Mechanical Systems and Signal Processing 140 (2020) 106683

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

A deep learning method for bearing fault diagnosis based on Cyclic Spectral Coherence and Convolutional Neural Networks

Zhuyun Chen [a], Alexandre Mauricio [b,c], Weihua Li [a,*], Konstantinos Gryllias [b,c]

[a] School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China
[b] Department of Mechanical Engineering, KU Leuven, Celestijnenlaan 300, Leuven 3001, Belgium
[c] Dynamics of Mechanical and Mechatronic Systems, Flanders Make, Belgium

ARTICLE INFO

Article history:
Received 31 May 2019
Received in revised form 31 October 2019
Accepted 26 January 2020

Keywords:
Fault diagnosis
Cyclic Spectral Coherence
Convolutional Neural Networks
Rolling element bearings
Group normalization

ABSTRACT

Accurate fault diagnosis is critical to ensure the safe and reliable operation of rotating machinery. Data-driven fault diagnosis techniques based on Deep Learning (DL) have recently gained increasing attention due to theirs powerful feature learning capacity. However, one of the critical challenges lies in how to embed domain diagnosis knowledge into DL to obtain suitable features that correlate well with the health conditions and to generate better predictors. In this paper, a novel DL-based fault diagnosis method, based on 2D map representations of Cyclic Spectral Coherence (CSCoh) and Convolutional Neural Networks (CNN), is proposed to improve the recognition performance of rolling element bearing faults. Firstly, the 2D CSCoh maps of vibration signals are estimated by cyclic spectral analysis to provide bearing discriminative patterns for specific type of faults. The motivation for using CSCoh-based preprocessing scheme is that the valuable health condition information can be revealed by exploiting the second-order cyclostationary behavior of bearing vibration signals. Thus, the difficulty of feature learning in deep diagnosis model is reduced by leveraging domain-related diagnosis knowledge. Secondly, a CNN model is constructed to learn high-level feature representations and conduct fault classification. More specifically, Group Normalization (GN) is employed in CNN to normalize the feature maps of network, which can reduce the internal covariant shift induced by data distribution discrepancy. The proposed method is tested and evaluated on two experimental datasets, including data category imbalances and data collected under different operating conditions. Experimental results demonstrate that the proposed method can achieve high diagnosis accuracy under different datasets and present better generalization ability, compared to state-of-the-art fault diagnosis techniques.

1. Introduction

Condition monitoring and fault diagnosis are essential for normal and safe operation of rotating machinery. Rolling element bearings, being one of the key components of industrial machines, often suffer from structure damages due to long-time running and harsh working conditions. A sudden failure usually leads to equipment breakdown, to economic loss or

* Corresponding author.
E-mail address: whlee@scut.edu.cn (W. Li).

https://doi.org/10.1016/j.ymssp.2020.106683
0888-3270/

even to human casualties [1,2]. Therefore, it is necessary to develop fault diagnosis techniques to monitor the health condition, guarantee the reliability of machines and support the decision-making tasks.

Data-driven based intelligent fault diagnosis methods usually involve two steps, namely feature extraction realized by signal processing techniques and fault recognition implemented by classification algorithms. Feature extraction is important to reduce the input dimension and the random noise by extracting meaningful health indicators, based on spectral analysis and time-frequency analysis [3,4]. Traditional classification techniques such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and *k*-NN are widely adopted due to the fast implementation and good classification performance [5–9]. However, being limited to their shallow architectures, they have difficulty in learning effectively discriminative features from raw high-dimensional inputs. The diagnosis performance relies strongly on the diagnosis expertise and the diagnostic feature quality.

In recent years, DL-based intelligent fault diagnosis approaches have received increasing attention. DL models refer to stacking multiple non-linear processing layers to construct hierarchical architectures, making it easy to learn representations from raw or preprocessed mechanical data such as time series [10–12], spectra [13,14], and time frequency maps [15–17]. Various deep network architectures and variants, such as Auto Encoders (AE) [18,19], Deep Belief Networks (DBN) [20,21], Long Short-Term Memory (LSTM) networks [22,23], and Convolutional Neural Networks (CNN) [24–26], have been adopted for solving different diagnosis tasks.

Among the DL algorithms, CNNs have been widely used and achieved state-of-the-art diagnosis performance for their strong local feature extraction capability and flexible architectures. Guo [27] converted raw data into input matrices and then designed a CNN hierarchical framework with adaptive learning rate to recognize bearing fault patterns and fault sizes. Liu [28] proposed a Dislocated Time Series CNN (DTS-CNN) for fault diagnosis of an electric machine, where a dislocate layer was added to bridge the relationship between signals with different intervals in periodic mechanical signals. Islam [29] utilized the wavelet packet transform to preprocess raw acoustic emission signals and designed an adaptive CNN for multi-fault classification of bearings. Ince [30] developed a 1D CNN to conduct end-to-end motor fault diagnosis from raw signal input. Jia [31] proposed a normalized CNN for improving the bearing diagnosis performance under imbalanced data by embedding normalized layers and weighted Softmax loss. Additionally, a Neuron Activation Maximization (NAM) algorithm is further adopted to explore the properties of the learned filters. Moreover, batch normalization with good domain adaptation ability has been introduced into CNNs to process the raw vibration signal inputs, and further improve the bearing diagnosis performance under variable operating conditions [32,33].

Although different feature extraction and classification methods based on CNNs have been successfully used for different fault diagnosis issues, there are still a number of open questions and the following situations should be considered.

1) The development of faults in rolling element bearings lead to the generation of repeated impacts due to the passing of rolling elements over the defect. The collected time domain signals usually cannot directly reveal the intrinsic characteristics of bearing health conditions. Valuable information, i.e. periodicities and correlations related to the failure modes, are easily masked by strong background noise and other components. Many existing feature extraction methods do not leverage well the diagnosis knowledge to obtain enhanced feature representations, limiting the improvement of the final diagnosis performance.
2) A number of already proposed diagnosis models yield satisfactory performance based on the basic assumption that the training and the testing samples follow the same data distribution, which does not take the domain adaptation ability into consideration. The methods may suffer a significant loss in performance when applied on a new diagnosis task due to the data distribution discrepancy induced by imbalanced samples [31] or different operating conditions [32], even if the new task is similar to the original one.

To overcome the first weakness, in the preprocessing step, by considering the bearing health conditions and the physical characteristics of the defects, the cyclic spectral analysis is proposed to obtain 2D CSCoh images/maps. CSCoh constructs a frequency-frequency domain map to exhibit the amplitude levels of the modulation frequency of excitations by exploiting the second-order cyclostationary behavior. Thus, the hidden periodic behaviors of each fault type in the vibration signals are revealed with a unique fault pattern, which is regarded as enhanced feature representations for the input of diagnosis networks. In addition, to conquer the second shortcoming and to achieve good domain adaptation for the diagnosis model, CNN is developed to learn high-level feature representations and conduct fault classification. More specifically, a Group Normalization (GN) technique is embedded into the CNN to normalize the feature maps of network and to adaptively learn distribution for each group data. This modification can reduce the internal covariant shift induced by data distribution discrepancy so as to improve the network generalization performance.

The main contributions of this paper can be summarized as follows:

(1) A novel CSCoh-CNN intelligent diagnosis method is proposed for vibration-based fault diagnosis of motor bearings. The proposed method not only leverages the excellent feature extraction ability of CSCoh, but also exploits the outstanding classification performance of CNN.

(2) In the proposed method, the CSCoh is introduced to preprocess the vibration signals, which provides superior discriminative feature representations of bearing health conditions. In additional, GN is further embedded into the CNN to reduce the internal covariant shift induced by data distribution discrepancy, which is able to obtain better generalization performance.
(3) In real industrial cases, the vibration characteristics of bearings are easily affected by various factors such as high background noise and variable operating conditions, which is a big challenge for accurately and reliable fault diagnosis. The proposed CSCoh-CNN provides a potential solution to handle this issue.

The rest of the paper is organized as follows: In Section 2, a brief theory introduction to CSCoh and CNN is provided. The proposed rolling element bearing fault diagnosis framework is presented analytically in Section 3. In Section 4, the details of the experimental datasets are described, and a comprehensive evaluation and method comparisons are carried out. Finally, some conclusions are extracted in Section 5.

2. Cyclic Spectral Coherence and Convolutional Neural Networks

2.1. Cyclic Spectral Coherence (CSCoh)

Cyclostationary signals present some statistical properties that vary cyclically with time. Opposed to stationary signals, cyclostationary signals, though not periodic, are generated by periodic mechanism and contain extra information, which is carried by hidden periodicities. Such signals can be characterized by the second-order of cyclostationarity [34,35]. In rotating machines, the bearing defects usually generate modulated signals by the characteristic frequencies of the bearings. Such signals are usually second-order periodicity processes and are able to be separated from other interfering signals to detect and identify their hidden periodic behavior [36–38].

For a cyclostationary signal $x(t)$, the second-order moment of cyclostationarity can be defined as an instantaneous AutoCorrelation Function (ACF) with a cyclic period T, which is described as:

$$R_{xx}(t,\tau) = R_{xx}(t+T,\tau) = E\{x(t+\tau/2)x(t-\tau/2)^*\} \tag{1}$$

where the star (*) denotes complex conjugation, τ is the time-lag, and E is the statistic mean. The Fourier coefficients of the ACF correspond to the cyclic ACF and are obtained by:

$$R_{xx}(\tau,\alpha) = \int R(t,\tau)e^{-j2\pi\alpha t}dt \tag{2}$$

where α is defined as the cyclic frequency. From the Eq. (2), it can be observed that the cyclic ACF represents the Fourier coefficients of α with respect to a time-lag signal $R(t,\tau)$.

The second-order statistical descriptor of cyclostationarity, called Cyclic Spectral Correlation (CSC), can be estimated by implementing the Fourier transform on the Cyclic ACF, which is given by:

$$\mathrm{CSC}(\alpha,f) = \int R_{xx}(\tau,\alpha)e^{-j2\pi f\tau}d\tau = \int\int R(t,\tau)e^{-j2\pi(\alpha t+f\tau)}dtd\tau \tag{3}$$

The CSC can also be defined as the double Fourier transform of the signal, being a function of the spectral frequency f and the cyclic frequency α. Contrary to the classic spectral analysis, it provides an additional frequency dimension, revealing both the carriers and their modulations. Spectral frequency f is linked to the carrier component, and the cyclic frequency α is linked to its modulation. When α is equal to zero, it corresponds to the classical power spectrum. On the other hand, when α is not equal to zero, it indicates the power spectrum for that specific cyclic component.

Then the CSCoh can be used to measure the degree of correlation between two spectral components estimated by:

$$\mathrm{CSCoh}(\alpha,f) = \frac{\mathrm{CSC}(\alpha,f)}{\mathrm{CSC}(0,f)\mathrm{CSC}(0,f-\alpha)} \tag{4}$$

The CSCoh can be interpreted as the CSC of a whitened signal, which tends to equalize regions with very different energy levels, magnifying weak cyclostationary signals [37].

2.2. Convolutional Neural network (CNN)

CNN is a variant of multi-layer neural network and usually consists of alternating convolutional and pooling layers [39]. In contrast to fully-connected network, CNN presents three basic characteristics, i.e. local receptive fields, shared weights and pooling, reducing the number of trainable parameters without the loss of the expressive power. This also makes CNN easy to train with a backpropagation algorithm. CNN is a hierarchical structure, stacking multiple layers. The layer types considered in this paper are introduced in the next subsections.

2.2.1. Convolutional layer

In the convolutional layer, each neuron is connected to a small region of the input neurons. The region in the input is called local receptive field. Then the convolution operation is conducted on the local receptive field to extract local features with a learned convolution kernel/weight. For each input $\mathbf{x}_i$ and convolution kernel $\mathbf{k}_j$, the output feature map is calculated as follows:

$$\mathbf{y}_{i,j} = f(\mathbf{b}_j + \sum_i \mathbf{k}_j * \mathbf{x}_i) \tag{5}$$

$$f(\mathbf{x}) = \max(0,\mathbf{x}),\ \mathbf{x} > 0 \tag{6}$$

where, $*$ denotes the convolution operation, $\mathbf{k}$ and $\mathbf{b}$ are the shared value of the kernel and the bias, which means that all the neurons in this layer detect the same feature, just at different locations in the input. $f(\cdot)$ is the neural activation function, which is usually selected as the Rectified Linear Unit (ReLU) to accelerate the convergence of CNN.

2.2.2. Pooling layer

The pooling layer is used to obtain a representation that is invariant against small translations and distortions. This is achieved by summarizing the feature responses in a region of neurons in the previous layer. For an input feature map $\mathbf{x}_i$, the output feature map $\mathbf{y}_i$ is obtained:

$$\mathbf{y}_i = \max_{r\times r}(\mathbf{x}_i) \tag{11}$$

where r is the pooling size and the common pooling operation adopted is known as max-pooling, which simply outputs the maximum activation in the input region.

3. The proposed CSCoh-CNN model

The appearance of a bearing failure results in the generation of a sequence of repetitive shocks, which in turn may excite one or more natural frequencies of the structure. Moreover, in industrial machinery, due to speed fluctuations, random slip of the rolling elements and load variations, the impacts are not exactly periodic and their amplitude present often a random fluctuation as well [40,41]. In this case, classic spectral analysis method generally fails to reveal the failure behavior in the presence of masking noise, especially in the case of the incipient faults [41]. Therefore, the CSCoh as a data pre-processing step is firstly introduced to deal with the bearing cyclostationary signals to obtain good feature representations. The vibration signals from different fault types and operating conditions are segmented. Then, the CSCoh are calculated to analyze the vibration signals and extract the unique fault patterns. After the feature learning by the CSCoh, the CNN is designed and adopted for fault classification and diagnosis, as explained in details in the next subsections.

3.1. The architecture of the proposed CNN

In this section, a novel CNN based on LeNet-5 is constructed to automatically learn features from the input. Inspired by the typical LeNet-5, the proposed CNN architecture is designed, based on the same strategy by stacking two convolutional, pooling, fully-connected layers and one Softmax layer. The dropout technique is adopted in each fully-connected layer to reduce the overfitting during network training. More specifically, the GN technique is embedded in each convolutional and fully-connected layer.

In the convolutional layer configurations, following the same principle used in [24,29], a small receptive field (3×3) is applied in each convolution layer to capture the detail information and reduce the number of parameters. The convolution stride is fixed to 1 pixel. The number of filters is set to 16 in the first convolutional layer and the second one is doubled (32) to increase the feature learning capability. The zero padding pads the input volume with zeros around the border. It is adopted to keep the same spatial dimensions between input and output volume, which helps to preserve as much information as possible about the previous input. In the pooling layers, the spatial pooling is carried on the input over a (2×2) pixel window with stride 2. Therefore, the feature map size is halved to reduce the time complexity. Then the preceding layer is followed by two fully-connected architectures. The first one has 256 neurons and the second one contains 126 neurons, respectively. The finally layer is the Softmax layer.

In addition, GN is employed in the network to normalize activations by computing the mean and the standard deviation over each group. The dropout layer, as a simple and effective regularization technique, is adopted to reduce the network overfitting. It is only embedded into the two fully-connected layers, where large parameters will increase the overfitting risk. As GN itself has some regularization effect and the network size is relative small, the parameter of dropout rate is determined by a grid search through the range of 0 to 0.5. The optimal value is set to 0.2. The details of the layer type and the parameter used are given in Table 1.

Table 1
Parameters of the proposed CNN architecture.

Layer	Layer type	Kernel	Number of filters	Filter size	Padding	Output Size
1	Input	/	/	/	/	(112, 112)
2	Conv	Kernels	16	3 × 3	Yes	(112, 112, 16)
3	Group Norm	/	/	/	/	(112, 112, 16)
4	MaxPool	Pooling size	/	2 × 2	/	(56, 56, 16)
5	Conv	Kernels	32	3 × 3	Yes	(56, 56, 32)
6	Group Norm	/	/	/	/	(56, 56, 32)
7	MaxPool	Pooling size	32	2 × 2	/	(28, 28, 32)
8	FC	/	256	/	/	(256, 1)
9	Group Norm	/	/	/	/	(256, 1)
10	Dropout	Dropout rate		0.2	/	(256, 1)
11	FC	/	126	/	/	(126, 1)
13	Group Norm	/	/	/	/	(126, 1)
14	Dropout	Dropout rate	/	0.2	/	(126, 1)
17	Softmax	/	/	/	/	(10, 1)

3.2. Group normalization (GN)

Normalizing the input data on neural networks has been widely adopted to adjust the activations of neural networks, which play a crucial role in getting the model trained effectively. Previously, the Local Response Normalization (LRN) was used to obtain local normalized features in a small neighborhood for each pixel. Recently, Batch Normalization (BN) has been proposed to perform a global normalization along the channel dimension [42]. It has become a cornerstone of DL and has been employed in recent DL fault diagnosis framework to improve the domain adaptation ability of the network [32,33].

Despite its simplicity, for CNN with fixed depth, BN requires to store the batch statistics separately for each hidden layer. Furthermore, BN cannot be applied to online learning tasks where the batch sizes have to be small. However, small batch sizes can lead to an inaccurate batch statistics estimation inciting data distribution discrepancy. This inconsistent estimation can negatively impact training and finally may lead to poor diagnosis performance. Thus push us to explore another normalization technique named Group Normalization (GN) [43] to remove the effect of BN on the statistics on mini-batch examples and to enhance the domain adaptation capability. Unlike BN, GN normalizes the features per sample within each of G groups, along the channel axis. GN can be used with small batch sizes by following a simple idea, which divides the channels into multiple groups and computes the mean and the variance of each group for normalization, thus it is independent of the batch sizes.

In the GN, the input of a mini-batch is $\mathbf{X} = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N\}$, $\mathbf{X} \in R^{N\times H\times W\times C}$, where N denotes the batch size, H and W are the spatial height and width, respectively, while C is the channel number. The parameter M, equaling to C/G, is defined as the number of channels of each group. G is the number of groups, which should not be larger than the number of the channels. For each group, GN calculates the mean $\mathbf{u}_G$ and the variance $\boldsymbol{\sigma}_G^2$ along the (H, W, M) axes, given by:

$$\mathbf{u}_G = \frac{1}{N\times H\times W} = \sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{k=(G-1)*M+1}^{G\cdot M}\mathbf{X}_{h,w,k} \tag{12}$$

$$\boldsymbol{\sigma}_G^2 = \frac{1}{N\times H\times W}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{k=(G-1)\cdot M+1}^{G*M}\left(\mathbf{X}_{h,w,k}-\mathbf{u}_G\right)^2 \tag{13}$$

Then, the output of the previous layer is normalized by subtracting the mean and dividing the standard deviation of the previous statistics as follows:

$$\widehat{\boldsymbol{X}}_{h,w,k} = \frac{\boldsymbol{X}_{h,w,k}-\boldsymbol{u}_G}{\sqrt{\boldsymbol{\sigma}_G^2+\varepsilon}} \tag{14}$$

$$\boldsymbol{Y}_{h,w,k} = \gamma\widehat{\boldsymbol{X}}_{h,w,k}+\beta \tag{15}$$

where γ and β are, respectively, the trainable scale and shift parameter, introduced to preserve the expressiveness of the network. During the testing stage, BN normalizes the input with the mean and the variance, inherited from the training stage. A divergence of the distribution may occur, if the training and testing distributions are different [43]. GN boosts the flexibility of learning a different distribution for each group data and it normalizes the group data of each layer using the statistics of its own. Thus, the normalization of each layer is able to receive data from similar distributions adaptively, without exception if it comes from the training or the testing set, ensuring good classification and generalization capability.

Due to the advantages of GN, it is nature to embed the GN into the proposed CNN to obtain better diagnosis accuracy since the classification tasks are more related to the divergence of data distributions.

3.3. Fault diagnosis framework based on CSCoh and CNN

In this section, the proposed CNN combined with the CSCoh is applied for the fault diagnosis of rolling element bearings. Firstly, the raw vibration signals are preprocessed based on cyclic spectral analysis to obtain the CSCoh maps. Thus the fault patterns of each category are enhanced before the network input, which is able to reduce the learning difficulties of the CNN. Then, the CNN is trained based on the CSCoh input to learn from low-level to high-level features layer by layer. In addition, the network introduces the GN technique to normalize the features, which can further improve the network learning capability. The overall diagnosis procedure is presented in Fig. 1 and can be summarized as follows, (1) Firstly, the vibration signals of bearings, operating under different conditions, are collected and divided into many segments to obtain training and testing data. Then, the CSCoh maps are estimated and are downsized to 112 × 112 size. Those maps are regarded as the CNN input. (2) A CNN is constructed by alternatively stacking multiple convolutional and pooling layers. More specifically, GN is embedded into the CNN to normalize the features to further reduce the internal covariant shift induced by the data distribution discrepancy. Additionally, the network weights in each layer are randomly initialized, being prepared for the update in the training stage. (3) Training data are used to train the constructed CNN model. Back propagation is utilized to optimize the network parameter of each layer by minimizing the cross entropy loss error between the output and the true label. Finally, the bearing fault diagnosis framework based on the CNN and the CSCoh is obtained. (4) Testing data are then fed into the trained CNN model to automatically extract features. The classification and the visualization results are presented to provide comprehensive diagnosis analysis.

4. Experimental validation

4.1. Experiment setup and data description

The experimental data of rolling element bearings have been acquired from the public bearing data center provided by Case Western Reserve University (CWRU) [44], which is often used as a benchmark dataset. As shown in Fig. 2, the bearing test rig is mainly composed of an induction motor, a transducer and a dynamometer. The vibration data at the driven end of the motor have been obtained by an accelerometer mounted on the housing of the induction motor with a sampling frequency of 48 *k*Hz. In addition to the Normal Condition (NC), three types of single point faults including Inner-race Faults (IFs), Outer-race Faults (OFs) and Ball Faults (BFs) have been introduced into SKF deep-groove ball bearings using an Electro-Discharge Machining (EDM). For the fault bearings, each fault type has three levels of severity with fault diameters of 7 mils, 14 mils and 21 mils (1 mil = 0.001 inches), respectively, and the fault depths are 11 miles. Thus, there are in total

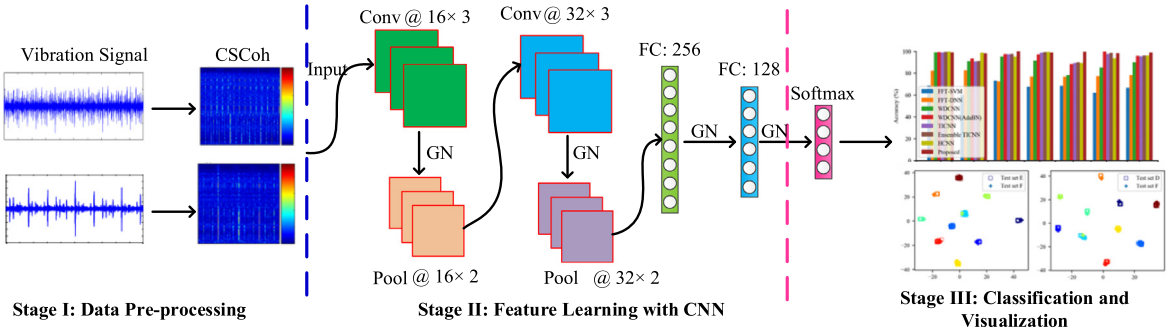


Fig. 1. The proposed CSCoh-CNN fault diagnosis framework.

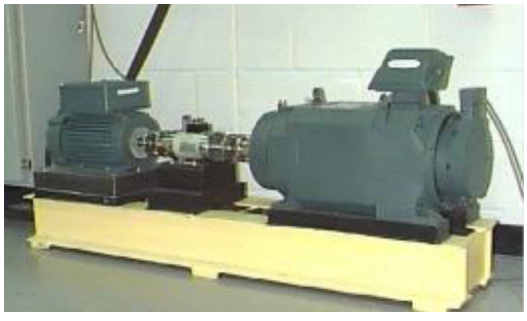


Fig. 2. The CWRU bearing fault test rig.

ten types of bearing health conditions (NC, BF7, BF14, BF21, IF7, IF14, IF21, OF7, OF14 and OF21) under three different shaft speeds of 1772 rpm, 1750 rpm and 1730 rpm, corresponding to the loads of 1, 2, and 3 hp, respectively.

In the pre-processing stage of CSCoh, usually, a large sampling point is helpful to obtain a high frequency resolution, which may bring higher diagnosis accuracy. However, it may also lead to increase of the computational cost. A small sampling point will bring less computational cost, but it will lead to large information loss, which may deteriorate the classification performance. In order to obtain the 2D CSCoh maps from the vibration measurements, 24,000 data points (time duration of 0.5 s) are considered from the raw time-series signals to form one sample by balancing the tradeoff between computational cost and testing accuracy. Therefore, 20 samples are obtained from each health condition under each operating load. The health conditions from the different operating loads are regarded as the same category. In addition, it should be noted that, due to the limited sampling time, there are only 14 samples obtained for the IF14 under the load 1. This can be regarded as an extra data imbalance, which is present in the dataset. The details of the data description are listed in Table 2.

4.2. Analysis of Cyclic Spectral Coherence (CSCoh)

When a defect occurs in a rolling element bearing, repeat impacts are generated due to the passing of rolling elements over the defect. An effective preprocessing technique is able to reveal the fault nature for different fault types, i.e. the Fundamental Train Frequency (FTF), the Ball Spin Frequency (BSF), the Ball Pass Frequency of Inner-race (BPFI) and the Ball Pass Frequency of Outer-race (BPFO). To investigate the effectiveness of CSCoh in revealing the fault patterns of rolling element bearings, the 2D CSCoh maps of different health conditions are presented in Fig. 3.

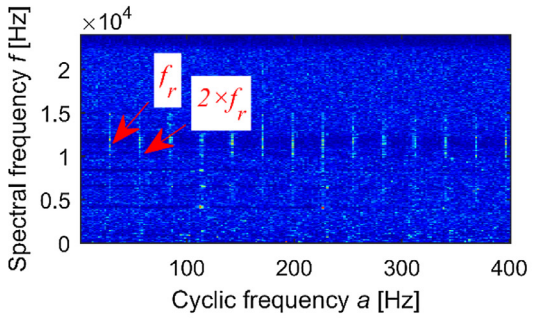
It can be observed that the CSCoh maps provide a unique feature for given fault types. In Fig. 3 (a), there are only a clear shaft frequency (f_r) and its harmonic present, corresponding to the normal condition. In Fig. 3 (b) and Fig. 3 (c), the inner-race defect frequency (BPFI) and the outer race defect frequency (BPFO) and its harmonic can be clearly captured, demonstrating the occurrence of the specific faults. It should be noted that in the case of BFs, the (weak) amplitude of the FTF and the BSF, which reveal the existence of the ball fault, can be detected only in a few of samples of BF14, such as the one presented in Fig. 3 (d). In other cases, such as the BF7 shown in Fig. 3 (e), only the shaft frequency and its harmonics are present, while the fault characteristic frequency and its harmonic are indistinguishable. Similarly, in OF14 shown in Fig. 3 (f), no obvious bearing fault frequency can be identified.

The failure of the method in diagnosing the abovementioned faults is not really a weakness of the method, as it can be logically explained. Firstly, the limited sampling time of 24,000 data points (0.5 s) may lead to low resolution, which may not be enough to capture the cage frequency FTF. Secondly, the data acquisition may have not been correctly realized at these particular cases, as the test rig assembly seems to affect the diagnosis results more than the fault itself. Smith and Randall [44] provided a benchmark study for the CWRU bearing data with three state-of-the-art diagnostic techniques, analyzing and categorizing the diagnostic performance for each signal independently. In their work, all three diagnostic methods failed to detect the BSF in a number of BF cases. More specifically, they are classified as "Data potentially diagnosable" or "Data not diagnosable for the specified bearing fault". In the OF14 case, only few data samples are clearly diagnosable, showing in parallel non-stationary characteristics in the time and/or the frequency domain, while the rest samples are categorized as "Data not diagnosable".

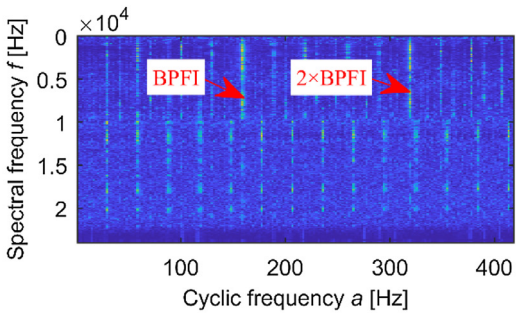
From the analysis above, it can be concluded that the CSCoh maps provide nearly consistent diagnosis results with [44], showing its feasibility in extracting discriminative features of bearings. Those features can be regarded as good input representations of CNN. In addition, for the "not diagnosable cases", the fault characteristics are not presented directly, which may increase the learning difficulty of CNN. However, from the prospect of pattern recognition, the CNN attempts to learn hierarchical features to distinguish different classes ignoring the specific physical meaning. For example, although the samples of BF7 and OF14 are regarded as "not diagnosable cases", their CSCoh maps present different fault properties, which can be possibly learned by the CNN to be recognized as two different categories. On the other hand, CNN does not just exploit the characteristic information like humans do, but additionally exploits the whole CSCoh maps, where abundant diagnosis information such as carrier components and modulated components are contained. This information is helpful for CNN to

Table 2
Description of the rolling element bearing dataset.

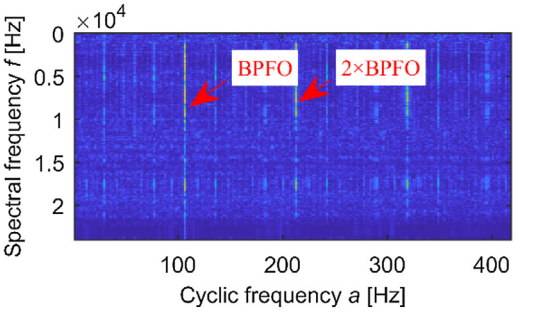
Load (hp)	Speeds (rpm)	Fault Types	Fault Diameters (mil)	Number of Samples	Class Label
1 & 2 & 3	1772 & 1750 & 1730	NC	0	20 & 20 & 20	1
1 & 2 & 3	1772 & 1750 & 1730	BF	7	20 & 20 & 20	2
1 & 2 & 3	1772 & 1750 & 1730	BF	14	20 & 20 & 20	3
1 & 2 & 3	1772 & 1750 & 1730	BF	21	20 & 20 & 20	4
1 & 2 & 3	1772 & 1750 & 1730	IF	7	20 & 20 & 20	5
1 & 2 & 3	1772 & 1750 & 1730	IF	14	14 & 20 & 20	6
1 & 2 & 3	1772 & 1750 & 1730	IF	21	20 & 20 & 20	7
1 & 2 & 3	1772 & 1750 & 1730	OF	7	20 & 20 & 20	8
1 & 2 & 3	1772 & 1750 & 1730	OF	14	20 & 20 & 20	9
1 & 2 & 3	1772 & 1750 & 1730	OF	21	20 & 20 & 20	10



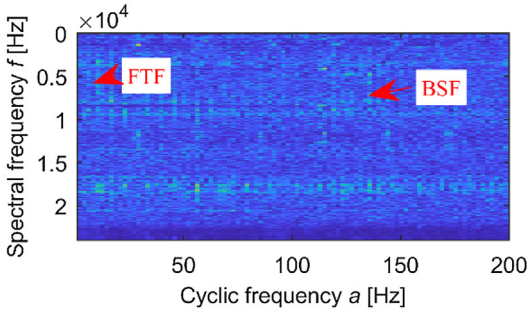
(a) Normal Condition (NC)



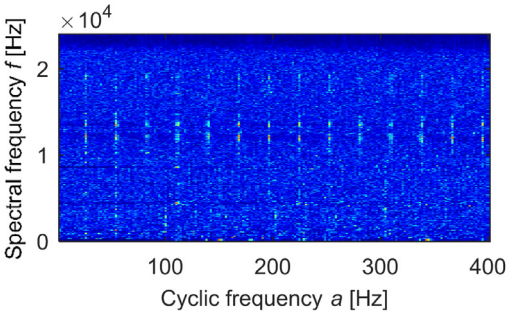
(b) Inner-race Fault with diameter of 7 mil (IF 7)



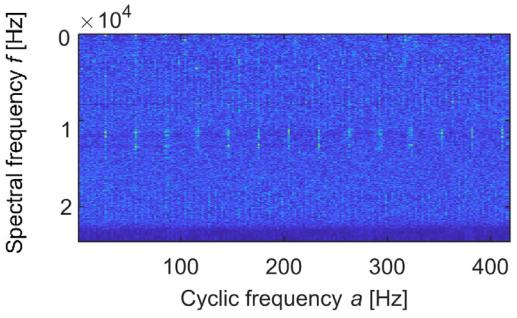
(c) Outer-race Fault with diameter of 7 mil (OF7)



(d) Ball Fault with diameter of 7 mil (BF7)



(e) Ball Fault with diameter of 7 mil (BF7)



(f) Outer-race Fault with diameter of 14 mil (OF14)

Fig. 3. The 2D CSCoh maps of the different bearing health conditions.

provide comprehensive evaluation and discrimination. Thus, it is expected to obtain more automatic and intelligent diagnosis and fault classification of rolling element bearings by combining CSCoh maps with CNN.

4.3. Fault diagnosis under imbalanced dataset

4.3.1. Imbalanced dataset

In real industrial environments, the normal condition is the major pattern, and therefore, the samples of normal condition are usually more than those of the fault types. This fact appears as a data imbalanced problem, where the correct diagnosis and fault classification of a dataset is relatively more difficult due to the inconsistent data distribution of each health condition. In this section, in order to better evaluate the performance of the proposed method, referred to [31], four different datasets are constructed to simulate various imbalanced data distributions. In dataset A, all the samples from different health conditions are split into equal pieces. 50% of the data are used for training and the rest for testing. Thus dataset A can be considered as a balanced dataset. In dataset B, the BFs, the IFs and the OFs are randomly segmented into different partitions, in which 30%, 20% and 10% of the samples are used for training, while the testing samples are still 50% to facilitate the comparisons. Following the strategy of dataset B, Dataset C and D are then prepared by changing the rate order of BFs, IFs and OFs. The details of each dataset are listed in Table 3.

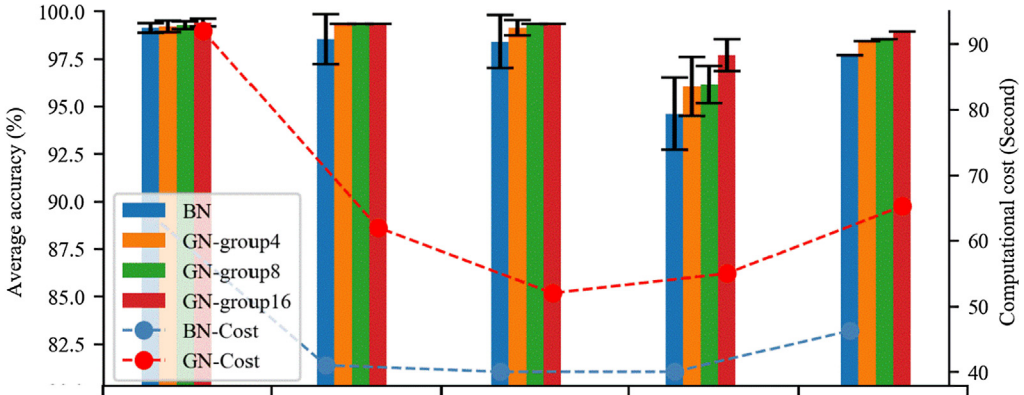
Table 3
Description of datasets with different sample imbalance ratios.

Health condition	The percent of training samples				The percent of testing samples
	Dataset A	Dataset B	Dataset C	Dataset D	Dataset A/B/C/D
NC	50%	50%	50%	50%	50%/50%/50%/50%
BF7	50%	30%	20%	10%	50%/50%/50%/50%
BF14	50%	30%	20%	10%	50%/50%/50%/50%
BF21	50%	30%	20%	10%	50%/50%/50%/50%
IF7	50%	20%	30%	20%	50%/50%/50%/50%
IF14	50%	20%	30%	20%	50%/50%/50%/50%
IF21	50%	20%	30%	20%	50%/50%/50%/50%
OF7	50%	10%	10%	30%	50%/50%/50%/50%
OF14	50%	10%	10%	30%	50%/50%/50%/50%
OF21	50%	10%	10%	30%	50%/50%/50%/50%

4.3.2. Comparison between BN and GN with different group sizes

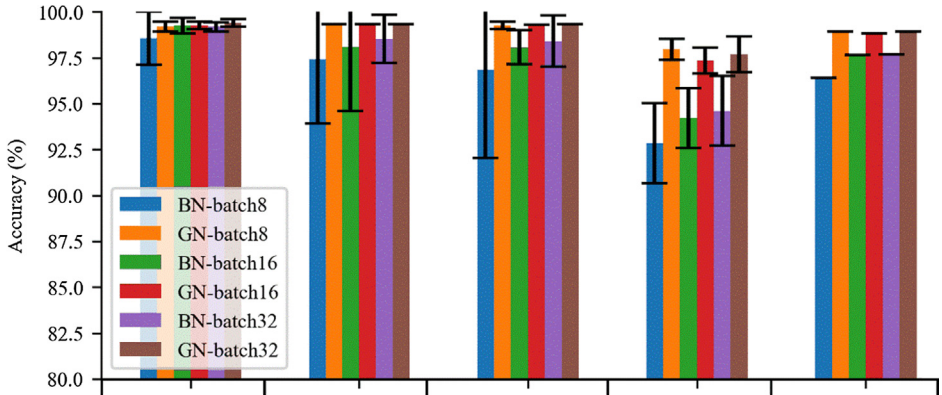
In this section, we firstly evaluate the performance of the proposed method on the four datasets. Two different normalization techniques including BN and GN are adopted in the network to make a comparison. For GN, there is one parameter for the group number, which should be taken into consideration. To investigate the effect on the classification performance, GN with three different group numbers (group = 4, 8 and 16, respectively) are considered. Adam optimization algorithm is employed to update the parameters due to its computational effectiveness. The max training epoch is set to 200 with a batch size of 32. All computations throughout the whole experiments have been performed on a computer with Intel Xeon E5-262v3 (2.4 GHz), 64 GB RAM, 4 TITAN X graphics cards and Google TensorFlow framework. The classification process is repeated 10 times to reduce the influence of randomness introduced by network initialization. The identification accuracy, which is defined as the number of correctly classified health conditions of bearings to the total number of health conditions is introduced to assess the prediction performance of the proposed method. The average results including the accuracy and the standard deviation are shown in Fig. 4.

It can be observed that GN with different group numbers works well in all datasets and outperforms BN. The average accuracy of BN is 97.68%, while that of GN with a group number of 4 and 8 is 98.42% and 98.51%, respectively. The high 98.93% accuracy of GN is achieved with a group number of 16, indicating that the detailed subdivision of groups leads to the best performance. In addition, the standard deviation of GN with a small group number of 4 is larger than what GN obtained with larger group numbers, but is still lower than that of BN. It is demonstrated that GN is more robust and toler-



	Dataset A	Dataset B	Dataset C	Dataset D	Average
BN	99.19±0.26	98.53±1.31	98.40±1.39	94.6±1.90	97.68
GN-group4	99.19±0.30	99.33±0	99.13±0.40	96.03±1.54	98.42
GN-group8	99.26± 0.20	99.33±0	99.33±0	96.13 ±0.98	98.51
GN-group16	99.39±0.22	99.33±0	99.33±0	97.68+0.83	98.93

Fig. 4. Comparison results between BN and GN with different group sizes from 4 to 16.



	Dataset A	Dataset B	Dataset C	Dataset D	Average
BN-batch8	98.55±1.44	97.43±3.51	96.83±4.80	92.85±2.19	96.42
GN-batch8	99.20±0.27	99.33±0	99.26±0.20	97.94±0.57	98.93
BN-batch16	99.25 ± 0.22	98.08±3.50	98.06±0.92	95.21 ±1.62	97.65
GN-batch16	99.26±0.20	99.33±0	99.33±0	97.34±0.71	98.82
BN-batch32	99.19±0.26	98.53±1.31	98.40±1.39	94.6±1.90	97.68
GN-batch32	99.39±0.20	99.33±0	99.33±0	97.68±0.98	98.93

Fig. 5. Comparison results between BN and GN with different batch sizes from 8 to 32.

ated. However, one of the drawbacks is that the computational cost of GN is a bit longer than that of BN, which can be alleviated with the development of advanced computational devices. It should be noted that the results in both methods are relatively poor with dataset D compared to those of the other datasets. It is possible that the feature patterns of BFs cannot be extracted in the preprocessing procedure, as analyzed in Section 4.2. Thus reducing the number of training data with a large rate leads to the loss of the more useful information, which makes it difficult to learn a good classification models.

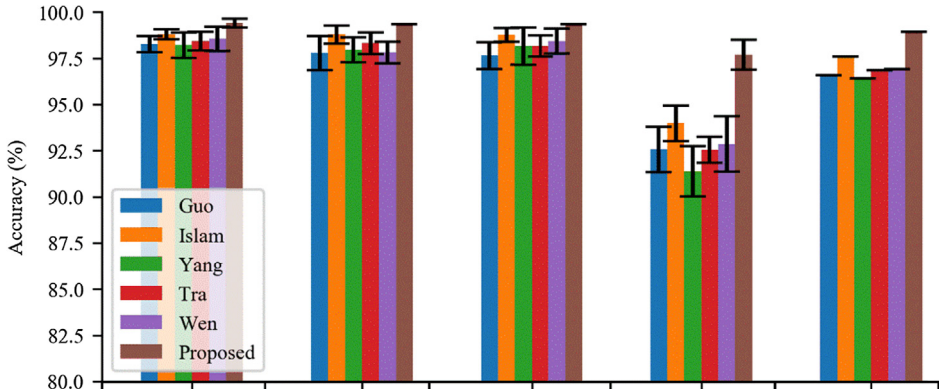
4.3.3. Comparison between BN and GN with different batch sizes

In addition, the performance of BN and GN is further evaluated using different batch sizes. Three different batch sizes of 8, 16 and 32 are set, respectively, referring to small, medium and large batch size. The evaluation results are presented in Fig. 5.

It can be first seen that the accuracy of BN with a large batch size of 32 is competitive to that of GN in dataset A, B, and C. On the other hand, there is only 94.60% accuracy in dataset D, which is worse than that obtained from GN. When the batch size becomes smaller, the classification performance of BN deteriorates. On the contrary, GN obtains stable and accurate results under different batch sizes. The average accuracy of GN with three different batch sizes is 98.93%, 98.82% and 98.93%, respectively, indicating that GN is less sensitive to the change of batch sizes. This is due to the fact that for BN, a small batch leads to inaccurate estimation of the batch statistics, which deteriorates the classification performance during the testing stage. In the imbalanced dataset B, C and D, a small batch size would further increase the uncertainty of the data distribution in each training stage, thus leading to a large standard deviation and to a decline in accuracy. While, for the use of GN, it gets rid of the constraints of batch size by dividing the channel into groups and normalizing the features within each group. Therefore, GN allows to adopt different batch sizes for different datasets in a large range without affecting its accuracy too much, which is more generally applicable in a real industrial environment.

4.3.4. Comparison with other CNN architectures

In order to further demonstrate the advantages of the proposed approach, five different CNN architectures are considered for comparison. The CNN architectures have been respectively proposed by Guo [27], Islam [29], Yang [24], Tra [25] and Wen [26], for bearing fault classification, respectively. For a fair comparison, all the five algorithms are trained with the CSCoh map input, and the network parameters, such as the batch size and the number of the epochs, are kept consistent with the proposed approach. The results, averaged by 10 trials, are shown in Fig. 6.



	Dataset A	Dataset B	Dataset C	Dataset D	Average
Guo	98.25± 0.44	97.78±0.93	97.64±0.72	92.56±1.24	96.56
Islam	98.79±0.28	98.78±0.48	98.76±0.38	93.97±0.96	97.58
Yang	98.21±0.68	97.95±0.66	98.15±1.00	91.38±1.35	96.42
Tra	98.42 ±0.50	98.31±0.59	98.15±0.58	92.53±0.71	96.85
Wen	98.55±0.66	97.81±0.58	98.43±0.67	92.86±1.51	96.91
Proposed	99.39±0.20	99.33±0	99.33±0	97.68±0.98	98.93

Fig. 6. Comparison result with five different CNN architectures.

From the results, it can be seen that all the algorithms provide relative good classification accuracy and low standard deviation in the Datasets A, B and C. This demonstrates that on one hand, the CSCoh maps provide useful discriminative feature input, contributing to a consistent good performance even under imbalanced datasets. On the other hand, the CNNs have the ability to learn effective discriminative features, keeping relative robust to the small imbalanced dataset training. In addition, in dataset D, there is less discriminative information (fault characteristics) provided from the CSCoh maps, as mentioned above. The reduced training samples of the ball faults further lead to the increase of the uncertainty of data distribution. Therefore, the health conditions are relatively more difficult to diagnosis. In this dataset, all five compared algorithms perform relative poor, while the proposed method still obtains relative high accuracy, attributing to the good performance of the GN in reducing the divergence of data distribution.

4.4. Fault diagnosis under different operating conditions

4.4.1. Dataset description

In this section, the generalization performance of the proposed method is further verified under different operating loads. The experimental scenario referred to Zhang [33] is considered, and the descriptions are given in detail in Table 4. In this experiment, the dataset from one load is used for training and that from the other two loads are used for testing. Total six group diagnosis experiments are conducted. For example D → E and D → F denote that dataset D is used for training, and dataset E and dataset F are respectively considered for testing. Each dataset contains ten fault types under one load, and the details have been presented previously in Table 1. More specifically, dataset D, E and F are linked to load 1, 2 and 3, respectively.

4.4.2. Comparison with time-frequency analysis methods

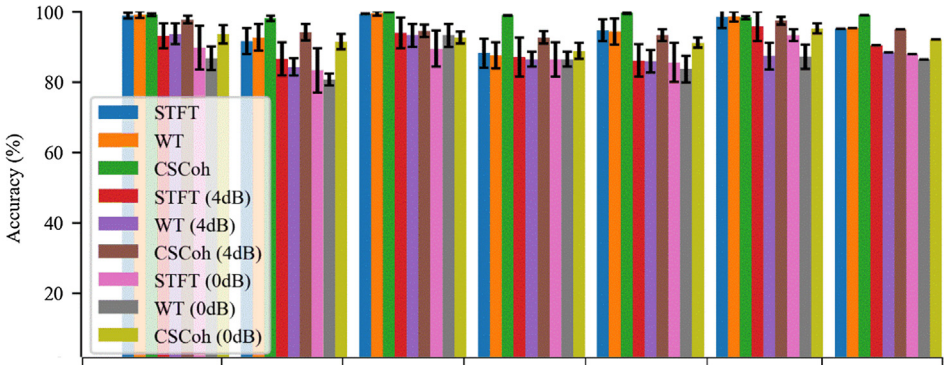
The time-frequency analysis such as the Short-Time Fourier Transform and the Wavelet Transform are usually considered as effective tools for preprocessing nonstationary and transient signals. However, the characteristic frequencies may be unobservable in the spectra when vibrations produced by other components mask the repetitive impacts in the signals. On the other hand, the main advantage of CSCoh lies in its ability to reveal hidden periodicities of second-order cyclostationarity processes, like the signals of defective bearings that are masked under stronger background noise [37].

Table 4
Description of datasets under different operating conditions.

Dataset type Description	Training dataset under one load Labeled data	Testing data under another load Unlabeled data	
1	Dataset D	Dataset E	Dataset F
2	Dataset E	Dataset D	Dataset F
3	Dataset F	Dataset D	Dataset E

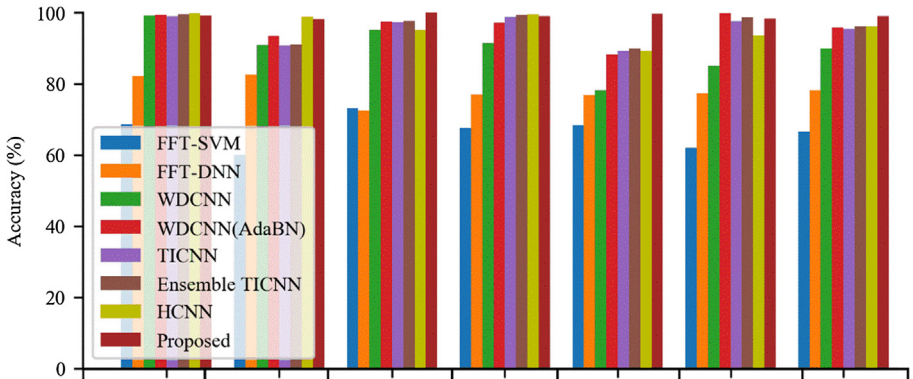
In CWRU dataset, the raw vibration data inherently includes a level of background noise. However, in the real-world industrial applications, machinery may operate on a harsh environment, the vibration signals collected from rotating machinery may contain a higher level of noise. Therefore, it is meaningful to investigate the effect of Signal-to-Noise (SNR) of signals on classification performance among different methods. In our experiments, white Gaussian noise with a SNR of 4 dB and 0 dB are added on the raw signals to simulate the different noise levels in an industrial production environment. Then the signals with additive noise are further converted into 2D representations, which are fed into the proposed CNN for training and testing. A comparison among STFT spectrograms, WT scalograms with Morlet wavelet [16] and the CSCoh maps is carried out. All images are put into constructed CNN for training and testing. The accuracies and standard deviations are shown in Fig. 7.

In regard to the three map inputs under different noise levels, it can be observed that 1) the diagnosis accuracy of all methods decrease with the increase of noise levels. This appears due to the fact that a low SNR means that signals are contaminated in a significant manner. Thus, it is difficult to extract features reflecting the health conditions of rotating machinery. 2) In the lower SNR (SNR = 0, 4), the classification accuracies of all methods are obviously lower than those with original data. This is mostly because bearing signals collected from the test rig present different fault types and especially different



	D→E	D→F	E→D	E→F	F→D	F→E	Average
STFT	98.80±0.84	91.60±3.68	99.42±0.15	88.20±4.17	94.70±3.07	98.45±3.15	95.20
WT	99.19±0.98	92.70±3.84	99.53±0.67	87.56±3.64	94.33±3.66	98.60±1.44	95.32
CSCoh	99.15±0.39	98.10±0.73	99.95±0.15	98.95±0.15	99.69±0.41	98.30±0.40	99.02
STFT (4dB)	93.15±3.52	86.55±4.71	93.97±4.43	87.05±5.53	86.13±4.67	95.80±4.16	90.44
WT (4dB)	93.70±2.93	84.30±2.38	93.29±3.27	86.45±2.07	85.95±3.17	87.35±3.80	88.51
CSCoh (4dB)	97.75±1.10	94.10±2.36	94.50±1.75	92.68±1.73	93.35±1.65	97.45±1.06	94.97
STFT (0dB)	89.75±6.2	83.35±6.32	89.48±5.10	86.40±4.88	85.57±5.59	93.35±1.65	87.98
WT (0dB)	86.70±3.38	80.75±1.66	93.29±3.27	86.45±2.07	83.66±3.84	87.20±3.44	86.34
CSCoh (0dB)	93.56±2.63	91.50±2.17	92.65±1.64	88.81±2.25	91.19±1.50	95.20±1.46	92.15

Fig. 7. Comparison result with time-frequency analysis methods under different operating conditions.



	D → E	D → F	E → D	E → F	F → D	F → E	Average
FFT-SVM	68.6	60	73.2	67.6	68.4	62.0	66.6
FFT-DNN	82.2	82.6	72.5	77.0	76.9	77.3	78.1
WDCNN	99.2	91.0	95.1	91.5	78.1	85.1	90.0
WDCNN(AdaBN)	99.4	93.4	97.5	97.2	88.3	99.9	95.9
TICNN	99.1	90.7	97.4	98.8	89.2	97.6	95.5
Ensemble TICNN	99.5	91.1	97.6	99.4	90.0	98.7	96.1
HCNN	99.93	98.79	95.15	99.45	89.33	93.69	96.1
Proposed	99.15	98.1	99.95	98.95	99.69	98.30	99.02

Fig. 8. Comparison of the proposed method with other state-of-the-art methods.

fault severities. In the small fault sizes, the characteristic frequencies are weak, which are easily submerged into noise making it hard to discriminate. 3) In addition, it can be observed that a CNN with CSCoh representations presents high classification results under different noise levels. Compared to the other two methods, the proposed method achieves better classification performance in all six diagnosis cases including the original data and the two noise levels in terms of the average accuracies, which shows a robust and superior feature representation ability.

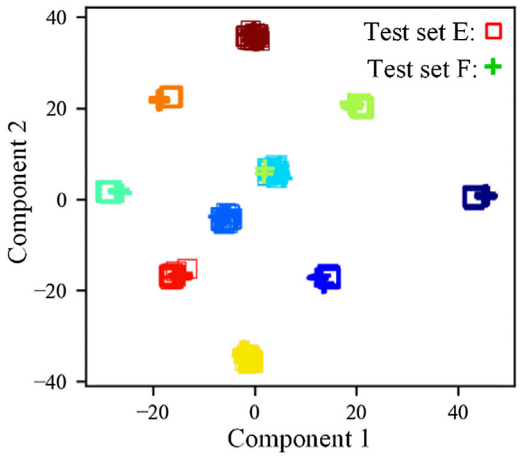
4.4.3. Comparison with other state-of-the-art methods

In this section, the results of the proposed method are further compared with published studies on the same dataset, which are shown in Fig. 8. In those methods, the WDCNN and the improved WDCNN (AdaBN) have been proposed by Zhang [32]. WDCNN adopts a wide convolutional kernel and BN for improving the classification. The domain adaptation capacity of the WDCNN (AdaBN) is extended by embedding the adaptive BN in the test stage. They obtain an average accuracy of 90.0% and 95.9%, respectively in the six group experiments. TICNN and the improved Ensemble TICNN have been proposed by Zhang [33], which adopts a small batch size and kernel dropout to boost the classification. They obtain an average accuracy of 95.5% and 96.1%, respectively. HCNN have been proposed by Wen [11], which employs a two-level hierarchical diagnosis network for fault classification. It achieves an average accuracy of 96.1%. In addition, FFT-SVM and FFT-DNN [13] have spectra as input, and are selected as the baseline for this comparison. The results are taken from [32].

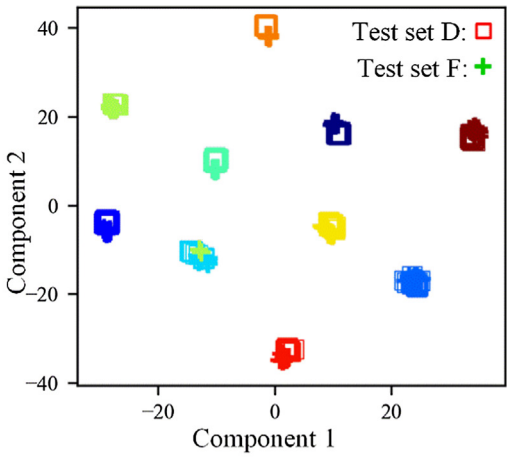
It can be seen that the proposed method achieves an equivalent or an even better diagnosis accuracy compared to the other state-of-the-art methods in the six group cases. Each diagnosis case obtains more than 98% testing accuracy, which shows better stability and generalization performance than the other methods. In addition, the average accuracy is 99.02%, which also outperforms other diagnosis techniques. The results demonstrate that the proposed method is effective in improving the diagnosis of bearing under different operating conditions.

4.4.4. Feature visualization analysis

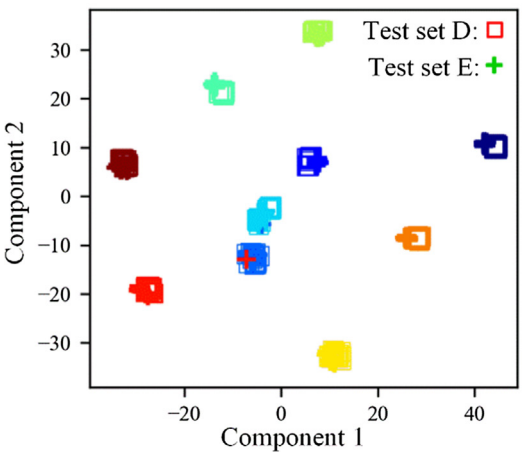
In order to further verify the feature learning effectiveness of the proposed method under different operating conditions, the t-distributed Stochastic Neighbor Embedding (t-SNE) technique [45] is considered for feature visualization by mapping



(a) Feature visualization for test sets E and F



(b) Feature visualization for test sets D and F



(c) Feature visualization for test sets D and E

Fig. 9. Feature visualization via t-SNE reduced from the learned representations using the proposed method.

the high-level output in the layer 14 from 256 into 2 dimensions. The results are shown in Fig. 9 (a)-(c). In Fig. 9 (a), the model is trained on the dataset D, and the features from the test set E and test set F are extracted for visualization. The square and the plus sign represent the testing samples from two different datasets. The ten different colors denote the ten different classes. It is expected that samples sharing the same colors even from two different datasets can be clustered together, since their cluster performance can directly reveal the domain adaptation capability under different operating conditions.

From the visual results, it is clear that the data points of different health conditions are well separated and that the same fault types of dataset E and F, sharing the same colors are clustered together, which demonstrated that the two feature representation distributions are consistent with each other. In addition, in data set F, small quantities of data points are overlapped between IF14 and BF21, denoting that the samples from the IF14 are misclassified as samples of BF21. This will lead to a low diagnosis accuracy compared to that of test set E. This observation is consistent with the results obtained in Fig. 7. In addition, Fig. 9 (b) and (c) also perform excellent, where the different categories can be separated clearly, and the same categories from different datasets are also well clustered together. Therefore, it can be concluded that the proposed method is able to learn good discriminative features and present strong domain adaptation capability.

5. Conclusion

A new DL-based fault diagnosis framework, combining CSCoh and CNN, has been presented, in order to improve the fault recognition performance of rolling element bearing. Firstly, CSCoh is considered, as a preprocessing step, to reveal the fault nature of each fault type, thus different discriminative features are preliminary obtained. Then a CNN is constructed for further feature learning and classification. In addition, GN is adopted in CNN to reduce the divergence of data distribution by

implementing normalization for each independent group. Thus, a CNN with a strong domain adaptation capability is obtained. The proposed method is applied and evaluated on the CWRU motor bearing dataset. Different cases including imbalanced data and data collected under different operating conditions are used to evaluate the effectiveness of the methodology. In addition, the effects of the group numbers and the batch sizes are also investigated. It has been demonstrated that the proposed method not only achieves high classification performance, but also presents better generalization performance compared to other state of the art fault diagnosis methods.

Fault diagnosis based DL model presents powerful capabilities in feature extraction and classification. However, its performance is affected, as normally expected, by the scale and the quality of datasets. Domain knowledge combined with signal processing techniques can be used to extract meaningful features, contributing to the success of applying DL model on machine health monitoring.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (Grant No. 51875208) and National Key Research and Development Program of China (Grant No. 2018YFB1702400).

References

[1] R. Liu, B. Yang, E. Zio, X. Chen, Artificial intelligence for fault diagnosis of rotating machinery: a review, Mech. Syst. Signal Process. 108 (2018) 33–47.
[2] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Process. 115 (2019) 213–237.
[3] Y. Li, K. Ding, G. He, X. Jiao, Non-stationary vibration feature extraction method based on sparse decomposition and order tracking for gearbox fault diagnosis, Measurement 124 (2018) 453–469.
[4] L. Song, H. Wang, P. Chen, Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery, IEEE Trans. Instrum. Meas. 67 (2018) 1887–1899.
[5] W. Li, S. Zhang, S. Rakheja, Feature denoising and nearest-farthest distance preserving projection for machine fault diagnosis, IEEE Trans. Ind. Inf. 12 (2016) 393–404.
[6] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R.E. Vásquez, Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals, Mech. Syst. Signal Process. 76–77 (2016) 283–293.
[7] Y. Liao, L. Zhang, W. Li, C. Li, J.V. de Oliveira, Regrouping particle swarm optimization based variable neural network for gearbox fault diagnosis, J. Intell. Fuzzy Syst. 34 (2018) 3671–3680.
[8] K.C. Gryllias, I.A. Antoniadis, A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments, Eng. Appl. Artif. Intell. 25 (2012) 326–344.
[9] C.T. Yiakopoulos, K.C. Gryllias, I.A. Antoniadis, Rolling element bearing fault detection in industrial environments based on a K-means clustering approach, Expert Syst. Appl. 38 (2011) 2888–2911.
[10] C. Sun, M. Ma, Z. Zhao, X. Chen, Sparse deep stacking network for fault diagnosis of motor, IEEE Trans. Ind. Inf. 14 (2018) 3261–3270.
[11] L. Wen, X. Li, L. Gao, A new two-level hierarchical diagnosis network based on convolutional neural network, IEEE Trans. Instrum. Meas. (2019) 1–9.
[12] R. Huang, Y. Liao, S. Zhang, W. Li, Deep decoupling convolutional neural network for intelligent compound fault diagnosis, IEEE Access 7 (2019) 1848–1858.
[13] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mech. Syst. Signal Process. 72–73 (2016) 303–315.
[14] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, S. Van Hoecke, Convolutional neural network based fault detection for rotating machinery, J. Sound Vibrat. 377 (2016) 331–345.
[15] M. Zhao, M. Kang, B. Tang, M. Pecht, Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes, IEEE Trans. Ind. Electron. 65 (2018) 4290–4300.
[16] D. Verstraete, A. Ferrada, E.L. Droguett, V. Meruane, M. Modarres, Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings, Shock Vibrat. 2017 (2017) 1–17.
[17] Z. Chen, K. Gryllias, W. Li, Mechanical fault diagnosis using convolutional neural networks and extreme learning machine, Mech. Syst. Signal Process. 133 (2019) 106272.
[18] X. Hong, B. Zhang, Y. Liu, Z. Zhou, M. Ye, H. Qi, Liquid level detection in porcelain bushing type terminals using piezoelectric transducers based on auto-encoder networks, Measurement 141 (2019) 12–23.
[19] H. Shao, H. Jiang, H. Zhao, F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mech. Syst. Signal Process. 95 (2017) 187–204.
[20] V.T. Tran, F. AlThobiani, A. Ball, An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks, Expert Syst. Appl. 41 (2014) 4113–4122.
[21] Z. Chen, W. Li, Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network, IEEE Trans. Instrum. Meas. 66 (2017) 1693–1702.
[22] B. Zhang, S. Zhang, W. Li, Bearing performance degradation assessment using long short-term memory recurrent network, Comput. Ind. 106 (2019) 14–29.
[23] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, J. Wang, Machine health monitoring using local feature-based gated recurrent unit networks, IEEE Trans. Ind. Electron. 65 (2018) 1539–1548.
[24] Y. Yang, H. Zheng, Y. Li, M. Xu, Y. Chen, A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network, ISA Trans. (2019).
[25] V. Tra, J. Kim, S.A. Khan, J.M. Kim, Bearing fault diagnosis under variable speed using convolutional neural networks and the stochastic diagonal Levenberg-Marquardt algorithm, Sensors (Basel) 17 (2017).
[26] L. Wen, X. Li, L. Gao, Y. Zhang, A new convolutional neural network-based data-driven fault diagnosis method, IEEE Trans. Ind. Electron. 65 (2018) 5990–5998.
[27] X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement 93 (2016) 490–502.
[28] R. Liu, G. Meng, B. Yang, C. Sun, X. Chen, Dislocated time series convolutional neural architecture: an intelligent fault diagnosis approach for electric machine, IEEE Trans. Ind. Inf. 13 (2017) 1310–1320.
[29] M.M.M. Islam, J.-M. Kim, Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network, Comput. Ind. 106 (2019) 142–153.

[30] T. Ince, S. Kiranyaz, L. Eren, M. Askar, M. Gabbouj, Real-time motor fault detection by 1-D convolutional neural networks, IEEE Trans. Ind. Electron. 63 (2016) 7067–7075.
[31] F. Jia, Y. Lei, N. Lu, S. Xing, Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization, Mech. Syst. Signal Process. 110 (2018) 349–367.
[32] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, Sensors (Basel) 17 (2017).
[33] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Signal Process. 100 (2018) 439–453.
[34] P. Borghesani, J. Antoni, A faster algorithm for the calculation of the fast spectral correlation, Mech. Syst. Signal Process. 111 (2018) 113–118.
[35] A. Mauricio, J. Qi, K. Gryllias, Vibration-based condition monitoring of wind turbine gearboxes based on cyclostationary analysis, J. Eng. Gas Turbines Power 141 (3) (2019) 031026.
[36] K. Gryllias, A. Mauricio, J. Qi, Advanced cyclostationary-based analysis for condition monitoring of complex systems, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 385–389.
[37] A. Mauricio, J. Qi, W. Smith, R. Randall, K. Gryllias, Vibration based condition monitoring of planetary gearboxes operating under speed varying operating conditions based on cyclo-non-stationary, Analysis 61 (2019) 265–279.
[38] J. Antoni, Cyclic spectral analysis in practice, Mech. Syst. Signal Process. 21 (2007) 597–630.
[39] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[40] G. He, K. Ding, H. Lin, Fault feature extraction of rolling element bearings using sparse representation, J. Sound Vibrat. 366 (2016) 514–527.
[41] J. Antoni, Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions, J. Sound Vibrat. 304 (2007) 497–529.
[42] S. Ioffe, C. Szegedy, 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.
[43] Y. Wu, K. He, Group normalization, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[44] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study, Mech. Syst. Signal Process. 64–65 (2015) 100–131.
[45] L.V.D. Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (Nov) (2008) 2579–2605.