

Bearings Fault Diagnosis Based on Convolutional Neural Networks with 2-D Representation of Vibration Signals as Input

Zhang Wei, Peng Gaoliang and Li Chuanhao

School of mechanics and electronics, Harbin Institute of Technology, Harbin, P. R. China

Abstract. Periodic vibration signals captured by the accelerometers carry rich information for bearing fault diagnosis. Existing methods mostly rely on hand-crafted time-consuming preprocessing of data to acquire suitable features. In this paper, we use an easy and effective method to transform the 1-D temporal vibration signal into a 2-D image. With the signal image, convolutional Neural Network (CNN) is used to train the raw vibration data. As powerful feature extractor and classifier for image recognition, CNN can learn to acquire features most suitable for the classification task by being trained. With the image format of vibration signals, the neuron in fully-connected layer of CNN can see farther and capture the periodic feature of signals. According to the results of the experiments, when fed in enough training samples, the proposed method outperforms other common methods. The proposed method can also be applied to solve intelligent diagnosis problems of other machine systems.

1. Introduction

As core components of rotating mechanism, the health conditions of rolling element bearings, for example, the fault diameters in different places under different loads could have a direct impact on the performance, stability and life span of the mechanism. To prevent possible damages, a real-time monitoring of vibration is needed during the operation of rotating mechanism. The condition signals of the rolling element bearings are collected by the sensors, and then intelligent fault diagnosis methods are implemented to recognize the fault types [1]-[3]. Most intelligent fault diagnosis methods can be divided into two steps, namely, feature extraction and classification. Features such as mean, median, kurtosis, peak to peak, minimum, maximum, standard deviation, absolute mean, skewness and root mean square of the vibration signals are common used to describe the condition of bearings [4]. In recent years, machine learning methods such as support vector machine (SVM) [5] and neural networks have become dominant methods to predict the fault types with the features extracted from the raw temporal vibration signals.

Over the last five years, with the fast development of deep learning especially the convolutional neural networks [6] (CNNs), image classification has achieved incredible success. This success came from the efficient use of GPUs, ReLUs, dropout and data augment. CNNs are now the dominant approach for almost all recognition and detection tasks and are comparable to human beings on some tasks [7]. In order to implement CNN on the 1-D signal, common method is to treat the signal as an image with height equals 1, which is already used in speech

recognition [8]. However, the vibration signal is different from the speech signal. The vibration signal is periodic, which means the data in time t is not only correlated to the neighboring data but also the data collected in several cycles' intervals. In order to capture the features of periodic signals, a 2-D representation of these signals is used in this paper. This representation is first used to analyze and compress the power-quality event data [9], this 2-D transformation method is quite easy and intuitive, which will be shown in next section. With a signal image, CNN can easily extract the periodic features with small convolution and pooling kernel. We show that a segment of long periodic signals from any initialization can be classified with CNNs with the transformed image input format.

The remainder of this paper is organized as follows. The intelligent diagnosis method with new input format based on CNNs is introduced in Section 2. Some experiments are conducted to compare our methods with some common methods. Then discussions about the results of the experiments are presented in Section 3. We draw the conclusions and summarize the future work in Section 4.

2 Bearing fault intelligent diagnosis using the proposed method

2.1 A brief introduction to CNNs

The architecture of CNNs is briefly introduced in this section, more details for CNNs can be found in [6].

The convolutional neural network is a multi-stage neural network which is composed of some filter stages and one classification stage. The filter stage is designed to extract features from the inputs, which contains the convolutional layer and the pooling layer. The classification stage is a multi-layer perceptron, which is composed of several fully-connected layers. The function of each type of layer will be described as follows.

2.1.1 Convolutional layer

The convolutional layer convolves the input local regions with filter kernels, then generates the output features by the activation unit. Each filter uses the same kernel which is also known as weight sharing, to extract the local feature of the input local region. One filter corresponds to one frame in the next layer, and the number of frames is called the depth of this layer. We use K_i^l and b_i^l to denote the weights and bias of the i -th filter kernel in layer l , respectively, and use $X^l(j)$ to denote the j -th local region in layer l . The convolutional process is described as follows:

$$y_i^{l+1}(j) = K_i^l * X^l(j) + b_i^l \quad (1)$$

where the notation $*$ computes the dot product of the kernel and the local regions, and $y_i^{l+1}(j)$ denotes the input of the j -th neuron in frame i of layer $l + 1$.

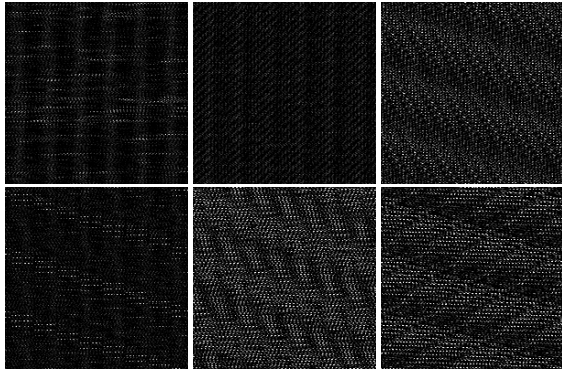


Figure 1. Images of the vibration signals for different bear faults

After the convolutional operation, the Rectified Linear Unit (ReLU) which computes the function $f(x) = \max\{0, x\}$ is used as the activation unit of our model to accelerate the convergence of the CNNs.

2.1.2 Pooling layer

Pooling layer is usually placed after a convolutional layer in the CNN architecture. It functions as a downsampling operation which merges semantically similar features into one and reduces the parameters of the network. The most commonly used pooling layer is max-pooling layer, which performs the local max operation over the input features. It can reduce the parameters and obtain location-invariant features at the same time. The max-pooling transformation is described as follows:

$$P_i^{l+1}(j) = \max_{\substack{(j-1)W+1 \leq t_x \leq jW \\ (j-1)H+1 \leq t_y \leq jH}} \{Q_i^l(t)\} \quad (2)$$

where $Q_i^l(t)$ denotes the value of t -th neuron in the i -th frame of layer l , W and H are the width and height of the pooling region respectively, and $P_i^{l+1}(j)$ denotes the corresponding value of the neuron in layer $l + 1$ of the pooling operation.

2.2 2-D representation of vibration signals

It is easy to transform the 1-D temporal signal into a 2-D image for recognition of CNN. Firstly, raw signal x is divided into n equal parts. Secondly, each part is aligned as the row of the signal image in sequence. This transformation can be described in mathematical format as follows:

$$I = \begin{bmatrix} x(t) \cdots x(t+n-1) \\ \vdots \\ x(t+(m-1)n) \cdots x(t+mn-1) \end{bmatrix} \quad (5)$$

where I denotes the signal image and $x(t)$ is the vibration data of time t .

In order to show the image clearly, we transform a section signal with length of 40000 into a 200×200 image. The images with different faults are shown in Fig. 1. As shown in Fig. 1, the images of signals look just like texture images. People may have problem in recognizing each fault type directly from the 1-D signal wave. However, with simple transformation of each signal, we can easily classify each signal from the signal images. With this intuition, CNN may perform better on the signal image than the 1-D raw signal.

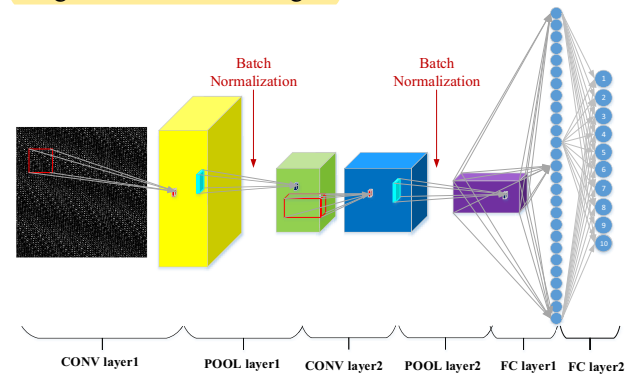


Figure 2. Architecture of the proposed CNN model

2.3 CNN-based intelligent diagnosis method

The architecture of the proposed CNN model is shown in Fig. 2. It is composed of two filter stages and one classification stage. The input of the CNN is a signal image of bearing fault vibration signal. The first convolutional layer extracts features from the input image directly without any other transformation. The features of the images are extracted from two convolutional layers and two pooling layers. As shown in Fig. 2, with the increasing of the number of layers, the depth of the layers becomes larger while the width of each frame becomes smaller. The classification stage is composed of two fully-connected layers to accomplish the classification process. In the output layer, the softmax function is used to make

the logits of the ten neurons accord with the probability distribution for the ten different bearing health conditions. Softmax function is:

$$f(z_j) = \frac{e^{z_j}}{\sum_k^{10} e^{z_k}} \quad (3)$$

where z_j denotes the logits of the j -th output neuron.

The loss function of our CNN model is cross-entropy between the estimated softmax output probability distribution and the target class probability distribution. Let $p(x)$ denotes the target distribution and $q(x)$ denotes estimated distribution, so the cross-entropy between $p(x)$ and $q(x)$ is:

$$H(p, q) = - \sum_x p(x) \log q(x) \quad (4)$$

In order to minimize the loss function, the Adam Stochastic optimization algorithm is applied to train our CNN models. Adam implements straightforwardly, computes efficiently and requires little memory, which is quite suitable for models with big data or many parameters. The detail of this optimization algorithm can be found in [10]. In order to make CNN more robust, Batch Normalization [11] is used to cope with problem of bad initialization. As a common way to control overfitting, dropout is used in the training process.

Table 1. Description of rolling element bearing datasets

Fault location		none		Ball		Inner race		Outer race		Load (hp)	
Fault		0.007		0.014		0.021		0.007		0.014	
diameter(mm)		0.007		0.014		0.021		0.007		0.014	
Data A	Train	3000	3000	3000	3000	3000	3000	3000	3000	3000	1,2,3
no.	Test	750	750	750	750	750	750	750	750	750	
Data	Train	150	150	150	150	150	150	150	150	150	1,2,3
B no.	Test	750	750	750	750	750	750	750	750	750	

3 Experiments and discussion

3.1 Data description

In order to train the CNN model sufficiently, huge numbers of training samples are prepared. The original experiments data was obtained from the accelerometers on a motor driving mechanical system at a sampling frequency of 12 kHz from the Case Western Reserve University Bearing Data center. There are four fault types of the bearing health. Namely, they are normal, ball fault, inner race fault and out race fault, and sizes of the fault diameters include 0.007mm, 0.014mm and 0.021mm, so there are all together ten conditions. In this experiment, each samples contain 2400 data points. The 2400×1 1-D signal is transformed into an image with size of 60×40 . Dataset A contains 30000 training data and 7500 testing data of ten different fault conditions under three loads. In order to evaluate the necessity of large training data for the CNN, dataset B containing 1500 training is also prepared. The details of all the datasets are described in Table 1.

3.2 Experimental Setup

3.2.1 Baseline system

We compare our methods with the standard ANN system with frequency features. The ANN system have one single hidden layer. The input of the networks is the normalized 2400 Fourier coefficients transformed from the raw temporal signals using fast Fourier transformation (FFT).

3.2.2 Hyper-parameters of the proposed CNN

The architecture of the proposed CNN is composed of 2 convolutional layers, 2 pooling layers, and also one fully-connected hidden layer. The parameters of the convolutional and pooling layers are shown in Table 2. The number of the neurons in the fully-connected hidden layer is 500. The experiments were implemented using *tensorflow* toolbox of Google.

Table 2. Parameters of the convolutional and pooling layers

Layer	Kernel width	Kernel height	Kernel depth	Stride	Frames
CONV1	5	5	1	1	25
POOL1	2	2	1	2	25
CONV2	5	5	25	1	50
POOL1	2	2	1	2	50

3.3 Results

After taking into account the sample quantity of each dataset, 3000 samples and 100 samples are randomly selected as the training validation dataset of A and B, respectively. Twenty trials were implemented for the diagnosis of each dataset. Fig. 3 and Table 3 show the fault recognition results of datasets A and B, and compares this to raw 1-D signal recognition using CNN directly and the baseline system. As shown in Fig. 3(a), the diagnosis accuracies of all twenty trials are over 99.7%. Actually, except for the sixth trial, the diagnosis accuracies of rest trials on dataset A are higher than 99.9%. This means that by feeding CNN with vibration signal images, a very good result can be achieved when diagnosing the fault conditions of the rolling element bearings. The accuracy of the CNN declined when trained with dataset B, which indicates that more training data will result in higher accuracy. The reason for this would be discussed later.

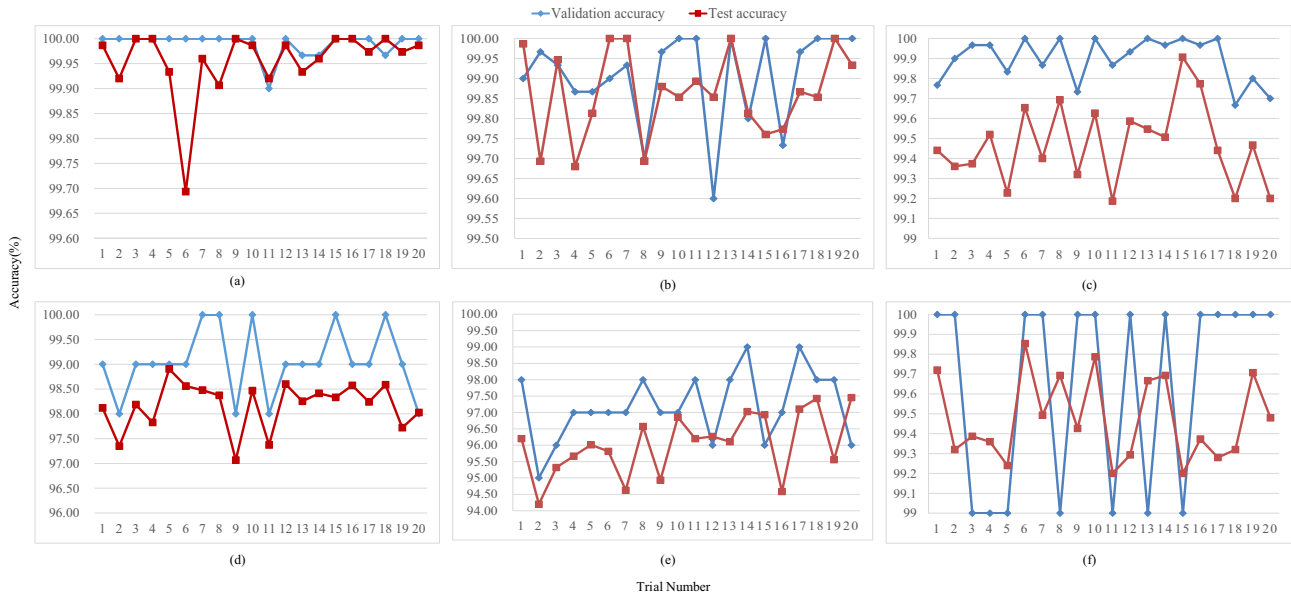


Figure 3. Fault recognition results of the proposed method and the baseline system in 20 trials. (a) and (d) are the validation and test accuracy of proposed method on dataset A and B, respectively. (b) and (e) are the validation and test accuracy of method using CNN recognition the 1-D signals, respectively. (c) and (f) are the validation and test accuracy of baseline system, respectively

Compared with other methods, the accuracies of our method outperform the others. The standard deviations of 20 trials of the proposed method on both the validation samples and test samples of dataset A are the smallest among all three methods. According to the compare above, our method is the most precise and stable one. The results on dataset B are quite different. The test accuracy of the proposed method declines near 1.8%, while the standard deviation on test samples increases by seven times. However, the results of the baseline system remain almost unchanged in terms of test accuracy and standard deviation.

Table 3. Comparison results between our method and other methods

Method	Validation accuracy	Test accuracy	Dataset
Proposed	99.99±0.02	99.95±0.07	A
Method	99.05±0.69	98.17±0.48	B
Raw 1D+CNN	99.91±0.12	99.86±0.11	A
	97.2±1.06	96.04±0.96	B
FFT+ANN	99.90±0.11	99.47±0.20	A
	99.65±0.49	99.47±0.21	B

3.4 Discussion

Some interesting points acquired from the experiments above are as follows:

The results above show that CNN performs excellently when dealing with the bearing vibration signal images. Although the performance of the FFT+ANN method is not bad, it could be quite tricky to choose between FFT or WT as the preprocessing method of the data to get a very high accuracy, because the most suitable frequency features may vary from different dataset. The advantage of our method is that it can see much further than the method which uses CNN directly on the 1-D signals. We now explore how far a neuron in the first

fully-connected layer can see using the two methods. In order to ensure that the numbers of parameters remain the same, we compress the 5×5 convolution kernel into the size of 25×1 and 2×2 pooling kernel into the size of 4×1 for the CNN architecture performed on 1-D signals. The stride for convolutional layers equals 1 for both methods. Assume the first signal point the neuron in fully-connected layer can see is $x(t)$, for the raw 1-D CNN, the farthest point it can see is $x(t + 136)$, while for the proposed method, the farthest point it can see is $x(t + 976)$. According to the analysis, the neurons in the fully-connected layer of the proposed method see much farther than the other method. **With this feature of the proposed method, CNN can better capture the periodic features than directly performing on the 1-D signals.**

When the size of the dataset is large enough to train the CNN sufficiently, CNN has better and more stable performance than common methods. The major drawback of CNN is the need for huge amount of data, as can be seen from Table 3. This may result in overfitting and it could also make it hard to be applied to industries where the labeled data is hard to acquire. However, the **labeled data of the rolling element bearing vibration is easy to acquire, so data augmentation is unnecessary.** For example, when the sample frequency is 12 kHz, the length of the data is 2400 and the stride is 1, it takes only 1 second to acquire 9601 different labeled samples. So CNN is quite suitable for the bearing fault diagnosis since huge amount of data are available.

What's more, it is shown that it is possible to recognize the bearing fault type with periodic vibration signal images as input to the CNNs. What we want to emphasize here is that the bearing vibration signal is just a representative of the periodic signals, and this method may be good at recognizing many other kinds of periodic signals, which may provide us a new approach to deal with these kinds of signals.

4 Conclusion

In this paper, we proposed a CNN architecture with signal image as input to intelligently diagnose the bearing faults. As powerful feature extractor and classifier, CNN can learn to acquire features most suitable for the classification task by being trained. Fed with signal images, CNN can learn the periodic features of the vibration signals very well. According to the results of the experiments, the proposed method is able to mine suitable fault features adaptively from the raw bearing vibration data and classify the fault classes with high accuracy and stability. Different from the deep neural networks trained by frequency spectra, the proposed CNNs can be trained directly using raw temporal signals, which means this method can be applied to solve intelligent diagnosis problems of other machine systems.

Acknowledgment

The support of National High-tech R&D Program of China (863 Program, No. 2015AA042201), National Natural Science Foundation of China (No. 51275119) in carrying out this research is gratefully acknowledged.

References

1. Jayaswal, P., Verma, S.N. and Wadhwani, A.K., 2011. Development of EBP-Artificial neural network expert system for rolling element bearing fault diagnosis. *Journal of Vibration and Control*, **17**(8), pp.1131-1148.
2. Yiakopoulos, C.T., Gryllias, K.C. and Antoniadis, I.A., 2011. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Systems with Applications*, **38**(3), pp.2888-2911.
3. Li, Y., Xu, M., Wei, Y. and Huang, W., 2016. A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement*, **77**, pp.80-94.
4. Shao, H., Jiang, H., Zhang, X. and Niu, M., 2015. Rolling bearing fault diagnosis using an optimization deep belief network. *Measurement Science and Technology*, **26**(11), p.115002.
5. Widodo, A. and Yang, B.S., 2007. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical systems and signal processing*, **21**(6), pp. 2560-2574.
6. Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
7. LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, **521**(7553), pp.436-444.
8. Palaz, D., Collobert, R. and Doss, M.M., 2013. Estimating phoneme class conditional probabilities from raw speech signal using convolutional neural networks. *arXiv preprint arXiv:1304.1018*.
9. Gerek, O.N. and Ece, D.G., 2004. 2-D analysis and compression of power-quality event data. *IEEE Transactions on Power Delivery*, **19**(2), pp.791-798.
10. Kingma, D. and Ba, J., Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
11. Ioffe, S. and Szegedy, C., Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, (2015).