IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 2, 2021, accepted February 13, 2021, date of publication February 16, 2021, date of current version February 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3059761

Rolling Bearing Fault Diagnosis Based on One-Dimensional Dilated Convolution Network With Residual Connection

HAOPENG LIANG[1] **AND XIAOQIANG ZHAO**[1,2,3]

[1]College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China
[2]Key Laboratory of Gansu Advanced Control for Industrial Process, Lanzhou University of Technology, Lanzhou 730050, China
[3]National Experimental Teaching Center of Electrical and Control Engineering, Lanzhou University of Technology, Lanzhou 730050, China

Corresponding author: Xiaoqiang Zhao (xqzhao@lut.edu.cn)

This work was supported in part by the Project funded by the open fund of State Key Laboratory of Large Electric Drive System and Equipment Technology under Grant SKLLDJ012016020, in part by the National Nature Science Foundation of China under Grant 61763029, in part by the Open Fund Project of Key Laboratory of Gansu Advanced Control for Industrial Process under Grant 2019KFJJ01, in part by the National Defense Basic Scientific Research Project under Grant JCKY2018427C002, and in part by the Industrial Support and Guidance Project of Colleges and University of Gansu Province under Grant 2019C-05.

ABSTRACT As the rolling bearing is the most important part of rotating machinery, its fault diagnosis has been a research hotspot. In order to diagnose the faults of rolling bearing under different noisy environments and different load domains, a new method named one-dimensional dilated convolution network with residual connection is proposed in this paper. The proposed method uses the one-dimensional time-domain signals of rolling bearing as input. Zigzag dilated convolution is introduced into convolution neural network, which can effectively improve the receptive field of the convolutional layer. A multi-level residual connection structure with different weight coefficients is constructed, so that the lower layer features of convolution neural network can be transferred to the upper layer, which improves the feature learning ability. Moreover, in order to enhance the useful features and weaken the useless features, we add the attention module Squeeze-and-Excitation (SE) block after each sub-residual structure. By using the rolling bearing datasets, the experimental results show that the proposed method can effectively diagnose faults of rolling bearing under different noisy environments and different load domains. Compared with other methods, the proposed method has higher accuracy.

INDEX TERMS Different load domains, different noisy environments, dilated convolution, one-dimensional convolution neural network, rolling bearing fault diagnosis, residual connection.

I. INTRODUCTION

As the most important part of rotating machinery, the health status of rolling bearing could affect the performance, stability and life cycle of the whole rotating machinery. Because the working environment of rolling bearing in rotating machinery is generally complex, the rolling bearing has problems, such as aging and damage during operation, which would cause safety accidents and huge economic losses. Therefore, many people have paid more and more attention to the fault diagnosis of rolling bearing [1], [2]. In recent years, with the rapid development of computer technology, a large amount of condition monitoring data of mechanical equipment have been stored and analyzed. The data-driven fault diagnosis methods can extract and detect useful fault information from a large amount of monitoring data and do not need to establish an accurate system model, and they are suitable for complex systems that are difficult to establish an explicit model. Therefore, the researches on data-driven fault diagnosis methods of rolling bearing have been paid more and more attention[3]. For example, Yan and Jia [4] proposed an optimized SVM fault classification algorithm based on multi-domain features to improve the accuracy of fault classification of rolling bearing. Lu *et al.* [5] used the genetic algorithm and empirical mode decomposition (EMD) to extract features and then used SVM to classify and identify faults. Zhang *et al.* [6] used local mean decomposition to eliminate the noise in vibration signals. Mao *et al.* [7] proposed a classification method combining multi-hole permutation entropy and support vector machine (SVM) to classify bearing fault types. Although the

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

above-mentioned methods can extract fault features, they are affected by human subjective factors in the process of fault diagnosis. By relying on existing fault diagnosis experience, they are very difficult to deal with fault problems under complex working conditions [8], [9].

In recent years, with the development of artificial intelligence, fault diagnosis methods based on deep learning have received more and more attention [10], [11]. These methods combine fault feature extraction and feature classification, and automatically extract representative features from the original signal data, thereby eliminating the influence of artificial experience on feature extraction.

In deep learning, commonly used methods include SAE [12], CNN [13], DBN [14] and RNN [15], and then based on Convolutional Neural Network (CNN), VGG [16], Resnet [17], Inception-v4 [18], Capsule and many other neural networks are developed. These methods have achieved great success in the field of computer vision. In the field of mechanical fault diagnosis, many experts and scholars have introduced CNN and made progress. Zhang *et al.* [19] used short-time Fourier Transform to transform one-dimensional time-domain signals of rolling bearing into two-dimensional data and used hierarchical regularization to enhance the training results of CNN. Jiang *et al.* [20] proposed a new multi-scale convolutional neural network (MSCNN), which could simultaneously extract and classify multi-scale features of gearbox vibration signals. Chen *et al.* [21] used continuous wavelet transforms to preprocess the original vibration signals, then used a square pool architecture CNN to extract high level features, finally used the Extreme Learning Machine (ELM) classifier to implement fault classification. The above methods transform the original one-dimensional vibration data into two-dimensional data, and then input them into convolution network for fault diagnosis. However, most of the rolling bearing data are one-dimensional time series or frequency series. Compared with using the original one-dimensional data directly, transforming the original one-dimensional data into two-dimensional data may cause poor fault diagnosis results. The one-dimensional neural network directly uses the original data, which is more convenient for fault diagnosis. Abdeljaber *et al.* [22] applied one-dimensional CNN to the normalized vibration signals for damage detection and real-time location of structural damage, which reduced the dependence on manual feature extraction. Zhang *et al.* [23] proposed a one-dimensional CNN with a wide convolutional layer, which could effectively suppress high-frequency noise interference in bearing signals. Li *et al.* [24] proposed a method combining residual neural network (ResNet) with one-dimensional separable convolution to effectively classify gear pitting faults. Li *et al.* [25] applied the attention mechanism to help the CNN locate fault information, which had a high fault diagnosis accuracy rate under limited data samples. Xue *et al.* [26] proposed a deep convolutional network to extract frequency domain features, then used a support vector machine to classify multiple faults, with an average accuracy of 90.29%. Chen *et al.* [27] proposed a fault diagnosis method of deep capsule network based on random delta rule, which had strong robustness for vibration signal noise interference. Wang *et al.* [28] proposed a method of stacking multiple separable convolution residual connection blocks to learn the advanced features of data and provided accurate prediction results of bearing residual life. Hao *et al.* [29] put forward one-dimensional convolutional long short-term memory network method, which had better adaptability to rolling bearing data of different loads. Zhang *et al.* [30] proposed a fault diagnosis method based on deep residual learning, which could effectively process bearing vibration signals of different sequence lengths.

In the abovementioned literatures, the fault diagnosis methods based on deep learning have advantages compared with the traditional methods, and they have achieved satisfactory results in the fault diagnosis task of rolling bearing [31], [32]. However, the network model of deep learning is complicated, the network structure and parameter selection would affect the fault diagnosis performance of such methods. Many methods usually assume that the distributions of the training and test datasets are same, but the actual loads always change, the training data cannot show a good fitting effect in the test data, resulting in poor performance of fault diagnosis. Besides, due to the strong time-varying feature of vibration signals, the network model is easy to fall into the trap of local false features during training process. Therefore, it is necessary to ingeniously design a deep learning method to produce good fault diagnosis effects. In recent years, many scholars have designed many excellent deep learning models for bearing data, which have made great progress in bearing fault diagnosis under variable load conditions. For example, Peng *et al.* [33] proposed a multi-branch and multi-scale convolutional neural network model, which processed bearing vibration signals into three different signals, and then inputted them into the proposed model to diagnose the faults of train bearing with variable load. Jiao *et al.* [34] proposed a fault diagnosis method based on residual joint adversarial network and introduced an adversarial adaptive discriminator into the residual neural network, and achieved good diagnostic results on planetary gearboxes and rolling bearings with different load domains. Xu *et al.* [35] proposed a feature fusion process with attention mechanism and combined this process with deep learning model, finally achieved good generalization performance in the actual bearing variable load environment. Wang *et al.* [36] proposed a multi-scale domain adaptive network model, which had good adaptability to bearing data characteristics under different working conditions. Dong *et al.* [37] used Fourier transform to convert time-domain signals into frequency-domain signals and inputted them into an autoencoder for adaptive feature extraction. Their methods improve the fault diagnosis effect under variable working conditions.

Rolling bearings have complex fault characteristics under variable working conditions, and it is difficult to make effective diagnosis. In order to improve the anti-noise and generalization ability of the fault diagnosis method, we propose a one-dimensional dilated convolution network with residual connection for fault diagnosis of rolling bearing. Firstly, the one-dimensional time-domain signals are used as the input. Then we construct the wide convolution residual block and the zigzag dilated residual block to learn feature information, and use the global residual and sub-residual weight coefficient structure to transfer the features. Finally, the SE block is added after each sub-residual structure. Experimental results demonstrate that the proposed method has excellent anti-noise and generalization ability.

The main contributions of this paper are as follows:

(1) A zigzag dilated residual connection block with optimal dilation rate is proposed. The signals of rolling bearing have strong time-varying characteristics. In order to extract the characteristic information, we construct a zigzag dilated residual connection block, which can not only expand the receiving field of the convolutional layer, but also avoid the grid effect. Then we analyze the influence of different dilation rate combinations and determine the optimal dilation rate combination. Finally, we apply this residual block in the proposed method and obtain excellent experimental results.

(2) A residual connection structure with optimal weight coefficients is proposed. In order to enhance the fault diagnosis effect of the proposed method, we construct the global residual and sub-residual weight coefficient structure in the network, so that the upper and lower convolutional layers of the network transfer appropriate feature information, and finally analyze the influence of different weight coefficients.

(3) A neural network fault diagnosis model with attention mechanism is proposed. In order to improve the feature information recognition ability of the proposed method, we add an attention mechanism module (SE block) after each sub-residual structure and use this module to learn the output of useful information features by each sub-residual block and suppress useless Information characteristics.

(4) The influences of different dilation rates and different residual connection weight coefficients for the fault diagnosis effect are analyzed through experiments, and experiments in different noise environments and different load domains are carried out. The results show that the proposed method has higher fault diagnosis accuracy than other methods.

The rest of this article consists of the following. The dilated convolution, residual connection and SE block are introduced in Section 2. The method of one-dimensional dilated convolution network with residual connection is proposed in detail in Section 3. By analyzing the influence of different dilation rates and residual connection weight coefficients, for noise environment and different load domains, experiments verify the effectiveness and superiority of the proposed method in Section 4. The conclusion is given in Section 5.

II. RELATED WORKS

A. ZIGZAG DILATED CONVOLUTION

Dilated convolution is a convolution method that can increase the receptive field of the convolution layer. Proposed by Yu and Koltun [38] in 2016, dilated convolution can obtain data features by jumping step size and can output more information and keep the parameter constant. Supposed that $\boldsymbol{x}(a)$ represents the one-dimensional input, the feature length is A, r represents the dilation rate, $w(i)$represents the one-dimensional convolution kernel, $\boldsymbol{y}(a)$ represents the output feature after the dilated convolution operation. The relationship equation between them is as follows:

$$\boldsymbol{y}(a) = \sum_{i=1}^{A} \boldsymbol{x}(a + r \times i) \times w(i) \tag{1}$$

For normal convolution with a convolution kernel size K, after performing a dilated convolution operation with a dilate rate of r, the convolution kernel size will be equal to $K + (K - 1)(r - 1)$. For example, after a dilated convolution operation with a dilation rate of 2, a normal convolution with a convolution kernel size of 3 is equivalent to a dilated convolution kernel with a size of 5. The one-dimensional dilated convolution process is shown in Fig.1.

Although dilated convolution can expand the receptive field, simply stacking dilated convolution layers with the same dilation rate would cause some problems. Because the features of the dilated convolution obtain data by jumping, the obtained feature information would be uneven, that is, some feature locations are accessed frequently, and some feature locations are not accessed. This kind of problem is defined as the grid effect. However, the grid effect can be effectively avoided by constructing zigzag dilated convolution. That is, different dilation rates are set for the dilated convolution of each layer, and the superimposed dilation rates have not a common divisor greater than 1, and the following equation is

$$M_i = \max[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i] \tag{2}$$

r_i represents the dilation rate of the ith layer, M_i represents the maximum dilation rate of the ith layer, assumed that the number of convolutional layers is n, the default value is $M_n = r_n$.

B. RESIDUAL CONNECTION

To solve the problems of gradient disappearance and degradation in the training process of deep convolution neural network, He *et al.* [17] proposed the concept of residual connection in 2016 and designed a residual block with residual connection lines shown in Fig.2.

Assumed that the input of the residual block is $\boldsymbol{x}$, $\{\boldsymbol{W}_i\}$ represents the weight obtained when the input passes through the ith convolutional layer, $F(\boldsymbol{x})$ represents the residual mapping function, $H(\boldsymbol{x})$ represents the output of the residual block, and the input $\boldsymbol{x}$ is directly connected to the output through the identity connection line. The relationship equation between them is as follows:

$$H(\boldsymbol{x}) = F(\boldsymbol{x}, \{\boldsymbol{W}_i\}) + \boldsymbol{x} \tag{3}$$

A simple deformation of Equation (1) is obtained as follows:

$$F(\boldsymbol{x}, \{\boldsymbol{W}_i\}) = H(\boldsymbol{x}) - \boldsymbol{x} \tag{4}$$

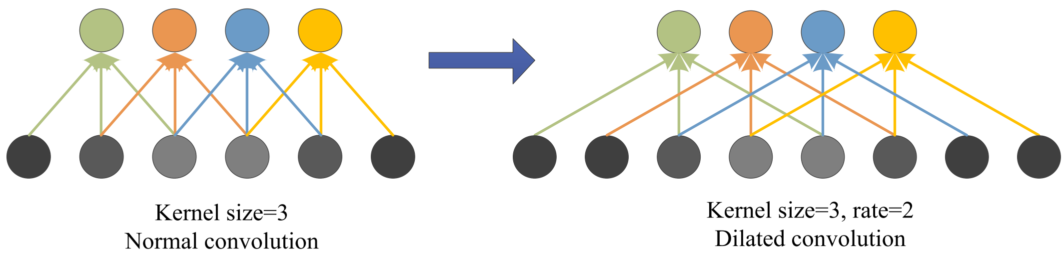


FIGURE 1. One-dimensional dilated convolution.

The convolutional layer in the residual block only needs to learn the difference between input and output. He *et al.* proved that fitting the residual mapping function is easier than fitting the identity mapping function through experiments, and during the training processing of neural network, the errors of the lower layer can be transmitted to the upper layer through the identity connection line, which effectively solves the disappearance of the network gradient with the increase of depth.

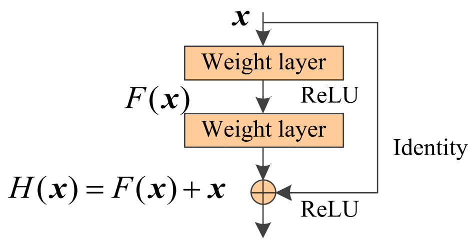


FIGURE 2. Residual block structure.

C. SE BLOCK

Since CNN can generate a multi-channel feature map after feature extraction, different channels of feature maps represent different feature information. The core point of SE block [39] is to use the relationship among these feature channels. SE block obtains the weight information of each feature channel through network learning, and then multiplies this weight information to the feature information of each layer, so that the network can selectively enhance useful feature channels and suppress useless feature channels, thereby realizing the feature channel self-adapt to calibration. The SE block structure is shown in Fig.3.

$\boldsymbol{X}_i$ represents the feature matrix of the corresponding layer, H, W, C represent the three-dimensional information of the feature matrix, Conv represents the normal convolution operation. F_{sq} represents the squeezing operation, the squeezing operation can extract the information of C_2 feature channels by using the global average pool operation, and then a set of weight information with a size of $1 \times 1 \times C_2$ is obtained. u_c represents the cth feature after convolution transformation, z_c represents the cth feature map after the extrusion operation. The relationship equation between them is as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{5}$$

F_{ex} represents the excitation operation, which refers to the generation of corresponding weights for each feature channel through the Sigmoid function. $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ represent the weight information of the two fully connected layers, δ represents the ReLU function, and σ represents the Sigmoid function. The relationship equation between them is as follows:

$$s = F_{ex}(z_c, \boldsymbol{W}) = \sigma(\boldsymbol{W}_2 \delta(\boldsymbol{W}_1 z_c)) \tag{6}$$

F_{scale} represents the rescaling operation, which refers to multiplying the weight generated by each channel with each original feature, s_c represents the weight of the cth feature map after the incentive operation, $\bullet$ represents a scalar and vector multiplication. The relationship equation between them is as follows:

$$\boldsymbol{X}_3 = F_{scale}(u_c, s_c) = s_c \bullet u_c \tag{7}$$

III. THE PROPOSED METHOD

A. RESIDUAL CONNECTION BLOCK

The residual connection blocks are constructed by using two convolutional layers and a pooling layer, which are placed at the front end of the network to extract advanced features from the data. The structure is shown in Fig.4. The first convolutional layer uses a wide convolution kernel of 64×1, and the second convolutional layer uses a large convolution kernel of the size of 7×1. Activation functions BN (Batch Normalization) and ReLU are used after each convolutional layer. Two large convolutional layers can not only obtain longer time-domain sequence information in convolution operation, but also suppress false feature interference in the fault signals. The two maximum pooling layers are responsible for reducing the dimensionality of features and parameter calculation. The construction of the residual connection makes the input features pass backward and avoids the problem of overfitting in the training process.

B. DILATED RESIDUAL CONNECTION BLOCK

The dilated residual connection blocks are constructed by using three consecutive zigzag dilated convolutional layers and are used to effectively learn feature information. The structure is shown in Fig.5. The convolution kernel size of the three empty convolutional layers is set as 3×3, activation

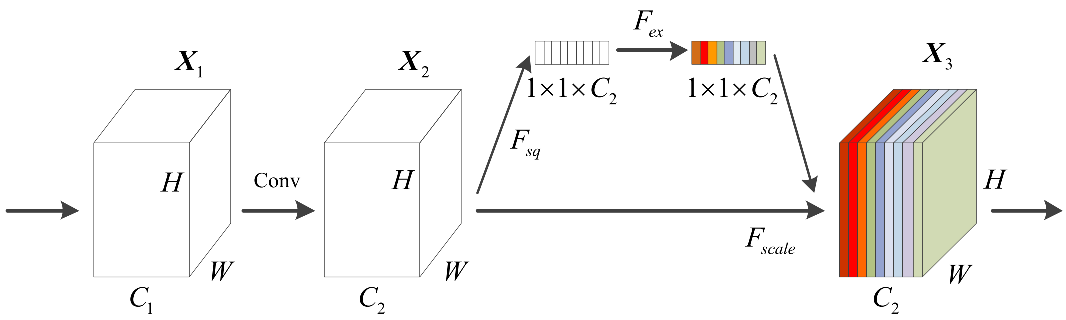


FIGURE 3. SE block structure.

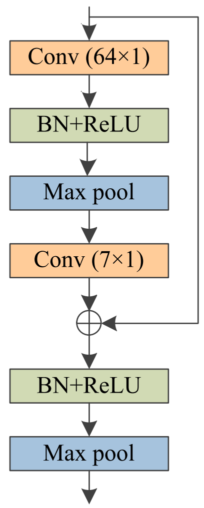


FIGURE 4. Residual connection block.

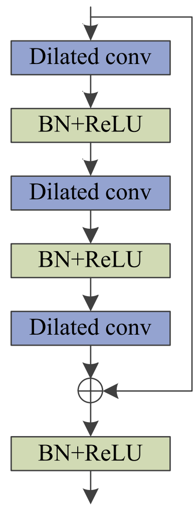


FIGURE 5. Dilated residual connection block.

functions BN (Batch Normalization) and ReLU are used after each layer. Besides, we construct residual connections for dilated convolution to transfer features. Then, in order to avoid the grid effect, different combinations of dilation rates are set up to construct the zigzag dilated convolutional layer. We analyze the effects of different combinations through experiments and then obtain the optimal dilation rate combination. In addition, as the pooling layer enlarges the receptive field, it also changes the feature size and causes partial information loss. However, dilated convolution can overcome this shortcoming. During the downsampling operation, the feature size of each layer remains unchanged, and the receptive field of the convolutional layer can be effectively expanded. Therefore, the maximum pooling layer is only used after the dilated convolutional layer and is not used between the dilated convolutional layers.

C. THE NETWORK STRUCTURE OF THE PROPOSED METHOD

Due to the load change in the working environment of rolling bearing, noise interference and other factors are inevitable, which lead to poor generalization ability and noise resistance. To solve this problem, a one-dimensional dilated convolutional neural network with residual connection method is proposed. It uses the original bearing time-domain data as input, and its network structure is mainly composed of residual connection block, dilated residual connection block, SE block, residual connection and full connection layer (FC), which shows as Fig.6.

First of all, the feature information of input data can be extracted and learned effectively through the residual connection block and the void residual connection block. Then, the SE block is added to the end of the two residual blocks, the beneficial output features of each residual block are adaptively selected and passed to the next residual block or subsequent processing layer. In addition, in order to transfer the feature information between the top layer and the bottom layer, the global residual connection is constructed between the feature input and the last pool layer. The residual connection can improve the learning efficiency of neural network. When the feature information is transmitted to the bottom layer through the residual connection, the default weight coefficient of the feature is 1. However, for neural network with fewer layers, reducing the residual connection feature weight can make the residual block learn more feature information, which is easy to be ignored. Therefore, the proposed method adds λ times of the weight coefficient to each residual connection line and analyzes the influence of

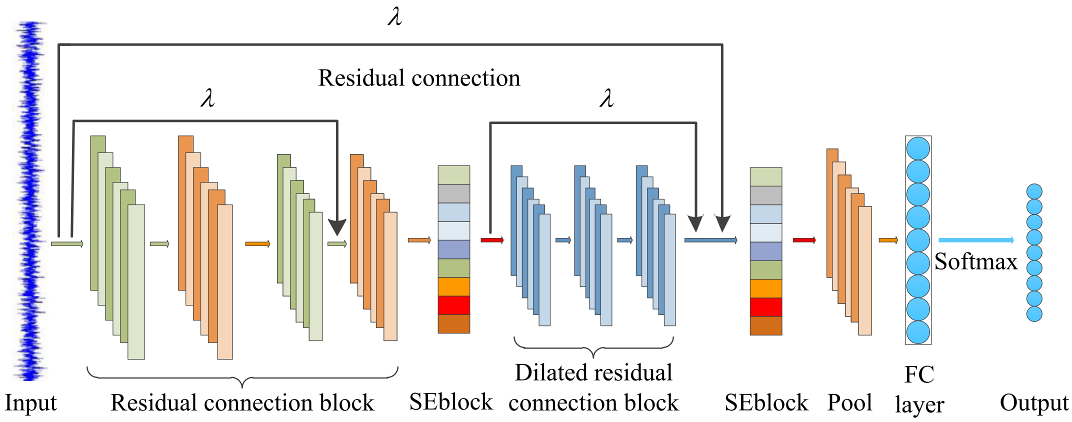


FIGURE 6. Network structure of the proposed method.

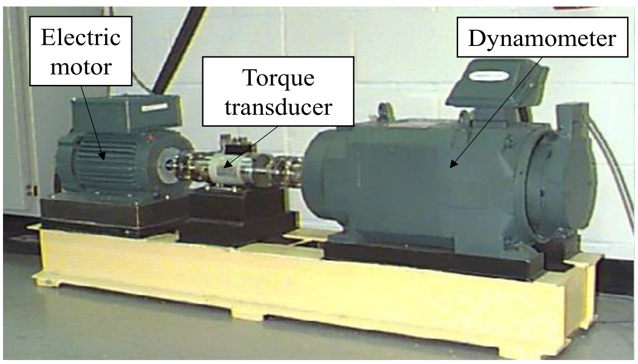


FIGURE 7. Rolling bearing test bench.

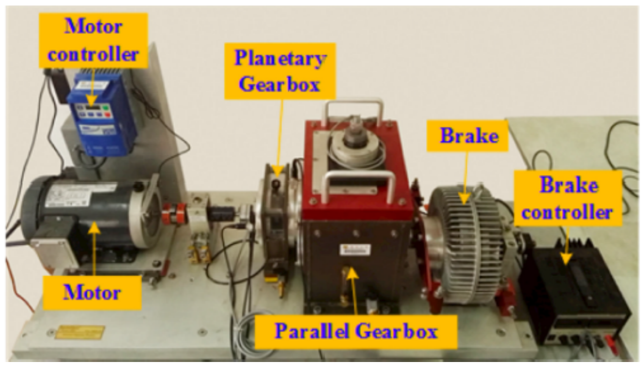


FIGURE 8. DDS test bench.

different weight coefficients through experiments, thus the best weight coefficient is obtained. Finally, through the full connection layer and softmax classifier, the fault diagnosis results are output.

IV. EXPERIMENTAL VERIFICATION

A. DATA DESCRIPTION

1) CWRU DATASET

The bearing dataset based on Case Western Reserve University (CWRU) [40] is used in the experiment. The dataset is widely used for bearing fault diagnosis experiments. The rolling bearing test bench is shown in Fig.7. From left to right, there are the electric motor, torque transducer and dynamometer. The bearing of the motor drive end SKF6205 is selected as the experimental bearing, and the fault is caused by the electric spark discharge method. The faults are located at the rolling body, inner ring, outer ring 3, 6 and 12 o'clock, respectively, and the fault diameters are 0.007, 0.014, 0.021 and 0.028in, so as to simulate pitting faults of different degrees of rolling bearing. According to the faults with different damage diameters in different positions, there are 16 kinds of vibration signals.

16 kinds of vibration signals of rolling bearing are detected by an acceleration sensor and collected by a data recorder. The sampling frequency of the data recorder is 12kHz and there are 16 sampling channels. For the vibration signals, a moving time window with a step length of 784 points is used to intercept them without overlap. The step length of the moving window determines the input length of the deep learning fault diagnosis method. Some scholars choose the step length of the moving window of 1024 or 2048, but the above two lengths are larger than 784, which would generate more calculations. In addition, a smaller step length would also get more data samples, so we choose 784 as the moving window step size. Dataset A, B and C represent the bearing data under three loads at the motor drive end. The three loads are 1, 2 and 3hp, respectively. The corresponding speeds of the electric shock drive end are 1772, 1750 and 1730r/min. Set the ratio of training samples to test samples be 5:1. The specific components of the bearing dataset are shown in Table 1.

2) SU DATASET

SU dataset of rolling bearing is from Southeast University (SU) in China. Drivetrain Dynamic Simulator (DDS) is shown in Fig.8. The test bench is mainly composed of motor, motor controller, planetary gearbox, parallel gearbox, brake and brake controller. The dataset includes bearing data and gearbox data. The speed-load condition is 20HZ-0V or 30HZ-0V. We use bearing data as the experimental dataset

TABLE 1. Description of rolling bearing dataset.

Fault location		Normal	Ball				Inner race				Load(hp)
Condition labels		1	2	3	4	5	6	7	8	9	
Fault diameter (in.)		0	0.007	0.014	0.021	0.028	0.007	0.014	0.021	0.028	
Dataset A	Train	514	128	129	129	129	130	129	129	129	1
	Test	103	26	26	26	26	26	26	26	26	
Dataset B	Train	514	129	129	129	129	128	129	129	129	2
	Test	103	26	26	26	26	26	26	26	26	
Dataset C	Train	515	129	130	129	129	130	129	129	129	3
	Test	103	26	26	26	26	26	26	26	26	

Fault location		Outer race 3 o'clock		Outer race 6 o'clock			Outer race 12 o'clock		Load(hp)
Condition labels		10	11	12	13	14	15	16	
Fault diameter (in.)		0.007	0.021	0.007	0.014	0.021	0.007	0.014	
Dataset A	Train	129	129	130	129	129	128	128	1
	Test	26	26	26	26	26	26	26	
Dataset B	Train	129	129	130	129	129	128	128	2
	Test	26	26	26	26	26	26	26	
Dataset C	Train	129	129	129	129	129	128	129	3
	Test	26	26	26	26	26	26	26	

and merge the data under the two conditions into a composite dataset. The composite dataset includes 5 condition labels, namely health, ball failure, inner failure, outer failure, combined inner and outer failure. 600 samples are selected for each type of failure, and there are 3000 samples in total. The ratio of training set and testing set is 4:1.

B. PARAMETERS OF THE PROPOSED METHOD

The structure of the proposed method is mainly composed of residual connection block and dilated residual connection block. The residual connection block consists of two normal convolution layers and the maximum pooling layer, in which the convolution Kernel size of the normal convolution layer is 64×1 and 7×1, the stride is 8×1 and 1×1, the number of Kernel channels is 16 and 32, and the kernel size and stride of the maximum pooling layer are both 2×1. The dilated residual connection block consists of three dilated convolution layers, in which the Kernel size of the dilated convolution layer is 3×1, the stride is 1×1 and the number of Kernel channels is 64. Detailed parameter descriptions are shown in Table 2. In addition, when the feature sizes and channels of the input and output of the residual block are inconsistent, the residual connection needs to use a 1×1 convolutional layer for matching. For example, the first sub-residual line uses a 1×1 convolution, the number of channels is 32, and the stride size is 16. The global residual connection line uses a 1×1 convolution, the number of channels is 64, and the stride size is 32.

The deep learning framework used in the simulation experiment is TensorFlow, the programming language is Python, the computer operating system is Ubuntu, the CPU is I9-9900K and the GPU is RTX2080Ti. During the experiment, the Adam optimization algorithm is used to update the network training parameters, the number of iteration batches is 2000, and the number of samples in each batch is 64. The network is trained with a dynamic learning rate, the initial learning rate is set to 0.001, the decay rate is 0.9, and the decay is performed every 1000 iterations. To prevent overfitting, the dropout abstention value is set to 0.5. The average of 10 experimental results is selected as the final result.

C. THE EFFECT OF DILATION RATE AND RESIDUAL CONNECTION WEIGHT COEFFICIENT

1) THE EFFECT OF DILATION RATE

Compared with normal convolution, dilated convolution has a larger receptive field. However, stack dilated convolution with the same dilation rate would lead to grid effect and affect the ability of network feature learning. Therefore, a zigzag dilated convolutional layer is constructed, that is, different combinations of dilation rates are set for three continuous dilated convolutional layers. Assumed that the combination of dilation rate is expressed as (A, B, C), according to Equation (7), there are the following four combinations (1, 2, 2), (1, 2, 3), (1, 2, 5) and (1, 2, 7). When both dilation rates are 1, dilated convolution is equivalent to normal convolution, so a group of normal convolutions is selected for experimental comparison, and its combination is (1,1,1). In order to more intuitively analyze the experimental results, the residual connection weight coefficient is set as 1, and the training samples in dataset A and testing samples in dataset B are selected to verify the influence of different dilation rate combinations. The experimental results are shown in Fig.9. It can be seen from Fig.9 that the fault diagnosis accuracy of the dilated convolution experiment group is higher than that of the normal convolution experiment group, which indicates that the dilated convolution can expand the receiving field and improve the generalization ability. In addition, under the premise that all four groups of dilation rate combinations satisfy Equation (2), the fault diagnosis accuracy of the (1,2,3) combination is 95.33%, which is higher than the results of the other three groups. This means that the (1,2,3) combination can maximize the fault diagnosis performance of the proposed method. Therefore, in the subsequent experiments, the combination of (1,2,3) is used as the optimal combination of zigzag dilated convolution.

TABLE 2. Parameters of the proposed method.

No.	Layer type	Kernel size/stride	Kernel channel number	Output size (Width × Depth)
1	Conv layer 1	64×1/8×1	16	98×16
2	Pool layer 1	2×1/2×1	16	49×16
3	Conv layer 2	7×1/1×1	32	49×32
4	Pool layer 2	2×1/2×1	32	25×32
5	Dilated Conv layer 1	3×1/1×1	64	25×64
6	Dilated Conv layer 2	3×1/1×1	64	25×64
7	Dilated Conv layer 3	3×1/1×1	64	25×64
8	Pool layer 3	2×1/2×1	64	13×64
9	Fully connected layer	100	1	100×1
10	Softmax	16	1	16

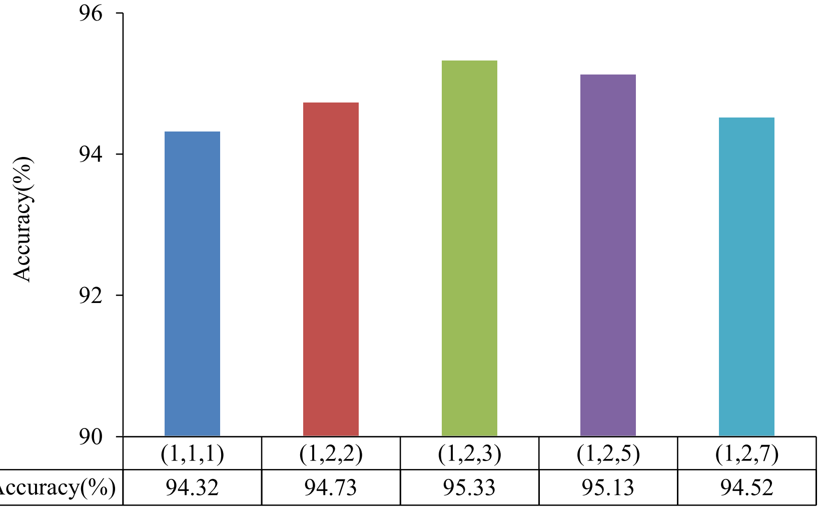


FIGURE 9. Diagnosis results using different dilation rate combinations.

2) THE EFFECT OF RESIDUAL CONNECTION WEIGHT COEFFICIENT

The residual connection can transfer feature information between the upper and lower layers of neural network. In order to make the residual block learn more feature information, the proposed method adds a weight coefficient of λ times to each residual connection line, and $\lambda = 0$ means that the method does not include residual connection. Furthermore, we select training samples in dataset A and the test samples in dataset B as experimental data. The influence of different residual connection weight coefficients is shown in Fig.10. It can be seen from Fig.10 (a) that the fault diagnosis accuracy of the first group is significantly lower than that of the other four groups, so this means that the residual connection can effectively improve the feature learning ability. In addition, it can be seen from the other four experiments that the fault diagnosis accuracy increases with the reduction of the residual connection weight. When $\lambda = 1$, the fault diagnosis accuracy is the lowest, which is 95.33%, and when $\lambda = 0.2$, the fault diagnosis accuracy is the highest, which is 96.74%. We can see that as the value of λ decreases, the accuracy is gradually increasing. In order to obtain the optimal parameter, the value of λ is respectively selected as 0.05, 0.1, 0.15, 0.2 and 0.3 for experiments, and the results are shown in Fig.10 (b). It can be seen that when λ is between 0.05 and 0.2, the accuracy increases gradually, and when λ is 0.3, the accuracy decreases. Therefore, we can draw the following conclusion that the effect of the proposed method can be improved by appropriately reducing the weight of residual connection, but when the weight is too low, residual connection plays a small role and does not improve the effect of the proposed method. In the residual connection weight coefficient experiment, the global residual connection and the sub-residual connection use the same weight coefficient, which can reduce the complexity of the network and facilitate the adjustment of network parameters. In addition, although the optimization algorithm can find the optimal parameter, it also increases the algorithm complexity of the proposed method. It can be seen from Fig.10 (b) that the accuracy difference between 0.15, 0.2 and 0.3 is small and is in an acceptable range, so we set $\lambda = 0.2$ as the optimal residual connection coefficient in our method.

Through the experiments in this section, we confirm the optimal dilation rate and residual connection weight coefficient. The function of the dilation rate is to increase the receptive field of the convolutional layer and avoid the grid effect. The residual connection weight coefficient can reduce the feature weight of the residual connection, so these two parameters are not directly related. In addition, these two parameters are simultaneously optimized in the experiment,

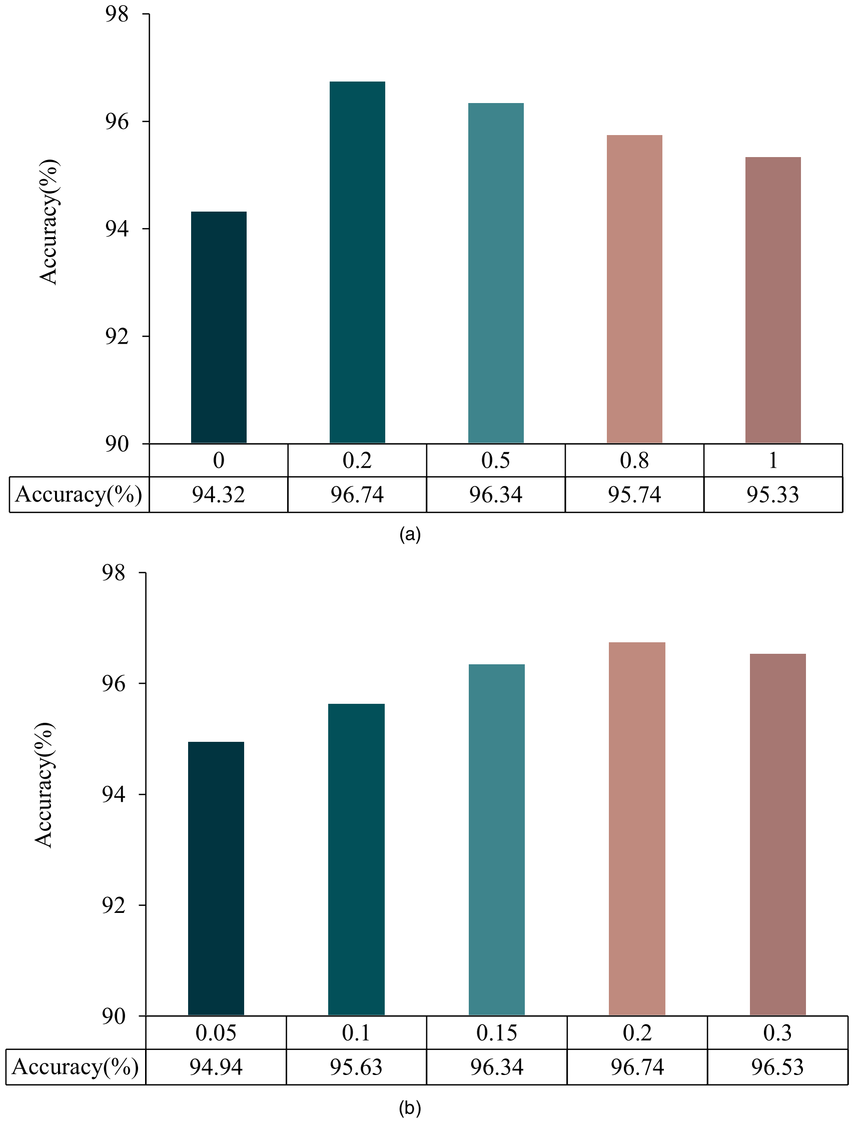


FIGURE 10. Diagnosis results using different residual connection weight coefficients.

the effect of optimizing these two parameters at the same time is the same as optimizing the dilation rate first, and then optimizing the residual connection weight coefficient. Therefore, the adopted optimization method is to first assume that the residual connection weight coefficient is 1, the dilation rate is analyzed, and after the optimal dilation rate is determined, the optimal weight coefficient is analyzed and determined. For other parameters of the proposed method, according to a large number of references, the approximate ranges of the parameters are determined, so the parameters are determined by using the enumeration method.

D. PERFORMANCE UNDER DIFFERENT EXPERIMENT ENVIRONMENTS

1) CONTRAST METHODS

In order to verify the effectiveness of the proposed method, we select SVM-EMD envelope spectrum and BPNN-EMD envelope spectrum methods based on artificial filtering [41], LeNet-5 [13], ResNet [42] and WDCNN [23] methods based on deep learning as comparison methods. Among them, the first 5 IMFS Hilbert envelope spectra decomposed by EMD algorithm are used as the input of the methods based on artificial filtering. SVM uses the "one-to-many" classification method, and the kernel function adopts the "Gaussian kernel function", BPNN uses 3920 dimensions and 3920-300-16 structures. Based on the deep learning method, the inputs of LeNet-5 and ResNet are the two-dimensional gray map after the transformation of one-dimensional bearing data by matrix. LeNet-5 adopts the classical structure, the convolution kernel size is 5×5, the pooling step size is 2, and the parameter of the full connection layer is 120. The network structure of ResNet is composed of 5 residual blocks, two pooling layers and a full connection layer. In the residual block, the convolution kernel size is 3×3, the pooling step

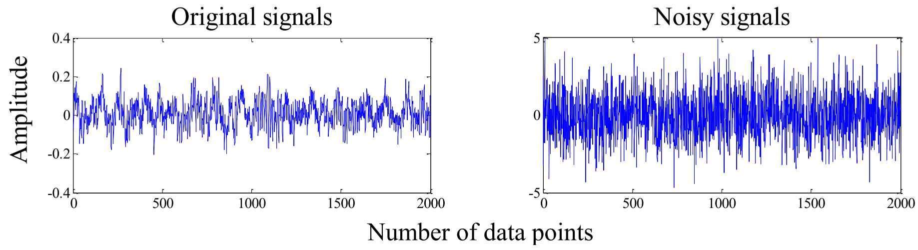


FIGURE 11. Original signals and Noisy signals.

size is 2, and the parameter of the full connection layer is 1024. The input of WDCNN is one-dimensional bearing data. The network structure is composed of 5 convolutional layers and 5 maximum pooling layers, and BN is added after each convolution layer. The size of the convolution kernel of the first layer is 64×1, the size of the convolution kernel of the last four layers is 3×1, the pooling step is 2, and the number of parameters of the full connection layer is 100. Among the above methods, SVM-EMD and BPNN-EMD represent traditional machine learning fault diagnosis methods. LeNet-5 represents the classic convolutional neural network. ResNet is generated by constructing residual connections in convolutional neural networks, and it represents an important development of convolutional neural network. WDCNN represents a rolling bearing fault diagnosis method based on one-dimensional convolutional neural network. So we consider that these methods chosen as the comparison methods in this paper are convincing.

2) PERFORMANCE UNDER DIFFERENT NOISY ENVIRONMENTS

In the working process of rolling bearing, the vibration and friction of parts would produce noise, which not only affects the health state of bearing, but also contaminates the collected vibration data and covers the fault information in vibration data. Therefore, it is required that the fault diagnosis method of rolling bearing can overcome the interference of noise. In this paper, Gaussian white noise with different SNR is added to the original signals to construct different composite noise signals, so as to simulate different noise environments, which show in Fig.11, the left side is the original signals, and the right side is the noisy signals, which is Gaussian white noise with the SNR of 1dB added to the original signals.

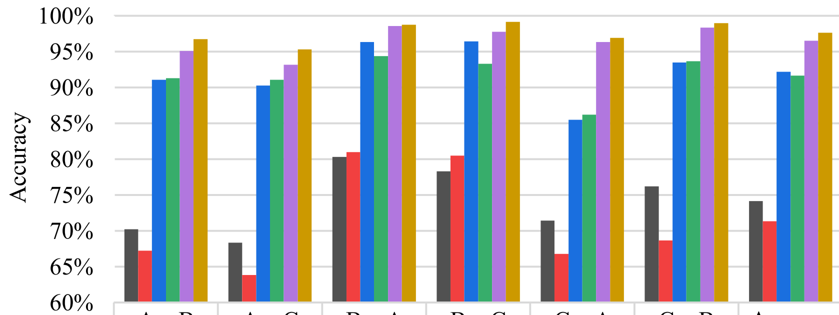
The definition of SNR is as follows:

$$\mathrm{SNR_{dB}} = 10\log_{10}(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}) \tag{8}$$

where P_{signal} is represents signal power, and P_{noise} is represents noise power. Among them, the lower the SNR is, the more complex the composite noise signals are. The experimental data for fault diagnosis in different noise environments select the training samples and test samples in dataset A, and add Gaussian white noise with SNR of -1, 0, 1, 3, 5, 7, 9, 11dB to the test samples, thereby constructing 6 kinds of compound noise test samples. These samples are used to test the noise immunity of each method. The experimental results are shown in Table 3. It can be seen from Table 3 that the accuracies of the proposed method in the six noise environments are higher than other methods, and all have reached more than 82%. The two methods, SVM-EMD envelope spectrum and BPNN-EMD envelope spectrum based on artificial filtering, both have the problem of EMD's end effect, which limits the ability of these two methods to process noise data. Based on the methods of deep learning LeNet-5 and ResNet, the accuracy is more than 98% in the environment with the SNR of 9, 11 dB, but in the environment with the SNR of -1, 0, 1, 3 dB, the accuracy is lower than 90%, because these two methods use two-dimensional bearing data as input, they are difficult to effectively extract feature information from the original data of bearing, so they cannot effectively process strong noise data. Although the fault diagnosis accuracy of WDCNN method is relatively high, when the SNR is 3 dB, the accuracy drops significantly, when the SNR is 1 dB, the accuracy is only 91.68%, which shows that the anti-noise capability of this method has certain limitations. When the SNR is 1 dB, the accuracy of the proposed method is 93.30%, and when the SNR is 7, 9, 11 dB, the accuracies reach more than 99%, and the average accuracy is 94.99%, which is significantly higher than other methods. Because the proposed method constructs the wide convolution kernel residual connection which can maximize to extraction the feature information of the noise data, the zigzag dilated residual connection block can deeply learn the effective feature information. At the same time, we insert SE block after each residual block to enhance the feature recognition ability, so that our method has better anti-noise ability in different noise environments. In addition, when the SNR is 0, -1dB, the accuracies of the six methods are all lower than 90%. Although the accuracy of the proposed method is relatively high, the highest accuracy is only 89.25%. It can be seen that the proposed method has a strong anti-noise ability, but the performance would be limited when the noise is strong. Therefore, when the SNR range is 3-11dB, it is the reliable SNR range for the four deep learning methods, when the SNR range is 1-11dB, it is the reliable SNR range for the proposed method.

3) PERFORMANCE ACROSS DIFFERENT LOAD DOMAINS

The loading domain of rolling bearing would inevitably change in the actual working condition, so that the fault diagnosis method must have a good generalization ability. In order



	A→B	A→C	B→A	B→C	C→A	C→B	Average
SVM-EMD	70.23%	68.34%	80.32%	78.34%	71.45%	76.21%	74.15%
BPNN-EMD	67.23%	63.83%	81.02%	80.50%	66.80%	68.67%	71.34%
LeNet-5	91.08%	90.27%	96.35%	96.43%	85.52%	93.51%	92.19%
ResNet	91.32%	91.07%	94.40%	93.33%	86.24%	93.70%	91.68%
WDCNN	95.13%	93.17%	98.58%	97.77%	96.34%	98.37%	96.56%
Our method	96.74%	95.34%	98.78%	99.19%	96.96%	98.99%	97.67%

FIGURE 12. Fault diagnosis results of 6 methods in different load domains.

TABLE 3. The accuracy of six methods under different noisy environments.

Algorithms	Accuracy (%)								
	-1 dB	0 dB	1 dB	3 dB	5 dB	7 dB	9 dB	11 dB	Average
SVM-EMD	74.85	78.34	80.35	81.32	82.46	82.88	84.67	85.80	81.32
BPNN-EMD	70.23	70.76	72.15	73.43	75.32	76.50	76.58	79.20	74.26
LeNet-5	76.85	81.10	83.85	89.12	92.75	96.36	98.27	98.55	89.60
ResNet	77.76	81.56	84.26	89.56	93.95	97.76	98.85	99.13	90.66
WDCNN	81.18	88.38	91.68	94.93	98.18	98.38	98.78	98.98	93.80
Our method	82.50	89.25	93.30	97.65	98.98	99.18	99.49	99.59	94.99

to verify the generalization ability of our method under different load domain conditions, experiments are carried out under the conditions of different load domain, and a comparative analysis is carried out with the five selected fault diagnosis methods. The load domain change means that in the 3 different load domain data of dataset A, B, C, one of the load domain data is used as the training sample, and the other two load domain data are used as the test samples. In the experiment, A→B and A→C represent the training set of dataset A as the training sample, the test set of dataset B and C as the test sample for two groups of experiments, respectively, and other groups of experiments are the same way. The fault diagnosis results of 6 methods in different load domains are shown in Fig.12. It can be seen that in each group of different load domain experiments, the accuracy of the proposed method is higher than the other five methods. Taken C→A and C→B as examples: Due to envelope errors and modal confusion, fault diagnosis accuracy in different load domains is below 80% based on artificial filtering SVM-EMD envelope spectrum and BPNN-EMD envelope spectrum. The LeNet-5 and ResNet methods based on deep learning use two-dimensional data as input, because the original one-dimensional data is converted into two-dimensional data, there could be loss of feature information, resulting in the highest accuracies of the different load domains of the two methods are only 93.51% and 93.70%, which is the experimental result of the C→B group. Although the accuracies of the two groups of experiments based on WDCNN reach 96.34% and 98.37%, due to its relatively simple structure, the data features could not be fully extracted, resulting in the failure diagnosis effect of different load domains is not as good as the proposed method. The fault diagnosis accuracies of the proposed method in C→A and C→B group experiments reach more than 96%, in addition, the average accuracy of all groups of experiments is 97.67%. This is because the global residual and sub-residual connections increase the learning process, and then we set the residual connection weight coefficients to make the transfer of features more efficient. Moreover, the zigzag dilated convolution makes the perception range of the dilated residual connection block wider, so as to deeply dig out the internal feature information in the data of different load domains, which make the proposed method have advantages in the aspect of feature learning. Therefore, although the features of test samples and training samples in different load domains are quite different, this method can still well adapt to the negative effects of data feature differences and has better generalization ability.

4) PERFORMANCE UNDER NOISE AND DIFFERENT LOAD DOMAINS ENVIRONMENTS

Since the working environment of rolling bearing is often accompanied by noise and load domain changes, we add

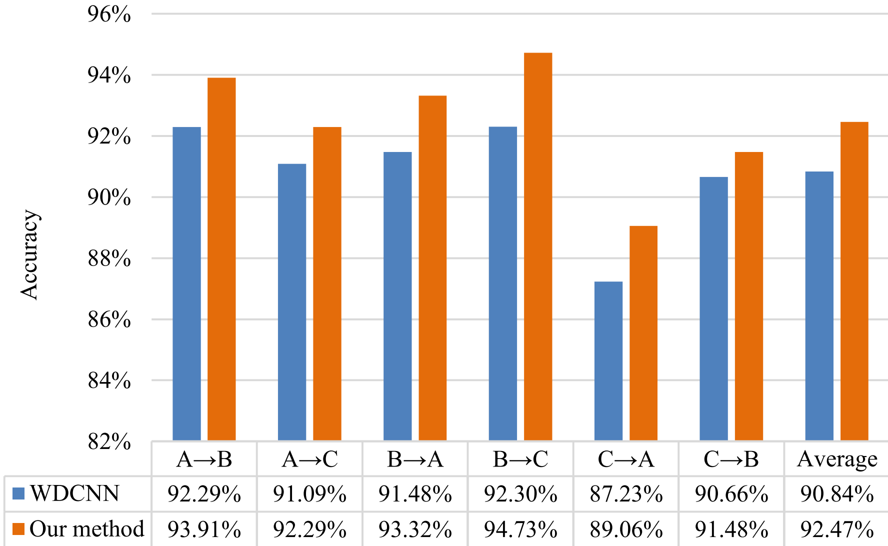


FIGURE 13. Fault diagnosis results under noise and different load domains environments.

TABLE 4. Calculation time of the four methods.

Method	LeNet-5	ResNet	WDCNN	Our method
Time (s)	179	193	108	120

Gaussian white noise with an SNR of 3dB in the test set and set the load domain change. It is used to verify that the proposed method has both anti-noise and generalization ability. Compared with WDCNN method, the experimental results are shown in Fig.13. The highest fault diagnosis accuracy of WDCNN method is 92.30%, the lowest accuracy is 87.23%, and the average accuracy is 90.84%. The highest fault diagnosis accuracy of the proposed method is 94.73%, the lowest accuracy is 89.06%, and the average accuracy is 92.47%. In addition, the accuracy of the proposed method in each experiment is higher than that of WDCNN method. It can be seen that the proposed method has better fault diagnosis results than WDCNN, and can still maintain better noise resistance and generalization ability in complex working conditions.

5) CALCULATION TIME

The calculation times of the four deep learning methods are shown in Table 4. It can be seen that WDCNN has the least calculation time. The calculation time of the proposed method is higher than that of WDCNN, but lower than LeNet-5 and ResNet, because the proposed method uses more convolutional layers than WDCNN, and the structure is more complex. In the future, the proposed method should be optimized to reduce the calculation time.

6) CROSS-VALIDATION SCHEME

The cross-validation algorithm of machine learning is an evaluation algorithm used to verify model parameters and evaluate model classifier. Since the ratio of training set and test set is 5:1, we divide dataset A into 6 parts, and take turns to use 5 parts as training set and 1 part as test set. After

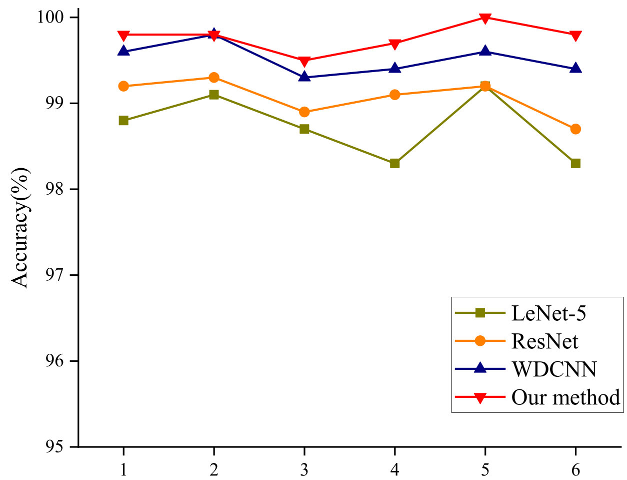


FIGURE 14. Cross-validation results of the four deep learning methods.

TABLE 5. Error of the four deep learning methods.

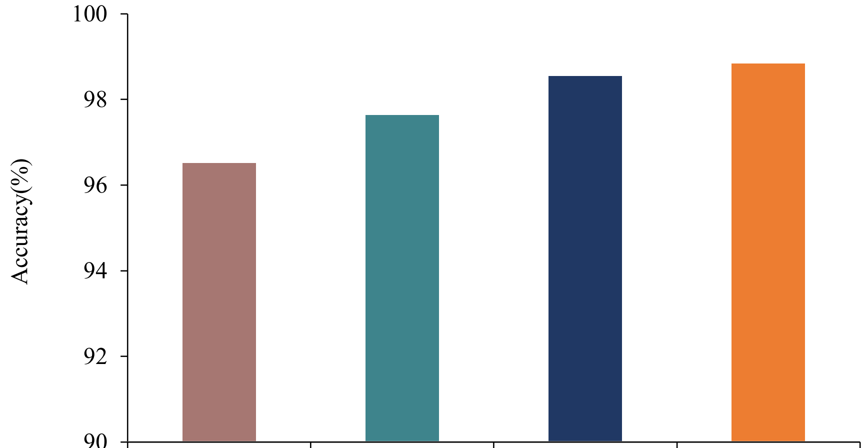
Method	LeNet-5	ResNet	WDCNN	Our method
Error	±0.43%	±0.37%	±0.28%	±0.23%

repeated verification and testing, the performance evaluation results of the model are finally obtained.

The cross-validation results of the four deep learning methods are shown in Fig.14, and the error is shown in Table 5. It can be seen from the figure and table that the accuracy of WDCNN and the proposed method reached 99.3%, the classification accuracy of the proposed method is higher, and the error is smaller, only about 0.23%. Therefore, the proposed method has the advantages of high classification accuracy, stable calculation results, small errors, and no over-fitting or under-fitting.

E. FAULT DIAGNOSIS PERFORMANCE UNDER THE SU DATASET

We conduct a fault diagnosis experiment under the SU dataset and compare with LeNet-5, ResNet, and WDCNN. The



	LeNet-5	ResNet	WDCNN	Our method
Accuracy(%)	96.51	97.63	98.55	98.83

FIGURE 15. Fault diagnosis results under the SU dataset.

results are shown in Fig.15. From Fig.15, we can see that the fault classification accuracy of the proposed method is significantly higher than the other three methods, the classification accuracy of WDCNN is 98.55%, and the classification accuracy of the proposed method is 98.83%, which is 0.28% higher than the former. The experimental result verifies the effectiveness of the zigzag dilated residual block and residual weight coefficient structure. Therefore, the proposed method can effectively learn the feature information in the data, has better generalization ability, and has excellent fault diagnosis effect for different datasets. Both the CWRU data set and the SU data set are collected through the rolling bearing test bench. The characteristics of the two data sets are similar. In addition, we found in experiments that $\lambda = 0.2$ is also the optimal value in the SU data set.

V. CONCLUSION

In order to improve the anti-noise performance and generalization ability of fault diagnosis method of rolling bearing in different load domains and noise environments, a one-dimensional dilated convolution network with residual connection method is proposed. In this paper, we construct a zigzag dilated convolution, which has a larger receptive field and can obtain more feature information in data samples. The fault features are transmitted in the upper and lower layers of the network through global residual connection and sub-residual connection, which can effectively avoid the risk of overfitting. After each sub-residual structure is obtained, the SE block module is inserted to adaptively select beneficial features, so that the proposed method can efficiently obtain the feature information in data. CWRU and SU datasets are used for the simulation experiments. In the experiment of the noise environment, the average fault diagnosis accuracy of the proposed method is higher than other methods, which verifies that the proposed method has satisfactory anti-noise capability. In experiments of different load domains, the fault diagnosis accuracy of the proposed method in each group experiment is higher than that of WDCNN method, which verifies that the proposed method has better generalization performance. In the noise and different load domain environment experiment, the proposed method has a higher fault diagnosis accuracy than that of WDCNN method, which verifies the superiority of the proposed method. In the future, to further improve diagnostic performance, how to determine the optimal parameters will be studied. The experimental data set used in our current work is manually processed data set. The data features of bearings in real work are more complex, so the proposed method has limitations for bearing failures in real work.

REFERENCES

[1] J. Jiao, M. Zhao, J. Lin, and C. Ding, "Deep coupled dense convolutional network with complementary data for intelligent fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9858–9867, Dec. 2019.

[2] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.

[3] P. Chen, X. Zhao, and Q. Zhu, "A novel classification method based on ICGOA-KELM for fault diagnosis of rolling bearing," *Int. J. Speech Technol.*, vol. 50, no. 9, pp. 2833–2847, Sep. 2020.

[4] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, Nov. 2018.

[5] L. Lu, J. Yan, and C. W. de Silva, "Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition," *J. Sound Vibrat.*, vol. 344, pp. 464–483, May 2015.

[6] K. Zhang, J. S. Cheng, and Y. Yang, "Roller bearing fault diagnosis based on local mean decomposition and morphological fractal dimension," *J. Vibrat. Shock*, vol. 32, no. 9, pp. 90–94, Sep. 2013.

[7] W. Mao, L. He, Y. Yan, and J. Wang, "Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine," *Mech. Syst. Signal Process.*, vol. 83, pp. 450–473, Jan. 2017.

[8] Z. Zhu, G. Peng, Y. Chen, and H. Gao, "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis," *Neurocomputing*, vol. 323, pp. 62–75, Jan. 2019.

[9] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017.

[10] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.

[11] Z.-X. Hu, Y. Wang, M.-F. Ge, and J. Liu, "Data-driven fault diagnosis method based on compressed sensing and improved multiscale network," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3216–3225, Apr. 2020.

[12] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. ICML*, Jan. 2011, pp. 833–840.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[14] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[15] J. L. Elman, "Finding structure in time," *Cognit. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[18] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, and Aaai, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI*, 2017, pp. 1–12.

[19] Y. Zhang, K. Xing, R. Bai, D. Sun, and Z. Meng, "An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image," *Measurement*, vol. 157, Jun. 2020, Art. no. 107667.

[20] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196–3207, Apr. 2019.

[21] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mech. Syst. Signal Process.*, vol. 133, Nov. 2019, Art. no. 106272.

[22] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *J. Sound Vibrat.*, vol. 388, pp. 154–170, Feb. 2017.

[23] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, pp. 1–21, Feb. 2017.

[24] X. Li, J. Li, C. Zhao, Y. Qu, and D. He, "Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1D separable convolution with residual connection," *Mech. Syst. Signal Process.*, vol. 142, Aug. 2020, Art. no. 106740.

[25] X. Li, W. Zhang, and Q. Ding, "Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism," *Signal Process.*, vol. 161, pp. 136–154, Aug. 2019.

[26] Y. Xue, D. Dou, and J. Yang, "Multi-fault diagnosis of rotating machinery based on deep convolution neural network and support vector machine," *Measurement*, vol. 156, May 2020, Art. no. 107571.

[27] T. Chen, Z. Wang, X. Yang, and K. Jiang, "A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals," *Measurement*, vol. 148, Dec. 2019, Art. no. 106857.

[28] B. Wang, Y. Lei, N. Li, and T. Yan, "Deep separable convolutional network for remaining useful life prediction of machinery," *Mech. Syst. Signal Process.*, vol. 134, Dec. 2019, Art. no. 106330.

[29] S. Hao, F.-X. Ge, Y. Li, and J. Jiang, "Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks," *Measurement*, vol. 159, Jul. 2020, Art. no. 107802.

[30] W. Zhang, X. Li, and Q. Ding, "Deep residual learning-based fault diagnosis method for rotating machinery," *ISA Trans.*, vol. 95, pp. 295–305, Dec. 2019.

[31] 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.*, vols. 72–73, pp. 303–315, May 2016.

[32] Z. Xiang, X. Zhang, W. Zhang, and X. Xia, "Fault diagnosis of rolling bearing under fluctuating speed and variable load based on TCO spectrum and stacking auto-encoder," *Measurement*, vol. 138, pp. 162–174, May 2019.

[33] D. Peng, H. Wang, Z. Liu, W. Zhang, M. J. Zuo, and J. Chen, "Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4949–4960, Jul. 2020.

[34] J. Jiao, M. Zhao, J. Lin, and K. Liang, "Residual joint adaptation adversarial network for intelligent transfer fault diagnosis," *Mech. Syst. Signal Process.*, vol. 145, Nov. 2020, Art. no. 106962.

[35] Z. Xu, C. Li, and Y. Yang, "Fault diagnosis of rolling bearings using an improved multi-scale convolutional neural network with feature attention mechanism," *ISA Trans.*, to be published, doi: 10.1016/j.isatra.2020.10.054.

[36] K. Wang, W. Zhao, A. Xu, P. Zeng, and S. Yang, "One-dimensional multi-scale domain adaptive network for bearing-fault diagnosis under varying working conditions," *Sensors*, vol. 20, no. 11, pp. 6039–6056, Nov. 2020.

[37] S. Dong, K. He, and B. Tang, "The fault diagnosis method of rolling bearing under variable working conditions based on deep transfer learning," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 42, no. 11, pp. 585–601, Oct. 2020.

[38] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. ICLR*, 2016, pp. 1–13.

[39] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[40] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.

[41] X. Zhang, Y. Liang, J. Zhou, and Y. Zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, Jun. 2015.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 630–645.

HAOPENG LIANG received the B.Sc. degree from the Lanzhou University of Technology, in 2018, where he is currently pursuing the master's degree. His main research interests include deep learning and fault diagnosis.

XIAOQIANG ZHAO received the Ph.D. degree from Zhejiang University, in 2006. He is currently a Professor and a Ph.D. Supervisor with the Lanzhou University of Technology. His main research interests include process monitoring and fault diagnosis, production scheduling, and data mining.