ISA Transactions 95 (2019) 295–305

Contents lists available at ScienceDirect

ISA Transactions

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

Practice article

Deep residual learning-based fault diagnosis method for rotating machinery

Wei Zhang [a], Xiang Li [b,c,*], Qian Ding [d]

[a] School of Aerospace Engineering, Shenyang Aerospace University, Shenyang 110136, China
[b] College of Sciences, Northeastern University, Shenyang 110819, China
[c] Key Laboratory of Vibration and Control of Aero-Propulsion System Ministry of Education, Northeastern University, Shenyang 110819, China
[d] Department of Mechanics, Tianjin University, Tianjin 300072, China

HIGHLIGHTS

- Large neural network can be efficiently trained in fault diagnosis.
- Machinery signal with variable sequential length can be processed.
- Accurate testing diagnosis results are obtained with small variance.
- Little prior expertise on signal processing and fault diagnosis is required.

ARTICLE INFO

Article history:
Received 17 September 2018
Received in revised form 25 November 2018
Accepted 17 December 2018
Available online 24 December 2018

Keywords:
Fault diagnosis
Rotating machinery
Residual learning
Rolling bearing
Convolutional neural network

ABSTRACT

Effective fault diagnosis of rotating machinery has always been an important issue in real industries. In the recent years, data-driven fault diagnosis methods such as neural networks have been receiving increasing attention due to their great merits of high diagnosis accuracy and easy implementation. However, it is mostly difficult to fully train a deep neural network since gradients in optimization may vanish or explode during back-propagation, which results in deterioration and noticeable variance in model performance. In fault diagnosis researches, larger data sequence of machinery vibration signal containing sufficient information is usually preferred and consequently, deep models with large capacity are generally adopted. In order to improve network training, a residual learning algorithm is proposed in this paper. The proposed architecture significantly improves the information flow throughout the network, which is well suited for processing machinery vibration signal with variable sequential length. Little prior expertise on fault diagnosis and signal processing is required, that facilitates industrial applications of the proposed method. Experiments on a popular rolling bearing dataset are implemented to validate the proposed method. The results of this study suggest that the proposed intelligent fault diagnosis method for rotating machinery offers a new and promising approach.

1. Introduction

Rotating machinery has been widely applied in modern industries. Unexpected failures of the components of bearing assembly result in serious losses of safety and large costs of maintenance [1]. Effective and accurate fault diagnosis is always demanded to enhance machine reliability, and reduce operation costs. This paper proposes an intelligent data-driven fault diagnosis method for rotating machinery.

In the past years, many signal processing methods have been applied to fault signal analysis, including wavelet analysis [2,3], stochastic resonance techniques [4–6] and so forth [7–11]. However, sufficient expert knowledge on signal processing is usually required for the traditional methods, which is not convenient for industrial applications. Therefore, intelligent fault diagnosis methods, which are able to rapidly and efficiently process collected signals, provide reliable fault diagnosis results and require little prior expertise, are becoming more and more popular nowadays. A large number of studies have been carried out based on machine learning and statistical inference techniques, such as artificial neural networks (ANN) [12–14], random forest (RF) [15], support vector machines (SVM) [16,17], fuzzy inference and other improved algorithms [18,19]. In general, neural networks are one of the most popular methods to identify faulty and healthy machine conditions. Fault diagnosis is considered a classification problem through feature extraction. First, raw input signal is mapped into

* Corresponding author at: College of Sciences, Northeastern University, Shenyang 110819, China.
E-mail address: xiangli@mail.neu.edu.cn (X. Li).

https://doi.org/10.1016/j.isatra.2018.12.025
0019-0578/

Nomenclature

$\mathbf{x}^{(i)}$	ith input sample
$\mathbf{x}_i^{in}$, $\mathbf{x}_i^{out}$	Input and output of the ith building block
y	Model ouput
N	Length of input
N_s	Number of samples
$\mathbf{w}$	Network weight
$\mathbf{W}$	Weight in the building block
b	Network bias
$\mathbf{z}$	Feature map
$\mathbf{p}$	Feature extracted by pooling
g	Pooling length
L	Loss function
F_L	Filter length
F_N	Filter number
$\mathcal{F}$	Residual function
φ	Activation function
K	Number of classes
θ	Softmax parameter
N_{train}	Number of training samples for each condition under one load
N_{test}	Number of testing samples for each condition under one load
N_{input}	Network input dimension
P_{signal}, P_{noise}	Powers of original signal and noise

high-level representative features. The health condition is classified based on the extracted features afterwards. Despite the success achieved by ANN on fault diagnosis in the past years [12,13,20], fault classification problems with highly complex nonlinear signals are difficult to be addressed with shallow network structures [21].

Recently, deep learning network is emerging as a highly effective network structure for feature extraction and pattern recognition, which holds the potential to overcome the existing obstacles [21–24]. Deep learning is characterized by the deep network architecture where multiple layers are stacked in the network to fully capture the representative information from raw input data [25]. High-level abstractions of data can be modeled well using the complex deep structures, leading to more efficient feature extraction compared with the shallow networks. Deep learning methods have attracted great attention and achieved significant results in many fields, including image recognition [26], speech recognition [27] *etc*. Since the raw data obtained from rotating machinery component monitoring share similar high dimensionality with those in image processing field, deep learning architecture has great potential in machinery health condition recognition. A recurrent neural network was proposed by Liu et al. [28] using the auto-encoder architecture for signal denoising and bearing fault diagnosis. While promising diagnosis performance is obtained, the training difficulty of neural networks remains a problem. Stacked denoising auto-encoder was used by Lu et al. [29] for certain machinery health state identifications considering environmental noise and working condition variations. Sun et al. [23] adopted sparse auto-encoders in unsupervised deep learning for induction motor fault diagnosis. The robustness of feature representation is improved by imposing partial corruption on the input data. An auto-encoder-ELM-based method was proposed by Mao et al. [30] for diagnosing faults of bearings to overcome the deficiencies of the traditional neural network methods, such as long training time.

Within the deep learning architecture, convolutional neural networks (CNNs), that are specifically designed for highly nonlinear and complex signals, are further utilized in this study. CNNs have shown remarkable success in various applications in the past few years. CNN was first proposed for image processing. Its ability to maintain data information regardless of scale and shift has been shown. A large number of researches on computer vision, speech processing *etc*. have benefited from CNN's merits of local receptive fields, shared weights *etc*. Guo et al. [31] proposed a hierarchical learning rate adaptive deep convolutional neural network to diagnose bearing faults and determine the severity. Three convolutional and pooling layers are stacked in the network, as well as the top fully-connected classifier. Sun and colleagues [32] proposed a convolutional discriminative feature learning method for induction motor fault diagnosis. Convolutional pooling architecture is used to extract the discriminative and invariant features from the raw vibration data. A support vector machine is attached on the top layer as classifier. While common machinery vibration data can be transformed into 2D format such as time-frequency spectrum in accordance with traditional CNN, 1D format such as time-series is more straight-forward for analysis in fault diagnosis field [33–35]. In this paper, 1D CNN is employed to extract the local data features through the network.

While great success has been achieved by deep neural networks, it is usually difficult to fully train a deep model since the gradients in optimization may vanish or explode during back-propagation. Recently, deep residual networks have emerged as a family of extremely deep architectures showing compelling accuracy and excellent convergence behaviors [36,37]. State-of-the-art performance on many popular machine learning tasks has been achieved by residual learning structure. The identity skip-connections are introduced in the framework which allows the data information to propagate directly throughout the network. The high-level representative features can be better extracted in this way. Therefore, residual learning is very promising to process the mechanical signals.

In this paper, a new deep learning method for fault diagnosis of rotating machinery components is proposed. Residual learning algorithm is adopted, and the deep model is expected to be efficiently trained. 1-dimensional convolutional layers are employed throughout the network in order to capture the local sequence features of the data information flow. The proposed fault diagnosis method is tested in a popular rolling bearing case study. Comprehensive analysis of the proposed approach and comparisons with other existing methods and state-of-the-art results on the same dataset are presented. The experimental results justify the effectiveness and superiority of the proposed method.

The remainder of this paper starts with the description of the proposed network in Section 2, along with brief introductions of CNN and residual learning structure. The proposed method is experimentally validated using a rolling bearing dataset in Section 3. We close the paper with conclusions in Section 4.

2. Proposed fault diagnosis framework

2.1. Convolutional neural network

Convolutional neural networks (CNNs) are generally used for processing highly nonlinear and complex signals. In the past few years, a large number of researches [31,32,34,38,39] have benefited from CNN's merits of shared weights, local receptive fields *etc*. In this study, CNN is adopted for feature extraction.

The convolutional layers convolve multiple filters with the input sample and generate features, and pooling layers are usually adopted to extract the most significant local features afterwards. Since the input data in this research are a sequence of machinery

vibration or frequency signal, the 1-dimensional (1D) CNN is used and briefly introduced.

The input sequential data is assumed to be $\mathbf{x} = [x_1, x_2, \ldots, x_N]$ where N denotes the length of the sequence. CNN can be defined as a multiply operation between a filter $\mathbf{w}$, $\mathbf{w} \in R^{F_L}$, and a concatenation vector representation $\mathbf{x}_{i:i+F_L-1}$, which can be expressed as,

$$\mathbf{x}_{i:i+F_L-1} = x_i \oplus x_{i+1} \oplus \cdots \oplus x_{i+F_L-1}, \tag{1}$$

where $x_{i:i+F_L-1}$ denotes a window of F_L length sequential data starting from the ith point, and $\oplus$ concatenates the data samples into a longer embedding. The final convolution can be defined as,

$$z_i = \varphi(\mathbf{w}^T \mathbf{x}_{i:i+F_L-1} + b), \tag{2}$$

where $*^T$ denotes the transpose of a matrix $*$, and b and φ are the bias term and activation function, respectively. The output z_i can be considered as the learned high-level feature of the filter $\mathbf{w}$ on the data sequence $\mathbf{x}_{i:i+F_L-1}$. By sliding the filter on the input data, the feature map of the jth filter can be obtained, which can be denoted as,

$$\mathbf{z}_j = [z_j^1, z_j^2, \ldots, z_j^{N-F_L+1}]. \tag{3}$$

In CNNs, multiple filters can be applied in the convolutional layer with different length F_L.

In this paper, the max-pooling function is applied, which is implemented in the feature maps with a pooling length of g. The extracted feature can be expressed as,

$$\mathbf{p}_j = [p_j^1, p_j^2, \ldots, p_j^s], \tag{4}$$

$$p_j^k = \max\left(z_j^{(k-1)g+1}, z_j^{(k-1)g+2}, \ldots, z_j^{kg}\right), \tag{5}$$

where $\mathbf{p}_j$ denotes the output of the pooling operation implemented on the jth feature map and has s dimensions.

2.2. Residual learning

The residual network was first proposed by He et al. [36] in 2015 and has achieved great success in image processing field in the past two years [37]. In general, the residual network has three significant features. First, the identity skip-connections are introduced, which allow data flowing from other layers directly to the subsequent layers. Second, the depth of the network structure can be largely extended using the skip connections. Finally, removing single layers from residual networks basically has no remarkable influence on the testing performance [36].

A residual learning block is shown in Fig. 1, that can be defined as,

$$\mathbf{z} = \mathcal{F}(\mathbf{x}, \{\mathbf{w}_i\}) + \mathbf{x}, \tag{6}$$

where $\mathbf{x}$ and $\mathbf{z}$ denote the input and output vectors of the layer, respectively. $\mathcal{F}$ represents the residual function. Take the structure in Fig. 1 for instance, $\mathcal{F} = \mathbf{w}_2\varphi(\mathbf{w}_1^T\mathbf{x})$ where φ is the nonlinear activation function, and the biases are not shown for simplicity. The operation $\mathcal{F}(\mathbf{x}, \{\mathbf{w}_i\})+\mathbf{x}$ is performed by a shortcut connection and element-wise addition. It should be noted that the dimensions of $\mathcal{F}$ and $\mathbf{x}$ are supposed to be the same. When they are not matched, linear projection of $\mathbf{x}$ can be used for the addition [37].

2.3. Softmax classifier

A softmax classifier is usually implemented on the top layer in a deep neural network for classification [40]. The training samples are denoted as $\mathbf{x}^{(i)}$ and the corresponding label set is $y^{(i)}$ where $i = 1, 2, \ldots, N_{train}$ and N_{train} is the number of the training samples. $\mathbf{x}^{(i)} \in R^{N\times 1}$ and $y^{(i)} \in \{1, 2, \ldots, K\}$ where K denotes the number of the labeled categories. With respect to an input sample $\mathbf{x}^{(i)}$, the softmax classifier is able to estimate the probability $p(y^{(i)} = j \mid \mathbf{x}^{(i)})$ for the label j ($j = 1, 2, \ldots, K$). The estimated probabilities of $\mathbf{x}^{(i)}$ belonging to each label can be obtained based on the hypothesis function,

$$J_\theta(\mathbf{x}^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 \mid \mathbf{x}^{(i)}; \theta) \\ p(y^{(i)} = 2 \mid \mathbf{x}^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = K \mid \mathbf{x}^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{k=1}^{K} e^{\theta_k^T \mathbf{x}^{(i)}}} \begin{bmatrix} e^{\theta_1^T \mathbf{x}^{(i)}} \\ e^{\theta_2^T \mathbf{x}^{(i)}} \\ \vdots \\ e^{\theta_K^T \mathbf{x}^{(i)}} \end{bmatrix}, \tag{7}$$

where $\theta = [\theta_1, \theta_2, \ldots, \theta_K]^T$ denotes the model parameters.

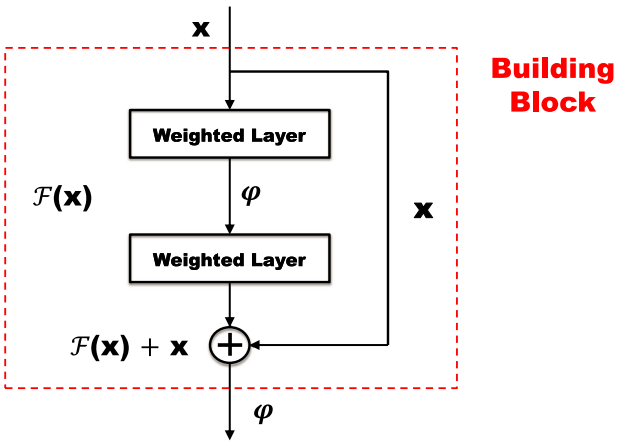

Fig. 1. One building block in residual learning.

2.4. Network architecture

A deep residual learning architecture is proposed in this study, which is shown in Fig. 2.

The raw collected data are directly used as the model inputs without pre-processing, that indicates little prior expertise on fault diagnosis and signal processing is required. A 1-dimensional convolutional layer is first designed with F_N local filter kernels of F_L length window size. The raw input data is processed for feature extraction initially. Next, multiple residual building blocks are stacked to learn high-level representations. By default, two building blocks are employed. In each block, two convolutional layers are adopted for the residual learning. For each convolutional layer in the proposed network, zeros-padding operation is implemented to keep the feature map dimension from changing [41]. Pooling layers can be used in the network to reduce the number of parameters and accelerate the training process while keeping the significant feature information. In this study, one max-pooling layer is adopted between the two residual building blocks. Finally, the high-level representations are put into a fully-connected layer and a softmax regression for machine health condition classification.

Batch normalization (BN) is able to accelerate the training process especially for deep network and has achieved good performance in deep learning recently [42]. In this study, BN is used after each convolutional layer and before activation. In addition, the rectified linear units (ReLU) activation functions are generally used in the network [43].

Corresponding with the softmax classifier, the cross-entropy function is used as the loss function of the network [44], which is defined as,

$$L = -\frac{1}{N_s}\left[\sum_{i=1}^{N_s}\sum_{k=1}^{K} 1\{y_i = k\} \log \frac{e^{\theta_k^T \mathbf{x}^{(i)}}}{\sum_{j=1}^{K} e^{\theta_j^T \mathbf{x}^{(i)}}}\right], \tag{8}$$

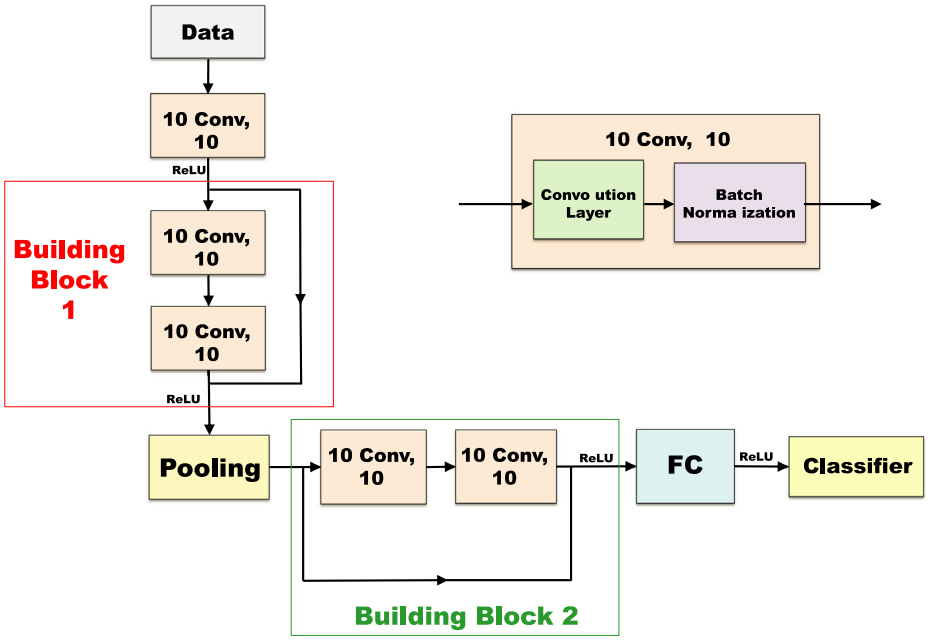

Fig. 2. Proposed deep learning architecture for machinery fault diagnosis.

where N_s denotes the number of the samples. L is supposed to be minimized in order to reduce the empirical risk of the supervised training data.

2.5. Flow chart

The flow chart of the proposed fault diagnosis method is presented in Fig. 3. First, the raw machinery vibration signals are collected by sensors, and the training and testing samples are prepared. The raw vibration data are used directly as the model input. No hand-crafted signal processing feature is needed, such as skewness, kurtosis *etc*. Therefore, little prior expertise on fault diagnosis is required in the proposed method, that facilitates the industrial applications.

Next, based on the concerned fault diagnosis problem and the dataset information, the network configuration is determined including the number of neurons in the hidden layers, number of residual building blocks, convolutional filter number and length *etc*. To start model learning, the training data are delivered to the proposed network. Features of the vibration signals are extracted through the convolutional layers and the residual building blocks. At the end of the network, the softmax module is applied to classify the rolling bearing health conditions using the learned high-level features from the deep network. The back-propagation (BP) algorithm [45] is applied for the updates the weights in the network, and the stochastic gradient descent optimization method is adopted for training [46]. The testing samples will be fed to the proposed network when the training process is finished, and the testing results can be obtained.

In this study, the mini-batch stochastic gradient descent optimization is used in the training process. First, in each training epoch, the training samples are randomly divided into multiple mini-batches with each batch containing 128 samples, and fed into the proposed network. Next, the network information, i.e. the weights and biases in each layer, are optimized based on the mean loss function in each mini-batch.

Two residual blocks are generally adopted, and all the convolutional layers share the same configuration in this paper. The fully-connected layer has 100 neurons for classification. After training for 100 epochs, the loss of the proposed network basically converges. The default parameters of the proposed method are presented in Table 1, which are determined from validations in the case studies presented in the following section.

Table 1
Default parameters of the proposed method and the experimental setting.

Parameter	Value	Parameter	Value
Batch size	128	N_{train}	100
Epoch number	100	N_{test}	100
F_N	10	N_{input}	500
F_L	10		

3. Experimental study

3.1. Experimental setup

The rolling bearing dataset used in this study is provided by the Bearing Data Center of Case Western Reserve University [47]. The dataset contains multivariate vibration signals generated by a bearing test-rig, as presented in Fig. 4.

The main components of the experimental apparatus are a 2-horsepower (hp) motor, a torque transducer/encoder and a dynamometer. The bearing data are measured by acceleration transducers on four conditions (load 0, 1, 2 and 3 hp) and the sampling rate is 12 kHz. The rotating speed changes between 1730 and 1797 rpm based on the applied load.

The vibration signals used in this study were collected from the drive end of the motor in the test rig on four bearing health conditions: (1) healthy condition (H), (2) outer race fault (OF), (3) inner race fault (IF) and (4) ball fault (BF). All the three kinds of faults are generated by electro-discharge machining with diameters of 7 mils, 14 mils and 21 mils, respectively. Therefore, the dataset includes 10 bearing health conditions under the four loads, where the same health condition under different loads is considered as 1 class.

In this study, N_{train} and N_{test} samples for each health condition under one load are assumed to be prepared for training and testing respectively, and each sample is a vibration signal sequence of bearings containing N_{input} data points. The detailed descriptions

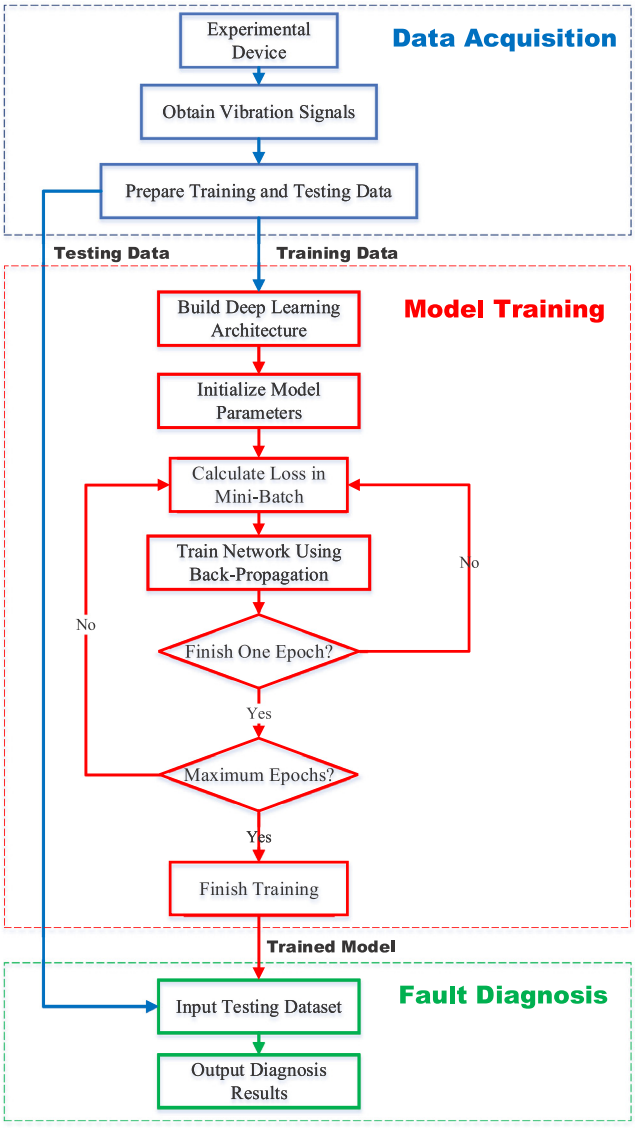

Fig. 3. Flow chart of the proposed method for machinery fault diagnosis.

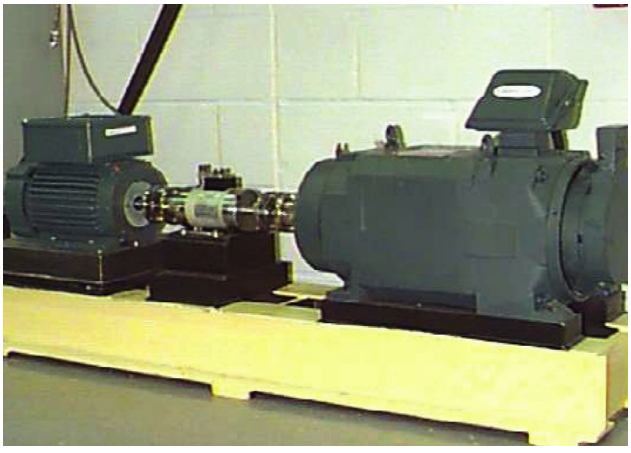

Fig. 4. The bearing test rig used in the experiments.

of the dataset used in this paper are presented in Table 2. For the convenience of classification, the 10 health conditions with different fault location and fault size are artificially set as class label 1 to 10, respectively.

3.2. Compared approaches

In order to validate the proposed method, the testing performance is compared with those of other popular approaches, including basic neural network, deep neural network, basic convolutional neural network, deep convolutional neural network and stacked auto-encoder.

(1) NN
The basic *neural network* (NN), which is also known as multilayer perceptron (MLP), is used for comparison, which has 1 hidden layer of 1000 neurons. Dropout rate of 0.5 and a softmax classifier are applied. The input and output layers are the same with the proposed network for all the compared methods.

(2) DNN
The *deep neural network* (DNN) has 3 hidden layers. The numbers of neurons in the hidden layers are 1000, 500 and 100, respectively. Dropout is also employed.

(3) SAE
The *stacked auto-encoder network* (SAE) is a popular deep learning method and has been receiving increasing attention recently due to its high efficiency in feature extractions [29]. In this study, 3 auto-encoders are stacked in the network and trained using the greedy layer-wise unsupervised learning. After initialization of the auto-encoder layers, a fully-connected layer and softmax classifier are attached on the top for supervised learning, which have the similar configuration with that of the proposed network. The numbers of neurons in the hidden layers are 1000, 500 and 100, respectively.

(4) CNN
There is one 1D convolutional layer and pooling layer in the *convolutional neural network* (CNN). 10 filters are used and the filter length is 10, that are the same with the CNN structure in the proposed method. A fully-connected layer and a softmax classifier are attached on the top.

(5) DCNN
The *deep convolutional neural network* (DCNN) consists of 5 stacked 1D CNN structures, a max-pooling layer, a fully-connected layer and a softmax classifier, which has the same depth with the proposed network. The 5 convolutional layers share the same configuration with the CNN structure presented above. Batch normalization is also used after each convolutional layer. It is noted that the proposed network will become the DCNN if the direct information propagation channel in each residual learning block is removed. Therefore, DCNN is implemented as comparison to show the improvements of the residual learning architecture.

3.3. Experimental results and performance analysis

In this section, the testing performance of the proposed fault diagnosis method on the rolling bearing dataset is presented. The effects of different parameter selections on the results are investigated, including the filter number, the filter length, the receptive input size and so forth. Five other popular methods described in Section 3.2 are implemented for comparisons. The experimental results are further compared with the related state-of-the-art performance on the same dataset to show the effectiveness and superiority of the proposed approach. In this paper, the reported experimental results are averaged by 30 trials to reduce the effect of randomness, and the mean values and standard deviations are

Table 2
The rolling bearing dataset information.

Class label	Fault location	Fault size (mil)	Load (hp)	Training sample	Test sample	Sample length
1	N/A (H)	0	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}
2	IF	7	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}
3	IF	14	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}
4	IF	21	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}
5	BF	7	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}
6	BF	14	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}
7	BF	21	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}
8	OF	7	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}
9	OF	14	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}
10	OF	21	0, 1, 2, 3	$4 \times N_{train}$	$4 \times N_{test}$	N_{input}

provided. All the experiments are carried out on a PC with Intel Core i7 CPU, 8-GB RAM and GEFORCE GTX 950M GPU.

The comparison results of the classification performance with different methods are presented in Table 3. The effectiveness of the proposed method is examined with the default experimental setting in this case. It can be observed that the proposed deep learning method achieves the best performance in all scenarios. The average testing accuracy is as high as 99.98%. In fact, Only 1 out of 4000 testing samples are mis-classified with the proposed method, which belongs to Class 7. The DCNN method achieves the second best results with 99.93% testing accuracy, where 3 samples are not correctly classified. In addition, the CNN method performs better than the rest, i.e. SAE, DNN and NN. Only 84.75% testing accuracy can be obtained with the shallow basic neural network structure.

3.3.1. Effect of training samples

In this section, the effect of training samples on the classification performance is investigated. Fig. 5 shows the fault diagnosis accuracies using different networks with various number of training samples N_{train}. The number of testing samples is fixed at $N_{test} = 100$ for comparison, that indicates $4 \times 10 \times N_{test} = 4000$ samples are randomly selected for testing in all the scenarios. The testing data samples for each compared method are the same with identical N_{train}.

It can be observed that the testing accuracy increases and the standard deviation decreases with the rise of the number of training samples for all the methods. More training data lead to higher diagnosis accuracy in general. With identical training samples, the proposed method always achieves the highest testing accuracy, and the second best results are obtained by the DCNN method. That shows the effectiveness of the deep learning architecture with convolutional neural networks for machinery fault diagnosis problem, and the performance improvements by the residual learning are also confirmed.

It can be explained that with respect to the deep learning architecture, while the training problem of vanishing and exploding gradients can be addressed by data normalization [42], a degradation problem inevitably occurs. As the network goes deeper, accuracy becomes saturated and degrades quickly. Adding more layers into an appropriate structure results in higher training error [48]. The degradation problem has been effectively addressed by the identity skip-connection structures in the literature [36,37]. Therefore, better classification results can be obtained with the proposed method than with the DCNN.

The proposed method is able to perform well even with limited number of training samples. The testing accuracy with the proposed method using only 400 samples ($N_{train} = 10$) is over 80%, while for the other methods, the accuracies are usually lower than 50%. More than 99% of the testing samples can be accurately classified if the training samples are more than 1000. The fault diagnosis classification results of the proposed method are promising considering 10 categories of bearing conditions are focused on and 4000 samples are tested.

Compared with other approaches, the performance improvement achieved by the proposed method with limited training samples is more significant than that with more training data. The shallow networks are not able to diagnose bearing health condition well with a small number of training samples. Therefore, the proposed method is robust on the quantity of training data. For instance, the CNN method performs well with a large number of training data, but its performance deteriorates fast when N_{train} becomes smaller. Similar display patterns are also observed with the NN, DNN and SAE methods.

3.3.2. Analysis of residual learning

In this section, the reason of the improvements in network training by the residual learning algorithm is investigated. With respect to the residual learning scheme shown in Fig. 1, let $\mathbf{x}_i^{in}$ denote the input of the ith residual block, and $\mathbf{x}_i^{out}$ represents the output of the block. The residual learning module performs the operation,

$$\mathbf{x}_i^{out} = \mathbf{x}_i^{in} + \mathcal{F}(\mathbf{x}_i^{in}, \mathbf{W}_i), \tag{9}$$

where $\mathbf{W}_i$ denotes the corresponding weights within the block. When multiple blocks are stacked together, the computation between the kth and ith blocks becomes,

$$\mathbf{x}_i^{out} = \mathbf{x}_k^{in} + \sum_{j=k}^{i} \mathcal{F}(\mathbf{x}_j^{in}, \mathbf{W}_j), \tag{10}$$

where the input of one block is the output of the lower block, i.e. $\mathbf{x}_i^{in} = \mathbf{x}_{i-1}^{out}$. In this way, the backward propagation can be much improved, and the gradients in the network optimization are,

$$\frac{\partial L}{\partial \mathbf{x}_k^{in}} = \frac{\partial L}{\partial \mathbf{x}_i^{out}} \frac{\partial \mathbf{x}_i^{out}}{\partial \mathbf{x}_k^{in}} = \frac{\partial L}{\partial \mathbf{x}_i^{out}} \left(1 + \frac{\partial}{\partial \mathbf{x}_k^{in}} \sum_{j=k}^{i} \mathcal{F}(\mathbf{x}_j^{in}, \mathbf{W}_j)\right). \tag{11}$$

Eq. (11) suggests that when the network parameters are updated, the gradients consist of two components. The item $\frac{\partial L}{\partial \mathbf{x}_i^{out}}$ ensures the information flow throughout the network regardless of the weights of the layers, that effectively addresses the problem of gradient vanishing and exploding. Specifically, all the lower layers are able to receive this gradient without information loss, which enhances fully training of deep neural network. On the other hand, the weights of the layers are considered by the other item $\frac{\partial}{\partial \mathbf{x}_k^{in}} \sum_{j=k}^{i} \mathcal{F}(\mathbf{x}_j^{in}, \mathbf{W}_j)$. Therefore, the deep neural network significantly benefits from the residual learning algorithm, that facilitates the machinery vibration signal processing especially with large sequential length.

3.3.3. Effect of filter number and length

The filter number and length are two main parameters in convolutional network. In general, more filters and longer window size lead to better network performance with higher computational burden [32]. The long distance information in the sequential vibration signal cannot be captured with small window size in

Table 3
Comprehensive comparison of classification accuracy of different fault diagnosis method.

Classifier	Class label										Average (%)
	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)	8 (%)	9 (%)	10 (%)	
NN	100	80.75	65.50	100	82.25	87.25	79.75	76.50	99.50	76	84.75
DNN	100	97.50	92	100	94.50	94.50	95.50	92.75	99.50	83.25	94.95
SAE	100	97.50	90.75	100	95.50	93.75	97.25	95.75	99.25	91.25	96.13
CNN	100	97.75	90.50	100	99.50	94.25	98.25	99.75	100	96.25	97.63
DCNN	100	100	99.50	100	100	100	100	100	100	99.75	99.93
Proposed	100	100	100	100	100	100	99.75	100	100	100	99.98

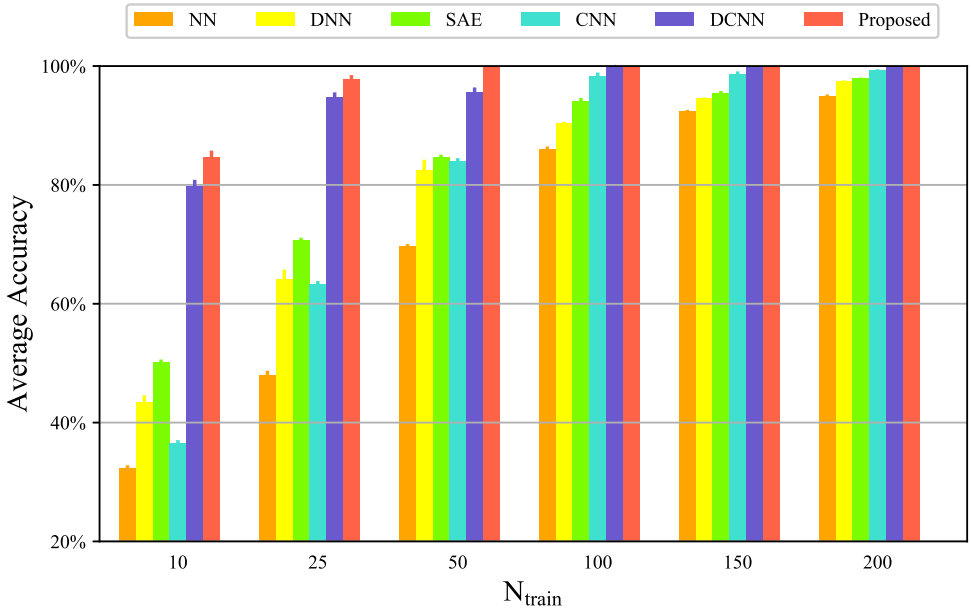

Fig. 5. The effect of the training data number on the network classification accuracy for various fault diagnosis methods.

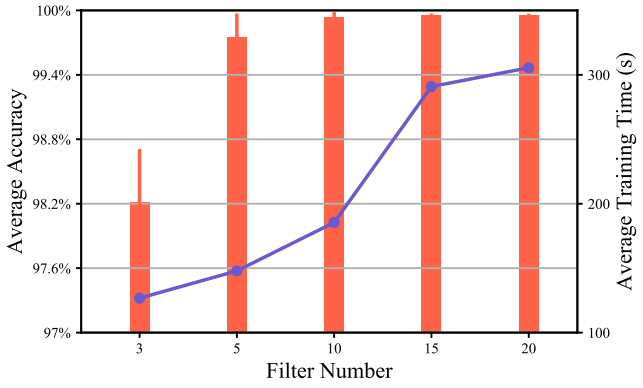

Fig. 6. The effect of the filter number in the convolution layers on the network classification accuracy. Bar: Average accuracy. Dot: Average training time.

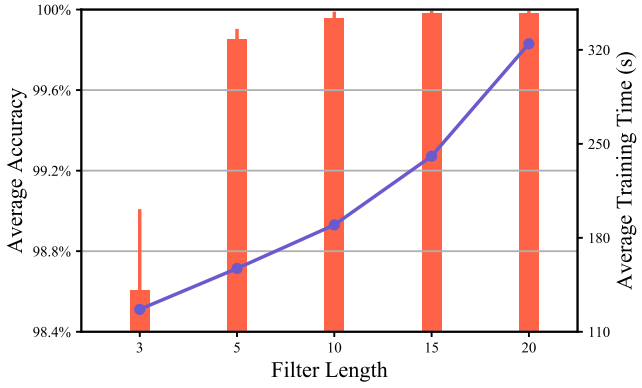

Fig. 7. The effect of the filter length in the convolution layers on the network classification accuracy. Bar: Average accuracy. Dot: Average training time.

some cases [32]. Therefore, the influence of filter number F_N and filter length F_L on the network performance is investigated in this section, and the experimental results are presented in Figs. 6 and 7.

The testing accuracies and training time of the proposed method are exhibited with different F_N and F_L. It can be observed in Fig. 6 that generally, more filters in each convolutional layer lead to higher testing accuracy. More features of the raw input data can be extracted by the network with larger F_N, and better feature representation can be obtained in this way. However, it is noted that larger computational cost of the proposed method is required with the increase in the number of filters in each convolutional layer, while larger F_N leads to higher fault diagnosis accuracy. Therefore, tradeoff between computational burden and testing classification accuracy has to be made.

The experimental results show that when 10 filters are used in each convolutional layer, high classification accuracy can be obtained with medium training time. No significant improvements in accuracy will be made by larger F_N when F_N exceeds 10. Therefore, 10 filters are used in each convolutional layer in the proposed network by default.

Similar with the effect of the filter number on the network performance, the testing accuracy displays the same pattern with different filter length for each kernel as presented in Fig. 7. Based

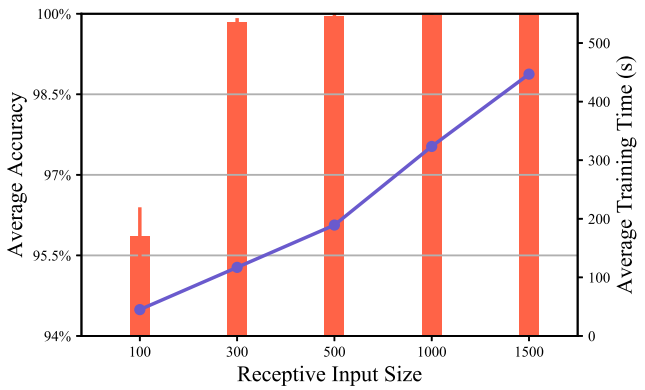

Fig. 8. The effect of the receptive input dimension on the network classification accuracy. Bar: Average accuracy. Dot: Average training time.

on the experimental results, the filter window size of 10 is used as the default value for the proposed network.

3.3.4. Effect of receptive input size

Different from other tasks such as image processing, the dimension of the samples prepared from machinery vibration data can be artificially determined based on the specific demands. Basically, more data points included in the samples contain more information of the machine health state, which generally leads to better prediction. However, larger model will be correspondingly required, as well as heavier computational burden.

Fig. 8 shows the effect of the receptive input dimension N_{input} on the network classification accuracy. Similar with the effect of the filter configurations, better data representation and feature extraction can be obtained with larger receptive input dimension generally. However, the receptive input dimension used in practice should also be determined considering the network computational burden.

It can be observed that small input size as 100 is not able to provide enough information for rolling bearing fault diagnosis, resulting in testing accuracy around 95.5%. While higher classification accuracy can be obtained with larger input dimension, more model parameters are introduced in the network with larger N_{input}, that increases the training computational burden. The experiments indicate that sufficient machinery signal information can be covered by the input data dimension of 500, ensuring computing efficiency for bearing fault diagnosis.

3.3.5. Visualization of learned representation

In this section, the effectiveness of the proposed method for fault diagnosis is illustrated qualitatively based on visualization of learned representation. An effective technique "t-SNE" is employed to visualize the high-dimensional data representation by mapping the samples from the original data space into a 2-dimensional space map [49]. The principal component analysis (PCA) is first employed to reduce the feature dimensionality to 50. Afterwards, "t-SNE" is used to transform the 50-dimensional learned representation into a 2-dimensional sample.

Fig. 9 shows the resulting maps of the raw data and the learned representations of DNN, DCNN and the proposed method. It can be seen that the features extracted by the proposed method cluster the best where all the samples in different bearing health conditions are separated well. The good separation of features extracted by the DCNN method is also observed, despite the existence of a small amount of point overlappings. As comparison, the features learned by the DNN method do not cluster well. Significant point overlappings of different classes are observed, that indicates an effective classification cannot be achieved by the DNN method. The feature visualization results are consistent with the conclusions from previous sections, that the fault diagnosis problem can be effectively solved by the deep learning architecture with convolutional neural networks.

Moreover, noticeable data samples of Class 3 and 10 are observed located in the cluster regions of other classes, which indicates the classifications of the two bearing health conditions are difficult. This observation corresponds with the testing results presented in Table 3.

It should be pointed out that the final classification is carried out in a high-dimensional space nonlinearly. Therefore, acceptable point overlappings for different health conditions in visualization agree with the high classification accuracies presented in Table 3.

3.3.6. Testing performance under additional noise

In the real operating scenarios of rotating machinery, environmental noise inevitably exists, and the collected vibration data are thus corrupted. That significantly deteriorates the effectiveness of the data-driven fault diagnosis methods. In this study, additional Gaussian noises are added to the original vibration data to evaluate the testing performance of the proposed method under noisy environment. Specifically, the noisy samples are created with different signal-to-noise ratio (SNR), that is defined as,

$$\mathrm{SNR\,(dB)} = 10\log_{10}(P_{signal}/P_{noise}), \tag{12}$$

where P_{signal} and P_{noise} denote the powers of the original signal and the additional Gaussian noise, respectively. In this study, the training data are also corrupted by noise, and the noisy signal ranging from 0 to 8 dB is adopted to examine the proposed method.

Fig. 10 shows the testing accuracies of the proposed method with different level of additional noise. The effects of variable input sample size are investigated, and the DCNN method is also examined for comparison, since competitive results are obtained by the DCNN approach based on previous experiments. $N_{train} = 100$ is used with $F_L = 10$ and $F_N = 10$.

It can be observed that generally, the additional noise significantly deteriorates the diagnosis performance in different scenarios. Stronger noise basically leads to lower testing accuracies. Specifically, the display patterns of the average testing accuracies by the two methods are similar with each other, and the proposed method outperforms the DCNN method in most cases. High testing accuracies can be still achieved by the proposed method with additional environmental noise.

Furthermore, the variances of the results by the proposed method are remarkably smaller than those by the DCNN approach, that indicates the network training is more stable using the residual learning method in noisy environment.

Moreover, the improvements on the testing results by larger input size of the samples are further confirmed. When more data points are included in the samples, better feature extraction can be implemented using more information by the deep neural network and consequently, better diagnosis results are obtained. While higher input dimension also leads to higher testing accuracy by the DCNN method, the network performance is generally not stable and larger variance is obtained. That mostly results from the training process where the propagation of gradients is of lower efficiency compared with that of the residual learning algorithm.

Therefore, the proposed method is well suited for processing the machinery vibration signal using deep neural network, where the samples can be prepared in different dimensions based on the specific demands.

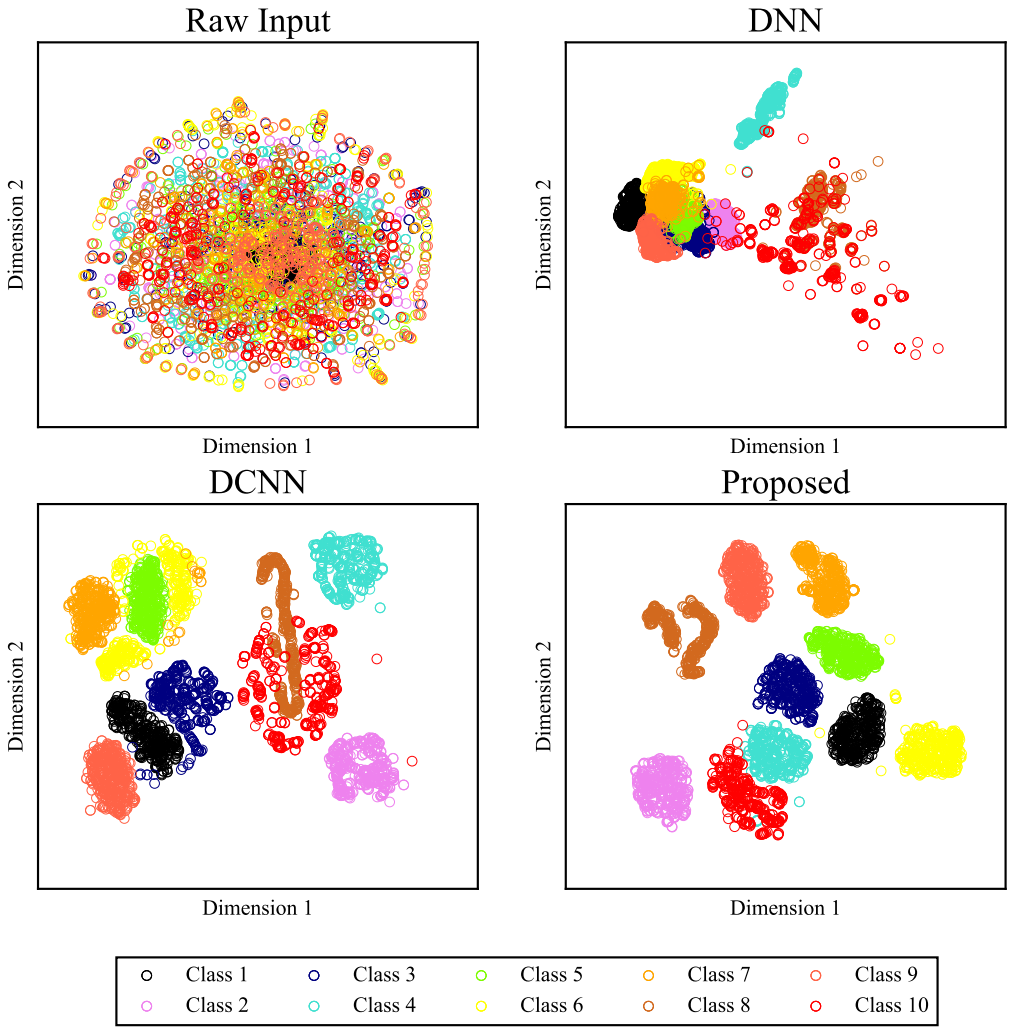

Fig. 9. Feature visualization maps.

3.3.7. Comparisons with related works

The rolling bearing dataset used in this paper is very popular in machinery fault diagnosis researches, and many state-of-the-art classification results have been reported in the past years. In [50–52], 95% and higher testing accuracies were achieved using this dataset. However, only 4 rolling bearing health conditions or fewer were considered. When 10 bearing conditions are classified as studied in this paper, 88.9%, 92.5% and 97.9% testing accuracies were obtained in [53,54] and [55], respectively. In [40], a two-stage machine learning method was proposed based on unsupervised feature learning and sparse filtering. Fairly high diagnosis accuracy of 99.66% were achieved with this dataset.

Using the default network configuration, the proposed method in this paper can obtain the testing accuracy of **99**.**98**% as presented in Table 3. It should be noted that, as a deep learning approach, the defect of the proposed method is the comparative longer training time than most shallow networks such as that in [40]. Furthermore, if the optimal network configuration of the proposed method is used regardless of the off-line computational burden and sample overlapping, higher classification accuracy up to **99**.**99**% can also be obtained. For instance, if the receptive input size of 1000 is used with the proposed method, all the 4000 testing samples can be accurately classified in 25 out of 30 experimental trials. For the rest 5 trials, only 1 testing sample is mis-classified respectively, that makes the mean classification accuracy 99.996%. The detailed comparison results are summarized in Table 4.

Table 4
Comparisons of testing diagnosis accuracies of related researches on the same rolling bearing dataset.

Method	Classes	Testing accuracy
[50]	4	95.8%
[53]	10	88.9%
[54]	10	92.5%
[55]	11	97.91%
[40]	10	99.66%
Proposed	10	99.99%

4. Conclusions

In this paper, a novel deep learning method for rotating machinery fault diagnosis is proposed using residual learning algorithm. Experiments on a popular rolling bearing dataset are carried out to verify the effectiveness of the proposed method. Existing popular neural network-based approaches for fault diagnosis, including basic neural network, deep neural network, stacked auto-encoders, convolutional neural network and deep convolutional neural network, are implemented for comparisons. The state-of-the-art results on the same dataset are also presented to show the superiority of the proposed method.

In general, large neural networks can be effectively trained and higher classification accuracy can be achieved with the proposed

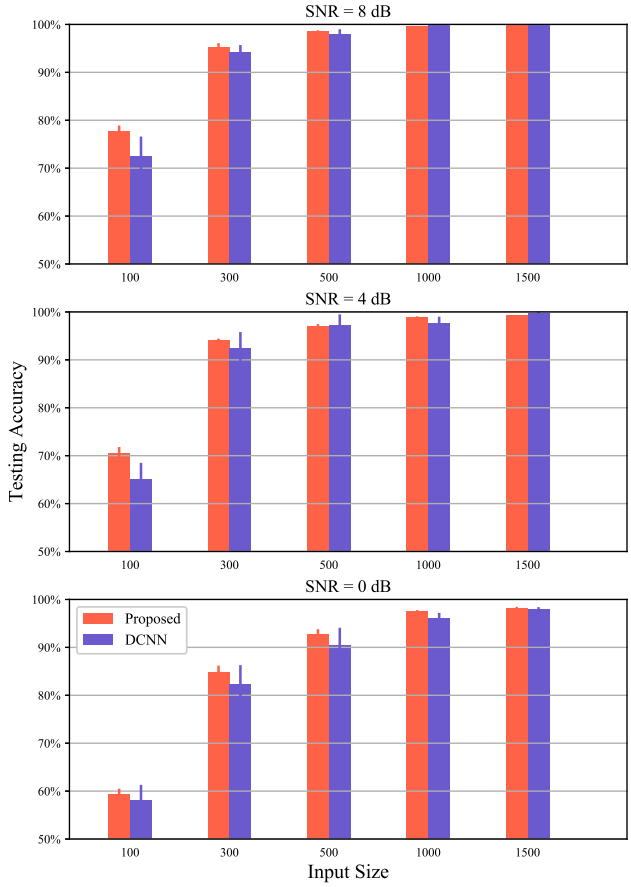

Fig. 10. Testing accuracies under different level of additional noise.

approach compared with other popular methods using similar configurations. With the default experimental setting, the testing classification accuracy is 99.98% over 4000 samples. If the enhanced network configuration is used regardless of the off-line computing load, up to 99.99% testing accuracy can also be obtained. The proposed method is able to perform well even with limited training data. The effects of multiple network parameters are investigated, such as the filter number and length, receptive input size *etc.*

While good fault diagnosis results can be obtained by the proposed method, its disadvantage is the comparatively longer training time than most shallow neural network-based approaches. Deep learning methods commonly suffer from heavy computational load. However, since most of the training tasks can be finished within a few minutes in the off-line training part, the computational burden of the proposed method is acceptable.

It should be noted that since the neural network is very complex with large amounts of parameters, improving its performance is always necessary. In the recent years, evolutionary algorithm-based approaches and multi-objective extremal optimization methods have shown promising results in enhancing network abilities [56–60], which will be considered to be applied in machinery fault diagnosis tasks in the following works. Furthermore, more experiments are supposed to be used to validate the proposed method, especially in more practical industrial scenarios. That will also be investigated in future research.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, China through the grants (N170503012, N170308028), Scientific Research Fund of Liaoning Provincial Education Department, China through the grant (No. L201737), and the National Natural Science Foundation of China through the grant (61871107).

References

[1] Sun H, He Z, Zi Y, Yuan J, Wang X, Chen J, et al. Multiwavelet transform and its applications in mechanical fault diagnosis - A review. Mech Syst Signal Process 2014;43(1–2):1–24.
[2] Tse PW, Peng YH, Yam R. Wavelet analysis and envelope detection for rolling element bearing fault diagnosis - Their effectiveness and flexibilities. J Vib Acoust 2001;123(3):303–10.
[3] Ren Z, Zhou S, E C, Gong M, Li B, Wen B. Crack fault diagnosis of rotor systems using wavelet transforms. Comput Electr Eng 2015;45:33–41.
[4] Chen XH, Cheng G, Shan XL, Hu X, Guo Q, Liu HG. Research of weak fault feature information extraction of planetary gear based on ensemble empirical mode decomposition and adaptive stochastic resonance. Measurement 2015;73:55–67.
[5] He HL, Wang TY, Leng YG, Zhang Y, Li Q. Study on non-linear filter characteristic and engineering application of cascaded bistable stochastic resonance system. Mech Syst Signal Process 2007;21(7):2740–9.
[6] Li Q, Wang T, Leng Y, Wang W, Wang G. Engineering signal processing based on adaptive step-changed stochastic resonance. Mech Syst Signal Process 2007;21(5):2267–79.
[7] Yan X, Jia M, Zhang W, Zhu L. Fault diagnosis of rolling element bearing using a new optimal scale morphology analysis method. ISA Trans 2018;73:165–80.
[8] Zhou P, Lu S, Liu F, Liu Y, Li G, Zhao J. Novel synthetic index-based adaptive stochastic resonance method and its application in bearing fault diagnosis. J Sound Vib 2017;391:194–210.
[9] He G, Ding K, Lin H. Fault feature extraction of rolling element bearings using sparse representation. J Sound Vib 2016;366:514–27.
[10] Žvokelj M, Zupan S, Prebil I. EEMD-based multiscale ICA method for slewing bearing fault detection and diagnosis. J Sound Vib 2016;370:394–423.
[11] Lu S, Wang X, He Q, Liu F, Liu Y. Fault diagnosis of motor bearing with speed fluctuation via angular resampling of transient sound signals. J Sound Vib 2016;385:16–32.
[12] Bin GF, Gao JJ, Li XJ, Dhillon BS. Early fault diagnosis of rotating machinery based on wavelet packets - Empirical mode decomposition feature extraction and neural network. Mech Syst Signal Process 2012;27:696–711.
[13] Samanta B, Nataraj C. Use of particle swarm optimization for machinery fault detection. Eng Appl Artif Intell 2009;22(2):308–16.
[14] Tran VT, Yang BS, Gu F, Ball A. Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis. Mech Syst Signal Process 2013;38(2):601–14.
[15] Yang BS, Di X, Han T. Random forests classifier for machine fault diagnosis. J Mech Sci Technol 2008;22(9):1716–25.
[16] Zhang X, Chen W, Wang B, Chen X. Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization. Neurocomputing 2015;167:260–79.
[17] Jegadeeshwaran R, Sugumaran V. Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines. Mech Syst Signal Process 2015;52–53:436–46.
[18] Li Z, Fang H, Huang M. Diversified learning for continuous hidden Markov models with application to fault diagnosis. Expert Syst Appl 2015;42(23):9165–73.
[19] Youssef A, Delpha C, Diallo D. An optimal fault detection threshold for early detection using Kullback-Leibler divergence for unknown distribution data. Signal Process 2016;120:266–79.
[20] Samanta B. Artificial neural networks and genetic algorithms for gear fault detection. Mech Syst Signal Process 2004;18(5):1273–82.
[21] Jia F, Lei Y, Lin J, Zhou X, Lu N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Signal Process 2016;72–73:303–15.
[22] Zhang W, Li C, Peng G, Chen Y, Zhang Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. Mech Syst Signal Process 2018;100:439–53.
[23] Sun W, Shao S, Zhao R, Yan R, Zhang X, Chen X. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. Measurement 2016;89:171–8.
[24] Li X, Zhang W, Ding Q. A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning. Neurocomputing 2018;310:77–95.
[25] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504.
[26] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proceedings of 26th annual conference on neural information processing systems, vol. 2. 2012, p. 1097–105.

[27] Hinton GE, Deng L, Yu D, Dahl G, Mohamed AR, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Process Mag 2012;29(6):82–97.
[28] Liu H, Zhou J, Zheng Y, Jiang W, Zhang Y. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. ISA Trans 2018;77:167–78.
[29] Lu C, Wang ZY, Qin WL, Ma J. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. Signal Process 2017;130:377–88.
[30] Mao WT, He JL, Li Y, Yan YJ. Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study. Proc Inst Mech Eng C 2016. 0954406216675896.
[31] Guo X, Chen L, Shen C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement 2016;93:490–502.
[32] Sun W, Zhao R, Yan R, Shao S, Chen X. Convolutional discriminative feature learning for induction motor fault diagnosis. IEEE Trans Ind Inf 2017;13(3):1350–9.
[33] 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.
[34] Ince T, Kiranyaz S, Eren L, Askar M, Gabbouj M. Real-time motor fault detection by 1-D convolutional neural networks. IEEE Trans Ind Electron 2016;63(11):7067–75.
[35] Li X, Zhang W, Ding Q. Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks. IEEE Trans Ind Electron 2018. 1–1.
[36] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition, arXiv preprint arXiv:1512.03385, 2015.
[37] He KM, Zhang XY, Ren SQ, Sun J. Identity Mappings in Deep Residual Networks. In: Proceedings of european conference on computer vision. Cham, Switzerland; 2016, p. 630–45.
[38] Li X, Ding Q, Sun J-Q. Remaining useful life estimation in prognostics using deep convolution neural networks. Reliab Eng Syst Safety 2018;172:1–11.
[39] Li X, Zhang W, Ding Q. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliab Eng Syst Safety 2019;182:208–18.
[40] Lei Y, Jia F, Lin J, Xing S, Ding SX. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. IEEE Trans Ind Electron 2016;63(5):3137–47.
[41] Liu B, Liu J, Bai X, Lu H. Regularized Hierarchical Feature Learning with Non-negative Sparsity and Selectivity for Image Classification. In: Proceedings of 22nd international conference on pattern recognition. 2014, p. 4293–8.
[42] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of 32nd international conference on machine learning, vol. 1. Lile, France; 2015, p. 448–56.
[43] Dahl GE, Sainath TN, Hinton GE. Improving deep neural networks for LVCSR using rectified linear units and dropout. In: Proceedings of IEEE international conference on acoustics, speech and signal processing. 2013, p. 8609–13.
[44] Hinton G, Vinyals O, Dean J. Distilling the Knowledge in a Neural Network, arXiv preprint arXiv:1503.02531, 2015.
[45] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323(6088):533–6.
[46] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
[47] Smith WA, Randall RB. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. Mech Syst Signal Process 2015;64–65:100–31.
[48] Srivastava RK, Greff K, Schmidhuber J. Highway Networks, 2015, arXiv preprint arXiv:1505.00387.
[49] Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579–625.
[50] Li W, Zhang S, He G. Semisupervised distance-preserving self-organizing map for machine-defect detection and classification. IEEE Trans Instrum Meas 2013;62(5):869–79.
[51] van Wyk BJ, van Wyk MA, Qi G. Difference Histograms: A new tool for time series analysis applied to bearing fault diagnosis. Pattern Recognit Lett 2009;30(6):595–9.
[52] Muruganatham B, Sanjith MA, Krishnakumar B, Satya Murty SA V. Roller element bearing fault diagnosis using singular spectrum analysis. Mech Syst Signal Process 2013;35(1–2):150–66.
[53] Du W, Tao J, Li Y, Liu C. Wavelet leaders multifractal features based fault diagnosis of rotating mechanism. Mech Syst Signal Process 2014;43(1–2):57–75.
[54] Jin X, Zhao M, Chow TWS, Pecht M. Motor bearing fault diagnosis using trace ratio linear discriminant analysis. IEEE Trans Ind Electron 2014;61(5):2441–51.
[55] Zhang X, Liang Y, Zhou J, Zang Y. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. Measurement 2015;69:164–79.
[56] Zeng G-Q, Xie X-Q, Chen M-R, Weng J. Adaptive population extremal optimization-based PID neural network for multivariable nonlinear control systems. Swarm Evol Comput 2018.
[57] Wang S, Zhang N, Wu L, Wang Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. Renew Energy 2016;94:629–36.
[58] Lu K, Zhou W, Zeng G, Du W. Design of PID controller based on a self-adaptive state-space predictive functional control using extremal optimization method. J Franklin Inst B 2018;355(5):2197–220.
[59] Lu K, Zhou W, Zeng G, Zheng Y. Constrained population extremal optimization-based robust load frequency control of multi-area interconnected power system. Int J Electr Power Energy Syst 2019;105:249–71.
[60] Zeng G-Q, Chen J, Dai Y-X, Li L-M, Zheng C-W, Chen M-R. Design of fractional order PID controller for automatic regulator voltage system based on multi-objective extremal optimization. Neurocomputing 2015;160:173–84.