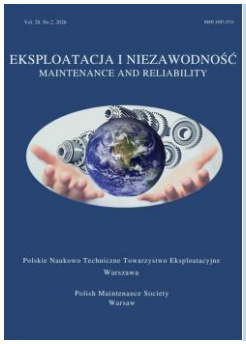

Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 28 (2026), Issue 2

journal homepage: http://www.ein.org.pl

Article citation info:
Wei Q, Wang X, Sun Y, Tian X, Cui L, A signal structure priori enhanced deep compressed sensing method for vibration signals and an improved reconstruction evaluation method in the PHM framework, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(2) http:/10.17531/ein/215926

A signal structure priori enhanced deep compressed sensing method for vibration signals and an improved reconstruction evaluation method in the PHM framework

Indexed by: Web of Science Group

Qingzhe Wei[a,b], Xiuyu Wang[a,b], Yujie Sun[a,c], Xincheng Tian[a,b,*], Long Cui[a,b,]**

[a] The Center for Robotics, School of Control Science and Engineering, Shandong University, China
[b] The Engineering Research Center of Intelligent Unmanned System, Ministry of Education, China
[c] School of Mechanical, Electrical and Information Engineering, Shandong University, China

Highlights

- A deep compressed sensing method with the locality/globality and the structure prior.
- The output layer weights are initialized by the data structure information.
- Subsequent PHM tasks accuracy is considered to better evaluate the PHM tasks fitness.
- The method has superior adaptability in the PHM framework.

Abstract

The compressed sensing (CS) methods have two issues when processing vibration signals of rotating machinery: (1) the data-driven method and signal priors (e.g., global/local and structure priors) are not combined well, resulting in the reconstruction not converging to high-precision results; (2) the current metrics only focus on numerical accuracy and do not consider PHM scene fitness. We propose a deep compressed sensing framework (RMVS-DCS): (1) Global-Local Feature Reconstruction (GLFR) blocks alternately reconstruct features at global and local scales. Pre-train the output layer by dictionary learning, introducing the signal structure prior to guiding the model to converge to a better one. (2) A dual evaluation system, combining the numerical indicators and prognostics and health management (PHM) task accuracy, helps the model selection in PHM scenarios. On the rotor unbalance data set, its reconstruction and prediction error are reduced by 82.60% and 89.90% (compression ratio is 0.5), demonstrating the advanced reconstruction precision and PHM compatibility of the framework.

Keywords
deep compressed sensing, signal structure prior, prognostics and health management, rotating machinery vibration signal, reconstruction evaluation method

1. Introduction

With the theory development and hardware upgrading, prognostics and health management (PHM) [1] has become critical to equipment maintenance [2, 3]. The vibration signal of rotating machinery is used in PHM tasks such as fault diagnosis [4], anomaly detection [5], and life prediction [6] because it contains a lot of running state information [7, 8].

Long-time and high-frequency sampling will produce amounts of data that significantly burden the sampling, transmission, and storage devices [9], especially after adopting multi-channel [10], multi-modal [11], and large model [12] technologies. Many data compression methods can be used to alleviate the problem. Still, almost all of them need to follow the Nyquist sampling theorem and then use complex algorithms to compress the signal on the sampling device.

(*) (**) Corresponding author.
E-mail addresses: Q. Wei (ORCID: 0009-0002-0192-3045) 451021634@qq.com, X. Wang (ORCID: 0009-0009-3360-1199) ytwangxiuyu@gmail.com, Y. Sun (ORCID: 0000-0002-1397-8302) sunyujie@sdu.edu.cn, X. Tian (ORCID: 0000-0001-9814-9521) txch@sdu.edu.cn, L. Cui (ORCID: 0000-0001-5450-5221) cuilong@sdu.edu.cn

Compressed sensing (CS) [13] is a method combining signal sampling, compression, and reconstruction, by which people can obtain the compressed signal under the condition of less than Nyquist sampling frequency [14, 15]. Only one measurement matrix is needed for compression on the sampling device, and the signal is recovered from the compressed data on the server device. Therefore, compressed sensing has attracted the attention of academic and industrial fields since it was proposed [16, 17].

In classical compressed sensing, the signal can be compressed if it is sparse enough in some transform domain. The signal is downsampled with the matrix satisfying certain conditions. Some random matrices, such as the Gaussian and Bernoulli random matrices, are proven to fit the task [13]. Finally, the original signal is recovered using the reconstruction algorithm. There are three main problems in compressed sensing: sparse representation, measurement matrix selection, and reconstruction algorithm design. Among the issues, reconstruction algorithm design is most important because it directly decides reconstruction quality. The sparse signal in specific domains can be recovered firstly by some methods, such as SAMPVSS [18] and other greedy algorithms; fast iterative parametric improved threshold algorithm (FIPITA) [19] and other convex relaxation methods; covariance-free sparse Bayesian learning [20] and other reconstruction algorithms based on Bayesian frameworks. Mathematical models support these algorithms and have good interpretability, but the reconstruction accuracy is usually insufficient.

In recent years, deep learning has shown excellent learning ability in many tasks. Some studies introduce deep learning into compressed sensing reconstruction tasks [21, 22]. Kulkarni et al. [23] reconstruct the original signal/image from measured values end-to-end. Zhang et al. [24] combine the optimization method with the network to unfold the iterative process, and methods like it are called deep unfolding networks (DUN) [25].

In the PHM field, some scholars try to apply compressed sensing to fault diagnosis tasks [26]. Shao et al. [27] use the compressed sensing measurement process to reduce the vibration data amount for analysis efficiency. Yuan et al. [28] use compressed sensing for fault diagnosis based on reconstruction. However, only a few compressed sensing studies [29, 30] aim to get more accurate reconstructed vibration signals of rotating machinery. Sun et al. [31] introduce the block sparse Bayesian learning (BSBL) algorithm, which uses the block characteristics and inherent structure of the bearing signal to reconstruct the sparse coefficient and restore the original signal. Pan et al. [32] propose an algorithm named variable stepsize forward-backward pursuit (VSFBP), which uses two fuzzy parameters to control the two-stage matching pursuit algorithm to improve the reconstruction accuracy as well as speed, and updates the observation matrix after each iteration to prevent excessive backtracking. In [33] and [34], the researchers try to overcome the defects existing in the iterative threshold shrinkage algorithms and apply the improved methods to the vibration signal compressed sensing. Zhang et al. [35] use the end-to-end learning method to build the measurement matrix and the high-precision reconstruction network, designing a deep compressed sensing (DCS) network named DCSNet.

The methods demonstrate the potential of compressed sensing for vibration signals, but there are still some problems in the studies:

(1) the methods do not combine the data-driven method's powerful learning ability well with the signal's priors (e.g., the locality/globality of signal features and the signal structure prior), resulting in the reconstruction not converging to high-precision results.

(2) the suitability for PHM tasks is not comprehensively discussed, which affects the optimization and selection of the model.

Actually, the latter may be attributed to the fact that most evaluation methods only consider the statistical properties of the reconstructed signals and do not take the particularity of the compressed sensing in the PHM framework into account: the reconstruction results are prepared for subsequent health status analysis.

To obtain a method that can produce more accurate and suitable reconstruction results in the PHM framework, we propose a deep compression sensing framework named **rotating machinery vibration signal-deep compressed sensing** (RMVS-DCS). The measurement is accomplished by only one learnable linear measurement layer to reduce the burden of the sampling device and preserve the key features. The reconstruction network is a multi-layer neural network,

including several blocks for cooperative and alternative global information fusion and local feature learning. We use an improved optimization strategy to acquire the parameters of the output layer. The network training obtained an excellent initial state by giving the initial weights that can sparsely represent the original signal, improving the signal reconstruction ability.

In addition to the numerical accuracy, we also use reconstructed signals and raw data as inputs for health diagnosis tasks, incorporating their accuracies into the evaluation criteria to better measure the fitness of compressed sensing methods in various rotating machinery PHM scenarios.

The contributions of the paper are as follows:

(1) An deep compressed sensing framework for PHM named RMVS-DCS is proposed. Global-Local Feature Reconstruction (GLFR) blocks can alternately reconstruct features at both global and local scales by the powerful nonlinear fitting capability of attention-convolution mixed networks. An improved output layer optimizing strategy is proposed. The output layer weights are initialized with a set of standard sparse bases that can redundantly represent the signal, introducing the signal structure prior for guiding the model to converge to a better one.

(2) A dual evaluation system for reconstruction helps the evaluation and selection of CS models in PHM scenarios. Besides the numerical accuracy, the accuracy of subsequent various PHM tasks should also be considered compared with some control groups.

The remainder of the paper will give a more detailed description of our research. Section 2 introduces classical compressed sensing. Section 3 describes the algorithm and reconstruction evaluation method of the RMVS-DCS framework. Section 4 presents the experiments on a private rotor data set and two bearing data sets. Section 5 is the conclusions and prospects.

2. Compressed sensing

Compressed sensing includes two stages: measurement and reconstruction.

The measurement is a process combining sampling and compression. The quality of the measurement is vital for perfect recovery of the signal, so it is usually impossible to ignore the measurement when discussing the reconstruction. Suppose the raw signal is $x \in \mathbb{R}^{N\times 1}$, the measurement process can be expressed as

$$y = \Phi x \tag{1}$$

where $y \in \mathbb{R}^{M\times 1}$ is the measurement result and $\Phi \in \mathbb{R}^{M\times N}(M < N)$ is a measurement matrix. We define the compression ratio CR in Eq. (2)

$$CR = M/N \tag{2}$$

The reconstruction is the process of getting x by y and Φ. It is an ill-conditioned problem that makes it challenging to obtain a solution directly. However, if a signal can be sparsely represented by a sparse basis matrix (dictionary) and some sparse coefficients, it can be recovered from the measurement result according to classical compressed sensing [13]. Suppose $x = \Psi s$, where $\Psi \in \mathbb{R}^{N\times N}$ is the sparse basis matrix and $s \in \mathbb{R}^{N\times 1}$ is the sparse coefficients reconstructed first. If s has K ($K < M$) non-zero elements, we call s a K-sparse vector. Eq. (1) can be transformed into

$$y = \Phi x = \Phi\Psi s = \Theta s \tag{3}$$

where Θ is called the sensing matrix.

To ensure that the raw signal is successfully reconstructed, Θ must meet the restricted isometry property (RIP)

$$(1-\delta_K)\|x\|_{l_2}^2 \le \|\Theta x\|_{l_2}^2 \le (1+\delta_K)\|x\|_{l_2}^2 \tag{4}$$

where δ_K is restricted isometry constraint (RIC), $\delta_K < 1$. Confirming whether a matrix satisfies RIP is complicated. In [36], it is proved that the equivalence condition for RIP is that Φ and Ψ are incoherent, and some random measurement matrices (Gaussian random matrix, etc.) satisfy the condition.

The reconstruction problem can be described as

$$s' = argmin\|s'\|_0 \text{ subject to } \|y - \Theta s'\|_2^2 \le \varepsilon \tag{5}$$

where ε is the parameter depending on the noise variance. The problem is still NP-hard [37]. Under certain conditions, the ℓ_1-norm optimization objective is equivalent to the ℓ_0-norm optimization objective [13]

$$s' = argmin\|s'\|_1 \text{ subject to } \|y - \Theta s'\|_2^2 \le \varepsilon \tag{6}$$

Eq. (6) can be solved by the convex optimization theory

3. RMVS-DCS and improved CS reconstruction evaluation method

3.1. RMVS-DCS

Under the abovementioned constraints, classical CS frameworks realize signal reconstruction through greedy algorithms, convex relaxation methods, and others. However, they might have mediocre reconstruction accuracy, lose important status information [22], and provide poor inputs for

health state evaluation. There is the possibility of further improvement in these aspects. We propose a more effective deep compressed sensing framework named RMVS-DCS, accomplishing the end-to-end measurement-reconstruction process through neural networks, as shown in Fig. 1. It can improve the reconstruction accuracy and fitness in PHM tasks by the designed structure.

3.1.1. Signal Preprocess

First, we need to preprocess the raw signal. An original signal sample $x_i = [x_i^1, x_i^2, \ldots, x_i^L]$ (L is the length of the sample, i is the index in the data set or batch X) is divided into s sub-segments with length N. The dimensions of the batch change from $X \in \mathbb{R}^{b\times L}$ to $X \in \mathbb{R}^{bs\times N}$, where b is the batch size.

3.1.2. Measurement

The measurement module (MM) contains only one learnable linear layer. The reason for not using a deeper network is that the measurement is carried out at the sampling end, where the computational ability is usually limited, so complex networks are unsuitable. The expression is

$$Y = (W_\Phi X^{\mathrm{T}})^{\mathrm{T}}, W_\Phi \in \mathbb{R}^{M\times N} \tag{7}$$

where W_Φ is the $M \times N$ learnable measurement matrix, X is the input signal, and $Y \in \mathbb{R}^{bs\times M}$ is the measurement result. The measurement and the reconstruction modules (RM) are connected during network training and its parameters are updated by back-propagation. Compared with classic measurement matrices (e.g., Gaussian random matrix), the learning measurement matrix ignores the constraint of RIP and can project the signal into the suitable space for reconstruction. The information required is preserved as much as possible.

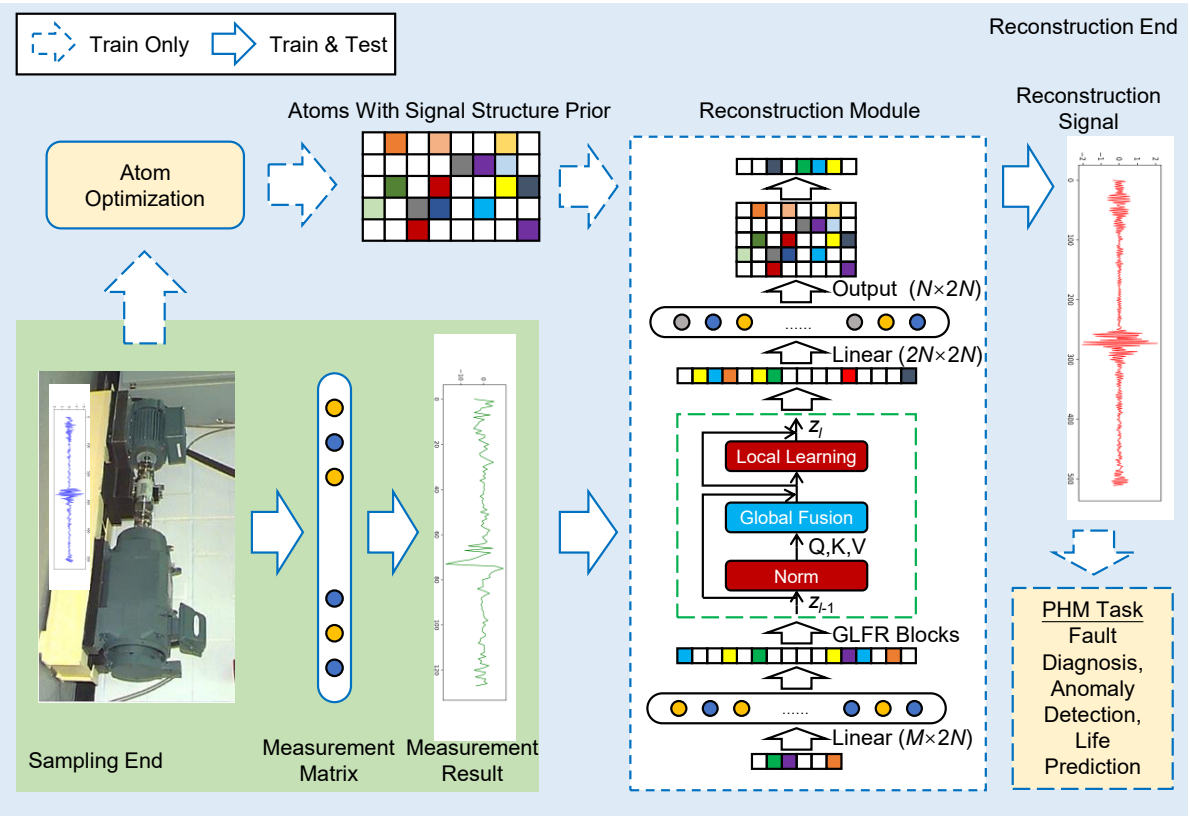

Fig. 1. Schematic diagram of RMVS-DCS.

3.1.3. Reconstruction

As shown in Fig. 2, the reconstruction module consists of the 1st linear projection layer, several GLFR blocks, the 2nd linear projection layer, and an output layer that acquires the prior knowledge about the signal structure.

1st linear projection layer In the reconstruction stage, the dimensions of the measurement result should be increased first

$$Z_{\mathrm{LL1}} = (W_{\mathrm{LL1}} Y^{\mathrm{T}})^{\mathrm{T}} + B_{\mathrm{LL1}}, W_{\mathrm{LL1}} \in \mathbb{R}^{2N\times M} \tag{8}$$

where Z_{LL1} is the output after increasing dimensions, and Y is the measurement result. W_{LL1}, B_{LL1} are the weights and biases of the layer, respectively. The design of increasing the dimensions to $2N$ is not a simple expansion. Instead, it is aimed at constructing a more comprehensive feature expression space:

(1) In the low-dimensional measurement space, the features of different vibration modes are prone to overlap. After dimensionality increase, the larger spatial freedom can disperse the originally compact measurement vectors to higher-dimensional coordinates, highlighting the subtle differences of vibration signals under different conditions.

(2) The subsequent module needs to alternatively perform global information fusion and local feature learning. The high-dimensional feature space can provide more sufficient information carriers for these two operations and reduce the attenuation of effective information caused by multiple feature transformations.

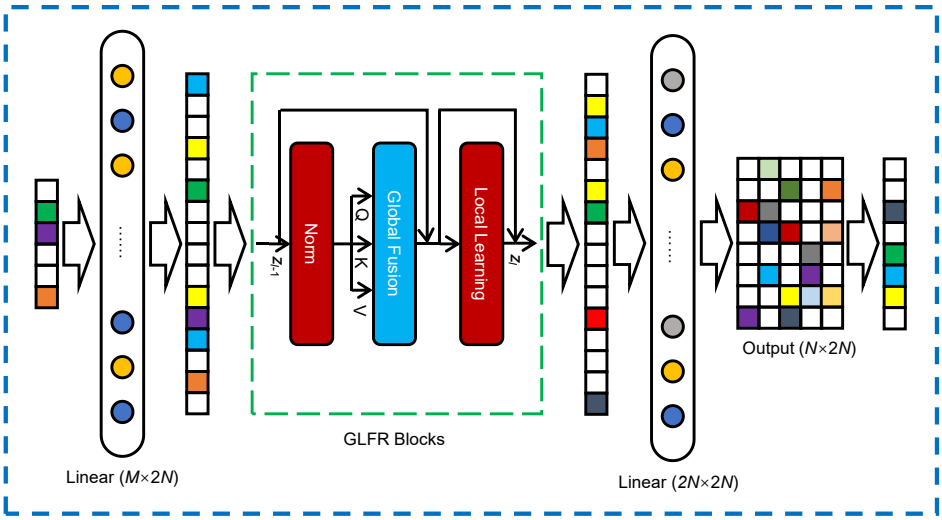

Fig. 2. Schematic diagram of the reconstruction module.

GLFR block The signal recovery quality depends very much on the feature extraction ability of the reconstruction network. Global-local feature reconstruction is more suitable for rotating machinery vibration signals than a single operation. The core reason is that rotating machinery vibration signals inherently contain both global-level information and local-level information (such as transient impact pulses). A single global operation is prone to smoothing out local fault details, resulting in a deterioration of reconstruction quality, while a single local operation would disrupt the long-range dependency relationships of the signal.

We design a block that can alternatively extract both types of features, named GLFR block. A global information fusion component first models the dependencies between sub-segments to provide an overall context, then extracts local fine-grained features through a local feature learning component, and simultaneously retains both types of information through residual connections. The details are as follows.

The GLFR block comprises a normalization layer, a global information fusion component, and a local feature learning component. In the block, layer normalization LN(·) is applied to the input first. It performs normalization within the sample, mitigating overfitting and making the training process smoother

$$\begin{aligned} \mu &= \sum_{i=0,j=0}^{s,N} x_{i,j} \\ \sigma &= \sqrt{\sum_{i=0,j=0}^{s,N} (x_{i,j}-\mu)^2 + \varepsilon} \\ \mathrm{LN}(x) &= (x-\mu)/\sigma * \gamma + \beta \end{aligned} \tag{9}$$

where $x \in \mathbb{R}^{s\times N}$ are the representations of the sub-segment, μ is the mean, σ is the standard deviation, γ and β are the learnable scaling and bias parameters, respectively, ε is a small value that prevents σ from being zero.

The global information fusion component performs information fusion among sub-segments by the self-attention operation

$$Attn(Q,K,V) = \mathrm{Softmax}(QK^{\mathrm{T}}/\sqrt{d_k})V \tag{10}$$

where $\sqrt{d_k}$ is a scale factor. Q, K, and V are query, key, and value matrices. Because they are the mapping results of the same input in linear layers, the multiplication of Q and K can capture the dependence between sub-segments in pairs [38, 39], by which the relationships among sub-segments are learned. If using the convolutional network, it needs to increase the depth to expand the receptive field to extract global features, while the self-attention structure can accomplish this without the addition of extra layers.

The global feature fusion component abandons the multi-head mechanism because the core design logic of the GLFR block is the synergy of global information fusion (attention) and local feature learning (convolution). If multi-head attention further splits the features in multiple dimensions, it would separate the originally continuous features into multiple sub-dimensions, resulting in the dispersion of global correlations between sub-segments; single-head attention would retain globality. Besides, multi-head attention, due to its more parameters and more complex optimization space, is prone to training oscillations and difficult to converge to the optimal solution; while single-head attention has a simpler optimization objective and a smoother model training process, and can converge to a better one.

After that, there should be a component for in-segment learning. We design a local feature learning component, which is composed of convolutional layers, batch normalization layers, and a GELU (Gaussian error linear unit) [40] activation layer for in-segment feature extraction

$$\begin{aligned} Z_l^i &= \mathrm{BN}_i(\mathrm{Conv}_i(Z_l^{i\prime})), i = 1,2,3 \\ Z_l^{\mathrm{output}} &= \mathrm{GELU}(Z_l^3) \end{aligned} \tag{11}$$

where $Z_l^{i\prime}$ and Z_l^i are the input and output of the ith unit of the local feature learning component in the lth GLFR block. $\mathrm{Conv}_i(\cdot)$ is the convolutional layer. The key local features of vibration signals have local correlation and translation invariance. The convolution layer inherently possesses these two inductive biases. By sliding the convolution kernel, it can accurately extract local details within sub-fragments, avoiding missed detections due to changes in feature positions. Additionally, the convolutional structure can avoid overfitting during the signal reconstruction process.

In the convolutional structure, each kernel is equivalent to a filter. The filters in the shallower layer are responsible for extracting low-frequency features and learning the general outline. These types of features are relatively few in number but have a larger size. With a small number of kernels, they can learn relatively fully. Therefore, the first convolutional layer has a smaller number of channels (32 in our experiments) and a larger kernel size (7). The filters at the deeper layer tend to extract high-frequency features, i.e., the details of the signal. These types of features are more numerous and have a smaller size. More kernels are needed for learning. Therefore, the second convolutional layer has a larger number of channels (64) and a smaller kernel size (3). The third convolutional layer is a single-channel structure mainly to summarize all the feature maps and obtain a feature map that best reflects the general characteristics.

$\mathrm{BN}_i(\cdot)$ represents the batch normalization layer in the ith unit.

$$\begin{aligned} \mu_j &= \frac{1}{B}\sum_{i=0,k=0}^{B,N} X_{i,j,k} \\ \sigma_j &= \sqrt{\frac{1}{B}\sum_{i=0,k=0}^{B,N}(X_{i,j,k}-\mu)^2 + \varepsilon} \\ \mathrm{BN}(X) &= (X-\mu)/\sigma * \gamma + \beta \end{aligned} \tag{12}$$

where j is the channel index, and B is the current batch size. The expression of GELU is

$$\begin{aligned} \Phi(x) &= [1 + \mathrm{erf}(x/\sqrt{2})]/2 \\ \mathrm{GELU}(x) &= x\Phi(x) \end{aligned} \tag{13}$$

where $\mathrm{erf}(\cdot) = (2/\sqrt{\pi})\int_0^x e^{-t^2}\,dt$. Using GELU as an activation function has three advantages: (1) it has a smoother derivative curve; (2) it alleviates the vanishing gradient; (3) it accelerates convergence. The component reduces the risk of overfitting, making the network easier to train by introducing inductive biases and normalization.

The global information fusion and local feature learning components are connected by shortcuts, respectively. The expression is

$$\begin{aligned} Z_l{}' &= F_{\mathrm{global}}(\mathrm{LN}(Z_{l-1})) + Z_{l-1} \\ Z_l &= F_{\mathrm{local}}(Z_l{}') + Z_l{}' \end{aligned} \tag{14}$$

where $Z_l{}'$ is the output of the global information fusion component in the lth block, $F_{\mathrm{global}}(\cdot)$ is the global information fusion component, $F_{\mathrm{local}}(\cdot)$ is the local feature learning component, $\mathrm{LN}(\cdot)$ is layer normalization, Z_{l-1} is the output of the l-1th GLFR block and Z_l is the output of the lth GLFR block. Not only do the characteristics of the original vibration signals possess both global and local properties, but we also hope that the reconstruction network can integrate the global and local features at different depths. Therefore, the two processes are completed alternately, rather than in other orders.

2nd linear projection layer After multiple GLFR blocks, a linear layer maps the output once

$$Z_{\mathrm{LL2}} = (W_{\mathrm{LL2}}Z_l{}^{\mathrm{T}})^{\mathrm{T}} + B_{\mathrm{LL2}}, W_{\mathrm{LL2}} \in \mathbb{R}^{2N\times 2N} \tag{15}$$

where Z_l and Z_{LL2} are the input and output of the 2nd linear layer, W_{LL2} and B_{LL2} are the weights and biases. The shape of

W_{LL2} indicates that the output's dimensions are not changed in the layer.

Signal structure prior enhanced output layer The deep compressed sensing completely optimized by back-propagation largely depends on data quality. Its generalization performance needs to be further improved. In other words, the reconstruction quality of the new sample is not as good as expected.

To deal with the problem, we propose an output layer optimizing strategy for RMVS-DCS. The strategy tries to find an overcomplete (a matrix with columns more than *N*) set of atoms $\Psi \in \mathbb{R}^{N\times C}$ $(N < C)$ to initial weights, which can perform sparse signal reconstruction. The optimization objective of the weight initialization is

$$X \approx \Psi S \tag{16}$$

where the original signals $X = \{x_i\}_{i=1}^{n}$, the sparse coefficients $S = \{s_i\}_{i=1}^{n}$. S should be as sparse as possible. The constrained optimization objective can be expressed as

$$\min_{\Psi,S} \|X - \Psi S\|_F^2 \text{ subject to } \forall i \ \|s_i\|_0 \le K_0 \tag{17}$$

where K_0 is the sparsity threshold. Eq. (17) also can be transformed into an unconstrained form using the Lagrange multiplier method

$$\min_{\Psi,S} \|X - \Psi S\|_F^2 + \lambda \|s_i\|_1 \tag{18}$$

where λ is the Lagrange multiplier. $\|s_i\|_0$ is substituted for $\|s_i\|_1$ to facilitate the solution. Eq. (17) can be solved by K-SVD. After that, the acquired atoms are transferred to the output layer as the initial weights W_Ψ. Since the initial weights are a set of overcomplete atoms that can sparsely represent the original signal, the subsequent optimization process is in a better initial state. In contrast, the ordinary weight initialization method does not have this property. Finally, we train the network using the back-propagation algorithm.

During inference, the output of the 2nd linear layer Z_{LL2} and W_Ψ are multiplied to get the reconstructed sub-segments

$$X_i' = (W_\Psi {Z_{\mathrm{LL2}}}^{\mathrm{T}})^{\mathrm{T}} \tag{19}$$

The layer projects the previous output into the corresponding space. The prior knowledge of the signal structure is transferred to the network, and the range of solutions is constrained, improving its generalization and reconstruction ability. Because the introduced signal structure prior is supported by rigorous mathematical theory, the interpretability of the model is also enhanced.

After several sub-segments are reconstituted to form the complete signal, the reconstruction result with higher accuracy is acquired.

3.1.4. Training process

The training process mainly includes two stages: initializing the weights of the output layer and training the network.

Algorithm 1: the optimizing process of the output layer in RMVS-DCS

Task: initialize and optimize the output layer parameters of RMVS-DCS

Hyperparameters: max iterations of K-SVD n_{iter}, tolerance of K-SVD t

Reshape $X \in \mathbb{R}^{n\times L} \to X \in \mathbb{R}^{ns\times N}$

Initialize Select C signal sub-segments in X to initialize $\Psi^{(0)} = \{\psi_i^{(0)}\}_{i=1}^{C}, \Psi^{(0)} \in \mathbb{R}^{N\times C}$. Iterator $J = 1$

Normalize $\psi_i^{(0)} = \psi_i^{(0)} / \sqrt{\sum_{j=1}^{N} (\psi_{j,i}^{(0)})^2}, i = 1,2,3,\dots,C$

Repeat until $J = n_{\mathrm{iter}}$ or $\|X - \Psi S\|_F^2 < t$

K-SVD Dictionary Learning Obtain S by match pursuit $\mathrm{MP}(X, \Psi)$. Update Ψ by decomposing the error ${E_k}'$ using SVD. $J = J + 1$

end

Normalization $\psi_i = \psi_i / \sqrt{\sum_{j=1}^{N} \psi_{j,i}^2}, i = 1,2,3,\dots,C$

If $\|X - \Psi S\|_F^2$ does not converge, increase n_{iter} and train again.

// Network training stage

Initialize the weights of the output layer in RMVS-DCS $W_\Psi = \Psi$

Forward $X_i' = (W_\Psi {Z_{\mathrm{LL2}}}^{T})^{T}$

Batch loss $loss(X, X';\ W_\Psi)$

If W_Ψ need to be updated

Calculate gradients $\frac{\delta loss}{\delta W_\Psi}$

Update weights $W_\Psi \leftarrow W_\Psi - \eta \frac{\delta loss}{\delta W_\Psi}$

Else do not calculate gradients

end

Before using back-propagation to optimize the neural network, the initial weights of the output layer should be obtained by optimizing Eq. (17). We choose the K-SVD algorithm for the task. K-SVD is a dictionary (i.e., a set of atoms) learning method that can utilize data structure information to design an overcomplete dictionary. Let *C* sub-segments of the samples as atoms to initialize a matrix $\Psi^{(0)} = \{\psi_i^{(0)}\}_{i=1}^{C}, \Psi^{(0)} \in \mathbb{R}^{N\times C}$. Ψ and S are updated alternately during the optimization process. Fixing Ψ, it is also NP-hard. Therefore, we can use matching pursuit (MP) to approximate *S*. When updating each column ψ_k of the dictionary Ψ in turn, Eq. (17) can be transformed into

$$\left\|X-\Psi S\right\|_F^2=\left\|X-\sum_{j=1}^{C}\psi_j s_{\mathrm{T}}^j\right\|_F^2$$
$$=\left\|(X-\sum_{j\neq k}\psi_j s_{\mathrm{T}}^j)-\psi_k s_{\mathrm{T}}^k\right\|_F^2 \quad (20)$$
$$=\left\|E_k-\psi_k s_{\mathrm{T}}^k\right\|_F^2$$

where s_{T}^j, s_{T}^k are the jth and kth rows of S. The form of the optimization problem becomes

$$\min_{\psi_k,s_{\mathrm{T}}^k}\left\|E_k-\psi_k s_{\mathrm{T}}^k\right\|_F^2 \quad (21)$$

However, decomposing E_k directly does not result in a sparse result. The columns in E_k, which correspond to the non-zero positions in s_{T}^k, need to be extracted to build $E_k{}'$

$$\min_{\psi_k,s'^k_{\mathrm{T}}}\left\|E_k{}'-\psi_k s'^k_{\mathrm{T}}\right\|_F^2 \quad (22)$$

We can use singular value decomposition (SVD) to decompose the error to update the column

$$E_k{}'=U\Sigma V^{\mathrm{T}} \quad (23)$$

Let $\psi_k=u_1$, where u_1 is the first column of U. Let $s'^k_{\mathrm{T}}=\Sigma_{1,1}v_{\mathrm{T}}^1$, where $\Sigma_{1,1}$ is the first singular value and v_{T}^1 is the first row of V. At last, update s_{T}^k with s'^k_{T}. The weights Ψ are used as output layer weights W_Ψ and trained with the other layers using back-propagation. Algorithm 1 describes the above output layer optimizing process in pseudo-code.

Algorithm 2: the training process of RMVS-DCS

Input: training set $X=\{x_i\}_{i=1}^n, x_i\in\mathbb{R}^L$

Output: reconstruction result $X'=\{x_i'\}_{i=1}^n, x_i'\in\mathbb{R}^L$

Task: find a parameter group $\{W^{(l)},B^{(l)}\}$ by solving $\min_{W^{(l)},B^{(l)}}\mathrm{MSE}(X,X')$

Hyperparameters: epoch of RMVS-DCS n_{epoch}, batch size b,

max iterations of K-SVD n_{iter}, tolerance of K-SVD t

Reshape $X\in\mathbb{R}^{n\times L}\rightarrow X\in\mathbb{R}^{ns\times N}$

Initialize atoms $\Psi,\Psi\in\mathbb{R}^{N\times 2N}$

Update atoms $\Psi_{\mathrm{best}}=update(X,\Psi,n_{\mathrm{iter}},t)$

Initialize network parameters $\{W^{(l-1)},B^{(l-1)}\},W_\Psi=\Psi_{\mathrm{best}}$

For epoch $=1,2,\ldots,n_{\mathrm{epoch}}$ do

  For batch $=1,2,\ldots,n/b$ do

    For each $x_i\in\mathbb{R}^{s\times N}$ in X do

      Measurement matrix $y_i=(W_\Phi x_i{}^{\mathrm{T}})^{\mathrm{T}},W_\Phi\in\mathbb{R}^{M\times N}$

      Linear layer 1 of RM $z_{\mathrm{LL1}}=(W_{\mathrm{LL1}}y_i{}^{\mathrm{T}})^{\mathrm{T}}+B_{\mathrm{LL1}},W_{\mathrm{LL1}}\in\mathbb{R}^{2N\times M}$

      // GLFR blocks

      For block $l=1,2,\ldots,n_{\mathrm{SR\_Block}}$ do

        Self-attention $z_l{}'=F_{\mathrm{SA}}(\mathrm{LN}(z_{l-1}))+z_{l-1}$

        CNN $z_l=F_{\mathrm{CNN}}(z_l{}')+z_l{}'$

      end

      Linear layer 2 of RM $z_{\mathrm{LL2}}=(W_{\mathrm{LL2}}z_l{}^{\mathrm{T}})^{\mathrm{T}}+B_{\mathrm{LL2}},W_{\mathrm{LL2}}\in\mathbb{R}^{2N\times 2N}$

      Dictionary layer $x_i'=(W_\Psi z_{\mathrm{LL2}}{}^{\mathrm{T}})^{\mathrm{T}}$

    end

    Reshape $X'\in\mathbb{R}^{bs\times N}\rightarrow X'\in\mathbb{R}^{b\times L}$

    // Calculate loss and gradient descent

    MSE $L_{\mathrm{MSE}}(X,X'),X'=\{\hat{x}_i\}_{i=1}^b$

    Batch loss $loss(X,X';W^{(l)},B^{(l)})$

    Calculate gradients $\frac{\delta loss}{\delta W^{(l)}},\frac{\delta loss}{\delta B^{(l)}}$

    Update weights $W^{(l)}\leftarrow W^{(l)}-\eta\frac{\delta loss}{\delta W^{(l)}},B^{(l)}\leftarrow B^{(l)}-\eta\frac{\delta loss}{\delta B^{(l)}}$

  end

end

In the network training stage, the back-propagation algorithm optimizes the parameters. We select the mean square error (MSE) as the loss function

$$L_{\mathrm{MSE}}(X,X')=\frac{1}{b\times L}\sum_{i=1}^{b}\sum_{j=1}^{L}(x_{i,j}-x_{i,j}{}')^2, x_{i,j}\in X, x_{i,j}{}'\in X' \quad (24)$$

where $x_{i,j}$ and $x_{i,j}{}'$ are the elements of labels $X\in\mathbb{R}^{b\times L}$ and reconstructed results $X'\in\mathbb{R}^{b\times L}$. Adam is used as the optimizer.

The whole training process is shown in Algorithm 2:

(1) the original data $X\in\mathbb{R}^{n\times L}$ is split into $X\in\mathbb{R}^{ns\times N}$;

(2) use C sub-segments of the samples as atoms to initialize a matrix $\Psi^{(0)}\in\mathbb{R}^{N\times C}$ and update the weights by optimizing Eq. (17);

(3) initialize the parameters of the output layer W_Ψ by Ψ and the other parameters of the network;

(4) the training data X passes through the measurement module and the reconstruction module in turn to obtain the prediction results $X'\in\mathbb{R}^{b\times L}$;

(5) the loss is calculated by $L_{\mathrm{MSE}}(X,X')$, acquiring the gradient $\frac{\delta loss}{\delta W^{(l)}},\frac{\delta loss}{\delta B^{(l)}}$.

(6) update the weights $W^{(l)}\leftarrow W^{(l)}-\eta\frac{\delta loss}{\delta W^{(l)}},B^{(l)}\leftarrow B^{(l)}-\eta\frac{\delta loss}{\delta B^{(l)}}$;

(8) repeat (4), (5), (6) until the loss converges.

3.2. Improved CS reconstruction evaluation method combined with the accuracy of PHM tasks

The traditional CS reconstruction evaluation method, which is still used by most related studies today, only considers the numerical accuracy of the reconstructed signal and uses some statistical characteristics to measure the quality of the

compressed sensing result. It might bring potential dangers to the health status analysis without considering whether the result can effectively adapt to PHM tasks, even causing severe accidents due to misjudgments.

We improve the CS reconstruction evaluation method in the PHM framework to combine statistical characteristics with the accuracy of various PHM tasks. The statistical characteristics used to measure the value accuracy include mean square error (MSE), signal-to-noise ratio (SNR), and cosine similarity (Similarity).

MSE is calculated by Eq. (25)

$$\mathrm{MSE}(X,X')=\frac{1}{n\times L}\sum_{i=1}^{n}\sum_{j=1}^{L}(x_{i,j}-x_{i,j}')^2, x_{i,j}\in X, x_{i,j}'\in X' \tag{25}$$

It can measure the deviation between the predicted and actual values. The smaller the MSE is, the closer the prediction result is to the actual result. Conversely, a larger MSE indicates a more significant prediction error. Since MSE squares the error, it is more sensitive to significant errors. In other words, MSE will increase significantly if some reconstruction values deviate from the true value. The property helps to make it more obvious when evaluating reconstruction accuracy.

SNR is an essential criterion for measuring the quality of the reconstruction signal. It represents the ratio of the actual signal power to the reconstruction noise (error) power

$$\mathrm{SNR}(X,X')=10\log_{10}\left(\sum_{i=1}^{n}\sum_{j=1}^{L}x_{i,j}^2/\sum_{i=1}^{n}\sum_{j=1}^{L}(x_{i,j}-x_{i,j}')^2\right), x_{i,j}\in X, x_{i,j}'\in X' \tag{26}$$

High SNR means the signal quality is good, and the reconstruction noise has little interference with the signal. Low SNR indicates more reconstruction noise components in the signal and poor reconstruction quality.

Similarity is the cosine of the angle between two vectors in the vector space, which is a criterion of the difference between the reconstructed result and the original signal

$$\mathrm{Similarity}(X,X')=\sum_{i=1}^{n}\sum_{j=1}^{L}x_{i,j}x_{i,j}'\Big/\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{L}x_{i,j}^2}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{L}x_{i,j}'^2}, x_{i,j}\in X, x_{i,j}'\in X' \tag{27}$$

The range is [-1, 1]. The closer the result is to 1, the smaller the angle between the reconstructed result and the original signal is in the vector space. When it is closer to -1, there is a larger angle. The cosine distance distinguishes the difference more from the direction and is not sensitive to the absolute value, which is the main reason for choosing it as the criterion.

In the PHM scenario, the collected signals are used to analyze the operating status of the equipment. This is the most essential difference between vibration signal compressed sensing in PHM scenarios and ordinary compressed sensing. In addition to the above evaluation criteria, the reconstruction result is input into the diagnostic model of the subsequent PHM task, and the diagnostic accuracy is included in the criteria to measure the fitness, such as the prediction error of the rotor imbalance value and the classification accuracy of bearing fault diagnosis. The scenario where the compressed sensing algorithm is used determines what PHM task should be used as the evaluation criterion. Assuming that *Acc* is the accuracy of the PHM task

$$Acc=F_{\mathrm{PHMTask}}(X') \tag{28}$$

where F_{PHMTask} is the diagnostic algorithm for the PHM task. We think the following points need to be noted when using the criterion: (1) if necessary, discuss *Acc* for more than one type of the tasks. (2) Compare with a control group that uses raw data as the input. (3) Compare with other CS reconstruction methods.

The improved evaluation method measures the numerical accuracy as well as adaptability of the PHM framework. We will use the method to evaluate the performance of RMVS-DCS on a private rotor imbalance data set and two bearing fault diagnosis data set to prove its advantages.

4. Experimental verification

To evaluate the proposed method, we conduct experiments on a private rotor data set and two bearing data sets.

4.1. Experiments on a private rotor imbalance data set

4.1.1. Experiment platform and data set

Rotor imbalance is a common equipment problem. Some scholars have studied it through vibration signals [41]. Fig. 3 is an experiment platform of the rotor, which includes a motor, a rotor, a sensor, and data acquisition equipment. We acquire a private data set of the vibration signal by it. The rotor speeds are 1200 RPM and 1500 RPM. The values of imbalance are from 1 to 12.1 g. The sampling rate is 5.12 kHz, and sampling time is 0.8 s.

We get the samples of all working conditions without overlapping. Based on the rotational speed and the sampling frequency, a single period can contain a maximum of 256 sampling points. To reduce the feature loss caused by incomplete periods, the sample length must be greater than 256, and we select 512 in the experiment. Then, divide them into the train set, validation set, and test set according to the proportion of 6:2:2. There are 6691, 2230, and 2231 samples, respectively.

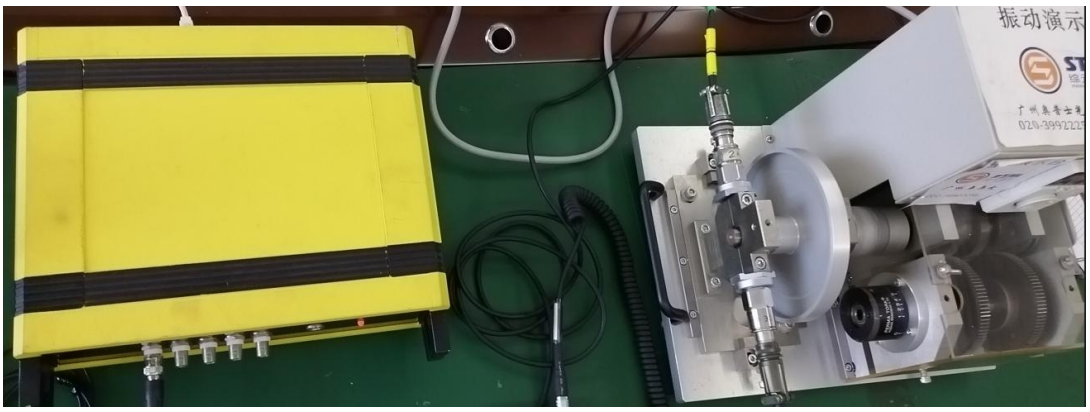

Fig. 3 Rotor test platform.

All the experiments are performed on a workstation. The CPU is 11th Gen Intel (R) Core (TM) i5-11500 @ 2.70 GHz 2.71 GHz, and the GPU is NVIDIA GeForce RTX 3060.

4.1.2. Ablation experiment

Before the experiment, we tune the primary hyperparameters of RMVS-DCS. The model converges when the epochs are set to 150 during training. The basic settings are shown in Table 1.

Table 1. Basic settings of RMVS-DCS.

Component name	Main parameters	Output dimensions
Measurement matrix	In channels = 32, bias = false	(64, 16, 32×*CR*)
1st linear projection layer	In channels = 32×*CR*	(64, 16, 64)
GLFR block	Self attention head = 1, Convolutional kernel size = [7, 3, 1] Convolutional channels = [32, 64, 1]	(64, 16, 64)
2nd linear projection layer	In channels = 64	(64, 16, 64)
Output layer optimized by new strategy	epochs of train = 20	(64, 16, 32)

Table 2. Basic information of each group in the ablation experiment.

Group	Measurement	Reconstruction	Output layer
RMVS-DCS	Learnable	GLFR block	New optimizing strategy
A	Gaussian		
B	Bernoulli		
C	Hadamard		
D		TEL dim_feedforward = 128	
E		TEL dim_feedforward = 256	
F		TEL dim_feedforward = 512	
G			Back-propagation

As shown in Table 2, to prove the effectiveness of the components in the model, we conduct an ablation experiment on the learnable measurement matrix, GLFR block, and output layer optimized by the improved strategy:

(1) select the measurement matrices commonly used in classic compressed sensing (Gaussian, Bernoulli, and Hadamard matrices) to compare with the learnable measurement matrix;

(2) replace GLFR blocks with several Transformer encoder layers (TEL) [39] that have different dimensions of the feed-forward layers;

(3) use the learned weights obtained by back-propagation to replace the current output layer.

Fig. 4 shows each group's MSE, SNR, and Similarity in the

ablation experiment. Compared with A, B, and C, the reconstruction accuracy of RMVS-DCS is significantly higher. Similarity is 0.7010 when *CR* is 0.125 and 0.9947 when *CR* is 0.875, indicating that the weights obtained by back-propagation can generate the measurement vector preserving more vital reconstruction information than those random matrices, which is more suitable for the model. The Hadamard matrix performs the worst, and its Similarity is 0.4734 when *CR* is 0.125. When *CR* is larger than 0.375, the reconstruction accuracies of D, E, and F decrease substantially (e.g., $\mathrm{SNR}_{0.375}^{\mathrm{E}} = 7.7620 \rightarrow \mathrm{SNR}_{0.5}^{\mathrm{E}} = 0.0559$). The results indicate that training the Transformer encoder layer as the main structure at the reconstruction stage is challenging. At the same time, the GLFR block is more accessible for training and obtaining stable reconstruction results due to alternative global information fusion and local feature learning. When comparing the output layers, the reconstruction accuracies of RMVS-DCS are better than those of G (e.g., $\mathrm{SNR}_{0.125}^{\text{RMVS-DCS}}\text{-}\mathrm{SNR}_{0.125}^{\mathrm{G}} = 0.5197$, $\mathrm{SNR}_{0.875}^{\text{RMVS-DCS}}\text{-}\mathrm{SNR}_{0.875}^{\mathrm{G}} = 1.9930$). It shows that introducing the signal structure prior by the proposed optimizing strategy is better than training entirely relying on back-propagation. We speculate the strategy can give better initial parameters to the network and restrict the space of solutions, improving generalization performance to get a better effect.

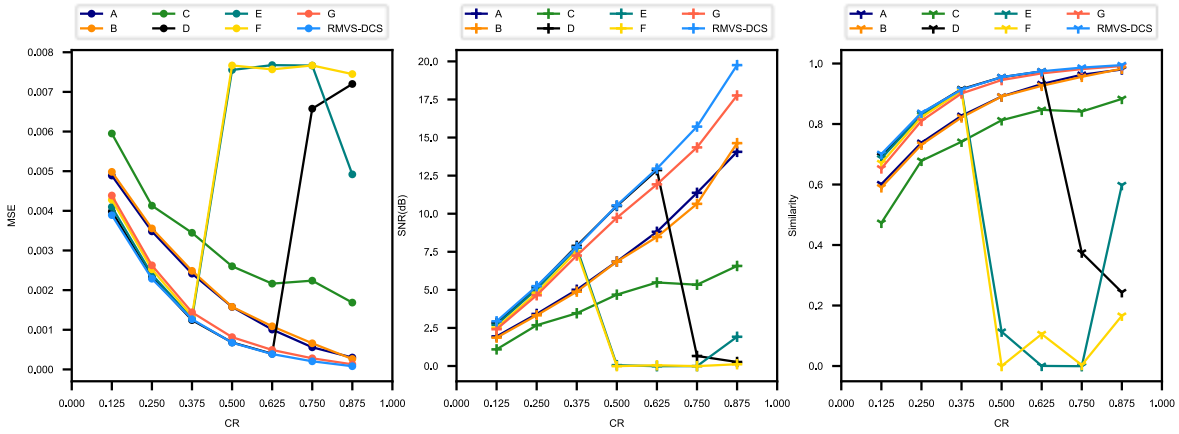

Fig. 4. Results of the ablation experiment on the rotor data set.

4.1.3. Comparison experiment

To prove the superiority of the method, we choose several classical and state-of-the-art (SOTA) methods from greedy, convex optimization, Bayesian, and DCS methods as comparison groups: (1) OMP [42], (2) VSFBP [32], (3) BP [13], (4) IISTA [43], (5) BSBL [31], (6) ISTA-Net [24], (7) DCSNet [35]. The results of the reconstruction are shown in Fig. 5.

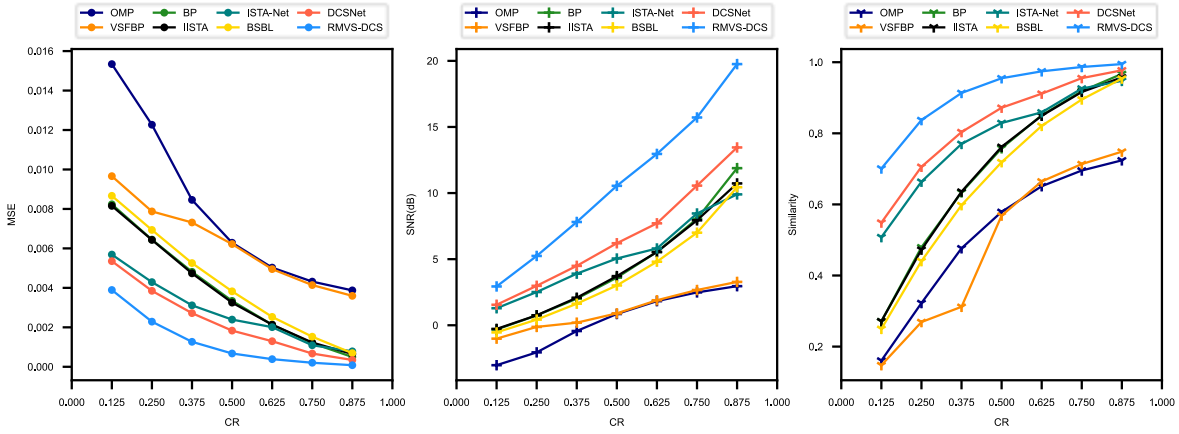

Fig. 5. Results of the comparison experiment on the rotor data set.

The results show that the greedy methods (OMP, VSFBP) do not perform well in the task. Similarity of OMP is 0.1607 when *CR* is 0.125, which is lower than $\mathrm{Similarity}_{0.125}^{\text{RMVS-DCS}}$. The convex optimization methods (BP and IISTA) are in the middle.

SNR are 3.6007 and 3.7160 when *CR* is 0.5. The best performance is achieved by the DCS methods: ISTA-Net, DCSNet, and RMVS-DCS. ISTA-Net performs worse than the other two methods, although it has better interpretability. Between the other two methods, RMVS-DCS performs better than DCSNet ($\mathrm{SNR}_{0.125}^{\text{RMVS-DCS}}$-$\mathrm{SNR}_{0.125}^{\text{DCSNet}} = 1.3853$, $\mathrm{SNR}_{0.875}^{\text{RMVS-DCS}}$-$\mathrm{SNR}_{0.875}^{\text{DCSNet}} = 6.3056$), which proves that the proposed method has superior reconstruction accuracy in the task.

Fig. 6 shows some visual reconstruction results of RMVS-DCS. Its reconstruction results not only perform superior in the statistical characteristics but also have minimal reconstruction errors for most sample points.

To evaluate their suitability in PHM scenarios, we use the raw data and the reconstruction results of the above methods as samples, inputting them into the multi-layer perceptron to predict the imbalance value of the rotor. The network structure and hyperparameters are shown in Table 3.

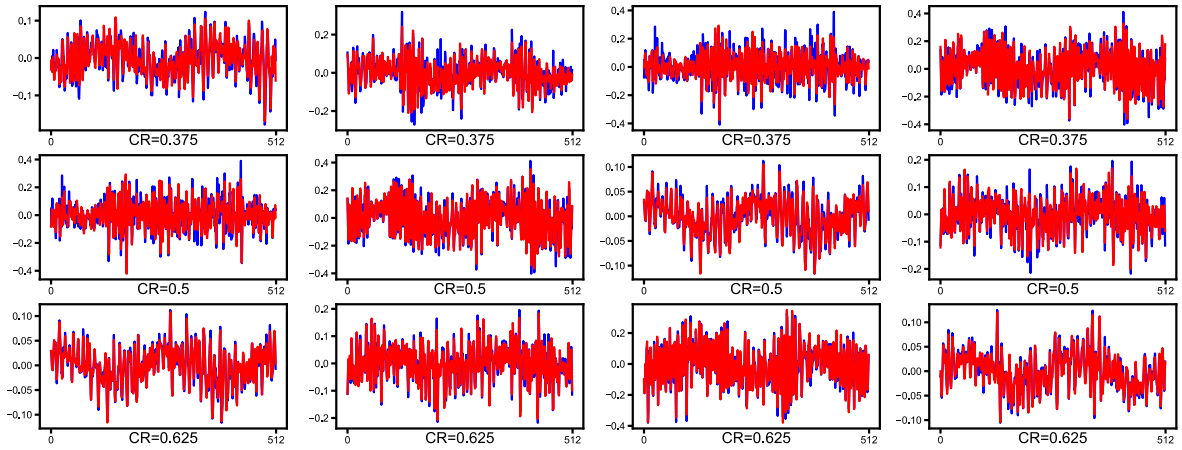

Fig. 6. Some reconstruction results of RMVS-DCS (red) and their original signals (blue) on the rotor data set.

The experiment results are shown in Fig. 7. The raw data prediction error is 0.0944 g. The prediction errors of RMVS-DCS under *CR* in the figure are 0.1913 g, 0.1717 g, 0.1518 g, and 0.1092 g. When *CR* is 0.875, the prediction gap between the reconstructed signal of the proposed method and the actual signal is only 0.0147 g. It shows that the proposed method has less critical information for the imbalance value prediction lost in the compression-reconstruction process. At low *CR*, it has obvious advantages over the second model (DCSNet) which has a prediction error of 1.1137 g when *CR* is 0.5. Meanwhile, OMP has the most significant prediction error (2.9346 g). ISTA-Net is behind RMVS-DCS and DCSNet in error (1.3945 g). Its accuracy is only higher than OMP (1.5382 g) and VSFBP (1.3391 g) when *CR* is 0.875 and exceeded by BP (0.3658 g), IISTA (0.3806 g) and BSBL (0.3646 g). We think that the manual design in ISTA-Net might lose more critical information about the rotor running status.

Table 3. The network structure and hyperparameters in the PHM tasks.

Components	Parameters	Other settings
1st linear layer	Input size = 512	Batchnorm = True
	Output size = 256	Activation = RELU
2nd linear layer	Input size = 256	Batchnorm = True
	Output size = 128	Activation = RELU
3rd linear layer	Input size = 128	Batchnorm = True
	Output size = 64	Activation = RELU
4th linear layer	Input size = 64	Batchnorm = True
	Output size = 32	Activation = RELU
5th linear layer	Input size = 32	Batchnorm = True
	Output size = 8	Activation = RELU
Output layer	Input size = 8	
	Output size = 1	
Training settings Epochs = 100, Batch size = 64, Optimizer = Adam		

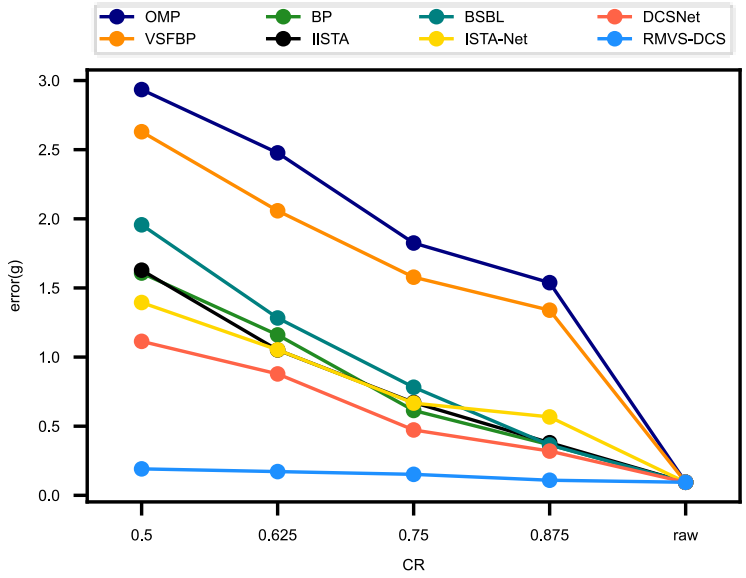

Fig. 7. Experiment results of the rotor imbalance prediction using original and reconstructed signals

4.2. Experiments on the Case Western Reserve University bearing data set

Rotating machinery fault diagnosis is an important PHM task in actual industrial scenarios, and it often uses vibration signals as the input for diagnosis [44]. In this section, we perform several experiments using the Case Western Reserve University (CWRU) bearing data set.

4.2.1. Experiment platform and data set

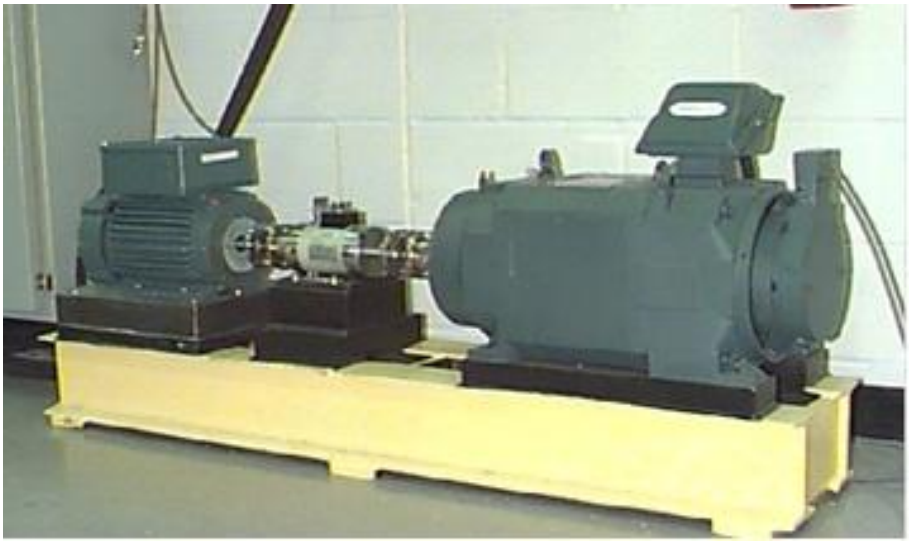

Fig. 8. CWRU bearing data set acquisition platform.

The CWRU bearing data set is commonly used in compressed sensing and PHM tasks of rotating machinery [2, 45, 46]. Its acquisition platform is shown in Fig. 8. The platform is driven by a 2 hp motor. SKF6205 and SKF6203 bearings are installed at the drive end (DE) and fan end (FE) with sampling frequencies of 12kHz/48kHz and 12kHz. Different loads (0 hp, 1 hp, 2 hp, 3 hp) and speeds (from 1730 RPM to 1797 RPM) are set during the sampling to simulate the variable operating condition.

We choose FE data, which contains inner ring fault (IF), rolling element fault (REF), and outer ring fault (OF). The faults are 0.007 inches, 0.014 inches, and 0.021 inches. The samples in the data set have a length of 512 (a single period can contain a maximum of 423 sampling points) and are divided into train, validation, and test sets with a proportion of 6:2:2.

4.2.2. Comparison Experiment

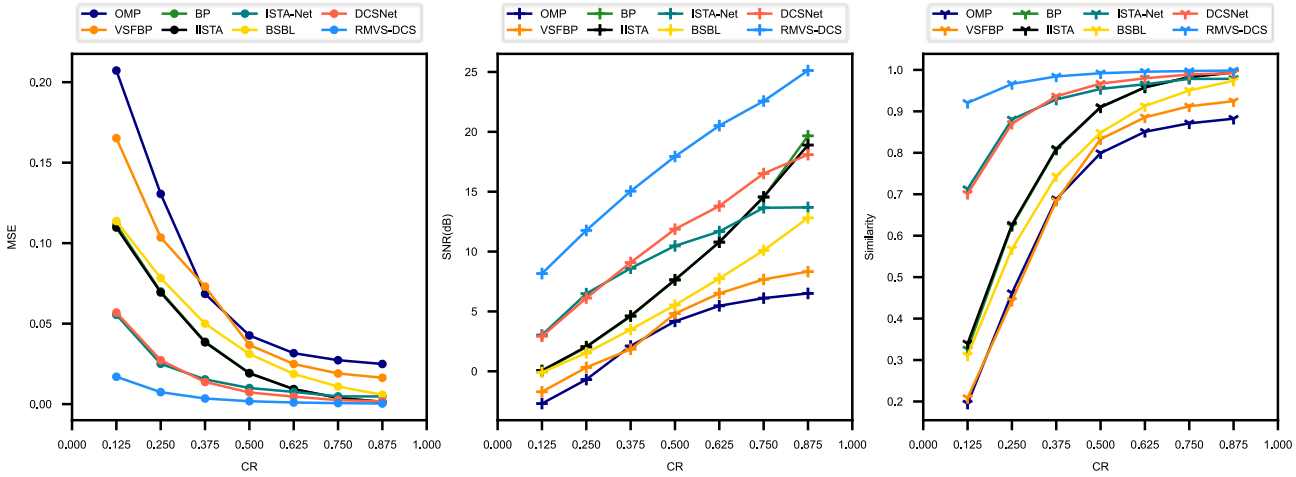

Fig. 9. Comparison experiment results on the CWRU bearing data set.

In the experiment, the models used for comparison are the same as in Section 4: (1) OMP, (2) VSFBP, (3) BP, (4) IISTA, (5) BSBL, (6) ISTA-Net, (7) DCSNet.

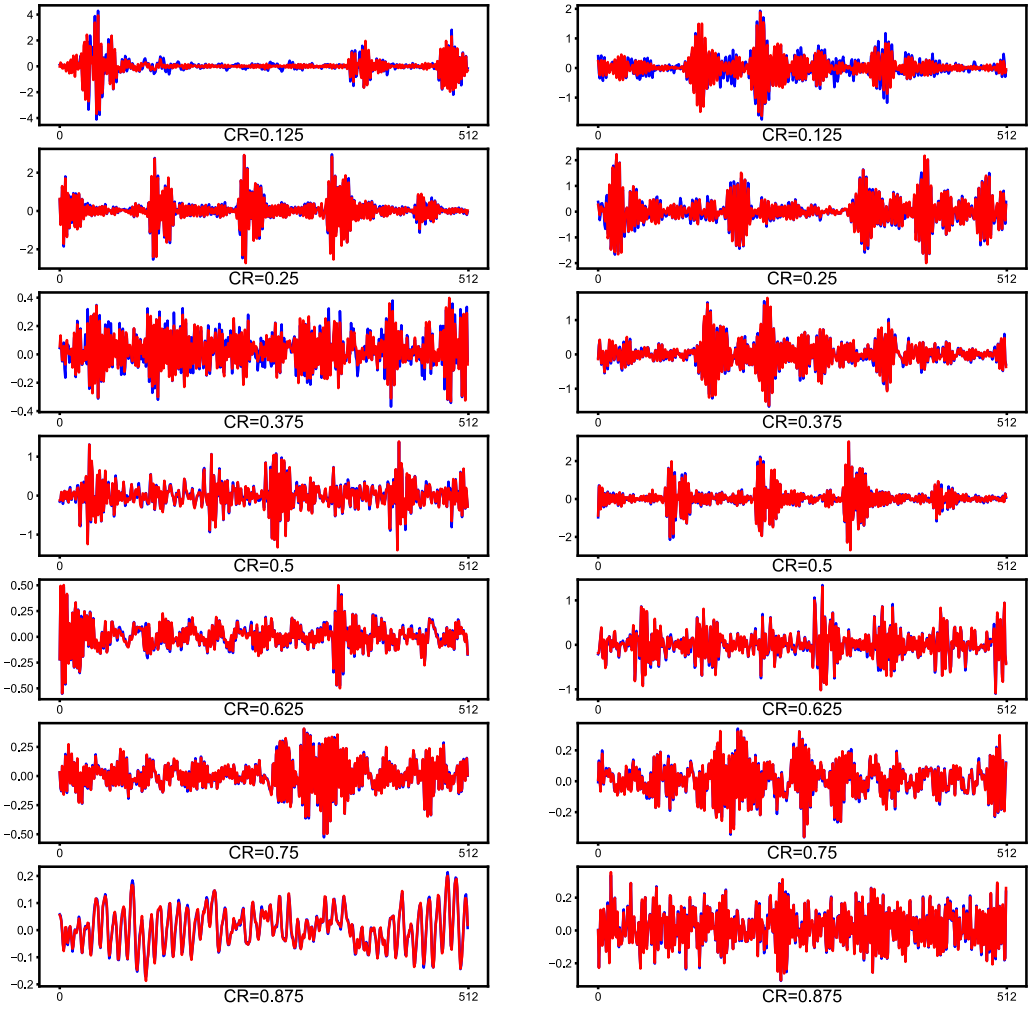

Fig. 10. Some reconstruction results of RMVS-DCS (red) and their original signals (blue) on the CWRU bearing data set.

MSE, SNR, and Similarity of the bearing vibration signal reconstruction are shown in Fig. 9. The numerical reconstruction accuracies of the matching pursuit methods are still not as good as those of the other methods. Similarity of VSFBP is 0.2085 when *CR* is 0.125 and 0.9245 when *CR* is 0.875. The reconstruction accuracies of BP and IISTA are worse than those of ISTA-Net and DCSNet, except when *CR* is 0.875, MSE of the former two are 0.0012 and 0.0014, which are better than those of the latter two (0.0048 and 0.0017). Among the three DCS methods, RMVS-DCS has more obvious advantages than ISTA-Net and DCSNet. Similarity is 0.9205, 0.9661, 0.9842, 0.9919, 0.9956, 0.9972, 0.9985, respectively. It is higher than 0.7126 of ISTA-Net and 0.7010 of DCSNet when *CR* is 0.125. These experiment results show that RMVS-DCS has better numerical reconstruction accuracy than the above matching pursuit, convex relaxation, Bayesian, and DCS methods in the compressed sensing task of the bearing vibration signal.

Fig. 10 shows the reconstructed signals by RMVS-DCS compared to the original signals under various *CR*. When *CR* < 0.25, there is a small gap between the original and reconstructed waveforms, but for larger *CR*, the gap almost disappears. Therefore, if it is necessary to save more accurate bearing vibration data under actual production conditions, a value larger than 0.125 can be used, such as 0.25, which is an ideal value to achieve the balance between *CR* and accuracy.

Unlike the prediction task in Section 4, a fault classification task is introduced in the following. We input the raw data and the reconstruction results generated by each method into the multi-layer perceptron, and the experiment results are shown in Table 4.

Table 4. Results of the bearing fault diagnosis experiment on the CWRU data set.

Method	Accuracy under various *CR* (%)							
	0.125	0.25	0.375	0.5	0.625	0.75	0.875	Raw
OMP	75.63	91.48	97.95	99.26	99.26	99.60	99.83	100.00
VSFBP	74.77	90.57	98.18	99.32	99.60	99.77	99.94	
BP	84.49	96.99	99.38	99.60	100.00	100.00	100.00	
IISTA	86.19	96.70	99.09	99.83	99.94	100.00	100.00	
BSBL	84.49	96.48	99.15	99.77	100.00	100.00	100.00	
ISTA-Net	89.60	98.30	99.55	100.00	100.00	100.00	100.00	
DCSNet	92.33	99.15	99.89	100.00	100.00	100.00	100.00	
RMVS-DCS	98.69	99.83	100.00	100.00	100.00	100.00	100.00	

In the fault classification experiment, our diagnosis network architecture is the same as in Table 3, except we change the last two layer dimensions (the output size of the 5th linear layer is 16, while the input and output sizes of the output layer are 16 and 9) and add a softmax layer. The classification accuracy of the raw data is 100.00%. For OMP and VSFBP, the fault classification accuracies are 75.63% and 74.77% when *CR* is 0.125. At the same *CR*, the recognition accuracies of BP, IISTA, BSBL, and ISTA-Net are between 80.00% and 90.00%, while DCSNet and RMVS-DCS exceed 90.00%. For RMVS-DCS, the recognition accuracies of the reconstruction results can reach 100.00% when $CR \geq 0.375$, DCSNet and ISTA-Net need to meet the condition of $CR \geq 0.5$. The other models require a larger *CR*. Even they are out of reach. The experiment results prove that the compression-reconstruction bearing vibration data of RMVS-DCS can provide inputs for fault diagnosis models to obtain high-accuracy fault classification results under various *CR* conditions, which is suitable for long-term bearing monitoring.

4.3. Experiments on the Shandong University bearing data set

4.3.1. Experiment platform and data set

The acquisition platform for the Shandong University (SDU) bearing data set consists of the following devices: a motor with its control system, a shaft, bearing housings (including bearings), a radial force loading device, a data acquisition instrument, a laptop, etc. Two acceleration sensors are installed in horizontal and vertical directions on the bearing housing.

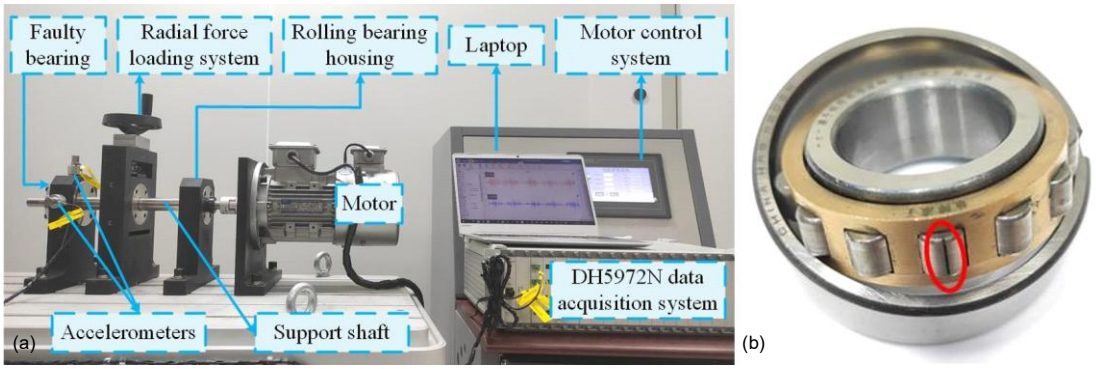

Fig. 11. SDU bearing data set acquisition platform and one of the faulty bearings. (a) platform, (b) faulty bearing.

Different from the CWRU bearing data set, it contains single fault conditions (OF, IF, REF, IF') as well as compound fault conditions (IOF: IF+OF, IRF: IF+REF, ORF: OF+REF, IORF: IF+OF+REF) with speed at 1750 RPM, 2000 RPM, and 2250 RPM, respectively. The sampling frequency is 12.8 kHz. The sampling segment length is 16384 points, and the sampling time is 1 minute. The SDU bearing data set acquisition platform and one of the faulty bearings are shown in Fig. 11.

4.3.2. Comparison Experiments

We conduct comparison experiments on the SDU bearing data set. We divide the samples, and each has a length of 512, because the maximum number of sampling points for one period is approximately 439. There are 13398 train samples, 4666 validation samples, and 4666 test samples. Table 5 shows MSE of the reconstruction results under various *CR*. When *CR* is 0.125, Only RMVS-DCS has an MSE below 5.0000. Those of OMP, VSFBP, BP, IISTA, and BSBL even exceed 10.0000. When *CR* is 0.875, MSE of RMVS-DCS is 0.0258, which is 11.81% of the 2nd best method.

Table 5. MSE of the reconstruction on the SDU bearing data set.

Method	MSE under various *CR*						
	0.125	0.25	0.375	0.5	0.625	0.75	0.875
OMP	20.2147	14.0797	9.2335	6.8766	5.6841	4.8839	4.4814
VSFBP	15.6834	10.4459	8.6243	7.5775	7.1295	6.8635	6.7020
BP	10.6958	7.9562	5.5665	3.7573	2.3650	1.3129	0.5072
IISTA	10.7820	7.8359	5.5530	3.7014	2.3796	1.2884	0.4995
BSBL	11.0212	8.5829	6.4266	4.6434	3.2082	2.0660	1.0343
ISTA-Net	7.1695	4.5189	3.0434	2.0983	2.0243	1.1813	0.7301
DCSNet	6.5915	4.3402	2.9354	1.9725	1.1992	0.6248	0.2185
RMVS-DCS	4.1242	2.3145	1.2425	0.7648	0.4178	0.1859	0.0258

Table 6. SNR of the reconstruction on the SDU bearing data set.

Method	SNR under various *CR*						
	0.125	0.25	0.375	0.5	0.625	0.75	0.875
OMP	-2.8688	-1.2980	0.5342	1.8141	2.6412	3.3002	3.6737
VSFBP	-1.7665	-0.0016	0.8306	1.3926	1.6573	1.8224	1.9258
BP	-0.1043	1.1808	2.7320	4.4391	6.4495	9.0055	13.1362
IISTA	-0.1391	1.2470	2.7426	4.5043	6.4229	9.0873	13.2027
BSBL	-0.2344	0.8515	2.1081	3.5195	5.1253	7.0365	10.0412
ISTA-Net	1.6330	3.6376	5.3543	6.9693	7.1252	9.4643	11.5539
DCSNet	1.9981	3.8128	5.5112	7.2377	9.3991	12.2307	16.7940
RMVS-DCS	4.0345	6.5434	9.2450	11.3523	13.9779	17.4949	26.0742

Table 6 shows SNR of the reconstruction results under various *CR*. SNR of OMP and VSFBP have negative values at

CR = 0.125 and *CR* = 0.25. SNR of BP, IISTA, and BSBL have negative values at *CR* = 0.125. The occurrence of negative values indicates that the reconstruction noise's power even exceeds the signal's power. Compared with ISTA-Net and DCS-Net, RMVS-DCS has apparent advantages, and its SNR range is from 4.0345 to 26.0742 under various *CR*. It indicates that there are fewer noise components in the signal reconstructed by RVMS-DCS, and the energy of the reconstructed result is closer to the raw signal.

Table 7 shows Similarity of the reconstruction results under various *CR*. When *CR* is 0.125, the angle between the reconstructed result of OMP and the raw signal in the vector space is about 79.39°. The direction gap is relatively large. When *CR* is 0.5, the angles of the other methods are larger than 25°, while that of the proposed method is only 15.70°. The angle is reduced to 2.81° when CR is 0.875. This indicates that the reconstructed result of the proposed method is similar to the raw signal in the vector space.

We randomly visualize two reconstructed results for each *CR* condition and compare them with the raw signals. The results are shown in Fig. 12. The reconstruction accuracy of the signal is positively correlated with *CR*. The lower *CR* causes more loss of structure information. When *CR* is 0.125, the proposed method still recovers most of the waveform, but there is a certain error in predicting the large peaks. When $CR \geq 0.25$, the recovery accuracy becomes very reliable. 0.25 is recommended to balance *CR* and reconstruction accuracies.

Table 7. Similarity of the reconstruction on the SDU bearing data set.

Method	Similarity under various *CR*						
	0.125	0.25	0.375	0.5	0.625	0.75	0.875
OMP	0.1842	0.3932	0.5565	0.6532	0.7070	0.7457	0.7658
VSFBP	0.2052	0.3977	0.4869	0.5489	0.5762	0.5935	0.6039
BP	0.3123	0.5337	0.6941	0.8029	0.8800	0.9351	0.9754
IISTA	0.3000	0.5414	0.6949	0.8060	0.8792	0.9363	0.9758
BSBL	0.2838	0.4858	0.6369	0.7493	0.8330	0.8957	0.9492
ISTA-Net	0.5618	0.7540	0.8418	0.8940	0.8979	0.9418	0.9644
DCSNet	0.6074	0.7645	0.8479	0.9011	0.9409	0.9696	0.9895
RMVS-DCS	0.7778	0.8822	0.9386	0.9627	0.9798	0.9911	0.9988

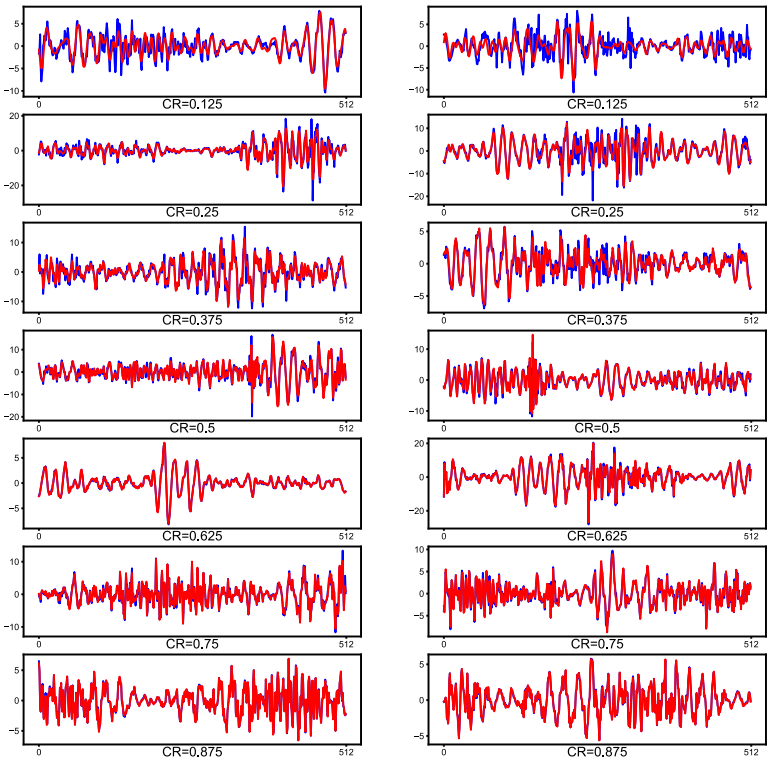

Fig. 12. Some reconstruction results of RMVS-DCS (red) and their original signals (blue) on the SDU bearing data set.

We send the results of the reconstruction experiment to the diagnostic model for fault detection. The reconstruction results are shown in Table 8. When *CR* is 0.125, RMVS-DCS has the highest accuracy (76.66%) and 17.27% higher than the next-best model. At the same time, those of OMP, VSFBP, BP, IISTA, and BSBL do not exceed 50.00%, which indicates that these methods do not restore the fault features of the highly compressed signal well. When *CR* is between 0.25 and 0.75, the accuracy of the proposed method is still the highest. When the *CR* is 0.875, it (98.65%) even exceeds the raw signal group (98.39%). We infer this is because the proposed model actually forms a set of filters. They filter the irrelevant noise, making the fault features easier to identify. This is not the case for the other models. Through the experiment, we can conclude that the proposed framework is suitable for single bearing fault scenarios and has a solid recovery ability for the signal with mixed faults.

Table 8. Results of the bearing fault diagnosis experiment on the SDU data set.

Method	Accuracy under various *CR* (%)							
	0.125	0.25	0.375	0.5	0.625	0.75	0.875	Raw
OMP	34.46	54.37	68.11	78.20	82.62	84.91	88.04	98.39
VSFBP	33.99	52.89	62.34	68.28	70.47	71.24	74.58	
BP	42.09	64.92	79.45	88.51	92.54	95.61	97.45	
IISTA	42.65	64.62	79.88	88.36	92.76	96.10	97.32	
BSBL	41.04	61.83	76.32	85.98	90.89	93.87	96.12	
ISTA-Net	52.57	75.44	86.76	91.60	93.01	95.20	96.66	
DCSNet	59.39	77.95	88.13	92.35	95.33	96.91	98.29	
RMVS-DCS	76.66	89.20	94.45	95.48	97.21	98.03	98.65	

5. Conclusions and prospects

We propose a DCS framework (RMVS-DCS) based on global-local feature alternating reconstruction and signal structure prior, and a dual evaluation method that combines reconstruction and PHM task accuracy. We evaluate the framework on three data sets, and the conclusions are as follows:

(1) The reconstruction SNRs on the rotor data set are 2.9357, 5.2431, 7.8123, 10.5474, 12.9601, 15.7178, and 19.7556 (compression ratio: from 0.125 to 0.875). When using other modules replace the learnable measurement matrix, GLFR blocks, and the pre-trained output layer, the numerical accuracy decreased to varying degrees, even does not converge. It proves the effectiveness of the modules in improving reconstruction accuracy and convergence stability.

(2) The proposed method has advantages over the comparison method in numerical and PHM task accuracy, e.g., on the rotor unbalance data set, its reconstruction and prediction error are reduced by 82.60% and 89.90% (compression ratio is 0.5). The results demonstrate its advanced reconstruction ability and PHM compatibility.

(3) The improved evaluation method indicates the quality of the reconstruction results for the PHM task and is more suitable for evaluating and selecting the CS method in PHM scenarios.

Next, we will extend the framework to multi-modal data fusion compressed sensing, such as vibration-sound and vibration-temperature. This will further broaden its industrial applicability range.

References

1. Liu Y, Huang J, Jia M. Knowledge Distillation - Based Zero - Shot Learning for Process Fault Diagnosis. Advanced Intelligent Systems. 2025;7(6):2400828, https://doi.org/10.1002/aisy.202400828.
2. Russell M, Wang P. Physics-informed deep learning for signal compression and reconstruction of big data in industrial condition monitoring. Mechanical Systems and Signal Processing. 2022;168:108709, https://doi.org/10.1016/j.ymssp.2021.108709.
3. Cui L, Tian X, Shi X, Wang X, Cui Y. A Semi-Supervised Fault Diagnosis Method Based on Improved Bidirectional Generative Adversarial Network. Applied Sciences. 2021;11(20), https://doi.org/10.3390/app11209401.
4. Wang Z, Liang P, Bai R, Liu Y, Zhao J, Yao L, et al. Few-shot fault diagnosis for machinery using multi-scale perception multi-level feature

fusion image quadrant entropy. Advanced Engineering Informatics. 2025;63:102972, https://doi.org/10.1016/j.aei.2024.102972.
5. Shu X, Zhang S, Li Y, Chen M. An anomaly detection method based on random convolutional kernel and isolation forest for equipment state monitoring. Eksploatacja i Niezawodność – Maintenance and Reliability. 2022;24(4):758-70, https://doi.org/10.17531/ein.2022.4.16.
6. Zhang T, Wang H. Explainable remaining useful life uncertainty prediction method for rolling bearing. Engineering Applications of Artificial Intelligence. 2025;161:112114, https://doi.org/10.1016/j.engappai.2025.112114.
7. Lei X, Lu N, Jiang B, Wang C, Chen C. A Multi-scale Attention Mechanism Diagnosis Method with Adaptive Online Updating Based on Deep Learning under Variable Working Conditions. Eksploatacja i Niezawodność – Maintenance and Reliability. 2025;27(1), https://doi.org/10.17531/ein/192975.
8. Zhu J, Li O, Chen M, Hu B, Ma E. Rolling bearing fault diagnosis method based on adaptive signal diagnosis network and its application. Eksploatacja i Niezawodność – Maintenance and Reliability. 2025;27(2), https://doi.org/10.17531/ein/194673.
9. Deng L, Lin H, Liu Z, Wang H. Compressed feature reconstruction for localized fault diagnosis with generalized minimax-concave penalty. Measurement. 2022;200, https://doi.org/10.1016/j.measurement.2022.111622.
10. Han P, Huang Z, Li W, He W, Cao Y. Multi-sensor bearing fault diagnosis based on evidential neural network with sensor weights and reliability. Expert Systems with Applications. 2025;269:126533, https://doi.org/10.1016/j.eswa.2025.126533.
11. Wang Z, Liu Y, Bai R, Chen H, Li J, Chen X, et al. Multi-modal multi-scale multi-level fusion quadrant entropy for mechanical fault diagnosis. Expert Systems with Applications. 2025;281:127715, https://doi.org/10.1016/j.eswa.2025.127715.
12. Tao L, Liu H, Ning G, Cao W, Huang B, Lu C. LLM-based framework for bearing fault diagnosis. Mechanical Systems and Signal Processing. 2025;224:112127, https://doi.org/10.1016/j.ymssp.2024.112127.
13. Donoho DL. Compressed sensing. IEEE Transactions on Information Theory. 2006;52(4):1289-306, https://doi.org/10.1109/TIT.2006.871582.
14. Xiao C, Tang H, editors. Adaptive compressive sensing method based on optimal classification for bearing vibration signals in axial piston pump. 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC); 2019: IEEE, https://doi.org/10.1109/SDPC.2019.00051.
15. Shi W, Jiang F, Liu S, Zhao D. Image Compressed Sensing using Convolutional Neural Network. IEEE Trans Image Process. 2019, https://doi.org/10.1109/TIP.2019.2928136.
16. Grzonka J, Marqueses-Rodríguez J, Fernández-García S, Chen X, Calvino JJ, López-Haro M. Combining Deep Learning and Compressed Sensing Methods for the 3D Characterization of Ultra‐Thin Epitaxial Layers Grown on Controlled‐Shape Nano‐Oxides. Advanced Intelligent Systems. 2023;5(3):2200231, https://doi.org/10.1002/aisy.202200231.
17. Yu Y, Jiang S, Timmerman R, Peng H. Leveraging Compressed Sensing and Radiomics for Robust Feature Selection for Outcome Prediction in Personalized Ultra‐Fractionated Stereotactic Adaptive Radiotherapy. Advanced Intelligent Systems. 2025:2500116, https://doi.org/10.1002/aisy.202500116.
18. Wang C, Zhang Y, Sun L, Han J, Chao L, Yan L. Improved sparsity adaptive matching pursuit algorithm based on compressed sensing. Displays. 2023;77:102396, https://doi.org/10.1016/j.displa.2023.102396.
19. Wei J, Mao S, Dai J, Wang Z, Huang W, Yu Y. A faster and more accurate iterative threshold algorithm for signal reconstruction in compressed sensing. Sensors. 2022;22(11):4218, https://doi.org/10.3390/s22114218.
20. Lin A, Song AH, Bilgic B, Ba D. Covariance-Free Sparse Bayesian Learning. IEEE Transactions on Signal Processing. 2022;70:3818-31, https://doi.org/10.1109/TSP.2022.3186185.
21. Bora A, Jalal A, Price E, Dimakis AG. Compressed Sensing using Generative Models. In: Doina P, Yee Whye T, editors. Proceedings of the 34th International Conference on Machine Learning; Proceedings of Machine Learning Research: PMLR; 2017. p. 537--46, https://doi.org/10.48550/arXiv.1703.03208.
22. Wu Y, Rosca M, Lillicrap T. Deep Compressed Sensing. In: Kamalika C, Ruslan S, editors. Proceedings of the 36th International Conference on Machine Learning; Proceedings of Machine Learning Research: PMLR; 2019. p. 6850--60, https://doi.org/10.48550/arXiv.1905.06723.
23. Kulkarni K, Lohit S, Turaga P, Kerviche R, Ashok A, editors. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016,

https://doi.org/10.48550/arXiv.1601.06892.
24. Zhang J, Ghanem B, editors. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018, https://doi.org/10.48550/arXiv.1706.07929.
25. Cui S, Dai J, Zhao S, Luo H. ISTOF-Net: An ISTA-Based Deep Unfolding Network With Optimized Feature Aggregation Architecture for Image Compressed Sensing. IEEE Sensors Journal. 2025;25(15):29271-83, https://doi.org/10.1109/JSEN.2025.3583291.
26. Hu ZX, Wang Y, Ge MF, Liu J. Data-Driven Fault Diagnosis Method Based on Compressed Sensing and Improved Multiscale Network. IEEE Transactions on Industrial Electronics. 2020;67(4):3216-25, https://doi.org/10.1109/TIE.2019.2912763.
27. Shao H, Jiang H, Zhang H, Duan W, Liang T, Wu S. Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing. Mechanical Systems and Signal Processing. 2018;100:743-65, https://doi.org/10.1016/j.ymssp.2017.08.002.
28. Yuan H, Lu C. Rolling bearing fault diagnosis under fluctuant conditions based on compressed sensing. Structural Control and Health Monitoring. 2017;24(5), https://doi.org/10.1002/stc.1918.
29. Guo W, Tse PW. A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals. Journal of Sound and Vibration. 2013;332(2):423-41, https://doi.org/10.1016/j.jsv.2012.08.017.
30. Li X, Bossmann F, Ma J. Optimal Compressed Sensing Reconstruction for Vibration Monitor Data Using Deep Learning. IEEE Transactions on Instrumentation and Measurement. 2025;74:1-14, https://doi.org/10.1109/TIM.2025.3571153.
31. Sun J, Yu Y, Wen J. Compressed-Sensing Reconstruction Based on Block Sparse Bayesian Learning in Bearing-Condition Monitoring. Sensors. 2017;17(6):1454, https://doi.org/10.3390/s17061454.
32. Pan Z, Meng Z, Zhang Y, Zhang G, Pang X. High-precision bearing signal recovery based on signal fusion and variable stepsize forward-backward pursuit. Mechanical Systems and Signal Processing. 2021;157:107647, https://doi.org/10.1016/j.ymssp.2021.107647.
33. Wang Q, Meng C, Ma WN, Wang C, Yu L. Compressive sensing reconstruction for vibration signals based on the improved fast iterative shrinkage-thresholding algorithm. Measurement. 2019;142:68-78, https://doi.org/10.1016/j.measurement.2019.04.012.
34. Wang H, Yang S, Liu Y, Li Q. Compressive sensing reconstruction for rolling bearing vibration signal based on improved iterative soft thresholding algorithm. Measurement. 2023;210, https://doi.org/10.1016/j.measurement.2023.112528.
35. Zhang M, Zhang H, Yuan D, Zhang M. Learning-Based Sparse Data Reconstruction for Compressed Data Aggregation in IoT Networks. IEEE Internet of Things Journal. 2021;8(14):11732-42, https://doi.org/10.1109/JIOT.2021.3059735.
36. Baraniuk R, Davenport M, DeVore R, Wakin M. A simple proof of the restricted isometry property for random matrices. Constructive approximation. 2008;28:253-63, https://doi.org/10.1007/s00365-007-9003-x.
37. Ma Y, Jia X, Bai H, Liu G, Wang G, Guo C, et al. A new fault diagnosis method based on convolutional neural network and compressive sensing. Journal of Mechanical Science and Technology. 2019;33(11):5177-88, https://doi.org/10.1007/s12206-019-1007-5.
38. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. Advances in Neural Information Processing Systems 30 (Nips 2017). 2017;30, https://doi.org/10.48550/arXiv.1706.03762.
39. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020, https://doi.org/10.48550/arXiv.2010.11929.
40. Hendrycks D, Gimpel K. Gaussian error linear units (gelus). arXiv preprint arXiv:160608415. 2016, https://doi.org/10.48550/arXiv.1606.08415.
41. Xu J, Ding X, Gong Y, Wu N, Yan H. Rotor imbalance detection and quantification in wind turbines via vibration analysis. Wind Engineering. 2022;46(1):3-11, https://doi.org/10.1177/0309524X21999841.
42. Tropp JA, Gilbert AC. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. IEEE Transactions on Information Theory. 2007;53(12):4655-66, https://doi.org/10.1109/TIT.2007.909108.
43. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences. 2009;2(1):183-202, https://doi.org/10.1137/080716542.
44. Wang Z, Zhang M, Chen H, Li J, Li G, Zhao J, et al. A generalized fault diagnosis framework for rotating machinery based on phase entropy. Reliability Engineering & System Safety. 2025;256:110745, https://doi.org/10.1016/j.ress.2024.110745.
45. Su H, Xiang L, Hu A, Xu Y, Yang X. A novel method based on meta-learning for bearing fault diagnosis with small sample learning under

different working conditions. Mechanical Systems and Signal Processing. 2022;169, https://doi.org/10.1016/j.ymssp.2021.108765.

46. Liu J, Zhang C, Jiang X. Imbalanced fault diagnosis of rolling bearing using improved MsR-GAN and feature enhancement-driven CapsNet. Mechanical Systems and Signal Processing. 2022;168, https://doi.org/10.1016/j.ymssp.2021.108664.