Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 53 (2019) 42–50

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Rolling element bearing fault diagnosis using convolutional neural network and vibration image

Duy-Tang Hoang [a], Hee-Jun Kang [b,*]

[a] Graduate School of Electrical Engineering, University of Ulsan, Ulsan, South Korea
[b] School of Electrical Engineering, University of Ulsan, Ulsan, South Korea

Received 20 November 2017; received in revised form 4 January 2018; accepted 4 March 2018
Available online 14 March 2018

Abstract

Detecting in prior bearing faults is an essential task of machine health monitoring because bearings are the vital components of rotary machines. The performance of traditional intelligent fault diagnosis methods depend on feature extraction of fault signals, which requires signal processing techniques, expert knowledge, and human labor. Recently, deep learning algorithms have been applied widely in machine health monitoring. With the capacity of automatically learning complex features of input data, deep learning architectures have great potential to overcome drawbacks of traditional intelligent fault diagnosis. This paper proposes a method for diagnosing bearing faults based on a deep structure of convolutional neural network. Using vibration signals directly as input data, the proposed method is an automatic fault diagnosis system which does not require any feature extraction techniques and achieves very high accuracy and robustness under noisy environments.

Keywords: Bearing fault diagnosis; Convolutional neural network; Deep learning; Machine learning

1. Introduction

Rolling Element Bearings (REBs) are the essential components of rotary machines. Health conditions of REBs have considerable impacts on machines. According to a literature review, 45–55% of broken machines are caused by bearing faults (Nandi, Toliyat, & Li, 2005). Thus, condition monitoring and fault diagnosis of bearings are significant tasks in the industry. The most popular way to diagnose faults of bearings is intelligent diagnosis which employs machine learning (ML) algorithms. A typical intelligent diagnosis method includes four steps as follows: data acquisition, feature extraction, feature selection, and feature classification (Huang, 1996).

Data acquisition step collects signal from machines by sensor systems. Until now, for bearing fault diagnosis, current methodologies often use acoustic emission signals (Chacon, Kappatos, Balachandran, & Gan, 2015), motor current signals (Singh, Kumar, & Kumar, 2014), and vibration signals (Zarei, Tajeddini, & Karimi, 2014). Among those types of bearing data, vibration signal based method is the most popular approach because vibration signals are easy to measure and can provide rich dynamic information reflecting bearing health status (Kharche & Kshirsagar, 2014).

Measured vibration signals from machines contain not only useful information reflecting machine conditions but also futile noisy signals. Consequently, it is important to extract only helpful features and avoid useless information. Originally, vibration signals are temporal signals in time domain, but they can be represented in frequency domain

* Corresponding author.
E-mail address: hjkang@ulsan.ac.kr (H.-J. Kang).

https://doi.org/10.1016/j.cogsys.2018.03.002
1389-0417/

and time-frequency domain. Correspondingly, features of vibration signals can be extracted from time domain (Samanta & Al-Balushi, 2003), frequency domain (Malhi & Gao, 2004), and time-frequency domain (Lou & Loparo, 2004; Yen & Lin, 2000). In time domain, features can be extracted by Root Mean Square, Kernel Density Estimation, Crest factor, Crest-Crest Value and Kurtosis (Prieto, Cirrincione, Espinosa, Ortega, & Henao, 2013). In frequency domain, Fourier Transform is the most popular tool (Lin, Ye, Huang, & Su, 2016), while in time-frequency domain, besides Short-time Fourier Transform method (Li, Sanchez, Zurita, Lozada, & Cabrera, 2016), features can be extracted by Wavelet Packet Transform (Hemmati, Orfali, & Gadala, 2016), Dual-tree Complex Wavelet Transform (Van & Kang, 2016). In addition, there are some other methods for extracting features such as Intrinsic Mode Function (Pandya, Upadhyay, & Harsha, 2013), Hilbert Huang Transform (HHT) (Osman & Wang, 2016), and Empirical Mode Decomposition (Van & Kang, 2015).

After the feature extraction step, the dimensionality of the feature set should be reduced because there is no guarantee that all features are equally useful in reflecting machine health (Shen, Wang, Kong, & Peter, 2013) and a high dimensional feature set not only weakens the performance but also slows down the learning process of the adopted classifier in the system. To address this problem, there are a lot of discriminant feature analysis techniques have been proposed. In general, there are two approaches can be used to select discriminant features (Van & Kang, 2016). In the first way, a subset of the original feature set will be generated based on the transformation of the existing features. Some well-known methods of this approach are Principal Component Analysis (PCA) (Malhi & Gao, 2004), Independent Component Analysis (ICA) (Hyvärinen & Oja, 2000). In the second method, based on some evaluation criteria, the original features will be evaluated to select most discriminant features. In this approach, a lot of algorithms can be applied such as Sequential Forward Selection (SFS), Sequential Backward Selection (SBS) (Guyon & Elisseeff, 2003), or we can exploit Genetic Algorithm (GA) (Yang & Honavar, 1998) and Particle Swarm Optimization (PSO) (Xue, Zhang, & Browne, 2013).

ML is often exploited to solve pattern recognition problems. There are two major types of ML, includes unsupervised learning and supervised learning. Unsupervised learning algorithms such as PCA and ICA are often used in feature extraction step of pattern recognition problems. Among supervised learning algorithms, Artificial Neural Network (ANN) is the most popular method with a huge number of models such as Recurrent Neural Network (Zaremba, Sutskever, & Vinyals, 2014), Flexible Neural Tree (Bao, Chen, & Wang, 2014; Bao, Wang, & Chen, 2017), Radial Basis Neural Network (Huang & Du, 2008), and Wavelet Neural Network (Pindoriya, Singh, & Singh, 2008). Besides ANN, Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Hidden Markov Model (HMM), and ensemble learning (Bao, Huang, Yuan, & Huang, 2017; Brown, 2011) are also popular supervised algorithms. In the last step of a general intelligent diagnosis, those supervised learning algorithms are often employed. Amaral et al. (2013) made spectral images from vibration signals, then these images were enhanced by a 2-D averaging filter and binary image conversion. Finally, an ANN was exploited to classify those enhanced spectral images. Li, Meng, Ye, and Chen (2008) proposed the method using higher-order statistics features based on Discrete Wavelet Transform. After obtaining the higher-order statistics features, a kNN was used to identify types of bearing faults.

Traditional ML algorithms have been employed widely for a long time in machine fault diagnosis. However, in this approach, some drawbacks are existing. First, the classification accuracy mainly depends on the feature extraction step, which requires signal processing techniques and expert knowledge (Li et al., 2015). Thus, for every specific fault diagnosis task, feature extractor must be redesigned. The second drawback is that the traditional ML algorithms have shallow architectures with simple structures which limit the capacity of the classifiers to learn complex non-linear relationship in fault diagnosis issues (Jia, Lei, Lin, Zhou, & Lu, 2016).

Recently, deep learning (DL) emerged as the hottest trend in ML research. DL are algorithms which employ deep architectures that can learn multiple levels of data representations that correspond to different levels of abstraction (Deng et al., 2014). With the capacity of automatically learning multiple complex features from the input data, DL algorithms have great potential to overcome the disadvantages of traditional ML as mentioned above. There are a lot of DL models have been proposed such as Recurrent Neural Network (RNN), Deep Belief Network (DBN), Deep Boltzmann Machine (DBM), Stacked Auto Encoder (SAE), and Convolutional Neural Network (CNN). DL has great promise in many practical application (Dean et al., 2012), including computer vision (Krizhevsky, Sutskever, & Hinton, 2012; Szegedy et al., 2015), natural language processing (Mikolov, Karafiát, Burget, Cernocky‘, & Khudanpur, 2010), medical image analysis (Brosch et al., 2016; Greenspan, van Ginneken, & Summers, 2016), and machine health monitoring (Tamilselvan & Wang, 2013). In machine health monitoring, since 2015, a lot of researchers have tried to exploit DL models to diagnose bearing faults. Jia et al. used SAE to extracted features from signals in frequency domain (Jia et al., 2016). In an introductory paper, Chen et al. proposed three deep models, includes DBN, DBM and SAE for bearing fault diagnosis (Chen et al., 2017). They applied two approaches to learning fault features, the first one is to use deep model with raw signals in time domain directly. In the second approaches, low-level features were extracted from time domain, frequency domain, or time-frequency domain, then deep models were employed to learn higher-level features from those low-level features. L. Eren used a one-dimensional CNN to diagnose bearing faults, using raw signals directly in time domain (Eren, 2017).

In the family of DL algorithms, DBM and DBN are based on Restricted Boltzmann Machine (RBM), SAE is based on Autoencoder (AE), all of them are unsupervised learning algorithms, while CNN is a supervised learning method. Initially, with three key architectural ideas: local receptive fields, weight sharing, and sub-sampling in spatial domain, CNN is suitable for processing 2-D data (Phung & Bouzerdoum, 2009). In machine health monitoring, some researchers have tried to apply one-dimensional CNN models (Chen et al., 2017; Eren, 2017; Jia et al., 2016). However, it is much easier to extract information from data in a high dimension (Ding et al., 2017). Being motivated by this fact, in this paper, we propose a CNN model (VI-CNN), for diagnosing bearing faults using 2-D form of vibration signals. First, vibration signals in time domain are transformed into 2-D form, called vibration images. After that, a CNN will be used to identify faults of bearing through vibration image classification. To verify the effectiveness of the proposed method, we conduct experiments with bearing data from Case Western Reverse University (Loparo, 2005).

The remainder of this paper is structured as follows. Section 2 shortly introduces about CNN. The proposed fault diagnosis method is explained in Section 3. Section 4 describes experiments. Finally, conclusions are drawn in Section 5.

2. Convolutional neural network

CNN is a NN with feed-forward structure. CNN has three important characteristics that make its strength in 2-D analysis, includes local receptive fields, weight sharing, and sub-sampling in the spatial domain (Phung & Bouzerdoum, 2009). A typical CNN consists of three types of layers: convolutional layer (CL), sub-sampling layer (SL), and fully-connected layer (FL). This section describes each layer in the architecture of a CNN and its mathematical model.

2.1. Convolutional layer

The CL convolves the input with its kernels. The input of a CL is the output of the previous layer. Every kernel has the same size and extract the local feature of the input local region, which is often called as weight-sharing (Phung & Bouzerdoum, 2009). The results of convolutional operations are put through the activation function to obtain the output. Recently, Rectified Linear Unit (ReLU) has been widely used as activation function because of it low computation requirement and high-speed training. In general, the mathematic model of CL can be described by the following equation:

$$x_j^l = \mathbf{f}\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{1}$$

where $(*)$ represents the convolution operation; M_j represents a selection of input maps; l is the l_{th} layer in the network; k is the kernel matrix with size $S \times S$; $\mathbf{f}$ is a nonlinear activation function.

2.2. Sub-sampling layer

Right behind each CL, a SL is applied. SL reduces the size of input features. It also decreases the number of parameters in the network. Moreover, SL makes the representation become invariant with a small translation of the input. The SL can be described mathematically by the following equation:

$$x_j^l = \mathbf{f}\left(\beta_j^l \mathbf{down}\left(x_j^{l-1} + b_j^l\right)\right) \tag{2}$$

where **down(.)** represents a sub-sampling function. Typically, this function sums over each distinct n-by-n block in the input image so that the output image is n-times smaller along both spatial dimensions. Each output map is given its own multiplicative bias β and additive bias b. The down-sampling function can be a max-sampling or an average-sampling. The max-sampling partitions the input image into a set of non-overlapping rectangle and, for each such sub-region, outputs the maximum value. In the case of average-sampling, the average value will be released as output.

2.3. Fully-connected layer

FL is a traditional feed-forward neural network that uses softmax function as activation function in the output. All neurons in this layer are connected to all activations in the previous layer. The purpose of the FL is to collect all features from the previous feature map for classification. Softmax function is used as activation function for the output layer. Softmax function takes a vector of arbitrary real value scores and squashed it to a vector of values between zero and one. Softmax function is defined as follow:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, \quad \text{for } j = 1, \ldots, K \tag{3}$$

3. Proposed CNN based bearing fault diagnosis

As mentioned in Section 1, CNN has been employed widely and successfully in image classification tasks. In this section, we propose a deep model which exploited the ability of CNN in image classification for fault diagnosis. At first, the vibration signals are converted into gray images by a simple method proposed in (Nguyen, Kang, Kim, & Kim, 2013). A deep convolutional model automatically learns high abstract features from the gray-scale vibration images. Finally, feature classification is done by a softmax function to recognize the type of bearing faults. Fig. 1 shows the flowchart of the proposed method.

3.1. Vibration image construction

The vibration signals of bearings are 1-D data form. In this section, we explain in detail the method to transform the vibration signals into gray-scale images. We call this type of image vibration image. Fig. 2 shows the process of vibration image construction. The amplitude of each sample in the vibration signal is normalized into the range $[-1, 1]$. After that, the normalized amplitude of each sample becomes the intensity of the corresponding pixel in the corresponding image. The conversion between the normalized amplitude of sample and corresponding pixel can be described by the following equation (Nguyen et al., 2013).

$$P[i,j] = A[(i-1)*M+j] \quad (4)$$

where $i = 1:N; j = 1:M; P[i,j]$ is the intensity of the corresponding pixel (i,j) in the $M \times N$ vibration image. $A[.]$ is the normalized amplitude of the sample in the vibration signal. The number of pixels in the vibration image equals to number of samples in the vibration signal.

3.2. Image classification with CNN

Fig. 3 shows the architecture of the VI-CNN for bearing fault diagnosis. Vibration images are given into CNN, firstly, successive CLs and SLs extract features from vibration images. Multiple layers help acquire good representations of the input image and improve the performance of the network. Features extracted from previous layers are classified by a full connection layer with the softmax function.

Loss function of the proposed model is the cross-entropy between the estimated softmax output probability distribution and the target class probability distribution. Let denote the target distribution as $p(x)$ and the estimated distribution as $q(x)$. The cross entropy between $p(x)$ and $q(x)$ is defined as follow:

$$H(p,q) = -\sum_{x} p(x) \log q(x) \quad (5)$$

Once the loss function is derived, numerous optimization algorithms can be applied to train the network, such as stochastic gradient descent, gradient descent with momentum and variable learning rate, Adam Stochastic optimization (Kinga & Adam, 2015). In this paper, stochastic gradient descent is applied to train our CNN because it is easy to implement, memory-saving and computationally effective.

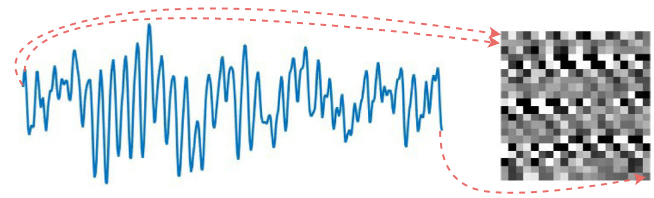

Fig. 2. Vibration image construction.

4. Experimental implementation

4.1. Test-bed description

To evaluate the performance of the proposed method, real bearing data are used. The data are obtained from the Case Western Reserve University Bearing Fault Database (Loparo, 2005). The motivation of this choice is the fact that this data has been analyzed by a number of other researchers as a benchmark data set in the field. Moreover, a public database which is accessible to the research community allows a fair comparison of the performance of the proposed algorithms. The test-bed shown in Fig. 4 includes a dynamometer (right), a 2 HP motor (left), a torque transducer/encoder (center). The test-bed also consists of a control electronics but was not shown in the figure. The motor shaft is supported by the test bearings. Single point faults were introduced to these bearings using electro-discharge machining with fault diameters of 7 mils, (1 mil = 0.001 in.). Vibration data are collected by using accelerometers, which are attached to the housing with magnetic bases. Accelerometers are placed at the 12 oclock position at both the drive housing. Vibration signals are collected using a 16 channel DAT recorder including three operating conditions. The operating conditions are considered with bearing 6205-2RS JEM SKM, which is a deep groove ball bearing type.

4.2. Data preprocessing

In our experiments, we consider four health status of bearing, includes normal status, fault in inner race, fault in ball, and fault in outer race as shown in Fig. 5. The bearing data for each bearing condition is supplied a single Matlab file. In order to have enough samples for training and testing classifiers, vibration signals are split into

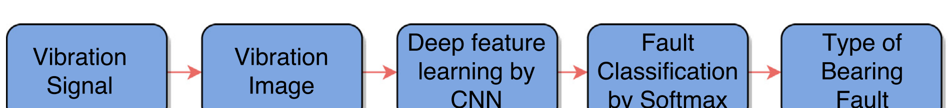

Fig. 1. Flowchart of the proposed model.

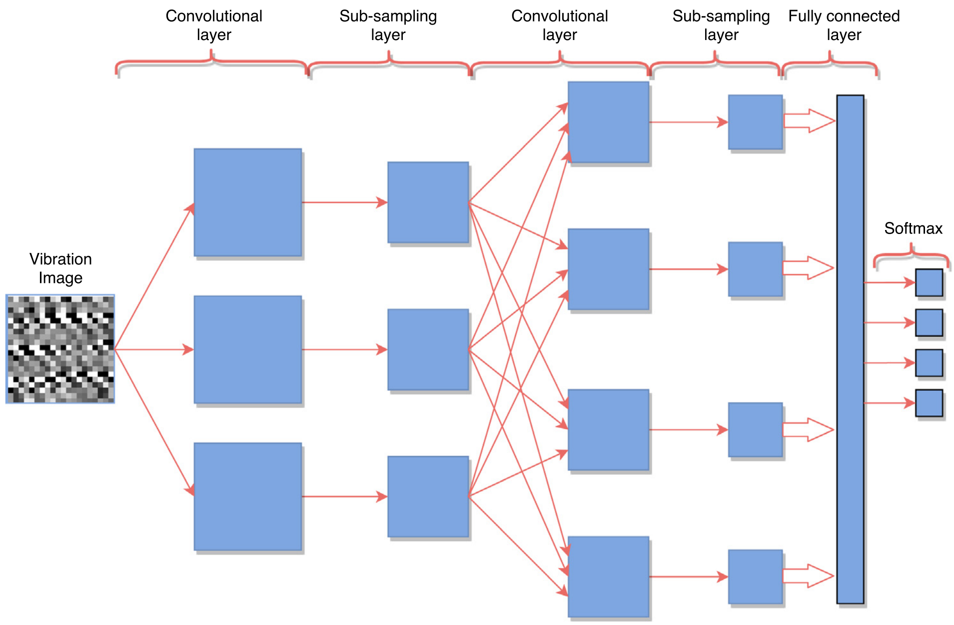

Fig. 3. Convolutional neural network.

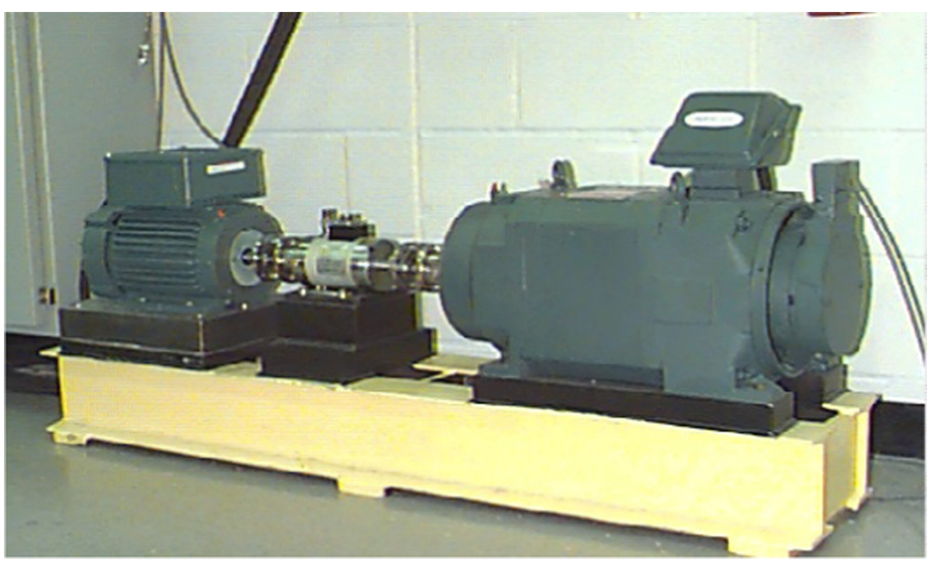

Fig. 4. Experiment setup.

segments with the same length. The process of dividing signals into segments is shown in Fig. 6. Vibration images are constructed by the method described Section 3.1. At first, each segment is normalized to have zero mean in the range $[-1, 1]$. One segment of vibration signal will be transformed into one corresponding vibration image with size $M = 20, N = 20$ by the equation as in Section 3.1. Correspondingly, we obtain four types of vibration image as shown in Fig. 7. The test-bed is operated under four load conditions, includes $0\ hp, 1\ hp, 2\ hp$ and $3\ hp$. We build four corresponding datasets $A, B, C, and\ D$ as described in Table 1. Each dataset corresponds to one operating condition of the test-beb and contains four types of vibration image, with the number of images in each type is 606.

4.3. Hyper-parameters selection

Each hyper-parameter has a big effect on not only the classification accuracy but also the training time of CNN model. Unfortunately, until now there is no standard method for selecting appropriate hyper-parameters. In this section, we explain our simple way to select hyper-parameters based on experiments.

The first parameter should be select is the size of signal segments. The segment size must be long enough to capture localized features of vibration signals. However, too long segments will make the classifier model more complex. Moreover, in order to easily transform signal segments into vibration images, the length of segments must be a square

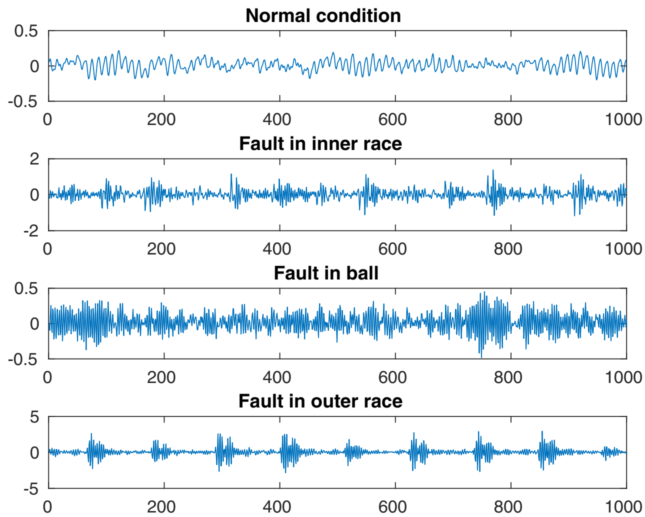

Fig. 5. Vibration signals.

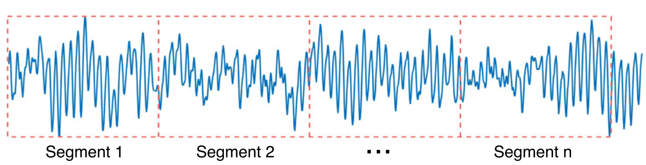

Fig. 6. Signal segmentation.

number. We choose segment size is 400, i.e., each segment contains 400 samples. Correspondingly, each vibration image has a size of 20×20 and contains 400 pixels.

A deeper structure does not ensure a better classification performance, and a two shallow structure limits the capacity of the model in learning complex features of data. Because the size of vibration images is small (20×20), we only use two CLs and two SLs. Kernel size is selected as follows. The next CL has kernel size smaller than that of the previous CL because after each CL and SL, the size of output data is reduced. The size of kernels in the first CL is 5×5, for the next CL, kernels have a size of 3×3. In the FC we fix the number of neurons equal to the number of fault types. The output layer uses softmax function.

After fixing the number of layers and the kernel size in each layer, we select the number of kernels in each layer by experiments as follows. We start with a small number of kernels. In the second CL, the number of kernels is two times higher than in the first CL. The initial CNN model will be trained by 2024 vibration images and tested by 400 vibration image from dataset A (load 0 hp). After each time of training-testing, the number of kernels will be increased by 5 and retrained with the same dataset. This process is continuously conducted until the satisfied performance is reached. Table 2 shown the classification accuracy of CNN with different kernel sizes.

Table 1
Vibration image dataset.

Name	Number of images	Load condition
A	2424	0 hp
B	2424	1 hp
C	2424	2 hp
D	2424	3 hp

Table 2
Classification accuracy with different kernel sizes.

CL1 kernel size	CL2 kernel size	Accuracy (%)
10	20	96.75
15	30	96.75
20	40	99
25	50	99.75
30	60	100

Based on the result of this experiment, we can see that the CNN model with 30 kernels in the first CL and 60 kernels in the second CL achieves 100% classification accuracy. As a result, we select the configuration of CNN model as shown in Table 3.

4.4. Evaluate under different load conditions

In Table 2, we can see that when being trained and tested with vibration images from the same load conditions (dataset A), the CNN classifiers can achieve 100% classification accuracy. However, in the real application or in the industry, machines and also their bearings have to work under various types of condition. When the working condition changes, the measured vibration signals also change. That makes the fault diagnosis more difficult, and the classifier usually must be retrained before being applied in a different working condition. In this section, we conduct experiments to evaluate the performance of the proposed diagnosis method under different load conditions, without retraining diagnosis system.

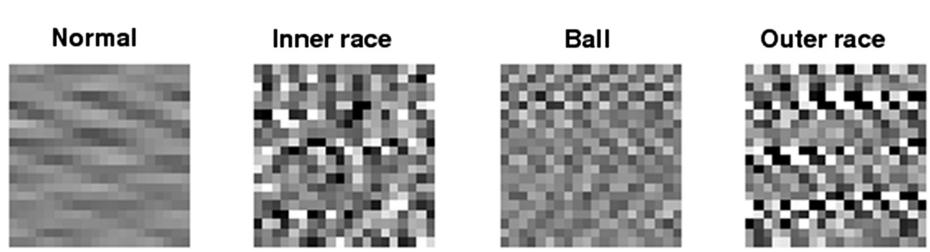

Fig. 7. Vibration images.

Table 3
Structure of CNN model.

Layer	Kernel size	Kernel number	Padding type	Output size
CL	5 × 5	30	SAME	20 × 20
SL	2 × 2	30	no	10 × 10
CL	3 × 3	60	SAME	10 × 10
SL	2 × 2	60	no	5 × 5

As mentioned in Section 4.2, we build four datasets of vibration images corresponding to four working conditions of the test-bed: 0 hp (dataset A), 1 hp (dataset B), 2 hp (dataset C), and 3 hp (dataset D). The selected CNN model as in Section 4.3 is trained by 2424 images from one dataset and then is evaluated by 1200 images from three other datasets. We compare our proposed method with the one-dimensional CNN proposed in (Eren, 2017) and the SAE model proposed in (Jia et al., 2016). Fig. 8 shows the comparison results.

In case 1, we use 2424 vibration images from dataset A to train three models, includes 1-D CNN (Eren, 2017), SAE (Jia et al., 2016) and our proposed VI-CNN models. After that, 400 vibration images samples from each data set B, C, and D (1200 in total) are used to test the trained models. Case 2, 3 and 4 are conducted by the same way with case 1 but the datasets are alternated.

In all four cases, the SAE model has the worst performance with lowest classification accuracy. In case 2, the 1-D CNN model achieves the best performance with the classification accuracy is 99.83%. In case 1, 3 and 4, the VI-CNN model achieves the best performance with the highest classification accuracy. Especially, in case 1 and 3, VI-CNN model reaches 100% classification accuracy. Through the comparisons, we can see that without retraining the whole model, our proposed method still work well when the working condition changes.

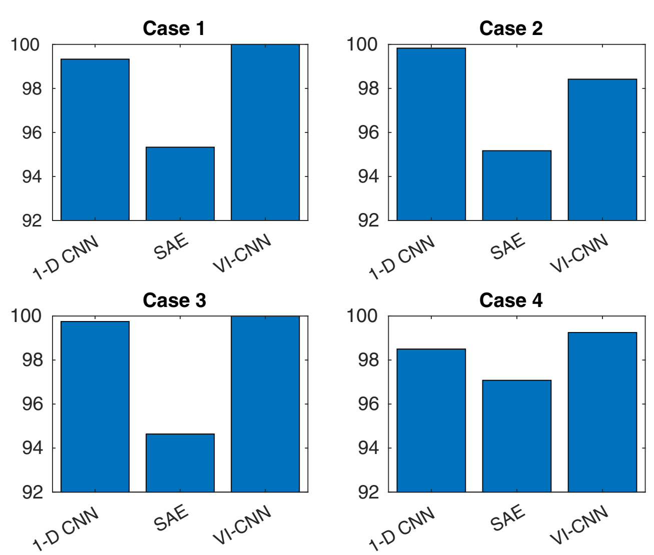

Fig. 8. Performance under different loads.

4.5. Evaluate under different noise conditions

Besides the change of working conditions, effect of noise is also a big problem which decreases the performance of fault diagnosis. In the real industrial environments, the sensory signals are contaminated by noise. In this section we consider the robustness and the accuracy of the proposed scheme under low signal-to-noise ratio (SNR) conditions, i.e., we try to detect faults of bearings with noisy signals. The additive Gaussian white noise (AGWN) with various standard variances are added to the original vibration signals to mimic the low SNR. The SNR is defined as follows:

$$SNR = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \quad (6)$$

where P_{signal} and P_{noise} are respectively the power of signal and noise in that signal. Fig. 9 shows the noisy signal which made by adding an original signal with Gaussian white noise.

The experiment is conducted as follows. The original vibration signal is added Gaussian white noise to form the noisy signals. Then these noisy vibration signals are pre-processed, split into segments and transformed into vibration images by the same way as mentioned in Section 4.2. The classifier models will work with vibration image datasets build from noisy signals.

Fig. 10 shows the fault classification accuracy of the proposed scheme with the SNR value in decibel (dB) varying from −10 dB to 0 dB. Obviously, the bigger noise power is, the more difficult to diagnose faults. In these experiments with noisy signals, the 1-D CNN shows the worst performance with very low accuracy. Under noise condition $SNR = -4$ dB, three diagnosis model still achieve 100% accuracy. But when the noise condition becomes $SNR = -10$ dB, the performances degrade significantly. With SAE model, the accuracy is 95.5% and in the case of 1-D CNN is 90.75%. However, our proposed CNN

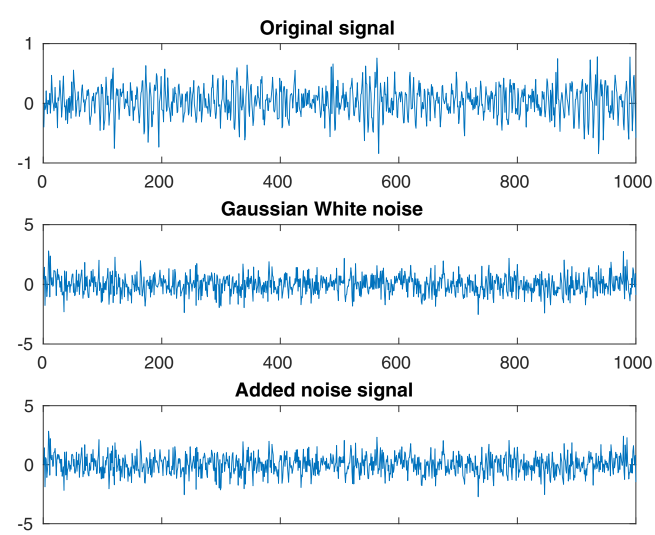

Fig. 9. Noisy signal with $SNR = -10$ dB.

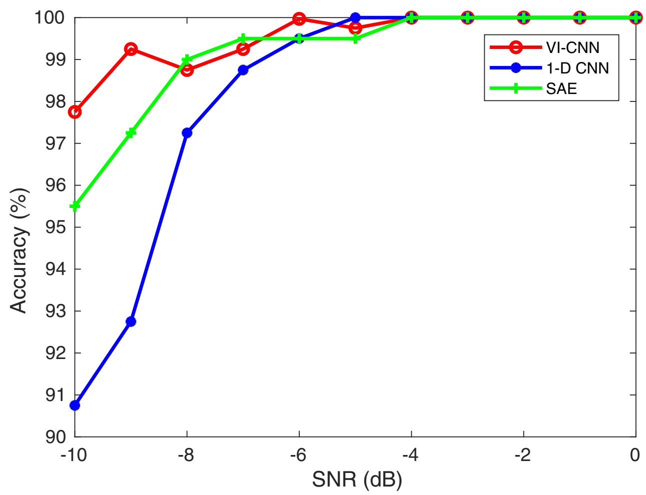

Fig. 10. Performance under different noise conditions.

model using vibration image data still achieve satisfied performance with the accuracy is 97.74%. The comparison results in this experiment show that the proposed VI-CNN model has good robustness and can work well with noisy vibration signals.

5. Conclusion

In this study, we proposed a new approach based on CNN for diagnosing faults of rolling element bearings. By transforming 1-D vibration signals into 2-D images and exploiting the effectiveness of CNN in image classification, the proposed method can achieve 100% accuracy in CWRU bearing data set.

Compared to traditional machine learning based fault diagnosis, the main advantage of the proposed methods is that it does not require the feature extraction step, but still achieve high classification accuracy. Furthermore, when the working load condition is changed, without retraining the classifier, our proposed method still achieves satisfied performance with high accuracy. Moreover, without any denoising process, the proposed method has the robustness and the capacity of tolerating noisy environment.

A simple procedure based on experiments was used to select hyper-parameters of the CNN model. However, selecting appropriate hyper-parameters to design DL algorithms for fault diagnosis is still a challenge.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF-2016R1D1A3B03930496) funded by the Ministry of Education.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cogsys.2018.03.002.

References

Amaral, T., Silva, L. M., Alexandre, L. A., Kandaswamy, C., Santos, J. M., & de Sá, J. M. (2013). Using different cost functions to train stacked auto-encoders. In *2th Mexican International Conference on Artificial Intelligence (MICAI), 2013* (pp. 114–120). IEEE.

Bao, W., Chen, Y., & Wang, D. (2014). Prediction of protein structure classes with flexible neural tree. *Bio-Medical Materials and Engineering, 24*(6), 3797–3806.

Bao, W., Huang, Z., Yuan, C.-A., & Huang, D.-S. (2017). Pupylation sites prediction with ensemble classification model. *International Journal of Data Mining and Bioinformatics, 18*(2), 91–104.

Bao, W., Wang, D., & Chen, Y. (2017). Classification of protein structure classes on flexible neutral tree. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Brosch, T., Tang, L. Y., Yoo, Y., Li, D. K., Traboulsee, A., & Tam, R. (2016). Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Transactions on Medical Imaging, 35*(5), 1229–1239.

Brown, G. (2011). Ensemble learning. In *Encyclopedia of machine learning* (pp. 312–320). Springer.

Chacon, J. L. F., Kappatos, V., Balachandran, W., & Gan, T.-H. (2015). A novel approach for incipient defect detection in rolling bearings using acoustic emission technique. *Applied Acoustics, 89*, 88–100.

Chen, Z., Deng, S., Chen, X., Li, C., Sanchez, R.-V., & Qin, H. (2017). Deep neural networks-based rolling bearing fault diagnosis. *Microelectronics Reliability*.

Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., ...Ng, A. Y. (2012). Large scale distributed deep networks. In *Advances in neural information processing systems, 2012* (pp. 1223–1231).

Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing, 3*.

Ding, X., & He, Q. (2017). Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*.

Eren, L. (2017). Bearing fault detection by one-dimensional convolutional neural networks. *Mathematical Problems in Engineering*.

Greenspan, H., van Ginneken, B., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging, 35*(5), 1153–1159.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*(Mar), 1157–1182.

Hemmati, F., Orfali, W., & Gadala, M. S. (2016). Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation. *Applied Acoustics, 104*, 101–118.

Huang, D.-S. (1996). *Systematic theory of neural networks for pattern recognition*. Beijing 201: Publishing House of Electronic Industry of China.

Huang, D.-S., & Du, J.-X. (2008). A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks. *IEEE Transactions on Neural Networks, 19*(12), 2099–2115.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*(4), 411–430.

Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing, 72*, 303–315.

Kharche, P. P., & Kshirsagar, S. V. (2014). Review of fault detection in rolling element bearing. *International Journal of Innovative Research in Advanced Engineering, 1*(5), 169–174.

Kinga, D., & Adam, J. B. (2015). A method for stochastic optimization. In *International conference on learning representations (ICLR), 2015*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems, 2012* (pp. 1097–1105).

Li, F., Meng, G., Ye, L., & Chen, P. (2008). Wavelet transform-based higher-order statistics for fault diagnosis in rolling element bearings. *Journal of Vibration and Control, 14*(11), 1691–1709.

Lin, H.-C., Ye, Y.-C., Huang, B.-J., & Su, J.-L. (2016). Bearing vibration detection and analysis using enhanced fast fourier transform algorithm. *Advances in Mechanical Engineering, 8*(10), 1687814016675080.

Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., & Vásquez, R. E. (2015). Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. *Neurocomputing, 168*, 119–127.

Li, C., Sanchez, V., Zurita, G., Lozada, M. C., & Cabrera, D. (2016). Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time–frequency ridge enhancement. *ISA Transactions, 60*, 274–284.

Loparo, K. A. (2005). *Bearing data center*. Case Western Reserve University.

Lou, X., & Loparo, K. A. (2004). Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing, 18*(5), 1077–1095.

Malhi, A., & Gao, R. X. (2004). PCA-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation and Measurement, 53*(6), 1517–1525.

Mikolov, T., Karafiát, M., Burget, L., Cernocky‘, J., & Khudanpur, S. (2010). Recurrent neural network based language model. In *Interspeech* (Vol. 2, pp. 3).

Nandi, S., Toliyat, H. A., & Li, X. (2005). Condition monitoring and fault diagnosis of electrical motors – a review. *IEEE Transactions on Energy Conversion, 20*(4), 719–729.

Nguyen, D., Kang, M., Kim, C.-H., & Kim, J.-M. (2013). Highly reliable state monitoring system for induction motors using dominant features in a two-dimension vibration signal. *New Review of Hypermedia and Multimedia, 19*(3–4), 248–258.

Osman, S., & Wang, W. (2016). A morphological Hilbert-Huang transform technique for bearing fault detection. *IEEE Transactions on Instrumentation and Measurement, 65*(11), 2646–2656.

Pandya, D., Upadhyay, S., & Harsha, S. P. (2013). Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN. *Expert Systems with Applications, 40*(10), 4137–4145.

Phung, S. L., & Bouzerdoum, A. (2009). *Matlab library for convolutional neural networks. Tech. Rep*. ICT Research Institute, Visual and Audio Signal Processing Laboratory, University of Wollongong.

Pindoriya, N., Singh, S., & Singh, S. (2008). An adaptive wavelet neural network-based energy price forecasting in electricity markets. *IEEE Transactions on Power Systems, 23*(3), 1423–1432.

Prieto, M. D., Cirrincione, G., Espinosa, A. G., Ortega, J. A., & Henao, H. (2013). Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. *IEEE Transactions on Industrial Electronics, 60*(8), 3398–3407.

Samanta, B., & Al-Balushi, K. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical Systems and Signal Processing, 17*(2), 317–328.

Shen, C., Wang, D., Kong, F., & Peter, W. T. (2013). Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. *Measurement, 46*(4), 1551–1564.

Singh, S., Kumar, A., & Kumar, N. (2014). Motor current signature analysis for bearing fault detection in mechanical systems. *Procedia Materials Science, 6*, 171–177.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ...Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition, 2015* (pp. 1–9).

Tamilselvan, P., & Wang, P. (2013). Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering & System Safety, 115*, 124–135.

Van, M., & Kang, H.-J. (2015). Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection. *IET Science, Measurement & Technology, 9*(6), 671–680.

Van, M., & Kang, H.-J. (2016). Two-stage feature selection for bearing fault diagnosis based on dual-tree complex wavelet transform and empirical mode decomposition. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 230*(2), 291–302.

Xue, B., Zhang, M., & Browne, W. N. (2013). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Transactions on Cybernetics, 43*(6), 1656–1671.

Yang, J., & Honavar, V. (1998). Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems and their Applications, 13*(2), 44–49.

Yen, G. G., & Lin, K.-C. (2000). Wavelet packet feature extraction for vibration monitoring. *IEEE Transactions on Industrial Electronics, 47*(3), 650–667.

Zarei, J., Tajeddini, M. A., & Karimi, H. R. (2014). Vibration analysis for bearing fault detection and classification using an intelligent filter. *Mechatronics, 24*(2), 151–157.

Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent neural network regularization. Available from 1409.2329.