

Cascade Convolutional Neural Network With Progressive Optimization for Motor Fault Diagnosis Under Nonstationary Conditions

Fei Wang, Ruonan Liu , *Member, IEEE*, Qinghua Hu , *Senior Member, IEEE*, and Xuefeng Chen, *Member, IEEE*

Abstract—Recently, convolutional neural networks (CNNs) have been successfully used for motor fault diagnosis because of its powerful feature extraction ability. However, there are still some barriers of traditional CNNs. Due to the fact of the hierarchical structure, feature resolution of CNNs will be reduced with layer growth, which can lead to the information loss. In addition, the fixed kernel size makes traditional CNNs not suitable for fault diagnosis of motors, which are widely used in nonstationary conditions. Therefore, starting from the physical characteristics of nonstationary vibration signals, a cascade CNN (C-CNN) with progressive optimization is proposed in this article. First, a cascade structure is built to avoid the information loss caused by consecutive convolution striding or pooling. Then, dilated convolution operations are implemented, which can extract the feature maps from different scales and extend the applications of CNN to nonstationary conditions. Furthermore, taking the advantage of the cascade structure, a progressive optimization algorithm is proposed for divide-and-conquer parameters optimization, which enables the C-CNN to converge to a more optimum state and improve the diagnosis performance. The proposed method is verified by two motor fault diagnosis experiments, which are conducted under constant speed and variable speed, respectively. The results show that the proposed method can achieve better performance when rotating speed is either constant or changing than existing methods.

Index Terms—Convolutional neural network (CNN), deep learning, motor fault diagnosis, nonstationary conditions, progressive optimization.

Manuscript received March 2, 2020; revised May 26, 2020; accepted June 13, 2020. Date of publication June 18, 2020; date of current version January 4, 2021. This work was supported by the National Science and Technology Major Project (2017–0007-0008). Paper no. TII-20-1092. (Corresponding author: Ruonan Liu.)

Fei Wang is with the School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: wfei.cs@gmail.com).

Ruonan Liu and Qinghua Hu are with the Tianjin Key Laboratory of Machine Learning College of Intelligence and Computing Tianjin University, Tianjin 300350, China (e-mail: liuruonan04@163.com; huqinghua@tju.edu.cn).

Xuefeng Chen is with the State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: chenxf@xjtu.edu.cn).

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TII.2020.3003353

I. INTRODUCTION

AS ONE of the most important parts in modern industrial systems, electric motors have begun to be more intensively utilized in industrial applications due to their ruggedness, low price, and less maintenance cost [1]–[3]. Therefore, to ensure the reliable operation of industrial systems and decrease the economic loss, condition monitoring and fault diagnosis of motors play an increasing significant role in normal operation and maintenance of motors [4].

In the literature, lots of methods have been proposed for motor fault diagnosis under constant rotating speed. Usually traditional fault diagnosis methods are built based on two steps: feature extraction and fault recognition. Feature extraction is an important step in motor fault diagnosis because raw signals can be represented by low-dimensional vectors for easier comparison and analysis in this step. Many signal processing techniques have been proposed for feature extraction, including statistical analysis, short-term Fourier transform, wavelet packet transform [5], empirical model decomposition [6], and sparse representation methods [7]. In the second step, the fault can be diagnosed by the following pattern recognition methods [8], such as support vector machine [9] and artificial neural network (ANN) [10]. Effective as traditional diagnosis methods are, they still suffer from several weaknesses, including the requirement of enough prior knowledge for feature extractor design and the limitation when facing with so many changeable operational conditions. Such problems promote the applications of deep learning methods in the field.

Deep learning methods aim at learning feature hierarchies from higher levels of the hierarchy formed by the composition of lower level features, which show the potential to solve the problems of traditional diagnosis methods [11]. For now, autoencoders [12], [13], convolutional neural networks (CNNs) [14], [15], recurrent neural networks [16], [17], and deep belief network [18], [19] are the most commonly used networks that can be implemented by deep structure. Since Jia *et al.* [20] applied deep neural network to intelligent diagnosis of rotating machinery in 2016, lots of deep learning based motor fault diagnosis methods have been proposed because they provide an effective way to extract features from raw signals automatically. CNNs have recently become one of the most commonly used deep architecture for tasks, such as object recognition in large image achieves, as

well as fault diagnosis, and have achieved the state-of-the-art performances with a significant gap. 1-D CNN has been used for real-time motor fault detection by Ince *et al.* [21]. Based on LeNet-5, Wen *et al.* [22] proposed a new CNN framework, which has been proved to be effective for motor bearing fault diagnosis. Liu *et al.* [23] proposed a multiscale kernel algorithm that can be applied in the CNN architecture to capture the patterns of vibration signals in the changing operational conditions.

However, although it is convenient to diagnose motor failures by such end-to-end deep learning methods, there are still some limitations of CNN when dealing with industrial signals. First, as we all know, CNN is a hierarchical structure. Therefore, with the layer becomes higher, feature resolution will be reduced due to the consecutive convolution striding or pooling operations, which may further lead to little influence on image recognition tasks because the decrease of resolution in a certain range will not change the extracted features of an image a lot. However, industrial signals are different from images, the vibration response signals of faults may vanish quickly. As a result, the decrease of feature resolution here can make a difference on the final diagnostic performance if such vibration response signals are lost during the convolutional or pooling operations. Second, in a CNN, once the convolutional kernel size is defined, it will be used across the whole perspective field, which can reduce the risk of overfitting. Therefore, the kernel size should also be defined carefully to ensure the diagnostic performance. Additionally, the induction motors are usually working under variable conditions. The fixed kernel size makes the CNN extract feature maps in only one scale, which limits the application of CNN from fault diagnosis tasks under nonstationary conditions. Third, gradient descent algorithm and its variants are widely used for deep network optimization, which adjust all the parameters together from a global view. However, since the number of parameters in networks is large, such an optimization manner can lead to an unsatisfactory locally optimization solution.

Therefore, from the view of these two problems, a cascade CNN (C-CNN) with progressive optimization is proposed in this article. On the one hand, the proposed C-CNN implements a cascade structure instead of a traditional single hierarchical structure to capture the semantic information at all levels. Specifically, the C-CNN is constructed based on a traditional bottom-up pyramidal CNN pathway first, then followed by a top-down pathway with lateral connections. In this way, the specific features of high resolution and the general features of low resolution can be combined without the sacrifice of speed. On the other hand, the convolutional kernels with various strides are also applied, which can be regarded as a dilated convolution structure applied in each convolutional layer. Thus, it controls the perspective field in which the network analyzes. Therefore, multiscale information can be extracted from different sizes of perspective fields by such a dilated convolutional structure with various rates. Finally, taking the advantage of the cascade structure, a progressive algorithm is proposed for C-CNN optimization, which decomposes the parameters into several levels and adjusts the parameters from bottom to top in a progressive manner. Thus, the C-CNN could converge to a better state,

increasing the performance of the motor fault diagnose. Since the C-CNN is proposed to improve the natural property of CNN, it will retain the aforementioned advantages even when the backbone CNN is constructed with few layers.

The rest of this article is organized as follows. Section II describes the details of the proposed method. In Section III, a motor experiment under constant rotating speed is conducted. In Section IV, another motor experiment under nonstationary operational conditions are conducted in the same testbed to illustrate the effectiveness of the proposed method for nonstationary signal analysis. Finally, Section V concludes this article.

II. PROPOSED METHOD

Since the C-CNN is constructed based on a CNN structure, a brief introduction of CNN theory background is demonstrated first. Then, the cascade structure and dilated convolution structure are illustrated, respectively. Next, the overall framework of the proposed C-CNN is described. Finally, the progressive optimization algorithm is proposed to adjust the parameters of the C-CNN.

A. Convolutional Neural Network

CNN is a kind of variant neural network, which has been successfully applied in many fields because of its excellent ability of feature extraction. Essentially, a CNN is constructed by convolutional layer, pooling layer, and fully connected network. The function of a convolutional layer is to extract the feature maps from inputs by sliding the filter kernels and calculating the convolution of input local regions and the kernels [24], which can be mathematically described as

$$y^{l(i,y)} = \mathbf{K}_i^l \cdot \mathbf{x}^{l(\mathbf{r}^j)} = \sum_{j'=0}^W \mathbf{K}_i^l(j') \mathbf{x}^{l(j+j')} \quad (1)$$

where \mathbf{K}_i^l represents the weights of the i th filter kernel in layer l . $\mathbf{x}^{l(\mathbf{r}^j)}$ represents the j th local region in layer l . \cdot denotes the dot product. W is the width of the kernel. $\mathbf{K}_i^l(j')$ denotes the j' th weight of the kernel.

Pooling is also an important part of CNN. It is a form of subsampling that can reduce the feature dimensions, as well as enable the representation that become invariant to small translation of the inputs. Max pooling is one of the most commonly used subsampling operation, which can be described as

$$p^{l(i,j)} = \max_{(j-1)W+1 \leq t \leq jW} \{a^{l(i,t)}\} \quad (2)$$

where $p^{l(i,j)}$ represents the value of the neuron in layer l of the pooling operation. $a^{l(i,t)}$ denotes the value of t th neuron in the i th frame of layer l , and $t \in [(j-1)W+1, jW]$. W represents the width of the pooling region.

After several alternate convolutional and pooling layers, the fully connected layers are followed to compute the class scores. Usually, the structure of fully connected layers is kept the same with traditional ANN.

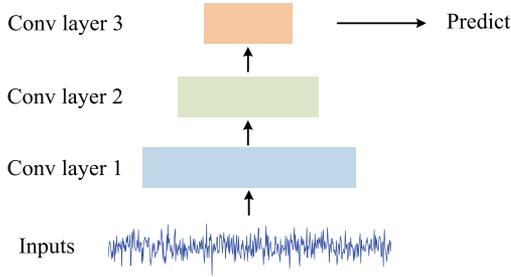


Fig. 1. Traditional CNN baseline.

B. Cascade Structure

Since CNN can extract feature from input automatically and directly without much requirement of prior knowledge, it has been widely used in fault diagnosis applications for now. However, effective as CNN is, it still suffers from several problems due to its essential structure. First, as a hierarchical structure, CNN is constructed by the stack of different layers. With the addition of layers, feature resolution will be reduced due to the consecutive convolution striding or pooling operation, which has been introduced in the Section II-A. The decrease of feature resolution in a certain range in image tasks may have little influence on the final results. Just like we human beings can recognize an object at different reasonable distance. However, industrial vibration signals are different from images. In fault diagnosis tasks, the main mission is to extract the fault response, which can vanish quickly. Therefore, the decrease of feature resolution may lead to some useful information loss and make a difference on the final diagnostic performance.

To overcome the aforementioned problem, the CNN's pyramidal hierarchical structure is utilized naturally to build a cascade structure in a backbone CNN, which can keep both the low-resolution and high-resolution feature maps without the growth of algorithm operational time and memory. There are three points in the cascade structure: traditional CNN baseline, top-down structure, and lateral connections.

1) **CNN Backbone:** The cascade structure starts from a traditional CNN backbone. A simple example is shown in Fig. 1. For each unit, to improve the generalization of the CNN model and allow us to use much higher learning rates, batch normalization (BN) is applied between each convolutional layer and activation layer. Given a minibatch $\{x_1, x_2, \dots, x_m\}$ with the parameters of γ and β , the transformation of BN layer can be described as

$$\mu = \frac{1}{m} \sum_{i=1}^m x_i \quad (3)$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu)^2 \quad (4)$$

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (5)$$

$$\text{BN}_{\gamma, \beta}(x_i) = \gamma \hat{x}_i + \beta. \quad (6)$$

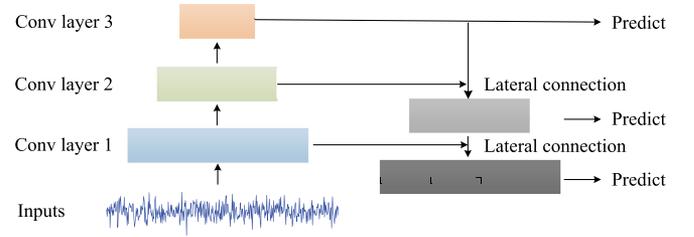


Fig. 2. Top-down structure.

Then, followed by the activation layer. In this article, rectified linear unit (ReLU) activation function is applied, which is defined as

$$a = \max \{0, \text{BN}_{\gamma, \beta}(x_i)\} \quad (7)$$

where $\text{BN}_{\gamma, \beta}(x_i)$ is the output of BN layer.

2) **Top-Down Structure and Lateral Connections:** In the last stage, it can be inferred that the feature resolution will be reduced with the increase of layers due to the convolutional striding or pooling operations. Yosinski *et al.* [25] have explored the generalization and specification of the extracted features by each layers. The experimental results show that the features of the first three layers are mostly general features. After that, the features become more and more specific. As a result, usually the features that used for final classification tasks are rather specific. However, for a motor system, the global features of the previous layers can also be useful if some fault information has been lost during the transmission of network. Therefore, a following top-down pathway is added in the backbone CNN structure to solve this problem [26]. Traditional multiresolution feature extraction will increase the computation considerably. Here, the top-down pathway structure is constructed naturally based on the CNN pyramidal hierarchical structure, which can keep both the low-resolution and high-resolution feature maps without the increase of computation. The graphical illustration is shown in Fig. 2 and the general procedures are summarized as follows.

- 1) The feature maps of the highest layer are used as the input of a fully connected network for prediction.
- 2) Second, the feature maps of the highest layer and the second highest layer are combined and used for prediction.
- 3) Then, repeat the last step until the lowest layer. If there are n layers in the network, there will be $n - 1$ prediction results.

Since there will be size difference between the feature maps of different layers, the higher resolution feature maps are up-sampled first to match with the lower resolution feature size. Then, the feature maps from different layers are merged by lateral connections, as illustrated in Fig. 3. There are many lateral connection methods that can be used. In this article, the elementwise addition is applied.

C. Dilated Convolution Structure

In a CNN, once the convolutional kernel size is defined, the corresponding kernel will be used across the whole perspective

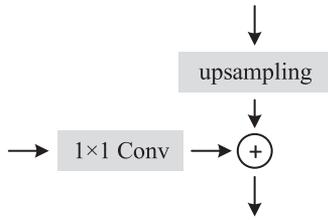


Fig. 3. Lateral connection.

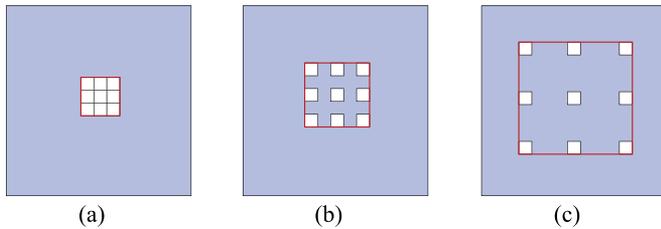


Fig. 4. Dilated convolution structure. (a) Rate = 1. (b) Rate = 2. (c) Rate = 4.

field of input. In this way, the risk of overfitting can be reduced. However, the flexibility of the convolutional kernel is also decreased, and therefore, the kernel size should be determined carefully and suitably, which may lead to a waste of human resources and time. Fixed kernel also limits the applications of CNN for multiscale data analysis, such as the vibration signals under variable rotating speed.

Therefore, to solve the aforementioned problem, a dilated convolution structure is applied based on atrous CNN [27]. A simple example with three rates is shown in Fig. 4. Consider a given input x of current layer, for each location i on the output y and a convolutional kernel (or it is also called as weights) w , dilated convolution is constructed over the input

$$y(i) = \sum_k x(i + r \cdot k)w(k) \quad (8)$$

where the rate r represents the stride we scale the kernel. As an example, the convolutional kernel of standard CNN is a special case for rate $r = 1$. It is found that the kernels with even number sizes can lead to the size difference between input and output feature maps [28]. Therefore, to ensure the kernel size to be odd number, the rate r should be even number here. Usually, r is set to be a power of two to make fully use of this dilated convolution structure.

Then, the scale operation and dilated convolution are implemented in C-CNN. In this stage, feature maps under different scale can be obtained by convolving the input with multiscale kernels. Since it may lead to computation waste if kernels with different sizes are applied directly here, the kernels are upsampled by inserting $r - 1$ zero elements between two kernel values at each dimension, instead of zooming the kernel directly. Thus, the perspective field of each scale kernel will not overlap with each other, which can avoid the computation waste effectively. On the other hand, since the kernel is scaled by inserting zero elements into original kernel, the risk of overfitting will not be increased.

D. C-CNN Diagnosis Framework

As shown in Fig. 5, the framework of the C-CNN is constructed based on a CNN architecture, and composed with dilated convolution layer and cascade structure.

First, because the framework is built based on a backbone CNN, it also inherits the advantages of traditional CNNs, including the powerful ability of feature extraction, invariant to small translations, and low risk of overfitting. Therefore, the proposed method can be used to build an intelligent end-to-end fault diagnosis framework, and the vibration signals can be analyzed directly without other feature extractor.

Then, it can be seen that the dilated convolution is applied at first convolutional layer to extract feature maps from different scales, which can make the diagnose framework more suitable for nonstationary vibration signal analysis.

The cascade operation is applied on the higher part of the framework via four units. The detailed structure of every unit is shown in the right-hand side of Fig. 5. The cascade structure in Fig. 5 is mainly referred to Resnet 18 [29]. There are four convolutional layers in each unit. The number of channels doubles with the unit increase, which are 64, 128, 256, and 512 in each unit, respectively. The length of each layer also halves with the unit increase, which are 64, 32, 16, and 8 in each unit, respectively. Since the cascade structure is built based on the natural CNN pyramid structure, there is no additional computation or memory requirement. The low-resolution feature maps and high-resolution feature maps can be merged together via the lateral connection and used in cascade for prediction, which can avoid information loss caused by the deep architecture.

Specially, the flowchart of the C-CNN fault diagnosis method is presented in Fig. 6. In the first step, the vibration signals are collected by acceleration sensors and cut as the samples of C-CNN. The samples are divided into training set and testing set randomly. Then, the training set are used for training C-CNN model until the terminal conditions are satisfied. Thus, the unknown test signals can be diagnosed to be normal or not by the C-CNN model.

E. Progressive Optimization Algorithm

In this article, cross-entropy loss is applied for predictions. For the target distribution $p(x)$ and the estimated distribution $q(x)$, the cross-entropy is defined as

$$\text{Loss}_i = H(p(x), q(x)) = - \sum_x p(x) \log q(x) \quad (9)$$

where Loss_i denotes the cross-entropy loss of unit i . Traditional gradient decrease algorithms generally use the overall loss $\text{Loss} = \sum_i (\text{Loss}_i)$ as the loss function. However, there is a large amount of parameters in a CNN. The optimization of all these parameters can make the network easier to get into the local optimum.

Therefore, taking the advantage of the cascade structure, a progressive optimization algorithm is proposed in this article. In the learning procedure, the parameters in cascade structure are divided into four parts first. Then, the algorithm optimizes the parameters progressively. That is, given a training epoch K , the C-CNN are optimized for K epochs. In each epoch, we first

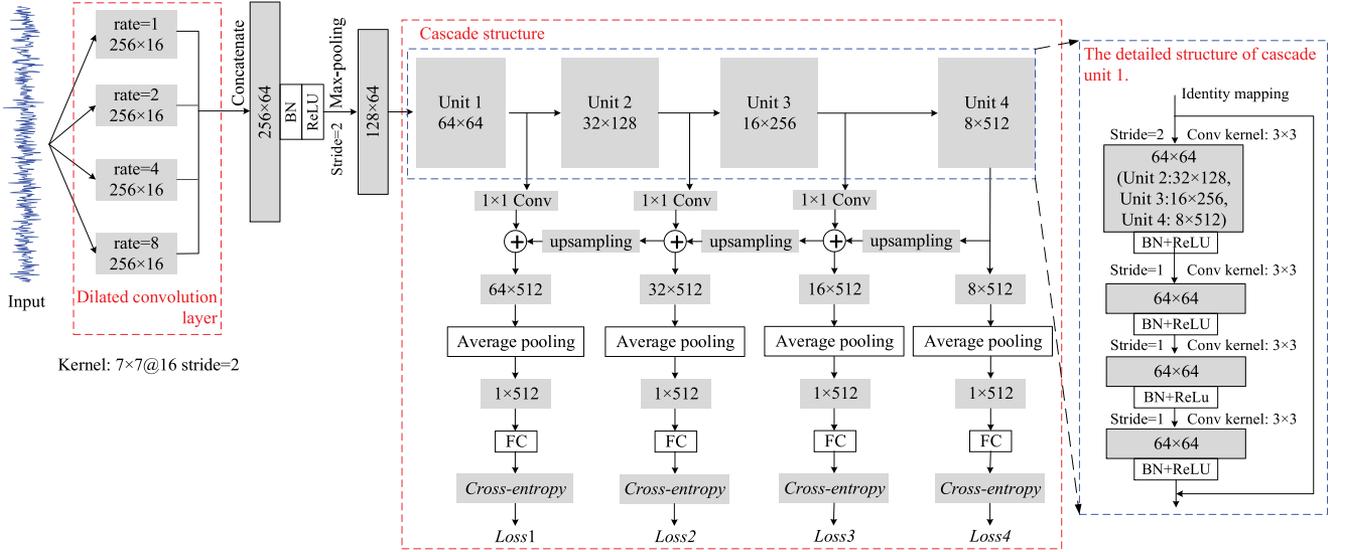


Fig. 5. C-CNN framework, which is composed with a dilated convolution layer and a cascade structure based on a CNN backbone, and the detail structure of cascade unit 1. The convolutional kernel sizes of unit 2, unit 3, and unit 4 are 32×128 , 16×256 , and 8×512 , respectively.

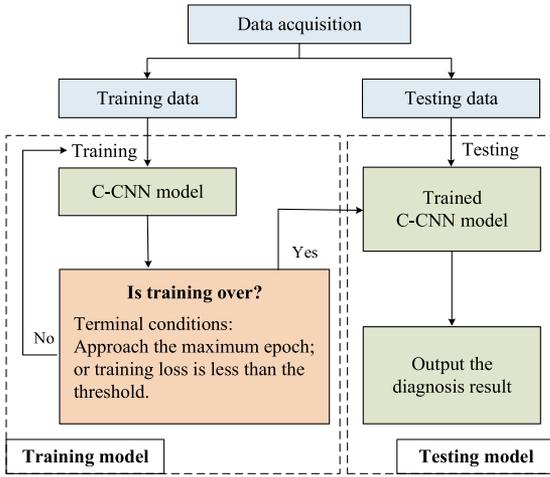


Fig. 6. C-CNN fault diagnosis method.

optimize the C-CNN with the high-level loss, i.e., $Loss_4$ shown in Fig. 5. Then, the $Loss_4 + Loss_3$ and $Loss_4 + Loss_3 + Loss_2$ are utilized to optimize the network, successively. Finally, all high-level and low-level losses, as shown in Fig. 5, are summed up to update the networks. The pseudocode of the progressive optimization algorithm is shown in Algorithm 1.

Compared with the traditional method that directly applies the sum of all losses to optimize the network, the proposed progressive optimization algorithm could learn better high-level features, then advance the low-level features via better high-level features. With this promotion, the C-CNN could converge to a better state, thereby increasing the performance of the motor fault diagnose. Theoretically, the progressive optimization can be treated as divide-and-conquer strategy, which makes optimization process easier via several divided steps.

Algorithm 1: Progressive Optimization Algorithm.

Input:

Training set $D = \{(x^{(i)}, y^{(i)})\}, i = 1, \dots, m$;

Output:

A CNN with optimized parameters.

Initialization:

Initialize parameters θ in the range of $(0,1)$;
 Initialize CNN cascade amount C ;
 Initialize training epoch K ;
 Initialize current epoch $k = 0$.

while $k < K$ do

Randomly shuffle $D \rightarrow D$;

Progressive Optimization:

Sequentially sample batch from $D \rightarrow (X, Y)$;
 Initialize current cascade $c \leftarrow C$;

while $c > 0$ do

Computer Cross Entropy Loss by (9);

$$Loss = \sum_{i=c}^C Loss_i$$

Optimize Loss with Adam optimization to get new

CNN $\theta^* \rightarrow \theta$

$c \leftarrow c - 1$;

end while

$k \leftarrow k + 1$

end while

III. MOTOR FAULT DIAGNOSIS UNDER CONSTANT ROTATING SPEED

The first experiment is conducted under constant rotating speed to simulate the stationary operation condition. The experiment setup is shown in Fig. 7, which is constructed by motor, tachometer, bearing, shaft, load disk, belt, bevel gearbox, magnetic load, variable speed controller, etc. In this experiment,

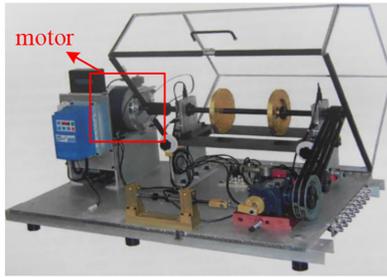


Fig. 7. Experimental setup.



Fig. 8. Locations of two accelerometer sensors.

the rotating speed is set to be a constant. In the first experiment, nine motors with different health or failure conditions are tested, including broken rotor bar, bowed rotor, faulted bearing, unbalanced rotor, normal motor, angular misalignment, parallel misalignment, low impedance, and high impedance, which are labeled as 1 to 9, respectively.

In addition, because vibration signals consist of plentiful information with the dynamical features, and the accelerometer is an inexpensive and immediate measurement, vibration signal analysis has been proven to be an effective way for motor fault diagnosis [30], [31]. Therefore, vibration signals are collected and analyzed by the proposed C-CNN method in this article. Two accelerometers are used for vibration data acquisition, as shown in Fig. 8. The sampling frequency is set to be 6400 Hz. Since we aim to build an end-to-end diagnosis method, the vibration signals are cut into samples directly after data acquisition, and the length of each segment is 512 points. 28 333 samples are selected randomly for training, and the rest 7084 samples are kept for testing.

Then, the proposed C-CNN (progressive C-CNN) method is applied to diagnose the nine different motor conditions. The confusion matrix of the analysis results is shown in Fig. 9(a). To illustrate the advantages of the proposed cascade structure and the dilated convolution operation, the same training samples are also used to train four different CNN models for comparison, which are: C-CNN without progressive optimization algorithm (C-CNN), C-CNN without dilated convolution (C-CNN), traditional CNN with dilated convolution (D-CNN), and traditional CNN without dilated convolution (CNN). The confusion matrices of the analysis results of the four comparison methods are

also shown in Fig. 9(b)–(e). The percentages are shown here instead of the sample numbers, which can be more reasonable because the sample number of each motor condition is different from each other. For a better data analysis, t-distributed stochastic neighbor embedding (t-SNE) technique [32] is applied for feature visualization, as shown in Fig. 10. In this experiment, the accuracy, recall, precision, and F1 score are used as the performance estimation indexes of the analysis results. The estimation indexes of different methods are shown in Table I.

It can be seen from Fig. 9 that the proposed method performs best at diagnosing different motor faults, with overall accuracy of 99.62%. Then, followed by the C-CNN without progressive optimization, C-CNN, D-CNN, and traditional CNN. It can be concluded that the progressive optimization algorithm, the cascade structure, and the dilated convolution operation can improve the diagnostic accuracy of CNN model. Cascade algorithm can be more useful than dilated convolution and progressive optimization in this experiment. It can also be found that classes 5 and 6 are the most confusing faults, which can also be verified in Table I. But the proposed method still can distinguish them better than the comparison methods, which is due to the fact that cascade structure can learn high-level features and low-level features from different scales, which can improve the diagnostic accuracy effectively. Except these two classes, the diagnostic accuracy of the proposed method are all higher than 99.90%. In Fig. 9(b) and (c), C-CNN without progressive optimization performs best in the four comparison methods. C-CNN performs more stable than other two comparison methods. Except classes 5 and 6, classes 4 and 6 are also confused by D-CNN in Fig. 9(d), which illustrates that with the growth of layer in CNN architecture, the information that can distinguish unbalanced rotor and angular misalignment is lost by the consecutive convolution striding or pooling operation. It can be seen from Fig. 9(e) and Table I that classes 1, 4–7, and 9 are confused by the traditional CNN, which indicate that the dilated convolution can extract the feature information of classes 1, 7, and 9 effectively. It can be seen from Fig. 10 that the interclass distance of the proposed C-CNN method is largest, as shown in Fig. 10(a). As a result, the diagnosis accuracy of the proposed method is highest. The interclass distances of C-CNNs in Fig. 10(b) and (c) are smaller than the proposed method, but larger than traditional CNNs. In Fig. 10(d) and (e), it can be seen that the interclass distances of traditional CNNs are smaller than C-CNNs, and the intraclass distances are larger, which increase the difficulty of fault diagnosis.

To further demonstrate the convergence process of the proposed method, the train accuracy and test accuracy of each epoch are shown in Fig. 11(a) and (b), respectively. It can be found that the training accuracy converges quickly, and get to 100% from epoch 14. After epoch 14, training accuracy stays in 100% except one epoch, which shows the stability of the proposed method. Similarly, test accuracy converges to 99.65% at epoch 19. Therefore, the proposed method can effectively diagnose different motor faults. The comparison results with state-of-the-art methods also illustrate the advantages of the cascade algorithm and dilated convolution.

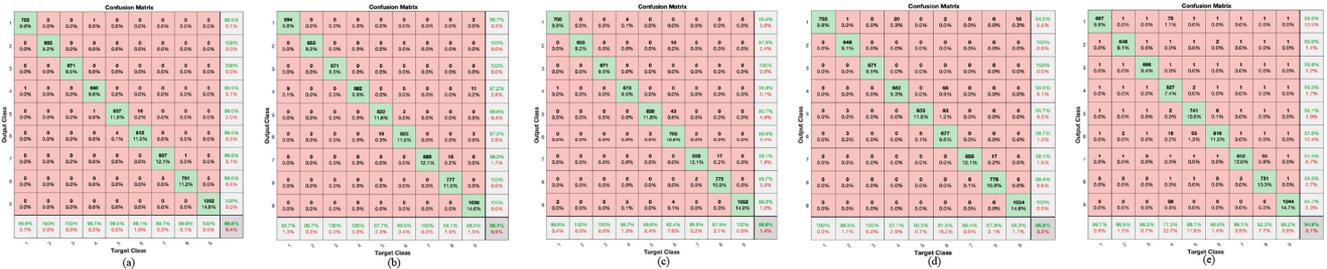


Fig. 9. Confusion matrix of (a) the proposed method, (b) the C-CNN without progressive optimization algorithm, (c) the C-CNN without dilated convolution, (d) traditional CNN with dilated convolution, and (e) traditional CNN without dilated convolution. (Best viewed in zoom up.)

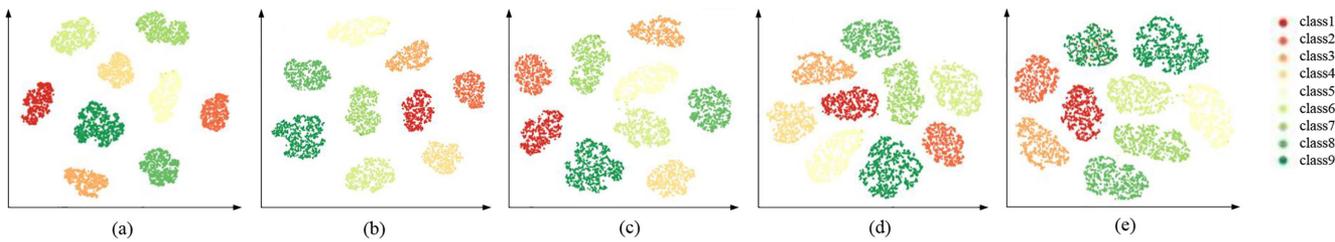


Fig. 10. Feature visualization via t-SNE of (a) the proposed method, (b) the C-CNN without progressive optimization algorithm, (c) the C-CNN without dilated convolution, (d) traditional CNN with dilated convolution, and (e) traditional CNN without dilated convolution.

TABLE I
THE ESTIMATION INDEXES OF THE PROPOSED METHOD AND FOUR COMPARISON METHODS

Index	1	2	3	4	5	6	7	8	9
C-CNN with progressive optimization, accuracy=99.62%									
<i>r</i>	99.86%	100%	100%	99.85%	97.98%	99.52%	99.88%	99.62%	100%
<i>p</i>	99.86%	100%	100%	99.71%	99.52%	98.10%	99.65%	99.87%	100%
F1	99.86%	100%	100%	99.78%	98.74%	98.80%	99.77%	99.75%	100%
C-CNN without progressive optimization, accuracy=99.11%									
<i>r</i>	98.72%	99.69%	100%	100%	97.74%	99.64%	100%	98.11%	98.57%
<i>p</i>	99.71%	100%	100%	97.29%	99.64%	97.17%	98.29%	100%	100%
F1	99.21%	99.85%	100%	98.63%	98.68%	98.39%	99.14%	99.04%	99.28%
Cascade CNN without dilated convolution, accuracy=98.56%									
<i>r</i>	99.43%	97.40%	100%	99.85%	94.53%	99.64%	98.02%	99.75%	98.86%
<i>p</i>	99.57%	100%	100%	98.70%	99.62%	92.39%	99.76%	97.89%	100%
F1	99.50%	98.69%	100%	99.27%	97.01%	95.87%	98.89%	98.81%	99.43%
Traditional CNN with dilated convolution, accuracy=96.75%									
<i>r</i>	95.59%	100%	100%	88.56%	87.99%	99.52%	98.37%	99.75%	100%
<i>p</i>	100%	99.39%	100%	97.42%	99.73%	82.24%	99.76%	98.26%	98.69%
F1	97.75%	99.70%	100%	92.78%	93.49%	90.05%	99.06%	99.00%	99.34%
Traditional CNN without dilated convolution, accuracy=94.90%									
<i>r</i>	86.49%	99.85%	100%	99.85%	99.29%	85.02%	93.14%	99.87%	92.97%
<i>p</i>	100%	99.70%	100%	78.10%	88.92%	99.29%	99.88%	93.17%	100%
F1	92.75%	99.77%	100%	87.64%	93.82%	91.61%	96.39%	96.40%	96.35%

IV. MOTOR FAULT DIAGNOSIS UNDER VARIABLE ROTATING SPEED

Due to the wide applications of motors, the operational conditions cannot always be guaranteed to be constant speed. Many factors can lead to the variate of operational conditions, such as the fluctuation of wind speed to wind turbine motors, the change of speed to vehicle motors, etc. On the other hand, as the applications of motors grow more and more popular, the operational conditions are also changing dramatically, which

lead to the use of motors extends to variable rotating speed conditions. Therefore, to further verify the effectiveness of the proposed method in modern industrial systems, the second experiment is conducted under variable speed. The rotating speed is controlled by the speed controller manually, which ranges from 0 to 3600 r/min. Six motors with different faults are tested in this experiment, namely, broken rotor bar, bowed rotor, faulted bearing, unbalanced rotor, normal motor, and high impedance.

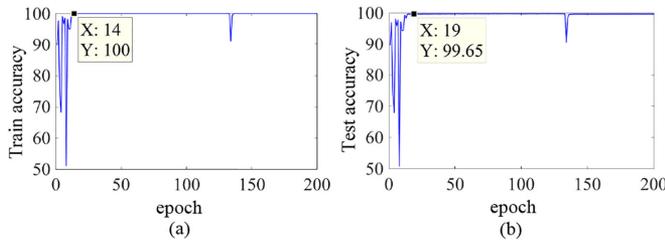


Fig. 11. (a) Training accuracy and (b) testing accuracy of every epoch.

TABLE II
THE ESTIMATION INDEXES OF THE PROPOSED METHOD
AND FOUR COMPARISON METHODS

Index	1	2	3	4	5	6
C-CNN with progressive optimization, accuracy=94.49%						
<i>r</i>	93.58%	85.30%	98.32%	100%	94.55%	94.77%
<i>p</i>	89.10%	95.18%	98.87%	100%	90.84%	93.48%
F1	91.28%	89.97%	98.60%	100%	92.66%	94.12%
C-CNN without progressive optimization, accuracy=92.63%						
<i>r</i>	86.59%	84.15%	96.08%	99.18%	95.91%	93.39%
<i>p</i>	89.60%	89.30%	96.89%	99.73%	86.70%	93.91%
F1	88.07%	86.65%	96.48%	99.45%	91.07%	93.65%
Cascade CNN without dilated convolution, accuracy=92.49%						
<i>r</i>	93.58%	84.73%	93.28%	100%	94.28%	88.71%
<i>p</i>	88.16%	85.47%	98.23%	99.46%	87.82%	96.70%
F1	90.79%	85.09%	95.69%	99.73%	90.93%	92.53%
Traditional CNN with dilated convolution, accuracy=90.92%						
<i>r</i>	91.34%	86.46%	86.27%	100%	96.73%	84.30%
<i>p</i>	89.34%	84.51%	100%	100%	85.54%	87.93%
F1	90.33%	85.47%	92.63%	100%	90.79%	86.08%
Traditional CNN without dilated convolution, accuracy=86.47%						
<i>r</i>	87.71%	74.06%	95.52%	100%	85.01%	76.03%
<i>p</i>	71.85%	83.71%	88.57%	96.83%	84.32%	98.22%
F1	78.99%	78.59%	91.91%	98.39%	84.67%	85.71%

The vibration signals are cut into samples directly as the first experiment, and the length of each segment is still 512 points. 8630 samples are selected randomly for training and the rest 2158 samples are used for testing.

Then, the proposed method is used to diagnose different conditions of the six motors under variable rotating speed. The confusion matrix of the proposed method is shown in Fig. 12(a). C-CNN without progressive optimization algorithm, C-CNN, D-CNN, and traditional CNN are also used for comparison, as shown in Fig. 12(b)–(e). t-SNE is applied for feature visualization, as shown in Fig. 13. The accuracy, recall, precision, and F1 score are also calculated to estimate the performance of different methods, as shown in Table II. It can be concluded from Table II and Fig. 12 that the proposed method performs better than the comparison methods, with overall accuracy of 94.49%. The accuracy of the proposed method has also been increased much larger from the traditional CNN (8.02%) than the first experiment (4.72%). This is because although traditional CNN shows advantages in extracting feature automatically, it is not suitable to analyze multiscale signals due to the fixed kernel size. While the proposed C-CNN method not only keeps fault information during the convolution striding and pooling layers, but also can

extract features from multiscales by the dilated convolution operation, which makes the C-CNN method more suitable for fault diagnosis under nonstationary conditions. In addition, because traditional gradient decrease algorithms enable the optimization with difference of convex objective functions converges to a local minimum or a saddle point, the iterative usage of gradient decrease algorithm in the proposed progressive optimization algorithm can guarantee the stability of learning procedure by decomposing the parameters into several parts and optimizing the C-CNN from bottom to top in a progressive manner, which can make the converge procedure more smooth and extend the C-CNN to motor fault diagnosis applications under nonstationary conditions. As a result, the progressive optimization algorithm also shows advantage in dealing with nonstationary signals, as shown in Fig. 12(b). In this experiment, classes 2 and 4 are the most confusion classes. It can be seen from Table II and Fig. 12(c) and (d) that the confusion between classes 1, 3, and 4 is increased in C-CNN, and the confusion between classes 1 and 5 is increased in D-CNN compared with the proposed method, which illustrates that some fault information of classes 1, 3, and 4 cannot be extracted from one scale, and the fault information of bowed rotor and high impedance are lost by the convolution striding and pooling. Then, it can be found from Fig. 12(e) that the first four classes are seriously confused by the traditional CNN, this is because the traditional CNN focuses its receptive field in a local area by a fixed kernel, which can be useful in image recognition. But vibration signals are different from images, the global perception is as important as local perception. Fault diagnosis based on a long vibration signal can be more accurate than a short vibration signal segment. It can be seen from Fig. 13 that the interclass distances in this experiment are larger than the first experiment under constant operational conditions, and the intraclass distances are smaller, which demonstrates that the fault diagnosis under variable operational conditions is more difficult. As shown in Fig. 13(a), the interclass distance of the proposed C-CNN method is largest, which shows that the proposed method can extract more representative and distinguishing information than traditional CNNs.

Then, to further illustrate the diagnose process of the proposed method, the train accuracy and test accuracy of each epoch are shown in Fig. 14. It can be seen from Fig. 14 that the convergence in this experiment is not as stable as the first experiment. This is a reasonable result. The oscillatory variation may result from the changeable operational conditions. Although the train and test accuracy begin to shock after epoch 50, the average accuracy is still better than 90%, which is an improvement for motor fault diagnosis under nonstationary conditions. Then, based on a GPU server with a NVIDIA Tesla P100, C-CNN is implemented repeatedly to predict the motor fault with the whole testing sets and record the processing time for 200 times, both on the cases of constant and variable rotating speed. The cumulative distribution function (CDF) figure is drawn in Fig. 15. It can be seen from the figure that the median processing time to make one fault prediction is between 7 and 9 ms. Therefore, the proposed method can perform better than traditional CNNs without the sacrifice of efficiency.

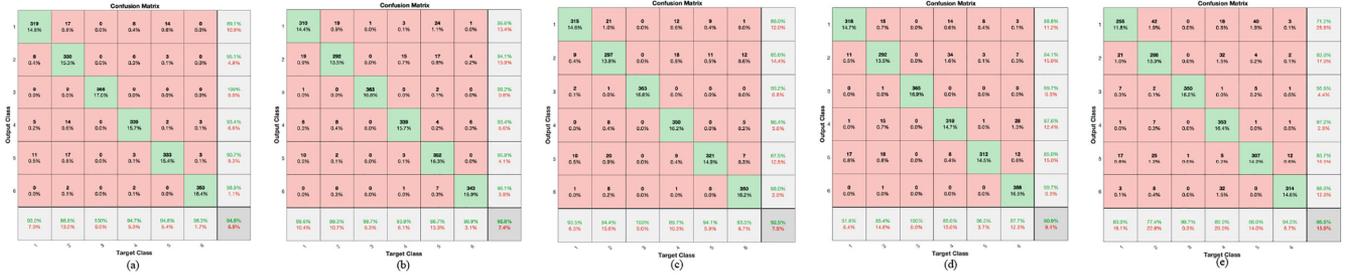


Fig. 12. Confusion matrix of (a) the proposed method, (b) the C-CNN without progressive optimization algorithm, (c) the C-CNN without dilated convolution, (d) traditional CNN with dilated convolution, and (e) traditional CNN without dilated convolution. (Best viewed in zoom up.)

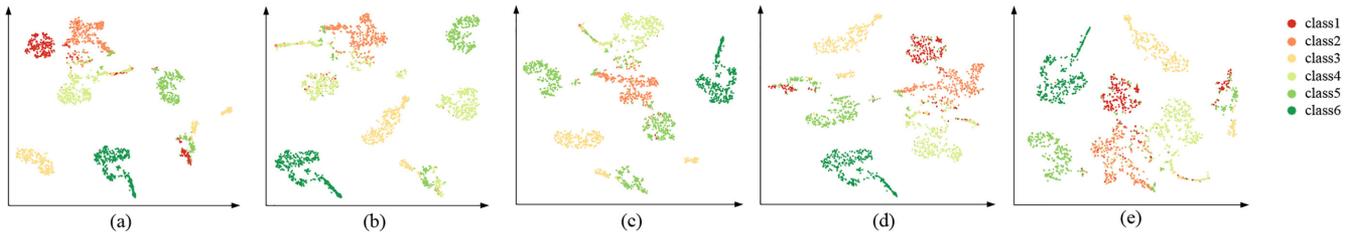


Fig. 13. Feature visualization via t-SNE of (a) the proposed method, (b) the C-CNN without progressive optimization algorithm, (c) the C-CNN without dilated convolution, (d) traditional CNN with dilated convolution, and (e) traditional CNN without dilated convolution.

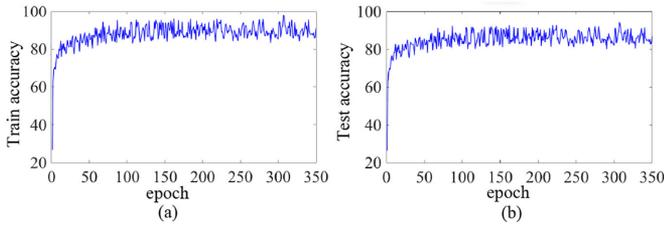


Fig. 14. (a) Training accuracy and (b) testing accuracy of every epoch.

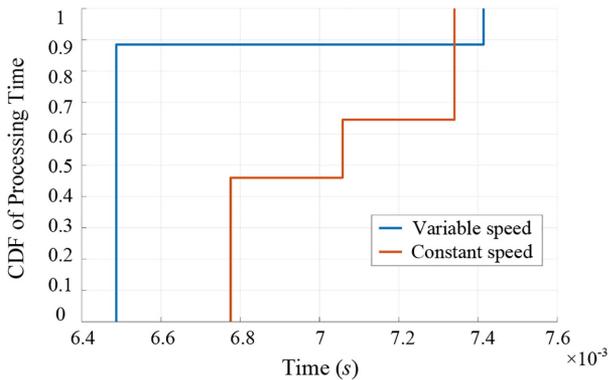


Fig. 15. CDF figure of the recorded time both under the constant and variable rotating speed.

Because of the dilated convolution structure, the proposed C-CNN method is also robust to missing data. To evaluate the performance of C-CNN on the case of missing, 0%, 10%, 20%, 30%, and 40% sampling points have been removed from the original samples, respectively. Then, linear interpolation is

TABLE III
DIAGNOSIS ACCURACY OF THE PROPOSED METHOD UNDER DIFFERENT MISSING DATA RATE

Missing data rate	constant rotating speed	variable rotating speed
0%	99.62%	94.49%
10%	98.66%	92.72%
20%	97.52%	91.33%
30%	94.31%	89.39%
40%	90.42%	86.14%

applied to scale the samples back to the time series with 512 points. After the operations, the C-CNN is utilized to analyze the removed-then-interpolated data. The results of the constant and variable rotating speed are shown in Table III. It can be seen that when the missing data rate is less than 20%, there is little performance degradation of the proposed method. When the missing data rate is 30% and 40%, the performance of the proposed method degrade for more than 5%. Therefore, it can be concluded that the proposed method has robust to missing data when the missing data rate is less than 20%. But when the missing rate is larger than 30%, the diagnosis accuracy will deteriorate drastically.

V. CONCLUSION

This article proposed a novel motor fault diagnosis method named C-CNN. Then, based on the natural structure of C-CNN, a progressive optimization algorithm was proposed for a better convergence. As an improved CNN model, C-CNN not only inherits the advantages of the traditional CNN that can extract features from vibration signals directly, but also increase the

stability and accuracy of the proposed method and make the proposed method more suitable for motor fault diagnosis under nonstationary conditions. Two experiments were conducted under constant speed and variable speed, respectively, to verify the effectiveness of C-CNN method. The experimental results showed that thanks to the cascade structure, the information loss during the convolution striding and pooling can be avoided. In addition, dilated convolution operation showed advantages in feature extraction from multiple scales, which extends the applications of CNN to fault diagnosis under nonstationary conditions. Finally, by decomposing the parameters into several parts and optimizing the C-CNN from bottom to top in a progressive manner, the progressive optimization algorithm makes the C-CNN more suitable for motor fault diagnosis under nonstationary conditions. The comparison results with four different kinds of CNNs also illustrated the improvement of diagnose accuracy by the proposed method. Note that early failures and multiple combined failures were two difficult issues needed to be addressed, future works can be focused on how to utilize the powerful feature extraction of C-CNN for weak or coupled fault feature extraction.

REFERENCES

- [1] S. Munikoti, L. Das, B. Natarajan, and B. Srinivasan, "Data-driven approaches for diagnosis of incipient faults in dc motors," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5299–5308, Sep. 2019.
- [2] J. L. Contreras-Hernandez, D. L. Almanza-Ojeda, S. Ledesma-Orozco, A. Garcia-Perez, R. J. Romero-Troncoso, and M. A. Ibarra-Manzano, "Quaternion signal analysis algorithm for induction motor fault detection," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8843–8850, Nov. 2019.
- [3] J. Song *et al.*, "Accurate demagnetization faults detection of dual-sided permanent magnet linear motor using enveloping and time-domain energy analysis," *IEEE Trans. Ind. Informat.*, to be published, doi: [10.1109/TII.2019.2962730](https://doi.org/10.1109/TII.2019.2962730).
- [4] J. Wang, P. Fu, L. Zhang, R. X. Gao, and R. Zhao, "Multilevel information fusion for induction motor fault diagnosis," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 5, pp. 2139–2150, Oct. 2019.
- [5] P. Liang, C. Deng, J. Wu, G. Li, Z. Yang, and Y. Wang, "Intelligent fault diagnosis via semi-supervised generative adversarial nets and wavelet transform," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4659–4671, Jul. 2020.
- [6] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, no. 1/2, pp. 108–126, 2013.
- [7] S. Wang, I. Selesnick, G. Cai, Y. Feng, X. Sui, and X. Chen, "Nonconvex sparse regularization and convex optimization for bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 9, pp. 7332–7342, Sep. 2018.
- [8] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, 2018.
- [9] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mech. Syst. Signal Process.*, vol. 85, pp. 746–759, 2017.
- [10] J. B. Ali, B. ChebelMorello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network," *Mech. Syst. Signal Process.*, vol. 56, pp. 150–172, 2015.
- [11] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9521–9530, Dec. 2019.
- [12] J. Yu, "Evolutionary manifold regularized stacked denoising autoencoders for gearbox fault diagnosis," *Knowl.-Based Syst.*, vol. 178, pp. 111–122, 2019.
- [13] S. J. Chang and J. B. Park, "Wire mismatch detection using a convolutional neural network and fault localization based on time–frequency-domain reflectometry," *IEEE Trans. Ind. Electron.*, vol. 66, no. 3, pp. 2102–2110, Mar. 2019.
- [14] S. Guo, B. Zhang, T. Yang, D. Lyu, and W. Gao, "Multi-task convolutional neural network with information fusion for bearing fault diagnosis and localization," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 8005–8015, Sep. 2020.
- [15] W. Yu and C. Zhao, "Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability," *IEEE Trans. Ind. Electron.*, vol. 67, no. 6, pp. 5081–5091, Jun. 2020.
- [16] Z. An, S. Li, J. Wang, and X. Jiang, "A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network," *ISA Trans.*, vol. 100, pp. 155–170, 2020.
- [17] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on long short-term memory networks," *Renewable Energy*, vol. 133, pp. 422–432, 2019.
- [18] C. Hu, H. Pei, X. Si, D. Du, Z. Pang, and X. Wang, "A prognostic model based on DBN and diffusion process for degrading bearing," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8767–8777, Oct. 2020, doi: [10.1109/TIE.2019.2947839](https://doi.org/10.1109/TIE.2019.2947839).
- [19] Y. Wang, Z. Pan, X. Yuan, C. Yang, and W. Gui, "A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network," *ISA Trans.*, vol. 96, pp. 457–467, 2020.
- [20] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vol. 72, pp. 303–315, 2016.
- [21] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.
- [22] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.
- [23] R. Liu, F. Wang, B. Yang, and S. J. Qin, "Multi-scale kernel based residual convolutional neural network for motor fault diagnosis under non-stationary conditions," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3797–3806, Jun. 2020.
- [24] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, 2018.
- [25] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
- [26] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.
- [27] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, [arXiv:1706.05587](https://arxiv.org/abs/1706.05587).
- [28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [30] W. Deng, S. Zhang, H. Zhao, and X. Yang, "A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing," *IEEE Access*, vol. 6, pp. 35042–35056, 2018.
- [31] Y. Shi, S. Ji, F. Zhang, F. Ren, L. Zhu, and L. Lv, "Multi-frequency acoustic signal under short-circuit transient and its application on the condition monitoring of transformer winding," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1666–1673, Aug. 2019.
- [32] L. v. d. Maaten and G. Hinton, "Visualizing data using T-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.



Fei Wang received the Ph.D. degree in computer science and technology from Xi'an Jiaotong University, Xi'an, China, in 2020.

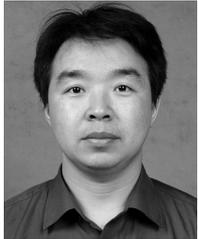
He is currently an Assistant Professor with the School of Software Engineering, Xi'an Jiaotong University. His current research interests include deep learning, time-series mining, human sensing, and robotic sensors.



Ruonan Liu (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2013, 2015, and 2019, respectively.

She was a Postdoctoral Researcher with the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, in 2019. She is currently an Associate Professor with the College of Intelligence and Computing, Tianjin University, Tianjin, China. Her current research interests include machine learning, fault diagnosis, and computer vision.

and computer vision.



Qinghua Hu (Senior Member, IEEE) received the B.S. and M.S. degrees in energy science and engineering from the Harbin Institute of Technology, Harbin, China, in 1999 and 2002, respectively, and the Ph.D. degree in aerospace from the Harbin Institute of Technology, Harbin, China, in 2008.

He was a Postdoctoral Fellow with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, from 2009 to 2011. He is currently the Dean of the School of Artificial Intelligence, the Vice Chairman of the Tianjin Branch of China Computer Federation, and the Vice Director of the SIG Granular Computing and Knowledge Discovery, Chinese Association of Artificial Intelligence. He is currently supported by the Key Program, National Natural Science Foundation of China. He has authored or co-authored more than 200 peer-reviewed papers. His current research interests include uncertainty modeling in big data, machine learning with multimodality data, and intelligent unmanned systems.

Dr. Hu is an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Acta Automatica Sinica*, and *Energies*.



Xuefeng Chen (Member, IEEE) received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2004.

He is currently a Full Professor and Dean of the School of Mechanical Engineering, Xi'an Jiaotong University. He was the Executive Director of the Fault Diagnosis Branch in China Mechanical Engineering Society. He has authored more than 100 SCI publications in areas of composite structure, aero-engine, wind power equipment, etc.

Prof. Chen was the recipient of the National Excellent Doctoral Thesis Award in 2007, First Technological Invention Award of Ministry of Education in 2008, Second National Technological Invention Award in 2009, First Provincial Teaching Achievement Award in 2013, Science and Technology Award for Chinese Youth in 2013, and First Technological Invention Award of Ministry of Education in 2015. Additionally, he hosted a National Key 973 Research Program of China as the Principal Scientist in 2015. He is a member of the American Society of Mechanical Engineers, and the Chair of IEEE Xian and Chengdu Joint Section Instrumentation and Measurement Society Chapter.