



Chinese Society of Aeronautics and Astronautics
& Beihang University

Chinese Journal of Aeronautics

cja@buaa.edu.cn
www.sciencedirect.com



Special Column: Application of PHM Technologies in Aerospace Industry—Research Article

A new bearing fault diagnosis method based on modified convolutional neural networks



Jiangquan ZHANG, Yi SUN, Liang GUO *, Hongli GAO, Xin HONG, Hongliang SONG

School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China

Received 2 January 2019; revised 16 March 2019; accepted 16 June 2019

Available online 2 August 2019

KEYWORDS

Bearing;
Convolutional neural networks;
Different load domains;
Fault identification;
Raw signals;
Fault diagnosis

Abstract Fault diagnosis is vital in manufacturing system. However, the first step of the traditional fault diagnosis method is to process the signal, extract the features and then put the features into a selected classifier for classification. The process of feature extraction depends on the experimenters' experience, and the classification rate of the shallow diagnostic model does not achieve satisfactory results. In view of these problems, this paper proposes a method of converting raw signals into two-dimensional images. This method can extract the features of the converted two-dimensional images and eliminate the impact of expert's experience on the feature extraction process. And it follows by proposing an intelligent diagnosis algorithm based on Convolution Neural Network (CNN), which can automatically accomplish the process of the feature extraction and fault diagnosis. The effect of this method is verified by bearing data. The influence of different sample sizes and different load conditions on the diagnostic capability of this method is analyzed. The results show that the proposed method is effective and can meet the timeliness requirements of fault diagnosis.

© 2019 Chinese Society of Aeronautics and Astronautics. Production and hosting by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

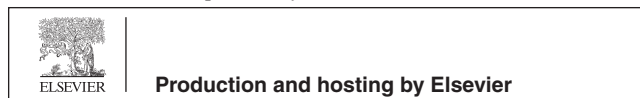
With the rapid development of industrial internet, it is possible to collect a large amount of sensor data from various machines. The availability of sensor data, which contains information about machine health, has attracted more and

more enterprises' attention. Enterprises want to use data-driven technology to complete the fault diagnosis of machines.¹ Therefore, in order to maintain industrial machinery working properly and reliably, the demand for intelligent machine health monitoring technique has never stopped.^{2–4} In different types of mechanical components, rolling bearing is the key components of the rotating mechanism, whose health conditions, such as failure diameters at different locations under different loads, can have a significant impact on the performance, stability and longevity of the mechanism.⁵ In order to avoid possible damage, the most common method is to monitor the health status of the mechanism in real time by analyzing the vibration signal.

* Corresponding author.

E-mail address: guoliang@swjtu.edu.cn (L. GUO).

Peer review under responsibility of Editorial Committee of CJA.



Aimed to monitor the health conditions of the machinery comprehensively, the installed sensors collect a large amount of signals after a long time. Because intelligent fault diagnosis methods can automatically process signals and identify the health status of the machine, a great deal of research has been made to study these methods. Asr et al.⁶ designed a feature extraction method based on Empirical Mode Decomposition (EMD). The extracted features are input into non-naive Bayesian classifier for intelligent fault diagnosis of rotating mechanism. Georgoulas et al.⁷ proposed a symbolic aggregate approximation framework, extracting features from bearing signals, and then classifying faults by using the nearest neighbor classifier. Xiong et al.⁸ considered that bearing vibration signals have multifractal properties, so multifractal detrended fluctuation analysis is used to extract multifractal features for intelligent fault diagnosis of bearings. Because continuous wavelet transform can overcome the shortcomings of the traditional Fourier transform, Wang et al.⁹ designed a Support Vector Machine (SVM) as a classifier for analyzing vibration signals.

Compared with the traditional machine learning method, deep learning method has achieved good results, but its application in fault diagnosis is still developing. Convolution Neural Network (CNN), as one of the most effective deep learning methods, is also used in fault diagnosis. Because the most common data type is time-domain signal, the one-dimensional CNN has been applied to the real-time motor fault diagnosis. In some cases, the data can be presented in two-dimensional format (such as time–frequency spectrum), and then these images can be classified by image processing methods.^{10–12} However, these methods also rely on the knowledge of experts.

Although many of the methods mentioned have achieved good results, there is still much room for improvement. For example, in many studies, classifiers train with very specific type of data, which means that it can achieve high accuracy on similar data while performing poorly on another type. This phenomenon is caused by extracting the error representation feature from the raw signal. In addition, the selection of suitable feature functions requires considerable mechanical expertise and a wealth of mathematical knowledge. On the other hand, due to manual feature extraction and selection, the accuracy of classification results will be unstable when processing different data. Therefore, some studies have suggested that classifiers should be able to classify the data in the original signal directly without feature extraction or manual selection.^{13–16} In other words, the method should be able to process the raw signal automatically and adaptively. And then it should extract the representation feature more precisely. At present, there are three main problems in intelligent fault diagnosis.

First, the unbalanced distribution of mechanical health is not considered. In the real world, the machine works under normal conditions during most of the operating phases.^{17–19} Therefore, data samples from mechanical failures are more difficult to collect than that from normal conditions. Data samples for different mechanical health conditions follow a long tail distribution which means that the data samples of normal conditions are abundant while the fault samples are relatively rare. The unbalanced distribution of data samples forces CNNs to be biased towards the majority health conditions.²⁰ As a result, the features of the minority health conditions are learned inadequately, leading to their misclassification.

Second, while many methods can achieve good results in fault diagnosis, there are few ways to directly process the time-domain signal. Most methods have the same classifier, such as SVM and Back Propagation Neural Network (BPNN). This paper focuses on improving feature representation and extraction.

Third, many diagnostic methods have poor domain adaptability. It is not rare for classifiers to be trained with data from one workload and fail to correctly classify samples from another workload.

Aimed to solve the above problems, a Convolution Neural Network with two Dropout layers and two Fully-connected layers (DFCNN) is proposed in this paper. The contribution of this paper is summarized as follows:

- (1) A new data preprocessing method is proposed, which converts the one-dimensional vibration signal into two-dimensional gray-scale images without any pre-determined parameter, and the method can eliminate the influence of expert's experience as far as possible.
- (2) We propose a novel and simple learning framework which uses the time-domain raw signal directly for diagnosis and does not need feature extraction process.
- (3) The method has good domain adaptability, and using two dropout layers and two Fully-connected (Fc) layers can easily improve performance.
- (4) By visualizing the features that CNN learned, this paper attempts to explore the intrinsic mechanism of CNN model in mechanical feature learning and classification.

2. A brief introduction to CNN

Due to the increasing number of equipment, devices monitoring points and acquisition frequency, a large amount of diagnostic data is obtained, which drives the fault diagnosis field into the “Big Data” era. CNN, with large data as engine, is one of the best classification methods. This section will introduce the differences between CNN and traditional methods, and briefly introduce the structure of CNN.

The convolutional neural network is a multi-stage neural network consisting of multiple filtering stages and classification stages, and more details on CNN can be found in Ref. ²¹. The purpose of the filtering stage is to extract features from the inputs, which contains two kinds of layers, the convolutional layer and the pooling layer. The classification stage is a multi-layer perceptron consisting of several Fc layers. The function of each type of layer will be described below.

2.1. Difference between CNN and traditional method

According to the research of bearing diagnosis method in recent years, all methods are combined with feature extraction and fault recognition, as shown in Fig. 1. When choosing the BPNN as the classifier, we initially reduce the dimensions of the data, in order to reduce the computational amount and avoid the occurrence of the fitting phenomenon. Using the data-driven feature extraction technology, most of the fault diagnosis models adopt the stacked auto-encoder. Although the step of dimension reduction is eliminated, the signal preprocessing is still needed.

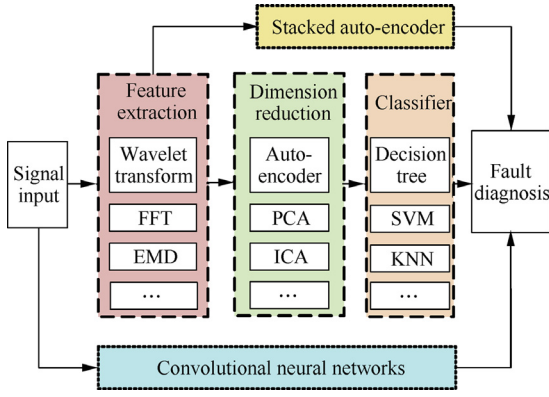


Fig. 1 Decomposition of steps of bearing diagnosis method.

In order to solve above problems, this paper proposes a fault diagnosis method based on CNN structure, which integrates the feature extraction and the fault classification without analyzing the intrinsic mechanism of mechanical device. The time-domain signal is used to diagnose directly, avoiding the problem of information loss caused by preprocessing.

2.2. Design of CNN for bearing fault diagnosis

2.2.1. Convolutional operation

Convolution layer uses convolution kernel to perform convolution operations on the local region of the input signal, resulting in corresponding characteristics. Weight sharing is the most important characteristics of the convolutional layer, which means that the parameters of the convolution window are fixed when each convolution window traverses the entire image. This avoids over-fitting phenomenon caused by the parametric explosions and reduces the memory required for the system to train the network. The first layer is taken as an example. As mentioned earlier, all the units in a feature map share the same set of weights and the same bias so they retain the same feature at all possible locations on the input. The convolution process is described as

$$y^{l(i,j)} = K_i^l x^{l(j)} = \sum_{j=0}^{W-1} K_i^{l(j)} x^{l(j+f)} \quad (1)$$

where $K_i^{l(j)}$ is the j weight of the i convolution kernel of the l layer; $x^{l(j)}$ is the j th convoluted local area of the l layer; W is the width of the convolution kernel.

2.2.2. Activation operation

After convolution layer processing, the activation layer is used to transform the logit value of each convolution output nonlinearly and to accelerate the convergence of the CNN. Leaky ReLU layers make the weights in the shallow layer more trainable when using back-propagation learning method to adjust the parameters. Select Leaky ReLU as the activation function, and the formula of Leaky ReLU is described as

$$a^{l(i,j)} = f(y^{l(i,j)}) = \begin{cases} k_i y^{l(i,j)} & y \leq 0 \\ y^{l(i,j)} & y > 0 \end{cases} \quad (2)$$

where k_i is a fixed parameter in (0, 1) interval and $y^{l(i,j)}$ is the activation value of convolution layer output y .

2.2.3. Batch normalization

The Batch Normalization (BN)²² layer is designed to reduce the shift of internal covariance, accelerate the training process of the deep neural network, improve the network training efficiency and enhance the network generalization ability. The BN layer process first subtracts the mean value μ_B of the mini-batch from the volume layer input and divides it by the standard deviation σ_B^2 . But this will cause the input value to be limited to a small range. Consequently, after standardization, it needs to multiply by a scaling amount γ , plus an offset value β . The input of the batch normalization layer is $y^l = [y^{l(i,1)}, y^{l(i,2)}, \dots, y^{l(i,j)}, \dots, y^{l(i,p)}]$. The max-pooling transformation is described as follows:

$$\hat{y}^{l(i,j)} = \frac{y^{l(i,j)} - \mu_B}{\sqrt{(\sigma_B^2 + \epsilon)}} \quad (3)$$

$$z^{l(i,j)} = \gamma^{l(i)} \hat{y}^{l(i,j)} + \beta^{l(i)} \quad (4)$$

where $\gamma^{l(i)}$ and $\beta^{l(i)}$ are the scaling and offset of the BN layer respectively, $z^{l(i,j)}$ is the output of the BN layer, and ϵ is a constant term that guarantees numerical stability.

2.2.4. Pooling layer

It is common to add a pooling layer after a batch normalization layer in the CNN architecture. The pooling layer is operated by the down-sampling operation, and the main purpose is to reduce the parameters of the neural network. In this paper, the largest pooled layer is chosen. For example, its input feature window size is 6×6 , and the input feature is down sampled to the 3×3 output feature by a pooling operation with the size of 2×2 and the step size of 2. The max-pooling transformation is described as

$$P^{l(i,j)} = \max_{(j-1)W+1 \leq t \leq jW} \{a^{l(i,t)}\} \quad (5)$$

where $a^{l(i,t)}$ is the activation value of t neuron in l layer of i frame and $P^{l(i,j)}$ is the width of the pool area.

3. Proposed intelligent diagnosis method

In this section, a novel fault diagnosis method based on CNN is proposed. First, a conversion method is proposed, which can transfer raw vibration signal to images. Then, two dropout layers and two Fc layers are added to the traditional CNN model.

3.1. Data processing

3.1.1. Data augmentation

In order to enhance the feature extraction capability of CNN, the structure is more and more complex, and subsequently, the training parameters are also increasing. However, this structure would easily get over-fitting without sufficient training samples. In the field of computer vision, data augmentation method is used to increase the number of training samples which are able to improve the generalization of CNN.^{23,24} Horizontal/vertical flips, random crops and color jitters are used to increase training samples in the field of computer vision. The augmentation of data is also essential for the high classification accuracy of CNN. It is easy to obtain a large number of training data by dividing overlapping training

samples. This process is shown in Fig. 2. For example, a vibration signal with 101,100 points can provide 200 training samples for the DFCNN. When the shift stride is 500, the length of each training sample is 1600.

3.1.2. Signal-to-image conversion method

The traditional intelligent fault diagnosis method is based on the statistical analysis, fuzzy logic expert system or genetic method to extract the features from the raw data. The process of feature extraction relies on experts' experience. In addition, the generality of extracted features is not good. Considering the impact of features on the final results, this paper proposes an effective data pre-processing method, which converts the raw signal of time-domain into an image. This method provides a way to explore two-dimensional features of raw signal.^{25,26} It should be noted that this data pre-processing method can be achieved without any predetermined parameter.

As shown in Fig. 2, in this data pre-processing method, the time-domain signals fill the pixels of the image sequentially. A segmentation sample of length K^2 is randomly extracted from the time-domain signals. Then, an image with pixel point of $K \times K$ are obtained by processing the sample. The intercepted signal segments are normalized from 0 to 255, which is just the pixel strength of the gray image. The choice of 160×160 in this paper is dependent on the volume of signal data. $L(i)$ ($i = 1, 2, \dots, K^2$) denotes the value of the segment signal. $P(j, k)$ ($j = 1, 2, \dots, K; k = 1, 2, \dots, K$) denotes the pixel strength of the image. The process is described as

$$P(j, k) = \frac{L(jK + k) - \text{Min } L}{\text{Max } L - \text{Min } L} \times 255 \quad (6)$$

3.2. Introduction of network structure

CNN has been applied to fault diagnosis. However, compared with the traditional models, the feature extraction ability of

these models has not been significantly improved. Most of the models have simple structure, which make them unable to extract the nonlinear expression of the input signal.

The overall framework of DFCNN is shown in Fig. 3. In order to improve the classification accuracy and the adaptability of the structure, two methods are used. First, 10×10 sampling points of kernel are adopted in the first convolutional layer. Then, max pooling layer is adopted to reduce the dimension of data. Secondly, two dropout layers and two Fc layers are adopted to improve the adaptability of the network.

3.2.1. Architecture of DFCNN model

The structure composition of DFCNN model is similar to normal CNN models. It is made up of several filtering stages and one classification stage. The main difference is that, in the first convolutional layer, the sampling points of kernels are wide, and the following kernels are with 3×3 sampling points. Then the max pooling layer is used to reduce the dimension of data. Finally, in the classification stage, the logits of the neurons are transformed by Softmax function, which match the probability distribution of different bearing health conditions.

The DFCNN uses wide kernels in the first convolutional layer which have 10×10 sampling points of kernels to extract features. Furthermore, the number of convolution cores in the first convolution layer is relatively less, so as to suppress high-frequency noise. The Visual Geometry Group (VGG) net²⁷ performs well with 3×3 sampling points of kernels. Therefore, in order to obtain the useful information of time-domain signals in the middle and low frequency bands, successive small kernels are adopted. And, there are a large number of convolution cores in the second to fourth convolution layers, which helps to obtain features of the time-domain signals and improve the performance of the network. Additionally, the structure composition of model is deeper than the CNN models mentioned before.

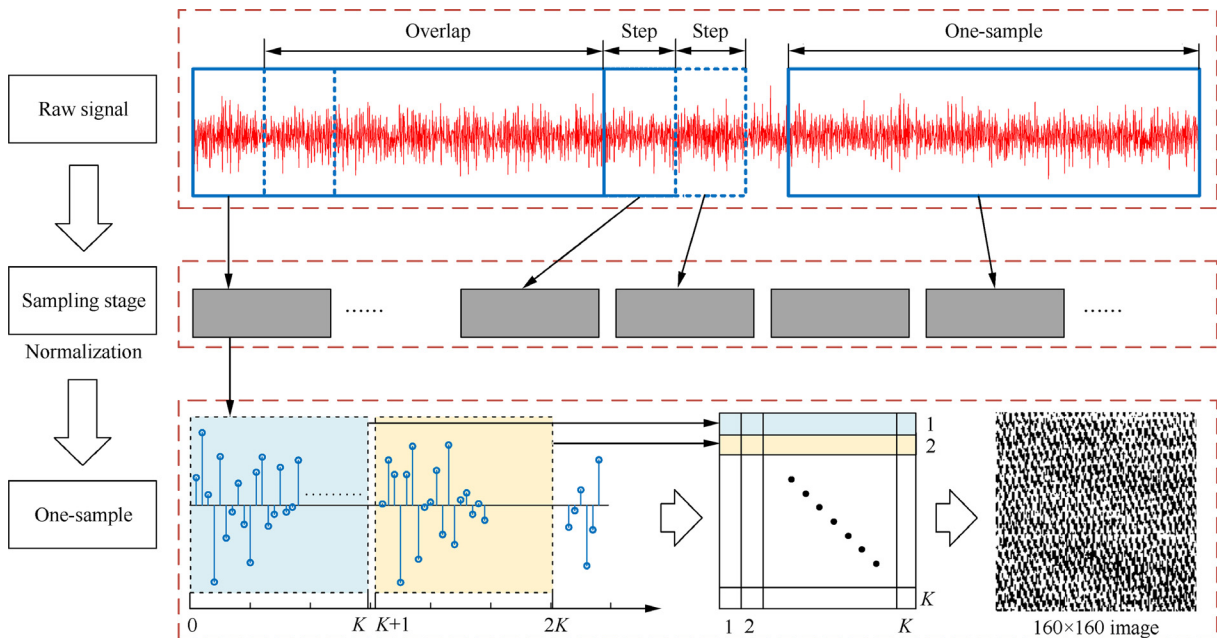


Fig. 2 Signal-to-image conversion process.

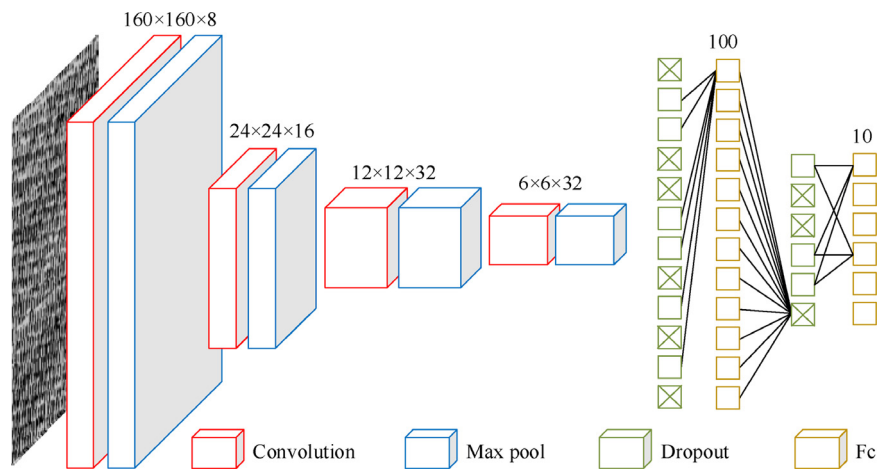


Fig. 3 Structure of DFCNN.

3.2.2. Two dropout layers and two Fc layers

The Fc layer classifies the data with features extracted by the filtering stage, which maps the “distributed feature representation” learned to the label space. The Fc layer uses many neurons to fit the data distribution, but it is sometimes impossible to solve the non-linear problem with only one Fc layer. As the number of Fc layers increase, the non-linear fitting ability of the network can be improved. Then, the dropout layer is added before the Fc layer. This operation means that a randomly selected unit and all its connections are temporarily removed from the model. Its purpose is to provide DFCNN with incomplete signals during training stage, so as to improve the generalization of model. For each new input element, it selects a randomly forming subset of neurons. These subsets use common weights. This operation does not depend on specific neurons and connections, thus solving the problem of over-fitting.^{28,29} The dropout layer is added to the CNN model, which can improve the generalization ability for various classification problems.

So two dropout layers and two Fc layers are adapted in the DFCNN. The length of the first Fc layer is 100, and the length of the second Fc layer is 10. Before using the first and the second Fc layer, dropout operation is performed.

4. Validation of proposed DFCNN method

The experimental data set comes from open source data of Case West Reserve University (CWRU) rolling bearing data center.²⁷ At present, the fault diagnosis methods develop

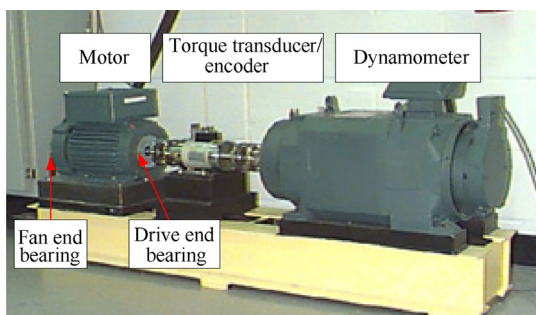


Fig. 4 CWRU bearing fault test-bed.

rapidly and new theories are constantly being proposed. In order to evaluate the performance of the method proposed in this paper, the standard database should be adopted and the mainstream methods should be used as a comparison.^{30,31} The CNN model is written in MATLAB 2018a with Neural Network Toolbox, and is implemented on a computer where the CPU is Ryzen 5 1600X, the memory is 16 GB, and the GPU is GTX1060.

4.1. Data description

In this experiment, the signal is collected from the drive end bearing as shown in Fig. 4, and the bearing fault is formed by Electrical Discharge Machining (EDM). The raw vibration data are obtained from the motor driving mechanical system by the accelerometers at a sampling frequency of 12 kHz. The test platform simulates four types of faults under different loads. The four different types are the normal state of the bearings, the ball fault, the inner race fault, and the outer race fault.

Data sets *A*, *B*, *C* and *D* are constructed in 10 different fault categories under different loads (0–3 hp (1 hp = 746 W)). Considering the actual situation, the bearing works under different loads, and the data set *E* is composed of data sets *A*, *B*, *C* and *D*. Table 1 shows the information of the experimental data sets. Different working conditions of bearings, the normal state of the bearing, the ball fault, the inner race fault and the outer race fault, are represented as Normal, B, IR and OR respectively. The depths of faults are represented as 0.18, 0.36 and 0.54 mm.

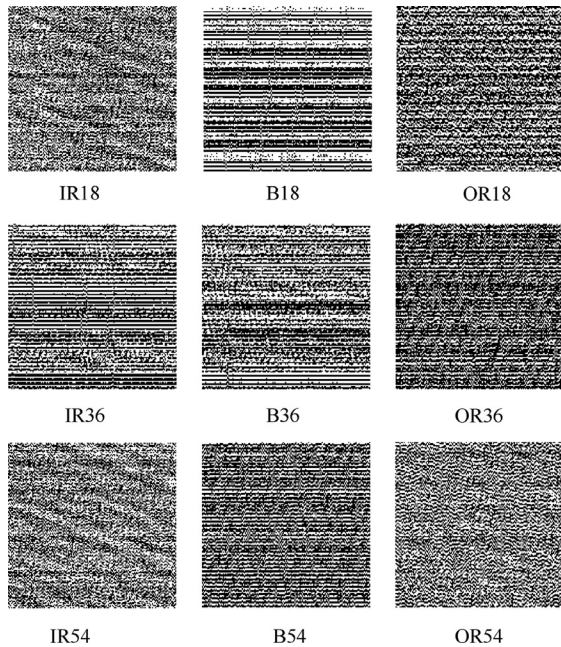
The size of the converted gray image is set as 160×160 , which are processed by signal-to-picture method from a sample signal with 25600 sampling points. Fig. 2 shows signal conversion result under normal work condition and Fig. 5 shows signal conversion results under other nine fault work conditions. As can be seen from the conversion results, the converted images under different work conditions look completely different.

4.2. Compared with other mainstream methods

Taking classification accuracy as an index, DFCNN is compared with several mainstream models. The result shows that,

Table 1 Description of data sets.

| Fault number | Fault category | Fault depth (mm) | Total number of sample | |
|--------------|----------------|------------------|--------------------------|-------------------|
| | | | Data sets <i>A/B/C/D</i> | Data set <i>E</i> |
| 1 | B18 | 0.18 | 50 | 200 |
| 2 | B36 | 0.36 | 50 | 200 |
| 3 | B54 | 0.54 | 50 | 200 |
| 4 | IR18 | 0.18 | 50 | 200 |
| 5 | IR36 | 0.36 | 50 | 200 |
| 6 | IR54 | 0.54 | 50 | 200 |
| 7 | Normal | 0 | 50 | 200 |
| 8 | OR18 | 0.18 | 50 | 200 |
| 9 | OR54 | 0.36 | 50 | 200 |
| 10 | OR36 | 0.54 | 50 | 200 |

**Fig. 5** Converted images under nine fault conditions.

in Table 2, DFCNN achieves better results than these methods.

In Ref. ³³, the multifractal is used to extract fault features, then the SVM classifier is used to classify the data from 10 kinds of fault signals under 0 load, and the classification accuracy of this method is 89.1%. In the same data set, the DFCNN gets 100% accuracy from the time-domain raw data. In Ref. ³⁴, statistics method is used to extract the time-domain features. After that, Multi-Layer Perceptron (MLP) is used to classify 6 kinds of faults from three different data sets, acquiring classification accuracy which are 95.7%, 99.6% and 99.4%, while the CNN could still get 100% accuracy from the same data sets. In Refs. ^{34,35}, the Deep Belief Network (DBN) method diagnoses faults directly from time-domain raw signals, and the classification accuracy of this method is 98.8% in data set *E*. The accuracy of the method proposed in this paper has achieved 99.8%, which is much better than the existing methods.

4.3. Effect of data number for training

As a branch of CNN, DFCNN has thousands of parameters to be trained. In order to avoid over-fitting and enhance the generalization ability of DFCNN model, a large number of training samples are needed. The experimental sample is randomly selected from the data set *E*. The total number of samples were 50, 100, 200, 400, 800, 1600 and the effect of data set size on CNN performance was observed. The initial weights of the CNN model are randomly generated, and the maximum number of iterations is set to 100 times. To verify the stability of the CNN, each experiment was repeated 10 times, and the results are shown in Fig. 6.

The gradient descent algorithm requires the learning rate to be set in an appropriate range. Excessive learning rate may cause instability, while too small learning rate will increase training time. On the premise of ensuring the stability of training, the adaptive learning rate can achieve a reasonable effect. At the same time, it can reduce the training time. Therefore, the initial learning rate is set to 0.0016, and after every 30 iterations, the rate drops to half of its original value.

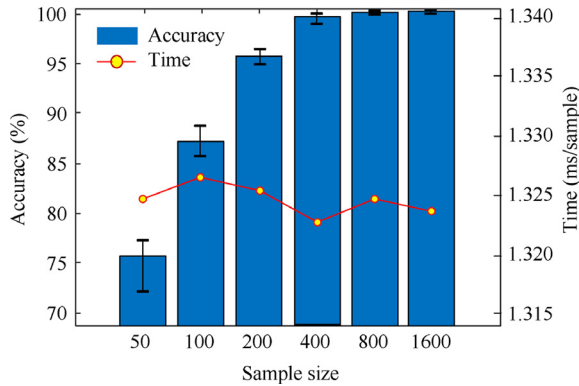
As shown in Fig. 6, as the number of training samples increases, the classification accuracy gradually increases, and the standard deviation of 10 times accuracy decreases. Even if training samples are relatively small, DFCNN can still achieve a high classification accuracy. The above results show that as the amount of input data increases, the classification accuracy of DFCNN is improved and the ability of feature extraction is gradually enhanced. The increase of training samples can improve the generalization ability of the model. ³⁶ And the size of training sample does not affect the time required to diagnose a sample. The time spent on diagnosis by DFCNN with different sizes of training samples is shown on the right side of Fig. 6, and the CNN model only needs 1.325 ms to diagnose a sample. DFCNN can meet the real-time requirement of fault diagnosis.

4.4. Performance across different load domains

In this study, the adaptive performance of DFCNN across different load domains was tested, compared with the SVM, MLP and DBN (the data is transformed by fast Fourier transformation). Fig. 7 describes the information of training set and test set, and shows the experimental results.

Table 2 Diagnostic results of bearing datasets.

| Method | Classification accuracy (%) | | | | |
|--|-----------------------------|-------------------|-------------------|-------------------|-------------------|
| | Data set <i>A</i> | Data set <i>B</i> | Data set <i>C</i> | Data set <i>D</i> | Data set <i>E</i> |
| Multifractal + SVM ³² | 89.1 ± 1.46 | | | | |
| Time + MLP ³³ | 95.7 ± 0.56 | 99.6 ± 0.34 | 99.4 ± 0.21 | | |
| PSO-WKLFDA ³⁴ | 97.2 ± 0.47 | | | | |
| Time-domain signal + DBN ³⁵ | 100 ± 0 | 100 ± 0 | 100 ± 0 | | 98.8 ± 0.26 |
| DFCNN | 100 ± 0 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 99.8 ± 0.11 |

**Fig. 6** Number of different training samples, CNN's accuracy for test set *E* and time to diagnose individual signals.

The methods of SVM, MLP and DBN have poor adaptability, the average accuracies of the six cases are 66.6%, 75.9% and 75.7% respectively. In contrast, the DFCNN method has higher accuracy than other three methods, with an average accuracy of 90.5%. The results show that the features learned by DFCNN from time-domain raw signals have more domain invariance than the traditional features, and the accuracy of each case is more than 80%.

As shown in Fig. 8, DFCNN is trained in the data set *A* and tested in the data set *C*. The feature representation of the four convolutional layers is reduced to a two-dimensional distribution by t-SNE, as shown in Fig. 8. DFCNN can effectively extract features of data sets with different loads, different fault categories and different fault depths. From Fig. 8(a) and (b), it can be seen that the signal features of different fault categories have a high degree of overlap. From Fig. 8(c), the features of different fault types can be clearly classified. However, the features of the same fault at different depths are still in an

aggregate state. From Fig. 8(d), the features are highly overlapped of ball fault with a depth of 0.54 mm and inner race fault with a depth of 0.18 mm. The rest of fault features are clearly divided into different areas.

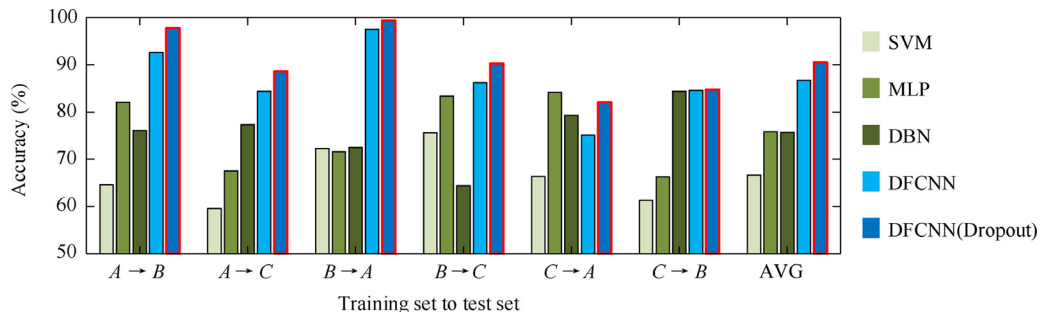
4.5. Necessity of two dropout layers and two Fc layers

In order to verify the effectiveness of the tricks used in the network, the following experiments were carried out to compare the network performance of two dropout layers and two Fc layers, without the first Fc layer and without two dropout layers. In this experiment, the small batch size was set to 20. Fig. 9(a) shows the effect of network training under normal data sets. Fig. 9(b) shows the effect of network training under data set *A* and testing under data set *B*.

From Fig. 9(a), it can be seen that under normal circumstances, the total number of training samples is 800, and the convergence speed of the network without the first layer of Fc layer is obviously higher than other models. The classification accuracy is much lower than the other two network structures, because the last BN layer outputs too many parameters, which cause a huge deviation in the automatic feature extraction of the network without the first Fc layer. From Fig. 9(b), it can be seen that when data set is across different load domains, the classification accuracy of the other two models are much lower than that of DFCNN. Without dropout layer, the generalization ability of various classification problems will be greatly reduced in the testing process.

5. Conclusions and future research work

The main contributions of this paper are application of the transformation from signals to images, and then design of a novel CNN network to solve the existing problems in fault diagnosis. DFCNN directly processes the time-domain raw

**Fig. 7** Comparison with SVM, MLP, DBN, DFCNN and DFCNN (Dropout) results under varying load domains.

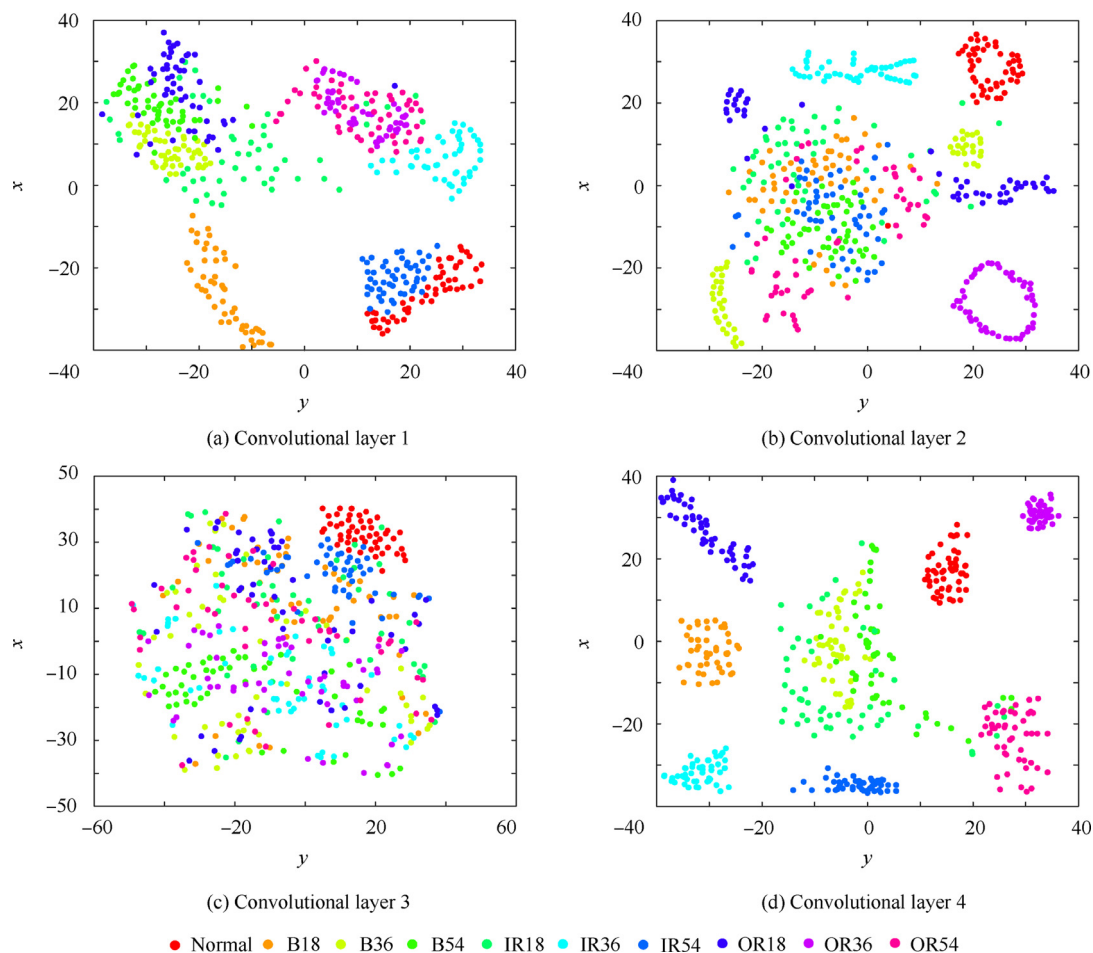


Fig. 8 Visualization under changing data set.

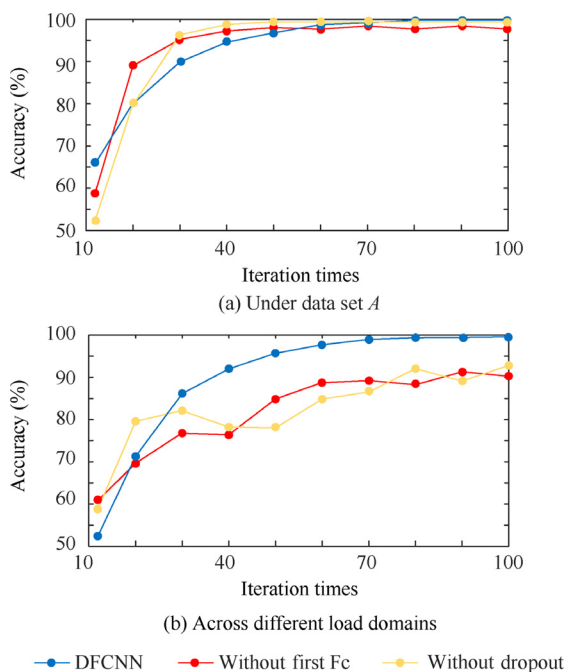


Fig. 9 Accuracy in different conditions.

signal without the need of any time-consuming feature extraction process, which reduces the dependence on experts' experience.

The results show that although the most advanced deep neural network model has high accuracy for normal data set, its performance decreases rapidly when the workload changes.^{37,38,13} However, DFCNN has a high classification accuracy on normal data, and also has a strong adaptability to workload changes.

Future research can be carried out in the following ways. Firstly, this method can be improved to achieve better adaptability. Secondly, the transfer learning based on CNN can be studied to reduce the training times.

Acknowledgements

Yi SUN and Jiangquan ZHANG contributed equally to this work, as co-first author of this article. Liang GUO and Hongli GAO contributed equally to this work, as co-corresponding author of this article. This research was co-supported by the National Natural Science Foundation of China (No. 51775452), Fundamental Research Funds for the Central Universities, China (Nos. 2682019CX35 and 2018GF02), and Planning Project of Science & Technology Department of Sichuan Province, China (No.2019YFG0353).

References

1. Yin S, Kaynak O. Big data for modern industry: Challenges and trends. *Proc IEEE* 2015;**103**(2):143–6.
2. Gao ZW, Cecati C, Ding SX. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE Trans Ind Electron* 2015;**62**(6):3757–67.
3. Lei YG, Lin J, Zuo MJ, He ZJ. Condition monitoring and fault diagnosis of planetary gearboxes: A review. *Measurement* 2014;**48**:292–305.
4. Fan L, Wang SP, Wang XJ, Han F, Lyu HW. Nonlinear dynamic modeling of a helicopter planetary gear train for carrier plate crack fault diagnosis. *Chinese J Aeronaut* 2016;**29**(3):675–87.
5. Ma ZH, Wang SP, Shi J, Li TY, Wang XJ. Fault diagnosis of an intelligent hydraulic pump based on a nonlinear unknown input observer. *Chinese J Aeronaut* 2018;**31**(2):385–94.
6. Asr MY, Etefagh MM, Hassannejad R, Razavi S. Diagnosis of combined faults in rotary machinery by non-naive bayesian approach. *Mech Syst Sig Process* 2017;**85**:56–70.
7. Georgoulas G, Karvelis P, Loutas T, Stylios CD. Rolling element bearings diagnostics using the Symbolic Aggregate approximation. *Mech Syst Sig Process* 2015;**60–61**:229–42.
8. Xiong Q, Zhang WH, Lu TW, Mei GM, Liang SL. A fault diagnosis method for rolling bearings based on feature fusion of multifractal detrended fluctuation analysis and Alpha stable distribution. *Shock Vib* 2016;**2016**:1–12.
9. Wang X, Zheng Y, Zhao ZZ, Wang JP. Bearing fault diagnosis based on statistical locally linear embedding. *Sensors-Basel* 2015;**15**(7):16225–47.
10. Liu Q, Kong DZ, Qin SJ, Xu Q. Map-reduce decentralized PCA for big data monitoring and diagnosis of faults in high-speed train bearings. *IFAC Papersonline* 2018;**51**(18):144–9.
11. He M, He D. Deep learning based approach for bearing fault diagnosis. *IEEE Trans Ind Appl* 2017;**53**(3):3057–65.
12. Zhou M, Wang K, Wang Y, Luo KJ, Fu HY, Si L. Online condition diagnosis for a two-stage gearbox machinery of an aerospace utilization system using an ensemble multi-fault features indexing approach. *Chinese J Aeronaut* 2019;**32**(5):1100–10.
13. Zhang L, Gao HL, Wen J, Li SC, Liu Q. A deep learning-based recognition method for degradation monitoring of ball screw with multi-sensor data fusion. *Microelectron Reliab* 2017;**75**:215–22.
14. Abdeljaber O, Avci O, Kiranyaz S, Gabbouj M, Inman DJ. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J Sound Vib* 2017;**388**:154–70.
15. Zhang W, Peng GL, Li CH. Rolling element bearings fault intelligent diagnosis based on convolutional neural networks using raw sensing signal. *Smart Innov Syst Tec.* 2017;**64**:77–84.
16. Lei YG, Jia F, Lin J, Xing SB, Ding SX. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Trans Ind Electron* 2016;**63**(5):3137–47.
17. Guo L, Lei YG, Xing SB, Yan T, Li NP. Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Trans Ind Electron* 2019;**66**(9):7316–25.
18. Martin-Diaz I, Morinigo-Sotelo D, Duque-Perez O, Romero-Troncoso RD. Early fault detection in induction motors using adaboost with imbalanced small data and optimized sampling. *IEEE Trans Ind Appl* 2017;**53**(3):3066–75.
19. Malik H, Mishra S. Proximal support vector machine (PSVM) based imbalance fault diagnosis of wind turbine using generator current signals. *Energy Procedia* 2016;**90**:593–603.
20. Tang YC, Zhang YQ, Chawla NV, Krasser S. SVMs modeling for highly imbalanced classification. *IEEE Trans Syst Man Cybernet B* 2009;**39**(1):281–8.
21. Cao R, Guo KH, Ma JH, Kang J. A deep convolutional neural network-based label completion and correction strategy for supervised medical image learning. 2018 *IEEE Smartworld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (Smartworld/Scalcom/UIC/ATC/CBDCCom/IOP/SCI)*; 2018 Oct 8–12; Guangzhou, China. Piscataway: IEEE Press; 2018. p. 1725–30.
22. Bai C, Huang L, Pan X, Zheng JW, Chen SY. Optimization of deep convolutional neural network for large scale image retrieval. *Neurocomputing*. 2018;**303**:60–7.
23. Mbo'o CP, Hameyer K. Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection. *IEEE Trans Ind Appl* 2016;**52**(5):3861–8.
24. Cui XD, Goel V, Kingsbury B. Data augmentation for deep convolutional neural network acoustic modeling. *Int Conf Acoust Speech* 2015;4545–9.
25. Do V, Chong UP. Signal model-based fault detection and diagnosis for induction motors using features of vibration signal in two-dimension domain. *Stroj Vestn-J Mech Eng* 2011;**57**(9):655–66.
26. Huang WY, Cheng JS, Yang Y. Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection. *Mech Syst Sig Process* 2019;**114**:165–88.
27. Zhou X, Lu F, Huang JQ. Fault diagnosis based on measurement reconstruction of HPT exit pressure for turbofan engine. *Chinese J Aeronaut* 2019;**32**(5):1156–70.
28. Ma XL, Dai Z, He ZB, Ma JH, Wang Y, Wang YP. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors-Basel* 2017;**17**(4):818–24.
29. Li T, Guo SH, Liu HJ. Research on fault diagnosis method of aluminum electrolytic cell based on feature extraction. 2017 32nd youth academic annual conference of Chinese association of automation (YAC); 2017 May 19–21; Hefei, China. Piscataway: IEEE Press; 2017. p. 127–31.
30. Maul T, Baba S. Unsupervised learning in second-order neural networks for motion analysis. *Neurocomputing* 2011;**74**(6):884–95.
31. Lou XS, Loparo KA. Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mech Syst Sig Process* 2004;**18**(5):1077–95.
32. Du WL, Tao JF, Li YM, Liu C. Wavelet leaders multifractal features based fault diagnosis of rotating mechanism. *Mech Syst Sig Process* 2014;**43**(1–2):57–75.
33. De Almeida LF, Bizarria JWP, Bizarria FCP, Mathias MH. Condition-based monitoring system for rolling element bearing using a generic multi-layer perceptron. *J Vib Control* 2015;**21**(16):3456–64.
34. Van M, Kang HJ. Bearing defect classification based on individual wavelet local fisher discriminant analysis with particle swarm optimization. *IEEE Trans Ind Inform* 2016;**12**(1):124–35.
35. Hong W, Cai WJ, Wang SP, Tomovic MM. Mechanical wear debris feature, detection, and diagnosis: A review. *Chinese J Aeronaut* 2018;**31**(5):867–82.
36. Wen L, Li X, Gao L, Zhang YY. A new convolutional neural network based data-driven fault diagnosis method. *IEEE Trans Ind Electron* 2018;**65**(7):5990–8.
37. Guo L, Lei YG, Li NP, Yan T, Li NB. Machinery health indicator construction based on convolutional neural networks considering trend burr. *Neurocomputing* 2018;**292**:142–50.
38. Guo L, Li NP, Jia F, Lei YG, Lin J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* 2017;**240**:98–109.