World Journal of Engineering Research
ISSN:2959-98731

RESEARCH ON ROLLING BEARING FAULT DIAGNOSIS BASED ON PARALLEL 1DCNN

Jianguang Yuan[1], Hongmin Liu[2], Yinghua Wang[1, 2, *]
[1.] *College of Mechanical Engineering, Yanshan University, Qinhuangdao 066004, Shandong, China;*
[2.] *TBEA Hengyang Transformer Co., Ltd, Hengyang 421007, Hunan, China.*

Abstract: Rolling bearings are commonly used parts in rotating machinery. Due to their harsh working environment, they are prone to failure. Therefore, a fault diagnosis method for rolling bearings based on parallel 1DCNN (one-dimensional convolutional neural network) is proposed. First, the vibration signal of the rolling bearing was processed and divided into training set and test set; then, a parallel 1DCNN model composed of two channels was constructed, which can obtain the time domain information and frequency domain information of the vibration signal respectively. A relatively small convolution kernel is used when extracting time domain information, and a relatively large convolution kernel is used when extracting frequency domain information, and the traditional fully connected layer is replaced by a global maximum pooling layer; finally, the trained The parallel 1DCNN model processed the rolling bearing test set data of Case Western Reserve University; at the same time, in order to verify the fault diagnosis effect of the parallel 1DCNN model, the model was compared with the traditional CNN model. The research results show that the fault diagnosis accuracy of the parallel 1DCNN model is higher than 0.996. Compared with the traditional single-channel CNN model, the parallel 1DCNN model can make full use of the extracted time domain and frequency domain feature information, and has better fault diagnosis ability.
Keywords: Rolling bearing; Fault diagnosis; Convolutional neural network; 1DCNN; Deep learning; Feature extraction

1 INTRODUCTION

Rolling bearings are modern mechanical equipment, especially commonly used components in rotating machinery. They usually work in harsh environments with high temperature, high pressure and high-speed rotation, so they are prone to failure. According to the statistics of relevant departments, in the industrial field, the failure of mechanical equipment caused by rolling bearing failure accounts for 41% [1].

The use of fault diagnosis technology can not only monitor the operating status of mechanical equipment in real time, but also identify the location of the fault and the degree of damage to improve the reliability and stability of mechanical equipment. Therefore, the development of fault diagnosis technology with rolling bearing as the research object is of great significance for improving the safety of the entire mechanical system.

With the gradual development of modern machinery and equipment in the direction of complexity and intelligence, the monitoring of the operating status of rolling bearings has also entered the era of "big data". In order to achieve better diagnostic results, traditional fault diagnosis techniques require complex signal processing methods and professional background knowledge. In the face of massive data in the "big data" era, this process usually requires a lot of manpower, which limits the application prospects of traditional technologies to a certain extent.

In recent years, with the rapid development of artificial intelligence technology, fault diagnosis technology based on deep learning theory has been widely used. This technique is especially suitable for solving bearing fault diagnosis problems in complex scenarios [2, 3]. In this context, many scholars have carried out a lot of research work.

Han Tao et al. [4] first processed the vibration signal of rolling bearing with multi-wavelet transform (MWT), and obtained the corresponding multi-wavelet coefficient branch. On this basis, a feature map was constructed and a convolutional neural network was established. Network (CNN) classifier model realizes the intelligent diagnosis of complex faults of rolling bearings. Wu Chunzhi et al. [5] proposed a gearbox fault diagnosis model based on one-dimensional CNN, which can learn features directly from the original vibration signal, and completed the gearbox fault diagnosis. Zhang Xiangyang et al. [6] aimed at the weak fault characteristics of the casing under the excitation of rolling bearing faults, using the continuous wavelet scale spectrum method to convert the one-dimensional original signal into a two-dimensional image signal, and input it into the CNN network to identify the rolling bearing fault. Wang Hailong et al. [7] combined empirical mode decomposition (empirical mode decomposition, EMD) and CNN methods, and constructed a two-dimensional feature map with the modal components obtained from the EMD processing of the rolling bearing signal and the original vibration signal. The feature maps are fed into CNN, which achieves better rolling bearing diagnosis results. Aiming at the problems of noise interference in the vibration signal of rolling bearings, Dong Shaojiang et al. [8] proposed an anti-noise multi-core convolutional neural network, and added a dropout layer to the network to improve the anti-interference ability of the model. Liu Hongjun et al. [9] converted the one-dimensional time-series vibration signal into a two-dimensional image through the Gram angle difference field, extracted its image features, and input it into the improved CNN model, and then passed the Adam mini-batch optimization method Iterative training is carried out, and finally the ideal fault detection accuracy is achieved.

The traditional CNN-based bearing fault diagnosis technology mainly uses the time domain information or frequency domain information of the vibration signal, but neither effectively combines the two and makes better use of them.

Based on the above reasons, the author proposes a rolling bearing fault diagnosis method based on parallel 1DCNN (one-dimensional convolutional neural network). First, the vibration signal of the rolling bearing is processed, and then a parallel 1DCNN model consisting of two channels is constructed, one channel learns the time domain information of the vibration signal, and the other channel learns the frequency domain information of the vibration signal, thereby speeding up the signal processing speed and reducing the amount of calculation , so as to effectively improve the utilization rate of fault feature information in bearing signals and enhance the fault identification accuracy of the model.

2 METHOD PRINCIPLE

2.1 Convolutional Neural Network

Convolutional neural network (CNN) is a multi-layer structure model, which consists of input layer, convolution layer, pooling layer, fully connected layer and output layer [10].

2.1.1 Input layer

As the first layer of CNN, the input layer is used to receive training and prediction samples, and to verify the size and format of the input samples.

2.1.2 Convolution layer

The convolutional layer is the core of CNN [11], and the feature information of the input image data can be effectively extracted through the convolution operation.

The specific operation of the convolutional layer is as follows:

$$X^{i}_{m} = \sum_{j \in C_n} X^{i}_{n-1} \times W^{i}_{nm} + b^{i}_{m} \quad (1)$$

In the formula: Cn—input feature vector; i—number of network layers; X^{i}_{m}—output of the i-th layer; X^{i}_{n-1}—input of the first layer; W—weight parameter; b—bias parameter.

It should be noted that the input to output of the convolutional layer is a linear mapping, which cannot approximate complex functions. In order to make the convolutional layer have nonlinear output capability, an activation function is usually introduced, whose function is to nonlinearly map the output of the convolutional layer. Generally, a rectified linear unit (ReLu) is used as the activation of a one-dimensional CNN. function.

2.1.3 Pooling layer

The role of the pooling layer is to downsample the network and reduce the dimension of the data, thereby reducing the parameters that need to be trained in CNN. Too many network parameters will not only reduce the training speed, but also easily cause overfitting [12]. Therefore, pooling layers are usually added after some convolutional layers for downsampling. There are generally three strategies for pooling, namely average pooling, maximum pooling, and minimum pooling. Currently, the latter two are widely used by CNN, namely:

$$C^{l}_{j+1} = \max_{(s-1)L+1 < m < sL}\{a^{l}_{j}(m)\},\ s = 1, 2, \ldots, M \quad (2) \qquad C^{l}_{j+1} = \operatorname{avg}_{(s-1)L+1 < m < sL}\{a^{l}_{j}(m)\},\ s = 1, 2, \ldots, M \quad (2)$$

In the formula: C + 1—the output of the pooling layer; s—the step size of the one-dimensional pooling kernel; L^{l}_{j}—the length of the one-dimensional pooling kernel; j^{l}Neurons.

2.1.4 Fully connected layer

The fully connected layer splices the two-dimensional features processed by the previous layers to obtain a one-dimensional feature, and uses it with the classifier to realize the classification function [13,14].

2.1.5 Output layer

The output layer contains a classifier, whose role is to classify the training features, and feed back the error with the objective function to the CNN to iteratively update the weights during training.

Since this model only uses the relevant information in the time domain of the signal, in order to make full use of the information in the frequency domain, the author proposes a parallel 1DCNN network model.

2.2 Parallel Convolutional Neural Network

Since the vibration signal generated by the rolling bearing is a one-dimensional time-series signal, the author mainly studies the one-dimensional CNN here.

The parallel 1DCNN consists of two channels, which can obtain the time-domain information and frequency-domain information of the vibration signal respectively. A relatively small convolution kernel is used when extracting time domain information, and a relatively large convolution kernel is used when extracting frequency domain information. The purpose is to ensure that the training speed of the network can be accelerated while obtaining fault feature information.

Compared with the traditional single-channel CNN model, the parallel 1DCNN model has better fault diagnosis ability because it can make full use of the extracted time-domain and frequency-domain feature information.

The model structure of parallel 1DCNN is shown in Figure 1.

In Figure 2, the first convolutional layer uses a small convolution kernel and a large convolution kernel for time domain information and frequency domain information, respectively, and then uses smaller convolution kernels; at the same time, in each convolution layer The zero padding operation is set to ensure that the input and output of the convolutional layer have the same dimension, which is beneficial to protect the edge information of the signal.

In the process of processing the bearing signal, the parameters of the small convolution kernel refer to the traditional 1DCNN model, and the parameters of the large convolution kernel are determined according to the relevant formula of the CNN receptive field. The schematic diagram of CNN's receptive field is shown in Figure 2.

Figure 2 , the black neurons in the input signal represent the receptive field of the neurons in the last pooling layer in the input signal. The core of constructing CNN is to determine the receptive field, that is, the perception range of a neuron in its underlying network [15]. Since the mechanical vibration signal is periodic, and the phase of each input signal is not necessarily the same. In order for the network to learn displacement-independent features, it should be ensured that the receptive field size of the neurons in the last pooling layer in the input signal is greater than one signal period.

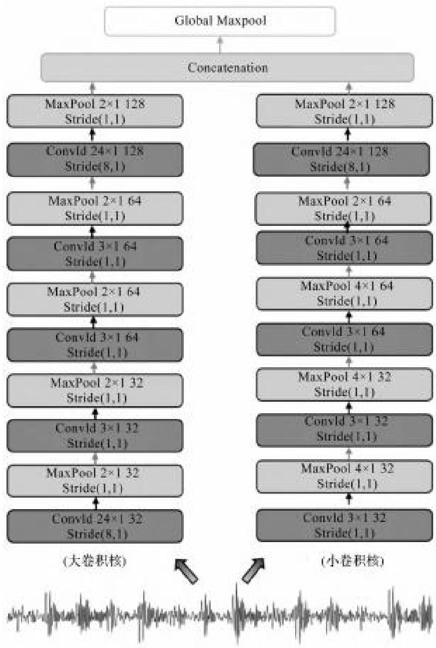

Fig. 1 Parallel 1DCNN network model

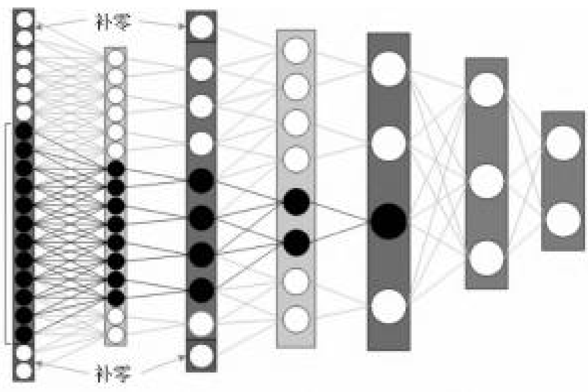

Fig. 2 Schematic diagram of receptive field

In parallel 1DCNN, R(0) is the receptive field of neurons in the last pooling layer in the input signal, T is the number of samples in one cycle when the signal is collected, and L is the length of the one-dimensional time series signal, then the receptive field R (0) should satisfy $T \le R(0) \le L$, and its calculation process is as follows:

The neurons in the last pooling layer satisfy the following relationship between the receptive field R(K) of the K-th pooling layer and the receptive field R(K-1) of the K-1th pooling layer:

$$R(K-1) = S(K)\ (P(K)\ R(K)-1) + W(K) \tag{3}$$

In the formula: S(K) —the step size of the Kth convolutional layer; W(K) —the width of the Kth convolutional kernel; P(K) —the number of downsampling points of the Kth pooling layer.

In this paper, when the number of layers K>1, S(K) = 1, W(K) = 3, P(K) = 2, then formula (4) can be simplified as:

$$R(K-1) = 2R(K) + 2 \tag{4}$$

When K = n, R(n) = 1, then the receptive field of the last pooling layer in the first pooling layer is:

$$R(1) = 2n-1 \times 3-2 \tag{5}$$

According to the above relationship, the receptive field of the neurons in the last pooling layer on the input signal can be calculated as:

$$R(0) = S(1)\ (P(1)\ R(1)\ -1) + W(1) = 2S(1)\ (2n-1 \times 3-2) + W(1)\ -S(1) \approx (7)\ S(1)\ (2n \times 3-4) \tag{6}$$

Since T≤R(0) ≤L, T≤S(1) (2n × 3－4) ≤L, meanwhile the step size S(1) should be able to divide the signal length L evenly. Here the length of the input signal is 1024, and the period T is 400. When the number of convolution layers is 5, S(1) is selected as 8, and the convolution width is not less than 3 times the step size. Therefore, the first convolution kernel width is selected as 24 here. In order to simultaneously extract the time domain and frequency domain information in the vibration signal of the rolling bearing, another channel of the parallel 1DCNN uses a smaller convolution kernel.

It should be noted that fully connected layers are usually used in traditional 1DCNN. However, the disadvantage of the fully connected layer is that there are many training parameters, which reduces the training speed of the model and easily leads to overfitting [16, 17]. For this reason, in the parallel 1DCNN model, the author uses the global maximum pooling layer to replace the fully connected layer.

In addition, in order to improve the efficiency of fault diagnosis, the author uses a callback function in the parallel 1DCNN model to ensure that the training stops when the loss no longer changes. The author combines the Model Checkpoint and Early Stopping callback functions. When the monitoring target index does not change within the set round, you can use EarlyStoppin to terminate the model training. At the same time, ModelCheckpoint can continuously save the model during the training process to obtain the best model..

3 ROLLING BEARING FAULT DIAGNOSIS SIMULATION

In order to verify the effectiveness of the parallel 1DCNN model, the author uses the bearing data set of Case Western Reserve University (WRU) in the United States for simulation.

In the used data set, the rotational speed of the workbench is 1772 r/min, and 400 samples can be collected in one cycle. Bearing faults include inner ring, outer ring and rolling element faults. Each fault corresponds to three different damage degrees, and the fault sizes are 0.007 ft, 0.014 ft, and 0.021 ft, respectively.

The author selects 3 sets of inner ring fault data with different damage levels, 3 sets of outer ring fault data with different damage levels, 3 sets of rolling element fault data with different damage levels and 1 set of normal data, and divides all the above data according to 56%, The ratio of 14% and 30% is divided into training set, verification set and test set. The detailed usage of the data is shown in Table 1.

Table 1 Dataset description

fault location	category label	Fault depth/ft	Number of samples
normal	0	0	500
	1	0.007	500
Inner ring failure	2	0.014	500
	3	0.021	500
	4	0.007	500
Outer ring failure	5	0.014	500
	6	0.021	500
	7	0.007	500
rolling element failure	8	0.014	500
	9	0.021	500

In order to prevent the underfitting phenomenon of the network model during the training process due to too few samples in the data set [18], in the experiment, the author uses the resampling method to increase the number of data samples [19].

Resampling is a method of data augmentation. It enhances the correlation between samples by overlapping reading vibration signals, so that the model can enhance its robustness through learning. The signal length selected here is 1024, and the sampling interval is 200.

3.1 Network Visual Analysis

3.1.1 Visualization of the middle layer of the network

The visualization of the middle layer can show the process of the network model extracting the fault characteristics of the bearing vibration signal, so as to better understand the parallel 1DCNN model. The author takes the inner circle fault signal as an example to show the visualization results of the middle layer of parallel 1DCNN.

The waveform diagram of the inner ring fault signal is shown in Figure 3.

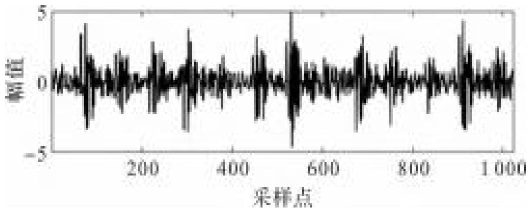

Fig. 3 inner ring fault signal

The author inputs the fault signal into the parallel 1DCNN model, then extracts the intermediate output results of different convolutional layers, and visualizes them : when the signal passes through the convolutional layer, different channels can extract different features (for example, the second convolutional layer The 6th channel of the multilayer is sensitive to the shock part of a certain area of the vibration signal; while the 13th channel of the third convolutional layer is sensitive to the shock part of another part of the vibration signal).

Since different channels can detect the impact signals of different sections of the bearing, the network model can learn the rules of different faults by identifying the impact of different areas of the signal, and thus identify the fault category and fault depth of the bearing.

3.1.2 T-SNE visualization

t-SNE (t-distributed stochastic neighbor embedding) is a nonlinear dimensionality reduction technique, mainly used to visualize high-dimensional data [20]. Through t-SNE dimension reduction, high-dimensional data can be reduced to a two-dimensional plane scatter plot, and the scatter plot can intuitively reflect the fault classification of the deep learning model.

In the parallel 1DCNN model, the classification results of the input layer are shown in Figure 4.

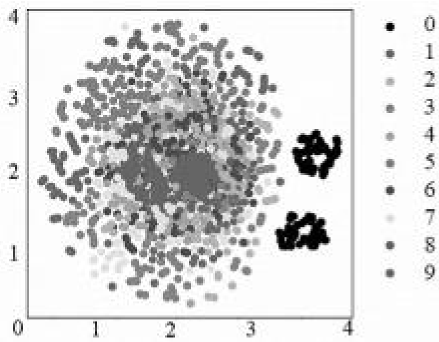

Fig. 4 Distribution of input samples

from Figure 4 that the normal data and fault data are clearly separated; but at the same time, the data of different fault types are completely superimposed together, which cannot be well distinguished. In the parallel 1DCNN model, the classification results of the output layer are shown in Figure 5.

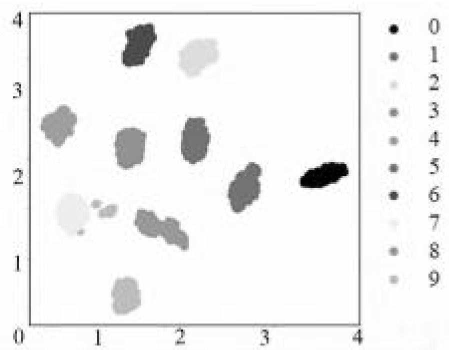

Fig. 5 Output of parallel 1DCNN model

From Figure 5 that after parallel 1DCNN processing, the 10 types of data can be better distinguished, and each data structure is more compact, only a few category 9 is wrongly predicted as category 7; In other words, the difference between the vibration signals caused by damage size is relatively similar.

3.2 Comparison of Results

In order to avoid chance, in the experiment, it is necessary to train the parallel 1DCNN model 5 times in parallel, and the fault diagnosis accuracy obtained on the test set is 0.9969.

In addition, the author also tested the diagnosis accuracy of the small convolution kernel network and the large convolution kernel network in the experiment, and compared the results with the diagnosis accuracy obtained by using the parallel 1DCNN model. The results are shown in Table 2.

Table 2 The model accuracy obtained by training the model in parallel 5 times

Model	model accuracy					
group	1	2	3	4	5	
small convolution kernel network	0.985 3	0.988	0.990 6	0.989 3	0.993	3
Large Convolutional Kernel Network	0.990 6	0.989 3	0.992	0.986 6	0.993	3
parallel network	0.996 7	0.996 7	0.998 0	0.996 7	0.996	7

It can be clearly seen from Table 2 that the fault diagnosis accuracy of the parallel 1DCNN model is higher than that of the other two models: the parallel 1DCNN model has achieved a better fault diagnosis effect, and the accuracy of the five runs is higher than 0.996, up to 0.998.

4 CONCLUSION

In order to make full use of the fault feature information in the time domain and frequency domain, the author proposes a new bearing fault diagnosis model, that is, a parallel 1DCNN (one-dimensional convolutional neural network) model.
When using this model to diagnose bearing faults, the vibration signal of the rolling bearing is processed first, and then a parallel 1DCNN model composed of two channels is constructed. One channel learns the time domain information of the vibration signal, and the other channel learns the frequency domain information of the vibration signal. This speeds up the signal processing speed, improves the utilization rate of the fault feature information in the bearing signal, and enhances the fault identification accuracy of the model.
Since 1DCNN is an improvement over traditional CNN, it has better fault diagnosis ability. Compared with other intelligent diagnosis models, parallel 1DCNN has the following advantages:
(1) Parallel 1DCNN learns the time domain information and frequency domain information of the rolling bearing vibration signal through two channels with different convolution kernel sizes, and can make full use of the fault information contained in the vibration signal;
(2) The input of the model is simpler, and the one-dimensional time-series signal is directly processed without converting it into a two-dimensional image, thereby speeding up the signal processing speed and reducing the amount of calculation;
(3) Compared with the traditional CNN model, parallel 1DCNN can effectively improve the utilization rate of fault feature information in bearing signals, and greatly improve the fault recognition accuracy of the model.
In future research, the author will focus on exploring the impact of environmental noise on the stability of parallel 1DCNN, and propose corresponding solutions.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Song Yuyu, Liu Zhaoliang. Research on Rolling Bearing Fault Diagnosis Based on ICEEMD-EFICA Combined Noise Reduction. Mechanical and Electrical Engineering, 2021, 38 (4): 494-497, 520.
[2] Zhu Guanlin, Wang Zhaoqiang, Wang Yifan. Research on hydraulic pump fault diagnosis based on neural network and evidence fusion. Mechanical and Electrical Engineering, 2020, 37(12): 1498- 1503.
[3] Qu Jianling, Yu Lu, Yuan Tao. Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network. Journal of Instrumentation, 2018, 39 (7): 134-143.
[4] Han Tao, Yuan Jianhu, Tang Jian. Intelligent composite fault diagnosis method for rolling bearings based on MWT and CNN. Mechanical Transmission, 2016, 40 (12): 139-143.
[5] Wu Chunzhi, Jiang Pengcheng, Feng Fuzhou. Gearbox fault diagnosis based on one-dimensional convolutional neural network. Vibration and Shock, 2018, 37(22): 51-56.
[6] Zhang Xiangyang, Chen Guo, Hao Tengfei. Convolutional neural network diagnosis method for rolling bearing faults based on casing signals. Journal of Aerodynamics, 2019, 34 (12): 2729-2737.
[7] Wang Hailong, Xia Xiaoyun, Sun Weitang. Rolling bearing fault diagnosis based on EMD and convolutional neural network. Combined Machine Tool and Automatic Processing Technology, 2019 (10): 46-48, 52.
[8] Dong Shaojiang, Yang Shuting, Wu Wenliang. Bearing fault diagnosis method based on anti-noise multi-core convolutional neural network. Journal of Beijing University of Chemical Technology: Natural Science Edition, 2020, 47(6): 100-106.

[9] Liu Hongjun, Wei Xuyang. Research on rolling bearing fault diagnosis based on GADF and convolutional neural network. Mechanical and Electrical Engineering, 2021, 38(5): 587-591.

[10] Zhao R, Yan RQ, Chen ZH. Deep learning and its applications to machine health monitoring. Mechani- cal Systems and Signal Processing, 2019, 115: 213-237.