

A dual attention module and convolutional neural network based bearing fault diagnosis

Yazhou Zhang^{1,a,*}

¹College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou
730050, People's Republic of China

^a1911599612@qq.com

*Corresponding author

Keywords: Fault diagnosis, Convolutional neural network, Dual attention module, Non-local feature extraction module

Abstract: Vibration signals of rolling bearings are affected by changing operating conditions and environmental noise, so they are characterized by a high degree of complexity. Although deep learning fault diagnosis methods have achieved considerable success in practical applications, the high complexity characteristics are ignored. To address this issue, we propose a dual attention module and convolutional neural network (DAM-CNN) for rolling bearing fault diagnosis. In this method, we designed a dual-attention module (DAM) by using a channel-attention module and a spatial-attention module. DAM can recode feature information in channel and spatial dimensions, so as to achieve adaptive enhancement of effective network information and suppression of interference information. In addition, to enhance the extraction of long-range features of the convolutional network, we introduce the non-local feature extraction module. This module can significantly expand the perceptual field of convolutional operations and enhance the generalization ability of the network. By verifying the effectiveness of the method in CWRU datasets, the results show that the method in this paper not only has good noise immunity in strong noise environment, but also has high diagnostic accuracy and good generalization performance in different load condition domains.

1. Introduction

With the development of science and technology, rotating machinery began to develop towards automatic, efficient and intelligent trends. Rotating machinery is the most important part of mechanical equipment, which is widely used in various industries of national economy. As a key component in rotating machinery, the operating condition of rolling bearings directly affects the working process of the whole rotating machinery [1]. Therefore, it is essential to perform more accurate and intelligent fault diagnosis of rolling bearings[2].

The fault diagnosis of rolling bearings includes data acquisition, pre-processing, feature extraction and fault classification. Among them, fault classification plays a key role in the diagnosis result. The traditional fault diagnosis technology relies on experts and technicians to complete the manual feature extraction of the collected data, which cannot meet the requirements of the "big data era". With the

development of artificial intelligence technology, researchers have used support vector machine (SVM)[3] and BP neural network [4] for rolling bearing fault diagnosis. Although these methods have certain nonlinear fitting ability and achieve better results in the field of fault diagnosis, they reduce the accuracy of rolling bearing fault diagnosis because of their shallow network structure which is difficult to extract the deep feature information.

In recent years, deep learning is a new research hotspot in the development of machine learning. It was proposed by Hinton et al [5] in 2006. Because of its powerful feature automatic extraction capability, deep learning has been applied to fault diagnosis by scholars. Convolutional neural network (CNN) in deep learning is a supervised deep learning algorithm. It enables end-to-end rolling bearing fault diagnosis without preprocessing the fault data. For example, Zhang et al [6] proposed a convolutional neural network model based on the Adaptive batch normalization (AdaBN) algorithm. Bearing fault diagnosis under variable operating conditions was achieved by self-feature extraction of convolutional neural networks. Lei et al [7] proposed a wind turbine fault diagnosis method based on long short-term memory (LSTM) network combined with convolutional neural network. Although the above methods achieve better fault diagnosis results in their respective fault diagnosis tasks and higher fault diagnosis accuracy compared with traditional diagnosis methods and machine learning methods, they are also limited by the following two factors. On the one hand, these methods do not use wide convolution for extracting features, while rolling bearings have different frequency variations of the same fault information in variable operating conditions, and the general convolutional networks cannot effectively extract these small fault features. On the other hand, the network structure of these methods is more complex, and the optimized network structure is not used to easily cause problems such as the network is difficult to train or even degrade.

To address the above shortcomings, we propose a DAM-CNN method for fault diagnosis of rolling bearings in variable operating conditions. The contributions of this paper are summarized as follows:

- (1) A dual attention module is designed by using a channel attention module and a spatial attention module. This module can recode feature information in channel and spatial dimensions to achieve adaptive enhancement of network effective information and suppression of interference information.
- (2) To enhance the extraction of long-range features of the convolutional network, we introduce the nonlocal feature extraction module. This module can significantly expand the perceptual field of convolutional operations and enhance the generalization ability of the network model.
- (3) Experimental validation is performed by using CWRU datasets. Experiments on the CWRU dataset show that the method has good noise immunity and generalization performance.

2. Theoretical background

Convolutional neural network (CNN) is a feed-forward neural network and it has a powerful feature extraction capability automatically. CNN extracts deep features from the input data by constructing multiple convolutional kernels, and uses a down-sampling operation to achieve a reduced input dimensionality[8].

2.1. Convolutional layers

The convolutional layer performs feature extraction on the input by using convolutional kernels. Each element of the convolution kernel contains a weight factor and a deviation. The specific formula can be described as:

$$X_{i,j}^{l+1} = f \left(\sum_{j=1}^L \sum_{i=1}^m (X_{i,j}^l \times w_{i,j}^l) + b \right) \quad (1)$$

where the input data $X_{i,j}^l$ is the l^{th} feature value of the i^{th} feature map in the j^{th} layer of the network, L is the convolution kernel size, $w_{i,j}^l$ is the weight coefficient, b is the deviation value, and $f(\square)$ is the activation function.

2.2. Batch standardization (BN)

Batch normalization is a way to standardize the data in a network model, thus speeding up the training of the network model. Batch normalization also preserves the expressiveness of the original data. The batch normalization can be described as follows:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

$$\delta^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (3)$$

$$\tilde{x}_i = \frac{x_i - \mu}{\sqrt{\delta^2 + \varepsilon}} \quad (4)$$

$$y_i = \alpha \tilde{x}_i + \beta \quad (5)$$

where N is the number of small batches of data, x_i is the i^{th} input, μ and δ^2 are the mean and variance of the small batches of data, respectively. ε denotes a constant close to 0 but greater than 0, \tilde{x}_i is the result of data normalization, α and β are the parameters that the network can learn, and y_i denotes the i^{th} output of the data after BN.

2.3. Activation function

Both convolution and deconvolution are linear operations, however, for vibrating signals with high complexity, linear relations cannot extract enough feature information. Therefore, an activation function with nonlinear learning capability should be added after the convolution operation. We choose Leak-ReLU as the activation function. Unlike ReLU, Leak-ReLU does not set the negative value to zero and does not cause huge information waste. The specific formula can be described as:

$$ReLU_{leaky} = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases} \quad (6)$$

where x is the input signal, α is the leakage.

3. The method DAM-CNN proposed in this paper

In this section, we introduce each module of the proposed method. Then, the whole framework of the proposed method is introduced.

3.1. Dual attention module

In order to further extract the features of the convolutional flow, we propose the dual attention module (DAM). DAM consists of a channel attention module and a spatial attention module. Its structure of DAM is shown in Figure 1. In this module, first, we perform dimensional compression by 1×1 convolution. 1×1 convolution can improve the feature extraction ability of the network without increasing the number of network parameters. Then, we connect channel attention and spatial

attention in parallel. Channel attention can focus on the information in the channel dimension, while spatial attention focuses on the information in the spatial dimension. By using channel attention and spatial attention, we can achieve feature re-calibration of the convolutional stream in both space and channel. Finally, we use the Concat operation to stitch the output of the attention mechanism and extract global contextual information by 1×1 convolution. Besides, the DAM also uses residual connectivity. The residual connection not only makes the DAM better inserted into the network, but also prevents overfitting when the network is back-propagated.

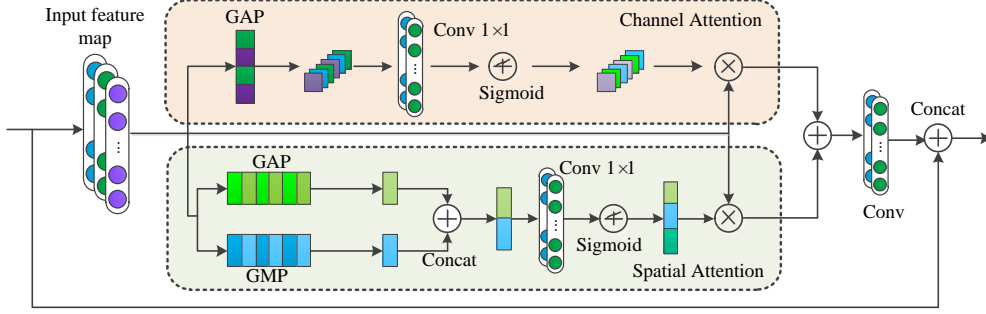


Figure 1: Dual attention module structure diagram

3.2. Non-local feature extraction module

In order to capture the length dependence of convolutional neural networks, Wang et al [9] proposed a nonlocal feature extraction method. The structure diagram of the nonlocal feature extraction method is shown in Figure 2.

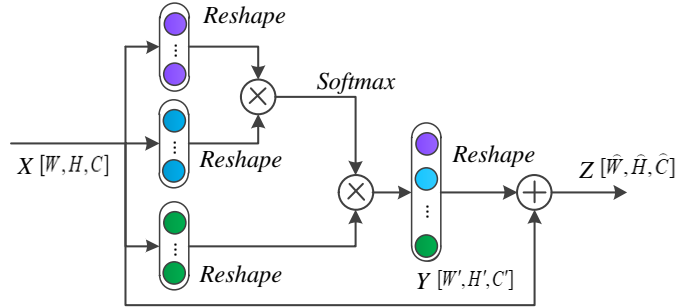


Figure 2: Non-local feature extraction module structure diagram

As can be seen from Figure 2, the non-local feature extraction module can be described as:

$$Z[\hat{W}, \hat{H}, \hat{C}] = X[W, H, C] + Y[W', H', C'] \quad (7)$$

where $Z[\hat{W}, \hat{H}, \hat{C}]$ is the output feature map of the nonlocal operation, $X[W, H, C]$ is the input feature map, and $Y[W', H', C']$ is the result of the nonlocal feature operation. Y can be described as:

$$Y[W', H', C'] = \text{Conv}(Q_s \bullet \text{Conv}(X)) \quad (8)$$

$$Q_s = \text{softmax}(R_1(X) \bullet R_2(X)) \quad (9)$$

where $\text{softmax}(\square)$ is the normalized exponential function, Q_s is the global information of the input feature mapping, $R_1(X)$ and $R_2(X)$ are the results of the convolution operation.

3.3. The proposed method in this paper

In order to solve the problem under strong noise and cross-domain conditions, a fault diagnosis method of DAM-CNN is proposed in this paper. It is shown in Figure 3. The method uses CNN to extract features from the input data, while employing dual attention for feature attention in spatial and channel dimensions, and a non-local feature extraction module for feature capture at long distances. Finally, fault classification is performed by cross-entropy damage function. The method takes the original one-dimensional vibration signal as input and does not rely on manual feature extraction and expert knowledge at all, thus making maximum use of the convolutional neural network learning capability.

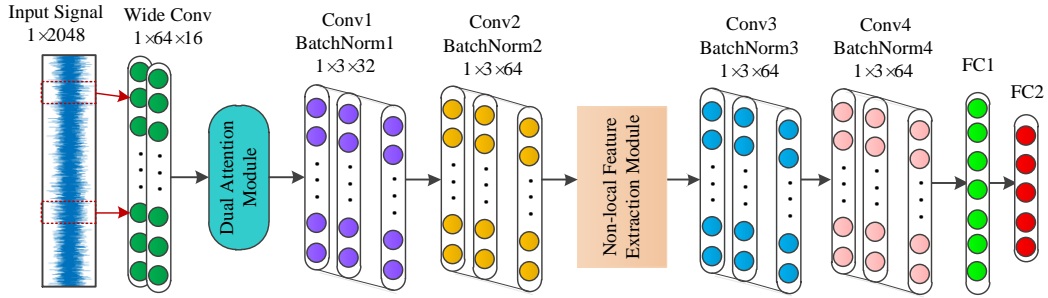


Figure 3: DAM-CNN network structure diagram

4. Experimental results and analysis

To evaluate the fault diagnosis performance of DAM-CNN, we conduct rolling bearing fault diagnosis experiments by using the Case Western Reserve University (CWRU) bearing dataset. We validate the method in terms of noise immunity, generalization, and fault diagnosis capability under cross-domain. All simulation experiments are done in the framework of deep learning Tensorflow.

4.1. Case 1

4.1.1 Dataset description

Table 1: Description of CWRU dataset

Fault type	Fault size(in)	Load (HP)	Training	Testing	Class labels
Ball	0.007	0,1,2	150	50	0
Inner race	0.007	0,1,2	150	50	1
Outer race6	0.007	0,1,2	150	50	2
Ball	0.014	0,1,2	150	50	3
Inner race	0.014	0,1,2	150	50	4
Outer race6	0.014	0,1,2	150	50	5
Ball	0.021	0,1,2	150	50	6
Inner race	0.021	0,1,2	150	50	7
Outer race6	0.021	0,1,2	150	50	8
Ball	0.028	0,1,2	150	50	9
Inner race	0.028	0,1,2	150	50	10
Normal	-	0,1,2	150	50	11

The experimental data is based on the rolling bearing dataset from Case Western Reserve University (CWRU), USA. This dataset is widely used for rolling bearing fault diagnosis. The tested bearing type is SKF6205. In this paper, the drive-side data is used which has a sampling frequency of

12 KHz. This dataset was collected for acceleration datasets at speeds of 1797 r/min, 1772 r/min, 1750 r/min and 1730r/min, corresponding to load states of 0HP, 1HP, 2HP and 3HP. The collected datasets were divided into 12 state labels according to different locations and the degree of damage, and the number of samples in each state label was approximately the same. The collected datasets were divided into training and test samples according to the ratio of 3:1, and the sampling points of each segment were set to 2048 points. The description of the experimental dataset is shown in Table 1.

4.1.2 Comparison method

To verify the superior performance of the proposed method, we chose to compare it with advanced deep learning methods, including WDCNN[6], ResNet[10], MRSCNN[11], and MSDARN[2]. WDCNN uses wide convolution in the first layer, which can effectively suppress the interference of strong noise. ResNet uses residual connectivity, which can prevent the network from overfitting. MRSCNN combines multi-scale and residual shrinkage blocks, which can effectively remove redundant information from the network. MSDARN can dynamically adjust the weights of different convolutional layers to improve the feature learning ability of the network.

4.1.3 Fault classification

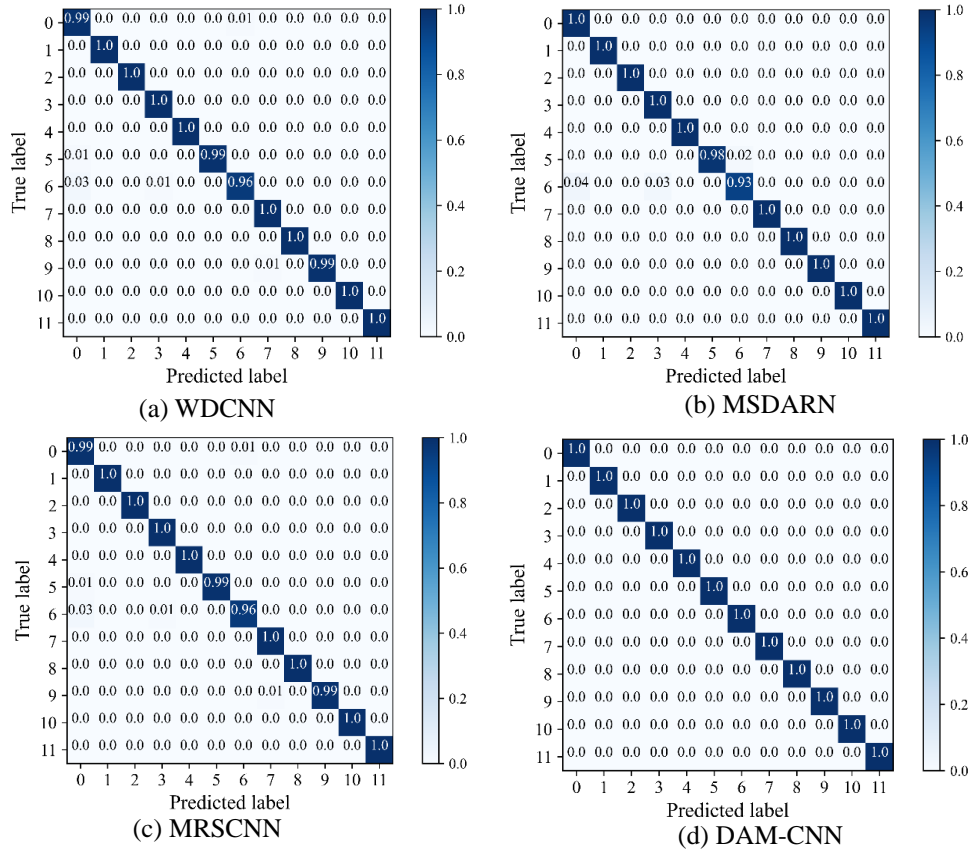


Figure 4: 1HP dataset fault diagnosis results

To verify the ability of DAM-CNN to identify faults, we quantified the diagnostic results of the 1HP dataset in Table 1 by using the misclassification quantified confusion matrix. Figure 4 shows the misclassification quantified confusion matrix for the diagnostic results of WDCNN, MRSCNN, MSDARN, and DAM-CNN. As seen in Figure 4, WDCNN, MRSCNN, and MSDARN have

misclassified and confused faults with labels 5 and 6. Finally, it leads to a diagnostic accuracy of 97.83% for WDCNN, 99.16% for MRSCNN, and 99.41% for MSDARN. However, the diagnostic accuracy of DAM-CNN is 100% without any misclassification. Therefore, DAM-CNN has a more superior diagnostic performance compared with other algorithms.

4.1.4 Visualization of learning representations

To further verify the ability of DAM-CNN adaptive feature extraction, we used the t-SNE method to analyze the adaptively extracted features and the original input data. t-SNE visualization results are shown in Figure 5. In Figure 5, the coordinates of each point indicate the location of the point in 2D space, and different labels indicate different fault types. As can be seen in Figure 5, DAT-CNN is able to fully classify the 12 fault types in the FC layer for the datasets 0HP and 1HP. For dataset 2HP, although there is a misclassification of label 6 into label 5, all other faults can be clustered accurately. Therefore, DAM-CNN has a strong feature extraction capability and can accurately classify multiple faults.

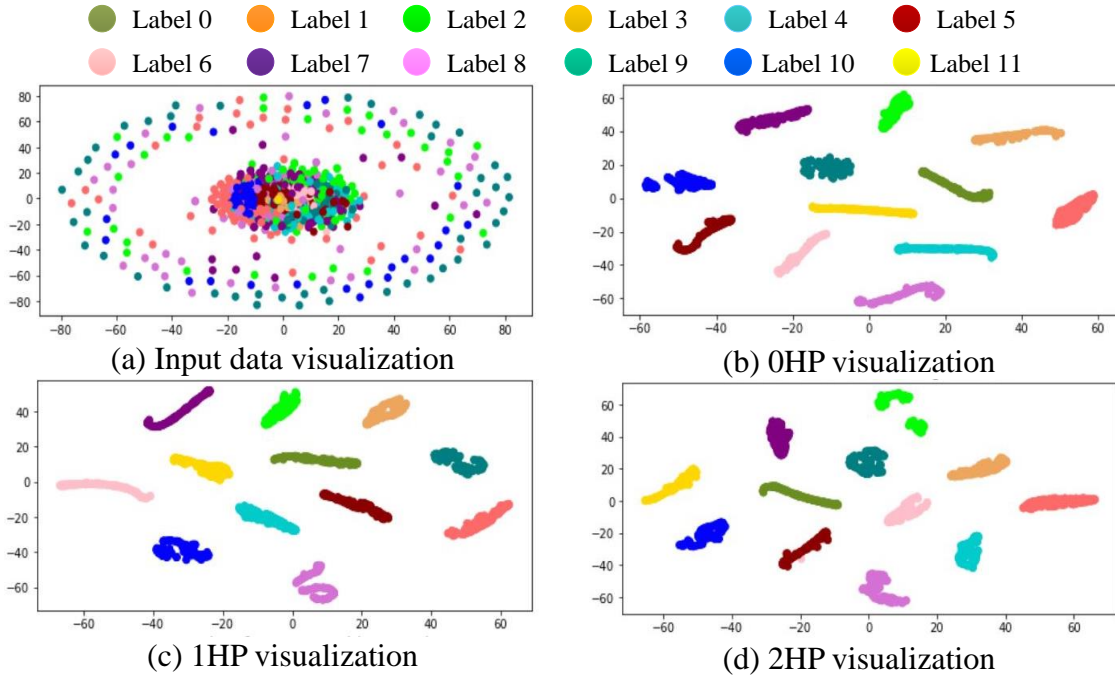


Figure 5: Visualization of learning features

4.1.5 Performance under noise environment

In this section, we discuss the diagnostic accuracy of the proposed method in a noisy environment. To simulate the noisy environment, we add Gaussian white noise to the test samples of CWRU for network performance verification[12]. The signal-to-noise ratio (SNR) is described as:

$$SNR = 10 \log_{10} (\hat{P}_S / \hat{P}_N) \quad (10)$$

where \hat{P}_S and \hat{P}_N denote the power of the original signal and the noise signal, respectively.

We add -2dB, 0dB, 2dB, 4dB, 6dB and 8dB Gaussian white noise to the 2HP test samples. The diagnosis results of DAM-CNN in the noisy environment are shown in Table 2. As can be seen from Table 3, the fault diagnosis rate of DAM-CNN is 87.50% when the SNR is -2dB. WDCNN and ResNet networks are simpler and have insufficient feature extraction capability, resulting in a diagnosis accuracy rate lower than 80%. Although MRSCNN and MSDARN use advanced deep

learning techniques, the network structure is more complex resulting in 2.00% and 4.69% lower diagnostic accuracy than DAT-CNN, respectively.

Table 2: Fault diagnosis results under noisy environment

Algorithms	Accuracy (%)					
	-2dB	0dB	2dB	4dB	6dB	8dB
WDCNN	73.72 \pm 3.19	82.76 \pm 2.78	90.99 \pm 2.09	96.02 \pm 1.31	97.39 \pm 1.19	98.82 \pm 0.54
ResNet	67.50 \pm 2.58	76.31 \pm 1.05	84.16 \pm 0.89	89.02 \pm 0.87	93.22 \pm 0.07	98.79 \pm 0.42
MRSCNN	84.50 \pm 1.91	90.50 \pm 2.13	93.50 \pm 1.79	97.50 \pm 1.30	99.10 \pm 0.12	99.81 \pm 0.07
MSDARN	82.81 \pm 1.39	90.73 \pm 0.71	94.75 \pm 0.64	98.10 \pm 0.26	98.88 \pm 0.30	99.10 \pm 0.13
DAM-CNN	87.50 \pm 2.30	93.30 \pm 0.18	96.87 \pm 0.28	99.55 \pm 0.12	100.0 \pm 0.00	100.0 \pm 0.00

4.1.6 Performance between different domains

In the industrial field, rolling bearings are often faced with the problem of cross-domain. Therefore, it is necessary to conduct experimental studies on rolling bearings with cross-domain. In order to simulate the cross-domain variation, we use one of 0HP, 1HP and 2HP as training samples and the other three load datasets as test samples. The results of the cross-domain experiments are shown in Table 3.

Table 3: Fault diagnosis results under different domains

Algorithms	Accuracy (%)					
	0HP-1HP	0HP-2HP	1HP-0HP	1HP-2HP	2HP-0HP	2HP-1HP
WDCNN	93.23 \pm 2.07	90.56 \pm 3.54	94.76 \pm 0.37	98.32 \pm 0.09	87.15 \pm 2.57	95.20 \pm 1.31
ResNet	94.36 \pm 0.20	95.97 \pm 1.27	94.15 \pm 1.99	96.39 \pm 2.26	92.29 \pm 1.19	93.05 \pm 2.05
MRSCNN	93.30 \pm 1.76	94.42 \pm 1.03	95.35 \pm 0.06	98.39 \pm 0.47	93.77 \pm 0.14	98.02 \pm 0.09
MSDARN	94.98 \pm 1.56	97.32 \pm 0.67	94.07 \pm 1.43	97.43 \pm 1.79	86.34 \pm 2.16	94.09 \pm 1.85
DAM-CNN	97.32 \pm 1.07	99.77 \pm 0.22	96.15 \pm 0.74	99.10 \pm 0.09	94.10 \pm 0.26	98.43 \pm 0.21

As can be seen from Table 3., the fault diagnosis accuracy of DAM-CNN is always better than the other four compared methods in different domain experiments. The diagnostic accuracy of DAM-CNN is 94.10% when 2HP is used as the training set and 0HP is the test set. the diagnostic accuracy of MSDARN is 7.76% lower than that of DAM-CNN.

5. Conclusion

In this paper, a fault diagnosis network DAM-CNN is proposed for noisy environments, load variations and cross-domain conditions. DAM-CNN directly uses the raw vibration signal as the input to the diagnosis network. Then, feature extraction is performed by wide convolution. The wide convolution can extract the high-frequency interference information in the vibration signal, thus improving the network diagnosis accuracy. Then, a dual-attention module is implemented to adaptively enhance the effective information and suppress the interference information of the network. Finally, DAM-CNN is experimentally validated by using CWRU bearing datasets.

Acknowledgements

The author would like to thank the referees and an editor for providing useful comments.

References

- [1] D.-T. Hoang and H.-J. Kang, "A survey on deep learning based bearing fault diagnosis," *Neurocomputing*, vol. 335, pp. 327-335, 2019.
- [2] H. Liang, J. Cao, and X. Zhao, "Multi-scale dynamic adaptive residual network for fault diagnosis," *Measurement*, vol. 188, p. 110397, 2022.
- [3] R. Jegadeeshwaran and V. Sugumaran, "Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines," *Mechanical Systems and Signal Processing*, vol. 52, pp. 436-446, 2015.
- [4] Y.-K. Gu, X.-Q. Zhou, D.-P. Yu, and Y.-J. Shen, "Fault diagnosis method of rolling bearing using principal component analysis and support vector machine," *Journal of Mechanical Science and Technology*, vol. 32, no. 11, pp. 5079-5088, 2018.
- [5] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504-507, 2006.
- [6] 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, p. 425, 2017.
- [7] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks," *Renewable energy*, vol. 133, pp. 422-432, 2019.
- [8] J. Zhao, S. Yang, Q. Li, Y. Liu, X. Gu, and W. Liu, "A new bearing fault diagnosis method based on signal-to-image mapping and convolutional neural network," *Measurement*, vol. 176, p. 109088, 2021.
- [9] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794-7803.
- [10] W. Zhang, X. Li, and Q. Ding, "Deep residual learning-based fault diagnosis method for rotating machinery," *ISA transactions*, vol. 95, pp. 295-305, 2019.
- [11] X. Zhao and Y. Zhang, "An intelligent diagnosis method of rolling bearing based on multi-scale residual shrinkage convolutional neural network," *Measurement Science and Technology*, vol. 33, no. 8, p. 085103, 2022.
- [12] H. Liang and X. Zhao, "Rolling bearing fault diagnosis based on one-dimensional dilated convolution network with residual connection," *IEEE Access*, vol. 9, pp. 31078-31091, 2021.