A novel low-cost bearing fault diagnosis method based on convolutional neural network with full stage optimization in strong noise environment

Li Jiang  
Wuhan University of Technology

Zhipeng Yu  
Wuhan University of Technology

Kejia Zhuang  
Wuhan University of Technology

Yibing Li  (ahlyb@whut.edu.cn)  
Wuhan University of Technology

Research Article

Keywords: Roll bearing, Fault diagnosis, Convolutional neural network, Support vector machine, Lightweight

Posted Date: November 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2287761/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
**Title**  
A novel low-cost bearing fault diagnosis method based on convolutional neural network with full stage optimization in strong noise environment

**Author information**  
Li Jiang ¹, ²(0000-0003-3644-436X), Zhipeng Yu ¹(0000-0002-8664-9163), Kejia Zhuang ¹, Yibing Li ¹, ²(0000-0002-6532-6461)  
(¹ School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070; ² Hubei Digital Manufacturing Key Laboratory, Wuhan University of Technology, Wuhan 430070)  
**Corresponding author:** Yibing Li (e-mail: ahlyb@whut.edu.cn).

**Abstract**  
In recent years, convolutional neural network (CNN) has been successfully applied in the field of bearing fault diagnosis. So as to improve the diagnosis performance in harsh environment with strong noise, the structure of CNN-based feature extractor becomes deeper and more complex. However, with the increase of depth, the model may lose shallow features and the training parameters will surge. Moreover, if the sample size is not large, it tends to over fit. It deviates from the concept of network lightweight. On the other hand, little attention will be paid to the optimization of model classifiers which can significantly improve the classification performance. Therefore, we proposed a CNN with full stage optimization (FSOCNN) model for bearing fault diagnosis in strong noise environment. In the feature extraction stage, the model is optimized with a novel multi-feature output structure connected with global average pooling to improve the feature extraction ability without any extra trainable parameters. In the classification stage, the traditional softmax layer will only participate in the parameter optimization of CNN model through gradient descent algorithm, and the diagnosis results will be output by support vector machine. The effectiveness of the proposed method is verified on the two bearing datasets under different levels of noise. Compared with the existing five fault diagnosis models, the results prove that the proposed method possesses higher accuracy, less training time, and better stability.

**Keywords**  
Roll bearing · Fault diagnosis · Convolutional neural network · Support vector machine · Lightweight

**Statements and Declarations**  
**Conflict of interest:** We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work. There is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, ‘A novel low-cost bearing fault diagnosis method based on convolutional neural network with full stage optimization in strong noise environment’.

**Funding:** This study was funded by the Fundamental Research Funds for Hubei Province Natural Science Foundation of China (2019CFB565) and the National Natural Science Foundation of China (Nos. 51775391, 51705384 and 51875430).
Acknowledgments
The authors would like to thank Case Western Reserve University and Universität Paderborn for providing free access to the bearing vibration experimental data from their websites.
1 Introduction

Rotating machinery plays a crucial role in industrial production, accounting for more than 80% of industrial machinery[1]. Rolling bearing is an indispensable component in all rotating machinery, which represents the main source of faults in such devices[2]. In case of failure, it will seriously affect the safe and stable operation of the machinery[3]. In serious cases, it will lead to huge economic losses and even casualties[4]. Thus, the accurate diagnosis of bearing faults is conducive to improve the safety and stability of industrial production.

Thanks to the development of sensor technology, a large amount of bearing state information is obtained easily. Therefore, the traditional data-driven intelligent diagnosis methods have developed rapidly[5], such as support vector machine (SVM)[6], extreme learning machine (ELM)[7], etc. However, it is difficult for these shallow models to learn nonlinear features from complex vibration signals, especially under severe operating conditions. As a result, the diagnosis performance of the shallow models will largely depend on the artificial features extracted by prior knowledge and expert experience[8]. So as to realize automatic fault feature extraction and excellent diagnosis performance, the deep learning method has become a priority alternative. At present, due to its superior feature extraction ability and performance in classification task, lots of researchers apply deep learning method to bearing fault diagnosis[9], such as auto-encoder (AE)[10], deep belief network (DBN)[11], convolutional neural network (CNN)[12,13], etc. Among them, CNN is the most widely utilized to solve various complex problems in bearing fault diagnosis field, thanks to its flexible structure and less trainable parameters brought by weight sharing mechanism. Chen et al.[14] combined CNN with extreme learning machine for gear and bearing fault diagnosis. Chen et al.[15] input two-dimensional time-frequency diagram into LeNet-5 network for bearing fault diagnosis under imbalanced data and different operating conditions. Xu et al.[16] proposed an improved multi-scale CNN with feature attention mechanism to extract and fused features for bearing fault classification.

The above CNN-based approaches have achieved great success in bearing fault diagnosis. However, the performance will decline dramatically if unexpected noise appears. Unfortunately, the actual operating environment of the fault bearings often contains a lot of white noise. The fault features of the collected vibration signals are usually submerged in the complex vibration signals with strong background noise and interference components, which makes the sensitive fault feature extraction extremely difficult[17]. Zhang et al.[18] and Wang et al.[19] proved that the signal contaminated by strong noise will lead poor diagnosis performance. To overcome this problem, researchers have made efforts. Han et al.[20] integrated the denoise algorithm called non-local means into auto-encoder to denoise. Zhang et al.[18] proposed a one-dimensional CNN (1D CNN) with a wide convolution kernel and a small batch training strategy to improve the performance of the model in a noisy environment. Huang et al.[21] utilized 1D CNN with multiscale convolution kernel to extract more features from vibration signals and prevent the model from over fitting in noisy environment. Similarly, Peng et al.[22] proposed a multibranch and multiscale CNN to improve the
feature extraction ability. Ko et al.[23] integrated classification and denoise into one model. This method connected LSTM with one output of CNN for signal noise reduction to improve the anti-noise ability of the model. Sun et al.[1] utilized frequency slice wavelet transform to generate time–frequency images and put them into an efficient multi-scale CNN for fault diagnosis in a noisy environment.

From the above anti-noise methods, it can be found that the model optimization method is often used to improve the anti-noise performance and enhance the feature extraction ability of the model, especially the optimization method with multi-scale convolution structure, such as [1,21,22]. However, these optimization methods increase a large number of trainable parameters and calculating costs. This deviates from the recent concept of model lightweight[24]. On the other hand, it can be found that a large number of methods consider optimizing the feature extraction stage, but the optimization classification stage is rarely considered. In the noisy environment, the sensitive fault features cannot be completely exploited in the feature extraction stage. This can be proved from the t-SNE figures of [1,18]. Thus, this makes it difficult for the traditional softmax algorithm to distinguish them correctly and limits the improvement of the traditional CNN-based model diagnosis accuracy.

To address those problems mentioned above, we proposed a CNN with full stage optimization (FSOCNN) for bearing fault diagnosis. Firstly, a novel multi-feature output (MFO) structure is designed to enhance the feature extraction ability of the model without any extra trainable parameters. Furthermore, we replace fully connected (FC) layer with global average pooling (GAP) layer to reduce the parameters and improve the anti-noise performance of the model. As a result, the feature extraction stage has been optimized. Then, we optimize the classification stage of the model by replacing softmax layer with the traditional machine learning algorithm SVM. The experimental results on the two bearing datasets show that the classification performance of the proposed FSOCNN algorithm is improved and the training time is further shortened.

The rest of this paper is organized as follows. Section 2 briefly introduces composition and theory of CNN. Section 3 introduces the proposed FSOCNN model and the details of the proposed bearing diagnosis method based on FSOCNN. The experimental process and the superiority validation of the proposed method are shown in Section 4. Finally, the conclusions are provided in Section 5.

2 Brief theory

2.1 One-dimensional convolutional neural network

Convolutional neural network consists of feature extraction stage and classification stage[18]. The feature extraction stage includes convolution layer, activation layer, and pooling layer. The classification stage is composed of several fully connected layers. The feature extraction stages will be described in detail in this section.
2.1.1 Convolution layer

The convolution layer convolutes the input signal with a certain stride through the convolution kernel to form the feature output after the activation operation. Each layer of convolution contains multiple channels. The convolution kernel of each channel is the same, which is usually called weight sharing. Each channel convolutes with all input signals and generates a feature output. Finally, the number of features output is the same as the number of channels. The calculation formula of convolution is as follows:

\[ y^{(m,n)} = w * x^{(m,n)} = \sum_i \sum_j w^{(i,j)} * x^{(m+i,n+j)} \]  

(1)

where \( w \) is the weight of the convolution kernel, \( x \) is the input signal, and \( (m,n) \) represents the position where the convolution operation starts.

2.1.2 Activation layer

The activation operation is set after the convolution operation. It is a nonlinear function, which can enhance the feature expression of convolution layer the nonlinear transformation of input. The commonly used activation functions behind the convolution layer are Sigmoid function, Tanh function and ReLU function. ReLU function can make the convolution neural unit zero, resulting in more sparse feature expression, which is helpful to accelerate the convergence of the model. Thus, in this paper, the ReLU function is used as the activation unit. The formula of the ReLU function is as follows:

\[ y = \max(0, x) \]  

(2)

where \( x \) is the output of convolution operation.

2.1.3 Pooling layer

The pooling layer is added after the convolution layer to compress the feature map. Maximum pooling is widely used in CNN. It will keep the maximum value in a specific range and discard other values to reduce the feature dimension. The calculation formula of maximum pooling is as follows:
\[ y = \max \{ x^i | j \leq i \leq j + w \} \]  

(3)

where \( j \) represents the position where the pooling operation starts, \( w \) is the kernel width of pool layer.

### 2.2 Support vector machine

SVM is rooted in the statistical learning theory and based on the structural risk minimization (SRM) principle, which is designed to minimize error on the training data set[25]. This makes SVM have good generalization ability. SVM attempts to find a hyperplane to segment two samples with different labels, so as to solve the problem of secondary classification. In the sample space, SVM will be trained by maximizing the distance from the sample boundary of different labels to the hyperplane. The selection of hyperplane is optimized in the following ways:

\[
\text{minimize } \frac{1}{2} ||w||^2 \tag{4}
\]

subject to

\[
y^i \ast (w^T \ast x^i + b) \geq 1, \quad i = 1, 2, ..., M \tag{5}
\]

where \( x^i \) is the sample data, \( y^i \) is the homologous label, \( w \) is a \( M \)-dimensional vector, \( b \) is a scaler.

### 3 The proposed FSOCNN model for bearing fault diagnosis

This section proposes a novel fault diagnosis method based on CNN with full stage optimization (FSOCNN) for bearing fault diagnosis. Optimized feature extraction stage can extract more features from noise signals for classification stage. It is worth noting that this optimization method does not bring too many trainable parameters. In the classification stage, we use SVM instead of the traditional softmax layer to improve the classification performance of the model. The optimization method of feature extraction and the training method of the proposed model will be introduced in detail.

**Fig. 2** The model structure of proposed FSOCNN

#### 3.1 Optimization method of feature extraction stage

1DCNN can deal with one-dimensional time series signals well. In the fault diagnosis
field, vibration signal samples can be directly input into 1DCNN without any conversion processing. Therefore, in this paper, a selected 1DCNN will be optimized. A typical 1DCNN is shown in Fig.1. The network consists of several convolution layers and pooling layers stacked with each other, and finally completes the classification task through several fully connected layers. However, in the noisy environment, the fault features in the collected time-domain signals are submerged in the noise, and leads to the poor performance of conventional 1DCNN. Therefore, we design the following optimization methods to improve the feature extraction ability of the model in noisy environment. Firstly, the wide convolution kernel is set at the first layer of the network; Secondly, the multi-feature output (MFO) structure is designed to enrich the features output in the feature extraction stage; Third, the globe average pooling (GAP) layer replaces the fully connected layer to further improve the anti-noise performance and reduce the trainable parameters of the model.

3.1.1 Wide convolution kernel

The first convolution layer with wide kernel was first utilized in the fault diagnosis field by [18]. It effectively improves the fault diagnosis performance of 1DCNN. This is because the fault pulse exists in the whole signal at a certain period. It leads that the vibration signal has strong correlation and a small convolution kernel will easily cause model overfit. On the other hand, wide convolution kernel can improve the receptive field of the model as well. This is also a reason why wide convolution can improve the performance of the model. Therefore, in this paper, we increased the kernel width of the first convolution layer to 64 for the first step of optimization.

3.1.2 Multi-feature output

A small size model will be under fitting when the data characteristics are inapparent due to the lack of nonlinear fitting ability of shallow network. In order to mine the deep features of the input signal and enhance the nonlinear fitting ability of the model, the network depth is increased. However, another problem arises. Within a certain range of sample number, with the increase of network layers, the model over fitting becomes more and more serious, especially in noisy environment. As a result, the accuracy of the test set is much lower than that of the training set. To solve this problem, optimization methods such as multi-scale convolution kernel, multi branch parallel structure and so on are proposed. Although these methods inhibit the over fitting of the model under the noisy condition to a certain extent, they introduce too many training parameters. Therefore, to improve the anti-noise ability and limit the size of the model, a novel multi-feature output (MFO) structure is proposed. The MFO method is shown in Fig.3.
As shown in Fig. 3, the feature map of each pooling layer is output to the feature fusion layer. This means that the features obtained by each convolution and pooling will be used to directly classified. This helps the classifier distinguish different faults from different feature levels, rather than just the deepest level, and suppress the model overfitting. As can be seen from Fig. 3, we only directly extracted the feature map of each pool layer without adding any trainable parameters, which maintains the model lightweight.

### 3.1.3 Feature fusion based on globe average pooling

For traditional CNN, fully connected layer is usually used to fuse the final output feature map of convolution and pooling stages via serial connection. Fig. 4(a) shows the fully connected method. However, MFO will obtain multiple feature maps, as shown in Fig. 3. If fully connected layer is adopted in this paper, it will lead to a sharp increase of the feature length, even much longer than the length of the original signal. This will indirectly produce a large number of trainable parameters when connecting with the dense layer, which is contrary to our original intention of lightweight. What's worse, the fully connected layers are prone to overfitting[26]. Fortunately, the global average pooling layer has become a good choice to solve this problem. It was proposed by Lin et al.[26], which connects the average of each feature map. Fig. 4(b) shows the GAP method. It limits the surge of feature length and has no trainable parameters to optimize. As a result, the overfitting is avoided. Thus, GAP is selected to fuse the feature map in this paper.
3.2 Fault classification based on SVM

After finishing feature fusion, one-dimensional features for classification are obtained. The fault categories represented by these features are identified by the hyperplane of SVM. The method further restrains the over fitting and accelerates the convergence of the model. This is because during training, the loss value will be 0 and the parameter update will stop, once the distance between the feature and the hyperplane is over 1. This avoids the continuous updating of parameters caused by softmax whose output probability of target label is difficult to reach 1, especially when the features are not significant in the noise environment. Thus, SVM is adopted in the classification stage.

3.3 Bearing fault diagnosis based on the proposed FSOCNN

After the optimization scheme of each stage of the model is determined, we establish the full stage optimization model FSOCNN displayed in Fig.2. One-dimensional vibration signals are input into the model. Four convolution and pooling layers connect each other to extract features. The parameters of convolution and pooling layers are shown in Table 1. The first, second and last feature maps of pooling layer are compressed by GAP. After that, these features are concatenated via dense layer. The number of dense nodes corresponding to MF1, MF2 and MF3 branches is 50, 50 and 32, respectively. Therefore, the output feature length of the feature extraction stage is 132. These features will be put into SVM for classification.

The diagnostic process based on the proposed FSOCNN is shown in Fig.5. The diagnostic steps are as follows:

Step1: Collect the vibration signals of bearings with different health condition via acceleration sensor.

Step2: The noise data is added to the collected raw signals, and then the synthesized signals will be divided into training set and testing set in certain proportion.

Step3: The training set is used to train FSOCNN model. The training parameters are set as follows. The batch size is 64. The learning rate of Adam optimizer is 0.05. The loss function is categorical cross entropy. The kernel function of SVM is linear. Note that the softmax layer is still used to optimize the feature extraction stage via back
propagation algorithm. This is because SVM cannot participate in the gradient descent process of CNN. SVM will be trained by the output features of feature extraction stage.

Step4: The testing set is used to verify the diagnosis performance of FSOCNN model, and the result will be displayed.

**Table 1** The parameters of convolution and pooling layers

<table>
<thead>
<tr>
<th>Layer</th>
<th>Kernel size</th>
<th>Stride</th>
<th>Filter number</th>
<th>Activation function</th>
<th>Output size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv1</td>
<td>64</td>
<td>4</td>
<td>16</td>
<td>ReLU</td>
<td>(256, 16)</td>
</tr>
<tr>
<td>MaxP1</td>
<td>3</td>
<td>2</td>
<td>\</td>
<td>\</td>
<td>(127, 16)</td>
</tr>
<tr>
<td>Conv2</td>
<td>7</td>
<td>2</td>
<td>32</td>
<td>ReLU</td>
<td>(64, 32)</td>
</tr>
<tr>
<td>MaxP2</td>
<td>3</td>
<td>2</td>
<td>\</td>
<td>\</td>
<td>(31, 32)</td>
</tr>
<tr>
<td>Conv3</td>
<td>5</td>
<td>1</td>
<td>64</td>
<td>ReLU</td>
<td>(31, 64)</td>
</tr>
<tr>
<td>MaxP3</td>
<td>3</td>
<td>2</td>
<td>\</td>
<td>\</td>
<td>(15, 64)</td>
</tr>
<tr>
<td>Conv4</td>
<td>3</td>
<td>1</td>
<td>128</td>
<td>ReLU</td>
<td>(15, 128)</td>
</tr>
<tr>
<td>MaxP4</td>
<td>3</td>
<td>1</td>
<td>\</td>
<td>\</td>
<td>(13, 128)</td>
</tr>
</tbody>
</table>

![Fig.5](image-url) The fault diagnosis method based on FSOCNN

**4 Experimental validation**

This section contains three parts. Firstly, two datasets used in the experiment are introduced. Second, one of the datasets is used as experimental data to verify the effectiveness of the proposed optimization methods. Thirdly, the proposed model is compared with several existing models.
4.1 Introduction of experimental data

Datasets from Case West Reserve University (CWRU)[28] and Universität Paderborn (UPB) [29] were used in this experiment. CWRU data will be used to verify the validity of MFO, GAP and SVM in the proposed FSOCNN algorithm. In fact, the CWRU dataset is used by a large number of researchers for model validation, so it is used appropriately to compare the proposed model with other existing research results and highlight the advantages of the proposed model. UPB dataset will be used in part 3 of this section for further model validation.

4.1.1 CWRU Data

The original experimental data was obtained from the accelerometers of the motor driving mechanical system at a sampling frequency of 12 kHz from the CWRU Bearing Data center. Fig.6 exhibits the test rig.

![Fig.6 The CWRU bearing fault test rig](image)

Besides the normal state, the test rig collected the bearing data of three fault categories under four different working conditions (1797 rpm, 1772 rpm, 1750 rpm and 1730 rpm). The fault categories are inner-race fault (IF), outer-race fault (OF) and rolling ball fault (BF). Three severity levels exist in each kind fault bearing. The fault diameter is 0.007 inches, 0.014 inches and 0.021 inches, respectively. Thus, there are ten health states, and we adopt all health states bearing vibration data under 1797r/min condition as sample data. In this experiment, 1024 data points are contained in each sample. Each health category contains 200 samples, of which 70% are divided into training samples and the rest are used as testing samples. Table 2 shows the details of all the datasets.

<table>
<thead>
<tr>
<th>Label</th>
<th>Health State</th>
<th>Train Number</th>
<th>Test Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>1</td>
<td>Ball Fault 007 (BF7)</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>Ball Fault 014 (BF14)</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>Ball Fault 021 (BF21)</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>Inner Fault 007 (IF7)</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>Inner Fault 014 (IF14)</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>Inner Fault 021 (IF21)</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>Outer Fault 007 (OF7)</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>Outer Fault 014 (OF14)</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>9</td>
<td>Outer Fault 021 (OF21)</td>
<td>140</td>
<td>60</td>
</tr>
</tbody>
</table>
4.1.2 UPB Data

The open bearing database of UPB is collected from the test rig consisted of six modules that are shown in Fig.7. The raw bearing vibration data are obtained from the piezoelectric accelerometer mounted at the top end of the rolling bearing module at a sampling frequency of 64 kHz. In this experiment, we select the test condition with 1500 rpm rotational speed, 0.1Nm load torque and 1000N radial force. The artificial fault bearings are made by the methods of electrical discharge machining (EDM), electric engraver and drilling separately or jointly. Three fault types are processed, and they are single fault with one artificial crack, repetitive fault with multiple identical artificial cracks and multiple faults with multiple cracks made by different methods separately.

Fig.7 The UPB bearing test rig

A normal state, two types of inner-race fault and two types of outer-race fault are carried out by the test rig. Each fault type contains 200 samples, and each sample contains 1024 points. The details of experimental data are displayed in Table 3. There are five fault types totally, including normal state (NS), multiple outer fault (MOF), single outer fault (SOF), repetitive inner fault (RIF) and single inner fault (SIF) separately. Fig.8 shows part of the time-domain signals of each health state. Significantly, the 64 kHz sampling frequency is far higher than the 12 kHz of CWRU rig at about the same motor speed. This means that the low-frequency fault feature existing in each fault sample is less. It leads the model over-fitting easily especially under noise condition that the fault features are masked by noise signals.

<table>
<thead>
<tr>
<th>Label</th>
<th>Fault types</th>
<th>Damage types</th>
<th>Train Number</th>
<th>Test Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal</td>
<td>/</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>1</td>
<td>Outer Fault</td>
<td>multiple</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>Outer Fault</td>
<td>single</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>Inner Fault</td>
<td>repetitive</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>Inner Fault</td>
<td>single</td>
<td>140</td>
<td>60</td>
</tr>
</tbody>
</table>
4.1.3 Generated Noise data

In order to verify the advantages of the model in noise environment, we add additive white Gaussian noise to the original dataset provided by CWRU and UPB to synthesize noise signals with different SNR. The definition of SNR is as follows:

$$SNR_{dB} = 10 \log_{10} \left( \frac{P_{signal}}{P_{noise}} \right)$$

where $P_{signal}$ and $P_{noise}$ are the power of signal and the noise respectively.

According to equation (4.1), we synthesized the noise signals ranging from -6dB to 8dB respectively. To speed up model training, z-scores normalization is used to preprocess sample data. Fig.9 exhibits the time domain normalization signal of each label sample at -4dB. It can be seen that in the high noise environment, the fault features are masked by the noise signal, so it is challenging to extract features quickly and correctly.
4.2 The efficiency of the proposed optimization methods

To expound the necessity of the proposed structure, five models are designed and employed on the CWRU datasets for a comparison. The structures of all models are displayed in Fig. 10. The details of each model are introduced as follow:

(1) 1DCNN with a fully connected layer. The parameters of convolution and pooling layers are same as shown in Table 1. It is the most basic framework of the proposed model, which serves as a reference model to reflect the specific gain of other models.
(2) 1DCNN with global average pooling layer (GCNN). To reveal the role of GAP separately, we only replace the fully connected layer of the first model with a GAP layer, and keep other parameters of the same as the first model.

(3) GCNN connected with SVM (GCNN-SVM). We replaced the softmax layer of GCNN with SVM.

(4) GCNN with MFO structure (MGCNN). We connect a GAP layer between each of the first two pooling layers and dense layer of GCNN, so that we can get MGCNN model to verify the effect of the proposed MFO structure.

To avoid the contingency of experiment, we repeat 20 times for each model on CWRU data. Fig. 11 displays the results of each model under different noise condition from -6dB to 8dB. All experiments are conducted in the equipment environment as shown in Table 4.

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Experimental equipment and environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>i5-6700U</td>
</tr>
<tr>
<td>GPU</td>
<td>\</td>
</tr>
<tr>
<td>Memory Size</td>
<td>4G</td>
</tr>
<tr>
<td>Framework</td>
<td>TensorFlow 2.3.1</td>
</tr>
<tr>
<td>Language</td>
<td>Python 3.8.5</td>
</tr>
</tbody>
</table>

From the performance of GCNN and 1DCNN in Fig.11, it can be seen that the anti-noise ability of GCNN with GAP structure far exceeds that of 1DCNN with fully connected structure. Especially under strong noise conditions, such as -4dB and -6dB, the accuracy of GCNN with GAP model is more than 10% higher than that of 1DCNN.

According to the comparison results between the GCNN model and MGCNN model separately, it looks like that MFO structure does not produce notable benefit due to almost the same performance. However, from the manifestation of GCNN-SVM model, it clearly shows that SVM has more ascendant classification capacity than softmax layer. It is due to the fact that the feature classification ability of softmax layer
is limited leading to the wastage of feature produced by MFO. Besides, in the light of the result of the proposed FSOCNN algorithm, it also proves that MFO can extract more significant feature from noise signals.

On model training time scale, although GCNN with GAP possess appreciable anti-noise effect, it cost too much time to get optimal solution of gradient descent. On the other hand, GCNN-SVM model cost less convergence time than GCNN model. This proves that SVM is quicker in convergence than softmax layer. What’s more, the more important trait of MFO is that it only costs about 2 more seconds, besides more features MFO can extract from different aspects of signals.

In addition to the accuracy, stability is also an important attribute of the model. Fig.12 shows the boxplot of accuracy obtained from 20 tests of each model under -4dB. It can be seen distinctly that the proposed shows the highest accuracy and excellent stability. Compared with the traditional 1DCNN, each structure we proposed brings ideal benefits. GAP represents strong noise resistance, but its stability is insufficient. MFO improve the stability and accuracy of the model by extracting multi-level features from noise signals. What’s more, SVM further improves the accuracy of the model owe to its excellent classification ability.

![Fig.12](image)

**Fig.12** Box point diagram of each model under -4dB

### 4.3 Comparison with existing methods

In order to further prove the superiority of the proposed model, five classification algorithms including a typical machine learning method (SVM), two typical CNN based methods (WDCNN and LeNet-5) and two existing anti-noise methods (MC-CNN and NLB-CNN) are applied on the same datasets for comprehensive comparison. We strictly follow the structure of the contrast model in the reference, and select the
appropriate parameters for each model. The detailed parameters of contrast models are introduced as follow:

(1) SVM. SVM with Radial Basis Function kernel is adopted to handle multiple classification. To obtain the optimal gamma value, multiple parameter values $[1, 0.1, 0.01, 0.001]$ are evaluated respectively. Finally, the gamma value is set equal to 0.01.

(2) LeNet-5. In recent, LeNet-5 proposed by LeCun et al.[27] has been employed by researchers in the field of fault diagnosis. The model parameters is the same as the citation[27].

(3) WDCNN. A classical 1DCNN has been widely applied to fault diagnosis. It is proposed by Zhang et al.[18]. The model parameters is the same as the citation[18].

(4) MC-CNN. The multi-scale convolution kernel is adopted to the first convolution layer of 1DCNN. [21] proposed MC-CNN that perform superiorly in four classification fault diagnosis under noise condition. The multi-scale convolution number of the first layer is $[126, 64, 16]$, and the parameters of other layers are the same as 1DCNN proposed in Section 4.2.

(5) NLB-CNN. The non-local block (NLB) structure is proposed by Han et al.[20]. Han puts NLB into the Auto-Encoder (AE) model for anti-noise and the effect is remarkable. We creatively utilize NLB in GCNN for fault diagnosis directly. The NLB is placed behind the first pooling layer of GCNN. We reduce a convolution and pooling layer of GCNN to ensure that NLB-CNN depth is the same as other contrast models. The CNN parameters of NLB-CNN are the same as the first three layers of the proposed model. And the parameters of NLB is the same as the citation[20].

The trainable parameter numbers of the deep learning models are shown in Table 5. The NLB-CNN has the least trainable parameters thanks to the replacement of NLB to a convolution and pooling layer. Multi-scale structure increases the memory size of model with 247994 trainable parameters. In contrast, only 48102 trainable exists the proposed model. This is because that MFO structure only fuses the feature output of each layer without introducing other parameters.

**Table 5** The trainable parameters of models

<table>
<thead>
<tr>
<th>Model</th>
<th>LeNet-5</th>
<th>WDCNN</th>
<th>MC-CNN</th>
<th>NLB-CNN</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainable Parameters</td>
<td>61,968</td>
<td>41,990</td>
<td>247,994</td>
<td>14,146</td>
<td>48,102</td>
</tr>
</tbody>
</table>

All models will be trained and tested under CWRU and UPB datasets respectively. The comparative analysis and visual analysis of the model will be introduced in this section.

### 4.3.1 Data: CWRU

To contrast the performance of each model in noise environment, all models are tested at -4dB respectively. Each model is trained and tested at the same sample data that is shown in Table 2. In addition, it is necessary to reshape sample size from $(1024, 1)$ into $(32, 32, 1)$ for LeNet-5 training and testing. When the training loss value of each model
no longer keeps dropping, the model training will be completed. To avoid the one-time occasionality in results and ensure the validity of the test, each experiment is repeated 20 times. The experimental results are shown in Fig.13.

As shown in Fig.13, SVM only achieves around 62.67% accuracy. However, methods based on deep learning perform superiorly under strong noise condition, with the highest accuracy being over 85%. Especially, it is the outstanding results that the proposed method achieves up to 97.08% accuracy. This proves that the proposed FSOCNN can extract more pivotal features from high noise signal than traditional methods. On the other hand, the performance of LeNet-5, WDCNN, MC-CNN and NLB-CNN that utilize softmax layer as the classification layer is instable in 20 experiments. In contrast, the accuracy variance of the proposed method with SVM is only 0.93, far less than that of the other models without SVM. This illustrates that SVM can always correctly distinguish the fault category from the in-depth features, although its feature extraction ability is not as excellent as the deep learning method.

![Fig.13 Diagnosis results of six models under -4dB](image)

As can be seen from Fig.13, the training time of LeNet-5 and MC-CNN is 38.29s and 32.68s which are nearly twice that of WDCNN (18.94s). This is because the size of the convolution kernel in the first layer of LeNet-5 is small, which leads to the increase of the amount of computation required for feature extraction in each layer. And compared with WDCNN, MC-CNN has three convolution kernels with different widths in the first layer to extract features in parallel, which leads to the multiple increase of the trainable parameters and the feature length in the later layers, thus increasing the cost of the model training. What’s more, NLB-CNN training time is 188.02s which is far more than other models due to large-dimensional matrix operation, despite its decent performance in the noise conditions. By comparison, the proposed model training only costs 13.2s and perform bests thanks to the splendid classification ability of SVM and
the low-cost feature extraction method of MFO.

The confusion matrixes are shown in Fig.14. The normal bearings are correctly distinguished by each model, while the performance of BF recognition is terrible. Only NLB-CNN and the proposed model have achieved superior performance with the average recognition accuracy being more than 90%. Moreover, the IF and OF recognition performance of anti-noise model MC-CNN and NLB-CNN are more sterling compared to that of the conventional CNN architecture LeNet-5 and WDCNN. In particular, the proposed models have reached 100% recognition accuracy except for BF recognition.

![Fig.14](imageestationary)
To show the extracted features of each model before classification, the t-SNE diagrams of the penultimate layer of each model are display in Fig. 15. In addition, the feature fusion effect of the proposed MFO is also shown in Fig. 15(b). It can be seen from Fig. 15(a) that only NLB-CNN in the comparison models shows preferable feature discrimination ability, and the various features have been basically identified before classification layer. Furthermore, the feature separation level of the proposed model is the most palmary in all the contrast models.

From the three feature outputs extracted by the proposed model in Fig. 15(b), it can be seen that part of category features are gradually correctly separated with the increase of the model depth, such as the features of the OF (label 7, 8&9). On the other hand, some features will be misidentified as other types owing to over fitting concurrently, such as the features of IF14 (label 5). However, the label5 features in the red box are separated out again after the features fusion of MFO, which can be seen from the feature cluster diagram of concatenate layer. This proves that the proposed MFO structure can effectively alleviate the over fitting problem.

Fig. 15: t-SNE cluster diagram. (a) the penultimate layer of contrast models; (b) each output of MFO and concatenate layer of the proposed

Fig. 16: Feature diagram of each fault type: (a) NLB-CNN; (b) Proposed FSOCNN.

To intuitively see the features before classification, we draw the feature map of the
penultimate layer of each fault type. As shown in Fig.16, the fault expression level of the features extracted by the proposed model is higher. The features of each fault are more recognizable. This is an important reason for the model to achieve high accuracy and excellent stability. Moreover, the feature length shown in Fig.16(b) is 132. Features [1, 50], [51, 100] and [101, 132] are obtained from the feature outputs of MF1, MF2 and MF3 respectively. The outputs of MF2 are included in the fault features composition of BF, IF21, OF7 and OF14. This proves that shallow features also have a certain value of fault feature expression.

4.3.2 Data: UPB

The signals at -6dB, -3dB and 0dB are tested by six contrast models. Data samples are shown in the Table 3. All of model parameters are the same as those introduced in section 4.3.1. We tested each model 20 times respectively under each noise condition. The experimental results are shown in Fig.17 and the confusion matrix of each model tested at -6dB is displayed in Fig.18.

The proposed model achieves the highest accuracy and stability under all noise conditions. From the confusion matrix in Fig.18, it can be seen that the accuracy of contrast models in IF recognition is unsatisfactory, while the recognition accuracy of the proposed model in all fault categories is more than 90%. As shown in Fig.16(b) and Fig.19(b), compared with CWRU data, shallow features account for more in the fault feature composition of UPB data.

![Fig.17 Diagnosis results of six models under each noise condition](image-url)
Fig. 18 The confusion matrixes: (a) SVM; (b) LeNet-5; (c) WDCNN; (d) MC-CNN; (e) NLB-CNN; (f) Proposed
In this paper, a novel bearing fault diagnosis method named FSOCNN is proposed. It can accurately detect the fault location and severity in noise environment. The diagnosis model is based on 1DCNN with full stage optimization. The wide convolutional kernel, MFO structure and GAP are utilized to optimize the feature extraction stage. Among them, the proposed MFO is a novel method to enrich the output features of CNN. The highlight is that MFO does not bring any trainable parameters but effectively suppresses the overfitting of deep model. Then, SVM is used to optimize the classification stage. The effectiveness of the proposed FSOCNN method is verified on the two bearing datasets under different levels of noise. The results demonstrate that the proposed method achieves excellent performance, with over 94% diagnosis accuracy in the noise environment from -6dB to 0dB and over 99% diagnosis accuracy in the noise environment from 2dB to 8dB. What’s more, the training time is only 13.2s. Compared with several existing anti-noise methods, the proposed method has better stability and diagnosis accuracy on CWRU and UPB data sets. In particular, its training time is the shortest.

The full stage optimization method provides excellent robustness and classification ability. However, traditional machine learning classification algorithms, such as SVM and KNN, cannot perform gradient descent optimization. As a result, traditional softmax is still needed to participate in model training. This increases the time-consuming of model training to a certain extent. In order to solve this problem, further reduce the training time and improve the diagnostic performance, a new classifier optimization method will be investigated in near future.

**Funding**

This study was funded by the Fundamental Research Funds for Hubei Province Natural Science Foundation of China (2019CFB565) and the National Natural Science Foundation of China (Nos. 51775391, 51705384 and 51875430).

**Conflict of interest**

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work. There is no professional
or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, ‘A novel low-cost bearing fault diagnosis method based on convolutional neural network with full stage optimization in strong noise environment’.

Acknowledgement

The authors would like to thank Case Western Reserve University and Universität Paderborn for providing free access to the bearing vibration experimental data from their websites.

Reference


