A Multi-Domain Data-Based Attentive Residual Network for Bearing Fault Diagnosis

Leqi Zhu*, Zhiqiang Lin[a], Liang Li[b]

School of Management, HeFei University of Technology, HeFei, China.

*zhuleqi2002@163.com

[a]zhiqiang__lin@163.com

[b]2020213464@mail.hfut.edu.cn

Abstract. Methods of data-driven fault diagnosis are of great significance to ensure the stability and reliability of bearings systems. However, the existing methods still encounter many challenges. From the perspective of data, a single domain signal cannot fully reveal complex industrial processes. From the perspective of the model, abstract features at different levels contain fault information with various importance, which affects the model performance. In this paper, an adaptive residual network based on multi-domain data is proposed to make up for the shortcomings of existing fault diagnosis methods. Firstly, FFT frequency domain analysis is conducted on time domain signals, and multi-domain data are constructed together. Secondly, an adaptive attention mechanism is introduced into the residual block based on one-dimensional convolution to fuse shallow and deep features, so as to extract features more effectively. Finally, experiments on rolling bearings at Case Western Reserve University manifest that the proposed method is superior to other comparative methods in fault diagnosis.

Keywords: Data-driven, fault diagnosis, bearing systems.

1. Introduction

In modern industrial manufacturing systems, as an integral part of rotating machinery, rolling bearings are vital to mechanical transmission, components coupling, and rotating efficiency improvement. However, as for common faults such as wear, indentation, and gap increase, it is necessary to diagnose and maintain faults in time to avoid unnecessary losses [1]. In recent years, effective and reliable methods of data-driven fault diagnosis have been widely concerned.

Fault diagnosis combining signal analysis technology with deep learning method is the mainstream research at present. Common signal analysis methods include the analysis of the time domain, frequency domain, and time-frequency domain [2]. Through the time domain analysis such as mean value, maximum value, and root mean square of vibration amplitude, frequency domain analysis such as Fast Fourier Transform (FFT), or time-frequency analysis such as Wavelet Packet Transform (WPT) and Short Time Fourier Transform (STFT), the original vibration signal is usually processed and then input into the model for fault diagnosis [3].

As a deep learning model, CNN can perceive more representative local fault information in signals and realize more efficient fault classification by extracting detailed features. Initially, in the application of fault diagnosis based on the deep learning method, one-dimensional signals are transformed into two-dimensional ones to adapt to the two-dimensional model structure and extract features. For example, Zhao et al. designed a depth residual network to extract features from WPT time-frequency domain images using a dynamic weighting strategy [4]. Jin et al. used STFT for time domain signals and input it into the proposed deep learning network to realize compound fault diagnosis [5]. However, on the one hand, the structure of two-dimensional CNN is relatively complex, which improves the model accuracy at the cost of its calculation time. On the other hand, rough signal processing easily leads to irreversible loss of key fault information [6].

With the development of research, one-dimensional CNN has been proposed to directly process one-dimensional signals. Thanks to its lightweight model structure and high computational efficiency, it has been widely used in fault diagnosis. For example, Chen et al. proposed improved CNN and LSTM to directly process one-dimensional time-domain signals and realized efficient bearing fault diagnosis by mining spatiotemporal features in signals [7]. Junior et al. designed a multi-head 1DCNN structure, which used a multi-channel structure to extract the signal features of multi-source time-domain vibration [8].

To sum up, the method based on deep learning has been applied to industrial fault diagnosis. However, the existing methods still face the following problems. For one, there are multi-angle signal representations for the machinery operation, and the model data input in a single domain such as the time domain or frequency domain lacks complementary information between different signal representations. For the other, although residual learning has been widely adopted to solve the network degradation caused by the deepening of model layers, the simple superposition of shallow and deep features tends to ignore key information and trigger information redundancy.

To solve the above problems, this paper proposes a attentive residual network based on homologous multi-domain data (MDARN). Firstly, the original time domain signal is analyzed by FFT to obtain the signal with frequency domain characteristics, and the more robust data is constructed. Then, an adaptive residual fusion block based on one-dimensional CNN is developed, which comprehensively considers the macro positioning of the shallow layer and the fault details of the deep layer. In addition, experiments on the CWRU data set show that the proposed method achieves high-precision fault diagnosis.

The main contributions of this paper are as follows:

(1) A fault diagnosis framework based on multi-domain data is proposed, which uses one-dimensional CNN to extract features from time domain and frequency domain signals simultaneously, considering the complementary information from different perspectives.

(2) A residual method of fusing relationship-aware attention is designed so that the network can be adaptive to select and fuse important parts of shallow and deep features in the feature extraction.

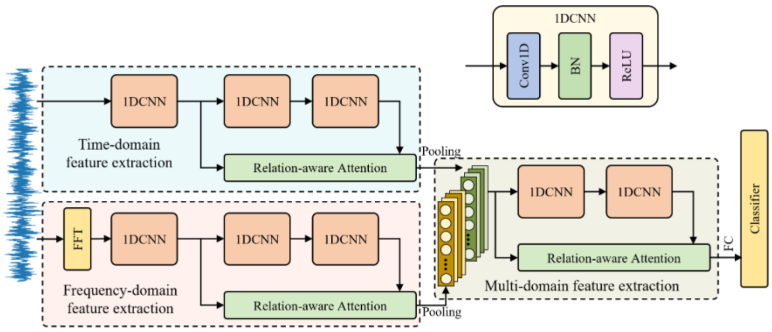


Figure 1 Overall Model Diagram

2. Proposed Methods

Figure 1 shows the overall model framework of MDARN proposed in this paper. The model is mainly composed of 1DCNN and adaptive residual blocks, whose core mainly includes time domain signal, frequency domain signal, and the fault information learning of mixed multi-domain deep features. The specific details of the model will be introduced in following sections.

2.1 Construction of Homologous Multi-Domain Data

Sensors installed on the bearings system capture its vibration according to a fixed frequency and obtain a large number of time series vibration signals for subsequent deep learning modeling through data transmission and database storage.

The time domain signal reflects the distribution of bearing vibration amplitude in the time dimension. However, industrial processes are usually complex, nonlinear, and non-stationary, thus data mining with single-domain representation is limited for fault prediction. Frequency domain analysis can map time domain signals to other angular representations, which can reveal how many signals are in a frequency band. FFT is a simple and fast frequency domain analysis used in this paper to obtain frequency domain signals.

In this paper, time-domain and frequency-domain signals are used as model inputs simultaneously to construct homologous multi-domain data, which supplements and enriches the original vibration signals from a more comprehensive perspective. Assuming there are T instances of $\{(x_t, y_t)\}_{t=1}^{T}$, $x_t \in \mathbb{R}^P$ refers to one-dimensional original vibration signals collected in the t period with P data points and y_t represents the health status label corresponding to the signal. The specific process of homologous multi-domain data construction is:

$$Z_t = Parallel\{x_t, FFT(x_t)\} \tag{1}$$

Parallel represents parallel combination; the one-dimensional time and frequency domain signals converted by FFT are directly used as the subsequent model input based on one-dimensional CNN.

2.2 One-Dimensional Convolution Neural Network

One-dimensional CNN uses a one-dimensional convolution kernel to slide on sequence data at a fixed step size to capture local receptive fields, which can be used to directly process one-dimensional signal input in the time domain and frequency domain. Meanwhile, CNN characterized by parallel computing can complete the parallel feature extraction of multi-domain data simultaneously. One-dimensional CNN refines abstract features at higher levels through the alternate operation of convolution and pooling. The mathematical calculation formula of convolution is:

$$y_j^{l+1} = f(\sum x_i^l \times k_{ij}^l + b_j^l) \tag{2}$$

x_i^l is the l-layer network input with i as the original parameter quantity; k_{ij}^l and b_j^l represent the parameter weight and error offset of the convolution kernel respectively. f is the activation function and ReLU is adopted in this paper. Finally, y_j^{l+1} is output with j as the parameter, which can be used as the input of the next layer of network. In addition, the number of channels output by the network is the same as that of convolution kernels.

2.3 Attention Residual Fusion Module

A residual learning strategy can more effectively solve the gradient disappearance or gradient explosion that may occur in the model training, which is widely used in fault diagnosis [9]. As shown in the upper left of Figure 2, after the shallow feature short circuit jumps to two-parameter layers in the traditional residual block, the output containing both macro fault positioning and local fault details is obtained. However, the multi-level features are blindly added, which makes the model effect vulnerable to redundant information. In recent years, the attention mechanism of plug for immediate usage has been introduced into the existing research.

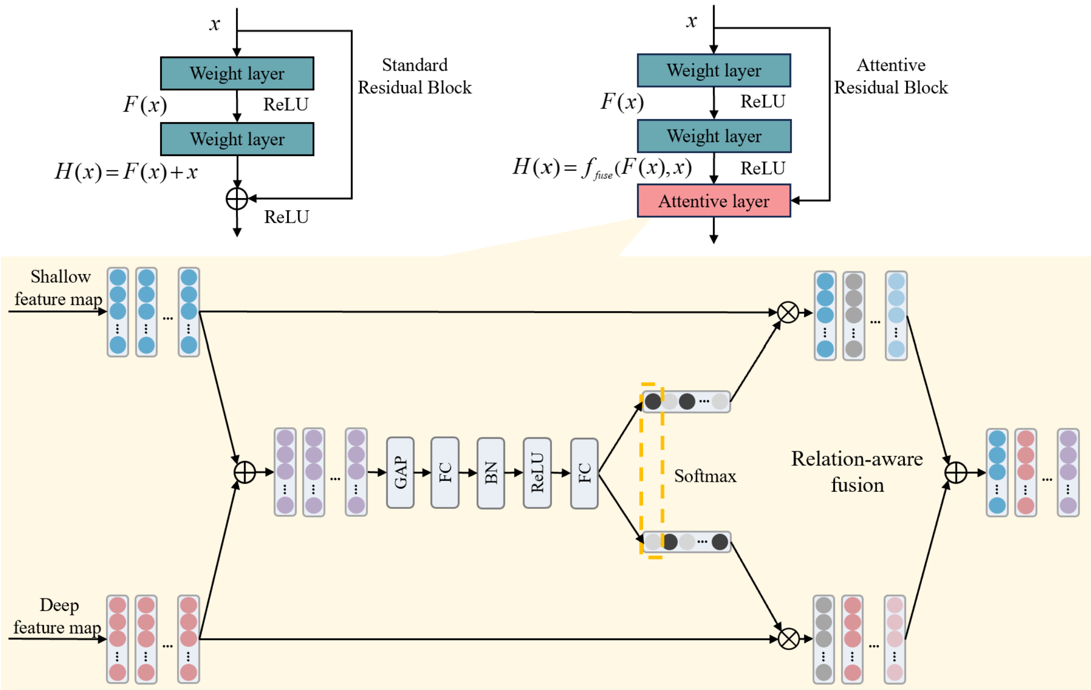


Figure 2 Residual Block and Diagram of Relation-Aware Attention Structure

The attention mechanism is inspired by human visual attention selection, which makes the model pay more attention to abstract features useful for final prediction in the training and gives them greater weights [6]. Thus, the model layers are deepened with increasing abstract features when avoiding the influence of redundant information on the model's robustness and generalization. Besides, channel attention such as SE-net, SK-net, etc. is common, which are used to emphasize the information about channel angle in feature mapping [10], [11]. According to Figure 2, this paper designs an attention mechanism that considers the correlation between abstract features at different levels, instead of skipping in the traditional residual process. Attention is input into shallow and deep features of $M^s, M^d \in \mathbb{R}^{N\times L}$ simultaneously, and channel representation based on common features $W \in \mathbb{R}^{N\times 1}$ is obtained by binary addition and average pooling. Then, the shared full connection layer is squeezed and the independent full connection layer is expanded. Then, the shared full connection layer is squeezed and the independent full connection layer is expanded. softmax function learns the importance weights $W^s, W^d \in \mathbb{R}^{N\times 1}$ of feature channels at different levels to model prediction and gives them to the original input. Finally, adaptive residual fusion is realized by binary addition. The mathematical calculation of Relation-aware attention is:

$$W = f_{GAP}(M^s + M^d) = \frac{1}{L}\sum_{l=1}^{L}(M^s + M^d)(l) \tag{3}$$

$$W^s, W^d = f_{soft\max}(w_1, w_2) = f_{soft\max}(\sigma_1(W), \sigma_2(W)) \tag{4}$$

$$F^{fuse} = M^s \times W^s + M^d \times W^d \tag{5}$$

f_{GAP} and $f_{soft\max}$ represent the channel-based global average pooling and softmax processing respectively. σ_1 and σ_2 represent the full connection layer with corresponding weights of output and input features. Finally, the $F^{fuse} \in \mathbb{R}^{N\times L}$ after adaptive feature selection and fusion is output.

3. Experiment and Result Analysis

To prove the effectiveness of the proposed method, experiments are conducted on the open data set of rolling bearings provided by CWRU [12].

3.1 Bearing Data Set

The test rig of CWRU Bearing Data Center shown in Figure 3 simulates the real operation of the bearing system, which collects vibration signal data under different working conditions. In this paper, four sub-datasets are used by setting the motor speed and motor load to various values at the sampling frequency of 12 kH$_Z$. Their specific information can be seen in Table 1. Each sample has its corresponding health status label, such as normal state, outer ring fault (OF), inner ring fault (IF), and ball fault (B), with each at varied fault degrees. In other words, fault diagnosis based on the CWRU data set is a problem about ten classifications.

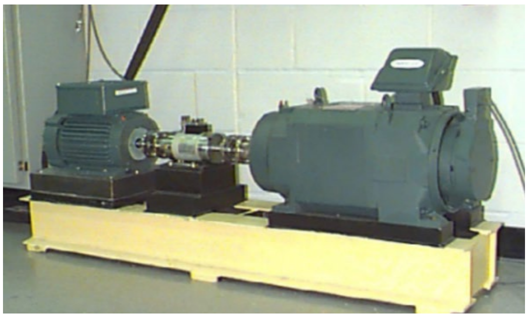


Figure 3 The Test-Rig

In each sub-data set, this paper extracts the data of every three sampling of the signal, and each signal sample contains 2048 data. In addition, each original time domain signal is transformed by FFT, and the frequency domain signal containing 2048 data is also obtained. The time domain and frequency domain signals are combined into homologous multi-domain data for experiments.

Table 1 CWRU Experimental Data Set

Datasets	Motor Speed（rpm）	Motor Load（hp）	Numbers of Instances	Description
CWRUA	1730	3	1200	Normal, B007, B014, B021, IR007, IR014, IR021, OR007, OR014, OR021
CWRUB	1750	2	1200	
CWRUC	1772	1	1200	
CWRUD	1797	0	1200	

3.2 Experimental Design

To prove the effectiveness and superiority of the proposed method, comparative experiments are conducted under four sub-datasets. The methods used in the experiment include MDARN, deep learning methods based on time domain signals such as WDCNN [13], ResCNN [9], one-dimensional SE-net-based CNN (SE-CNN) premised on channel attention, and one-dimensional FFT-based CNN (FFT-CNN) based on frequency domain signals. The evaluation index used in the experiment is the accuracy commonly used in classification problems.

The experiment uses a random number seed to divide the data set according to the 75% training set and 25% test set, repeating the experiment 20 times to reduce the error caused by random factors. Meanwhile, the early stop mechanism is introduced with its value set to 10 to avoid over-fitting in the model training. In addition, the validity and stability of fault classification are evaluated by calculating the mean and standard deviation of many experiments' accuracy.

3.3 Analysis of Experimental Results

Table 2 Experimental Results Under Four Data Sets

Accuracy (%)	CWRUA	CWRUB	CWRUC	CWRUD

Methods	Mean	Std	Mean	Std	Mean	Std	Mean	Std
WDCNN	97.77	0.99	98.80	0.93	98.52	0.85	98.32	0.64
FFT-CNN	94.38	1.49	95.35	1.52	95.70	1.54	93.60	1.69
ResCNN	98.70	0.66	99.43	0.58	99.28	0.59	99.17	0.87
SE-CNN	98.52	0.82	99.03	0.56	99.07	0.63	98.93	0.66
Proposed	99.23	0.54	99.77	0.26	99.67	0.30	99.55	0.41

The accuracy means and standard deviation of the proposed method and the comparative method are counted and listed as shown in Table 2. Firstly, as for single-mode input data, the method based on the time domain signal is more effective than that based on the frequency domain signal. For example, the accuracy of WDCNN is 97.77%, higher than that of FFT-CNN in the CWRUA data set, which proves that the original vibration signal can provide more abundant fault information than the converted frequency domain signal. Secondly, ResCNN and SE-CNN perform better than WDCNN. For example, the accuracy of ResCNN and SE-CNN under the CWRUB data set exceeds 99.00%, while WDCNN is only 98.80%. This proves that residual learning and attention mechanism is vital to promote the model's robustness and generalization. Finally, the proposed method has the highest fault classification accuracy under four data sets, which are 99.23%, 99.77%, 99.67%, and 99.55% respectively. Meanwhile, the standard deviation is lower than other comparison methods, which manifests that adopting multi-domain input data has more complementary information. Besides, a feature extraction strategy using parallel and adaptive residuals can significantly improve the prediction performance of the model with higher stability.

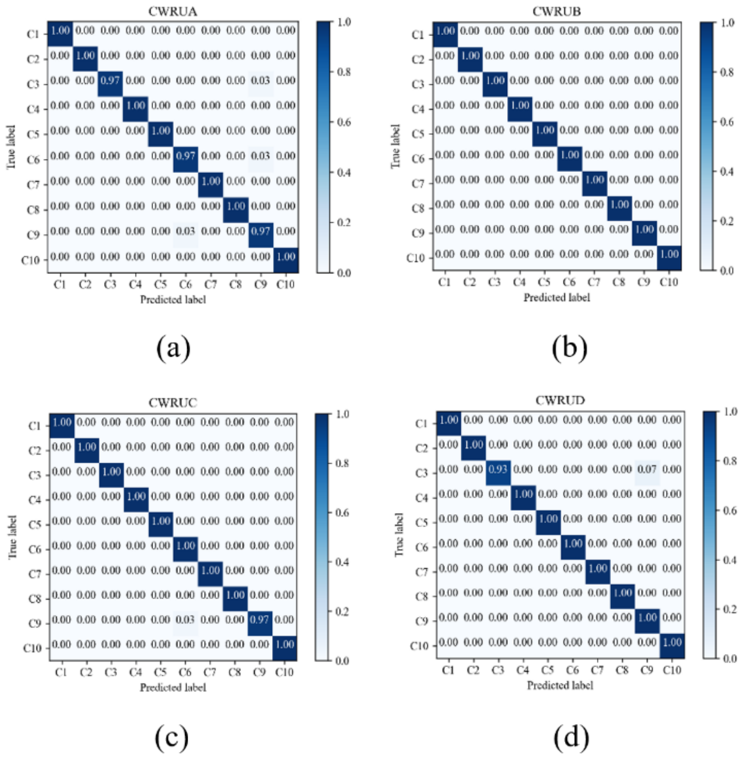
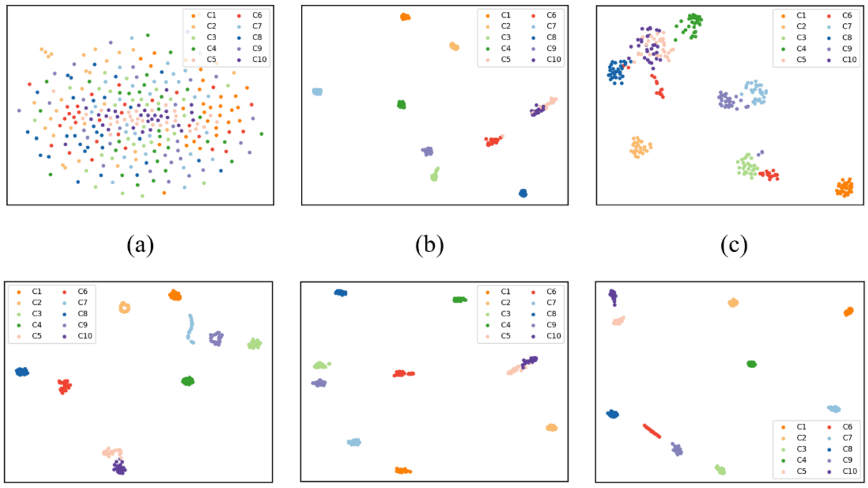


Figure 4 Confusion Matrix of the Proposed Method Under Four Data Sets

To observe the classification performance of MDARN, a confusion matrix is used to present the prediction of each health category. Figure 4 shows the classification effect of the proposed method under four data sets, and the rows and columns in each graph represent samples' real labels and the predicted labels of model output respectively. It can be seen that the proposed MDARN has a remarkable fault diagnosis effect. For example, the accuracy rate on oblique diagonal lines in CWRUB data sets has reached 100% with rare state classification errors in other data sets such as C3 and C9, manifesting that MDARN has achieved a high-level fault diagnosis.



(a) (b) (c)

(d) (e) (f)

Figure 5 Visualization of Feature Extraction.
(a)Original Vibration Signal. (b) WDCNN. (c) FFT-CNN.
(d) ResCNN. (e) SE-CNN. (f) MDARN.

To compare the extraction ability of models' abstract features, t-NSE is used to map high-dimensional features to low-dimensional planes to visualize different methods' intuitively. According to Figure 5 consisting of the original vibration signal, the four comparison methods, the proposed method, and the feature t-NSE scatter plot before the classifier, ten kinds of health states are scattered irregularly to the same aggregation. In Figure (c), different clusters are roughly divided by the single frequency domain data, but 5, 6, and 10 are still mixed together. In figures (b), (d), and (e), single-time domain data with good classification effects can clearly distinguish most health states, but 5 and 10 cannot be well distinguished among the three methods. It is found in Figure (f) that compared with the first four methods, the distance between different clusters is further increased and the inner class is further converged. Therefore, the proposed method can better mine the complementary and deep information in multi-domain data, achieving higher accuracy of fault prediction.

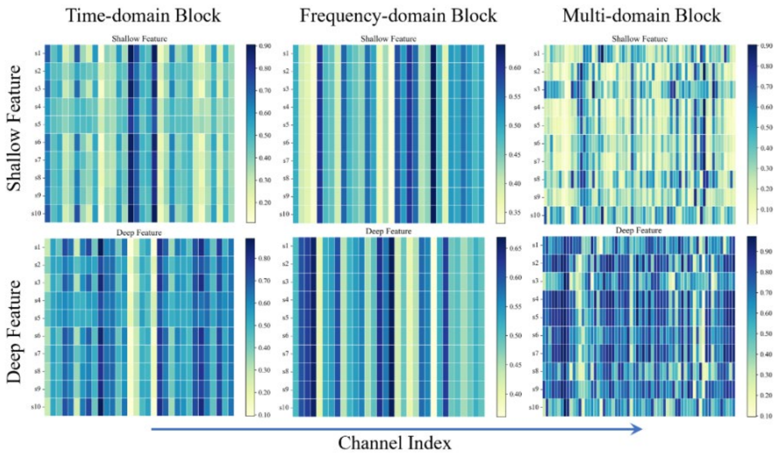


Figure 6 Weight Visualization of Relation-Aware Attention

Relation-aware attention can adaptively fuse shallow and deep abstract features in residual connection, emphasizing useful channels and weakening useless ones. Figure 6 displays the heat map of attention activation degree for selected ten samples of different health conditions, with each column representing the adaptive selection of shallow and deep features in the corresponding module. In each heat map, every square represents the weights of the s-th sample and the n-th channel, with the depth of color representing the weight value. Taking the lower left figure as an example, the weights of the 10th and 20th channels are close to 0.9, which indicates that more useful fault information exists in this channel. In addition, the weights of the 15th and 19th channels are about 0.1, which indicates that their characteristics are not conducive to fault prediction but suppressed. Moreover, the weight values of shallow and deep feature channels of the same module are correspondingly added to 1. This experiment proves the effectiveness of attention adaptive fusion, and there are some useful parts for fault prediction in both shallow and deep abstract features.

4. Summary and Future Work

Regarding the massive original vibration signals collected by sensors in industrial processes, the inputs in different domains have various sensitivities to faults, which is easy to ignore the complementary key information only by using signals in a single domain. To enhance the robustness of model training, the time-domain and frequency-domain signals are combined to construct homologous multi-domain data by using signal analysis technology. At the same time, given that with the deepening layers of deep learning model, multi-level feature mapping contributes to fault prediction to varying degrees. Based on the parallel feature extraction strategy of lightweight one-dimensional CNN, an adaptive residual fusion mechanism is introduced to better select important features at different levels in the residual process. Finally, through the comparative experiment on the rolling bearing data set of CWRU (including four sub-datasets), the effectiveness of multi-domain data and the new parallel residual network structure is verified. The accuracy rate on the four sub-data sets is higher than 99.20% with 99.77% as the highest.

Much related work still needs to be done in the future. Firstly, due to the complex and changeable industrial processes, the proposed method needs to be verified on more industrial examples and more massive industrial data to test the effectiveness and generalization ability of fault diagnosis. Secondly, the parallel computing technology of lightweight CNN is significant to improve the model efficiency, which needs further exploration. Finally, how to use efficient signal analysis technology to construct homologous multi-domain data and provide richer information for fault diagnosis is worthy of further experimental study in the future.

References

[1] R. Liu, B. Yang, E. Zio & X. Chen. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing, (108) 33–47, doi: 10.1016/j.ymssp. 2018.02.016.

[2] G. Wang, F. Zhang, B. Cheng & F. Fang. (2021). DAMER: a novel diagnosis aggregation method with evidential reasoning rule for bearing fault diagnosis. J Intell Manuf, 32(1), 1–20, doi: 10.1007/s10845-020-01554-5.

[3] L. Ciabattoni, F. Ferracuti, A. Freddi & A. Monteriu. (2018). Statistical spectral analysis for fault diagnosis of rotating machines. IEEE Trans. Ind. Electron., 65(5), 4301–4310, doi: 10.1109/TIE.2017.2762623.

[4] M. Zhao, M. Kang, B. Tang & M. Pecht. (2018). Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes. IEEE Trans. Ind. Electron., 65(5), 4290–4300, doi: 10.1109/TIE.2017.2762639.

[5] Y. Jin, C. Qin, Y. Huang & C. Liu. (2021). Actual bearing compound fault diagnosis based on active learning and decoupling attentional residual network. Measurement, 173, 108500, doi: 10.1016/j.measurement.2020.108500.

[6] H. Wang, Z. Liu., D. Peng & Y. Qin. (2020). Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis. IEEE Trans. Ind. Inf., 16(9), 5735–5745, doi: 10.1109/TII.2019.29555 40.

[7] X. Chen, B. Zhang & D. Gao. (2021). Bearing fault diagnosis base on multi-scale CNN and LSTM model. J Intell Manuf, 32(4), 971–987, doi: 10.1007/s10845-02 0-01600-2.

[8] R. F. R. Junior, I. A. D. S. Areias, M. M. Campos, C. E. Teixeira, L. E. B. Da Silva & G. F. Gomes. (2022). Fault detection and diagnosis in electric motors using 1d convolutional neural networks with multi-channel vibration signals. Measurement, 190, 110759, doi: 10.1016/j.measurement.2022.110759.

[9] W. Zhang, X. Li& Q. Ding. (2019). Deep residual learning-based fault diagnosis method for rotating machinery. ISA Transactions, (95), 295–305, doi: 10.1016/j.is atra.2018.12.025.

[10] J. Hu, L. Shen & G. Sun. Squeeze-and-Excitation Networks.

[11] X. Li, W. Wang, X. Hu & J. Yang. (2019). Selective kernel networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA: IEEE, 510–519. doi: 10.1109/CVPR.2019.00060.

[12] W. A. Smith & R. B. Randall. (2015). Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. Mechanical Systems and Signal Processing, (64–65), 100–131, doi: 10.1016/j.ymssp. 2015.04.021.

[13] W. Zhang, G. Peng, C. Li, Y. Chen & Z. Zhang. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors, 17(2), 425, doi: 10.3390/s17020425.