



Performance Analysis and Enhancement of Deep Convolutional Neural Network

Application to Gearbox Condition Monitoring

Jinjiang Wang · Yulin Ma · Zuguang Huang · Ruijuan Xue · Rui Zhao

Received: 7 June 2018 / Accepted: 6 February 2019 / Published online: 20 March 2019
© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2019

Abstract Convolutional neural network has been widely investigated for machinery condition monitoring, but its performance is highly affected by the learning of input signal representation and model structure. To address these issues, this paper presents a comprehensive deep convolutional neural network (DCNN) based condition monitoring framework to improve model performance. First, various signal representation techniques are investigated for better feature learning of the DCNN model by transforming the time series signal into different domains, such as the frequency domain, the time–frequency domain, and the reconstructed phase space. Next, the DCNN model is customized by taking into account the dimension of model, the depth of layers, and the convolutional kernel functions. The model parameters are then optimized by a mini-batch stochastic gradient descent algorithm. Experimental

studies on a gearbox test rig are utilized to evaluate the effectiveness of presented DCNN models, and the results show that the one-dimensional DCNN model with a frequency domain input outperforms the others in terms of fault classification accuracy and computational efficiency. Finally, the guidelines for choosing appropriate signal representation techniques and DCNN model structures are comprehensively discussed for machinery condition monitoring.

Keywords Machinery condition monitoring · Deep convolutional neural network · Model structure · Signal representation

1 Introduction

The recent advancement in the Industrial Internet of Things (IIoTs) (Wang et al. 2015; Tao et al. 2018a), Cloud Computing (Ren et al. 2017; Matt 2018; Ye et al. 2016) and the Cyber Physical System (CPS) (Lu et al. 2014; Richter et al. 2018) reshapes modern manufacturing. Manufacturing machines are increasingly equipped with sensors to increase system reliability and improve operational performance. Thus, an unprecedented volume of manufacturing data is generated. This not only creates new opportunities for machinery condition monitoring (Tao et al. 2018b; Ye et al. 2018), but also brings challenges for handling proliferated multi-source manufacturing data.

Much effort has been put on data analytics for machinery condition monitoring. To automatically extract fault features from monitored signal, a machine learning based condition monitoring approach has been widely adopted. Some previous applications of machine learning models have been conducted (Gangsar and Tiwari 2017;

Accepted after one revision by the editors of the special edition.

J. Wang (✉) · Y. Ma
Department of Mechanical and Transportation Engineering,
China University of Petroleum, Beijing 102249, China
e-mail: jwang@cup.edu.cn

Y. Ma
e-mail: yulimmacup@gmail.com

Prof. Z. Huang (✉) · R. Xue
Chinese Machine Quality Supervision and Inspection Center,
Beijing 100102, China
e-mail: hzg36@163.com

R. Xue
e-mail: sac_tc231@163.com

R. Zhao
School of Electrical and Electronic Engineering, Nanyang
Technological University, Singapore 639798, Singapore
e-mail: RZHAO001@e.ntu.edu.sg

Heinrich and Schwabe 2018). However, conventional machine learning models highly rely on handcrafted features. The empirical feature extraction in conventional machine learning highly relies on domain expertise, which has great influence on classification performance (Nalchi-gar and Eric 2018). Additionally, conventional machine learning usually has a shallow structure with at most three layers, performs feature extraction and model construction in a separated manner, and constructs each module step-by-step (Residual et al. 2018). Thus, they have difficulties to deal with the unprecedented volume of manufacturing data (Tao et al. 2018b; Ye et al. 2018) which renders conventional condition monitoring approaches impractical.

As a breakthrough in artificial intelligence, deep learning has achieved great success in various fields in recent years. It has distinctive differences from traditional machine learning regarding feature extraction and classification, as shown in Fig. 1. Deep learning incorporates feature extraction and classification into a deep hierarchical structure. Thus, high dimensional and abstracted representation is selected automatically from raw data by means of a multi-layered structure. Therefore, deep learning provides a new solution to intelligent condition monitoring of machinery. Different deep learning models including convolutional neural network (CNN), auto encoder (AE), restricted Boltzmann machine (RBM), and Recurrent Neural Network (RNN) have been investigated in the field of machinery condition monitoring (Zhao et al. 2019; Wang et al. 2018). As a typical model, CNN attracts much attention due to its superiority in the abstract feature representation learning and high classification accuracy (Le-Cun et al. 2005).

Since CNN models were originally designed for image analysis, two-dimensional CNN models are widely adopted for machinery condition monitoring by transforming the

one-dimensional time series signal into two-dimensional signal representation (Wang 2016; Ding and He 2017; Wang et al. 2017). However, the inherent information hidden in the raw signal may be lost, and even increase the computational complexity of the CNN model. To address this issue, a one-dimensional CNN model is developed in (Liu et al. 2016) and has been applied to machinery condition monitoring. However, the literature referred to is mostly devoted to obtaining better fault diagnosis accuracy through either changing signal representation techniques under the same CNN model or alerting the structures and hyper-parameters of a certain dimensional CNN model. A comprehensive analysis of the effects of signal representation and CNN model structure is still lacking. In general, two major challenges exist for CNN based fault diagnosis. First, choosing an appropriate signal representation technique mainly relies on experience and much valuable information hidden in the raw signal may be lost during the transformation. Second, it is hard to choose a suitable CNN model structure, and there has been no attempt to study the differences of feature extraction between one-dimensional and two-dimensional CNN models in depth.

To mitigate the research gap, this paper presents a comprehensively deep convolutional neural network (DCNN) based condition monitoring framework to improve model performance. Firstly, various signal representation techniques are investigated to improve the DCNN models feature learning by transforming the time series signal into different domains, such as frequency domain, time–frequency domain, and reconstructed phase space. Next, the DCNN models with different dimensions are studied for sensitive feature extraction. The model structures are custom designed and the model parameters are then optimized by mini-batch stochastic gradient descent algorithms. Experimental studies on a gearbox test rig

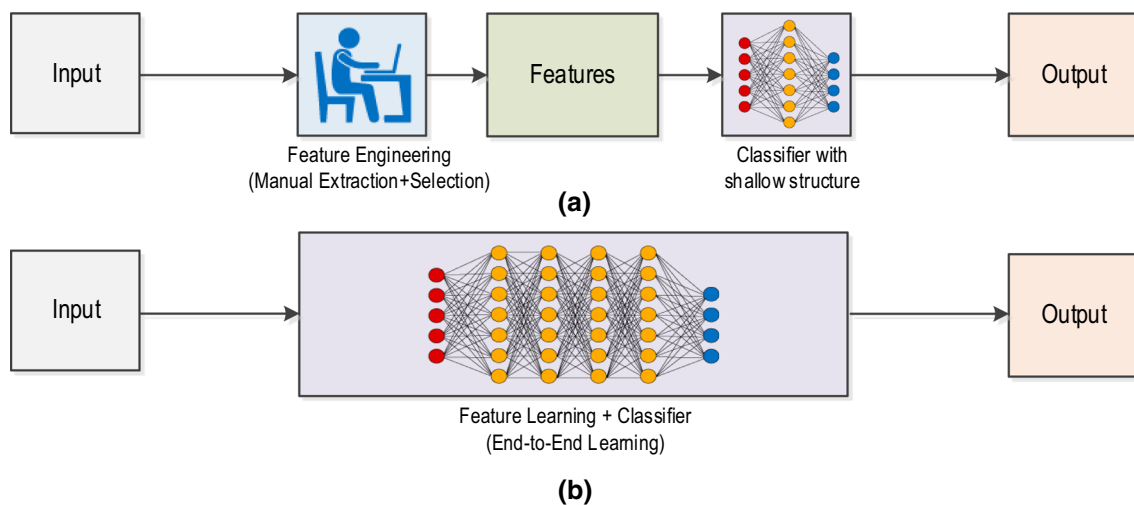


Fig. 1 Comparison between **a** conventional intelligent fault diagnosis and **b** deep learning based fault diagnosis

are utilized to evaluate the effectiveness of the presented DCNN models, and the results show that the one-dimensional DCNN model with frequency domain input outperforms the others in terms of fault classification accuracy and computational efficiency. In particular, this paper seeks to shed light on how different dimensional DCNNs work in machinery condition monitoring and provide the guidelines for choosing appropriate signal representation techniques and DCNN models.

The intellectual contribution of this paper includes: (1) a comprehensive machinery condition monitoring framework is proposed to enhance the performance of DCNN models, (2) the effects of different signal representation techniques on DCNN models are compared and the differences among different DCNN models are studied in depth, and (3) comprehensive experiments and a detailed analysis of a gearbox fault diagnosis have been performed. The rest of this paper is constructed as follows. First the theoretical background of one-dimensional and two-dimensional DCNN is briefly described in Sect. 2. A DCNN based machinery condition monitoring framework is formulated in Sect. 3. Then the framework is experimentally validated based on a gearbox fault dataset in Sect. 4, and the results are further discussed in Sect. 5. Finally, conclusions are drawn in Sect. 6.

2 Related Work

Inspired by visual neuroscience, CNN is initially designed to deal with the variability of two-dimensional shapes (LeCun et al. 1998). However, the time series signal measurements in machinery condition monitoring are usually one-dimensional. To introduce the CNN model for machinery condition monitoring, different techniques have been widely investigated which can be categorized as signal transformation and model structure alternation.

Various signal representation techniques have been investigated to generate two-dimensional model inputs. By decomposing a one-dimensional signal into a two-dimensional time-scale plane, wavelet coefficients are firstly employed as input of the two-dimensional CNN model for compressor fault diagnosis (Wang 2016). Then different wavelet transform techniques including wavelet packet energy (WPE) images and continuous wavelet transform (CWT) images have also been used as input of CNN models for machinery fault classification (Ding and He 2017; Wang et al. 2017). The image of a one-dimensional vibration time series signal is also investigated as the input of a two-dimensional CNN model for machinery fault classification (Fu et al. 2017). In order to obtain better representation of raw data, time–frequency images generated from short-time Fourier transform (STFT), wavelet

transform and Hilber–Huang transform (HHT) have been fed into a two-dimensional CNN for rolling element bearing fault diagnosis (Verstraete et al. 2017).

Moreover, the construction of CNN models has been studied with hyper-parameters optimization. An improved fault diagnosis model composed of multi two-dimensional CNN models has been investigated to classify bearing fault severity (Guo et al. 2016). Considering the difference of feature representation between images and time series signals, a specific DTS-CNN with a dislocated layer has been proposed for electric machine fault diagnosis (Liu et al. 2016). This model can extract periodic fault information between nonadjacent signals by continuously dislocating the input raw signals. Typical two-dimensional input requires a deep layer model structure and a large number of model parameters to train the CNN model, leading to a high computational complexity able to tackle high-dimensional input.

As a result, a one-dimensional CNN model has been proposed by incorporating one-dimensional convolutional kernels in the convolutional layers for machinery condition monitoring. A pre-trained one-dimensional CNN model fed with frequency domain signals has been tested regarding various types of input signals for gearbox fault diagnosis (Jing et al. 2017). A one-dimensional CNN model has also been investigated for fault diagnosis of rolling element bearing with the typical time series signal or its frequency spectrum as input (Zhang et al. 2017; Wang et al. 2018). An adaptive one-dimensional CNN model fed with pre-processed motor current signals has been investigated for motor anomaly detection, in which a variant learning rate was added to improve the feature extraction capability (Ince et al. 2016). A decentralized structural damage detection system using one-dimensional CNN with the input of raw acceleration signals has been proposed to identify the damaged joint, in which each one-dimensional CNN model is designed for a specific joint damage (Abdeljaber et al. 2017).

As discussed above, the majority of these studies has been devoted to achieving better fault diagnosis accuracy through either changing signal representation techniques or alternating the structures and hyper-parameters of a CNN model. Thus, it is of significance to develop a comprehensive framework to investigate the effects of signal representation and model construction for improved performance of machinery condition monitoring.

3 DCNN Based Condition Monitoring Framework

The developed DCNN based condition monitoring framework is shown in Fig. 2. Firstly, the different signal representation techniques are proposed by transforming the

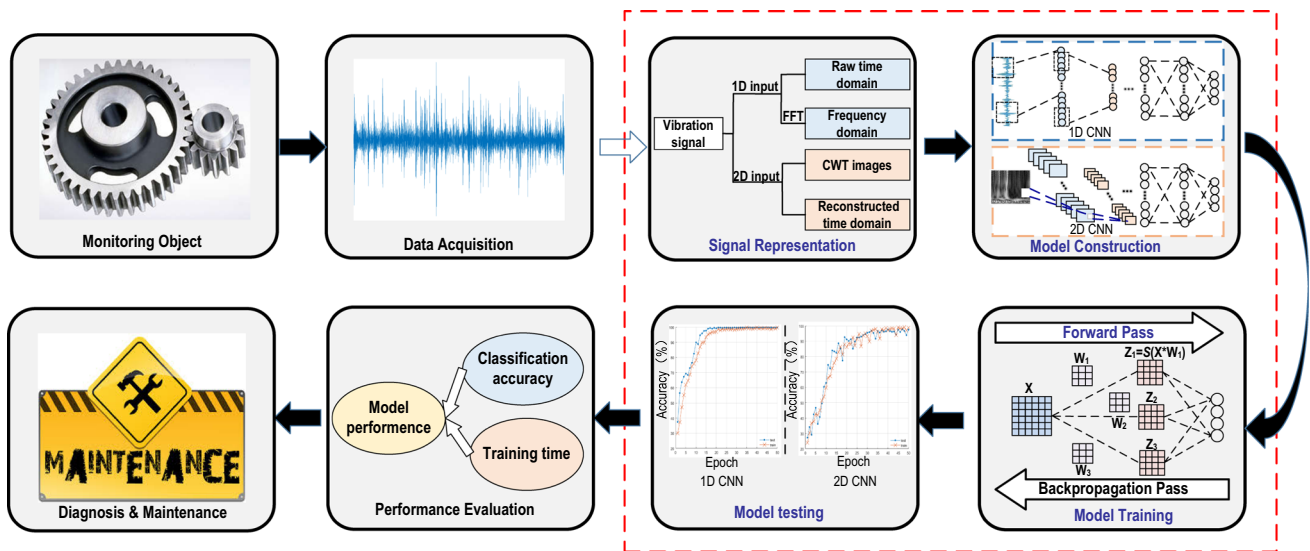


Fig. 2 DCNN based machinery condition monitoring framework

raw signals into different dimensions of model input for improving the fault sensitive features representation. Then a variety of DCNN model structures are constructed including one-dimensional and two-dimensional CNN models. The decisive factors for model construction as well as the differences in feature extraction between one-dimensional and two-dimensional DCNN models are also discussed. Finally, several typical performance metrics are designed to evaluate the performance of different DCNN models.

3.1 Signal Representation

To facilitate the feature learning of DCNN models, various signal representation techniques are employed to remove irrelevant and redundant information of raw time series signals. Frequency domain signals transformed through fast Fourier transform (FFT) are used as the input for one-dimensional DCNN. Two-dimensional inputs including time–frequency domain signals and reconstruction phase space signals are fed into two-dimensional DCNN.

Continuous wavelet transform (Konar and Chattopadhyay 2011) is used to transfer a one-dimensional vibration signal into a two-dimensional one, named wavelet scalogram. The obtained wavelet scalogram is fed into two-dimensional DCNN for machinery condition monitoring. The wavelet transform of a signal $x(t)$ with finite energy can be performed through convolution of $x(t)$ with the complex conjugate of a family of wavelets

$$wt(s, \tau) = |s|^{-1/2} \int x(t)\psi\left(\frac{t-\tau}{s}\right)dt \tag{1}$$

In the above equation, $\psi(\cdot)$ represents the complex conjugate of the scaled and shifted base wavelet. The

wavelet scalogram is defined as the square of wavelet coefficients $wt(s, \tau)$.

$$G_x(s, \tau) = |wt(s, \tau)|^2 \tag{2}$$

The wavelet scalogram can be regarded as a spectral map with constant relative bandwidth, which can reflect the time frequency information of the signal and is widely used used for machinery condition monitoring (Su et al. 2018; Yan et al. 2014).

In fact, the collected signals from machinery are generally time series signals with strong correlations between nonadjacent signals. For the sake of capturing time invariant features behind dynamic machinery monitored signals, the reconstructed phase space based on dynamical systems and chaos theory (Kliková and Raidl 2011) is employed for signal representation. The most commonly used method of reconstructed phase space is the method of time delay (Povinelli et al. 2004), in which the time series signal x_n can be constructed as follows:

$$X_n = [x_{n-(d-1)\tau}, \dots, x_{n-\tau}, x_n] \tag{3}$$

where X_n denotes the reconstructed vector in the phase space, $n = (I + (d - 1)\tau) \dots N$, τ represents the time delay. With the properly selected time delay steps and embedding dimension in phase space, valuable information of correlated machinery signals can be represented. The investigated signal representation techniques are summarized in Table 1.

3.2 DCNN Model Construction

DCNN is a powerful multi-layered neural network that leverages three architectural ideas including sparse connectivity, shared weights, and pooling (Lee and Kim 2017;

Table 1 Summary of signal representation techniques

Signal representation	Pros.	Cons.
Raw time domain	Strong periodicity	Irrelevant noisy and redundant information
Frequency domain	Important characteristics of frequency components	Only suitable for steady state analysis and time domain information is lost
Time–frequency domain	Capture of high-resolution information in time–frequency domain	Affected by wavelet base and wavelet scale
Reconstructed phase space	Capture deep correlations between nonadjacent signals	Hard to select the optimal parameters

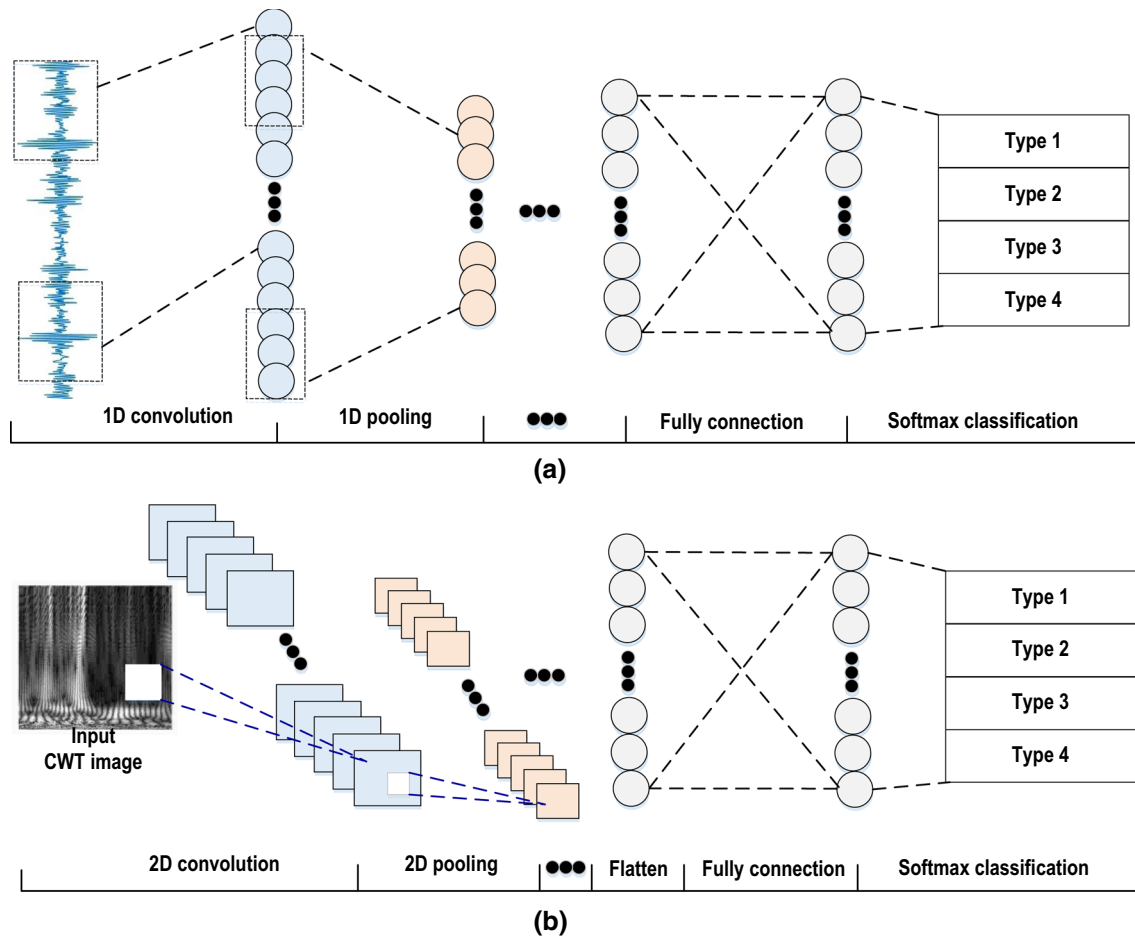


Fig. 3 Comparison of different model structures including **a** 1D DCNN and **b** 2D DCNN

Ko and Sim 2018). As shown in Fig. 3, a typical DCNN model consists of convolutional layers, pooling layers and fully connected layers. The convolutional layers with different convolutional kernels are used to extract features of input. Each convolutional kernel is specific to a feature map. After that, the pooling layers are used to reduce the size of extracted features and model parameters of the network. Finally, the fully connected layers perform as the classifier.

In one-dimensional CNN, one-dimensional counterparts (*conv1D*) are adopted to perform a convolutional operation. The output of the convolutional layer can be defined as:

$$x_{1D,k} = f \left(\sum_{i \in M} conv1D(x_{1D,i}, w_{1D}) + b \right) \tag{4}$$

where M represents the input feature map, and the “valid” is the type of convolutional operation. $x_{1D,i}$ is the one-dimensional input. w_{1D} denotes the convolutional kernel.

Each output feature map is given an additive bias b . In the one-dimensional DCNN, one-dimensional max pooling is used in the structure, which can be described as follows:

$$x_k^{l+1} = 1D \max(x_k^l) \quad (5)$$

The pooling layer takes the maximum value of distinct pooling regions with the width n , thus the pooling output is n -times shorter along the only one spatial dimension.

Similarly, the two-dimensional convolution (*conv2D*) is adopted in the two-dimensional CNN (Bouvier 2006). The output of the two-dimensional convolutional layer can be defined as:

$$x_{2D,k} = f \left(\sum_{i \in M} \text{conv2D}(x_{2D,i}, w_{2D}) + b \right) \quad (6)$$

The two-dimensional max pooling can be computed as:

$$x_k^{l+1} = 2D \max(x_k^l) \quad (7)$$

In the two-dimensional pooling layer, the input features are sub-sampled by distinct n -by- n blocks with a suitable factor n . Hence, the pooling output is n -times smaller along both spatial dimensions.

After extracting features through convolutional layers and pooling layers, several fully connected layers are employed as the final classifier. Each fully-connected layer is followed by a *ReLU* activation function. The softmax function is used for multiclass classification both in one-dimensional and two-dimensional DCNN models. The softmax activation of the k^{th} output unit is given as below:

$$o_k^{(n)} = \frac{\exp(z_k^{(n)})}{\sum_i \exp(z_i^{(n)})}, \quad (8)$$

$$x_k^l = \text{ReLU}(z_k^l), \text{ with } z_k^l = \sum_i w_{ik}^l x_k^{l-1} + b^l$$

where z_k is the layer $l-1$ input to unit k of the layer l .

3.3 Performance Evaluation Metrics

The performances of DCNN models are mainly evaluated in terms of classification accuracy and computational complexity. Two definitions including the recall ratio r and the precision ratio p should be considered before introducing the metric of classification accuracy. The recall ratio r and precision ratio p are defined as:

$$r = \frac{TP}{TP + FN} \quad (9)$$

$$p = \frac{TP}{TP + FP} \quad (10)$$

where true positive (TP) is correctly classified as positives samples, false positive (FP) is misclassified as positives

samples, true negative (TN) is correctly classified as negatives samples, and false negative (FN) is misclassified as negatives samples. Based on these two definitions, the classification accuracy of each class can be calculated as:

$$\text{accuracy} = \frac{FP + FN}{TP + FP + TN + FN} \times 100\% \quad (11)$$

In addition, the metric of computational complexity consists of the number of model parameters and total model training time. The total model parameters N_{total} is the sum of numbers of parameters in all convolutional layers and fully connected layers. It can be obtained by:

$$N_{total} = \sum_{i=1}^m (n_{oc} \times (n_{ic} \times n_w + 1)) + \sum_{j=1}^n (n_{os} \times (n_{is} + 1)) \quad (12)$$

where n_{oc} and n_{ic} denotes the number of output channels and input channels in convolutional layer respectively. n_w represents the convolutional kernel width. n_{os} and n_{is} represents the size of fully connected layer. m and n denote the total numbers of convolutional layers and fully connected layers respectively. The total model training time T is defined as follows:

$$T = \sum_{i=1}^N t_i \quad (13)$$

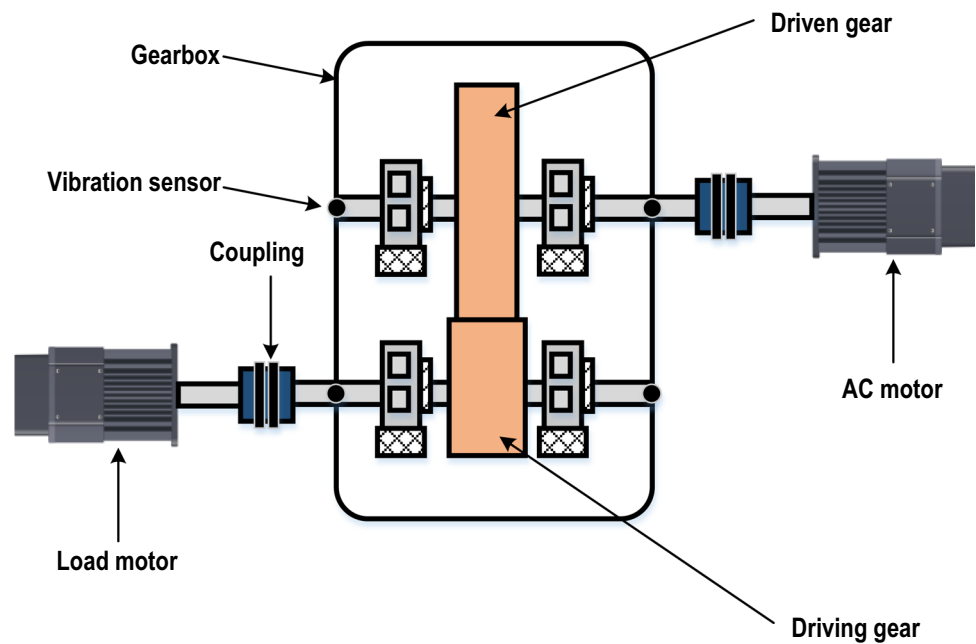
where t_i is the training time of each epoch, N is the total number of training epochs, and T is the model training time.

4 Experimental Study

In this section, experimental studies on a gearbox fault simulator are conducted to evaluate the performance of developed DCNN approach.

4.1 Experimental Setup

The gearbox fault simulator consists of four components including a speed controller, an altering current (AC) servo motor, a load motor and a gearbox. The experimental setup of the gearbox fault simulator is shown in Fig. 4. Three different levels of gear fault severities are created by adding different sizes of cracks to a gear tooth (Sun et al. 2016). In the first two cases, slight and medium cracks are added to the root of the gear tooth, while the third case is simulated with a broken tooth. Four vibration sensors are placed in the housings of two ends of input and output shafts to acquire the vibration signal of the gearbox. The vibration signals are collected at the sampling rate of

Fig. 4 Experimental setup of gearbox fault simulator

8192 Hz, and a total of 1920 data sets are generated while each data set contains 2048 points.

4.2 Data Processing

Different signal representation techniques including FFT, continuous wavelet transformation and reconstructed phase space are employed to transform time series signals into different domains. An illustration of a vibration signal under these representation techniques is shown in Fig. 5. The frequency domain signal is generated by transforming the time series signal into its spectrum by the FFT method. Thus, there are 1920 samples and each transformed sample contains 1024 points in frequency domain input. Both the raw time series signal and frequency domain signal are used as input for one-dimensional DCNN. As for the generation of inputs for two-dimensional DCNN, the Morlet wavelet is selected for base wavelet and the wavelet scale is set as 96. Hence, a total of 1920 time–frequency domain signals with the size of 96×96 are generated through continuous wavelet transformation. Before creating the reconstruction phase space signal for the two-dimensional DCNN model, the optimal value of the phase space parameters is estimated. The time lag τ is optimized as 40 and the embedding dimension d is selected as 12 (Kliková and Raidl 2011). Therefore, the reconstructed time domain input contains 1920 samples in total and the dimension of each sample is 1608×12 .

4.3 Model Structure

The DCNN models are customized considering the dimension of the models, the depth of layers, and the convolutional kernel functions. The convolutional kernel width and the number of convolutional feature maps are optimized to enhance model performance. Details about these DCNN models are illustrated in Fig. 6. In order to enhance the model performance of one-dimensional DCNN with raw time domain input, the convolutional kernel widths are set as 257×1 and 127×1 , and feature maps are selected as 24 and 48 respectively. Considering the sparsity of frequency domain input, the convolutional kernel widths are set as 3×1 and 4×1 , and feature maps in the one-dimensional DCNN model are selected as 12 and 24, respectively. To effectively extract features from the time–frequency domain input, the two-dimensional DCNN is composed of two groups of stacked convolutional layers. The convolutional kernel width of each convolutional layer is 3×3 and the feature maps of each convolutional group are 32 and 64, respectively. Since valuable information may be magnified in the reconstructed phase space signal, two-dimensional DCNN is designed with wide convolutional kernel width and large feature maps, in which the convolutional kernel widths are set as 127×3 and 66×2 , and feature maps are selected as 32 and 64 respectively.

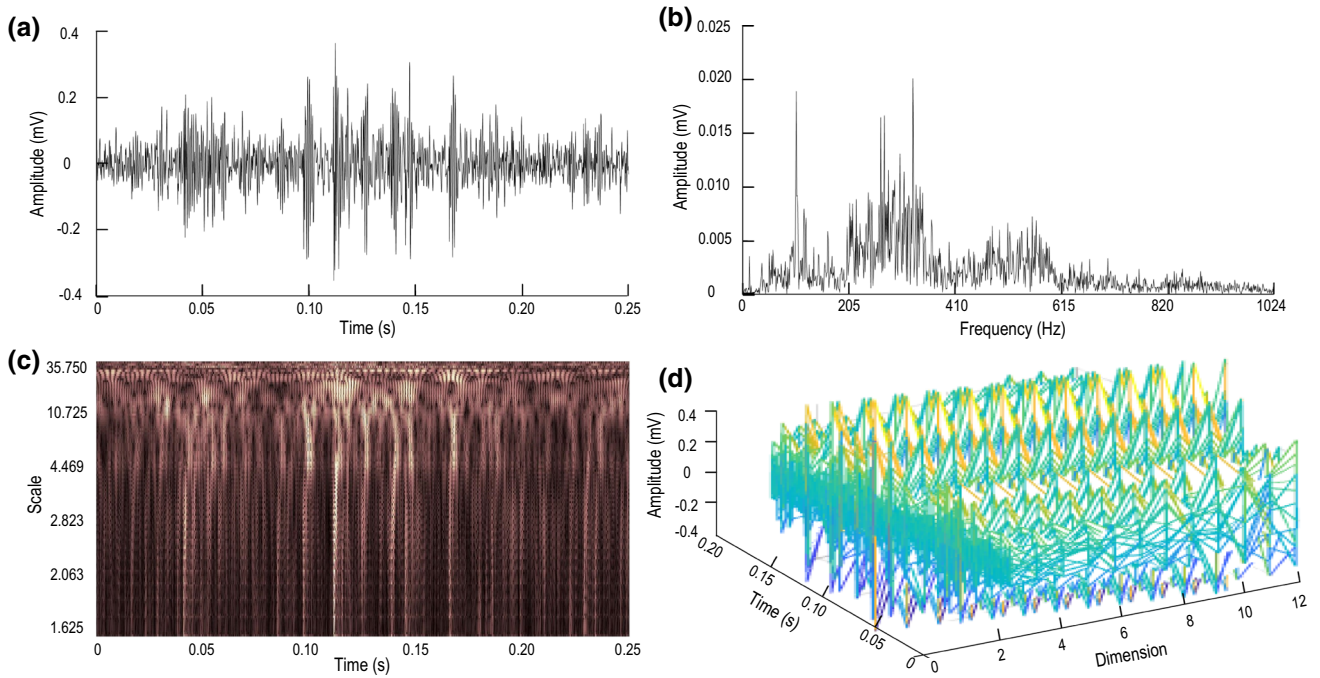


Fig. 5 Illustration of different signal representations from: **a** raw time domain, **b** frequency domain, **c** time–frequency domain, and **d** reconstructed phase space

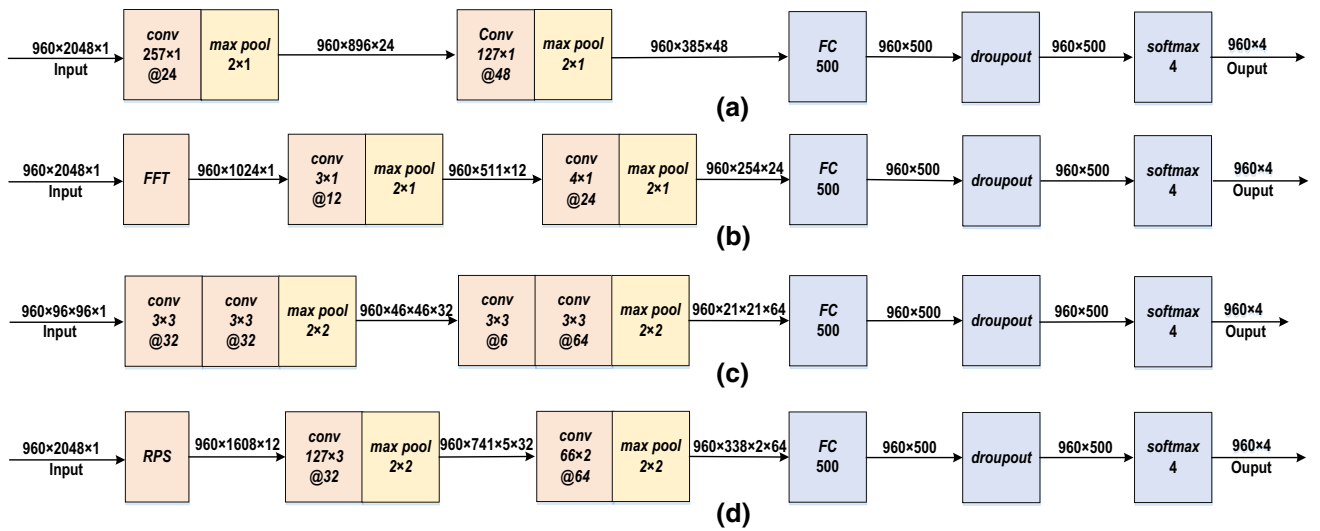


Fig. 6 Illustration of DCNN model structure: **a** one-dimensional DCNN with time domain signal, **b** one-dimensional DCNN with frequency domain signal, **c** two-dimensional DCNN with time–

frequency domain signal, and **d** two-dimensional DCNN with reconstructed phase space signal

4.4 Result Analysis

To confirm the effectiveness of presented DCNN models, two different neural networks including a Multi-layers Perceptron (MLP) and a Stacked Denoising Auto-encoder (SDA) (Leng and Jiang 2016) are used for comparison. Thus, a total of six different models are investigated and their detailed implementation are discussed below.

1. MLP: The MLP model consists of two fully connected layers for feature extraction and the size of each layer is set as 500.
2. SDA: The SDA model is composed of two stacked Auto-encoders, with each Auto-encoder containing an encoder layer and a decoder layer. The sizes of the first encoder layer and the second encoder layer are 500 and 100, respectively. These two Auto-encoders are trained

Table 2 Performance comparison of different models

Model	Signal representation	Accuracy (%)	Training time (s)	Total parameters
MLP	Raw time domain	75.14	11.98	1,277,004
	Frequency domain	64.89	11.29	765,004
	Time–frequency domain	49.28	89.26	4,861,004
	Reconstructed phase space	73.33	51.94	9,901,004
SDA	Raw time domain	79.87	41.08	1,075,004
	Frequency domain	99.79	38.88	563,004
	Time–frequency domain	65.31	68.45	4,659,004
	Reconstructed phase space	77.19	78.56	20,302,004
1D DCNN	Raw time domain	88.75	37.72	2,323,208
	Frequency domain	99.86	18.51	366,688
2D DCNN	Time–frequency domain	93.54	95.19	14,179,496
	Reconstructed phase space	98.43	216.03	21,917,128

with corrupted input and the learned parameters are used to fine-tune (Leng et al. 2018) the whole SDA model.

- One-dimensional DCNN with time series signal input: This one-dimensional DCNN model consists of two consecutive feature extraction couples, and each couple contains a convolutional layer and a max pooling layer. This model is designed with greater convolutional kernel width.
- One-dimensional DCNN with frequency domain input: This one-dimensional DCNN model consists of two consecutive feature extraction couples, and each couple contains a convolutional layer and a max pooling layer. This model is designed with convolutional kernel width.
- Two-dimensional DCNN with time–frequency domain input: The two-dimensional DCNN model containing two groups of stacked convolutional layers is employed to extract features from high-dimensional time–frequency domain input.
- Two-dimensional DCNN with reconstructed time domain input: The two-dimensional DCNN is designed with greater convolutional kernel width and large feature maps to effectively extract features from reconstructed phase space signals.

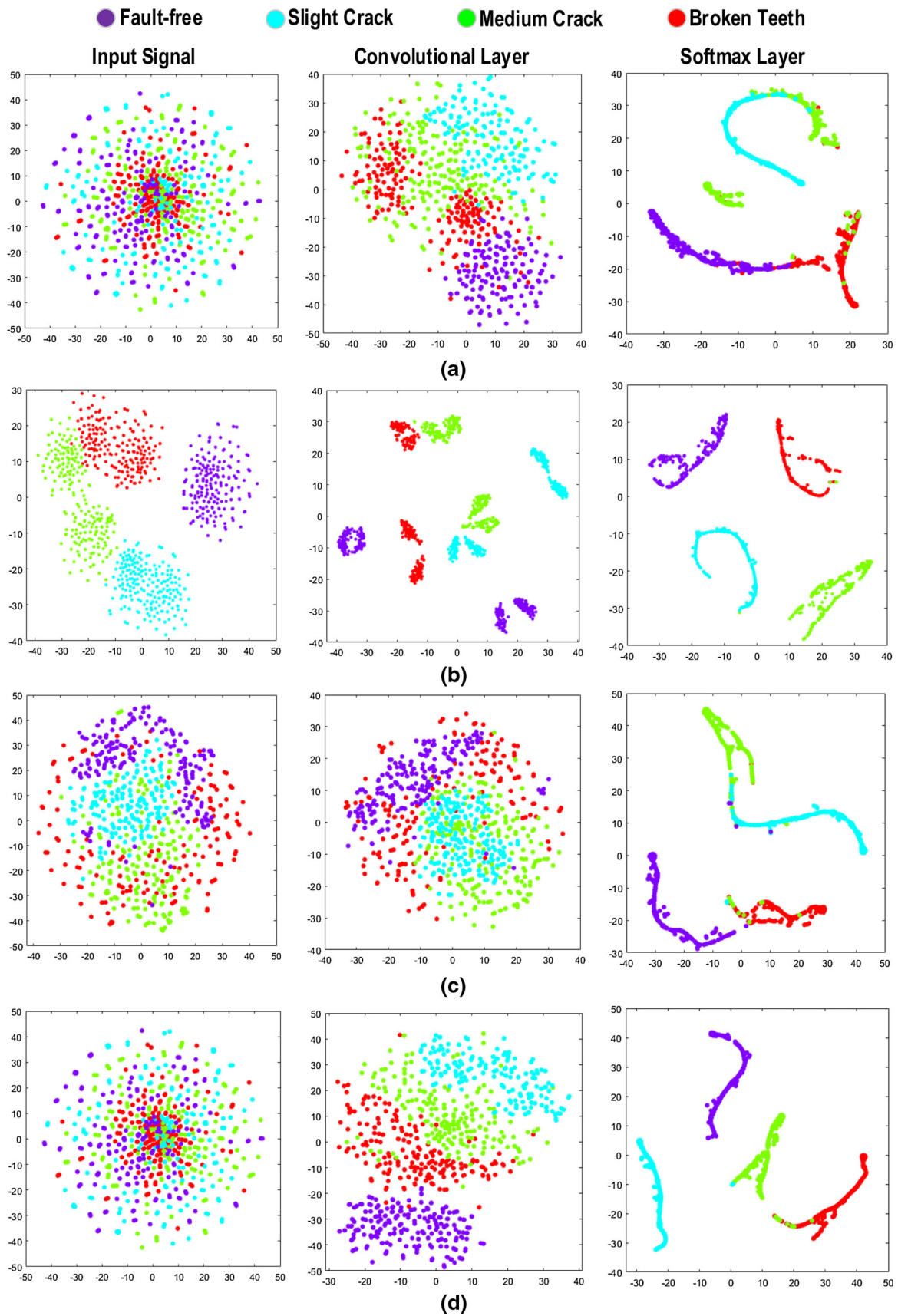
These models are trained using the same training strategy. Half the samples of each fault severity are used as training data, and the rest are set as testing data. A mini-batch stochastic gradient descent algorithm is used for model optimization. *Dropout* is a dropout layer added between the fully connected layer and the softmax layer to prevent the network from overfitting. The *ReLU* function is selected as the activation function and a softmax layer is used for fault classification. The training epoch is set as 50, and a batch size of 40 samples is used during training

optimization. The TensorFlow neural network toolbox is used for developing these models.

The classification results of these models are shown in Table 2. It is found that the classification accuracies of both the MLP and the SDA are lower than those of the DCNN models under these signal representations. The SDA model performs better than the MLP model in terms of classification accuracy, which indicates that the SDA model is more suitable to extract features than the MLP model. As for the DCNN models, classification accuracy of the one-dimensional DCNN model with frequency domain input achieves 99.86% and requires lower computational complexity, which is the best performing DCNN model. The two-dimensional DCNN with reconstructed time domain signal outperforms the two-dimensional DCNN with time–frequency domain input in terms of classification accuracy. However, the reconstructed time domain input requires more model parameters, resulting in increased model training time. In addition, it is worth noting that the classification accuracy of the SDA with frequency domain input reaches 99.79%, which is only 0.07% lower than that of the one-dimensional DCNN with the same input. However, model training time of the SDA is still higher than that of the 1D DCNN and more model parameters are generated.

5 Discussion

In this section, the effects of signal representation and model construction on DCNN model performance are discussed in detail. Then, an intuitive understanding of the learned features of the DCNN models are mapped to a low-dimensional space via t-SNE (van der Maaten and Hinton 2008). Finally, the model convergence of these DCNN models are analyzed.



◀ **Fig. 7** Feature visualization of the DCNN models via t-SNE: **a** one-dimensional DCNN with input of time domain signals, **b** one-dimensional DCNN with input of frequency domain signals, **c** two-dimensional DCNN with input of wavelet domain signals, and **d** two-dimensional DCNN with input of reconstructed phase space signals

5.1 Effect of Signal Representation

As shown in Fig. 7, to reach an intuitive understanding of the effectiveness of signal representation techniques for these DCNN models, learned features of input signals, convolutional layers and softmax layers are mapped to a low-dimensional space via t-SNE for visualization. The

Fault-free	97.41%	2.05%	0	0.54%
Slight Crack	4.71%	89.94%	4.41%	0.94%
Medium Crack	1.15%	3.03%	86.13%	9.69%
Broken Teeth	0	0.26%	4.29%	95.45
	Fault-free	Slight Crack	Medium Crack	Broken Teeth

(a)

Fault-free	100%	0	0	0
Slight Crack	0	100%	0	0
Medium Crack	0	0.78%	98.02%	1.20%
Broken Teeth	0	0	0	100%
	Fault-free	Slight Crack	Medium Crack	Broken Teeth

(b)

Fault-free	94.36%	1.31%	0.84%	3.49%
Slight Crack	0	94.79%	4.78%	0.43%
Medium Crack	0.91%	11.81%	81.83%	5.45%
Broken Teeth	2.86%	0	4.29%	92.85%
	Fault-free	Slight Crack	Medium Crack	Broken Teeth

(c)

Fault-free	100%	0	0	0
Slight Crack	0	99.16%	0.84%	0
Medium Crack	0	0	96.66%	3.34%
Broken Teeth	0	0	1.58%	98.42%
	Fault-free	Slight Crack	Medium Crack	Broken Teeth

(d)

Fig. 8 Classification results of the DCNN models: **a** one-dimensional DCNN with input of time domain signals, **b** one-dimensional DCNN with input of frequency domain signals, **c** two-dimensional DCNN with input of wavelet domain signals, and **d** two-dimensional DCNN with input of reconstructed phase space signals

with input of wavelet domain signals, and **d** two-dimensional DCNN with input of reconstructed phase space signals

graph shows that there are obvious differences in feature distribution between different signal representation inputs. Features of raw time series signal are intermingled. While the time domain signal is transformed into a frequency domain, features begin to cluster roughly. After transforming the time series signal into different domains, features of the three other inputs in the final softmax layer are separated well, fewer testing samples are clustered to wrong samples. These visualizations results confirm the

effectiveness of signal representation techniques on DCNN model performance on machinery condition monitoring.

Moreover, the classification results which use testing inputs of the DCNN models are shown in Fig. 8. It can be observed that the feature extraction capabilities between different DCNN models are different. The classification accuracy DCNN model with time domain input reaches 88.75%, which is lower than that of the other DCNN models. The reason may be that raw vibration signals contain noisy components, thus the feature extraction capability of one-dimensional DCNNs is restricted. In comparison, noises are suppressed in frequency domain and frequency domain representation becomes sparse (Deng et al. 2013). Thus, this signal representation can yield clearer features and the best performing one-dimensional DCNN achieves remarkable classification accuracy with low computational complexity. Moreover, as shown in Fig. 8, the two-dimensional DCNN with input of reconstructed time domain signals achieves 98.43% classification accuracy, which is 4.89% higher than that of two-

Table 3 Results of two-dimensional DCNN with time–frequency domain input

Signal representation	Wavelet scale	Test accuracy (%)	Training time (s)
Time–frequency domain	64	90.13	86.49
	96	93.54	95.19
	128	89.72	94.04

Table 4 Classification results of one-dimensional DCNN under different convolutional kernel widths and feature maps

Input	Index	Conv kernel width	Feature maps	Conv kernel width	Feature maps	Accuracy (%)	Training time (s)
Time domain	1	127	12	67	24	79.86	27.97
	2	257	12	35	24	81.98	25.39
	3	257	6	127	12	74.81	22.37
	5	257	12	127	24	83.54	25.41
	6	257	24	127	48	88.75	37.72
	7	257	32	127	64	74.11	44.69
	8	3	24	3	48	62.13	29.51
	Frequency domain	9	3	6	4	12	99.13
10		3	12	4	24	99.86	18.51
11		3	24	4	48	99.79	24.11
12		3	32	4	64	99.75	24.55
13		15	12	8	24	99.72	20.42
14		127	12	66	24	99.53	20.72
15		257	12	127	24	99.03	20.89

Table 5 Classification results of two-dimensional DCNN with reconstructed phase space signal under different convolutional kernel widths and feature maps

Conv kernel width	Feature maps	Conv kernel width	Feature maps	Accuracy (%)	Training time (s)	Total parameters
127 × 3	6	66 × 2	12	95.69	71.36	4,070,312
127 × 3	12	66 × 2	24	97.37	93.97	8,157,128
127 × 3	24	66 × 2	48	97.53	167.49	16,387,784
127 × 3	32	66 × 2	64	98.43	216.23	21,917,128
127 × 3	40	66 × 2	82	97.29	351.62	28,166,826
257 × 3	32	127 × 2	64	97.39	309.22	18,147,464

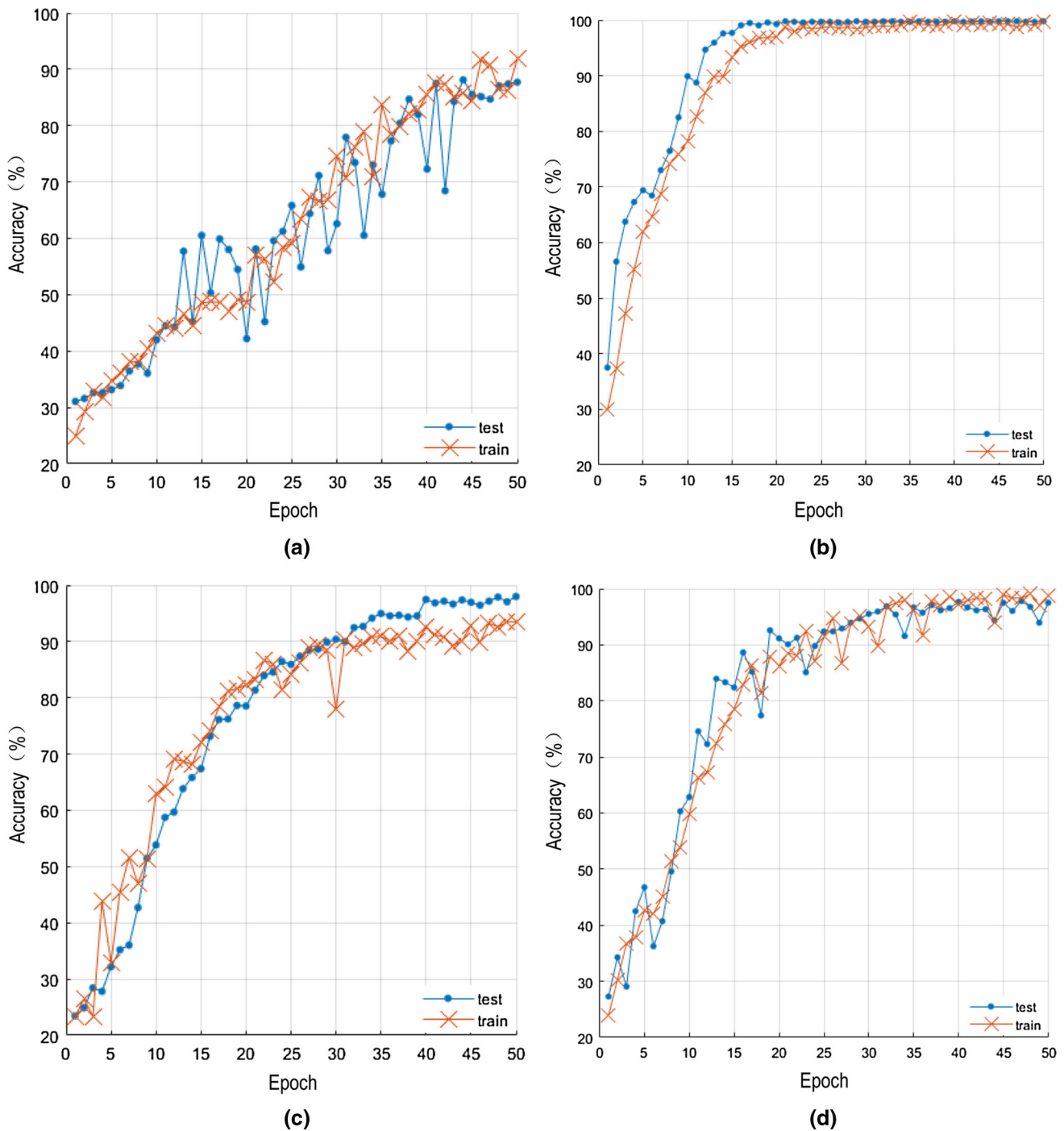


Fig. 9 Convergence of the DCNN models: **a** one-dimensional DCNN with time domain signals, **b** one-dimensional DCNN with frequency domain signals, **c** two-dimensional DCNN with time–frequency

dimensional DCNN with time–frequency domain input. The results indicate that this signal reconstruction technique is suitable for mining intrinsic correlations between machinery vibration signals.

As for the two-dimensional DCNN with time–frequency domain input, the wavelet scale influences the model classification performance. The more stretched the wavelet,

domain signals, and **d** two-dimensional DCNN with reconstructed phase space signals

the coarser the signal features measured by the wavelet coefficients. Thus, the effect of the wavelet scale is investigated in Table 3. It is observed that the classification accuracies are 86.49% and 90.14% under the wavelet scale of 64 and 128, respectively. Both of these classification accuracies are lower than when the wavelet scale is 96.

5.2 Comparison of Different Model Structures

To demonstrate the performance of the DCNN models, the effects of convolutional kernel width and number of feature maps on the classification results are investigated in Table 4. It can be seen that the one-dimensional DCNN model with time domain input can achieve an average classification accuracy of 80.51% when the convolutional kernel is large. However, a small convolutional kernel size 3×1 can only achieve a 62.13% classification accuracy. This may be explained by the fact that the large convolutional kernel in the convolutional layer is able to filter high frequency noise, while one-dimensional DCNN with frequency domain input is robust for convolutional kernel size.

Generally, classification accuracy can be rectified by increasing the number of convolutional feature maps. However, as shown in Table 4, classification accuracy is not always improved as the number of convolutional feature maps increase. The reason may lie in that redundant and useless information is extracted as features and causes the drop of classification accuracy. Additionally, larger numbers of feature maps also result in an increasing computational complexity. These results demonstrate the importance of the right selection of convolutional kernel width and feature maps so as to ensure the DCNN model performance.

As for the model structure of two-dimensional CNN with reconstructed time domain input, the effects of the convolutional kernel width and the number of feature maps are also investigated in Table 5. Since the reconstructed time domain input contains more valuable information than other signal representations, large number of feature maps are required to extract these features. As shown in Table 5, when the numbers of convolutional feature maps are set as 32 and 64 respectively, a maximum classification accuracy of 98.43% can be reached.

5.3 Model Convergence

To evaluate the convergence of different models, a computer is utilized with the setup of intel Xeon CPU (1.70 GHz), 16 GB DDR3 RAM, and GeForce GTX 1060 6 GB. The model convergence results are shown in Fig. 9. It can be seen that the classification accuracy of the one-dimensional DCNN with time domain input is hard to improve and that large fluctuations exist. However, two-dimensional DCNN models are almost saturated after 40 epochs with slight fluctuation. Compared to the one-dimensional DCNN, two-dimensional DCNN models have a large number of model parameters. Therefore, lower computational efficiency is generated. In comparison with the other three DCNN models, one-dimensional DCNN

with frequency domain input shows faster convergence and is almost saturated after 20 epochs, which indicates that this model is stable and possesses high computational efficiency.

6 Conclusions

This paper presents a comprehensive condition monitoring framework based on deep convolutional neural network (DCNN) which aims at improving model performance. The raw vibration signals are transformed by means of different signal representation techniques before being fed into DCNN models to classify health conditions. Experimental studies on a gearbox test rig are utilized to evaluate the effectiveness of the presented DCNN models, and major conclusions can be drawn as follows:

1. Classification performance of DCNN models for the classification of gearbox health conditions is investigated. The results indicate that classification accuracy of the one-dimensional DCNN fed with frequency domain signal outperforms the other DCNN models and shows lower computational complexity.
2. The effect of different signal representation techniques on classification performance of DCNN models is investigated. The results suggest that the classification accuracy of the one-dimensional DCNN with frequency domain signals outperforms that of raw time domain signals by over 11.11%. At the same time the classification accuracy of two-dimensional DCNN with reconstructed time domain signals is 4.89% higher than that of two-dimensional DCNN with time–frequency domain input.
3. The structure of the DCNN models has been investigated for selecting the optimal convolutional kernel width and the number of feature maps. Both one-dimensional and two-dimensional DCNN model structures with suitable convolutional kernel width and feature maps can achieve better classification performance.

There are still some challenges which need to be further investigated. First, the varying operating conditions may cause a variety of signal measurements which may deteriorate the performance of the presented DCNN model. On the other hand, the computational burden of DCNN models will increase with the size of input data. The new DCNN structures such as VGG-net (Simonyan and Zisserman 2014), Res-net (Wu et al. 2017) and inception-v4 (Szegedy et al. 2017) have been proposed. With the development of deep learning techniques, the effectiveness of the new variants of CNN models will be studied for machinery

condition monitoring under different operating conditions in our future research.

Acknowledgements This research acknowledges the financial support partially provided by Natural Science Foundation of China (No. U1862104), National Key Research and Development Program of China (No. 2016YFC0802103), and the Fundamental Research Funds for the Central Universities (No. ZX20180008). The constructive comments from the anonymous reviewers are greatly appreciated to help improve the paper.

References

- Abdeljaber O, Avci O, Kiranyaz S, Gabbouj M, Inman DJ (2017) Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J Sound Vib* 388:154–170
- Bouvier J (2006) Notes on convolutional neural networks. Unpublished, http://cogprints.org/5869/1/cnn_tutorial.pdf. Accessed 26 Feb 2019
- Deng L, Hinton G, Kingsbury B (2013) New types of deep neural network learning for speech recognition and related applications: an overview. In: *IEEE international conference on acoustic speech signal process*, pp 8599–8603. <https://doi.org/10.1109/icassp.2013.6639344>
- Ding X, He Q (2017) Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. *IEEE Trans Instrum Meas* 66(8):1926–1935
- Fu Y, Zhang Y, Gao Y, Gao H, Mao T, Zhou H, Li D (2017) Machining vibration states monitoring based on image representation using convolutional neural networks. *Eng Appl Artif Intell* 65:240–251
- Gangsar P, Tiwari R (2017) Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms. *Mech Syst Signal Process* 94:464–481
- Guo X, Chen L, Shen C (2016) Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* 93:490–502
- Heinrich P, Schwabe G (2018) Facilitating informed decision-making in financial service encounters. *Bus Inf Syst Eng* 60(4):317–329
- Ince T, Kiranyaz S, Eren L, Askar M, Gabbouj M (2016) Real-time motor fault detection by one-dimensional convolutional neural networks. *IEEE Trans Ind Electron* 63:7067–7075
- Jing L, Zhao M, Li P, Xu X (2017) A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* 111:1–10
- Kliková B, Raidl A (2011) Reconstruction of phase space of dynamical systems using method of time delay. In: *Proceedings of 20th annual conference Dr students—WDS 2011*, pp 83–87
- Ko KE, Sim KB (2018) Deep convolutional framework for abnormal behavior detection in a smart surveillance system. *Eng Appl Artif Intell* 67:226–234
- Konar P, Chattopadhyay P (2011) Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Appl Soft Comput J* 11(6):4203–4211
- LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2323
- LeCun Y, Bengio Y, Hinton G (2005) Deep learning. *Nature* 521:436–444
- Lee SJ, Kim SW (2017) Localization of the slab information in factory scenes using deep convolutional neural networks. *Expert Syst Appl* 77:34–43
- Leng J, Jiang P (2016) A deep learning approach for relationship extraction from interaction context in social manufacturing paradigm. *Knowl Based Syst* 100:188–199
- Leng J, Chen Q, Mao N, Jinag P (2018) Combining granular computing technique with deep learning for service planning under social manufacturing contexts. *Knowl Based Syst* 143:295–306
- Liu R, Meng G, Yang B, Sun C, Chen X (2016) Dislocated time series convolutional neural architecture: an intelligent fault diagnosis approach for electric machine. *IEEE Trans Ind Inform* 13(3):1310–1320
- Lu Y, Xu X, Xu J (2014) Development of a hybrid manufacturing cloud. *J Manuf Syst* 33:551–566
- Matt C (2018) Fog computing. *Bus Inf Syst Eng* 60(4):351–355
- Nalchigar S, Eric Y (2018) Designing business analytics solutions—a model-driven approach. *Bus Inf Syst Eng*. <https://doi.org/10.1007/s12599-018-0555-z>
- Povinelli RJ, Johnson MT, Lindgren AC, Ye J (2004) Time series classification using Gaussian mixture models of reconstructed phase spaces. *IEEE Trans Knowl Data Eng* 16:779–783
- Ren L, Zhang L, Wang L, Tao F, Chai X (2017) Cloud manufacturing: key characteristics and applications. *Int J Comput Integr Manuf* 30:501–515
- Residual F, Using V, Neural A (2018) Decision support for the automotive industry. *Bus Inf Syst Eng*. <https://doi.org/10.1007/s12599-018-0527-3>
- Richter A, Heinrich P, Stocker A, Schwabe G (2018) Digital work design. *Bus Inf Syst Eng* 60(3):259–264
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. In: *Proceedings of international conference on learning Representations*, pp 1–14. <https://arxiv.org/abs/1409.1556>. Accessed 14 Mar 2019
- Su H, Li X, Yang B, Wen Z (2018) Wavelet support vector machine-based prediction model of dam deformation. *Mech Syst Signal Process* 110:412–427
- Sun C, Wang P, Yan R, Gao RX (2016) A sparse approach to fault severity classification for gearbox monitoring. In: *Proceedings of the 19th international conference on information fusion. Heidelberg, IEEE*, pp 2303–2308
- Szegedy C, Ioffe S, Vanhoucke V, Alemi A (2017) Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: *31st AAAI conference on artificial intelligence, San Francisco*, pp 4278–4284
- Tao F, Cheng J, Qi Q (2018a) IIHub: an industrial Internet-of-Things hub toward smart manufacturing based on cyber-physical system. *IEEE Trans Ind Inform* 14(5):2271–2280
- Tao F, Qi Q, Liu A, Kusiak A (2018b) Data-driven smart manufacturing. *J Manuf Syst* 48:157–169. <https://doi.org/10.1016/j.jmsy.2018.01.006>
- van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 1(620):267–284
- Verstraete D, Engineering M, Engineering M (2017) Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Hindawi Shock Vib* 2017:1–29
- Wang J (2016) A multi-scale convolution neural network for featureless fault diagnosis. In: *Proceedings of the international symposium on flexible automation. IEEE, Cleveland*, pp 65–70
- Wang L, Torngren M, Onori M (2015) Current status and advancement of cyber-physical systems in manufacturing. *J Manuf Syst* 37:517–527
- Wang P, Ananya Yan R, Gao RX (2017) Virtualization and deep recognition for system fault classification. *J Manuf Syst* 44:310–316
- Wang J, Ma Y, Zhang L, Gao RX, Wu D (2018) Deep learning for smart manufacturing: methods and applications. *J Manuf Syst* 48:144–156. <https://doi.org/10.1016/j.jmsy.2018.01.003>

- Wu S, Zhong S, Liu Y (2017) Deep residual learning for image steganalysis. *Multimed Tools Appl* 77(9):10437–10453
- Yan R, Gao RX, Chen X (2014) Wavelets for fault diagnosis of rotary machines: a review with applications. *Sig Process* 96:1–15
- Ye Y, Hu T, Zhang C, Luo W (2016) Design and development of a CNC machining process knowledge base using cloud technology. *Int J Adv Manuf Technol* 94:3413–3425
- Ye Y, Hu T, Yang Y, Zhu W, Zhang C (2018) A knowledge based intelligent process planning method for controller of computer numerical control machine tools. *J Intell Manuf* 2018:1–17. <https://doi.org/10.1007/s10845-018-1401-3>
- Zhang W, Peng G, Li C (2017) Rolling element bearings fault intelligent diagnosis based on convolutional neural networks using raw sensing signal. In: *Proceeding of the twelfth international conference on intelligent information hiding and multimedia signal processing*, Springer, Cham, pp 77–84
- Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX (2019) Deep learning and its applications to machine health monitoring. *Mech Syst Signal Process* 100:439–453