Intelligent fault diagnosis of rolling bearing using one-dimensional Multi-Scale Deep Convolutional Neural Network based health state classification

Conference Paper · March 2018

DOI: 10.1109/ICNSC.2018.8361296

2 authors:

Zilong Zhuang

Wei Qin
Shanghai Jiao Tong University

Intelligent fault diagnosis of rolling bearing using one-dimensional Multi-Scale Deep Convolutional Neural Network based health state classification

Zhuang Zilong, Qin Wei
School of Mechanical Engineering
Shanghai Jiao Tong University, Shanghai, China
zhuangzilong910126, wqin@sjtu.edu.cn

***Abstract*—Fault diagnosis of rolling element bearings based on vibration signal is the most popular way to avoid underlying damage for any unexpected fault. In recent years, intelligent fault diagnosis algorithms using machine learning technique have achieved much success, and many deep learning techniques have also found their way into fault diagnosis of rotating machines. Considering that convolution is the most important method to analyze signals in digital signal processing, a novel deep convolutional neural networks is developed to operate directly on the raw vibration signal. The proposed MS-DCNN model could broaden and deepen the neural networks to learn better and more robust feature representations owing to multi-scale convolution layer, meanwhile, reduce the network parameters and the training time. Fault classification experiments of rolling element bearings have been undertaken to indicate the effectiveness of the MS-DCNN model. Compared with 1d-DCNN and 2d-DCNN, MS-DCNN can not only achieve higher accuracy rate in the testing set, but also run more smoothly in the training process.**

Keywords—fault diagnosis; convolutional neural networks; multi-scale

I. INTRODUCTION

Rolling element bearings (REBs) are one of the key components in rotating machines, whose any unexpected fault possibly causes great economical losses and serious casualties [1]. Fault diagnosis of REBs has been a hot topic of research for over 20 years [2]. The most popular way to avoid underlying damage is to implement a real-time monitoring of vibration when the rotating machines are in operation [3].

In recent years, intelligent fault diagnosis algorithms using machine learning technique have achieved much success. The vibration signals collected from rotating machines are raw temporal signals which contain the useful information of the machine, as well as useless noise. The common fault diagnosis algorithm is divided into three parts: useful features extraction and selection from raw signals, classifiers training based on these features, and generalization test [4]. Previous studies have shown that the accuracy of classifying results is largely dependent on the extracted features. Therefore, useful feature learning is becoming an important field in machine learning community in recent years [5].

As a novel theory of feature learning, deep learning was first proposed by Geoffrey Hinton et al. [6] as an effective way to imitate the learning process of human brain and has shown a great superiority in pattern recognition, and image processing. Deep learning does not need any hand-crafted features by people, instead it can learn a hierarchical feature representation from raw data automatically. In addition, the model doesn't rely on any domain adaptation algorithm or require any target domain information. The most surprising thing is that the model can achieve pretty high accuracy under noisy environment. Many deep learning techniques have already found their way into fault diagnosis of rotating machines.

The rest of the paper is organized as follows: the related work is reviewed in Section 2; Section 3 introduces the intelligent diagnosis method based on MS-DCNN; in Section 4, some experiments are conducted to evaluate our method against some other methods; and Conclusions are drawn in Section 5.

II. RELATED WORK

In recent years, there have been some attempts on feeding the frequency spectra of vibration signal into deep learning model. Jia et al. [7] fed the frequency spectra into Stacked AutoEncoders (SAE) to mine the useful information for fault diagnosis. Guo et al [8] and Verma et al [9] extracted time domain, frequency domain, and time-frequency domain features to represent characteristics of vibration signals, and then fed these features into SAE. All these researches need to obtain frequency spectra before conducting fault diagnosis, limiting the real-time performance of the algorithm.

In essence, the problem of fault diagnosis based on the raw vibration signal can be viewed as the classification task of time series, which is one of the most important fields of research for time series. To achieve end-to-end learning and real-time fault diagnosis, there are some studies work directly on the raw vibration signals. Lu et al. [10] presented a detailed empirical study of Stacked Denoising Autoencoders (SDA) for fault diagnosis of REBs. Lu et al. [11] mapped the raw monitoring signal to feature maps to make full use of the experience related to Convolutional Neural Networks (CNN) from image processing. Prasanna et al. [12] presented a novel multi-sensor health diagnosis method using a deep belief networks (DBN). Guo et al. [13] proposed hierarchical adaptive deep convolutional neural networks to diagnose bearing faults and determine their severity. Due to the fact that the algorithms themselves are designed for high-dimensional data, it is unfit

978-1-5386-5053-0/18/$31.00 ©2018 IEEE.

for them to be used directly to process on one-dimensional time series. Moreover, all of them have some drawbacks in both the execution efficiency and accuracy.

Convolution is the most important method to analyze signals in digital signal processing. Therefore, it appears natural to diagnose bearing faults using one-dimensional CNN. Zhang et al. [14] established one-dimensional Deep Convolutional Neural Networks (DCNN), which adopted an end-to-end learning method and could achieve pretty high accuracy under noisy environment. Zhang et al. [15] proposed DCNN with wide first-layer kernels, which is specially designed for extracting features and suppressing high frequency noise. Zheng et al. [16] proposed one-dimensional Multi-Channels Deep Convolution Neural Networks model, which is suitable for multivariate time series classification task. For fault diagnosis of REBs, multivariate time series means more sensors and higher cost. To achieve effective fault detection and diagnosis based on single data source, this paper proposes one-dimensional Multi-Scale Deep Convolutional Neural Networks (MS-DCNN) model. In this model, the concept of "inception" [17, 18] is combined closely with 1d convolutional neural networks, namely, convolution kernel of different sizes are used to extract different-scale features in parallel.

The Deep learning method based on MS-DCNN is exploited to mine discriminant feature representations by self-learning throughout the propagation steps. Through a large number of comparative tests, MS-DCNN shows great potential for fault diagnosis of REBs, compared with 1d-DCNN [14, 15] and 2d-DCNN [11].

III. MS-DCNN BASED INTELLIGENT DIAGNOSIS METHOD

Deep convolutional neural networks (DCNN) consists of some filter stages to extract features from the inputs and one classification stage to build relationship between inputs and outputs [14, 15]. Specifically, each filter stage contains the convolutional layer, batch normalization layer [19], activation layer, pooling layer, or any combination of them; the classification stage is usually composed of several fully-connected layers. Moreover, DCNN holds some degree of shift, scale, and distortion invariance by combining three architectural ideas: local receptive fields, shared weights and spatial sub-sampling [20]. On the one hand, MS-DCNN is a kind of DCNN with special architecture, therefore, MS-DCNN holds those characteristics as well, on the other hand, it can directly extract useful features of different sizes in parallel thanks to its special architecture—Multi-Scale convolution layer. This section details the proposed MS-DCNN based health state classification approach. Section 2.1 overviews the whole architecture and learning process. Section 2.2 analyzes the Multi-Scale convolution layer. Section 2.3 discusses the technique of data augmentation [21] for vibration signals.

A. *Architecture of the proposed MS-DCNN model*

Based on DCNNs, this study proposes a novel intelligent fault diagnosis method that adaptively mines the fault features from raw signals of REBs and automatically classifies machinery health conditions and severity with these fault features. The overall architecture of proposed MS-DCNN model resembles that of normal DCNN models, which is as shown in Fig. 1.

The vibration signal collected is usually a long time series. It is impossible for a long time series to model and mine the fault features. Therefore, the first step is cutting off time series into several short time series, at the same time, increasing the number of samples. This is named 'data augmentation' in the field of computer vision.

In the next several layers, the convolution layer and pooling layer will appear alternately. On the one hand, multiple fault features can be extracted automatically, on the other hand, the length of the time series could be reduced greatly.

After general convolution and pooling layer, the series will be put into Multi-Scale convolution layer. Convolutional kernels of different sizes will be used to extracted different-size features in parallel. Then it will be continually convolved and pooled separately before being concatenated into a set of neurons. Multi-Scale convolution layer can increase the depth and width of the networks without a significant increasing of the number of network parameters, which leaves enough much for continually deepening the number of layers of the neural network.

Finally, the values of neurons after Multi-Scale convolution layer will be put into the classification stage, which is composed of two fully-connected layers for classification. It is worth mentioning that softmax function transforms values of the neurons in output layer to satisfy the probability distribution, and the values after transformation represent the possibility of different bearing health conditions.

B. *Multi-Scale convolution layer*

The proposed MS-DCNN is just a new perspective, where the concept of "inception" is combined closely with 1d convolutional neural networks, namely, convolution kernel of different sizes are used to extract different-scale features in parallel, as shown in Fig. 2. In addition, 1 x 1 convolutions in Multi-Scale convolution layer have dual purpose: reducing dimensions and increasing the depth and width of the networks without a significant increasing needs for computing resources.

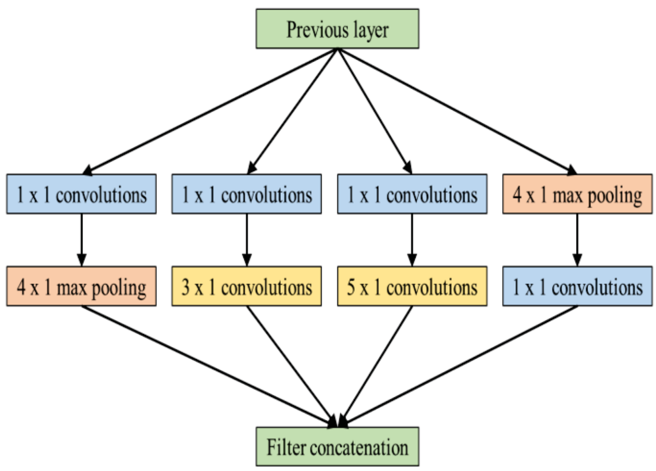

Fig. 2. *Multi-Scale convolution layer*

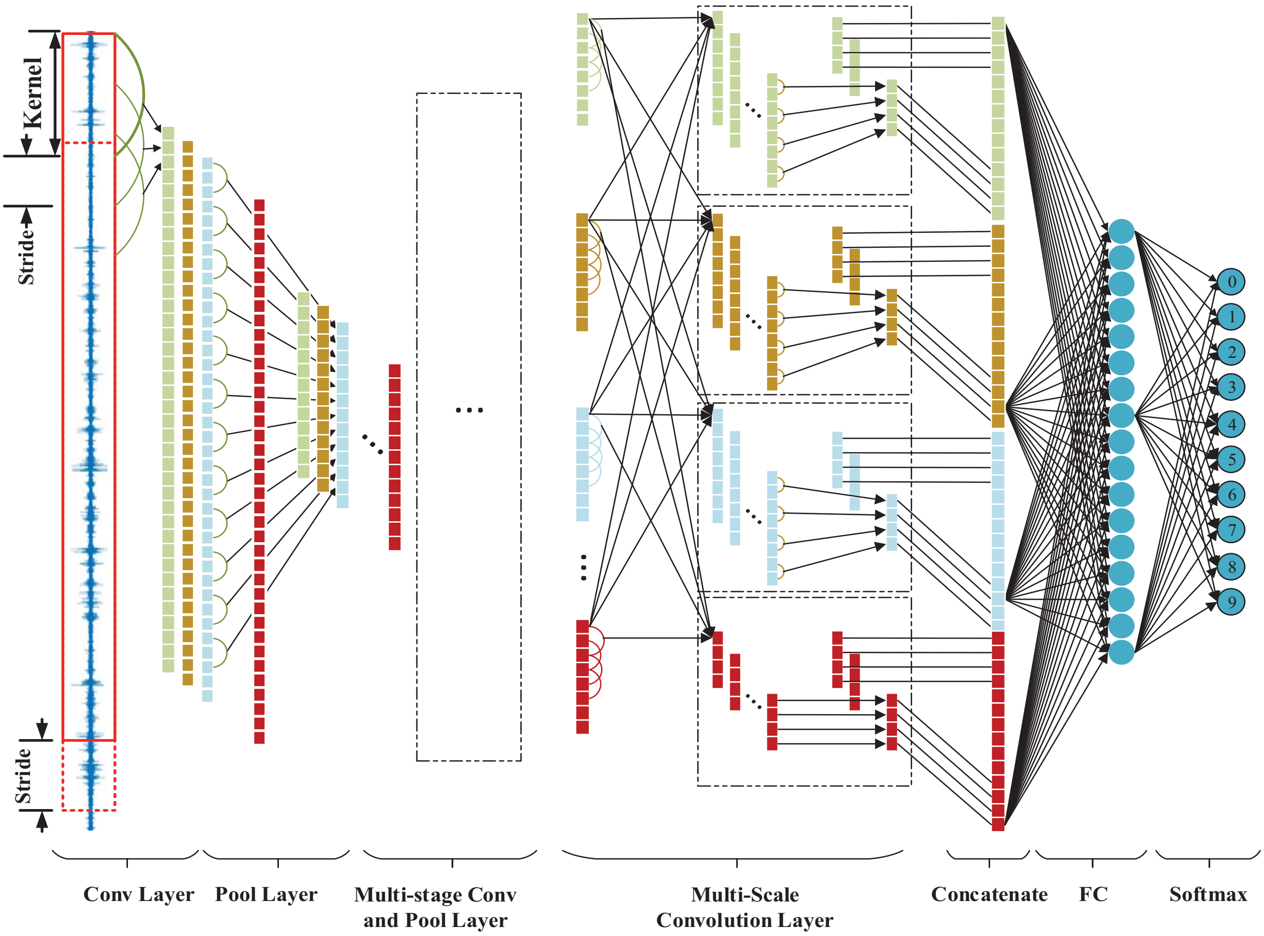

Fig. 1. *Architecture of the proposed MS-DCNN model*

C. *Data Augmentation*

With sufficient raw vibration signals with labels, the MS-DCNN model can be trained by supervised learning. Considering that MS-DCNN model has a mass of parameters to learn, it will easily get overfitting without sufficient training samples. As for fault diagnosis of REBs, the technique of data augmentation is also acquired to achieve higher classification accuracy for lack of fault samples. When the length of vibration signal is fixed, the number of samples is dependent on the sliding stride and the length of each sample, as shown in Fig. 3. In general, the length of each sample is fixed as 1~2 period of vibration. Many hyper-parameters in MS-DCNN model, such as kernel size, numbers and placement location need to be subtly adjusted.

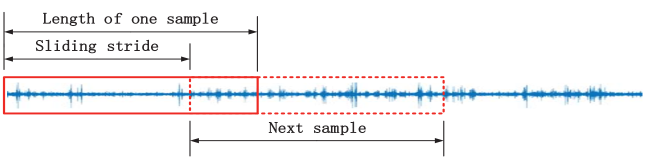

Fig. 3. *Diagram of the data augmentation*

IV. EXPERIMENTAL STUDY

A. *Data Description*

To investigate the effectiveness of the proposed MS-DCNN model, experiments were carried out based on bearing dataset from the Case Western Reserve University (CWRU) Bearing Data Center [22]. The original experiments data was collected at 12,000 samples per second from the accelerometers of the motor driving mechanical system, as shown in Fig. 4. The data point collected per circle can be inferred from frequency of data collection and rotating speed, and the formula can be concluded as: sampling count/per circle= sampling frequency *60/ rotating speed = 12000*60/ 1797 = 400. Therefore, the length of each sample can be fixed as 400, and the sliding stride can be flexibly adjusted according to the length of vibration signal and the number of samples.

The DE bearing data of the normal (ID 97), inner race fault (ID 105, 169 and 209), outer race fault (ID 130, 197 and 234), and rolling element fault (ID 118, 185 and 222) conditions were acquired for model training, where the fault diameters were selected to be 7 mils, 14 mils, and 21 mils, as shown in Table I.

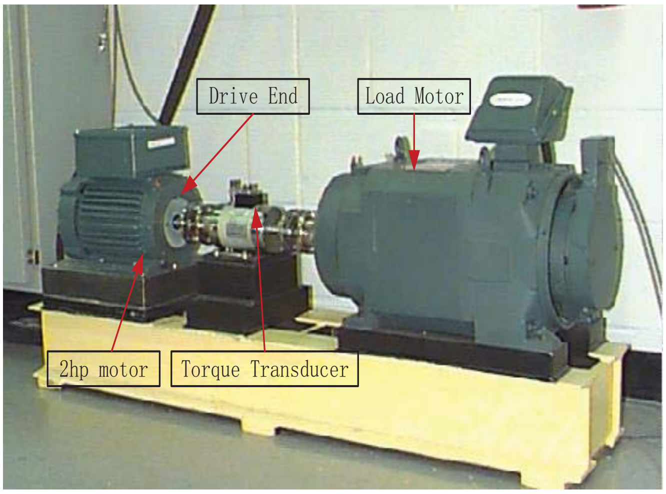

Fig. 4. *Test stand for acquiring vibration signals from REBs*

For multi-state multi-scale fault classification, each class is provided 10000 samples, among them, 9000 samples acts as training set, and remaining 1000 samples acts as testing set. Time Domain waveform of multi-state multi-scale vibration signals, including normal signal, B-fault signal (7, 14, 21), IR-fault signal (7, 14, 21), and OR-fault signal (7, 14, 21), can be seen in Fig. 5.

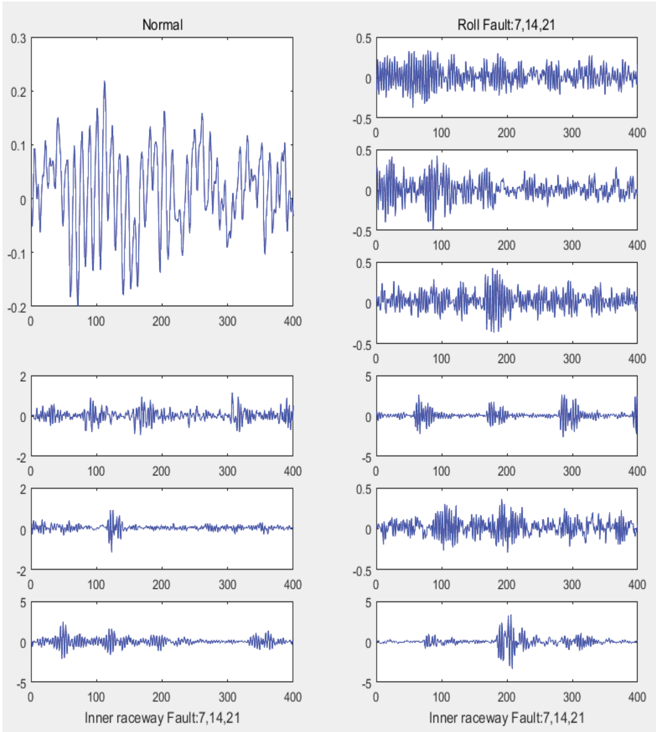

Fig. 5. *Time Domain waveform of multi-state multi-scale vibration signals*

TABLE I. THE DE BEARING DATA

Fault Location		*None*	*Ball*			*Inner Raceway*			*Outer Raceway*		
Category Labels		0	1	2	3	4	5	6	7	8	9
Fault Diameter(mil)		0	7	14	21	7	14	21	7	14	21
Dataset1	*Train*	9k(97)	9k(118)	9k(185)	9k(222)	9k(105)	9k(169)	9k(209)	9k(130)	9k(197)	9k(234)
	Test	1k(97)	1k(118)	1k(185)	1k(222)	1k(105)	1k(169)	1k(209)	1k(130)	1k(197)	1k(234)

B. Case Study: Multi-state Multi-size Fault Classification

For multi-state multi-size fault classification, the method based on 1d-DCNN, 2d-DCNN and MS-DCNN has been studied. The detailed structure parameter and classification accuracy is listed in Table II. It is observed that the basic components of these models are the same, but the structure organization mode of them have significant difference. The MS-DCNN model appears different because of its Multi-Scale convolution layer. Compared with 1d-DCNN and 2d-DCNN, MS-DCNN achieves better diagnostic results. The training process and the confusion matrix diagram of these methods are shown in Fig. 6~8. It can be observed that the mean accuracy rate of them are 98.57%, 98.25% and 99.27%, after 100 epochs. MS-DCNN can not only achieve higher accuracy rate in the testing set, but also run more smoothly in the training process. Moreover, the minimum standard deviation is also shared by MS-DCNN. All of them benefit from its multi-scale convolution layer. It can not only broaden and deepen the neural networks to learn better and more robust feature representation, but also reduce the network parameters and the training time.

TABLE II. DETAILED STRUCTURE PARAMETER AND CLASSIFICATION ACCURACY OF EACH METHOD

model	*1d-DCNN*	*2d-DCNN*	*MS-DCNN*			
Input	400*1	20*20	400*1			
Layer1	conv1d5-32	conv2d5-32	conv1d5-32			
Layer2	pooling4-32	pooling2-32	pooling4-32			
Layer3	conv1d5-64	conv2d5-64	conv1d1-5	conv1d1-5	conv1d1-5	pooling4-1
Layer4	pooling4-64	pooling2-64	pooling4-5	conv1d3-5	conv1d5-5	conv1d1-5
Layer5	flatten-1600	flatten-1600		pooling4-5	pooling4-5	

Layer6	dense-100	dense-100	flatten-125	flatten-125	flatten-125	flatten-125
Layer7	dense-10	dense-10	flatten-500			
Layer8			dense-100			
Layer9			dense-10			
parameters	171606	213206	52172			
Mean accuracy	0.9857	0.9825	0.9927			
Standard deviation	0.0073	0.0041	0.0023			

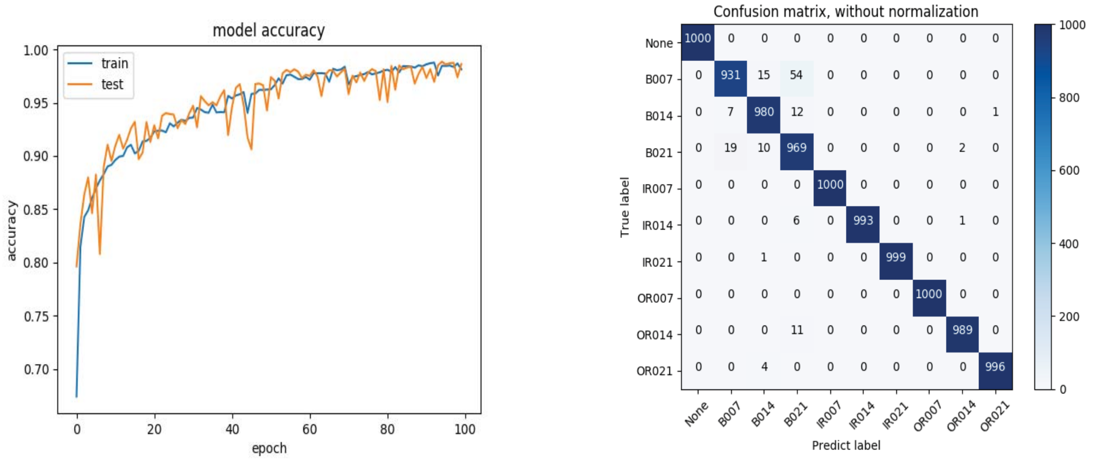

Fig. 6. *Training process and Confusion Matrix of 1d-DCNN model*

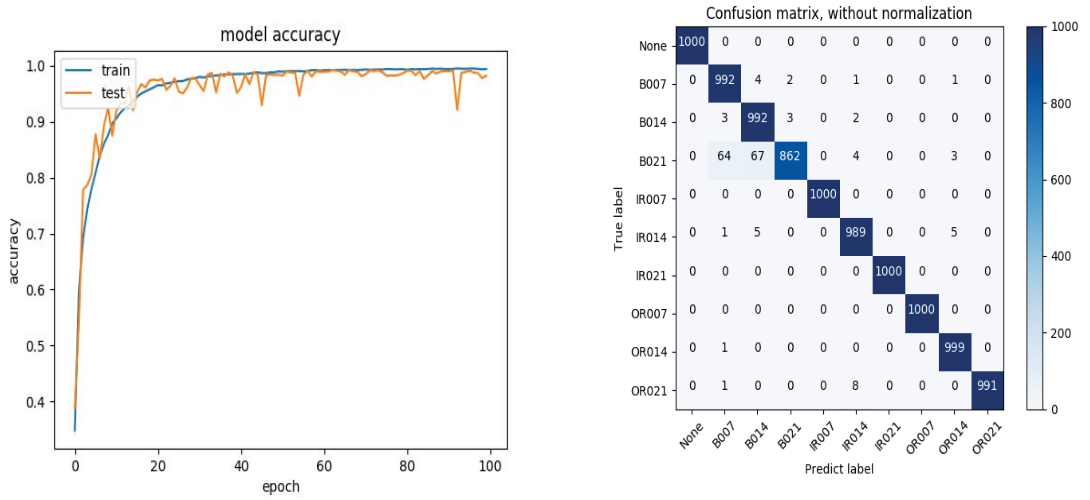

Fig. 7. *Training process and Confusion Matrix of 2d-DCNN model*

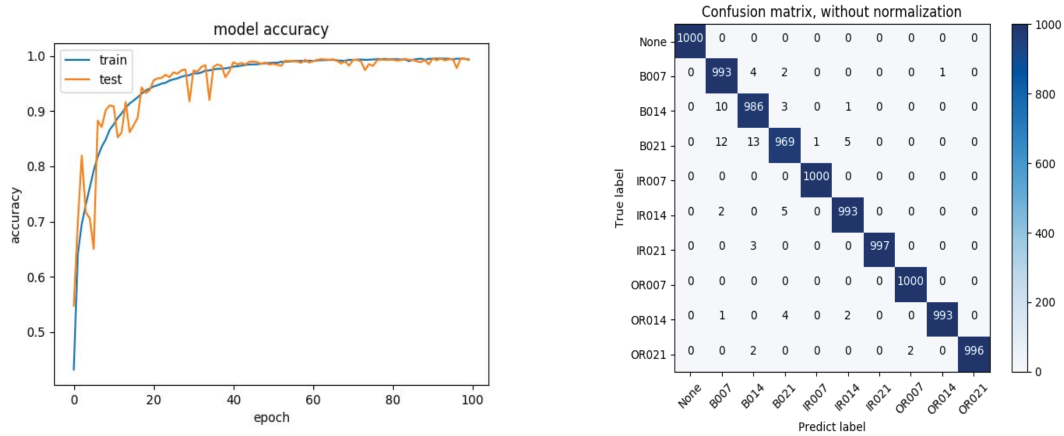

Fig. 8. *Training process and Confusion Matrix of MS-DCNN model*

V. Conclusion

The proposed MS-DCNN successfully combines the concept “inception” and one-dimensional deep convolutional neural networks. Compared with 1d-DCNN and 2d-DCNN, MS-DCNN can not only achieve higher accuracy rate in the testing set, but also run more smoothly in the training process. The merits of the proposed method are summarized as follows: (1) it can operate directly on the raw vibration signal, the real-time performance of the algorithm is enhanced to a certain extent. (2) Multi-Scale convolution layer could broaden and deepen the neural networks to learn better and more robust feature representations, meanwhile, reduce the network parameters and the training time. (3) It greatly enriched the processing method of time series.

Acknowledgment

This work was supported by financial support of National Science Foundation of China (No. 51435009), National Technology Support Program of China (No. 2015BAF12B02) and Shanghai Aerospace Science and Technology Innovation Fund (No. SAST2016048).

References

[1] Smith W A, Randall R B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study [J]. Mechanical Systems & Signal Processing, 2015, s 64–65:100-131.

[2] Qiu H, Lee J, Lin J, et al. Robust performance degradation assessment methods for enhanced rolling element bearing prognostics[J]. Advanced Engineering Informatics, 2003, 17(3):127-140.

[3] Li B, Chow M Y, Tipsuwan Y, et al. Neural-network-based motor rolling bearing fault diagnosis[J]. Industrial Electronics IEEE Transactions on, 2000, 47(5):1060-1069.

[4] Prieto M D, Cirrincione G, Espinosa A G, et al. Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical-Time Features and Neural Networks[J]. IEEE Transactions on Industrial Electronics, 2013, 60(8):3398-3407.

[5] Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(8):1798-828.

[6] Hinton G E, Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1527.

[7] Jia F, Lei Y, Lin J, et al. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data[J]. Mechanical Systems & Signal Processing, 2016, 72-73:303-315.

[8] Guo L, Gao H, Huang H, et al. Multifeatures Fusion and Nonlinear Dimension Reduction for Intelligent Bearing Condition Monitoring[J]. Shock and Vibration,2016,(2016-2-28), 2016, 2016:1-10.

[9] Verma N K, Gupta V K, Sharma M, et al. Intelligent condition based monitoring of rotating machines using sparse auto-encoders[C]// Prognostics and Health Management. IEEE, 2013:1-7.

[10] Lu C, Wang Z Y, Qin W L, et al. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification[J]. Signal Processing, 2017, 130(C):377-388.

[11] Lu C, Wang Z, Zhou B. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification [M]. Elsevier Science Publishers B. V. 2017.

[12] Tamilselvan P, Wang P. Failure diagnosis using deep belief learning based health state classification[J]. Reliability Engineering & System Safety, 2013, 115(7):124-135.

[13] Guo X, Chen L, Shen C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis[J]. Measurement, 2016, 93:490-502.

[14] Zhang W, Li C, Peng G, et al. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load[J]. Mechanical Systems & Signal Processing, 2018, 100:439-453.

[15] Zhang W, Peng G, Li C, et al. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals [J]. Sensors, 2017, 17(2):425.

[16] Zheng Y, Liu Q, Chen E, et al. Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks[M]// Web-Age Information Management. Springer International Publishing, 2014:298-310.

[17] Lin M, Chen Q, Yan S. Network In Network[J]. Computer Science, 2013.

[18] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// Computer Vision and Pattern Recognition. IEEE, 2015:1-9.

[19] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. 2015:448-456.

[20] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 2001, 86(11):2278-2324.

[21] Cui X, Goel V, Kingsbury B. Data augmentation for deep convolutional neural network acoustic modeling[C]// ICASSP 2015 - 2015 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2015:4545-4549.

[22] http://csegroups.case.edu/bearingdatacenter/pages/download-data-file