Computers, Materials & Continua CMC, vol.63, no.1, pp.283-299, 2020

A Rub-Impact Recognition Method Based on Improved Convolutional Neural Network

Weibo Yang[1, *], Jing Li[2], Wei Peng[2] and Aidong Deng[3]

Abstract: Based on the theory of modal acoustic emission (AE), when the convolutional neural network (CNN) is used to identify rotor rub-impact faults, the training data has a small sample size, and the AE sound segment belongs to a single channel signal with less pixel-level information and strong local correlation. Due to the convolutional pooling operations of CNN, coarse-grained and edge information are lost, and the top-level information dimension in CNN network is low, which can easily lead to overfitting. To solve the above problems, we first propose the use of sound spectrograms and their differential features to construct multi-channel image input features suitable for CNN and fully exploit the intrinsic characteristics of the sound spectra. Then, the traditional CNN network structure is improved, and the outputs of all convolutional layers are connected as one layer constitutes a fused feature that contains information at each layer, and is input into the network's fully connected layer for classification and identification. Experiments indicate that the improved CNN recognition algorithm has significantly improved recognition rate compared with CNN and dynamical neural network (DNN) algorithms.

Keywords: Acoustic emission signal, deep learning, convolutional neural network, spectral features, rub-impact.

1 Introduction

The effective recognition of acoustic emission (AE) from rotor rub-impact has a significant to early diagnosis of the rub-impact fault, analysis of rub-impact state and warning of fault development trend. However, it is hard to extract features from rotor rub-impact acoustic emission signal and recognize faults because of the noise interference, the complexity of rub-impact time-frequency signal in rotating machines and the waveform variation caused by dispersion effect [Deng, Zhao and Wei (2011); Deng, Cao, Tong et al. (2014)].

At present, some scholars have researched rub impact fault features through AE signal,

[1] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China.

[2] School of Information Engineering, Nanjing Audit University, Nanjing, 211815, China.

[3] National Engineering Research Center of Turbo-generator Vibration, Southeast University, Nanjing, 210009, China.

[*] Corresponding Author: Weibo Yang. Email: ywb1987@nuaa.edu.cn.

Received: 29 May 2019; Accepted: 01 July 2019.

which focuses on feature extraction, feature expression, recognition model and so on. In the aspect of feature extraction, many feature vectors including hit accumulation, amplitude distribution, frequency distribution, power spectral density and so on, their effectiveness has been verified. However, recognition algorithms combined with constructed feature vectors are widely used to solve faults diagnosis problem. Deng et al. [Deng, Zhao and Wei (2011); Deng, Cao, Tong et al. (2014)] researched waveform factual dimension algorithm and further to use support vector machine (SVM) to recognize the rubbing fault recognition in the rotatory machine [Abdel-Hamid, Mohamed and Jiang (2014)]. As an important recognition method using machine learning, it usually adopts statistical neural network model instead of the identified model to make the dynamic reconfiguration [Deng, Zhang, Tang et al. (2012); Deng, Tong, Tang et al. (2013)].

With the development of deep learning in recognition model, researchers have put forward many recognition models based on the deep learning, which were different from traditional machine learning frameworks, such as convolutional neural network (CNN) [Krizhevsky, Sutskever and Hinton (2012)], recurrent neural network (RNN) [Mikolov, Karafiát, Burget et al. (2010)], deep belief network [Mohamed, Dahl and Hinton (2011)]. Mostly, these methods used to solve image-processing tasks achieve great success [Abdel-Hamid, Mohamed and Jiang (2014); Cui, McIntosh and Sun (2018); Liu, Li, Shan et al. (2015); Perez, Karakus and Pellet (2017); Szegedy, Liu, Jia et al. (2015); Sun, Cao, Xu et al. (2015)]. In recent years, some researchers start to apply CNN in various acoustic recognition models and obtain some achievements. CNN applied in fault diagnosis improves the level and efficiency of fault diagnosis. However, deep CNN also has some drawbacks. Common problems include gradient disappearance, too many parameters and training data and so on. For the recognition problem on small training sets, the solutions would be different due to different fields.

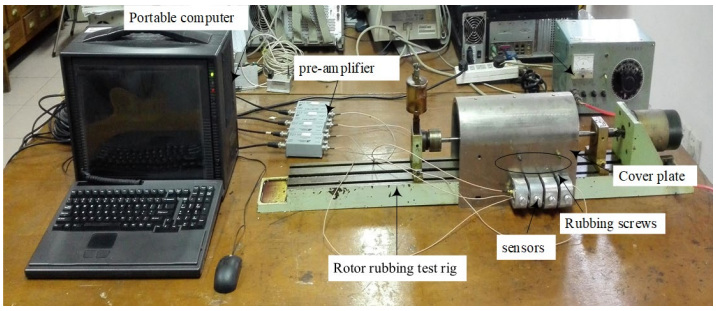
Taking decaying, delaying, frequency dispersion and multi-modal characteristics of the received AE signal into consideration, there is fewer useful AE signal for rub-impact recognition. In this paper, based on time-frequency analysis, an AE spectral features method is applied first. It can not only improve the effective data quantity of AE recognition but also use spectrogram as the characteristic of the rotor running state. It can show the detailed changes of AE signal from time, frequency and energy intensity, and then effectively describe the fault characteristics contained in the AE signal, which is of great significance for realizing fault diagnosis of rotating machines. Secondly, a more applicable CNN structure is also introduced to avoid overfitting under limited learning data. The experiment proves the effectiveness of this improved CNN structure.

The organization of this paper is as follows. Section 2 introduces the analysis method of AE spectral features. To solve the classification problem of the rub-impact signal, an improved CNN structure suitable for rub-impact recognition is constructed by referring to the connection mode of DeepID in Section 3. Then, the upper proposed method is verified by 700 AE signals for three conditions of normal operation, local rubbing, and rubbing in Section 4. Finally, Section 5 gives the conclusion.

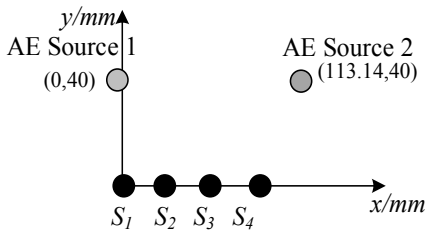
2 Study on analysis method of ae spectral features

The AE source signal has some characteristics like diversity, transient, non-stationary and time-frequency complexity and its characters are similar to the speech. Therefore, it can follow methods of speech signal processing. Assuming that the AE signal is short-time stationery, the AE signal can be spectrum analyzed in the vicinity of the time at any time. This continuous analysis can obtain two-dimensional spectral, called spectrogram.

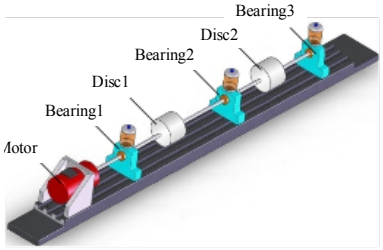
The rotor rubbing test bed is shown in Fig. 1. Regulate the speed of the rotor by adjusting the current of the alternating current motor. A retractable screw is installed on the screw hole of the arch guided wave plate. Rotor rubbing is produced by adjusting the screw along the radial direction of the rotating shaft to the center of the rotating shaft. The installation position of AE sensor S1 on the guided wave plate is shown in Fig. 1(a). the sampling rate is 1 MHz, which means 512000 points will be sampled after triggering and the time length is 0.512 seconds. A floating threshold trigger is set. The AE signals under non-rubbing, local (slight) rubbing and whole-period (severe) rubbing conditions are analyzed in the experiment. The frame length is 512 and the frame shift is 245. The time-frequency analysis of 512-point fast Fourier transform (FFT) is used to calculate the spectra The AE wave is transmitted to the guide plate through the screw and then to sensor S1 when rubbing. Combined with the location of the AE source, analyze the AE signal after multi-modal suppression and dispersion compensation. The abscissa of the spectrogram represents time and the ordinate is frequency. The gray level of each pixel reflects the energy density of a certain time and frequency. The spectrogram can not only help to find the characteristics of different fault types but also can be used as an index to evaluate the severity of rub-impact, which will contribute to the effective identification of rub-impact faults.



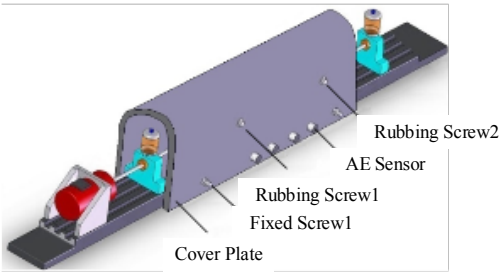
(a) Experiment system



(b) The location of rub-impact source



(c) Structure of rotor rub-impact test-bed



(d) Structure of rotor rub-impact locating test-bed

Figure 1: Rub-impact AE signal acquisition experimental system

Firstly, the AE signal is added with the window and processed into frames. Based on its short-time stationary property, the AE time series signal is divided into several short-time frames. The number of data points contained in each frame is the frame length and overlapping neighboring frames makes the transition between frames smooth and keeps its continuity. The frame separation is usually achieved by weighting a movable finite-length window. The window function $\omega(n)$ is multiplied with the signal received by AE sensors and the windowed signal will be achieved by $s_\omega(n) = s(n) \times \omega(n)$ [Chen, Li and Sanchez (2015)]. Because the Hamming window can improve the spectrum leakage, it is usually used as the window function to process the AE signal. The Hamming window is represented as Eq. (1).

$$w(n) = \begin{cases} 0.54 - 0.46\cos[2\pi n/(N-1)], & 0 \le n \le (N\text{-}1) \\ 0, & \text{others} \end{cases} \tag{1}$$

The signal $x_n(m)$ is the nth AE signal frame, its FFT result is:

$$X(n,k) = \sum_{m=0}^{N-1} x_n(m) e^{-j\frac{2\pi km}{N}} \tag{2}$$

$0 \le k \le N-1$, $|X(n,k)|$ is the short-time spectral magnitude estimation. The energy intensity function $P(n,k)$ will be acquired:

$$P(n,k) = |X(n,k)|^2 \tag{3}$$

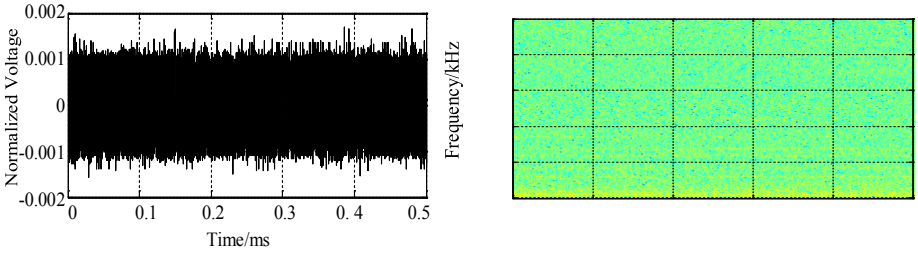
$P(n,k)$ is the two-dimensional non-negative real value function. The time n denotes the abscissa and the frequency k denotes the ordinate. The two-dimensional image constructed by $P(n,k)$ is the spectrogram. The spectrogram is a two-dimensional image that reflects the energy intensity of AE signal in time and frequency domain. It can describe the detailed changes from various angles. Then it will effectively and comprehensively describe the subtle fault characteristics of the signal. In order to measure various features effectively, the original features should be normalized before feature selection during the recognition of different rub-impact faults. The method is as follow:

$$\tilde{f}_{i,j} = \frac{f_{i,j} - \alpha_i}{\beta_i - \alpha_i} \tag{4}$$

In the formula, $f_{i,j}$ is the *i*th feature value of the *j*th sample. α_i is the minimum of the *i*th feature, $\alpha_i = \min_j(f_{i,j})$. β_i is the maximum of the *i*th feature, $\beta_i = \max_j(f_{i,j})$.

2.1 Non-rubbing state

The total gain of the AE driver is adjusted to 60 dB. The AE signal is acquired and the spectrogram is calculated in a state where the rotor test stand is stable and does not experience rubbing, as shown in Fig. 2. At this time, the AE waveform is a noise signal. From the time-frequency domain, the noise is concentrated at 0-20 kHz. The low-frequency part of the noise spectrum mainly comes from the vibration noise of the mechanical structure of the rotor table, and the high-frequency part is the noise caused by the weak rub inside the motor.

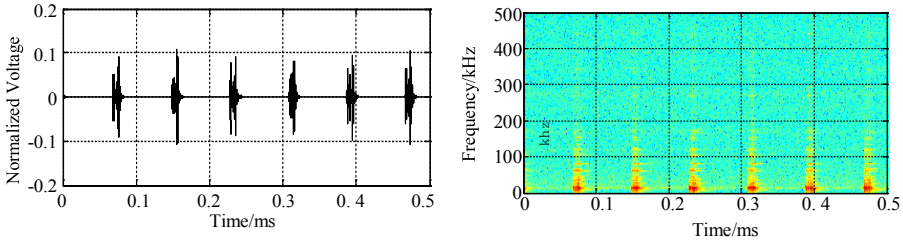


(a) Waveform (b) Spectrogram

Figure 2: AE signal without rubbing

2.2 Local rubbing state

Install a slightly curved rotor shaft, the center of the shaft deviates from the geometric center of the bearing, and the rubbing screw makes slight contact with the rubbing rotor. Every revolution of the rotor, rotary discs and rubbing screws generate multiple points or parts of arc contact, and local rubbing occurs at this time. Fig. 3 shows the AE signal and spectrogram. The obvious periodical impact phenomenon can be seen from the time-domain waveforms, which are expressed as several burst-type AE signals with a frequency range of 0-100 kHz and energy concentration near 25 kHz.

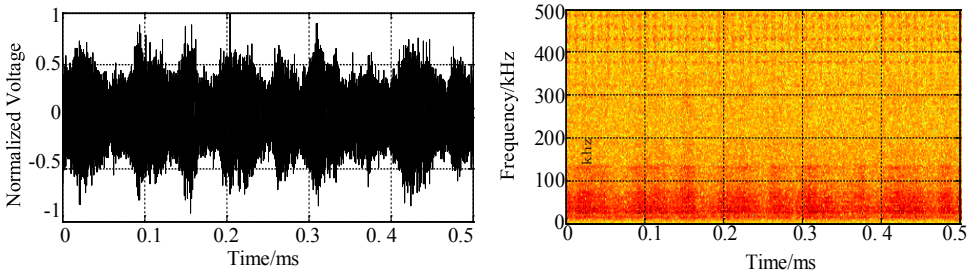


(a) Waveform (b) Spectrogram

Figure 3: AE signal for partial rubbing

2.3 The whole-period rubbing state

Adjusting rubbing screws to increase rubbing on the above experimental device, and the rotor is always in contact with the stator during one revolution. At this point, a full-week rub occurred. Fig. 4 shows the AE signal and sound spectrum during the rubbing of the whole-period. From the waveform of time domain Fig. 4(a), it can be seen that the amplitude of AE signal fluctuates significantly, and it is characterized by a continuous AE signal. A large number of AE events occur simultaneously and are indistinguishable in time. In Fig. 4(b), the frequency is more abundant, the high-frequency component is increased, and the energy amplitude is also greatly increased.



(a) Waveform (b) Spectrogram

Figure 4: AE signal of the whole-period rubbing status

3 Improving the structure of CNN

3.1 Convolutional neural network

A typical CNN consists of several different layers, which are cascaded in a certain order, including an input layer, a number of convolution layers and pooling layers, a certain number of fully-connected layers and the last output layer (loss layer). CNN introduces a special way to organize hidden layer units called convolution, which makes the network well extract local features of input data. The weight of the local connection between hidden layers is called the convolution kernel (convolution filter). All hidden layer units in the same layer share the same convolution kernel. Therefore, the output of this layer is the local feature extracted by the convolution kernel, also known as the feature map.

In the CNN, the features extracted from the lower convolution layer are closer to the original input data, so the level is relatively low. The features of the high-level convolution layer at the relatively top layer are derived from the features extracted from the low-level convolution layer, so the level is relatively high. CNN extract global features [Liu, Li, Shan et al. (2015)] by extracting the output (feature map) of the former layer.

3.2 DeepID network

The DeepID network [Sun, Chen, Wang et al. (2014)] breaks through 99% recognition rate in the field of face recognition with its special network connection mode and feature extraction method and achieves great success. DeepID learns a set of high-level feature representation (deep hidden identity feature) for face recognition through deep learning. Extract features by learning a multi-class (10,000 classes, each class has about 20 instances) task of face recognition and generalize features to face recognition and other

new recognition tasks. Meanwhile, the more types of faces to be predicted during training, the stronger the generalization ability of DeepID [Wang, Jiang, Shao et al. (2017)]. Along with the layers of feature extraction, decreasing the number of units will gradually form features of a few hidden units closely related to identity at a higher level.

DeepID network has achieved great success in face recognition, reaching 99.15% recognition rate, exceeding untrained experimenters. To address this problem, by connecting the output of all convolution layers in the network into one layer and connecting all feature maps in the network together [Hansong and Chuanyang (2008)], the features extracted by the convolution network contain the features from each layer. Thus, it can maintain the feature information to the greatest extent and avoid losing some edge information due to the dimension-reduction of the pooling layer in the traditional convolution network.

3.3 The improving method of CNN architecture

Traditional CNN adopts serial stacking structure but such structure will result in the lack of low-level features in the final features extracted by the network [Chan, Jia and Gao (2014)]. The features extracted by the convolution network finally contain the features from each layer, thus retaining the feature information to the greatest extent, and avoiding the loss of some edge information due to the dimensionality reduction of the pooling layer in the traditional convolution network.

Compared with the traditional imaging method, the spectrogram of sound has only one channel, relatively less information at the pixel level and strong local correlation [Li, Kang, Xu et al. (2014)]. While the convolution layers of traditional CNN are refining and extracting image features layer by layer, the size of feature maps of each layer decreases. Some coarse-grained information and edge information are losing, which makes the information dimension of the spectrogram with relatively less original pixel information lower in the top layer of CNN [Li, Kang, Xu et al. (2014)].

To better preserve the information of different dimensions and take into account the global characteristics of the top layer network and the local characteristics of the bottom layer network, this paper introduces a more applicable network connection method, that is, by adding branch connections, the output of each layer of the convolution layer in the network is directly connected to the last hidden layer of the network, which constitutes the final feature. This feature is called a multi-scale feature in other researches [Eberhardt, Stead and Stimpson (1999)]. It is widely used in unsupervised feature learning [Karl (2005)] and supervised feature extraction [Hyder and Mahata (2010)]. By adding such branch connections, the features of the last hidden layer contain features from all levels, so more information can be retained. However, the size of the bottom feature map is several times larger than that of the top feature map, which results in the information submergence in the top feature map. Meanwhile, the size of the bottom feature map is so large that also causes the problem of information redundancy [Dowling and Wald (1981)].

Therefore, in order to solve the classification problem of the rub-impact signal, a CNN structure suitable for rub-impact recognition constructed by referring to the connection mode of DeepID. The convolution kernel of the *l*th convolution layer of the network is represented by K^l, and the input of the layer is represented by I^l. Then the feature graph

Z of the layer can be represented by F^l, as shown in Eq. (5).

$$F^l(i,j) = (K^l \times I^l(i,j)) = \sum_m \sum_n I^l(i-m, j-n) K^l(m,n) \tag{5}$$

where, × denotes the convolution operation.

New feature maps can be obtained by connecting the feature maps extracted from the first three convolution layers. However, because the dimensions of the three-level feature map are different, it is necessary to flatten the feature map into one-dimensional features according to Eq. (6):

$$F^l_{\text{flat}}(i + j \times m + c \times m \times n) = F^l(i,j,c) \tag{6}$$

where, m and n represent the size of the lth convolution kernel.

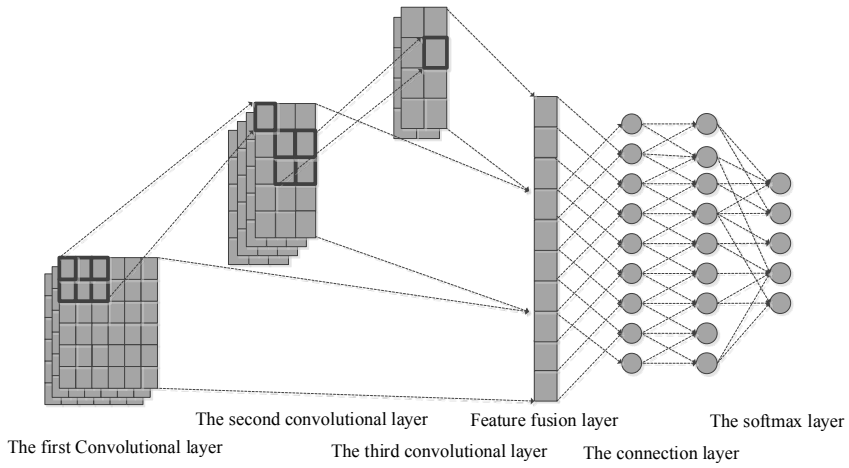


Figure 5: AE signal of the whole-period rubbing status

Then, accumulate the flattened feature maps to get the fusion feature maps. The network structure of this paper is shown in Fig. 5, including a convolution layer, a pooling layer, a convolution layer, a pooling layer, a fully-connected layer, and a classifier. The first convolution layer uses convolution kernel 64@2×2, the second convolution layer uses convolution kernel 96@2×2 and the third convolution layer uses convolution kernel 128@2×2. The average pooling size of the pooling layer is 2×2. The training parameters of the network are set as follows: batch=16, epoch=100, learning rate λ=0.00001. Random gradient descent algorithm is used to optimize the training parameters and cross-entropy is used as a loss function [Stoica, Babu and Li (2011)].

Because the amount of data used in this paper is small and the number of layers is large, serious over-fitting phenomenon occurs in the training process. Therefore, Dropout [Srivastava, Hinton, Krizhevsky et al. (2014)] proposed by Hinton is introduced in this paper and applied on all convolution layers and fully-connected layers to suppress over-fitting. At the same time, considering the different degree of over-fitting in each layer, the dropout rate is 0.2 for the convolution layer and 0.5 for the fully-connected layer.

4 The results of experiment

4.1 Input data processing

The experimental data is 700 AE signals for three conditions of normal operation, local rubbing, and rubbing during the entire cycle and each data lasts 512000 sampling points. The length of time is 0.512 seconds and 5-fold cross-validation is used. The time-series AE signal is characterized by frame length 512, 50% frame shift, and 512-point FFT spectrum. On this basis, the first-order and second-order differential features of both time and frequency axis direction [Mohimani, Babaie-Zadeh and Jutten (2009)] in Python, and the five-channel image input features are finally formed together with the original spectrogram.

There are two types of the differential spectrum: the differential spectrum along the time axis and along the frequency axis. The differential spectrum along the time axis can better represent the dynamic change of sound clips time [Hyder and Mahata (2010)]. Similarly, dynamic frequency characteristics can also be well represented by differential spectrum along the frequency axis. Tab. 1 describes the composition of multichannel spectrograms.

Table 1: Composition of the multichannel spectrograms

Channel	Composition of the Multichannel Spectrograms
1	Spectrogram without differential spectrum
2	Spectrogram with first-order differential spectrum
3	Spectrogram with first-order frequency and time differential spectrum
4	Spectrogram with first-order frequency differential spectrum, first-order time differential spectrum and second-order frequency differential spectrum
5	Spectrogram with first-order and second-order frequency differential spectrum and first-order and second-order time frequency differential spectrum

4.2 Selection of networks hyper parameters

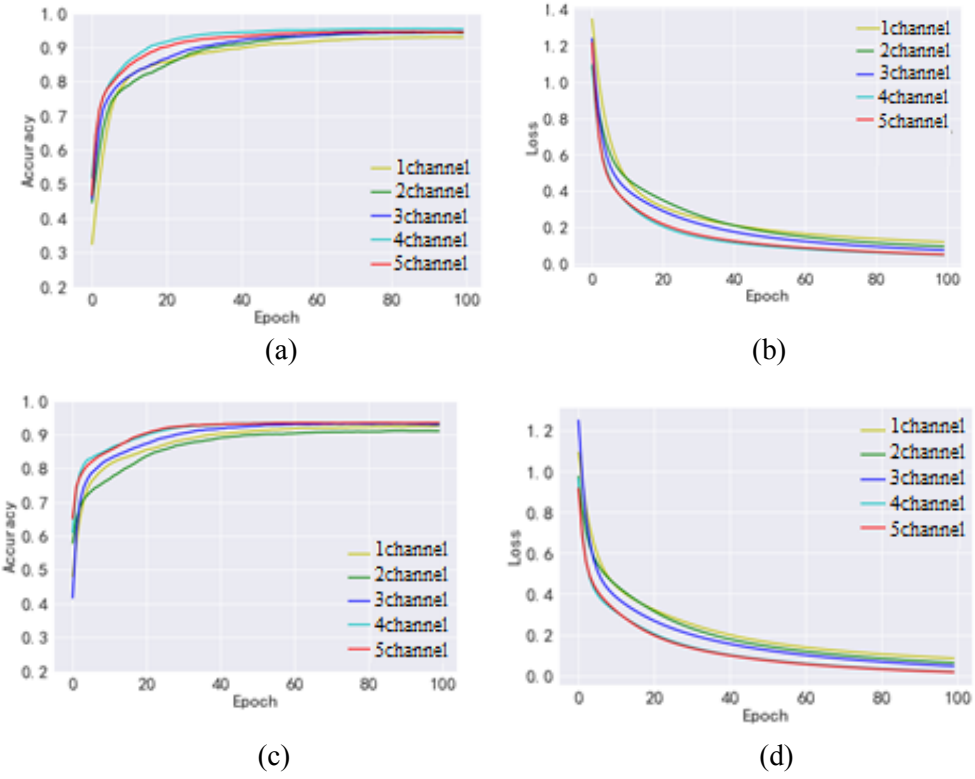
In the experiment, tensor flow, a general in-depth learning framework, was used to build and train the network. The selection of hyper parameters in neural networks has a great influence on the training of networks and even the convergence of the final networks. In order to ensure the generalization ability of networks, the final hyper parameters are determined through many experiments and comparisons, as shown in Tab. 2.

Table 2: Training hyper parameters table

Hyper Parameters	Value
Dropout Rate	0.2
Momentum	0.90
Learning Rate	0.0025
Regulation of Learning Rate	Fixed
Learning Algorithm	Stochastic Gradient Descent
Regularization	L2
Weight Decay	0.0005
Batch Size	64
Epoch	50

4.3 Analysis of recognition performance under noisy environment

Adding Gaussian white noise to the AE signal with the various signal to noise ratio (SNR), the recognition results of the improved CNN model and multichannel spectral features are shown in Fig. 6. The training curves of the recognition rate corresponding to the model are 20 dB, 10 dB, 0 dB, -10 dB and -20 dB respectively. The training curves of the loss function corresponding to the model are 20 dB, 10 dB, 0 dB, -10 dB and -20 dB respectively. The training curves of the model are 20 dB, 10 dB, 0 dB, 0 dB, -10 dB and -20 dB respectively. It can be found from figures that the accuracy of the model decreases and the loss function increases when reducing the SNR. Moreover, the accuracy of different channel features is gradually widened. When the SNR decreases gradually, the accuracy of spectrum features with more channels decreases less, such as 4-channel spectrum features and 5-channel spectrum features. However, when the SNR decreases gradually, the accuracy of spectrum features with fewer channels is higher. For example, single-channel spectrum features and 2-channel spectrum features illustrate the effectiveness of multichannel spectrum features, especially in lower SNR environments. In addition, the network is generally trained to converge after 40 epochs and the error tends to a smaller value and remains stable. The accuracy of model recognition rises with the increase in the number of channels in the spectrogram, but it does not continually increase. When the number of channels is around 5, it is basically stable.



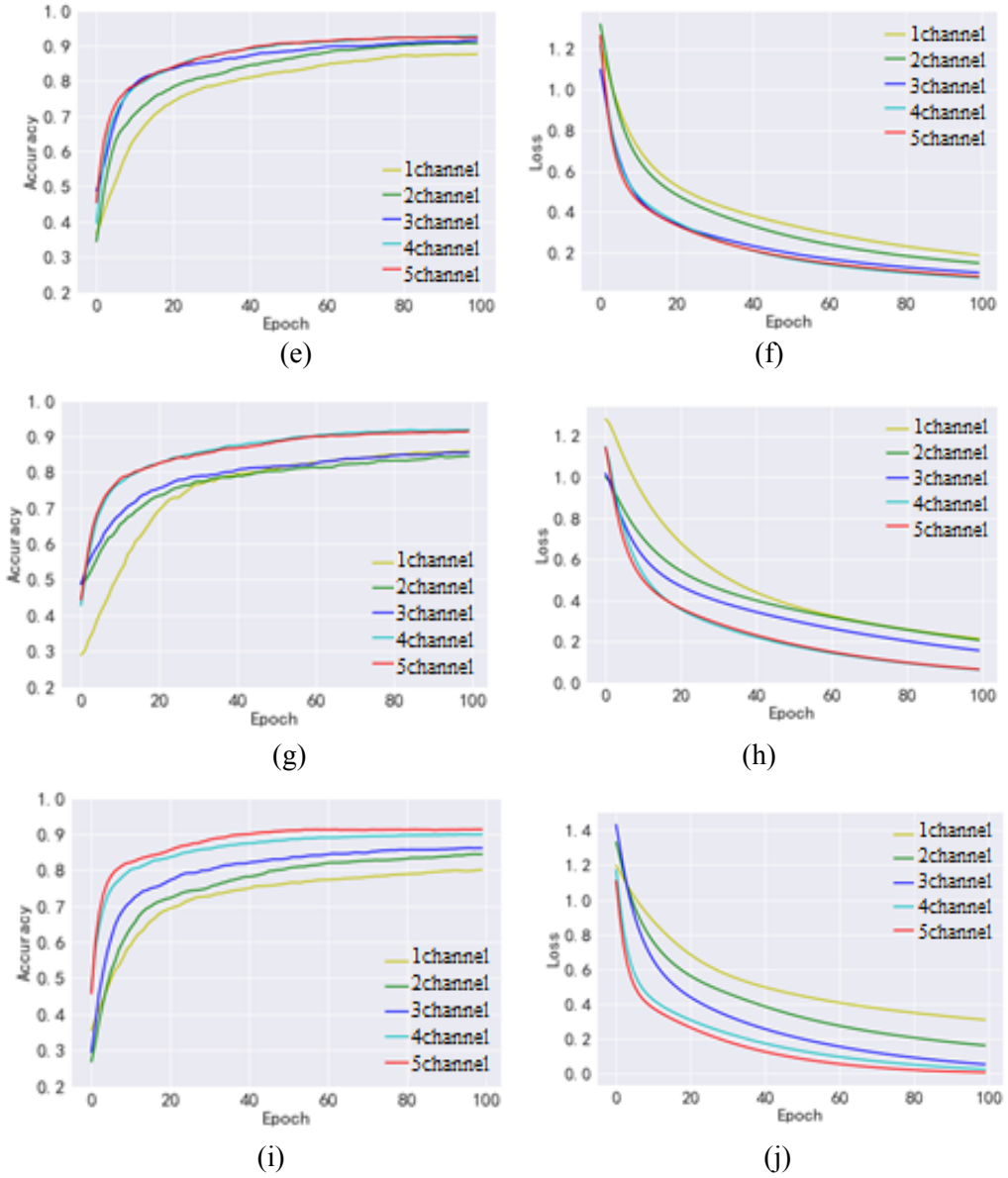


Figure 6: Convergence curves of 5 improved CNN models with different SNR during training

Tab. 3 gives the confusion matrix of AE recognition of rubbing faults for 5-channel spectral features at 0 dB. It can be seen that the recognition rate of the whole-period rubbing is 99.55%, close to 100%, which can effectively identify the whole-week rubbing; the recognition rate of the local rubbing fault is 90.36%, and the recognition rate of the non-rubbing state is 84.31%. It shows that under the normal environment, the model based on 5-channel spectral features can effectively recognize three rubbing states.

Table 3: Composition of the multichannel spectrogram

	Non-rubbing	Local Rubbing	Whole-period Rubbing
Non-rubbing	84.31%	13.53%	2.16%
Local Rubbing	7.86%	90.36%	1.78%
Whole-period Rubbing	0	0.45%	99.55%

Tab. 4 gives the confusion matrix of AE recognition of rub-impact faults for 5-channel spectral features at 10 dB. It can be seen that after the SNR is increased by 10 dB, the recognition rate of the whole-period rubbing is 100%, and the recognition rate of the whole-period rubbing is 100%. The recognition rate of the local rubbing fault is also improved, which is 91.72%. The recognition rate of the non-rubbing state is 87.85%.

Table 4: Confusion matrix of AE recognition of rubbing faults for 5-channel spectral features at 10 dB

	Non-rubbing	Local Rubbing	Whole-period Rubbing
Non-rubbing	87.85%	11.53%	0.62%
Local Rubbing	7.16%	91.72%	1.12%
Whole-period Rubbing	0	0	100%

Tab. 5 gives the confusion matrix of AE recognition of rub-impact faults for 5-channel spectral features at 20 dB. When the SNR is 20 dB, the recognition rate of the model is 100% and unchanged; the recognition rate of the local rubbing fault is 95.18%, which is 4.82% higher than that of the 0 dB; the recognition rate of the non-rubbing state is 88.74%, which is 5.03% higher than that of the 0 dB. Most of the error recognition occurs in the recognition of the non-rubbing state as the local rubbing state.

Table 5: Confusion matrix of AE recognition of rubbing faults for 5-channel spectral features at 20 dB

	Non-rubbing	Local Rubbing	Whole-period Rubbing
Non-rubbing	88.74%	10.92%	0.34%
Local Rubbing	4.12%	95.18%	0.70%
Whole-period Rubbing	0	0	100%

Tab. 6 gives the confusion matrix of AE recognition of rub-impact faults for 5-channel spectral features at -10dB. With the decrease of SNR, the recognition rate of whole-period rub-impact of the model decreases to 99.16%, slightly lower than that of 0 dB, but it can still recognize the whole-period rubbing effectively; the recognition rate of local rubbing is 84.90%, which is 5.46% lower than that of 0 dB; the recognition rate of non-rubbing state is 80.92%, which is 3.39% lower than that of 0 dB. Compared with 0dB, the overall reduction is 3.1%. Compared with the previous situation of lower SNR, the reduction is the largest. This shows that the recognition accuracy of the model decreases rapidly under the condition of the sharp deterioration of SNR. However, the model based on the 5-channel spectral features in this paper can still maintain an average recognition accuracy of 88.32%.

Table 6: Confusion matrix of AE recognition of rubbing faults for 5-channel spectral features at -10 dB

	Non-rubbing	Local Rubbing	Whole-period Rubbing
Non-rubbing	80.92%	16.15%	2.93%
Local Rubbing	12.89%	84.90%	2.75%
Whole-period Rubbing	0	0.84%	99.16%

Tab. 7 gives the AE confusion matrix for rub-impact fault recognition with 5-channel spectral features at -20 dB. It can be seen that, even in the low SNR environment, the recognition rate of the algorithm for the whole-period rubbing is 97.3%, which shows that the method can effectively recognize the whole-period rubbing; the recognition rate for the local rubbing is 75.24%, which is 15.12% lower than that for the 0 dB. Most of the local rubbing is recognized as the non-fault state, and the recognition accuracy for the non-rubbing state is 78.47%. There are many recognition confusion problems with local rubbing.

Table 7: Confusion matrix of AE recognition of rubbing faults for 5-channel spectral features at -20 dB

	Non-rubbing	Local Rubbing	Whole-period Rubbing
Non-rubbing	78.47%	15.84%	5.69%
Local Rubbing	21.74%	75.24%	4.0%
Whole-period Rubbing	1.12%	1.58%	97.3%

4.4 Model contrast experiments

In order to make the model evaluation more contrast, on the basis of the same features (5-channel spectrogram), the network is compared with a variety of classifiers on the rotor rub-impact AE signal data. In order to compare the effect of convolution layer feature extraction and the effectiveness of the network structure in this paper, a five-layer deep neural network and a traditional convolution neural network as well as some traditional classifiers are built and tested on the same data. The comparison classifiers and their parameters are as follows:

(1) Random forest. The number of decision trees is 400, the maximum depth of each decision tree is 6, the row sampling rate is 0.75, and the column sampling rate is 0.8.

(2) CNN. It has two convolution layers. The convolution nucleus is 128@3×3 and 156@2×2. There are two fully-connected layers with 5000 units and the Softmax layer as classifiers. In order to prevent over-fitting, dropout is used in both convolution layer and fully-connected layer, with a coefficient of 0.5.

(3) Dynamical neural network (DNN). It consists of five layers with fully-connected layers. The number of units in each layer is 384. Dropout is used in each layer with a coefficient of 0.5. The last layer is the Softmax layer.

(4) SVM. The penalty coefficient C=1, using radial basis function kernel and one *vs.* one multi-classification strategy.

(5) K-nearest neighbor algorithm. The number of nearest neighbors N=8, and K-

dimensional tree is used to search, where the metric distance of the tree is Euclidean distance.

(6) RNN. It has two layers with 1024 units in each layer, and the dropout rate of each layer is 0.5.

Tab. 8 shows the recognition rate of rotor rub-impact AE signal by various classifiers in an ideal noise-free environment. Compared with random forest, KNN and SVM, the recognition rate of the model in this paper increased by 9.8%, 21.3% and 14.5%, respectively. As a kind of strong anti-over-fitting classifier, the recognition rate of random forest is only 85.4%, which indicates that the random forest does not fully utilize the existing information in the sound spectrum data. And the recognition performance of SVM and KNN is only 80.7% and 73.9%, which proves that the CNN performs better than the traditional classifier in rotor rub-impact AE signal recognition when dealing with image multidimensional features (multi-channel spectral features).

Compared with DNN, the accuracy of the improved CNN model raised by 6.5%. Although both are deep neural networks, the model in this paper has a convolution layer and belongs to CNN structure. When dealing with two-dimensional image data, CNN can better adapt to the data on their spatial and local relationship, extract more effective features, so as to achieve better recognition rate.

Compared with traditional CNN, the specific structure of DeepID network integrates multi-level features by connecting the output of each convolution layer, thus retaining feature information to a greater extent. This model extends the DeepID network further. By adding two layers of a fully-connected layer, the network can better characterize the sample distribution in high-dimensional space and distinguish fusion features. Also, adding the Softmax layer makes the model can directly classify the rotor rub-impact AE signals, changing the pattern that DeepID only extracts features but does not classify them. Therefore, the recognition rate of this network is 4.4% higher than that of traditional CNN.

Table 8: Recognition rate of rotor rub-impact AE signal by various classifiers

Accuracy	Mean	Standard Deviation
Random Forest	85.4	0.9
Traditional CNN	90.8	1.2
RNN	91.5	1.6
KNN	73.9	2.1
SVM	80.7	1.4
DNN	88.7	1.7
Model in This Paper	95.2	1.3

5 Conclusion

In this paper, a rotor rub-impact fault recognition method based on improved CNN for AE signal is introduced. The time-frequency features of the spectrogram of the rub-impact AE signal are analyzed. The structure of CNN is improved by using DeepID

network with more applicable connection method. An improved CNN suitable for rotor rub-impact fault identification is proposed, which is summarized as follows:

(1) The problem of AE signal recognition for rotor rub-impact is different from that of face recognition. It's sound clips are one-dimensional time series data while the face image is two-dimensional data with a special topological structure. Firstly, this paper converts one-dimensional audio data into a two-dimensional spectrogram image and uses CNN as a classifier, using DeepID's unique network connection to organize the network.

(2) Compared with the traditional CNN, the improved CNN connects the output of all convolution layers into one layer and connects all feature maps together in the network so that the features extracted by the convolutional network contain the features from each layer, thus avoiding the loss of some edge information due to the pooling layer in the traditional CNN.

(3) In this paper, based on the spectrogram image, the extracted multi-order differential features of the spectrogram and the 5-channel image features are combined to form the original spectrogram. As the final input features, the static and dynamic characteristics of the spectrogram are effectively explored. Moreover, CNN is used as a classifier to make full use of the advantage of the multi-channel image, which makes the input data contain more static and dynamic audio features.

(4) Considering practical use conditions, the proposed method needs further improvements according to different rotate speed and loading. In addition, it is import to add more fault types of rotating machinery that the generalization ability of the proposed method could be continually improving its effectiveness.

Acknowledgement: The authors would like to acknowledge the Six Talent Peaks Project in Jiangsu Province [XCL-CXTD-007] and China Postdoctoral Science Foundation [2018M630559] for their financial support in this project.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Abdel-Hamid, O.; Mohamed, A. R.; Jiang, H. (2014): Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1533-1545.

Chan, T.; Jia, K.; Gao, S. (2014): PCANet: a simple deep learning baseline for image classification. *IEEE Transactions on Image Rocessing*, vol. 24, no. 12, pp. 5017-5032.

Chen, Z.; Li, C.; Sanchez, R. (2015): Gearbox fault identification and classification with convolutional neural networks. *Shock and Vibration*, vol. 15, no. 8, pp. 18-28.

Cui, Q.; McIntosh, S.; Sun, H. (2018): Identifying materials of photographic images and photorealistic computer generated graphics based on deep CNNs. *Computers, Materials & Continua*, vol. 55, no. 2, pp. 229-241.

Deng, A.; Cao, H.; Tong, H.; Zhao, L.; Qin, K. et al. (2014): Recognition of acoustic

emission signal based on the algorithms of TDNN and GMM. *Applied Mathematics & Information Sciences*, vol. 8, no. 2, pp. 907-916.

Deng, A.; Tong, H.; Tang, J.; Cao, H.; Qin, K. et al. (2013): Study on location algorithms of beamforming based on MVDR. *Applied Mathematics & Information Sciences*, vol. 7, no. 6, pp. 2455-2466.

Deng, A.; Zhao, L.; Wei, X. (2011): Application of quantum neural networks in localization of acoustic emission. *Journal of Systems Engineering and Electronics*, vol. 22, no. 3, pp. 507-512.

Deng, A.; Zhang, X.; Tang, J.; Zhao, L.; Qin, K. (2012): Localization of acoustic emission source based on chaotic neural networks. *Applied Mathematics & Information Sciences*, vol. 6, no. 3, pp. 713-719.

Dowling, J. E.; Wald, G. (1981): Proceedings of the national academy of sciences of the united states of america. *Nutrition Review*s, vol. 39, no. 3, pp. 135-138.

Eberhardt, E.; Stead, D.; Stimpson, B. (1999): Quantifying progressive pre-peak brittle fracture damage in rock. *International Journal of Rock Mechanics and Mining Sciences*, vol. 36, pp. 361-380.

Hansong, G.; Chuanyang, D. U. (2008): Rock mechanic analysis of earthquake prediction with acoustic emission. *Advances in Earth Science*, vol. 23, no. 12, pp. 1293-1298.

Hyder, M. M.; Mahata, K. (2010): An improved smoothed l0 approximation algorithm for sparse representation. *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2194-2205.

Hyder, M. M.; Mahata, K. (2010): Direction-of-arrival estimation using a mixed l2,0 norm approximation. *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646-4655.

Karl, W. C. (2005): Regularization in image restoration and reconstruction. *Handbook of Image and Video Processing (Second Edition)*, pp. 183-202.

Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012): Imagenet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems*, *Curran Associates Inc*, pp. 1097-1105.

Li, H. Y.; Kang, L.; Xu, Z.; Ji, Q.; Zhao, S. (2014): Precursor information analysis on acoustic emission of coal with different outburst proneness. *Journal of China Coal Society (In Chinese)*, vol. 39, no. 2, pp. 384-388.

Li, J. (2017): *Research on Condition Evaluation and Localization Technology Using Acoustic Emission (Ph.D. Thesis)*. Southeast University, China.

Liu, M.; Li, S.; Shan, S.; Chen, X. (2015): AU-inspired deep networks for facial expression feature learning. *Neurocomputing*, vol. 159, no. 8 pp. 126-136.

Mikolov, T.; Karafiát, M.; Burget, L.; Černocký, J.; Khudanpur, S. (2010): Recurrent neural network based language model. *Eleventh Annual Conference of the International Speech Communication Association*.

Mohamed, A.; Dahl, G. E.; Hinton, G. (2011): Acoustic modeling using deep belief networks. *IEEE Transactions on Audio Speech & Language Processing*, vol. 20, no. 1, pp. 14-22.

Mohimani, H.; Babaie-Zadeh, M.; Jutten, C. (2009): A fast approach for overcomplete sparse decomposition based on smoothed l0 norm. *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289-301.

Perez, S.; Karakus, M.; Pellet, F. (2017): Development of a tool condition monitoring system for impregnated diamond bits in rock drilling applications. *Rock Mechanics and Rock Engineering*, vol. 50, no. 5, pp. 1289-1301.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. (2014): Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929-1958.

Stoica, P.; Babu, P.; Li, J. (2011): SPICE: a sparse covariance-based estimation method for array processing. *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629-638.

Sun, J.; Cao, W.; Xu, Z.; Ponce, J. (2015): Learning a convolutional neural network for non-uniform motion blur removal. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 769-777.

Sun, Y.; Chen, Y.; Wang, X.; Tang, X. (2014): Deep learning face representation by joint identification-verification. *Advances in Neural Information Processing Systems*, pp. 1988-1996.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S. et al. (2015): Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9.

Wang, F.; Jiang, H.; Shao, H.; Duan, W.; Wu, S. (2017): An adaptive deep convolutional neural network for rolling bearing fault diagnosis. *Measurement Science and Technology*, vol. 28, no. 9, pp. 95-105.