Journal of Electronics and Information Science (2024)
Clausius Scientific Press, Canada
DOI: 10.23977/jeis.2024.090309
ISSN 2371-9524 Vol. 9 Num. 3

Multi-scale Adaptive Fusion for Rolling Bearing Fault Diagnosis

Ke Xu[1,a], Yongyong Hui[1,2,b,*]

[1]*College of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou, 730050, China*
[2]*National Experimental Teaching Center of Electrical and Control Engineering, Lanzhou University of Technology, Lanzhou, 730050, China*
[a]*2087601670@qq.com*, [b]*huiyy@lut.edu.cn*
[*]*Corresponding author*

Keywords: Rolling bearing, Fault diagnosis, Self-correcting convolution, Convolutional neural network, Attention mechanism

Abstract: In order to fully extract the bearing fault feature information under strong noise and variable load, a rolling bearing fault diagnosis method based on multi-scale adaptive fusion (MSAF) is proposed. Firstly, a multi-scale feature extraction module is designed, which uses convolutional layers of different scales to extract feature information, in order to better capture the characteristics of different fault signals. Secondly, a Self-Calibrated Convolution (SCC) module is constructed. This module automatically adjusts the weights of the convolutional kernels according to the characteristics of the input data, which enhances the network's perception of the input data. Thirdly, a lightweight channel attention residual module is established, which combines channel attention and residual connections, allowing the network to automatically select channels related to fault features, thereby reducing information redundancy. Finally, the Softmax probability distribution function is used as a classifier to achieve bearing fault classification. By using the bearing data set of CWRU for experiment and comparison, it is verified that the method still has strong fault diagnosis performance under variable load and variable noise.

1. Introduction

As a key component of rotating machinery equipment, rolling bearings are very vulnerable to vibration, impact, erosion and wear during operation and work in a complex environment with strong noise and variable load for a long time, resulting in frequent bearing failures [1]. Therefore, it is necessary to diagnose the fault of rolling bearings and identify different fault states [2]. At first, machine learning played an important role in the field of bearing fault diagnosis, including Support Vector Machine (SVM) [3], Random Forest (RF) [4] and Logistic Regression (LR) [5]. Xu et al. [6] combined empirical mode decomposition (EMD) and support vector machine (SVM) for fault diagnosis. Zhang et al. [7] used functional data analysis to extract the autocorrelation function fitting coefficients of bearing vibration signals, and used the random forest algorithm to diagnose and identify the features after dimensionality reduction. The algorithm constructs the fault feature set and optimizes the random forest parameters through the grid search method to obtain the importance ranking of features. Although the above method has a certain effect in fault diagnosis, the extracted features are almost shallow features, the ability to extract complex data features is poor, and the knowledge and experience of domain experts are needed.

In recent years, deep learning, as a relatively new and rapidly developing method, was proposed by Hinton [8] in 2006 and has been widely used in bearing fault diagnosis. It includes Convolutional Neural Network (CNN) [9], Long Short-Term Memory (LSTM) [10] and Deep Belief Network (DBN) [11]. In order to fully extract features and suppress the influence of high-frequency noise on fault diagnosis, Zhang et al. [12] proposed a deep convolutional neural network with a wide convolution kernel, which used wide convolution kernel for feature extraction and suppressed high-frequency noise. Qu et al. [13] proposed an adaptive one-dimensional convolutional neural network algorithm for bearing fault diagnosis to maximisze feature self-learning. All of the above methods are applied to CNNs, but they often suffer from training parameter complexity, which can lead to overfitting problems. In addition, in order to further obtain higher fault recognition accuracy under variable load, Tang et al. [14] used the two-dimensional time-frequency spectrum of vibration time series and adaptively extracted different fault features through CNN for fault diagnosis, which can also have higher fault diagnosis accuracy under variable load. Ye et al. [15] used empirical mode decomposition to process the extracted vibration signal and convert it into an image, and then used the convolution layer to extract image features to achieve fault diagnosis. Liang et al. [16] proposed a parallel convolutional neural network (P-CNN) fault diagnosis method, which has good stability in variable load environment. However, the above method does not consider the influence of environmental noise on the model diagnosis effect, thus reducing its accuracy and generalization ability in fault diagnosis.

Aiming at the above problems, a rolling bearing fault diagnosis method based on multi-scale adaptive fusion is proposed. Firstly, in order to extract the effective information of fault data to the greatest extent, a multi-scale feature extraction module is designed. Secondly, the SCC module is constructed to improve the network's perception of input data. Thirdly, the ECA-ResNet module is constructed, which combines channel attention and residual connection to adaptively enhance effective information. Finally, the Softmax function is used as a classifier to realize bearing fault classification.

2. Fault diagnosis model structure based on MSAF

The working environment of rolling bearings is relatively complex. In fault diagnosis, insufficient feature information extraction under strong noise and variable load will lead to low fault diagnosis accuracy and poor generalization performance.

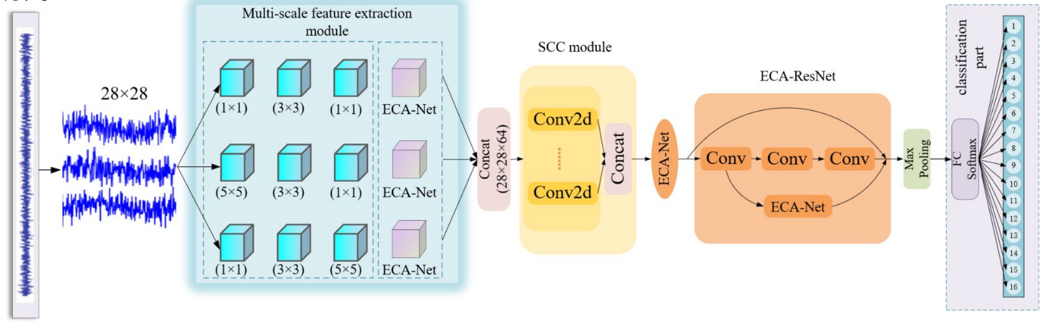
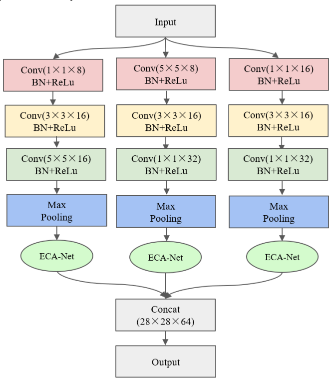


Figure 1: Total network structure model.

In this paper, the data preprocessing is carried out first, and the reconstructed two-dimensional data is used as the model input. Secondly, the local features of fault information are extracted by using the constructed multi-scale convolutional neural network. At the same time, the SCC module and ECA-ResNet module are designed to adaptively enhance effective information, improve the network's perception of input data, and suppress interference information. Finally, the Softmax probability distribution function is used as a classifier to realize bearing fault classification. The structure is shown in Figure 1.

2.1. Feature extraction module

The convolution kernel of traditional CNN extracts features in the local receptive field and then integrates the local features, which leads to the loss of some important information in the network [17]. The designed feature extraction module shows excellent feature extraction ability in bearing fault diagnosis, and its structure is shown in Figure 2.



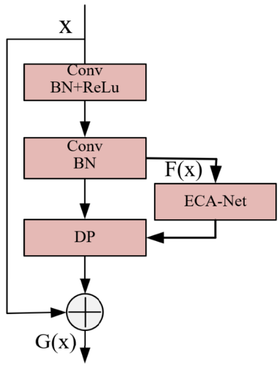


Figure 2: Structure diagram of multi-scale feature extraction. Figure 3: ECA_ResNet structure.

Firstly, for the input two-dimensional data, three convolutional layers of different scales are used for feature extraction. Secondly, in order to improve the diagnostic performance of the network, the output of each convolutional layer is batch standardized. Finally, by introducing the attention mechanism, the model's attention to key information is enhanced. The design of this feature extraction module has the potential to improve the accuracy and efficiency of bearing fault diagnosis.

2.2. ECA-ResNet module

The residual structure allows the training of deep neural networks[18], which makes it easier to train deep networks to capture and represent complex features, avoids the problem of gradient disappearance[19], and improves the performance of the model. ECA_ ResNet combines residual structure and channel attention mechanism to improve the performance of CNN, and its structure is shown in Figure 3. Firstly, the output F(x) is obtained by convolution kernels with sizes of 1×1 and 3×3, respectively. Then, the weights obtained by F(x) and ECA_Net are dot product, and the final result is added to the input x. The advantage of this structure is that it can better capture the correlation information between features, effectively suppress redundant information interference, and improve classification accuracy.

2.3. Self-calibrated convolution network structure

Self-calibrated convolution (SCC) introduces adaptive parameters to learn the feature representation and transformation of input data [20]. Traditional convolution uses fixed-size convolution kernels to extract features when processing input data. However, this fixed-size convolution kernel may not be able to adapt to different input data and has certain limitations on the invariance of the input data. Self-calibrated convolution allows the network to automatically adjust the weight of the convolution kernel according to the characteristics of the input data, thereby improving the network's perception and adaptability to the input data. This adaptability enables the network to better adapt to different input data and improve its generalization ability.

The structure is shown in Figure 4. SCC divides the input X evenly into two parts, and divides the convolution kernel K into four equal parts. The specific operation is as follows: Firstly, the convolution operation K_1 is performed on the input feature X_1 to generate the output feature Y_1. At the same time, the input feature X_2 is fed into channels with different resolutions, in which the low-resolution channel reduces the feature width in a certain proportion, and after average pooling downsampling, a low-dimensional embedding for correcting the high-resolution partial convolution kernel is generated. In order to improve the ability to extract fault features, the model performs feature extraction on the low-resolution scale and combines the feature information of the low-resolution channel and the high-resolution channel. Secondly, the convolution and upsampling operations are performed. The output calculated by the Sigmoid function is corrected with the features extracted by the K_3 convolution. The convolution operation K_4 is performed on the corrected features to obtain the output feature Y_2 of the self-correction part. The specific operation is shown in Equation (1) ~ (4).

$$T_1 = Avgpool_r(X_1) \tag{1}$$

$$X_1^{'} = Up(T_1 * K_2) \tag{2}$$

$$Y_1^{'} = X_1 * K_3 \bullet \sigma(X_1 + X_1^{'}) \tag{3}$$

$$Y_1 = Y_1^{'} * K_4 \tag{4}$$

Where r represents the sampling rate, Up(·) denotes upsampling, σ denotes the Sigmoid activation function. Finally, the output features from two scale spaces are fused to obtain the output feature Y. SCC establishes long-distance spatial and channel dependencies around each spatial location through self-correction, thereby generating more discriminative features.

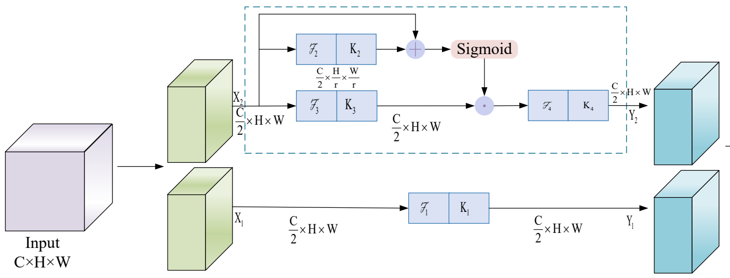


Figure 4: Self-correcting convolutional structure.

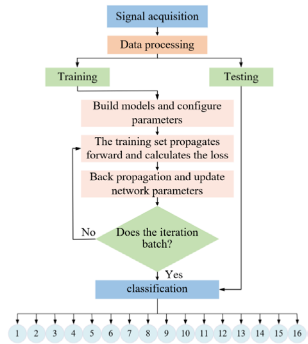


Figure 5: Diagnostic flow chart

2.4. Model diagnosis process

The diagnosis process of the bearing fault diagnosis model based on MSAF is shown in Figure 5.

As shown in Figure 5, it is mainly divided into data preprocessing, model training and fault diagnosis. Different types of bearing fault vibration signals and bearing vibration signal data under normal working conditions are collected by sensors or monitoring equipment. The vibration signals are divided into single samples and assigned labels to indicate the specific categories. The labeled samples are divided into training set and test set according to the proportion, and the bearing fault diagnosis model is constructed and the parameters are configured. The model parameters are optimized by backpropagation. If the iterative batch is not reached, the training is continued, otherwise the fault classification is performed.

3. Experimental verification and analysis

In order to verify the effectiveness of the proposed method, the bearing data set of Case Western Reserve University (CWRU) [21] is used for verification. The data set is divided into training set and test set. In the training process, the initial learning rate is set to 0.001, and the network is trained by exponential learning rate descent method. In order to prevent overfitting, the size of Dropout is set to 0.5 and the attenuation rate is 0.9. The software environment of this experiment is TensorFlow 1.13.1 version of PyCharm 2022.2.2, and the hardware environment is Intel (R) Core (TM) i9-12900H @ 2.50GHz.

3.1. Data set description

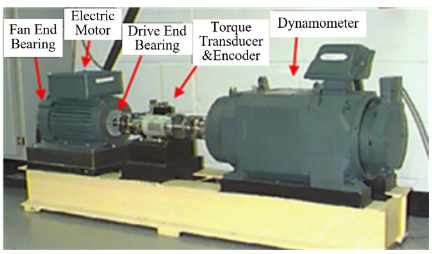
The bearing test bench of CWRU is shown in Figure 6, which is composed of a motor, torque sensor and power meter. The bearing model used in the experiment is SKF6205, and the sampling frequency is 12 kHz. The speed of the motor is 1797 r / min, 1772 r / min, 1750 r / min and 1730 r / min, respectively, and the corresponding loads are 0 hp, 1 hp, 2 hp and 3 hp, respectively. The data is collected by the acceleration sensor, and then marked as data sets A, B, C and D. The data set description is shown in Table 1. In order to better verify the diagnostic performance of the proposed method, variable noise, variable operating conditions and variable load experiments are carried out, and the proposed method is compared with the other five methods. The comparison methods include Inception Residual Block (IRB) [22], ResNet18 [23], LeNet-5 [24], Alexnet [25], and 1DCNN [26].

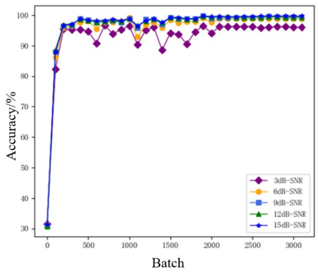
Table 1: Description of the data set

Data set	Rotating speed/ (r/ min)	Load/ (hp)
A	1797	0
B	1772	1
C	1750	2
D	1730	3

3.2. Variable noise fault diagnosis results and analysis

In order to test the noise immunity of the proposed method, a variable noise interference experiment was carried out. Gaussian white noise with different signal-to-noise ratios is added to the original vibration signal to simulate the actual noise environment.





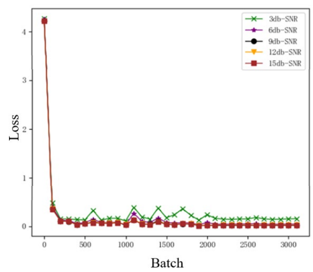


Figure 6: CWRU experimental platform Figure 7: Diagnostic accuracy Figure 8: Loss outcome map

The data with a load of 1hp are selected for experiments, and Gaussian white noises with signal-to-noise ratios of 3dB, 6dB, 9dB and 12dB are added to the test set to test the diagnostic ability of the proposed method. The diagnostic accuracy and loss values are shown in Figure 7 and Figure 8.

It can be seen from Figure 9 that the accuracy of the proposed method is higher than that of other methods. The average accuracy of the IRB method is lower than that of other methods. Although the IRB method improves the traditional residual structure and increases the number of convolution layers, the network cannot work well because it does not introduce the attention mechanism and has too many structural feature parameters. The structures of LeNet-5 and Alexnet methods are relatively simple, and the extracted shallow feature information cannot fully reflect the running state of the bearing, resulting in a poor diagnosis effect. ResNet introduces residual connections, which make it possible to effectively retain the feature information extracted from the shallow layer while extracting the deep feature information. However, when the signal-to-noise ratio is 3 dB, the accuracy of fault diagnosis can only reach 82.46 %, and the fault cannot be accurately classified. The accuracy of the 1DCNN method is only 84.41 % in the case of 3 dB strong noise, and the accuracy is low. When the signal-to-noise ratio is 3 dB, the fault accuracy of the proposed method can also reach 95.72 %, which further verifies that it has strong anti-noise performance.

3.3. Variable load fault diagnosis results and analysis

In the actual industrial process, the load of the bearing often needs to change, which requires the model to have strong generalization performance and strong fault diagnosis effects under different loads. The data with loads of 0Hp, 1Hp, 2Hp and 3Hp are selected for experiments. One data set is used as a training sample, while the other three data sets are used as test samples. 0-1, 0-2, 0-3 means that the 0Hp data set is used as the training set, while the 1Hp, 2Hp, and 3Hp data sets are used as the test sets respectively, and other numbers are used as the training set. The proposed method is compared with other methods, and the comparison results are shown in Figure 10.

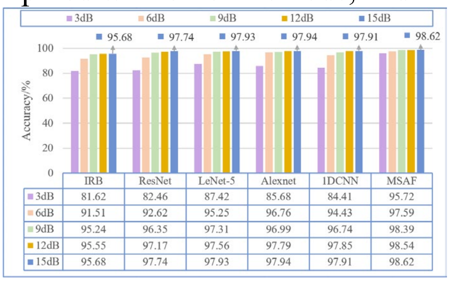


Figure 9: Results of variable noise

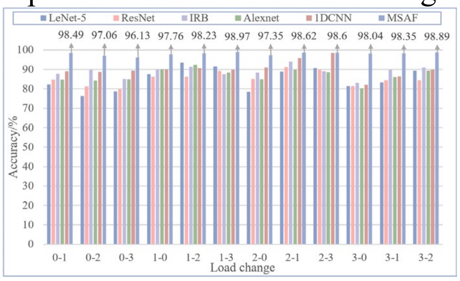
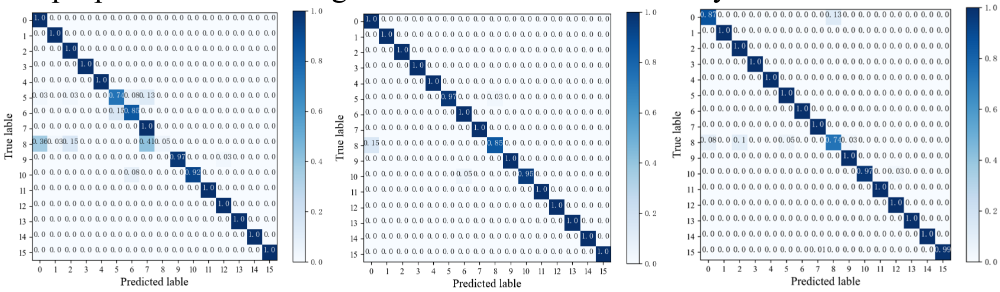


Figure 10: Variable load results

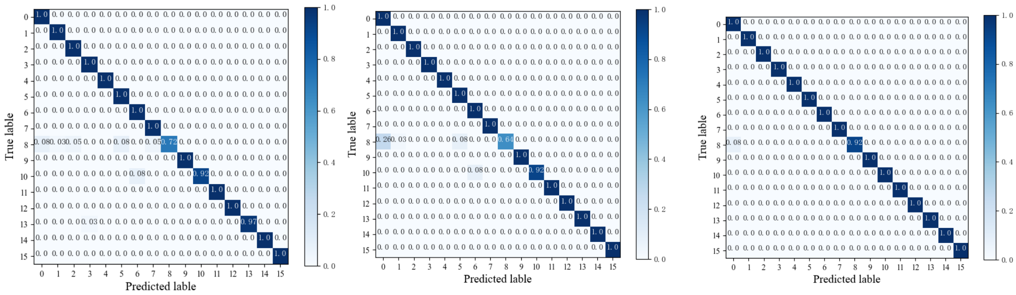
It can be seen from Fig.10 that the results of the proposed method are higher than those of the comparison method in the variable load experiment. The average accuracy of the proposed method can reach 98 %. The experimental results show that the proposed method has better generalization performance under variable load conditions. This is because under variable load conditions, the vibration signal of the bearing will change in different frequency and amplitude ranges, and the multi-scale feature extraction module enables the model to adapt to these changes and better extract useful features. In addition, the SCC module helps to improve the robustness of features. It can suppress noise and irrelevant features under both strong noise conditions and different loads, thereby improving the generalization performance of the model.

3.4. Confusion matrix results and analysis

In order to further verify the fault classification performance of the proposed method, the confusion matrix experiment is carried out on the test results. When the load is 1 HP and the signal-to-noise ratio is 3 dB, the confusion matrix experiment results are shown in Figure 11. It can be seen from Figure 11 that the comparison methods have different diagnostic errors, but the method proposed in this paper only has a deviation in the eighth type of fault. Therefore, this experiment shows that the proposed method has good fault classification ability.



(a) IRB method (b) LeNet-5 method (c) ResNet method



(d) Alexnet method (e) 1DCNN method (f) In proposed method

Figure 11: Comparison of confusion matrix

4. Conclusions

In order to improve the accuracy and generalization performance of fault diagnosis under strong noise conditions, a rolling bearing fault diagnosis method based on MSAF is proposed. Firstly, a multi-scale feature extraction module is designed, which uses multi-channel convolution layers of different scales to extract features from input data. The main purpose of the design is to ensure that the effective information in the fault data is extracted to the maximum extent. Secondly, the SCC module is introduced, which can automatically adjust the weight of the convolution kernel according to the characteristics of the input data, thereby improving the network's perception of the input data. At the same time, the ECA-ResNet module is constructed, which combines channel attention and residual connection to adaptively enhance effective information and suppress interference information. Finally, the Softmax probability distribution function is used as a classifier to realize bearing fault classification. Experiments on CWRU datasets verify that the proposed method has high fault accuracy in a strong noise environment, and has good anti-noise and generalization performance.

References

[1] Ni Q, Ji J C, Halkon B, et al. Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics[J]. Mechanical Systems and Signal Processing, 2023, 200: 110544.

[2] Khan M A, Asad B, Kudelina K, et al. The Bearing Faults Detection Methods for Electrical Machines—The State of the Art [J]. Energies, 2022, 16(1): 296.

[3] Fu W, Jiang X, Li B, et al. Rolling bearing fault diagnosis based on 2D time-frequency images and data augmentation technique [J]. Measurement Science and Technology, 2023, 34(4): 045005.

[4] Liu W, Zhang Z, Zhang J, et al. A novel fault diagnosis method of rolling bearings combining convolutional neural network and transformer [J]. Electronics, 2023, 12(8): 1838.

[5] Zhou Jianmin, Li Hui, Zhang Long, et al. Evaluation of bearing performance degradation based on EMD and logistic regression[J]. Mechanical Design and Research, 2016, 32(5): 72-75.
[6] Xu Ke, Chen Zonghai, Zhang Chenbin, et al. Fault Diagnosis of Rolling Bearings Based on Empirical Mode Decomposition and Support Vector Machine[J]. Control Theory & Applications, 2019, 36(6): 915-922.
[7] Zhang Xining, Zhang Wenwen, Zhou Rongtong, et al. Bearing Fault Diagnosis Method Based on Multi-dimensional Scale and Random Forest[J]. Journal of Xi'an Jiaotong University, 2019, 53(8): 1-7.
[8] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786): 504-507.
[9] Jia F, Lei Y, Lu N, et al. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization[J]. Mechanical Systems and Signal Processing, 2018, 110: 349-367.
[10] Zhou Xiangyu, Mao Shanjun, Li Mei. Bearing Fault Diagnosis Method Based on Frequency Domain Downsampling and CNN[J]. Journal of Peking University (Natural Science), 2023, 59(2): 251-260.
[11] Gan M, Wang C. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings[J]. Mechanical Systems and Signal Processing, 2016, 72: 92-104.
[12] Zhang W, Peng G, Li C, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals[J]. Sensors, 2017, 17(2): 425.
[13] Qu Jianling, Yu Lu, Yuan Tao, et al. Adaptive Fault Diagnosis Algorithm for Rolling Bearings Based on One-dimensional Convolutional Neural Network[J]. Chinese Journal of Scientific Instrument, 2018, 39(7): 134-143.
[14] Tang Bo, Chen Shenshen. Bearing fault diagnosis method based on deep convolutional neural network[J]. Journal of Electronic Measurement and Instrumentation, 2020, 32(3): 88-93.
[15] YE Zhuang, YU Jianbo. Gearbox Vibration Signal Feature Extraction Based on Multi-channel Weighted Convolutional Neural Network[J]. Journal of Mechanical Engineering, 2021, 57(1): 110-120.
[16] Liang M, Cao P, Tang J. Rolling bearing fault diagnosis based on feature fusion with parallel convolutional neural network[J]. The International Journal of Advanced Manufacturing Technology, 2021, 112: 819-831.
[17] LI Yandong, HAO Zongbo, LEI Hang. A Review of Convolutional Neural Networks[J]. Journal of Computer Applications, 2016, 36(9): 2508-2515.
[18] Kang Tao, Duan Rongkai, Yang Lei. Bearing Fault Diagnosis Method of Convolutional Neural Network Integrating Multi-attention Mechanism[J]. Journal of Xi'an Jiaotong University, 2022, 56(12): 68-77.
[19] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.
[20] Liu J J, Hou Q, Cheng M M, et al. Improving convolutional networks with self-calibrated convolutions[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10096-10105.
[21] Smith W A, Randall R B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study[J]. Mechanical systems and signal processing, 2015, 64: 100-131.
[22] Zhao Xiaoqiang, Liang Haopeng. Fault Diagnosis Method for Variable Working Conditions of Rolling Bearings Based on Improved Residual Neural Network [J]. Journal of Xi'an Jiaotong University, 2020, 54(9). 23-31.
[23] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016: 630-645.
[24] Zhao Xiaoqiang, Luo Weilan.Bearing fault diagnosis method for improved convolutional Lenet-5 neural network [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(06):113-125.
[25] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
[26] Abdeljaber O, Avci O, Kiranyaz S, et al. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks[J]. Journal of Sound and Vibration, 2017, 388: 154-170.