



## Practice article

## Contrastive learning and dynamics embedding neural network for label-free interpretable machine fault diagnosis

Shilong Sun <sup>a,\*</sup>, Tengyi Peng <sup>a</sup>, Yu Zhou <sup>b</sup>, Xiao Zhang <sup>c</sup>, Dong Wang <sup>d</sup><sup>a</sup> School of Mechanical Engineering and Automation, Guangdong Key Laboratory of Intelligent Morphing Mechanisms and Adaptive Robotics, Harbin Institute of Technology(Shenzhen), Shenzhen, China<sup>b</sup> College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China<sup>c</sup> College of Computer Science, South-Central Minzu University, Wuhan, Hubei 430074, China<sup>d</sup> State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China

## ARTICLE INFO

## Keywords:

Interpretable fault diagnosis  
Contrastive learning  
Dynamics embedding neural network  
Unlabelled vibrational data

## ABSTRACT

Industrial machinery often produces vibration signals that can serve as indicators of underlying faults. However, these signals often need to be labeled, presenting a challenge for accurate and interpretable fault diagnosis. While supervised learning methods, such as deep neural networks, have been applied for fault diagnosis, they need help in effectively distinguishing between different vibration-related faults. In response to this issue, our study introduces an innovative approach for automatic fault diagnosis through the application of the Bootstrap Your Own Latent and Dynamical Systems Model Discovery algorithm (BYOLDIS). This method not only addresses the challenge of unlabelled signals but also provides readily interpretable results. The proposed methodology consists of three fundamental steps. First, we derive a matrix of differential equations to capture the dynamic behavior of faulty bearings. Second, we employ a contrastive learning network alongside a time-delay embedding matrix to reconstruct the coordinates of the fault-dynamical system. Lastly, we construct a library of fault machine dynamic polynomial equations, incorporating prior constraints based on physical models. To assess the effectiveness and robustness of our proposed method, we conducted both simulations and experiments. The results of these case studies affirm that BYOLDIS can accurately diagnose bearing faults and offer dynamic explanations for the diagnostic outcomes. This suggests that BYOLDIS holds substantial promise as a diagnostic tool for processing unlabelled vibrational data.

## 1. Introduction

Bearings are widely used in rotating machinery, and their health status is critical for ensuring stable operations. Fault diagnosis is a crucial aspect of the mechatronic system design and maintenance [1]. With the emergence of Industry 4.0, intelligent fault diagnosis framework networks have gained much attention from enterprises. Intelligent fault diagnosis involves the application of artificial intelligence (AI) algorithms to traditional diagnosis methods, such as deep learning (DL), transfer learning (TL), contrastive learning (CL), meta-learning (ML), etc. DL-based fault diagnosis procedures typically involve two steps: data collection and health state recognition [2]. Due to its high feature extraction capability, DL can automatically extract meaningful features. Recently, numerous DL-based studies have been presented to diagnose bearing faults, such as autoencoder networks [3], convolutional neural

networks [4], and deep-belief networks [5]. However, the existing research largely overlooks two crucial issues regarding the neural network method in machine fault diagnosis: 1) the amount of labeled data significantly affects the performance of AI machine fault diagnosis methods, and 2) most neural network structures are black boxes and lack the interpretation of the diagnosis result.

In the context of industrial applications, acquiring sufficient labelled data to train neural networks for machine fault diagnosis is often challenging due to harsh working conditions. Scholars have attempted to address the issues of label and data insufficiency, including TL, CL and ML methods. TL can mitigate labelled data imperfections by utilizing prior knowledge learned from a sufficiently labelled dataset. Still, it relies on minimal distribution discrepancy between the source and target domain data, which may not be feasible in real industrial applications [6–8]. For example, Zhang [9] added the maximum mean

\* Corresponding author.

E-mail address: [sunshilong@hit.edu.cn](mailto:sunshilong@hit.edu.cn) (S. Sun).

discrepancy (MMD) and weight regularization to the loss function to achieve domain adaption. Xia [10] proposed a multisource domain adaptation network to fuse the multidomain features. Some other scholars used the adversarial domain network (ADN) to reduce the domain distribution discrepancy by making the domain discriminator unable to distinguish the domain to which the data belongs to learn the transferable feature of bearing faults in the target domain [11,12]. In contrast, CL is a powerful unlabelled data learning method that has been successfully used in self-supervised representation learning in computer vision. Still, its application in machine fault diagnosis is limited. Several studies have proposed CL-based approaches that can directly extract fault features from unlabelled vibration signals without labelled data [13–16] and have utilised the CL feature extractor to solve few-shot learning fault diagnosis tasks. Peng [17] proposed a contrastive learning approach based on the Bootstrap Your Own Latent network. This approach can directly extract disguised fault features from unlabelled vibration signals using a data augmentation method. Rombach [18] proposed a CL-based machine fault detection and diagnostics method by constructing reasonable positive and negative pairs to ensure that the model is sensitive to unseen fault data. Chen [19] proposed a supervised CL method for high-frequency machine vibration signal fault diagnosis in the few-shot learning situation. Lu [20] combined CL with a label diffusion module to deal with a few-shot learning fault diagnosis task.

Additionally, meta-learning (ML) models have been utilized to reduce data reliance in intelligent machine fault diagnosis [21–23]. Pei [24] proposed an enhanced few-shot Wasserstein auto-encoder (fs-WAE) to reverse the negative effect of imbalance. Li [25] proposed a novel meta-learning fault diagnosis method (MLFD) based on model-agnostic meta-learning to deal with limited data and complex working conditions. However, most studies require labelled information to guide the neural network learning classification for fault diagnosis. This is because most current neural network fault diagnosis studies are migration applications of the models proposed for solving computer vision and natural language processing problems. These neural networks do not have any prior constraints and must feed a large amount of data to converge to a posterior probability result independent of the prior. Fortunately, the prior constraints of the fault signal collected from the machine are expressible and governed by a multibody dynamics model. Therefore, a fault diagnosis model without label data can exist. One of the objectives of this study is to develop a neural network model that can achieve label-free fault diagnosis. We look forward to the network to analyse the dynamic features of the input signal as humans process the vibration signal. This neural network contains a multibody dynamics model that offers constraints on machinery fault signals prior to network training. This method can improve current neural network-based fault diagnosis studies that lack knowledge constraints and require a large amount of data to converge to new parameters for fault classification. Therefore, the primary goal of this study is to develop a neural network-based model that can achieve fault diagnosis of unlabelled data by incorporating the knowledge constraints of multibody dynamics.

In addition to the challenges of insufficient labelled data and imbalance, another issue in intelligent fault diagnosis is the need for more interpretability of neural networks. Current research on interpretable fault diagnosis focuses on two aspects: time series characteristics and signal morphology. The former is based on statistical, spectral, and time-frequency characteristics of the signals, while the latter is based on geometric feature extraction methods such as impact-impulse analysis [26,27]. Several studies have proposed novel approaches for interpretable fault diagnosis, including the use of Gramian-based noise reduction strategies [28], fully interpretable neural networks [29], deep wavelet transform approaches [30], and continuous wavelet convolutional layers [27]. For example, Yuan [28] first proposed a simple but efficient Gramian-based noise reduction strategy called Gramian Noise Reduction (GNR) based on the periodic self-similarity of vibrational signals. Wang [29] designed a fully interpretable neural network for

machine condition monitoring, which is a redesigned extreme learning machine combined with signal pre-processing methods, such as wavelet transform, square envelope, and Fourier transform. Michau [30] proposed a fully learnable deep wavelet transform approach that embeds the crucial properties of the fast discrete wavelet transform and also designed a new activation function to achieve industrial asset monitoring. Other studies have explored the use of physics-based models to achieve a trustworthy diagnosis by extracting crucial characteristics of the governed dynamic equation from the input fault-bearing signal [31, 32]. The signal morphology-based interpretable neural network is based on the impact-impulse geometric feature extraction method. Ye [26] proposed a deep morphological network to learn geometric features for interpretable gearbox fault diagnosis. Li [27] designed a continuous wavelet convolutional layer to impact impulse separation for interpretable fault diagnosis. Zhou [31] proposed a probabilistic Bayesian deep learning framework to achieve a trustworthy diagnosis via an uncertainty-aware model to understand the unknown fault information. Zuo [32] proposed a multi-layer spiking neural network for bearing fault diagnosis, which can provide interpretability transparency to the different bearing faults.

Inspired by the studies of data-driven dynamic equations discovered [33,34], another target is to find the crucial characteristic of the governed dynamic equation from the input fault-bearing signal, allowing the diagnosis result to have an excellent physical interpretation. However, developing physics-based models for diagnosis applications can be challenging due to the complexity of the systems and the large variability that may exist in system parameters [35]. Nonetheless, inspired by data-driven dynamic equations, researchers aim to find the crucial characteristics of the dynamic equation that can allow for a physical interpretation of the diagnosis results.

Therefore, this research work mainly focuses on two issues in the intelligent fault diagnosis area which are interpretability and insufficient label signal. One of the challenges is the need for more interpretability in most intelligent fault diagnosis methods. Interpretability is important because it allows engineers and operators to understand the reasoning behind the diagnosis and take appropriate actions. To address this issue, current research on interpretable fault diagnosis focuses on time series characteristics and signal morphology. These methods aim to extract meaningful features from the data that domain experts can easily interpret. Another challenge is the need for a large amount of labeled data to train neural networks. Labeling data can be time-consuming and expensive, especially in industrial settings where the data may be noisy and complex. To overcome this challenge, researchers have been exploring the use of unsupervised learning methods, which do not require labeled data. Unsupervised learning methods can learn patterns in the data and use them to make predictions without the need for human annotations.

In this context, the proposed study aims to diagnose machine faults directly from raw industrial vibration signals collected under industrial conditions. The authors use a physical model-based equation library to shape prior constraints and directly complete automatic fault diagnosis without labeling the input data. They also propose a contrastive learning algorithm based on Bootstrap Your Own Latent (BYOL) network to determine the dynamic system equation from the fault-bearing signal input. Finally, they present an automatic fault diagnosis method combining the BYOL network and the discovered dynamic system equation method (BYOLDIS) to offer a straightforward physical interpretation of the diagnosis results.

Overall, the proposed study aims to diagnose the fault type by directly handling the raw industrial vibration signals collected under industrial conditions. The main contributions of this study are as follows.

- (1) Building a physical model-based equation library to shape prior constraints, enabling automatic fault diagnosis without labelling input data.

- (2) Designing a contrastive learning algorithm based on Bootstrap

Your Own Latent network, consisting of a dual network, the online and target network, to determine the dynamic system equation from the fault-bearing signal input.

(3) Proposing an automatic fault diagnosis method combining the BYOL and BYOLDIS to offer a straightforward physical interpretation.

The outline of this paper is organized as follows. Section 2 introduces the related work of BYOL network and the coordination discovery method. Section 3 presents the principles of the proposed algorithm. Section 4 presents an experimental study of the proposed model, and Section 5 summarizes this work.

## 2. Preliminaries

The neural network-based intelligence fault diagnosis methods have strong feature extraction ability and achieve effective fault diagnosis. However, those methods require data for training. Most current neural network-based fault diagnosis studies are migration applications of the proposed initial models for solving computer vision and natural language processing problems. These neural networks do not know the task and need to feed a large amount of data to converge to the final task used knowledge. The fault signal knowledge is expressible and governed by the multibody dynamics model. Therefore, we suppose adding some dynamics knowledge to reduce the training data is feasible. This paper proposes to embed the dynamic system into a neural network for giving network knowledge of diagnosis, achieving label-free fault diagnosis based on the dynamic system discovery and reconstruction from input vibration signals. In this paper, we utilize BYOL contractive learning methods' characteristics of label-free learning and dynamics system discovery methods to reconstruct dynamic equations to realize fault diagnosis. This Section describes the crucial background knowledge about Bootstrap Your Own Latent (BYOL) network, sparse identification of nonlinear dynamics (SINDy), and delay-time embedding.

### 2.1. BYOL network

BYOL [36] network can extract the prominent features directly without labeled data, which means that the BYOL network can achieve effective feature learning for unseen unlabelled data. This feature of BYOL is handy for label-free fault diagnosis because it requires stable network convergence without any external supervision information. As shown in Fig. 1, BYOL [36] is a dual structure that consists of online and target networks. The online network is defined by a set of weights called  $\theta$ . The encoder  $f_\theta$  has a ResNet design [37], and the encoder output  $R_\theta$  is a network-learned representation of the input training samples. The projector  $q_\theta$  and predictor  $g_\theta$ ,  $g_\theta$  consist of a multi-layer perceptron (MLP), and the output dimensions are the same. The target network has the same structure as the online network without a predictor  $g_\theta$  but has a stop-gradient layer to block the weight  $\xi$  gradient update.  $\xi$  is a moving average of  $\theta$ .

BYOL processes two positive pairs generated from samples  $s$  and  $s'$ . The online network outputs representation  $y_\theta$  of the first processed sample  $s$  and the target network outputs  $y'_\xi$  from the second processed sample  $s'$ .

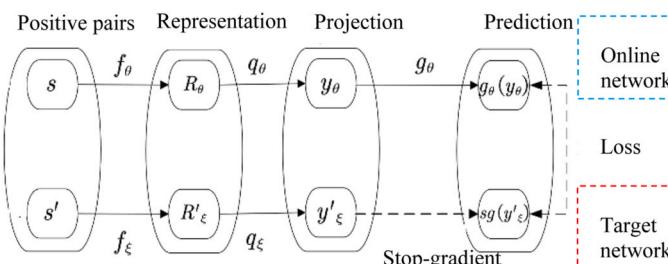


Fig. 1. Framework of the BYOL network.

Subsequently, BYOL utilizes  $g_\theta(y_\theta)$  and  $sg(y'_\xi)$  to calculate the network training loss:  $l_2$ -Normalized  $g_\theta(y_\theta)$  and  $sg(y'_\xi)$ . Then, defining  $\overline{g_\theta(y_\theta)} \triangleq g_\theta(y_\theta)/\|g_\theta(y_\theta)\|_2$  and  $\overline{sg(y'_\xi)} \triangleq sg(y'_\xi)/\|sg(y'_\xi)\|_2$ . At last, the loss function of the BYOL network is defined as the mean squared error between the normalized prediction  $\overline{g_\theta(y_\theta)}$  and normalized representation  $\overline{sg(y'_\xi)}$ , as shown in the following equation:

$$\mathcal{L}_{\theta,\xi} \triangleq \|\overline{g_\theta(y_\theta)} - \overline{sg(y'_\xi)}\|_2^2 \quad (1)$$

The loss  $\mathcal{L}_{\theta,\xi}$  in Eq. (1) requires  $s$  and  $s'$  to be fed separately into the online and target networks for computation. The BYOL updating progress can be summarized in two steps: Firstly, updating the weights  $\theta$  by using the gradient; Secondly, updating the weights  $\xi$  by using the updated  $\theta$ . The following two equations can describe this updating method:

$$\theta = \text{optimizer}(\theta, \delta\theta, \eta) \quad (2)$$

$$\xi = \tau\xi + (1 - \varepsilon)\theta \quad (3)$$

where  $\eta$  is the learning rate of the optimizer of the network and  $\varepsilon$  is the target network moving average to update the hyperparameter. The BYOL network relies on minimizing  $\mathcal{L}_{\theta,\xi}$  to reduce the representation distribution distance of inputs  $s$ ,  $s'$  and learning the representation of the input training data samples.

### 2.2. SINDy autoencoder and delay-time embedding

An ordinary differential equation system can be used to generate the simulated signal for a signal collected from the machine system. This ordinary differential equation describes the signal's crucial dynamics characteristic. It can define the governed equations with the analysis of the collected signal. The SINDy autoencoder [34] is an effective, hidden coordination discovery method for seeking the unknown and potential terms of the governed equations. For systems with full-state measured data, the SINDy autoencoder can discover the equations of the dynamic system. The network structure of the SINDy autoencoder is illustrated in Fig. 2.

The SINDy autoencoder can map the measurement data  $y$  into hidden coordination  $z$  and discover a nonlinear differential equation  $\dot{z} = \Theta(z)\Xi$ . In the SINDy model,  $\Theta$  is the candidate equation term library of the unknown system.  $\Xi$  is a sparse matrix representing the weight of each candidate term in the differential equation. In addition to the conventional reconstruction loss  $\mathcal{L}_{\text{recon}}$ , the SINDy autoencoder also contains other loss function terms to guide the encoder  $\phi$  and learn the reliable hidden coordination  $z$  via equation constraints:

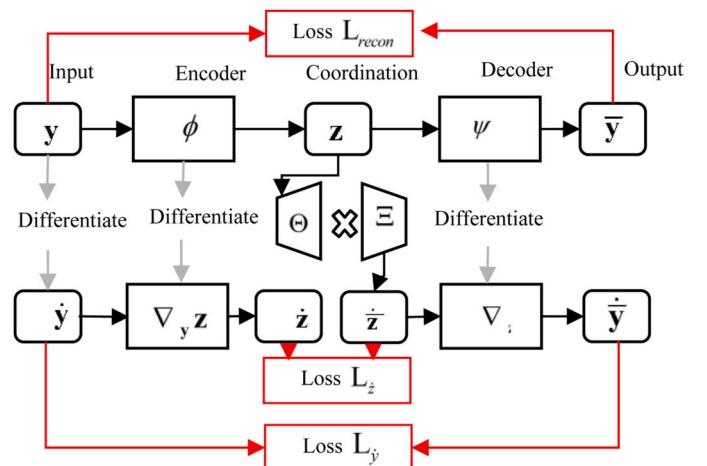


Fig. 2. Structure of the SINDy autoencoder algorithm.

$$\mathcal{L}_z \triangleq \|\nabla_y \phi(y) \dot{y} - \Theta(\phi(y)) \Xi\|_2^2 \quad (4)$$

$$\mathcal{L}_{\dot{y}} \triangleq \|\dot{y} - \nabla_z \psi(\phi(y)) \Theta(\phi(y)) \Xi\|_2^2 \quad (5)$$

Besides, a sparsity constraint is added to get sparse equations. The loss of the SINDy autoencoder is,  $\lambda_1, \lambda_2, \lambda_3$  are the hyperparameters to balance loss terms of the loss.

The state-space reconstruction technology can capture the dynamics systems information when the systems do not have full state-measured data. According to the state-space reconstruction in the delay coordinate method proposed by Packard [38], the unknown system characteristics can be analyzed by sampling this system's partial output time series. For systems with incomplete observations, the research in [39] showed that delay coordinate embedding could capture nonlinear systems' dynamics. After obtaining a measurement time series  $y$ , the delay coordinate embedding can be built as a Hankel matrix  $\mathbf{H}$ :

$$\mathbf{H} = \begin{bmatrix} y(\tau) & y(2\tau) & \cdots & y(q\tau) \\ y(2\tau) & y(3\tau) & \cdots & y((q+1)\tau) \\ \vdots & \vdots & \ddots & \vdots \\ y(n\tau) & y((n+1)\tau) & \cdots & y((n+q-1)\tau) \end{bmatrix} \quad (6)$$

where  $\tau$  is the delay time of the measured time series  $y$ , and  $n$  is the embedding dimension. Hankel matrix is a set  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_q]$ , and  $\mathbf{h}_i = [y(i\tau), y((i+1)\tau), \dots, y((i+n-1)\tau)]^T$ . According to Takens' theorem, the embedding dimension should satisfy the equation  $n \geq 2d + 1$ , which  $d$  is the unknown system's dimension. And in the research [40], the authors recommend the parameter  $n$  follow  $n\tau = 0.1$  for effectively reconstructing the critical characteristic of the anonymous dynamics system.

### 3. Built up the bootstrap your own latent and discovery system (BYOLDIS) network

In this Section, we describe the building process of the bootstrap your own latent and discovery system (BYOLDIS) network. This method adds

an equation library based on the dynamic equations to the neural network, forcing the network to choose one of the three fault modes (inner race fault, outer race fault, and roller fault) when reconstructing the dynamic system to realize fault diagnosis.

As shown in Fig. 3, BYOL helps the network discover the general characteristics of the positive pairs to achieve an effective dynamic coordinate reconstruction; DIS represents the dynamics system discovered for diagnosis. These two crucial parts combined into the BYOLDIS diagnosis of label-free conditions.

Generally, only one vibration sensor is assembled on the pedestal for bearing fault diagnosis. Only the acceleration information of one degree of freedom (DOF) is known. However, in building a fault-bearing dynamics system, two orthogonal DOF information (acceleration, velocity, displacement) of the shaft and pedestal is necessary. In reference [40], Bakarji proposed a deep delay autoencoder, which utilizes the universal approximation properties of neural networks to approximate the Phase space coordinates from the partial measurement signal input for signal governing equations discovery. In this study, for the 4DOF fault-bearing system, only 1DOF information is known, which is a typical partial measurement issue. Therefore, a delay time embedding-based coordinate encoder inspired by the Bakarji study is utilized to conduct 4DOF phase space reconstruction.

In summary, the first step of BYOLDIS is building delay time embedding positive pairs, which will be presented in Section 3.1. Section 3.2 offers the network structure and loss function for creating the dynamics system embedding a neural network. As mentioned above, for each DOF, the acceleration, velocity, and displacement information are necessary to build the 4DOF fault-bearing dynamics system. Therefore, the coordinate encoder needs to reconstruct the acceleration, velocity, and displacement information at the same time. Section 3.3 presents the coordinate encoder differential output calculation in the forwarding propagation process to rebuild the acceleration, velocity, and displacement information simultaneously. Section 3.4 introduces the embedding dynamics, fault-bearing equation library, and diagnosis method. This fault-bearing dynamics system is based on the multibody dynamics

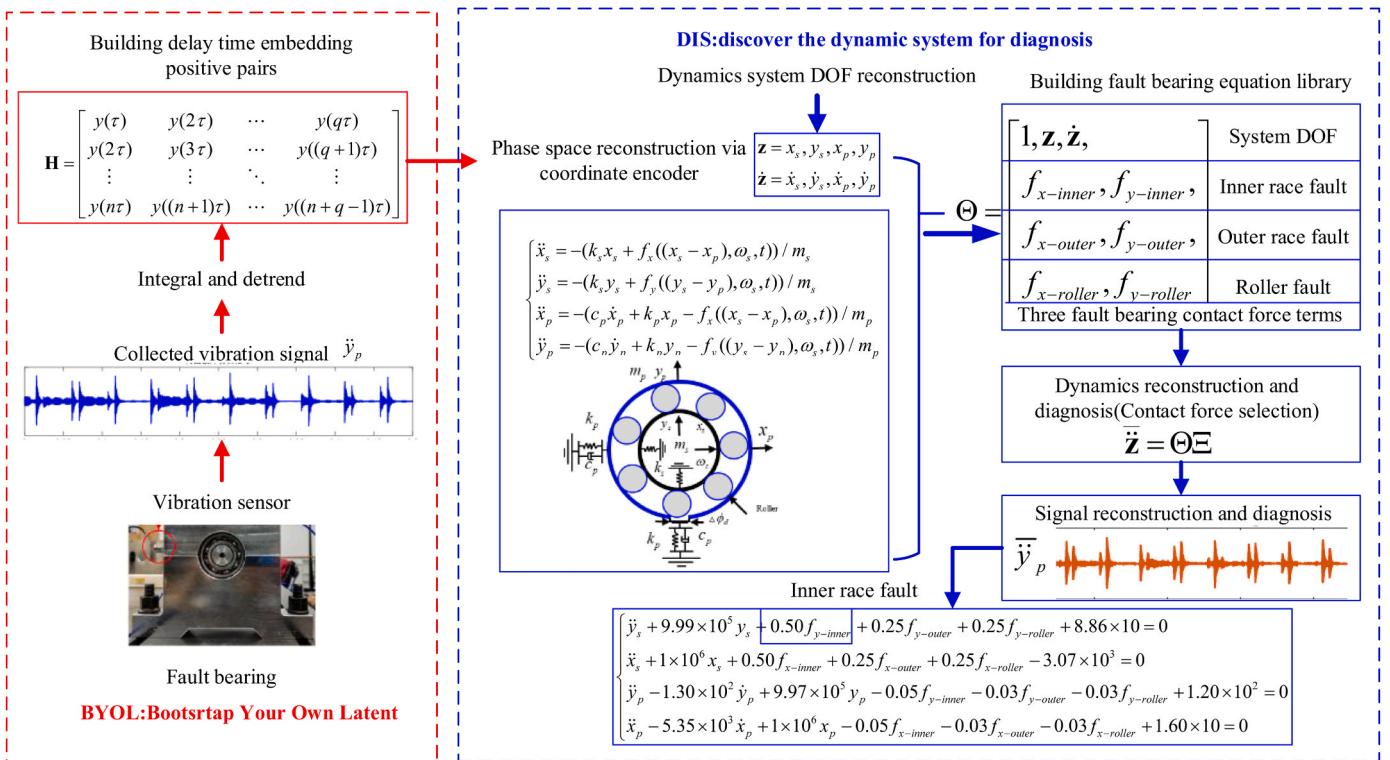
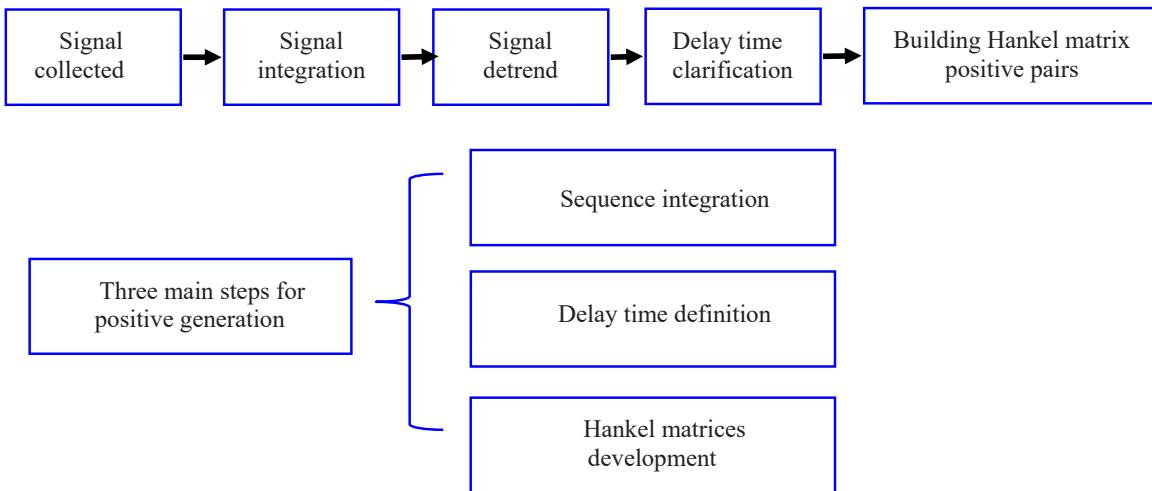


Fig. 3. Flowchart of the proposed BYOLDIS diagnosis method.



**Fig. 4.** Plot of the positive pairs generation.

methods, simulating the fault-bearing by constructing contact interference. Besides, the dynamics model of [Section 3.4](#) is specially designed to ensure that embedding dynamics can effectively conduct gradient information transmission. [Section 3.5](#) introduces the process of proposed BYOLDIS.

### 3.1. Positive pairs generation

The generation of positive pairs is crucial for the contrastive learning algorithm. This Section introduces how to build positive pairs with the Hankel matrix form. The vibration signal collected for the abnormal bearing is an acceleration signal. However, displacement and velocity signals are necessary for BYOLDIS to reconstruct the fault-bearing dynamic equation. Generally, there are three steps for the positive pairs generation. [Fig. 4](#).

The first step is sequence integration, assuming the collected vibration signal in time  $t$  is  $\ddot{x}(t)$ , the integration  $\dot{x}(t)$  and  $x(t)$  follow:

$$\begin{aligned} \dot{x}(t) &= \ddot{x}(t-1) + \ddot{x}(t)dt \\ x(t) &= x(t-1) + \dot{x}(t)dt \end{aligned} \quad (7)$$

where  $dt$  denotes the sampling time of the vibration signal. Because vibration signals  $\ddot{x}$  are often collected for only a few seconds, the integration sequence  $\dot{x}$  and  $x$  will have a certain trend. Therefore, detrending for sequences  $\dot{x}$  and  $x$  is necessary. In the latter part of this study, the symbols  $\dot{x}$  and  $x$  represent the detrend integration sequence (i.e.,  $\dot{x} \triangleq \text{detrend}(\dot{x})$ ).  $x(t)$  will be normalized (standard deviation set as 1);  $\dot{x}(t)$  and  $\ddot{x}(t)$  multiply by the same value utilized to normalize  $x(t)$ .

The second step is determining the delay time  $\tau$ . For the kind of state-space reconstruction issue that utilizes the coordinate delay method. If two coordinate components  $x(it), x((i+1)\tau)$  are similar with each other, then they cannot provide independent coordinate components. However, if the delay time  $\tau$  is too long, the two coordinate components will be completely independent, leading to each dimension in the reconstruction space not correlating. Therefore, this paper uses the autocorrelation coefficient method to fund the delay time  $\tau$ . The delay time  $\tau$  is ensured via the mutual information method. For the two signal sequences  $\{x_1^1, \dots, x_m^1\}$  and  $\{x_1^2, \dots, x_m^2\}$ , the information entropy is calculated by:

$$H(x) = - \sum_{i=1}^m P_x(x_i) \log_2 P_x(x_i) \quad (8)$$

where  $P_x(x_i)$  means the happened possibility of  $x_i$  in signal sequence  $x$ . Then the mutual information of  $x^1$  and  $x^2$  follows:

$$I(x^1, x^2) = H(x^1) + H(x^2) - H(x^1, x^2) \quad (9)$$

where  $H(x^1, x^2)$  can be calculated via  $H(x^1, x^2) = - \sum_{i=1}^m \sum_{j=1}^n P_{x^1, x^2}(x^1, x^2) \log_2 P_{x^1, x^2}(x^1, x^2)$ . When calculating the mutual information for ensuring delay time  $\tau$ , the original signal sequence with delay is  $x^1 = \{x_1, \dots, x_{m-\tau}\}$  and the delay time signal sequence is  $x^2 = \{x_{1+\tau}, \dots, x_m\}$ . After calculating the mutual information of  $x^1$  and  $x^2$ , we need to distinguish the first local minimum value of the mutual information of  $x^1, x^2$ . Then the  $\tau$  is the delay time.

The final step is building the two Hankel matrices via [Eq.\(6\)](#).  $n$  is equal to  $n = t_{cir}/\tau$ , where  $t_{cir}$  represents time required for bearing to rotate for one circle. The positive pairs ( $\mathbf{H}, \mathbf{H}_p$ ) are collected from the same sequences  $y$  but had different starting times, i.e., using  $\mathbf{x}(\tau : (n+q-1)\tau)$  to build  $\mathbf{H}$  and  $\mathbf{x}(0.5(n+q-1)\tau : 1.5(n+q-1)\tau)$  to build  $\mathbf{H}_p$ . It is noteworthy that  $\mathbf{H}$  and  $\mathbf{H}_p$  are also made in the same way.

### 3.2. Framework of the BYOLDIS network

This Section introduces the BYOLDIS network for raw signal-bearing fault diagnosis, where the architecture of the BYOLDIS network is shown in [Fig. 5](#). The positive pairs  $\mathbf{H}$  and  $\mathbf{H}_p$  are input into  $f_\theta$ ,  $f_\xi$ , and fault feature projectors  $q_\theta$  and  $q_\xi$ , respectively. Subsequently, the online network's projection  $p_\theta$  is fed to the predictor of the online network to realize the prediction  $g_\theta(p_\theta)$ .

Reference [41] indicates that the projector of the original BYOL network can improve the representation extraction performance of a contrastive learning network. In this study, fault feature projectors are used to enhance the representation extraction performance, filter out the influence of noise, and extract the common features of the positive pairs  $\mathbf{H}$  and  $\mathbf{H}_p$ .

The coordinate encoder  $f_\theta$  and  $f_\xi$  are essential to map the delay co-

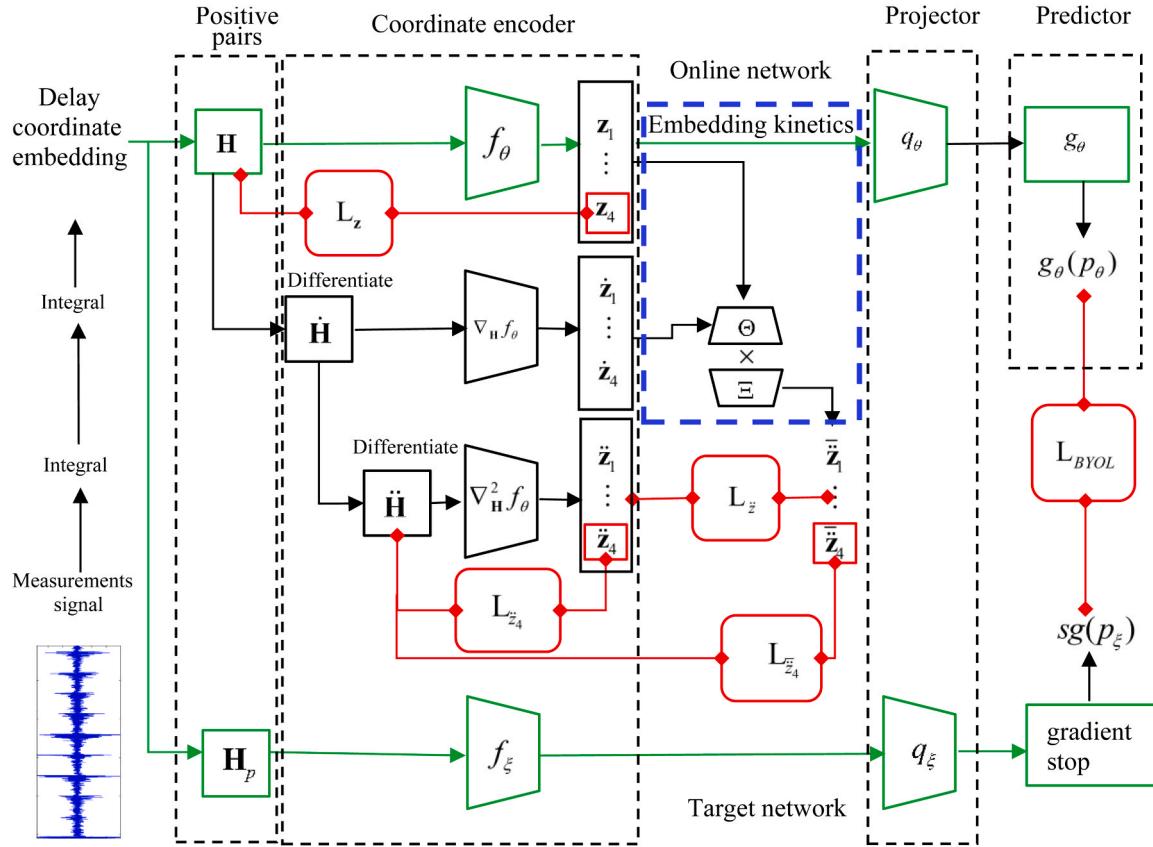


Fig. 5. Framework of the BYOLDIS network.

ordinate embedding positive pairs into the original dynamics DOF. According to Takens' theorem, a diffeomorphic map between the delay-embedded coordinate  $\mathbf{H}$  and the original kinematic system DOF  $\mathbf{z}$  exists. However, this diffeomorphic map is hard to determine. Therefore, in this study, we utilize a neural network to approximate the diffeomorphic map function to discover the original dynamics DOF in the fault-bearing dynamic differential equation because the neural network has universal approximation properties.

The coordinate encoders  $f_\theta$  and  $f_\xi$  consist of six fully connected layers, the density of the first fully connected layer is 128, and the density is halved layer by layer until it is equal to the DOF of the dynamic system. In this study, the density of the final layer is four because the fault-bearing simulation model is a 4-DOF model. The projector and predictor also consist of a multi-layer perceptron (MLP). The details of the parameter settings of BYOLDIS are shown in Table 1.

To automatically realize few-shot sample fault diagnosis, prior knowledge constraints  $\Theta$  are required. The equation library  $\Theta(\mathbf{z}, \dot{\mathbf{z}})$  is built using the fault-bearing signal generation simulation model, which contains the inner race, outer race, and three common roller fault types as candidate terms. The sparse matrix  $\Xi$  determines the candidate terms of the equation library  $\Theta$  that are active in the fault-bearing dynamic

model. The loss function of the BYOLDIS model consists of 5 loss terms ( $\mathcal{L}_{BYOL}, \mathcal{L}_{\bar{z}}, \mathcal{L}_{\dot{\bar{z}}_4}, \mathcal{L}_{\ddot{\bar{z}}_4}, \mathcal{L}_{\bar{z}_4}$ ).

The first loss  $\mathcal{L}_{BYOL}$  guides the network to learn the invariant information of the positive pairs  $\mathbf{H}$  and  $\mathbf{H}_p$ , which can filter the noise and construct a robust coordinate transfer. The following equation gives the loss term.

$$\mathcal{L}_{BYOL} = \left\| g_\theta(p_\theta) / \|g_\theta(p_\theta)\|_2 - sg(p_\xi) / \|sg(p_\xi)\|_2 \right\|_2^2 \quad (10)$$

The second loss term  $\mathcal{L}_{\bar{z}}$  can offer prior constraints to the coordinate encoder, usually built automatically from training with large amounts of data in the current neural network. Therefore, this loss term significantly reduces the dependence of the BYOLDIS network model on the data label,  $\bar{\mathbf{z}}$  and  $\dot{\mathbf{z}}$  are normalized in Eq.(11):

$$\mathcal{L}_{\bar{z}} = \sum_{i=1}^4 \left\| (\bar{\mathbf{z}}_i - \text{mean}(\bar{\mathbf{z}}_i)) / \text{std}(\bar{\mathbf{z}}_i) - (\ddot{\mathbf{z}}_i - \text{mean}(\ddot{\mathbf{z}}_i)) / \text{std}(\ddot{\mathbf{z}}_i) \right\|_2^2 \quad (11)$$

The third loss term  $\mathcal{L}_{\dot{\bar{z}}_4}$  is used to discover the accurate dynamic system for automatic fault diagnosis:

$$\mathcal{L}_{\dot{\bar{z}}_4} = \|\ddot{\bar{\mathbf{z}}}_4 - \dot{\bar{\mathbf{z}}}_4\|_2^2 \quad (12)$$

The last two-loss terms  $\mathcal{L}_{\ddot{\bar{z}}_4}$  and  $\mathcal{L}_{\bar{z}_4}$  let the 4th component of the coordinate encoder  $\mathbf{z}, \dot{\mathbf{z}}$  to be an exact reconstruction of the measurement signal  $\mathbf{x}, \dot{\mathbf{x}}$ . Those loss terms can guarantee that the dynamics of the coordinate (DOF of the system) are directly relevant to the measured quantity  $\mathbf{x}$ :

$$\mathcal{L}_{\ddot{\bar{z}}_4} = \|\ddot{\bar{\mathbf{z}}}_4 - \ddot{\mathbf{z}}_4\|_2^2 \quad (13)$$

$$\mathcal{L}_{\bar{z}_4} = \|\bar{\mathbf{z}}_4 - \bar{\mathbf{z}}_4\|_2^2 \quad (14)$$

**Table 1**  
Parameters setting of BYOLDIS.

Part	Layer	Number of layer	Activation function	Dense
Coordinate encoder	Fully connected	6	Elu (Except for the last layer)	[128,64,32, 16,8,4]
Projector	Fully connected	3	Elu	[512,128,64]
Predictor	Fully connected	2	Elu	[64,128]

where  $\ddot{\mathbf{w}}$  &  $\dot{\mathbf{w}}$  are the first row of the  $\ddot{\mathbf{H}}$  &  $\mathbf{H}$ , respectively.

Finally, all loss terms are combined:

$$\mathcal{L} = \mathcal{L}_{BYOL} + \lambda_1 \mathcal{L}_{\dot{\mathbf{z}}} + \lambda_2 \mathcal{L}_{\ddot{\mathbf{z}}_4} + \lambda_3 \mathcal{L}_{\ddot{\mathbf{z}}_4} + \lambda_4 \mathcal{L}_{\mathbf{z}_4} \quad (15)$$

The weight coefficient matrix  $\lambda = [\lambda_1, \lambda_2, \lambda_3, \lambda_4]$  is the hypothesis that includes the empirical parameters. The  $\lambda_2$  &  $\lambda_3$  should fulfill the requirement that the  $\lambda_2 \mathcal{L}_{\ddot{\mathbf{z}}_4}$  &  $\lambda_3 \mathcal{L}_{\ddot{\mathbf{z}}_4}$  in Eq.(15) should have two orders higher than the other two loss terms. The reason for making this requirement is because the  $\lambda_2 \mathcal{L}_{\ddot{\mathbf{z}}_4}$  &  $\lambda_3 \mathcal{L}_{\ddot{\mathbf{z}}_4}$  loss term directly imposing strong constraints on the reconstructed signal. As mentioned above, these two items can guarantee that the dynamics of the coordinate (DOF of the system) are directly relevant to the measured quantity  $\mathbf{x}$ . Although the loss terms  $\mathcal{L}_{\ddot{\mathbf{z}}_4}$  and  $\mathcal{L}_{\mathbf{z}_4}$  can offer the constraint information to the BYOLDIS network, the strong constraints which related to  $\mathcal{L}_{\ddot{\mathbf{z}}_4}$  and  $\mathcal{L}_{\mathbf{z}_4}$  are the key part of signal reconstruction and faults diagnosis. Otherwise, the high training weight of  $\mathcal{L}_{\ddot{\mathbf{z}}_4}$  and  $\mathcal{L}_{\mathbf{z}_4}$  can enable the network to achieve rapid convergence of input signals in the early training stage, ensuring the stability of training.

### 3.3. Coordinate derivatives computing

To compute the differentiated  $\dot{\mathbf{z}}$  and  $\ddot{\mathbf{z}}$  of the coordinate  $\mathbf{z}$  in the coordinate encoder forward propagation process, derivatives of the coordinate encoder variables ( $f_\theta$ ) are required when propagating the  $\mathbf{H}$  derivatives ( $\nabla_{\mathbf{H}} f_\theta, \nabla_{\mathbf{H}}^2 f_\theta$ ). Given the input  $\mathbf{H}$ , the forward propagation of the final layer of the coordinate encoder can be calculated as follows:

$$\mathbf{z} = f(\mathbf{l}_{L-1}) \mathbf{W}_L + \mathbf{b}_L \quad (16)$$

where  $L$  represents the number of coordinated layers,  $f(\cdot)$  represents the Elu activation function, and  $\mathbf{l}_{L-1}$  is the  $L-1$  layer's output. Then, the other layer's forward propagation in the coordinate encoder can be expressed the same as Eq.(16). For example, the forward propagation of the  $j$ th layer is  $\mathbf{l}_j = f(\mathbf{l}_{j-1}) \mathbf{W}_j + \mathbf{b}_j$ , and the first layer is  $\mathbf{l}_1 = \mathbf{H} \mathbf{W}_1 + \mathbf{b}_1$ . According to the chain rule of deviation, the first-order derivation of  $\mathbf{z}$  can be calculated as follows:

$$\begin{aligned} \frac{d\mathbf{z}}{dt} &= \left( f'(\mathbf{l}_{L-1}) \circ \frac{d\mathbf{l}_{L-1}}{dt} \right) \mathbf{W}_L \\ \frac{d\mathbf{l}_j}{dt} &= \left( f'(\mathbf{l}_{j-1}) \circ \frac{d\mathbf{l}_{j-1}}{dt} \right) \mathbf{W}_j \\ \frac{d\mathbf{l}_0}{dt} &= \frac{d\mathbf{H}}{dt} \mathbf{W}_0 = \dot{\mathbf{H}} \mathbf{W}_0 \end{aligned} \quad (17)$$

And the second-order derivation of  $\mathbf{z}$  can be calculated by

$$\begin{aligned} \frac{d^2\mathbf{z}}{dt^2} &= \left( f''(\mathbf{l}_{L-1}) \circ \frac{d\mathbf{l}_{L-1}}{dt} \circ \frac{d\mathbf{l}_{L-1}}{dt} + f'(\mathbf{l}_{L-1}) \circ \frac{d^2\mathbf{l}_{L-1}}{dt^2} \right) \mathbf{W}_L \\ \frac{d^2\mathbf{l}_j}{dt^2} &= \left( f''(\mathbf{l}_{j-1}) \circ \frac{d\mathbf{l}_{j-1}}{dt} \circ \frac{d\mathbf{l}_{j-1}}{dt} + f'(\mathbf{l}_{j-1}) \circ \frac{d^2\mathbf{l}_{j-1}}{dt^2} \right) \mathbf{W}_j \\ \frac{d^2\mathbf{l}_0}{dt^2} &= \frac{d^2\mathbf{H}}{dt^2} \mathbf{W}_0 = \ddot{\mathbf{H}} \mathbf{W}_0 \end{aligned} \quad (18)$$

For the activation function Elu, the first order and second derivations are as follows:

$$f'(\mathbf{l}) = \min(\mathbf{e}^{\mathbf{l}}, 1) \quad (19)$$

$$f''(\mathbf{l}) = \mathbf{e}^{\mathbf{l}} \circ \varepsilon(\mathbf{l}) \quad (20)$$

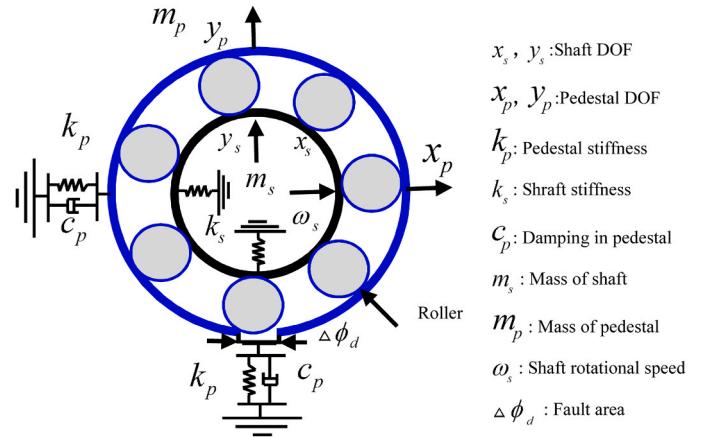


Fig. 6. Four-DOF fault signal generation simulating model.

where  $\varepsilon(\cdot)$  is the Heaviside function and  $\circ$  denotes Hadamard product.

### 3.4. Fault bearing equation library of BYOLDIS

A four-DOF fault-bearing signal generation simulation model is utilized to build the equation library  $\Theta$ . The model is summarized in Fig. 6. Four DOF ( $x_s, y_s, x_p$ , and  $y_p$ ) are related to the movement of the shaft and pedestal. The following equation governs the bearing system.

$$\begin{cases} \ddot{x}_s = -(k_s x_s + f_x((x_s - x_p), \omega_s, t)) / m_s \\ \ddot{y}_s = -(k_s y_s + f_y((y_s - y_p), \omega_s, t)) / m_s \\ \ddot{x}_p = -(c_p \dot{x}_p + k_p x_p - f_x((x_s - x_p), \omega_s, t)) / m_p \\ \ddot{y}_p = -(c_p \dot{y}_p + k_p y_p - f_y((y_s - y_p), \omega_s, t)) / m_p \end{cases} \quad (21)$$

In Eq.(21), the equation  $f_x(\cdot, \omega_s, dt)$  is used to describe the sum of the contact forces of each roller at the contact position, which can be written as follows based on the Hertzian contact relationship:

$$f_x = k_b \sum_{j=1}^{n_b} \text{relu}(\delta_j^{1.5}) \cos \phi_j \quad (22)$$

$$f_y = k_b \sum_{j=1}^{n_b} \text{relu}(\delta_j^{1.5}) \sin \phi_j$$

where  $k_b$  is the load-deflection factor, which depends on the contact geometry and the elastic contacts of the material, and  $\phi_j$  represents the angular positions of the  $j$ th rollers,

$$\phi_j = \frac{2\pi(j-1)}{n_b} + \omega_c t + \phi_0 \quad (23)$$

which depends on the cage speed  $\omega_c$  and the cage initial position  $\phi_0$  (a trainable value).  $n_b$  is the rollers' number. The relationship between the cage speed  $\omega_c$  and shaft speed  $\omega_s$  is given:

$$\omega_c = \frac{(D_p - D_b)\omega_s}{2D_p} \quad (24)$$

where  $D_p$  and  $D_b$  are the pitch and roller diameters of the bearing, respectively. Then, the contact interference of each roller is:

$$\delta_j = (x_s - x_p) \cos \phi_j + (y_s - y_p) \sin \phi_j - c_{f-d} \quad (25)$$

where  $c_{f-d}$  represents the spall depth caused by a fault.  $c_{f-d}$  determines the corresponding vibration signal forms of the different fault types. We then discuss the  $c_{f-d}$  construction methods for the three fault types.  $c_{f-d}$  has an approximate expression of the step function, which can be used to create a learnable equation library.

### 3.4.1. Outer race spall

A spall of a depth ( $c_{f-d}$ ) over an angular distance of ( $\Delta\phi_d$ ) is modeled to simulate the contact loss at a defined angular position  $\phi_{outer}$ :

$$c_{f-d-outer} = \left( -1 - \frac{1}{1 - e^{\frac{100(\phi_j - \phi_{outer} - 2\pi + \phi_d/20)}{\Delta\phi_d}}} + \frac{1}{1 - e^{\frac{100(\phi_j - \phi_{outer} + 19*\phi_d/20)}{\Delta\phi_d}}} \right) \times 10^{-5} \quad (26)$$

where  $\phi_{outer}$  is the outer-race angular failure location, which is in the direction of the external force, and  $\Delta\phi_d$  is the angular fault area.

### 3.4.2. Inner race spall

Different from the outer race spall is fixed in location, an inner race spall rotates at the same speed as the rotor, then,  $c_{f-d-inner}$  as followed:

$$c_{f-d-inner} = \left( -1 - \frac{1}{1 - e^{\frac{100(\phi_j - \phi_s - \phi_{inner} - 2\pi + \Delta\phi_d/20)}{\Delta\phi_d}}} + \frac{1}{1 - e^{\frac{100(\phi_j - \phi_s - \phi_{inner} + 19*\Delta\phi_d/20)}{\Delta\phi_d}}} \right) \times 10^{-5} \quad (27)$$

where  $\phi_{outer}$  is the inner race angular failure location because the inner race location is unknown.  $\phi_{outer}$  is a trainable variable and  $\phi_s$  is the angular position of the bearing's shaft (inner race).

### 3.4.3. Roller spall

The position of the spall is associated with the speed of the shaft speed. This means that the rolling element faults share some characteristics with faults on the inner race, which differs from the faults associated with the outer race (fixed location).

The loss of contact is detected for only the faulty rolling element  $k$ , and the contact loss appears twice for each complete rotation of that rolling element. And the curvature between the inner race and outer race is different. Therefore, the twice contact loss and contact periods are also different. The inner race will contact deeper and longer compared to the outer race. Therefore, the contact loss  $c_{f-d-roller}$  can be described as follows:

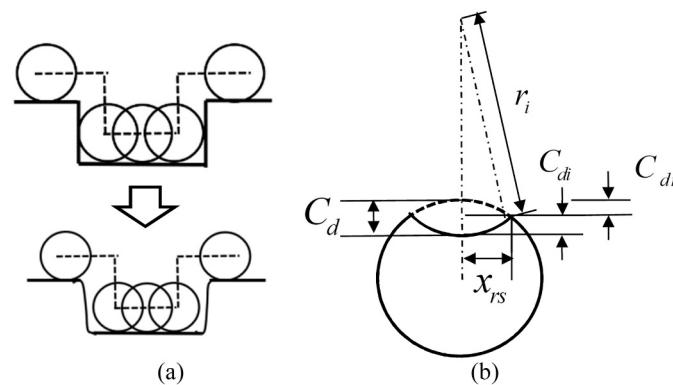


Fig. 7.  $c_{f-d}$  approximate construction and fault-roller geometric attribute.

$$c_{f-d-roller} = \left[ \left( -1 - \frac{1}{1 - e^{\frac{100(\phi_{spin} - 2\pi + \Delta\phi_{d-outer}/20)}{\Delta\phi_{d-outer}}}} + \frac{1}{1 - e^{\frac{100(\phi_{spin} + 19*\phi_{d-outer}/20)}{\Delta\phi_{d-outer}}}} \right) \times \left( -1 - \frac{1}{1 - e^{\frac{100(\phi_{spin} - 2\pi + \Delta\phi_{d-inner}/20)}{\Delta\phi_{d-inner}}}} + \frac{1}{1 - e^{\frac{100(\phi_{spin} + 19*\phi_{d-inner}/20)}{\Delta\phi_{d-inner}}}} \right) \right] \times 10^{-5} \quad (28)$$

where  $C_{dr} = D_b/2 - \sqrt{(D_b/2)^2 - x_{rs}^2}$ ,  $C_{di} = r_i - \sqrt{r_i^2 - x_{rs}^2}$ ,  $r_i = (D_p - D_b)/2$ ,  $C_{do} = r_o - \sqrt{r_o^2 - x_{rs}^2}$ ,  $\Delta\phi_{d-outer} = 2x_{rs}/r_o$ ,  $\Delta\phi_{d-inner} = 2x_{rs}/r_i$ , and  $x_{rs}$  represent half of the roller spalling area (as Fig. 7(b) shows).  $\phi_{spin}$  is the ball spall spin location, governed by  $\phi_{spin} = \omega_{spin}t + \phi_{spin0}$ .  $\phi_{spin0}$  is the roller spin initial position (trainable value).  $\omega_{spin}$  followed:

$$\omega_{spin} = \omega_s D_p / 2D_b \left( 1 - (D_b/D_p \cos \alpha)^2 \right) \quad (29)$$

where  $\alpha$  is bearing contact angle, in this paper,  $\alpha = 0$ .

In this study, the candidate terms of the equation library  $\Theta(\mathbf{z}, \dot{\mathbf{z}})$  are calculated from the coordinate encoder output  $\mathbf{z}$  and its deviation  $\dot{\mathbf{z}}$ .  $\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4]$  corresponds with  $y_s, x_s, y_p, x_p$  respectively. To achieve an effective fault diagnosis, we calculate the fault-bearing dynamic force terms, including inner race, outer race, and rollers' fault types. The fault-bearing dynamic force terms are calculated by Eqs. 10–18, symbolized as  $f_{x-outer}, f_{y-outer}, f_{x-inner}, f_{y-inner}, f_{x-roller}, f_{y-roller}$ . Then, the equations library  $\Theta$  is a 15 columns matrix:

$$\Theta = [1, \mathbf{z}, \dot{\mathbf{z}}, f_{x-outer}, f_{y-outer}, f_{x-inner}, f_{y-inner}, f_{x-roller}, f_{y-roller}] \quad (30)$$

After building the equation library  $\Theta$ , the fault-bearing dynamic differential equation can be constructed by combining it with the sparse matrix  $\Xi$ :

$$\bar{\mathbf{z}} = \Theta \Xi \quad (31)$$

[Eq.\(31\)](#) is a simplified form of [Eq.\(21\)](#). The details of [Eq.\(31\)](#) is:

$$\bar{\mathbf{z}} = [\bar{x}_s, \bar{y}_s, \bar{x}_p, \bar{y}_p]^T$$

$$\Theta = \begin{bmatrix} 0 & -\frac{k_s}{m_s}0 & 0 & 0 & 0 & 0 & \frac{w_{inner}}{m_s}0 & \frac{w_{outer}}{m_s}0 & \frac{w_{roller}}{m_s}0 \\ 0 & 0 & -\frac{k_s}{m_s}0 & 0 & 0 & 0 & \frac{w_{inner}}{m_s}0 & \frac{w_{outer}}{m_s}0 & \frac{w_{roller}}{m_s} \\ 0 & 0 & 0 & -\frac{k_s}{m_p}0 & 0 & 0 & \frac{w_{inner}}{m_p}0 & \frac{w_{outer}}{m_p}0 & \frac{w_{roller}}{m_p} \\ 0 & 0 & 0 & 0 & -\frac{c_p}{m_p}0 & 0 & \frac{w_{inner}}{m_p}0 & \frac{w_{outer}}{m_p}0 & \frac{w_{roller}}{m_p} \\ 0 & 0 & 0 & 0 & 0 & -\frac{c_p}{m_p}0 & 0 & \frac{w_{inner}}{m_p}0 & \frac{w_{outer}}{m_p}0 & \frac{w_{roller}}{m_p} \end{bmatrix} \quad (32)$$

We transmit the term  $\frac{1}{m_s} \& \frac{1}{m_p}$  to  $\Theta$ , to simplify the sparse matrix  $\Xi$  in the actual operation process. Finally, we can obtain the diagnosis results according to which fault-type bearing-fault dynamic force term in the sparse matrix  $\Xi$ . The larger the weight value, the larger the possible fault type corresponding to the weight.

For diagnosis directly, the weights' sum of fault-bearing dynamic force terms is one via SoftMax function. Therefore, we can obtain the fault type of input signal via the maximum weight of the fault-bearing dynamic force term.

### 3.5. The overall structure and optimization stareragay

The process of BYOLDIS training and bearing fault diagnosis can be summarized in the following Algorithm 1.

**Algorithm 1.** BYOLDIS training and diagnosis.

---

**Input:** input vibration signal  $\hat{\mathbf{x}}$  and initial sparse matrix  $\Xi$   
**Output:** output sparse matrix  $\Xi$  and diagnosis result

- 1 Calculate the detrended integral sequence  $\hat{\mathbf{x}}$  and  $\mathbf{x}$ ;
- 2 Determine the delay time  $\tau$ ;
- 3 Build the delay time embedding positive pairs (online network  $\mathbf{H}, \dot{\mathbf{H}}, \ddot{\mathbf{H}}$ ),(target network  $\mathbf{H}_p$ ) ;
- 4 Feed  $\mathbf{H}, \dot{\mathbf{H}}$ , and  $\ddot{\mathbf{H}}$  to online network, and feed  $\mathbf{H}_p$  into the target network ;
- 5 **for** iteration = 1,2,...,max iteration **do**
- 6   //Online network calculation;
- 7   Coordinate encoder calculates  $\mathbf{z}$ ,  $\dot{\mathbf{z}}$  and  $\ddot{\mathbf{z}}$  from  $\mathbf{H}, \dot{\mathbf{H}}, \ddot{\mathbf{H}}$ ;
- 8   Build kinetic equation library  $\Theta$  from  $\mathbf{z}$ ,  $\dot{\mathbf{z}}$ ;
- 9   Construct the second derivative  $\bar{\mathbf{z}}$  by  $\bar{\mathbf{z}} = \Theta\Xi$ ;
- 10   Projector and predictor calculate  $g_\theta(p_\theta)$  from  $\mathbf{z}$ ;
- 11   //Target network calculation;
- 12   Calculate  $sg(p_\xi)$  from  $\mathbf{H}_p$ ;
- 13   //Loss calculation;
- 14   Calculates each loss terms  $\mathcal{L}_{BYOL}$  ,  $\mathcal{L}_{\bar{\mathbf{z}}}$ , $\mathcal{L}_{\bar{\mathbf{z}}_4}$  and  $\mathcal{L}_{\mathbf{z}_4}$  ;
- 15   Obtain loss function  $\mathcal{L} = \mathcal{L}_{BYOL} + \lambda_1\mathcal{L}_{\bar{\mathbf{z}}} + \lambda_2\mathcal{L}_{\bar{\mathbf{z}}_4} + \lambda_3\mathcal{L}_{\bar{\mathbf{z}}_4} + \lambda_4\mathcal{L}_{\mathbf{z}_4}$ ;
- 16   //Network parameters updating;
- 17   Update the weight of the online network  $\theta$  and  $\Theta$  via backpropagation;
- 18   Utilizing the weights from the online network  $\theta$  to update the weights of the target network  $\xi$ by moving averaging;
- 19 **end**
- 20 Obtain the largest weight-fault-type bearing-fault dynamic force term in the sparse matrix  $\Xi$  ;
- 21 Obtain the diagnosis result from sparse matrix  $\Xi$ ;

---

#### 4. Simulation and experimental study

##### 4.1. Simulated and experimented with fault-bearing signal

Three fault-simulating signals (outer race, inner race, and roller faults) are utilized to verify the effect of the proposed BYOLDIS. The signal is generated using a five-DOF fault-bearing system [42,43]. Fig. 8 shows the five-DOF fault-bearing system. This five-DOF model adds a mass-spring system to simulate a typical high-frequency resonant response of the bearing (stiffness and damping selected to excite a 15 kHz frequency with 5% damping) and adds 3000 N external forces  $F$  in the shaft to simulate the loaded condition. The following equation governs this five-DOF fault-bearing system:

$$\begin{cases} m_s \ddot{x}_s + k_s(x_s) + f_x((x_s - x_p), \omega_s, dt) - F = 0 \\ m_s (\ddot{y}_s + 9.8) + k_s(y_s) + f_y((y_s - y_p), \omega_s, dt) = 0 \\ m_p \ddot{x}_p + (c_p + c_r) \dot{x}_p + (k_p + k_r)(x_p) - k_r x_b - f_x((x_s - x_p), \omega_s, dt) = 0 \\ m_p \ddot{y}_p + c_p \dot{y}_p + k_p(y_p) - f_y((y_s - y_p), \omega_s, dt) = 0 \\ m_r \ddot{x}_b + k_r(x_b - x_p) + c_r(\dot{x}_b - \dot{x}_p) = 0 \end{cases} \quad (33)$$

The detailed parameters of the simulation model are presented in Table 2. It should be noted that the vibration signal is collected for the output of the mass-spring system ( $\ddot{x}_b$ ).  $\ddot{x}_b$  which contains noise from high-frequency resonance, which is similar to the real fault-bearing vibration signal.  $dt$ denotes the simulated signal's sampling time. The vibration signal has a sampling rate of 25600 Hz and a shaft speed  $\omega_s = 1200\text{rpm}$ .

Besides, the experiment signal is supported by Case Western Reserve University [44]. The sampling frequency is 10000 Hz, the speed is 1797 rpm, 0 load, the bearing is SKF 6205, and the signal is collected from the drive end bearing.

##### 4.2. Implementation details of the experiment

The BYOLDIS network is implemented based on the TensorFlow framework. During the training network process, we utilize the same

**Table 2**

Parameter detail of the simulating model.

Statement of unit: $m(\text{kg}), k(\text{N/m}), c(\text{Ns/m}), \phi(\text{rad}), x(\text{m}), D(\text{m})$					
$m_p$	12.64	$k_p$	$1.51 \times 10^7$	$c_p$	$2.21 \times 10^3$
$m_r$	1	$k_r$	$8.88 \times 10^9$	$c_r$	$9.42 \times 10^3$
$m_s$	0.51	$k_s$	$4.24 \times 10^6$	$n_b$	9
$k_b$	$1.89 \times 10^{10}$	$\Delta\phi_d$	$1.92 \times 10^{-2}$	$x_{rs}$	.
$D_b$	$7.94 \times 10^{-3}$	$D_p$	$3.90 \times 10^{-2}$	$dt$	$3.91 \times 10^{-5}$

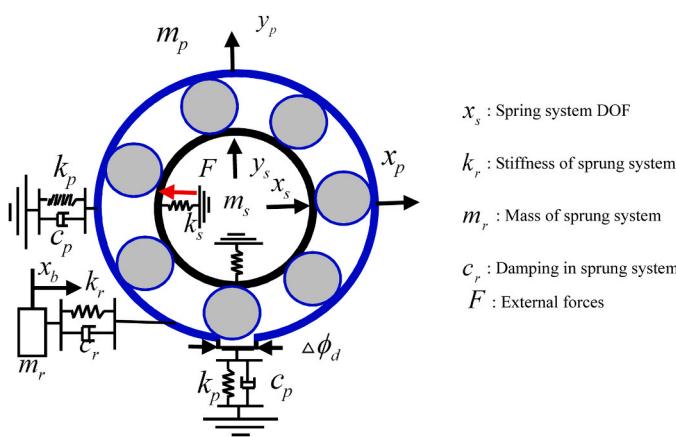


Fig. 8. Five-DOF fault signal generation simulating model.

value hyperparameters for all experiments described below.  $\tau$  is set using the cosine rate change equation in [36], as shown in Eq. (4).  $\varepsilon_{base}$  is set to 0.995.  $k$  denotes the current epoch and  $K$  denotes the maximum number of epoch.

$$\varepsilon = 1 - \frac{(1 - \varepsilon_{base}) \times \left( \cos\left(\frac{\pi k}{K}\right) + 1 \right)}{2} \quad (34)$$

The learning rate  $\eta$  of training includes the warm-up cosine decay schedule. The base learning rate  $\eta_{base}$  is set to 0.1 for the simulated signal and set to 0.5 for the collected signal, and the warm-up iteration period is set to 100. The learning rate decay schedule formula is shown in Eq. (35), which  $i$  denotes the iteration turn,  $K$  denotes the maximum number of iterations:

$$\eta = \begin{cases} \frac{\eta_{base}}{P} \times i, & i = 1, 2, \dots, P \\ \frac{\eta_{base} \times \left(1 + \cos\left(\frac{i\pi}{K}\right)\right)}{2}, & i = P + 1, P + 2, \dots, K \end{cases} \quad (35)$$

$$\begin{cases} \ddot{x}_s + 1 \times 10^6 x_s + 0.50f_{x-inner} + 0.25f_{x-outer} + 0.25f_{x-roller} - 3.07 \times 10^3 = 0 \\ \ddot{y}_s + 9.99 \times 10^5 y_s + 0.50f_{y-inner} + 0.25f_{y-outer} + 0.25f_{y-roller} + 8.86 \times 10 = 0 \\ \ddot{x}_p + 5.35 \times 10^3 \dot{x}_p + 1 \times 10^6 x_p - 0.05f_{x-inner} - 0.03f_{x-outer} - 0.03f_{x-roller} + 1.60 \times 10 = 0 \\ \ddot{y}_p - 1.30 \times 10^2 \dot{y}_p + 9.97 \times 10^5 y_p - 0.05f_{y-inner} - 0.03f_{y-outer} - 0.03f_{y-roller} + 1.20 \times 10^2 = 0 \end{cases} \quad (38)$$

The iteration number is set to 1000 and we are collecting signals at  $q_T = 0.5s$ . Eq.(22) is the initial value of the sparse matrix  $\Xi$ .  $F$  is the external forces for the simulated signal  $F = 3000, \lambda = [1, 10, 10, 1]$  while for the experiment signal  $F = 0, \lambda = [1, 100, 100, 1]$ . And the fault area is set to 1 mm,  $k_b$  which is set  $1.8 \times 10^{10}$  when calculating the contact forces.

$$\begin{bmatrix} F & -10^6 0 & 0 & 0 & 000 & 0 & 0.330 & 0.330 & 0.330 \\ 9.80 & -10^6 0 & 0 & 0 & 000 & 0 & 0 & 0.330 & 0.330 & 0.33 \\ 0 & 0 & 0 & -10^6 0 & 00 & -10^2 0 & 0.030 & 0.030 & 0.030 & 0.03 \\ 0 & 0 & 0 & 0 & -10^6 0 & 000 & -10^2 0 & 0.030 & 0.030 & 0.03 \end{bmatrix} \quad (36)$$

Considering that the exact values of the modulus of elasticity and other parameters are unknown in the real situation, the initial values of the sparse matrix  $\Xi$  are only accurate to the order of magnitude. BYOLDIS aims to realize fault diagnosis rather than accurate dynamic model discovery. The equation library  $\Theta$  is based on the four-DOF system, which is more conducive to the gradient update of the BYOLDIS than the five-DOF system library.

#### 4.3. Simulated signal diagnosis results and analysis

$$\begin{cases} \ddot{x}_s + 1 \times 10^6 x_s + 0.27f_{x-inner} + 0.43f_{x-outer} + 0.29f_{x-roller} - 3.03 \times 10^3 = 0 \\ \ddot{y}_s + 1 \times 10^6 y_s + 0.27f_{y-inner} + 0.43f_{y-outer} + 0.29f_{y-roller} - 1.47 \times 10 = 0 \\ \ddot{x}_p - 4.35 \times 10^3 \dot{x}_p + 9.96 \times 10^5 x_p - 0.03f_{x-inner} - 0.04f_{x-outer} - 0.03f_{x-roller} + 2.23 \times 10^2 = 0 \\ \ddot{y}_p - 4.82 \times 10^2 \dot{y}_p + 1 \times 10^6 y_p - 0.03f_{y-inner} - 0.04f_{y-outer} - 0.03f_{y-roller} + 3.66 \times 10 = 0 \end{cases} \quad (39)$$

In this Section, we demonstrate the ability of the proposed BYOLDIS network to automatic fault diagnosis, and we apply the algorithm to the following systems: the inner race fault-bearing system, the outer race fault-bearing system, and the roller fault-bearing system. These systems'

information follows the description of Section 4.1.  $\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4]$  corresponds to DOFs,  $x_s, y_s, x_p, y_p$  in the library  $\Theta$ , respectively. The second-order deviation  $\ddot{z}_4$  is from  $\ddot{x}_b$  in Eq.(33).

##### 4.3.1. Inner race fault-bearing diagnosis

The inner-race fault-bearing signal generated system of the equations  $\ddot{x}_b$  is given by substituting Table 2 into Eq.(37):

$$\begin{cases} \ddot{x}_s + 8.31 \times 10^6 x_s + 1.96f_{x-inner} - 5.88 \times 10^3 = 0 \\ \ddot{y}_s + 8.31 \times 10^6 y_s + 1.96f_{y-inner} - 19.21 = 0 \\ \ddot{x}_p + 9.18 \times 10^2 \dot{x}_p + 7.04 \times 10^8 x_p - 7.03 \times 10^8 x_b - 0.08f_{x-inner} = 0 \\ \ddot{y}_p + 1.75 \times 10^2 \dot{y}_p + 1.19 \times 10^6 y_p - 0.08f_{y-inner} = 0 \\ \ddot{x}_b + 8.88 \times 10^9 (x_b - x_p) + 9.42 \times 10^3 (\dot{x}_b - \dot{x}_p) = 0 \end{cases} \quad (37)$$

After BYOLDIS model training, the reconstructed system for the input vibration signal  $\ddot{x}$  is given by the sparse output matrix  $\Xi$ , the equation form of  $\Xi$  is:

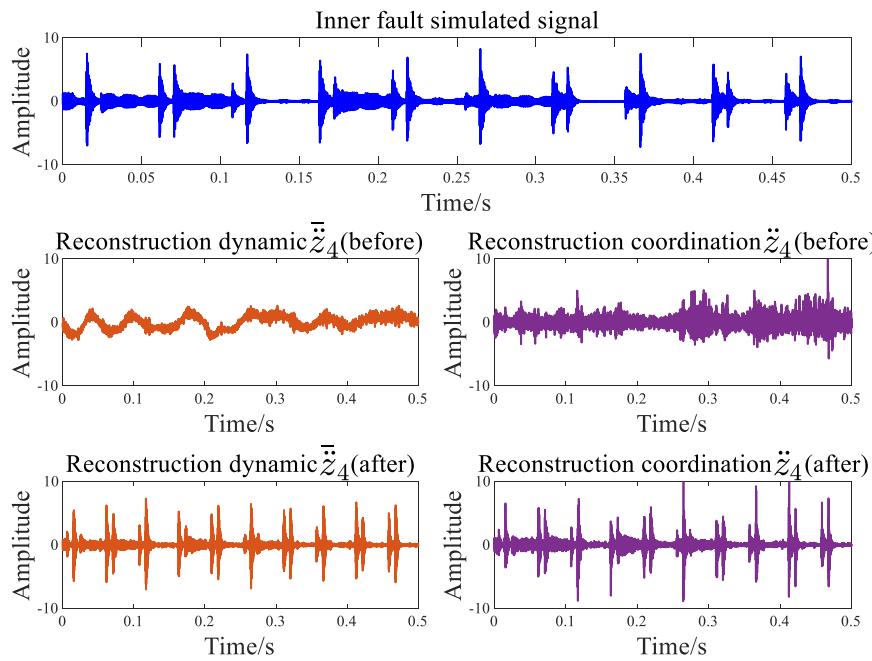
As shown in Eq. (38), the inner-race fault force term obtains the maximum weight in the reconstructed system. According to the reconstructed system, we can confirm that the diagnosis result is an inner-race fault for the vibration signal  $\ddot{x}$ . However, the reconstructed system's parameter details differ from the inner race fault-bearing signal-generated system. This shows that BYOLDIS cannot achieve accurate system reconstruction but is sufficient for bearing fault diagnosis.

Fig. 9 shows that input  $\ddot{x}$  is nearly perfectly reconstructed at the output of the coordination encoder's last dimension  $\ddot{z}_4$  and also can be nearly perfectly reconstructed by the last dimension of  $\bar{\mathbf{z}}$ , which is calculated via  $\bar{\mathbf{z}} = \Theta \Xi$ . It shows that the proposal can effectively capture the crucial dynamic system information of the inner race fault-bearing vibration signal, so the diagnosis result is reliable and interpretable.

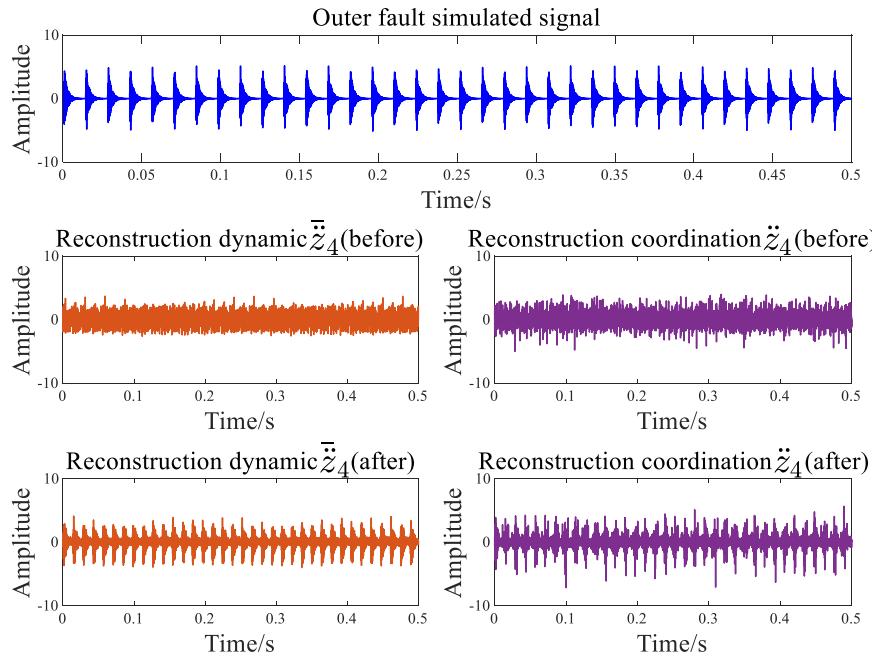
##### Outer race fault-bearing diagnosis

The outer-race fault-bearing signal generated system of the equations for  $\ddot{x}_b$  is given by Eq.(37), the dynamic force terms are  $f_{x-outer}$  and  $f_{y-outer}$ . After BYOLDIS model training, the reconstructed system for the input vibration signal  $\ddot{x}$  from the  $\ddot{x}_b$  is given by the sparse output matrix  $\Xi$ , the equation form of  $\Xi$  is:

As shown in Eq. (39), the outer race fault force term obtains the maximum weight in the reconstructed system; therefore, we can confirm that the diagnosis result is the outer race fault. Fig. 10 shows that BYOLDIS can also effectively capture the crucial dynamic system



**Fig. 9.** The reconstruction signal and the before->after training comparison of BYOLDIS from inner race fault-bearing vibration signal.



**Fig. 10.** The reconstruction signal and the before->after training comparison of BYOLDIS from outer race fault-bearing vibration signal.

information of the outer-race fault-bearing vibration signal.

#### 4.3.2. Roller fault-bearing diagnosis

The roller fault-bearing signal generated system of the equations for  $\ddot{x}_b$  is given by Eq.(31). The dynamic force are  $f_{x-roller}$  and  $f_{y-roller}$ . After

BYOLDIS model training, the reconstructed system for the input vibration signal  $\ddot{x}$  from the  $\ddot{x}_b$  is given by the sparse output matrix  $\Xi$ , the equation form of  $\Xi$  is:

$$\begin{cases} \ddot{x}_s + 1 \times 10^6 x_s + 0.28f_{x-inner} + 0.33f_{x-outer} + 0.39f_{x-roller} - 2.96 \times 10^3 = 0 \\ \ddot{y}_s + 9.99 \times 10^5 y_s + 0.28f_{y-inner} + 0.33f_{y-outer} + 0.39f_{y-roller} + 3.60 \times 10 = 0 \\ \ddot{x}_p - 5.31 \times 10^3 \dot{x}_p + 1 \times 10^6 x_p - 0.03f_{x-inner} - 0.03f_{x-outer} - 0.04f_{x-roller} - 2.79 \times 10^2 = 0 \\ \ddot{y}_p - 3.55 \times 10^2 \dot{y}_p + 9.99 \times 10^5 y_p - 0.03f_{y-inner} - 0.03f_{y-outer} - 0.04f_{y-roller} + 1.35 \times 10^2 = 0 \end{cases} \quad (40)$$

As shown in Eq.(40), the roller fault force term has the max weight, around 0.39 and 0.04, so the proposed method classified it as the roller fault. Fig. 11 shows that BYOLDIS can also effectively capture the crucial dynamic system information of the roller fault-bearing vibration signal.

#### 4.4. Experiment signal diagnosis results

Diagnosis results of the proposal are shown from the experiment signal: inner race fault, outer race fault, and roller fault.

##### 4.4.1. Inner race fault-bearing diagnosis

After BYOLDIS model training, the reconstructed system for the input vibration signal  $\bar{x}$  is given by the sparse output matrix  $\Xi$ ; the equation form  $\Xi$  is:

$$\begin{cases} \ddot{x}_s + 1 \times 10^6 x_s + 0.56 f_{x-\text{inner}} + 0.13 f_{x-\text{outer}} + 0.31 f_{x-\text{roller}} + 2.76 = 0 \\ \ddot{y}_s + 9.77 \times 10^5 y_s + 0.56 f_{y-\text{inner}} + 0.13 f_{y-\text{outer}} + 0.31 f_{y-\text{roller}} + 7.05 \times 10 = 0 \\ \ddot{x}_p + 6.87 \dot{x}_p + 9.87 \times 10^7 x_p - 0.06 f_{x-\text{inner}} - 0.01 f_{x-\text{outer}} - 0.03 f_{x-\text{roller}} + 1.31 \times 10 = 0 \\ \ddot{y}_p + 64.5 \dot{y}_p + 9.84 \times 10^5 y_p - 0.06 f_{y-\text{inner}} - 0.01 f_{y-\text{outer}} - 0.03 f_{y-\text{roller}} - 1.61 \times 10^2 = 0 \end{cases} \quad (41)$$

As shown in Eq.(41), the inner-race fault force term obtains the maximum weight in the reconstructed system. Fig. 12 shows that BYOLDIS can deal with the real vibration signal.

##### 4.4.2. Outer race fault-bearing diagnosis

After BYOLDIS model training, the reconstructed system for the input vibration signal  $\bar{x}$  from the  $\dot{x}_b$  is given by the output sparse matrix  $\Xi$ , the equation form of  $\Xi$  is:

$$\begin{cases} \ddot{x}_s + 9.92 \times 10^5 x_s + 0.13 f_{x-\text{inner}} + 0.63 f_{x-\text{outer}} + 0.25 f_{x-\text{roller}} + 7.21 = 0 \\ \ddot{y}_s + 9.80 \times 10^7 y_s + 0.13 f_{y-\text{inner}} + 0.63 f_{y-\text{outer}} + 0.25 f_{y-\text{roller}} + 7.78 \times 10 = 0 \\ \ddot{x}_p + 1.06 \times 10^3 \dot{x}_p + 1.01 \times 10^6 x_p - 0.01 f_{x-\text{inner}} - 0.06 f_{x-\text{outer}} - 0.03 f_{x-\text{roller}} + 2.73 \times 10 = 0 \\ \ddot{y}_p + 66.7 \dot{y}_p + 9.96 \times 10^5 y_p - 0.01 f_{y-\text{inner}} - 0.06 f_{y-\text{outer}} - 0.03 f_{y-\text{roller}} + 6.94 \times 10^2 = 0 \end{cases} \quad (42)$$

As shown in Eq.(42) & Fig. 13, the roller fault force term obtains the maximum weight in the reconstructed system; therefore, the diagnosis result is outer race fault.

##### 4.4.3. Roller fault-bearing diagnosis

After BYOLDIS model training, the reconstructed system for the input vibration signal  $\bar{x}$  from the  $\dot{x}_b$  is given by the output sparse matrix  $\Xi$ , the equation form of  $\Xi$  is:

$$\begin{cases} \ddot{x}_s + 1.01 \times 10^6 x_s + 0.25 f_{x-\text{inner}} + 0.14 f_{x-\text{outer}} + 0.61 f_{x-\text{roller}} + 3.55 \times 10 = 0 \\ \ddot{y}_s + 1.01 \times 10^6 y_s + 0.25 f_{y-\text{inner}} + 0.14 f_{y-\text{outer}} + 0.61 f_{y-\text{roller}} + 8.58 \times 10 = 0 \\ \ddot{x}_p + 9.13 \times 10 \dot{x}_p + 9.93 \times 10^5 x_p - 0.03 f_{x-\text{inner}} - 0.01 f_{x-\text{outer}} - 0.06 f_{x-\text{roller}} + 7.24 \times 10 = 0 \\ \ddot{y}_p + 6.63 \times 10^3 \dot{y}_p + 9.92 \times 10^5 y_p - 0.03 f_{y-\text{inner}} - 0.01 f_{y-\text{outer}} - 0.06 f_{y-\text{roller}} + 2.44 = 0 \end{cases} \quad (43)$$

As shown in Eq.(43) & Fig. 14, the roller fault force term obtains the maximum weight in the reconstructed system; therefore, the diagnosis result is roller fault.

#### 4.5. Discussions and limitations

Take the experiment bearing outer race failure signal as an example to discuss the function of the BYOLDIS each loss term. As Table 3 shows the inner race, outer race and the roller fault contact force weight of the sparse output matrix  $\Xi$  with different BYOLDIS structure. Table 3 indicates that the loss term  $\mathcal{L}_{\bar{z}_4}$  and  $\mathcal{L}_{\bar{z}_4}$  play an importance role on the BYOLDIS training. Without those two loss term, the diagnosis result from the reconstructed force weight is wrong. comparing with the loss term  $\mathcal{L}_{\bar{z}}$ ,  $\mathcal{L}_{z_4}$  and the BYOL network structure, loss terms  $\mathcal{L}_{\bar{z}}$  and  $\mathcal{L}_{z_4}$  have made the greatest contribution to the reconstruction fault force weight required for fault diagnosis. This is because these two loss items directly compare the reconstructed signal with the vibration signal, providing the most reliable information for the network to perform signal reconstruction. Beside,  $\mathcal{L}_{\bar{z}}$ ,  $\mathcal{L}_{z_4}$  and the BYOL network structure

also can help BYOLDIS network achieve effective signal reconstruction and diagnosis result output.

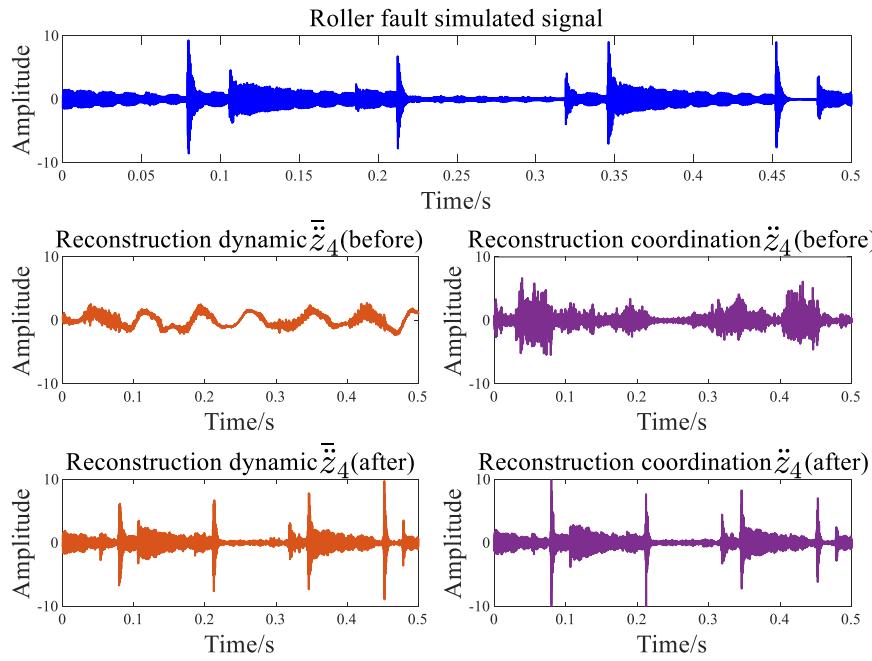
Taking the experiment bearing signal of outer race failure as an example, Fig. 15 shows the reconstruction dynamic signal has a similar signal shape with the input signal, and the spectral characteristics are the same as the input signal. It indicates that the BYOLDIS achieves input signal system reconstruction to capture dynamic information for bearing fault diagnosis. Besides, when the BYOLDIS have not the loss term  $\mathcal{L}_{\bar{z}_4}$ ,

the reconstruct dynamic signal spectrum can observe thehigh-frequency information loss and contains an extremely low frequency signal component. This indicates that insufficient reconstruction information was not obtained during the network reconstruction process, which can also explain why the fault diagnosis result obtained by reconstruction system after removing loss item  $\mathcal{L}_{\bar{z}_4}$  is incorrect.

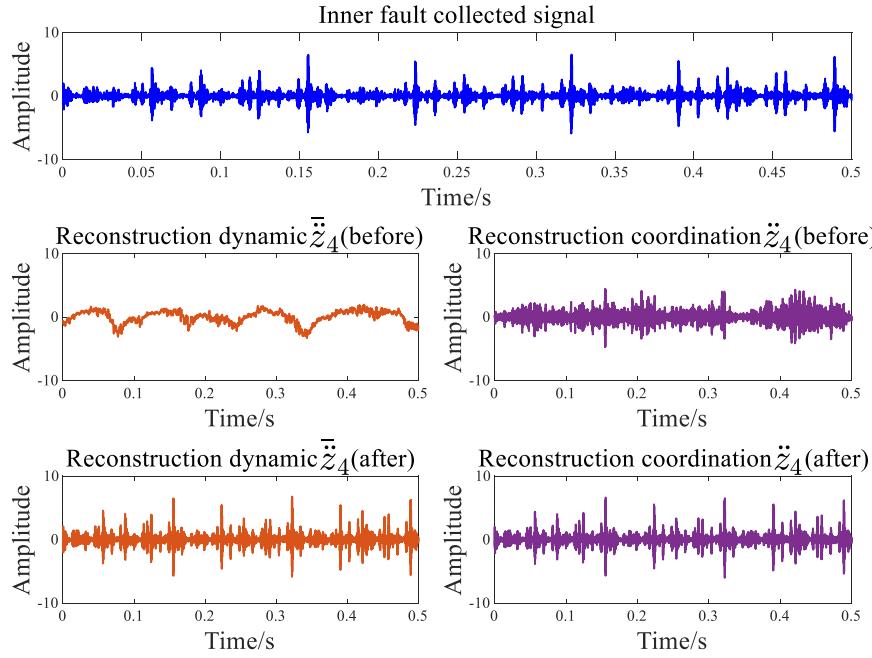
The experiments indicate that the dynamics embedding neural network is a feasible solution for dealing with the data lacking issue in machine fault diagnosis. Although this paper only discusses the bearing

fault diagnosis, we consider that the method can extrapolate to other mechanical systems that can be established as second-order differential equations.

At the same time, this method still has some limitations, which also



**Fig. 11.** The reconstruction signal the before->after training comparison of BYOLDIS from roller fault-bearing vibration signal.



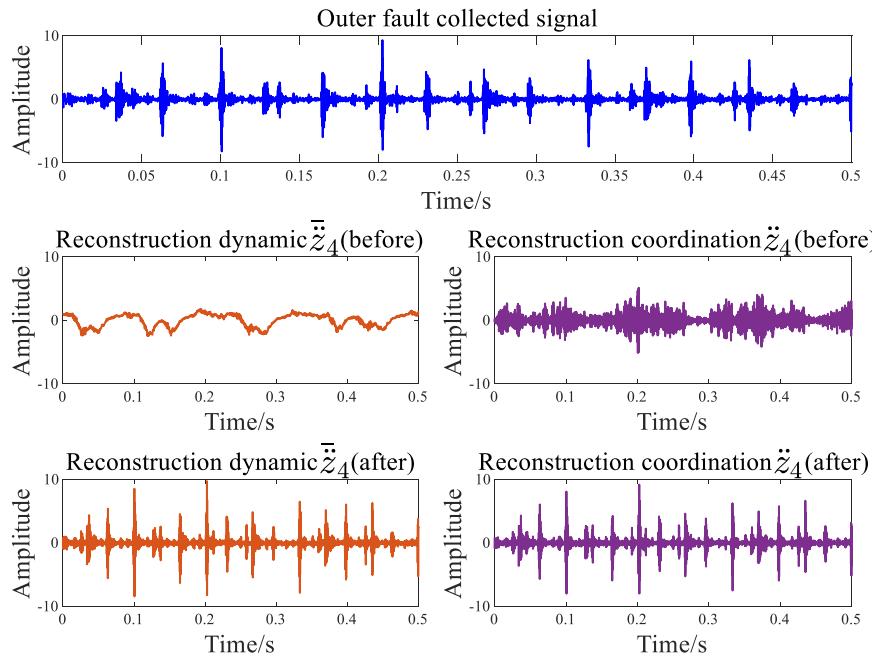
**Fig. 12.** The reconstruction signal the before->after training comparison of BYOLDIS from experiment inner race fault-bearing vibration signal.

are our future research direction. The first limitation is the gradient explosion. The coordinate encoder needs to generate the displacement, velocity, and acceleration information using the same network weights, and the magnitude order difference of displacement and acceleration is up to  $10^7$ . Therefore, this numerical difference can easily lead to gradient explosion in the training process. The second limitation is that it cannot diagnose a signal from a bearing that has a compound fault type. The reason is that the embedding fault-bearing multi-body contact dynamics system of BYOLDIS needs to be simplified to conduct accuracy dynamics system reconstruction. The principal summary of the proposed diagnosis method is to capture the critical information of the signal-governed dynamic system, which is achieved by selecting one of the three embedded fault-bearing models (SoftMax) to realize fault diag-

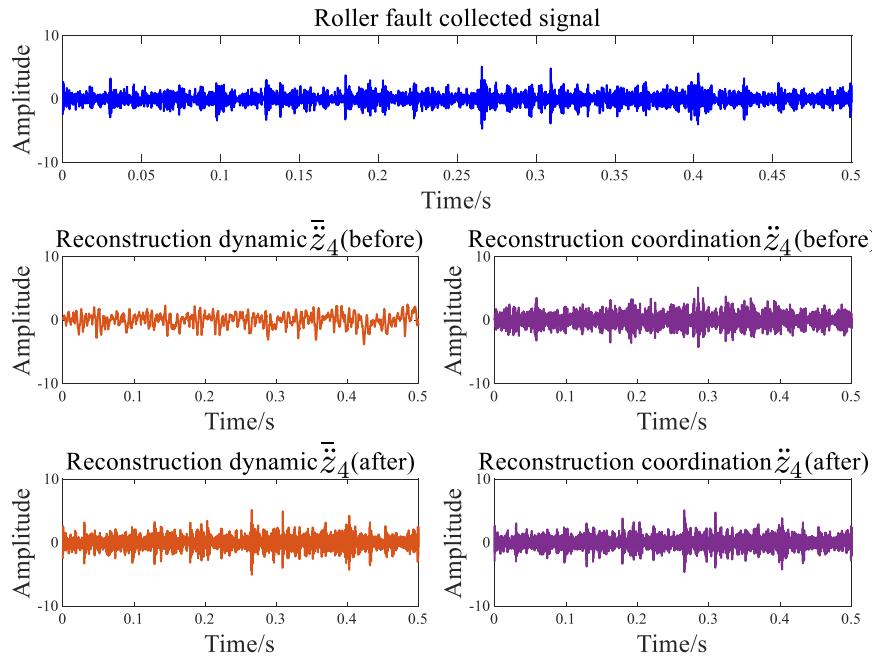
nosis. The third limitation is the time cost, the current training of the model for obtaining diagnosis results needs 6 min, and the time cost is high.

#### 4.6. Potential applications

The BYOLDIS (Bootstrap Your Own Latent and Dynamical Systems Model Discovery) method offers many potential applications, primarily in fault diagnosis and system monitoring within industrial and mechanical systems. This innovative approach presents a unique set of capabilities that can significantly enhance the reliability and efficiency of various operations across different industries. Below, we delve into some specific applications that exemplify the versatility and potential of



**Fig. 13.** The reconstruction signal the before->after training comparison of BYOLDIS from experiment outer race fault-bearing vibration signal.



**Fig. 14.** The reconstruction signal the before->after training comparison of BYOLDIS from experiment roller fault-bearing vibration signal.

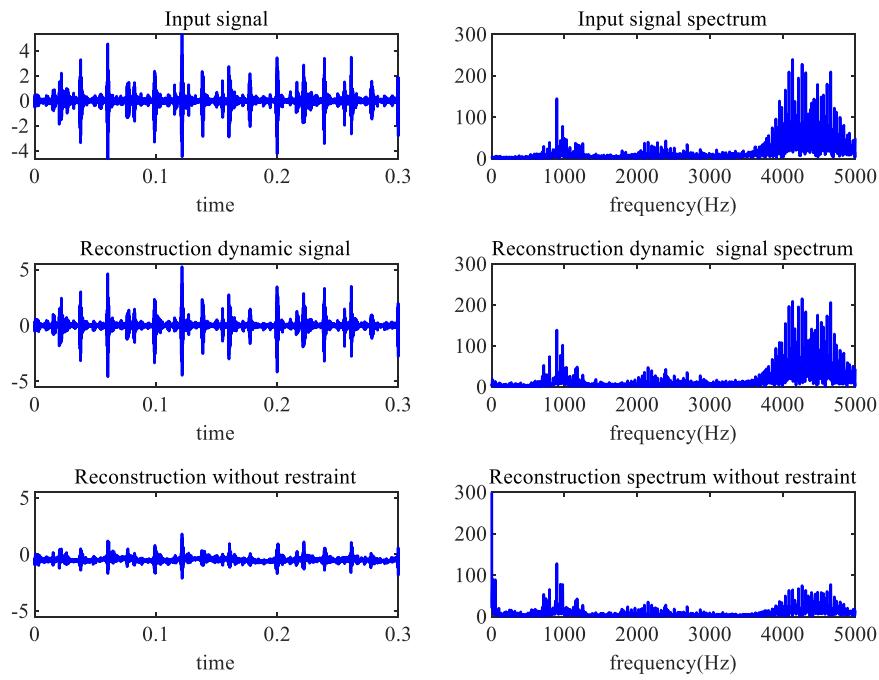
**Table 3**  
Influence of each loss term of BYOLDIS.

BYOLDIS structure/force weight	Inner race	Outer race	Roller
Orginal	0.13	0.63	0.25
Without BYOL structure	0.15	0.58	0.26
Without $\mathcal{L}_{\tilde{z}}$	0.14	0.52	0.33
Without $\mathcal{L}_{\tilde{z}_4}$	0.14	0.50	0.36
Without $\mathcal{L}_{\tilde{z}_4}$	0.17	0.27	0.56
Without $\mathcal{L}_{\tilde{z}_4}$	0.19	0.26	0.55
Only BYOL	0.33	0.33	0.33

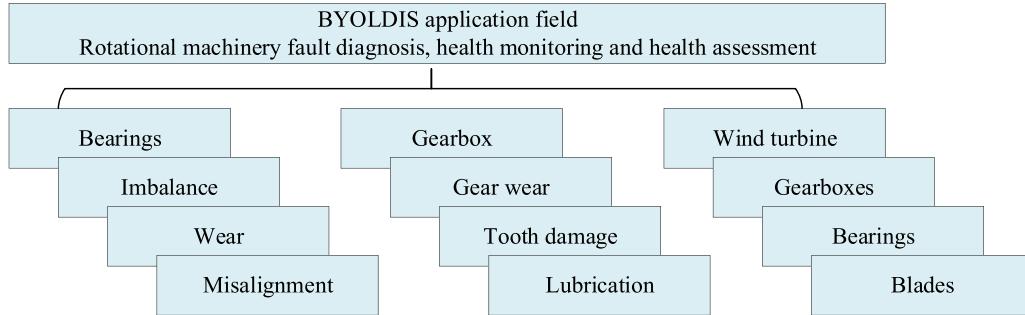
BYOLDIS, shown in Fig. 16. All of them are rotary machines, like bearings, gearboxes, wind turbines, etc:

#### 4.6.1. Bearings fault diagnosis

One of the primary applications of BYOLDIS lies in bearings fault diagnosis. In the context of rotating machinery, such as industrial motors, conveyor systems, and pumps, BYOLDIS proves invaluable. It excels at scrutinizing vibration signals emitted by bearings, effortlessly identifying and categorizing many faults. From wear and tear to misalignment and imbalance issues, BYOLDIS provides a comprehensive understanding of the bearing's health, allowing for timely maintenance and preventing costly breakdowns.



**Fig. 15.** The comparison of the input signal reconstruction dynamic signal and reconstruction dynamic signal without  $\mathcal{L}_{\bar{z}_i}$ .



**Fig. 16.** BYOLDIS potential application field.

#### 4.6.2. Gearbox health monitoring

Gears are ubiquitous components within machinery across industries, and their proper functioning is critical. BYOLDIS steps in as an adept gearbox health monitor. It diligently analyzes vibration data emanating from gearboxes, keeping a watchful eye on their condition. It can flag problems such as gear wear, tooth damage, and lubrication issues early in their development, mitigating potential catastrophes and minimizing downtime.

#### 4.6.3. Wind turbine health assessment

In the renewable energy sector, wind turbines are vital for power generation, and their health directly impacts energy output and efficiency. Continuous monitoring is paramount to ensure optimal performance. BYOLDIS can proficiently scrutinize vibration signals generated by wind turbines, making it adept at detecting and diagnosing faults within various components. From gearboxes to bearings and blades, BYOLDIS aids in the proactive maintenance of these critical assets, enhancing turbine longevity and energy production.

## 5. Conclusion

This paper proposes a BYOLDIS network for fault-bearing vibration signal diagnosis. This method utilizes a BYOL contrastive learning

framework combined with a library of fault-bearing dynamic polynomial equations that offer physical model-based prior constraints (dynamics embedding neural network). The findings can be summarized as follows: (1) Three experiments show that BYOLDIS can conduct label-free fault diagnosis well. The data dependence of the neural networks has been successfully removed. The proposed method can capture the vital fault physical characteristics of the input vibration signal for diagnosis. (2) BYOLDIS can effectively reconstruct the vibration impulse response from an input signal. (3) BYOLDIS can capture the essential physical characteristics of incoming vibration signals for conducting the diagnosis and the signals' physical characteristics can explain the fault occurrence principle. This indicates that the coordinate encoder of BYOLDIS can effectively map the partial dynamic information (1DOF) into the completed system dynamic information (4DOF). The reconstructed fault-bearing system equations contain crucial dynamic system information about the input signal. Therefore, the proposed BYOLDIS can effectively extract the fault dynamic equation characteristics, and the dynamic interpretation of the BYOLDIS diagnosis results is reliable. The proposed method can generally be considered promising for bearing fault diagnosis.

## 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 work was supported in part by the Natural Science Foundation of Guangdong Province (No. 2021A1515110615), in part by the Shenzhen Fundamental Research Program under Grants JCYJ20220810112354002 and JCYJ2020010911041013.

## References

- [1] Gawde S, Patil S, Kumar S, Kamat P, Kotecha K, Abraham A. Multi-fault diagnosis of Industrial Rotating Machines using Data-driven approach: a review of two decades of research. *Eng Appl Artif Intell* 2023;vol. 123:106139. /08/01/ 2023.
- [2] J. Sun, Z. Liu, J. Wen, and R. Fu, "Multiple hierarchical compression for deep neural network toward intelligent bearing fault diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 116, p. 105498, 2022/11/01/ 2022.
- [3] Haidong S, Hongkai J, Xingqiu L, Shuaipeng W. Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine. *Knowl-Based Syst* 2018;vol. 140:1–14.
- [4] Xu Y, et al. Cross-modal fusion convolutional neural networks with online soft label training strategy for mechanical fault diagnosis. *IEEE Trans Ind Inform* 2023;1–12.
- [5] Shao H, Jiang H, Zhang H, Liang T. Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network. *IEEE Trans Ind Electron* 2018;vol. 65(3):2727–36.
- [6] Sun M, Wang H, Liu P, Huang S, Wang P, Meng J. Stack autoencoder transfer learning algorithm for bearing fault diagnosis based on class separation and domain fusion. *IEEE Trans Ind Electron* 2022;vol. 69(3):3047–58.
- [7] Yang B, Lee C-G, Lei Y, Li N, Lu N. Deep partial transfer learning network: a method to selectively transfer diagnostic knowledge across related machines. *Mech Syst Signal Process* 2021;vol. 156:107618. /07/01/ 2021.
- [8] C. Huo, Q. Jiang, Y. Shen, Q. Zhu, and Q. Zhang, Enhanced transfer learning method for rolling bearing fault diagnosis based on linear superposition network, *Engineering Applications of Artificial Intelligence*, vol. 121, p. 105970, 2023/05/01/ 2023.
- [9] Lu W, Liang B, Cheng Y, Meng D, Yang J, Zhang T. Deep model based domain adaptation for fault diagnosis. *IEEE Trans Ind Electron* 2017;vol. 64(3):2296–305.
- [10] Y. Xia, C. Shen, D. Wang, Y. Shen, W. Huang, and Z. Zhu, "Moment matching-based intraclass multisource domain adaptation network for bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 168, p. 108697, 2022/04/01/ 2022.
- [11] Li X, Zhang W, Ding Q, Li X. Diagnosing rotating machines with weakly supervised data using deep transfer learning. *IEEE Trans Ind Inform* 2020;vol. 16(3):1688–97.
- [12] Zhu J, Chen N, Shen C. A new multiple source domain adaptation fault diagnosis method between different rotating machines. 1-1 *IEEE Trans Ind Inform* 2020. 1-1.
- [13] X. Liu et al., Self-supervised learning: Generative or contrastive," arXiv preprint arXiv:2006.08218, vol. 1, no. 2, 2020.
- [14] Oord A v d, Li Y, Vinyals O. Representation learning with contrastive predictive coding. *arXiv Prepr arXiv* 2018;1807:03748.
- [15] Songcan C, Chuanxing G. A comprehensive perspective of contrastive self-supervised learning. *Front Comput Sci* 2021;vol. 15(4).
- [16] J. Liu, H. Cao, S. Su, and X. Chen, Simulation-Driven Subdomain Adaptation Network for bearing fault diagnosis with missing samples," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106201, 2023/08/01/ 2023.
- [17] Peng T, Shen C, Sun S, Wang D. Fault feature extractor based on bootstrap your own latent and data augmentation algorithm for unlabeled vibration signals. *IEEE Trans Ind Electron* 2022;vol. 69(9):9547–55.
- [18] Rombach K, Michau G, Fink O. Contrastive learning for fault detection and diagnostics in the context of changing operating conditions and novel fault types. *Sens (Basel, Switz)* 2021;vol. 21.
- [19] X. Chen, C. Ge, M. Wang, and J. Wang, Supervised Contrastive Few-Shot Learning for High-Frequency Time Series," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 6, pp. 7069–7077, 06/26 2023.
- [20] Lu L, Wang J, Huang W, Shen C, Shi J, Zhu Z. Dual contrastive learning for semi-supervised fault diagnosis under extremely low label rate. *IEEE Trans Instrum Meas* 2023;vol. 72:1–12.
- [21] Zhang S, Ye F, Wang B, Haberler T. Few-shot bearing fault diagnosis based on model-agnostic meta-learning. 1-1 *IEEE Trans Ind Appl* 2021. 1–1.
- [22] Dixit S, Verma NK, Ghosh AK. Intelligent fault diagnosis of rotary machines: conditional auxiliary classifier gan coupled with meta learning using limited data. *IEEE Trans Instrum Meas* 2021;vol. 70:1–11.
- [23] D. Wang, M. Zhang, Y. Xu, W. Lu, J. Yang, and T. Zhang, Metric-based meta-learning model for few-shot fault diagnosis under multiple limited data conditions," *Mechanical Systems and Signal Processing*, vol. 155, p. 107510, 2021/06/16/ 2021.
- [24] Z. Pei, H. Jiang, X. Li, J. Zhang, and S. Liu, Data augmentation for rolling bearing fault diagnosis using an enhanced few-shot Wasserstein auto-encoder with meta-learning," *Measurement Science and Technology*, vol. 32, no. 8, p. 084007, 2021/05/17 2021.
- [25] Li C, Li S, Zhang A, He Q, Liao Z, Hu J. Meta-learning for few-shot bearing fault diagnosis under complex working conditions. *Neurocomputing* 2021;vol. 439: 197–211. /06/07/ 2021.
- [26] Z. Ye and J. Yu, Deep morphological convolutional network for feature learning of vibration signals and its applications to gearbox fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 161, p. 107984, 2021/12/01/ 2021.
- [27] T. Li et al., WaveletKernelNet: An Interpretable Deep Neural Network for Industrial Intelligent Diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–11, 2021.
- [28] Jia L, Chow TWS, Yuan Y. GTFE-Net: a Gramian Time Frequency Enhancement CNN for bearing fault diagnosis. *Eng Appl Artif Intell* 2023;vol. 119:105794. /03/01/ 2023.
- [29] Wang D, Chen Y, Shen C, Zhong J, Peng Z, Li C. Fully interpretable neural network for locating resonance frequency bands for machine condition monitoring. *Mech Syst Signal Process* 2022;vol. 168.
- [30] Michau G, Frusque G, Fink O. Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series. *Proc Natl Acad Sci* 2022; vol. 119(8):e2106598119.
- [31] T. Zhou, T. Han, and E.L. Drogue, Towards trustworthy machine fault diagnosis: A probabilistic Bayesian deep learning framework," *Reliability Engineering & System Safety*, vol. 224, p. 108525, 2022/08/01/ 2022.
- [32] L. Zuo, F. Xu, C. Zhang, T. Xiahou, and Y. Liu, A multi-layer spiking neural network-based approach to bearing fault diagnosis, *Reliability Engineering & System Safety*, vol. 225, p. 108561, 2022/09/01/ 2022.
- [33] Brunton SL, Proctor JL, Kutz JN. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Natl Acad Sci* 2016;vol. 113(15):3932–7.
- [34] Champion K, Lusch B, Kutz JN, Brunton SL. Data-driven discovery of coordinates and governing equations. *Proc Natl Acad Sci* 2019;vol. 116(45):22445.
- [35] Y. Xu, S. Kohtz, J. Boakye, P. Gardoni, and P. Wang, Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges," *Reliability Engineering & System Safety*, vol. 230, p. 108900, 2023/02/01/ 2023.
- [36] J.-B. Grill et al., Bootstrap your own latent-a new approach to self-supervised learning," *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- [37] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *IEEE Conf Comput Vis Pattern Recognit (CVPR)* 2016:770–8. 27–30 June 2016 2016.
- [38] Packard NH, Crutchfield JP, Farmer JD, Shaw RS. Geometry from a time series. *Phys Rev Lett* 1980;vol. 45(9):712.
- [39] Champion KP, Brunton SL, Kutz JN. Discovery of nonlinear multiscale systems: sampling strategies and embeddings. *SIAM J Appl Dyn Syst* 2019;vol. 18(1): 312–33.
- [40] J. Bakarji, K. Champion, J.N. Kutz, and S.L. Brunton, Discovering governing equations from partial measurements with deep delay autoencoders," *arXiv preprint arXiv:2201.05136*, 2022.
- [41] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, 2020: PMLR, pp. 1597–1607.
- [42] Sawalhi N, Randall RB. Simulating gear and bearing interactions in the presence of faults: part II: simulation of the vibrations produced by extended bearing faults,". *Mech Syst Signal Process* 2008;vol. 22(8):1952–66. /11/01/ 2008.
- [43] Sawalhi N, Randall RB. Simulating gear and bearing interactions in the presence of faults: part I. the combined gear bearing dynamic model and the simulation of localised bearing faults. *Mech Syst Signal Process* 2008;vol. 22(8):1924–51. /11/01/ 2008.
- [44] Smith WA, Randall RB. Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. *Mech Syst Signal Process* 2015;vol. 64–65:100–31.