Fault Diagnosis of Fan Bearing Based on Improved Convolution Neural Network

Boyang Ma*
Datang Northeast Electric Power Test & Research Institute Co. Ltd. Changchun City, Jilin Province
*Corresponding author: boyangma@china-cdt.com

Abstract. Because of its accuracy, Convolutional Neural Network (CNN) has become an important method in the field of fault diagnosis. However, the traditional CNN has a long time of training and diagnosis due to its complex structure. At the same time, due to many problems in the network, the detection accuracy is not high. Therefore, this paper proposes an improved CNN for fan bearing fault diagnosis, which speeds up the feature extraction of the network by improving the network structure; solves the problem of part of neurons not being activated by improving the activation function, and improves the accuracy of network detection. Finally, the network proposed in this paper is validated on the data set and compared with other advanced fault diagnosis algorithms. The results show that the accuracy of the algorithm proposed in this paper can reach 99.76%. Because of other algorithms, and the training and diagnosis time is relatively short, it has practical application value.

Keywords: Convolutional Neural Network, Fault diagnosis, deep learning.

1. Introduction
In recent years, China has made great achievements in the field of wind power [1-3], a large number of fans have been put into actual operation, and the economic loss caused by fan failure is increasing year by year. As an important part of fan, bearing will cause a lot of economic losses once it breaks down. Therefore, it has become a hot issue to find out the fault in time and avoid causing great economic loss. With the development of research, the bearing fault diagnosis method based on deep learning has become the mainstream research method because of its high precision. Traditional fault diagnosis methods, such as wavelet transform, support vector machine and spectrum analysis, transform and extract the features of vibration signals, analyze the extracted features, and classify the bearing state information contained in the signals [4]. These methods are often complex to operate and rely on human to provide features containing fault information, and the effect in practical application is often unsatisfactory.

In view of the shortcomings of traditional methods, this paper proposes an improved CNN and applies it to bearing fault diagnosis [5-8]. In this method, the traditional CNN is improved in two aspects: network structure and activation function. ROI align is used to speed up feature extraction, which greatly improves the training and calculation speed of the network. Linear swish function is used to further improve the number of neurons activated in the training process and improve the
accuracy of network diagnosis. Finally, the improved CNN algorithm is applied to the bearing experimental data set for fault diagnosis, and compared with the traditional SVM, KNN, BP network and DNN algorithm, which verifies the superiority and effectiveness of the improved CNN algorithm [9-10].

2. Related work

2.1. Fan bearing vibration data
In this paper, the main fault of the fan is the bearing fault of the fan. Usually, the most common signal to detect the fault of bearing is its vibration signal, set released by Case Western Reserve University.

Fig. 1 vibration signal of a bearing

Fig. 1 is the vibration signal of a bearing, which comes from the open source data. At present, bearing fault diagnosis algorithm has developed rapidly. CWRU has become the best third-party standard when verifying the performance of the algorithm [11-13].

2.2. Convolutional Neural Networks
Convolution neural network is a multi-level neural network, which can be divided into two parts to extract and classify the features of the target. The feature extraction part is composed of convolution layer, pooling layer and activation layer, and the classification part is composed of full connection layer. As shown in Fig. 2, CNN is generally composed of input layer, convolution layer, activation layer, pooling layer, full connection layer and output layer. Among them, the convolution layer, the activation layer, the pooling layer and the composition hidden layer can construct a deep CNN network by alternately stacking multiple groups of hidden layers.

Fig. 2 Structure of CNN

1 Convolution layer
Convolution operation is the core module of CNN. It uses rectangular convolution kernel to traverse every pixel on the image, uses parameter sharing mechanism, and the same input feature map uses the same convolution kernel weight. In general, the operation of convolution layer can be expressed in the form of Eq. 1[14]
In Eq. 1, \( x_i \) is the characteristic graph of the current layer; \( x_{i+1} \) is the characteristic graph after convolution; \( \odot \) is convolution operation; \( \mathcal{W}_i \) is convolution kernel weight; \( b_i \) is bias. The commonly used activation function in CNN is Relu.

2 Activation layers

The function of the activation layer is to enhance the nonlinear expression ability of the network through the activation function. For Eq. 1, after using activation function, it can be expressed as Eq. 2

\[
y_{i+1} = f(x_{i+1})
\] (2)

In Eq. 2, \( f(\cdot) \) is the function form of activation function relu, which can be expressed by Eq. 3

\[
f(x) = \begin{cases} 
x & x \geq 0 \\
0 & x < 0
\end{cases}
\] (3)

3 Pooling layers

The main function of pooling layer is to reduce the dimension of the convoluted high-dimensional feature map. On the premise of ensuring that the features are not lost as much as possible, the size dimension of the feature map is reduced, that is to find the representative features and reduce the computational complexity. CNN uses the maximum pooling layer to pool, that is to select the largest value in the region corresponding to the pool core as the representative value of the region. Although the numerical value obtained in this way has strong expression ability for texture features, it has lost some regional features. In general, the process of maximum pooling layer can be expressed by Eq. 4

\[
\text{max Pooling} = \max(f[i-1], f[i], f[i+1])
\] (4)

3. Improved Convolutional Neural Networks for Bearing Fault Detection

3.1. Improved Network Architecture

In traditional CNN network, the convolution feature of high dimension is reduced by pooling layer. However, pooling layer will produce some feature loss, which will reduce the accuracy of recognition. In order to solve the problem of feature loss, the maximum pooling layer after the convolution layer is deleted, so that the feature map with higher resolution can be obtained on the next convolution layer. Fig. 3 illustrates architecture of our model. The proposed network consists of fully convolutional layers for constructing a shared feature map, an adaptive RoI Align layer for extracting feature of each RoI, three fully connected layers for binary classification. Given a whole input matrix, the network computes a shared feature map of the input data through a single forward pass. A CNN feature corresponding to each RoI is extracted from the shared feature map using an adaptive RoI Align operation. Through this feature computation strategy, we reduce computational complexity significantly while improving quality of features [15].
In Figure 3, we can see the overall workflow of the network. Because CNN has better perception ability for two-dimensional signals, the better classification performance can be obtained by transforming one-dimensional vibration signals into two-dimensional image input network. Our network has an RoI Align layer to obtain object representations from a fully convolutional feature map constructed from a whole data. However, RoI Align are inherently coarse in features extracted. As Fig. 4 shown, it is the specific implementation method of ROI Align.

It can be seen from the Fig. 4 that compared with pooling operation, ROI Align gives up twice quantization and uses interpolation method to obtain regional pixel value, so as to better express the image. However, the traditional ROI Align is used to calculate the interpolation by using the grid points near the feature map. Once the sampling interval is too large, some useful information will be lost. For example, if only $2 \times 2 = 4$ feature points are selected on a $5 \times 5$ feature grid, it will inevitably lead to the loss of features. Therefore, an adaptive ROI Align is proposed to solve the problem of feature loss. Because the traditional ROI align selects a fixed size in the sampling process, this is the reason for the largest feature loss. To handle this issue, we adjust the interval of the grid points from the dense feature map adaptively. In specific, the bandwidth of the bilinear interpolation is determined by the size of RoIs; it is proportional to $\left\lfloor \frac{w}{w_o} \right\rfloor$, where $w$ and $w_o$ denote the width of RoI after the Conv3 layer and the width of RoI’s output feature in the RoIAlign layer, respectively, and $\lfloor \cdot \rfloor$ is a rounding operator. We call the ROI of this sampling interval related to the size of the feature map as adaptive ROI. Although the improved RoI Align makes minor changes, it can make the network in the overall classification performance has achieved better results, the main reason for the performance improvement is to reduce the small error, thus avoiding the global cumulative error.

3.2. Improved Activation Functions

Generally, CNN uses Relu function as the activation function of network [16], which can be expressed by Eq. 5 and Fig. 5:

$$\text{Relu}(x) = \max(0,x)$$  (5)
Relu solves the problem of gradient vanishing, and has a fast calculation speed, which is better than tanh function and Sigmoid function in convergence speed. However, because the negative half axis of relu function is always 0, some neurons will not be activated, that is, some parameters will not be updated during the training process with higher learning rate. In order to solve the problem of relu function, a new function which we called Swish is used as the activation function of neural network. Swish is a new activation function, which has great advantages in improving network accuracy. The function form is Eq. 6

\[ Swish(x) = x \cdot Sigmoid(x) \]  

Swish has the same problem as Sigmoid, and is more complex in calculation, so linear approximation is used to obtain the improved function shown in Fig. 6:

This linear approximation function is called \( h - Swish \). After testing, the improved activation function can improve the accuracy of classification without affecting the calculation speed. The form of \( h - Swish \) can be expressed by Eq. 7.

\[ h - Swish(x) = x \cdot \frac{\text{Relu}(x + 3)}{6} \]
3.3. Implementation Details
The outline of proposed network is shown in Algorithm 1. And the relevant implementation details are presented as follows.

**Algorithm 1** Proposed Classification Algorithm:
Iterate at Frame $t$

- **Input**: Vibration data of bearing
- **Output**: fault type
- **Repeat**
  - Resize one-dimensional data into two-dimensional data
  - **If** input data exists & Model exists
    - Input network as Fig. 3 for calculation
    - Updated Model
  - **Else**
    - Show missing data
  - **END**
- **Until**: End of the data

In the process of using the network, we first preprocess the data to convert one-dimensional data into two-dimensional data more suitable for CNN network.

1. **Network Initialization and Input Management**
   The weights of the three convolution layers are pre-trained by CWRU, while the full connection layer is randomly initialized. The size of the input data is adjusted to fit the size of the target object to $107 \times 107$, and is clipped to the smallest rectangle containing all sample ROI. The receiving field size of a single cell in the last convolution layer is equal to $75 \times 75$.

2. **Offline Pretraining**
   The parameters involved in the training process are shown in Tab. 1:

   | PARAMETERS     | VALUE |
   |----------------|-------|
   | Bitch_Size     | 8     |
   | Epoch          | 50    |
   | Number of threads | 1    |
   | learning rate  | 0.01  |
   | Optimizer      | Adam  |

3. **Online detection**
   In Offline Pretraining, we have obtained a model which can classify the vibration signals. In the off-line detection stage, we need to use the actual vibration signals to test the model. We select any 100 groups of data in CWRU to test, and draw the confusion matrix as shown in Fig. 7. Finally, according to the confusion matrix, the structure of the data set is modified and the number of samples is redistributed. Thirdly, based on the epoch $= 50$ model, the new data set is used to re-train the model.
4. Experiments

4.1. Experimental Data Set
In order to evaluate the effectiveness and accuracy of the improved CNN algorithm in the intelligent diagnosis of rolling bearing micro fault, the experimental verification is carried out in this paper. This experiment takes deep groove ball bearing as the object, and the fault data set is from Case Western Reserve University. Data can be divided into four categories, Normal, Ball, Inner Race, Outer Race.

4.2. Fault data preprocessing
Firstly, the vibration data is a typical one-dimensional signal, so it is converted into two-dimensional data, and then the following operations are performed.

1 Batch Normalization
Batch normalization (BN) is one of the most commonly used training techniques in deep learning. Its main function is to process the input data into standard data with the same order of magnitude, so as to effectively prevent the problems of "gradient disappearance" and "gradient explosion." BN makes each forward propagation output in the same order of magnitude, eliminates the difference of magnitude between layers, and makes the weight adjustment more reasonable in the back-propagation calculation. BN operation makes the calculated results of each layer conform to the standard normal distribution, as shown in the following Eq. 8:

$$x_i' = \frac{x_i - \bar{x}}{x_{var}} = \frac{x_i - \frac{1}{n} \sum_{j=1}^{n} x_j}{\sqrt{\frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2}}$$

2 Dropout
Dropout can be used as a kind of trick to train deep neural network. In each training batch, the over fitting phenomenon can be significantly reduced by ignoring half of the feature detectors (let half of the hidden layer node value is 0). This method can reduce the interaction between feature detectors (hidden layer nodes). Detector interaction means that some detectors rely on other detectors to work. As shown in Fig. 8, it is the working schematic diagram of Dropout.
4.3. Experimental results and evaluation

In this part, we use 600 sets of vibration data to test the improved model. As shown in Figure 9, it is the confusion matrix generated by the test results.

Through the statistical classification results, we can get Table 2:

|               | NORMAL | BALL  | INNER RACE | OUTER RACE |
|---------------|--------|-------|------------|------------|
| NORMAL        | 297    | 1     | 0          | 0          |
| BALL          | 0      | 148   | 1          | 0          |
| INNER RACE    | 3      | 1     | 49         | 2          |
| OUTER RACE    | 0      | 0     | 0          | 98         |

It can be seen from table 2 that the algorithm proposed in this paper has passed the test of 600 groups of data, and the accuracy rate is 98.67%, which is an ideal result. In order to further reflect the advantages of this algorithm, several advanced algorithms are selected for comparison. The comparison results are shown in Table 3.
### Tab. 3 Comparison of experimental results

|          | OURS  | CNN   | SVM   | BPNN  |
|----------|-------|-------|-------|-------|
| ACCURACY | 98.67%| 94.78%| 78.66%| 97.14%|

Through the comparison, we can see that the accuracy of this algorithm is not only better than SVN and other traditional algorithms, but also due to the advanced deep learning network algorithm [17-19].

### 5. Conclusions

This paper presents an improved CNN for bearing fault detection. In fault detection, the most important problem to be solved is the low detection accuracy. Our detector proposes an adaptive ROI Align method, which makes the network obtain the feature map more quickly and accurately. At the same time, the improved *Swish* function is used to further improve the network detection performance. On the other hand, we propose a new equation to segment the training samples. It can make the filter that we train more reliable. The experimental results show that the tracker shows its advantages in the test set and can effectively classify the fault data. A large number of experiments on a large benchmark platform show that the proposed algorithm has better performance than the existing algorithms.

### References

[1] Haidong Shao, Hongkai Jiang, Ying Lin, Xingqiu Li. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders [J]. Mechanical Systems and Signal Processing, 2018, 102.

[2] Weixiang Gu, Xianfa Wu, Jingnan Xu, et al. Fault Diagnosis of Fan Gearbox Bearing Based on DPSO-MKELM Algorithm [J]. Shandong Electric Power, 2019.

[3] Zhang Y H, Li C, Jing H, et al. Roller Element Bearing of Mine Ventilating Fan with Fault Diagnosis Based on Mechanics Properties and RBF Neural Network [J]. Advanced Materials Research, 2012, 625:125-129.

[4] Xie L, Miao Q, Chen Y, et al. Fan bearing fault diagnosis based on continuous wavelet transform and autocorrelation[J]. 2012.

[5] Zhang Y H, Li C, Jing H, et al. Roller Element Bearing of Mine Ventilating Fan with Fault Diagnosis Based on Mechanics Properties and RBF Neural Network [J]. Advanced Materials Research, 2012, 625:125-129.

[6] Bu Y, Wu J, Ma J, et al. The rolling bearing fault diagnosis based on LMD and LS-SVM [C]// Chinese Control & Decision Conference. IEEE, 2014.

[7] Levent E, Turker I, Serkan K. A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier [J]. Journal of Signal Processing Systems, 2018.

[8] Pan, Honghu, He. An Improved Bearing Fault Diagnosis Method using One-Dimensional CNN and LSTM [J]. Journal of Mechanical Engineering, 2018.

[9] Zhang Wenfeng, Zhou Jun. Fault Diagnosis Method of Rolling Bearing Based on Dropout-CNN [J]. Light Industry Machinery, 2019, 037(002):62-67.

[10] Wang D, Guo Q, Song Y, et al. Application of Multiscale Learning Neural Network Based on CNN in Bearing Fault Diagnosis [J]. Journal of Signal Processing Systems, 2019, 91(4).

[11] An F P. Rolling Bearing Fault Diagnosis Algorithm Based on FMCNN-Sparse Representation [J]. IEEE Access, 2019, 7:102249-102263.

[12] Qi-Yang W, Bang-Shu X, Xin-Min L I, et al. Fault Diagnosis for Rolling Bearing of Swashplate Based on Multi-noise Data Training CNN [J]. Journal of Nanchang Hangkong University(Natural Sciences), 2019.

[13] Xu J, Ma L, Zhang W, et al. An Improved Hybrid CNN-SVM based Method for BearingFault Diagnosis Under Noisy Environment [C]// Chinese Control & Decision Conference. 2019.
[14] Xu J , Ma L , Zhang W , et al. An Improved Hybrid CNN-SVM based Method for Bearing Fault Diagnosis Under Noisy Environment [C]// Chinese Control & Decision Conference. 2019.

[15] Fuan W , Hongkai J , Haidong S , et al. An adaptive deep convolutional neural network for rolling bearing fault diagnosis[J]. Measurement Science and Technology, 2017.

[16] Jian X , Li W , Guo X , et al. Fault Diagnosis of Motor Bearings Based on a One-Dimensional Fusion Neural Network [J]. Sensors, 2019, 19(1).

[17] Wang H , Xu J , Yan R , et al. A New Intelligent Bearing Fault Diagnosis Method Using SDP Representation and SE-CNN [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(5):2377-2389.

[18] Li X U , Ge L I , Liang Y U , et al. Roller Bearing Fault Diagnosis Based on Wavelet Packet and CNN [J]. Journal of Sichuan University of ence & Engineering(Natural ence Edition), 2018.