 4 dB, the curves of the proposed method are closest to the upper left corner, and their AUC values are 99.38%, 99.19%, 99.54%, and 99.91%, respectively, which are stable and better than other methods. As for the PCDF module, the AUC values of the HRCAE compared to the CAE-FDD are improved by 1.08%, 1.17%, 0.57%, and 0.87% with SNR of 0 dB, 1 dB, 2 dB, and 4 dB; the AUC values of the PCDF-MSE compared to the CAE-MSE are improved by 2.56%, 1.05%, 2.38%, and 1.99%, respectively. After replacing MSE with FDD loss function, the AUC values of DAE are improved by 5.84%, 4.25%, 4.14% and 2.60% respectively; SAE are improved by 4.62%, 4.18%, 3.82% and 2.77% respectively; CAE are improved by 2.37%, 1.28%, 2.92% and 2.28% respectively; PCDF module are improved by 0.89%, 1.40%, 1.11% and 1.16% respectively.

Fig.10 shows the confusion matrixes for methods on the C1 dataset with SNR of 1 dB, 2 dB, 4 dB, and 7 dB (P_n/P_s is approximately 80%, 60%, 40%, and 20%). Coordinate 0 represents normal samples, with the total number of 600, while coordinate 1 represents anomaly samples, with the total number of 700. The horizontal coordinate is the predicted value and the vertical coordinate is the true value. We can intuitively see that the proposed method HRCAE has good classification results under different noises. At the SNR values of 1 dB, 2 dB, 4 dB, and 7 dB, the proposed method achieves F-score of 96.50%, 98.30%, 97.68%, and 98.24%, with small volatility under different noises. Both the constructed PCDF module and the designed FDD loss function show enhancement effects, further demonstrating the robustness of the proposed method for anomaly detection under different noises.

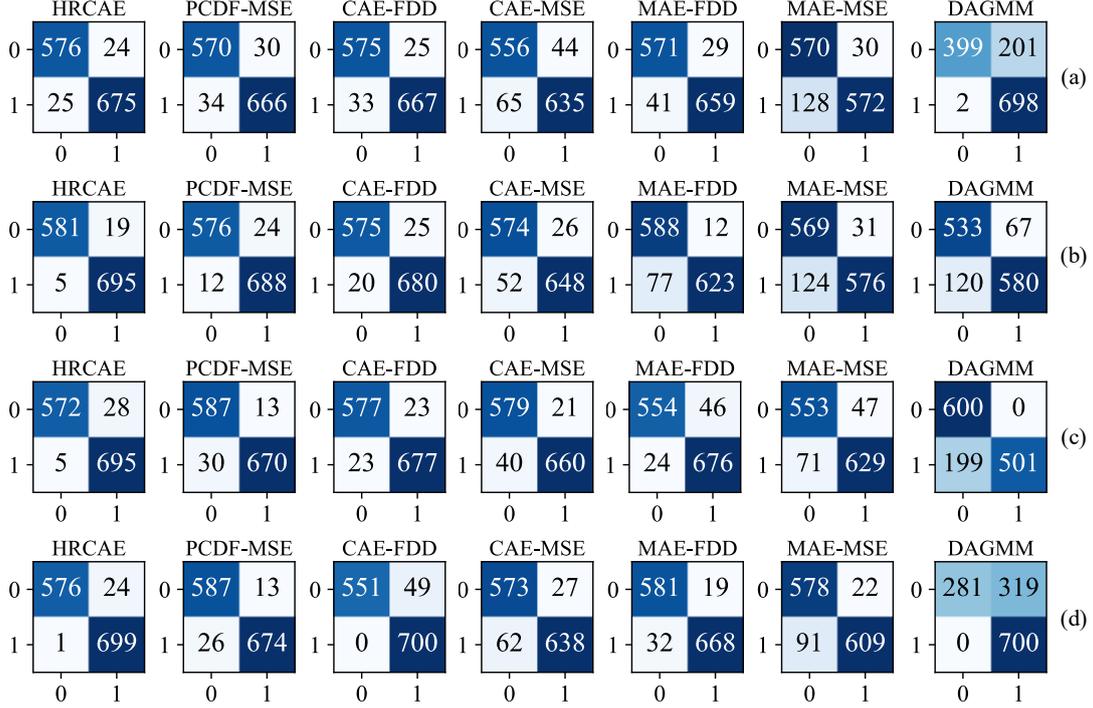


Fig.10. Confusion matrixes of methods under different noises on the C1 dataset: (a) SNR=1 dB; (b) SNR=2 dB; (c) SNR=4 dB; (d) SNR=7 dB.

4.3. Discussions on multi-sensor and λ values

Next, we verify the advantage of fusing multi-sensor information under different noises. **Fig.11** shows the accuracy and F-score of the proposed method under different sensor data. The figure legend XYZ indicates the multi-sensor data in three vibration directions, X, Y and Z indicate the single-sensor data in the vibration direction along that axis. Under the heavy noise with SNR = 0 dB on the C4 dataset, the average accuracies are 96.25%, 88.99%, 87.98%, and 79.49% for multi-sensor data and single-sensor data in X, Y, and Z directions, respectively; the F-values are 96.95%, 90.36%, 89.27%, and 80.64%, respectively. Under the influence of noise with SNR = 7 dB on the C1 dataset, the average accuracies are 97.60%, 86.66%, 86.85%, and 78.77% for multi-sensor data and single-sensor data in X, Y, and Z directions, respectively; the F-values are 97.80%, 86.08%, 86.73%, and 76.31%, respectively. Under the different noises in both datasets, the proposed method can effectively fuse multi-sensor information, and the anomaly detection results are significantly better than those of single-sensor data, reducing the requirements for sensor location arrangement and selection.

Finally, we investigate the values of the hyperparameter λ in the FDD loss function. **Fig.12** shows the evaluation indexes of the proposed method at different λ values under the heavy noise on the C1 and C4 datasets. To ensure that both distance and angle differences are useful, the MSE and CS loss

values need to be of the same order of magnitude. According to data preparation and Eqs.(3), (6-7), we choose λ values in the range of 0.05-0.3. With the increase of λ values, each evaluation index has a certain fluctuation. The fluctuation of recall is more intense compared with the other three evaluation indexes, which indicates that anomaly samples are relatively more sensitive to λ values; the accuracy and F-score first increase and then decrease. When $\lambda=0.2$, the proposed method is optimal in terms of accuracy, recall and F-score on both datasets.

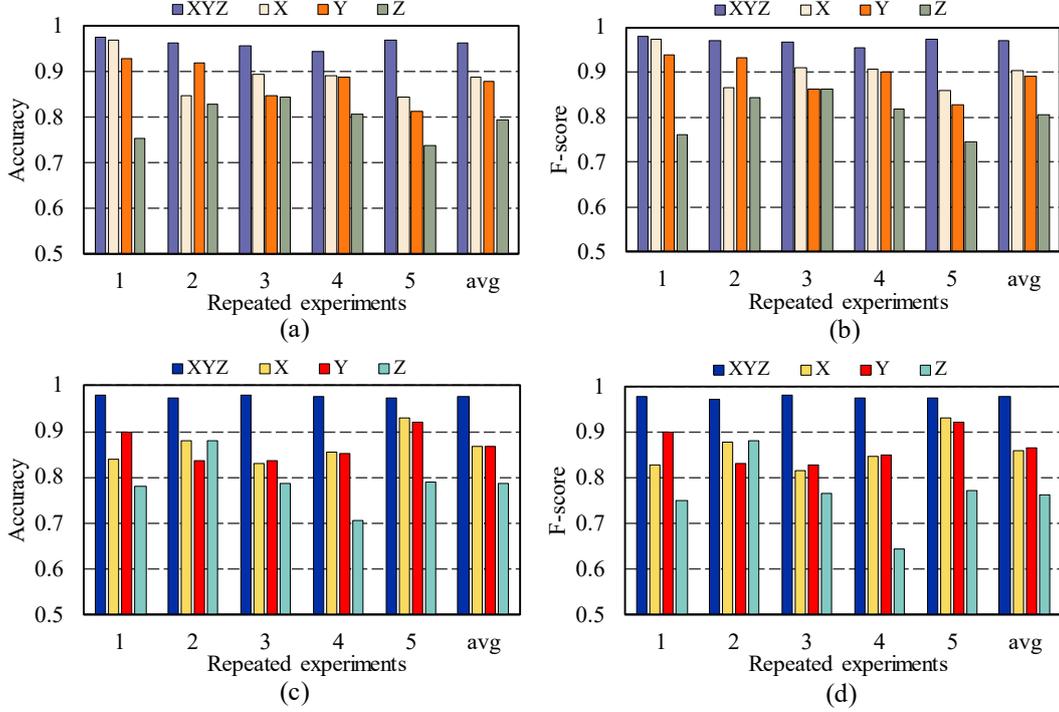


Fig.11. The evaluation indexes of the proposed method under different sensor data: (a) Accuracy on the C4 dataset with SNR=0 dB; (b) F-score on the C4 dataset with SNR=0 dB; (c) Accuracy on the C1 dataset with SNR=7 dB; (d) F-score on the C1 dataset with SNR=7 dB.

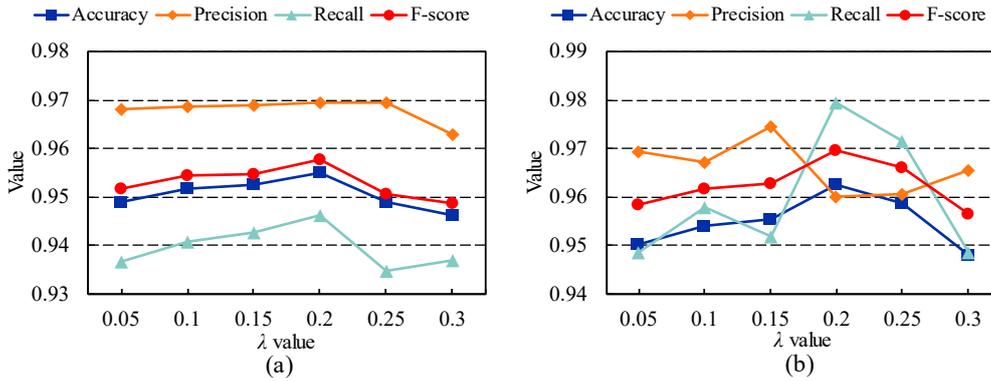


Fig.12. The evaluation indexes of the proposed method at different λ value: (a) SNR=1 dB on the C1 dataset; (b) SNR=0 dB on the C4 dataset.

5. Conclusions

To tackle challenges of difficulty to obtain labeled data and the presence of noises in anomaly detection of machine tools, a new unsupervised anomaly detection method named HRCAE is developed in this paper. The proposed method constructs a novel PCDF module and designs a FDD loss function to improve the robustness of anomaly detection. The constructed PCDF module can effectively fuse multi-sensor information and enhance the robustness of the network with parallel training to better learn the data distribution. The designed FDD loss function integrates the distance and

angle differences between the reconstructed data and the original data to further improve the robustness of the model.

The proposed method is validated by real CNC machine tool data and the comparison results show: (1) The constructed PCDF module can better fit the distribution characteristics of normal data for identification of anomaly data. (2) The designed FDD loss function can effectively suppress the effect of noises to improve the method reliability and has a certain degree of universality. (3) Compared with other commonly used unsupervised improved autoencoder methods, the proposed method can obtain better unsupervised anomaly detection under different noises. In future work, we will further evaluate the generalization of the proposed method to unsupervised anomaly detection on other mechanical devices. For the hyperparameters in the model, we will continue to explore a more general automatic hyperparameter determination approach to reduce the difficulty of model preparation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (No. 51905160) and the Natural Science Fund for Excellent Young Scholars of Hunan Province (No. 2021JJ20017).

References

- [1] Liu C, Vengayil H, Zhong R Y, *et al.* A systematic development method for cyber-physical machine tools, *Journal of manufacturing systems*, 48 (2018) 13-24.
- [2] Liu W, Kong C, Niu Q, *et al.* A method of NC machine tools intelligent monitoring system in smart factories, *Robotics and computer-integrated manufacturing*, 61 (2020) 101842.
- [3] Xu X. Machine Tool 4.0 for the new era of manufacturing, *The International Journal of Advanced Manufacturing Technology*, 92 (5) (2017) 1893-1900.
- [4] Wang L. From Intelligence Science to Intelligent Manufacturing, *Engineering*, 5 (4) (2019) 615-618.
- [5] Liu C, Zheng P, Xu X. Digitalisation and servitisation of machine tools in the era of Industry 4.0: a review, *International Journal of Production Research*, (2021) 1-33.
- [6] Ratava J, Lohtander M, Varis J. Tool condition monitoring in interrupted cutting with acceleration sensors, *Robotics and Computer-Integrated Manufacturing*, 47 (2017) 70-75.
- [7] Li G, Fu Y, Chen D, *et al.* Deep Anomaly Detection for CNC Machine Cutting Tool Using Spindle Current Signals, *Sensors*, 20 (17) (2020) 4896.
- [8] Zhou C, Guo K, Sun J. Sound singularity analysis for milling tool condition monitoring towards sustainable manufacturing, *Mechanical Systems and Signal Processing*, 157 (2021) 107738.
- [9] Li T, Shi T, Tang Z, *et al.* Real-time tool wear monitoring using thin-film thermocouple, *Journal of Materials Processing Technology*, 288 (2021) 116901.
- [10] Babouri M K, Ouelaa N, Djebala A. Experimental study of tool life transition and wear monitoring in turning operation using a hybrid method based on wavelet multi-resolution analysis and empirical mode decomposition, *The International Journal of Advanced Manufacturing Technology*, 82 (9) (2016) 2017-2028.
- [11] Niaki F A, Feng L, Ulutan D, *et al.* A wavelet-based data-driven modelling for tool wear assessment of difficult to machine materials, *International Journal of Mechatronics and Manufacturing Systems*, 9 (2) (2016) 97-121.
- [12] Yuan J, Liu L, Yang Z, *et al.* Tool Wear Condition Monitoring by Combining Variational Mode Decomposition and Ensemble Learning, *Sensors*, 20 (21) (2020) 6113.
- [13] Yu J, Cheng X, Lu L, *et al.* A machine vision method for measurement of machining tool wear, *Measurement*, 182 (2021) 109683.

- [14] Kong D, Chen Y, Li N. Hidden semi-Markov model-based method for tool wear estimation in milling process, *The International Journal of Advanced Manufacturing Technology*, 92 (9) (2017) 3647-3657.
- [15] Guo J, Li A, Zhang R. Tool condition monitoring in milling process using multifractal detrended fluctuation analysis and support vector machine, *The International Journal of Advanced Manufacturing Technology*, 110 (5) (2020) 1445-1456.
- [16] Wu D, Jennings C, Terpenney J, *et al.* A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests, *Journal of Manufacturing Science and Engineering*, 139 (7) (2017) 071018.
- [17] Duan J, Hu C, Zhan X, *et al.* MS-SSPCANet: A powerful deep learning framework for tool wear prediction, *Robotics and Computer-Integrated Manufacturing*, 78 (2022) 102391.
- [18] Chen X, Wang S, Qiao B, *et al.* Basic research on machinery fault diagnostics: Past, present, and future trends, *Frontiers of Mechanical Engineering*, 13 (2) (2018) 264-291.
- [19] Liang Y C, Wang S, Li W D, *et al.* Data-Driven Anomaly Diagnosis for Machining Processes, *Engineering*, 5 (4) (2019) 646-652.
- [20] Huang Z, Zhu J, Lei J, *et al.* Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations, *Journal of Intelligent Manufacturing*, 31 (4) (2020) 953-966.
- [21] Sun C, Ma M, Zhao Z, *et al.* Deep Transfer Learning Based on Sparse Autoencoder for Remaining Useful Life Prediction of Tool in Manufacturing, *IEEE Transactions on Industrial Informatics*, 15 (4) (2018) 2416-2425.
- [22] Cheng M, Jiao L, Yan P, *et al.* Intelligent tool wear monitoring and multi-step prediction based on deep learning model, *Journal of Manufacturing Systems*, 26 (2022) 286-300.
- [23] Deebak B D, Al-Turjman F. Digital-twin assisted: Fault diagnosis using deep transfer learning for machining tool condition, *International Journal of Intelligent Systems*, (2021).
- [24] Ou J, Li H, Huang G, *et al.* Tool Wear Recognition Based on Deep Kernel Autoencoder With Multichannel Signals Fusion, *IEEE Transactions on Instrumentation and Measurement*, 70 (2021) 1-9.
- [25] Shi C, Luo B, He S, *et al.* Tool Wear Prediction via Multidimensional Stacked Sparse Autoencoders With Feature Fusion, *IEEE Transactions on Industrial Informatics*, 16 (8) (2019) 5150-5159.
- [26] Wang G, Zhang F. A Sequence-to-Sequence Model With Attention and Monotonicity Loss for Tool Wear Monitoring and Prediction, *IEEE Transactions on Instrumentation and Measurement*, 70 (2021) 1-11.
- [27] Sun H, Zhang J, Mo R, *et al.* In-process tool condition forecasting based on a deep learning method, *Robotics and Computer-Integrated Manufacturing*, 64 (2020) 101924.
- [28] Yang Q, Pattipati K R, Awasthi U, *et al.* Hybrid data-driven and model-informed online tool wear detection in milling machines, *Journal of Manufacturing Systems*, 63 (2022) 329-343.
- [29] Li X, Zhang Z, Gao L, *et al.* A New Semi-Supervised Fault Diagnosis Method via Deep CORAL and Transfer Component Analysis, *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6 (3) (2021) 690-699.
- [30] Zong B, Song Q, Min M R, *et al.* Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection, *International Conference on Learning Representations*, (2018).
- [31] Gong D, Liu L, Le V, *et al.* Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019) 1705-1714.
- [32] Thill M, Konen W, Wang H, *et al.* Temporal convolutional autoencoder for unsupervised anomaly detection in time series, *Applied Soft Computing*, 112 (2021) 107751.

- [33] Li D, Guo H, Wang Z, *et al.* Unsupervised Fake News Detection Based on Autoencoder, IEEE Access, 9 (2021) 29356-29365.
- [34] Lyu P, Zhang K, Yu W, *et al.* A novel RSG-based intelligent bearing fault diagnosis method for motors in high-noise industrial environment, Advanced Engineering Informatics, 52 (2022) 101564.
- [35] Fang Z, Jia T, Chen Q, *et al.* Laser stripe image denoising using convolutional autoencoder, Results in Physics, 11 (2018) 96-104.
- [36] Jiang J, Ren H, Zhang M. A Convolutional Autoencoder Method for Simultaneous Seismic Data Reconstruction and Denoising, IEEE Geoscience and Remote Sensing Letters, 19 (2021) 1-5.
- [37] Zhu Y, Li L, Wu X. Stacked Convolutional Sparse Auto-Encoders for Representation Learning, ACM Transactions on Knowledge Discovery from Data, 15 (2) (2021) 1-21.
- [38] Yu J, Zhou X. One-Dimensional Residual Convolutional Autoencoder Based Feature Learning for Gearbox Fault Diagnosis, IEEE Transactions on Industrial Informatics, 16 (10) (2020) 6347-6358.
- [39] Szegedy C, Liu W, Jia Y, *et al.* Going Deeper With Convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015) 1-9.
- [40] Liu H, Han J, Hou S, *et al.* Single image super-resolution using a deep encoder–decoder symmetrical network with iterative back projection, Neurocomputing, 282 (2018) 52-59.
- [41] Ruderman A, Rabinowitz N C, Morcos A S, *et al.* Pooling is neither necessary nor sufficient for appropriate deformation stability in CNNs, arXiv:1804.04438, (2018).
- [42] Wu L, Wang Y, Gao J, *et al.* Deep adaptive feature embedding with local sample distributions for person re-identification, Pattern Recognition, 73 (2018) 275-288.
- [43] Sethi A, Singh M, Singh R, *et al.* Residual Codean Autoencoder for Facial Attribute Analysis, Pattern Recognition Letters, 119 (2019) 157-165.
- [44] Liu D, Liu Y, Wang L. Distance measure for Fermatean fuzzy linguistic term sets based on linguistic scale function: An illustration of the TODIM and TOPSIS methods, International Journal of Intelligent Systems, 34 (11) (2019) 2807-2834.
- [45] Cai L, Zhou S, Yan X, *et al.* A Stacked BiLSTM Neural Network Based on Coattention Mechanism for Question Answering, Computational Intelligence and Neuroscience, (2019).
- [46] Sheikh Fathollahi M, Razzazi F. Music similarity measurement and recommendation system using convolutional neural networks, International Journal of Multimedia Information Retrieval, 10 (1) (2021) 43-53.
- [47] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization, arXiv:1412.6980, (2014).
- [48] Li X, Lim B S, Zhou J H, *et al.* Fuzzy Neural Network Modelling for Tool Wear Estimation in Dry Milling Operation, Annual Conference of the PHM Society, 1 (1) (2009).
- [49] Zhao R, Yan R, Wang J, *et al.* Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks, Sensors, 17 (2) (2017) 273.
- [50] Wang J, Xie J, Zhao R, *et al.* Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing, Robotics and Computer-Integrated Manufacturing, 45 (2017) 47-58.
- [51] Xiao Y, Shao H, Han S Y, *et al.* Novel Joint Transfer Network for Unsupervised Bearing Fault Diagnosis From Simulation Domain to Experimental Domain, IEEE/ASME Transactions on Mechatronics, (2022).
- [52] Wen L, Wang Y, Li X. A new automatic convolutional neural network based on deep reinforcement learning for fault diagnosis, Frontiers of Mechanical Engineering, 17 (2) (2022) 1-12.
- [53] Sun W, Shao S, Zhao R, *et al.* A sparse auto-encoder-based deep neural network approach for induction motor faults classification, Measurement, 89 (2016) 171-178.
- [54] Vincent P, Larochelle H, Bengio Y, *et al.* Extracting and composing robust features with denoising autoencoders, Proceedings of the 25th international conference on Machine learning, (2008)

1096-1103.

- [55] Yan S, Shao H, Xiao Y, *et al.* Semi-supervised fault diagnosis of machinery using LPS-DGAT under speed fluctuation and extremely low labeled rates, *Advanced Engineering Informatics*, 53 (2022) 101648.