Measurement 146 (2019) 268–278

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes

Huipeng Chen [a,b], Niaoqing Hu [a,b,*], Zhe Cheng [a,b,*], Lun Zhang [a,b], Yu Zhang [a,b]

[a] College of Mechatronics and Automation, National University of Defense Technology, Changsha 410072, China
[b] Laboratory of Science and Technology on Integrated Logistics Support, NUDT, Changsha 410072, China

ARTICLE INFO

Article history:
Received 16 November 2018
Received in revised form 13 April 2019
Accepted 19 April 2019
Available online 16 May 2019

Keywords:
Planetary gearbox
Data fusion
Deep convolutional neural network
Feature extraction
Health state identification

ABSTRACT

With the great ability of transforming data into deep and abstract features adaptively through nonlinear mapping, deep learning is a promising tool to improve the intelligence and accuracy of diagnosis. On the other hand, one acceleration sensor is not sensitive enough to position-variable faults and the collected signal is usually nonstationary and noisy. As different measurement locations provide complementary information to the faults, the paper proposes a deep convolutional neural network (DCNN) based data fusion method for health state identification. This method fuses the raw data from the horizontal and the vertical vibration signals and extracts features automatically. The effectiveness of the novel method is validated through the data collected from a planetary gearbox test rig, and experiments using DCNN, SVM and BPNN based model in different data processing methods are also carried out. The results show that the proposed method could obtain better identification results than the other methods.

1. Introduction

The planetary gearbox is widely used for transmission in major machineries and equipment, such as wind turbines, helicopters, ships and so on, to change the speed and torque to match the power source and the load [1]. The significant advantages of planetary gearboxes include the light weight, the small size, the wide range of gear ratios, the strong bearing capacity, the high transmission efficiency, etc. [2]. In the key components of the planetary gearboxes, faults, such as severe pitting and fatigue crack etc., are frequently caused by low-speed and heavy-duty operation conditions [3]. If the failures and the damages of rotating parts of the equipment are ignored and proper maintenance work is not taken in time, the whole mechanical transmission system would probably be force to stop, causing economic losses and even catastrophic consequences. Therefore, it is essential to develop an efficient fault diagnosis method for planetary gearboxes, which is of great importance to maintaining the technical status of mechanical equipment and prolonging the service time.

Health state of the planetary gearbox can be identified by collecting several signals, including vibration, acoustic, temperature, driven motor current, speed, oil debris, etc. [4]. To improve the diagnosis accuracy, multi-signal has been employed at multiple levels which are data level, feature level and decision level. Among them, the data level fusion is mainly to integrate signals such as the vibration and temperature signals. The data level fusion usually needs various kinds of sensors and instruments in the step of signal acquisition [4,5], which leads to expensive monitoring cost and complicated manipulation. Feature level fusion method, such as the fault diagnosis method based on multi-sensors and kernel principal component analysis proposed in paper [6], requires complex signal analysis and weighted calculation. Jiang et al. [7] used SVM as a tool for feature level fusion, and eight gear vibration signals for fault diagnosis were investigated. However, this method needs to select and calculate time domain statistics, which has the shortcomings of poor real-time property. In the decision level, the intelligent approaches are often introduced into the fault diagnosis, for example, the expert systems, the decision tree, and SVM [7,8]. These methods all belong to the shallow learning method, and the learning ability is limited. In order to reduce costs, operational complexity and difficulties of information fusion, the type of sensor should be as small as possible. The vibration is sensitive to the evolution of the health state of the mechanical equipment, which has been widely used in condition monitoring. However, acceleration

* Corresponding authors at: College of Mechatronics and Automation, National University of Defense Technology, Changsha 410072, China.

E-mail addresses: chenhuipeng12@nudt.edu.cn (H. Chen), hnq@nudt.edu.cn (N. Hu), chengzhenudt@163.com (Z. Cheng).

https://doi.org/10.1016/j.measurement.2019.04.093
0263-2241/© 2019 Elsevier Ltd. All rights reserved.

sensor of a single direction is not always sensitive enough to faults which are position-variable. What is more, the fault characteristics are very weak and the useful information is limited. Therefore, it requires intricate signal processing and feature extraction, and the fault diagnosis accuracy is not stable. In order to obtain more information at each moment, and more complete description of the diagnostic object, it is necessary to set at least two sensors in both horizontal and vertical directions, and then use the fusion method to achieve more comprehensive and accurate diagnosis results. At the same time, as different mechanical equipment has different physical characteristics, the same feature is not possible for any system. Therefore, it is necessary and significant to develop a diagnosis method that can extract effective features adaptively and intelligently give the diagnosis results [9].

As an emerging method in the field of machine learning, deep learning has made great achievements in many fields including image identification, speech recognition, etc. [10–13] with its powerful capability of extracting features. Compared with the traditional methods, the application of deep learning in the field of fault diagnosis mainly has the following advantages: (1) Deep learning has the strong feature extraction ability to extract features from a large amount of data automatically, which reduces the dependence on the expert experience and the signal process technology, and thus the uncertainty due to manual participation could be reduced; (2) By establishing a deep model, the complex mapping relationship between monitoring data and fault conditions could be well characterized, which meets the requirements of the diagnosis with diverse, non-linear and high dimensional monitoring data. What is more, the deep neural network (DNN) has been utilized to fuse raw data and integrate features of signals in some studies [4,14]. Therefore, DNN should be an effective tool for fault diagnosis.

As a key model of DNN, deep convolutional neural network (DCNN) can extract the local features of input data and abstract high level features layer by layer, and finally obtains the feature representation of the input data with invariant translation, rotation and scaling. Due to its characteristics of local connection, weight sharing, spatially or temporally subsampling, DCNN can retain the spatial information of the input and the training parameters of the model are greatly reduced, and thus it is widely used in image recognition. DCNN has made great development that many architectures [15–17] are proposed since it was first proposed by LeCun [18] so far. Drawing on the successful application of DCNN, some scholars have already applied it for bearing fault diagnosis. Janssens et al. [19] applied CNN to learn features for fault detection of bearings, and the fault diagnosis without expert's experience is realized. Guo et al. [20] proposed a hierarchical adaptive deep CNN (ADCNN) with two combined CNNs to identify the health states and the degradation levels of rolling bearings. In addition, Wei Zhang et al. have done much research on the intelligent diagnosis methods based on DCNN with raw data, and some DCNN based models including WDCNN and TICNN, are proposed and successfully applied [21–23].

This paper focuses on the health state identification of a two-stage planetary gearbox and applies DCNN to fuse the horizontal and the vertical vibration signals to improve the representation information for health states. And the proposed method based on DCNN can extract features from raw data automatically with less dependence on the expert diagnosis experience and the signal processing techniques. The remainder of this paper is organized as follows. In Section 2, the methodology of DCNN is briefly introduced. In Section 3, the DCNN based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes is described. In Section 4, data from experiments are used to demonstrate the effectiveness and the superiority of the proposed method. After this, the obtained results with the discussion are given. And the conclusions are drawn in Section 5.

2. Deep convolutional neural network

2.1. Architecture of DCNN

DCNN is a typical feedforward neural network, which is to build a number of filters to extract features from input data. Through local receptive fields, shared weights and spatial subsampling, DCNN has the unique ability to maintain the initial information regardless of shift, scale and distortion invariance [13]. Fig. 1 shows a typical architecture of DCNN. The one-dimensional (1D) convolutional structure with a 1D filter bank as the kernel of the DCNN model is used in the paper, and the principle is the same to the typical architecture.

A typical DCNN model contains two kinds of parts, which are the filters and the classification [24]. The former part is constructed mainly by convolution and pooling alternately, and the classification part is built up with several fully-connected layers [25]. As shown in Fig. 2, the filter in this paper consists of four kinds of layers, which are the convolutional layer, the batch normalization layer, the activation layer and the pooling layer. In the following part, all kinds of layers will be described in detail.

The convolutional layer is one of the most important parts of DCNN. In the DCNN, convolution operation is performed on the feature map of the previous layer with each kernel of the convolutional layer repeatedly acting on the input local regions, and thus the features of the input are extracted. Note that each kernel shares the same parameters, including the same weight and bias [26]. In addition, one kernel corresponds to one feature map in the next layer, and the number of kernels is the depth of the convolutional layer. The process of the convolutional layer can be described as follows:

$$\mathbf{C}_k^{(m)} = \sum_{c=1}^{C} \boldsymbol{W}_k^{(c,m)} * \boldsymbol{X}_{k-1}^{(c)} + \boldsymbol{B}_k^{(m)} \tag{1}$$

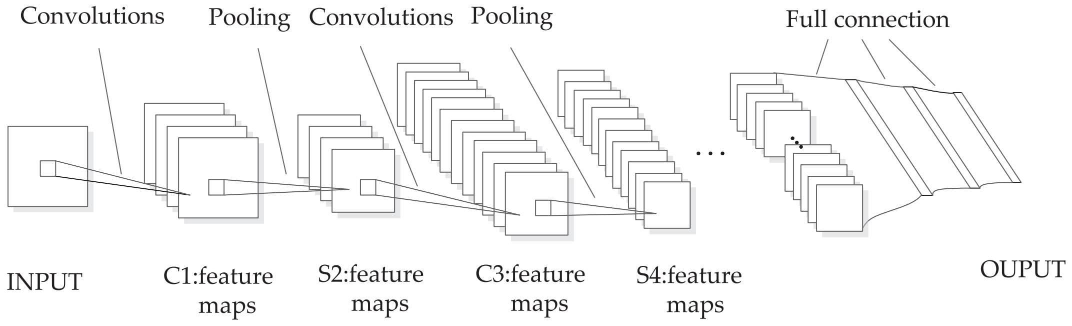

Fig. 1. A typical architecture of DCNN.

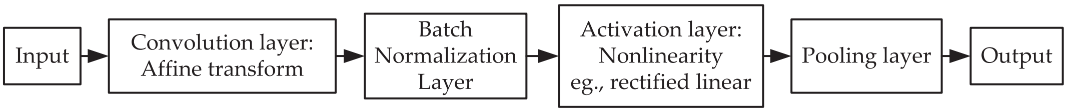
Input
Convolution layer: Affine transform
Batch Normalization Layer
Activation layer: Nonlinearity eg., rectified linear
Pooling layer
Output

Fig. 2. Structure of the filter.

where $\mathbf{C}_k^{(m)}$ is the m-th output of the convolutional layer, and $m = 1,\ldots, M$. M is equal to the kernel number; k denotes the k-th convolutional layer and the operator $*$ is used for the convolution of the input $\boldsymbol{X}_{k-1}^{(c)}$ and the kernel $\boldsymbol{W}_k^{(c,m)}$ of channel $c = 1,\ldots,C$. The matrix $\boldsymbol{B}_k^{(m)}$ is the bias weights.

The batch normalization (BN) [27] layer is used to improve the robustness and the learning efficiency of DCNN. The BN layer normalizes the output of the corresponding convolutional layer to prevent the phenomenon called "gradient diffusion". In DCNN, the BN layer should be added right before the activation layer. The transform of batch normalizing applied to activation $\mathbf{C}$ over a mini-batch is described as follows:

$$\begin{aligned} \widehat{c_i} &= \tfrac{c_i-\mu_\beta}{\sqrt{\sigma_\beta^2+\varepsilon}} \\ z_i &= \gamma\widehat{c_i} + \beta \equiv \mathrm{BN}_{\gamma,\beta}(c_i) \end{aligned} \tag{2}$$

where $\mathbf{C} = \{c_{1\ldots m}\}$ is the input of the BN layer over a mini-batch; $\mathbf{Z} = \{z_i = \mathrm{BN}_{\gamma,\beta}(c_i)\}$ is the output; $\mu_\beta = \frac{1}{m}\sum_{i=1}^{m} c_i$, $\sigma_\beta^2 = \frac{1}{m}\sum_{i=1}^{m}\left(c_i - \mu_\beta\right)^2$, and they are the mini-batch mean and the mini-batch variance respectively; γ, β are the parameters to be learned;

A nonlinear layer (or activation layer) is usually applied to introduce non-linear mapping to a network, which could improve the distinguishability of the learned features. In this paper, Rectified Linear Unit (ReLU) [28] is utilized as the activation function, which helps accelerating the convergence of the DCNN. The function f is applied to all values of the input, and all negative activation is transformed to zero, and the formula is:

$$q_i = f(z_i) = \max\{0, z_i\} \tag{3}$$

where $\mathbf{Z} = \{z_i\}$ is the output of the BN layer and q_i is the activation of z_i. In addition, the ReLU layer can also help to mitigate the gradient disappearance problem that as the gradient disappears exponentially in layers, the train at the bottom of network is very slow.

A pooling function is designed to replace the output of the net at a certain location with a summary statistic of the nearby outputs. The pooling layer is a scaled mapping of the previous layer to reduce the data dimension and the parameters of the network, which can both speed up calculation and prevent overfitting. The pooling layer using the maximum operation is called the max pooling layer [29], which is the most used pooling structure. Thus the pooling process is:

$$q_{j,m} = \max_{c=1}^{r}\left(q_{j,(m-1)\times s+c}\right) \tag{4}$$

where $q_{j,m}$ denotes the output of m-th pooling zone in the j-th feature map, s is the moving step of the pooling zone, and r denotes the size of the pooling zone, which indicates the lengths of data pooled together.

The classification of the DCNN model is composed of the fully-connected layer and the classifier, which is to classify the higher-level information from the previous layers. And the softmax regression model [30] is often used to get the probability distribution of the different types in the samples. The formula of the softmax regression can be expressed as:

$$O\left(\theta^{(i)}x\right) = P(y = i|x;\theta) = \frac{e^{\left(\theta^{(i)}x\right)}}{\sum_{j=1}^{K} e^{\left(\theta^{(j)}x\right)}} \tag{5}$$

where $\theta^{(i)}$, $(1 \le i \le N)$ are the parameters of the model, and the output $O\left(\theta^{(i)}x\right)$ is the estimated probability value for the category i with the input $\boldsymbol{x}$.

2.2. Training method

The cross-entropy between the estimated output probability distribution and the target class probability distribution is applied as the cost function of the DCNN based model:

$$L = -\sum_{x} p(x)\log q(x) \tag{6}$$

where $p(x)$ denotes the target class probability distribution and $q(x)$ denotes the estimated output probability distribution of the softmax layer.

In order to optimize the learnable parameters of the network, the Adaptive Moment Estimation (Adam) Stochastic optimization algorithm [31] is applied to minimize the cost function in the training process of the DCNN based model. Adam is able to compute adaptive learning rates for each parameter using the first moment (the mean) and the second moment (the uncentered variance) of the gradients. And it has succeeded in the optimizing the learning rate for DCNN faster than the similar algorithms.

3. The DCNN based fusion method of Two-direction vibration data for fault diagnosis

3.1. Flowchart of the proposed Method.

In this paper, a DCNN based fusion method of vibration signals is proposed for health state identification of the planetary gearbox. In this method, the horizontal and the vertical vibration signals are used to construct the input to enrich the operation information of the object, which improves the accuracy of diagnosis. And the DCNN model can fuse data and extract rich fault features from the raw data of two-direction vibration signals automatically, which improves the intelligence of diagnosis. In addition, as the vibration data is time series only correlated in time, 1D convolutional structure of DCNN is chosen for fault diagnosis.

The procedure of the novel method is shown in Fig. 3. And there are four main processes: two-direction vibration signal acquisition, data preprocessing, data level fusion, and fault recognition. Firstly, the vibration signals are collected by the accelerometers positioned horizontally and vertically. Secondly, the raw signals are sliced with overlap, which helps increase the volume of training data. The process is shown in Fig. 4. Each slice (the frame part of the figure) is used as a segment to construct input sample for DCNN model. In this way, a signal of length l can be divided into n segments:

$$n = (l - N)/m + 1 \tag{7}$$

where N is the number of data points of each segment; m is the length of the shift. Thirdly, the two segments of the horizontal

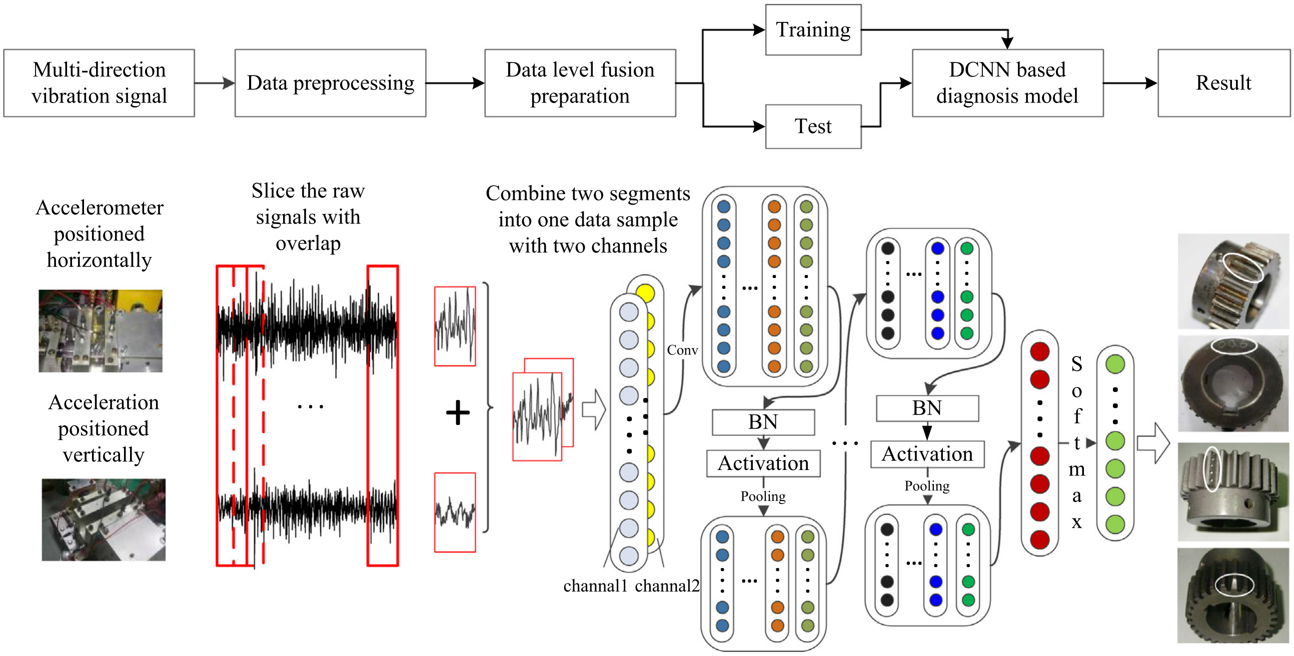
Multi-direction vibration signal
Data preprocessing
Data level fusion preparation
Training
Test
DCNN based diagnosis model
Result
Accelerometer positioned horizontally
Acceleration positioned vertically
Slice the raw signals with overlap
Combine two segments into one data sample with two channels
Conv
BN
Activation
Pooling
BN
Activation
Pooling
channal1 channal2
Softmax

Fig. 3. Flowchart illustrating of the proposed method.

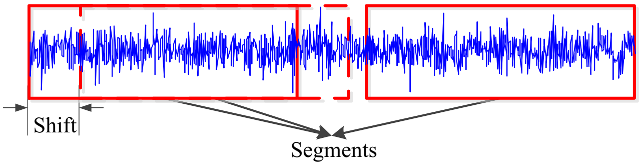

Fig. 4. Data augment with overlap.

and the vertical vibration signals at the same period of time are combined together as an input sample with two channels. And then without the tedious feature extraction process, the samples are put into the DCNN based model as input to train the diagnosis model. In addition, the test dataset is used to evaluate the effectiveness of the DCNN based diagnosis model. It should be noted that the third step is actually preparation for data level fusion which is implemented in the following DCNN model. In the following, the data processing method proposed in this paper is called channel overlay.

3.2. Model design of DCNN

The structure of the DCNN model is closely related to the number of the convolutional layers and the size of the kernels, and the model design of DCNN should match the characteristics of mechanical fault diagnosis. First of all, the learning ability of the DCNN is positively related to the number of layers. To a certain extent, the deeper the DCNN is, the stronger the expression ability of the network. If the DCNN structure is too simple, the learning ability is so poor that it cannot effectively integrate the information of the sensors. However, a deep structure of network means that more training data are needed to train the large number of parameters. Otherwise it may produce problems such as overfitting, local extremism and so on. Meanwhile, it is difficult to determine the most appropriate architecture of the model and it is time-consuming to train the model because of the complicated hyper-parameters [4]. Secondly, compared with the small convolutional kernels, the wide kernels in the first convolutional layer have a larger receptive field, which could provide more information for the latter layer. What is more, it is of great importance for mechanical fault diagnosis that the first convolutional layer with the wide kernel could better suppress high frequency noise [22]. In addition, the wide kernel could extract features from a relatively long time series, which might be the features in the frequency domain [4,32]. Although some basic principles for setting hyper-parameters are mentioned above, the specific values still need to be tested and adjusted.

4. Experiment and results

4.1. Methodology

4.1.1. Experiment setup

Experiment data is collected to evaluate the effectiveness of the method presented in this paper. As Fig. 5 shows, the planetary

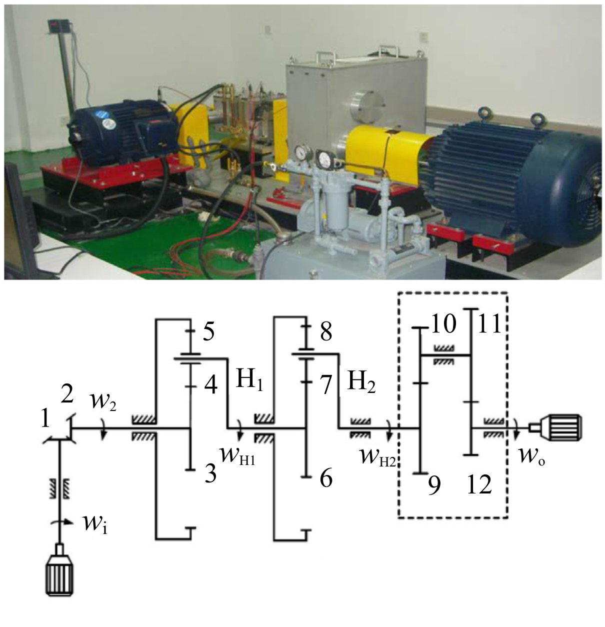

Fig. 5. Planetary gearbox test rig.

gearbox test rig consists of a driven motor, a bevel gearbox, a two-stage planetary gearbox, a straight gearbox, a loading motor, a lubrication system and a control system. The gear parameters of the planetary gearbox test rig are presented in Table 1.

Experiments are carried out on the sun gear of the 1st stage of the planetary gearbox under different health states, including normal, worn tooth, eccentric, chipped tooth and pitting tooth, and some images of the sun gears with manually created defect are shown in Fig. 6, where the worn tooth is manually induced by a file and the pitting tooth by machining. And each fault state contains two levels of fault, slight and severe respectively. Therefore there are nine health states in total. In the experiments, five sensors are employed, as shown in Fig. 7. Two accelerometers are mounted perpendicular to one another on the case of each stage to collect vibration data in the horizontal and the vertical direction synchronously. Tachometer is mounted over the shaft of the second stage planetary carrier. The technical specifications and settings of the experiment are summarized in Table 2. In addition, 3 experiments are performed for each condition and a component with defect is installed with the rest of components normal in each experiment. Therefore, the experimental dataset contains 9 different conditions and 3 sets of data for each condition. Each set of data is comprised of 4 channels of accelerometer signals and 1 channel of tachometer signal, with signal length of 102,400 points. In this study, only the two channels of vibration signals acquired from the 1st stage of the planetary gearbox are used. An example of the vertical vibration signals acquired under load-free conditions and their spectrums are shown in Fig. 8.

4.1.2. Data processing

As shown in Fig. 9, several methods are introduced to further evaluate the performance of the proposed method, and the proposed method is marked in dotted box.

At first, the collected signals are divided into segments. Considering the sampling frequency, the rotation speed and the convenience of the signal analysis in the frequency domain, each segment contains 2048 points. In addition, the segments for training samples are overlapped with shift of 300 points to augment data and there is no overlap among those for test samples. In this study, the segments for training samples are derived by the vibration signals obtained in the two of the experiments and those for test samples by the remaining experiment under each load and health state. Therefore, a vibration signal with 102,400 points can provide 335 ((102400–2048)/300 + 1 = 335.5) segments for training samples or 50 (102400/2048 = 50) segments for test samples, 330 or 25 segments of which are selected for follow-up use. Next, to test the performance of DCNN with various input formats, three datasets consisting of samples generated by three different data processing methods, namely no processing, data stitching [4] and channel overlay, are built, which are named Dataset A, B and C respectively. No processing considers the data segment of a single sensor as the input of the model, while data stitching and channel overlay combine the two segments of the horizontal and the vertical vibration signals at the same period of time together. Therefore, the number of samples in Dataset B and C is the same, which is half of Dataset A. Data stitching refers to splicing the segments of multi signals end to end into one data sample, and in this way the shape of each data sample will be (4096, 1, 1). It should be noted that in this study, the shape of the sample is (in_width, in_height, in_channels), representing the width, the height and the depth of the input sample respectively. Different from data stitching, the shape of data sample by channel overlay will be (2048, 1, 2). In short, three datasets containing 9 conditions under both two loads, 0 N m and 41.2 N m, are constructed, which with the category labels in details is presented in Table 3. In particular, the datasets are completely balanced in terms of gearbox con-

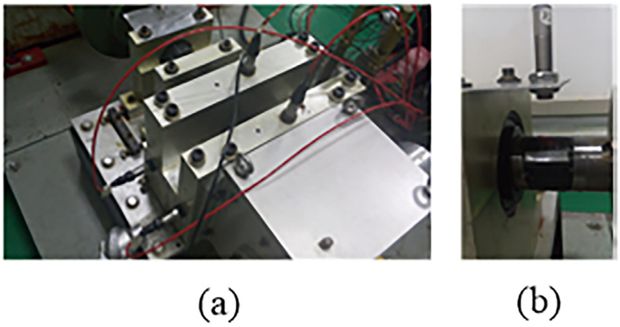

Fig. 7. Installations of sensors: (a) Accelerometers; (b) Tachometer.

Table 2
Technical specifications and settings of the experiment.

Property	Value
Accelerometer product type	SQI 601
Rotation speed	20 Hz
Sample frequency	5120 Hz
Sampling time	20 s
Load	0 N m, 41.2 N m

Table 1
Gear parameters of test rig.

Gear	Bevel gear		1st Stage of Planetary Gearbox			2nd Stage of Planetary Gearbox		
	Input	Output	Sun	Planet	Ring	Sun	Planet	Ring
Teeth Number	18	36	32	40(3)	112	28	34(4)	96

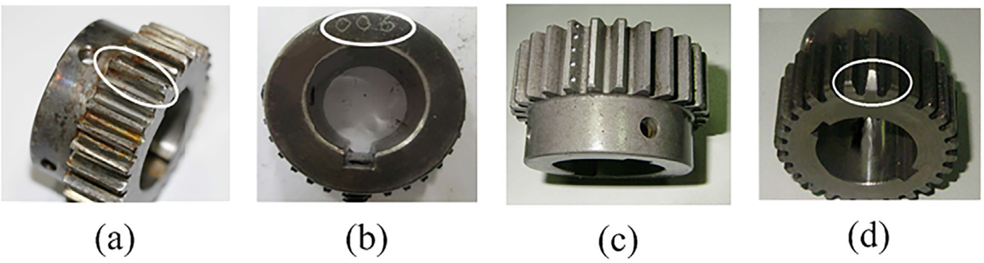

Fig. 6. Sun gears with defect: (a) Worn tooth; (b) Eccentric; (c) Pitting tooth; (d) Chipped tooth.

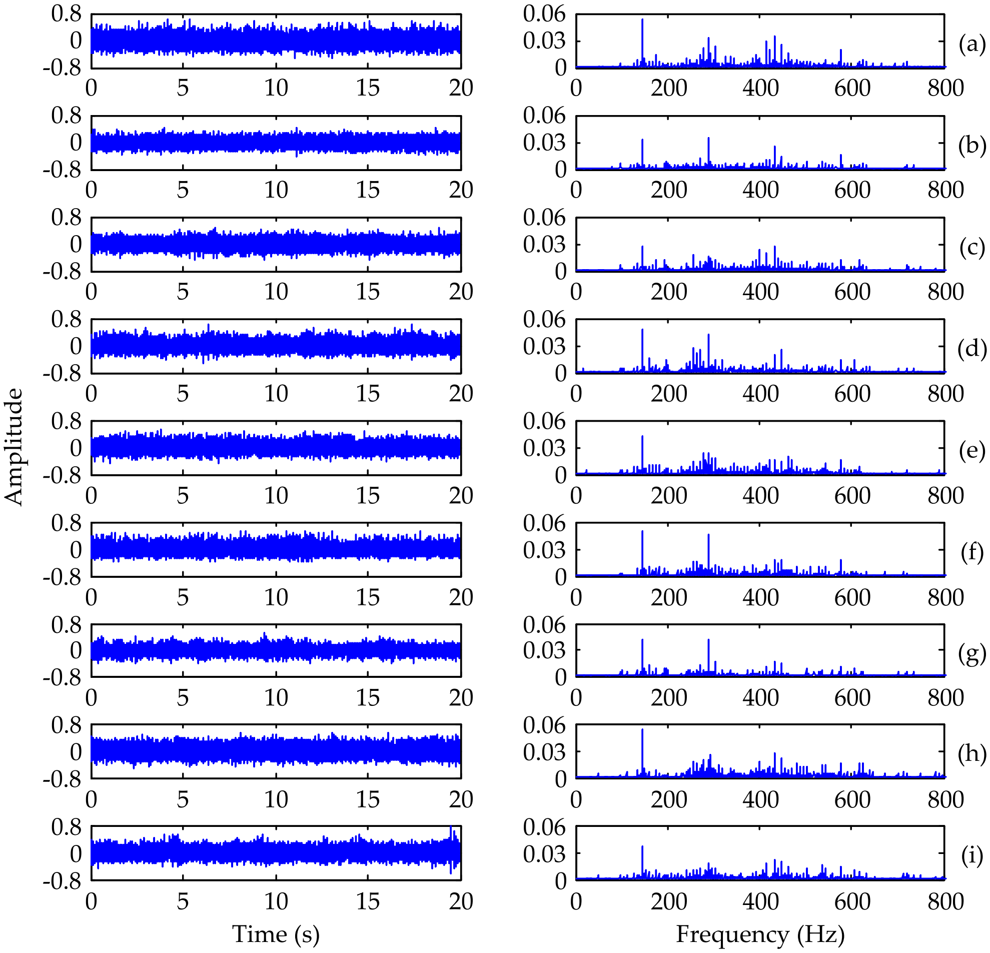

Fig. 8. Vibration signals and their spectrums of 9 health conditions: (a) normal, (b) slight worn tooth, (c) severe worn tooth, (d) slight pitting tooth, (e) severe pitting tooth, (f) slight chipped tooth, (g) severe chipped tooth, (h) slight eccentric and (i) severe eccentric.

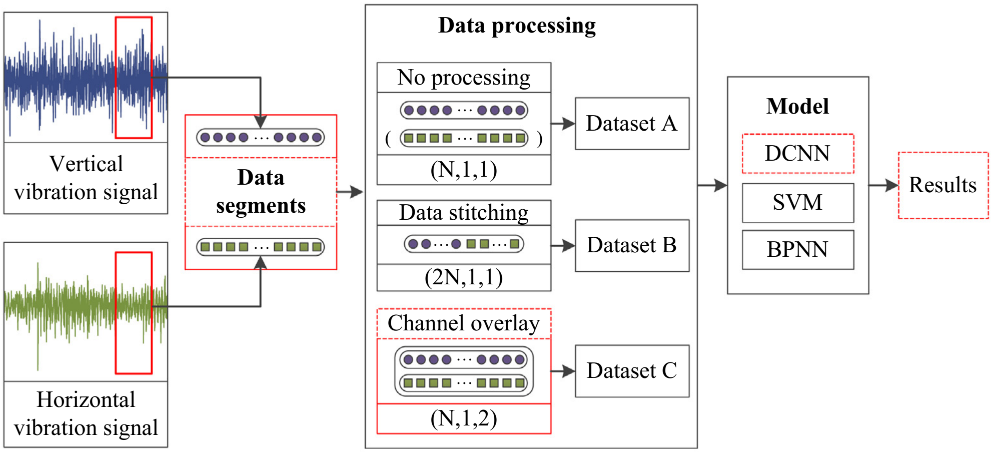

Fig. 9. Flowchart of experiments.

dition and load. Meanwhile, two commonly used models, SVM and BPNN, are utilized as comparisons of DCNN. In order to reduce the error of the result caused by particularity and contingency, the training set is divided into 10 equal parts by stratified sampling, and then 9 of them are used to train the model and the test set is input to get the corresponding result. Repeat the above process, and ten results of each method can be obtained. Finally, the average is taken as the final result. The process of the validation method is shown in Fig. 10.

4.2. Details of the models

The architecture of the DCNN model with a relatively good performance is displayed in Table 4, where the BN layer and the acti-

Table 3
Description of three planetary gearbox datasets; Load = 0, 41.2 Nm.

Gearbox condition		Normal	Worn tooth		Pitting tooth		Chipped tooth		Eccentric		Sample shape
Category label		1	2	3	4	5	6	7	8	9	
Fault level			slight	severe	slight	severe	slight	severe	slight	severe	
A no.	Train	2640	2640	2640	2640	2640	2640	2640	2640	2640	(2048,1,1)
	Test	100	100	100	100	100	100	100	100	100	
B no.	Train	1320	1320	1320	1320	1320	1320	1320	1320	1320	(4096,1,1)
	Test	50	50	50	50	50	50	50	50	50	
C no.	Train	1320	1320	1320	1320	1320	1320	1320	1320	1320	(2048,1,2)
	Test	50	50	50	50	50	50	50	50	50	

The samples in Dataset A and B will be reshaped to (2048, 1) and (4096, 1) when they are input to SVM and BPNN.

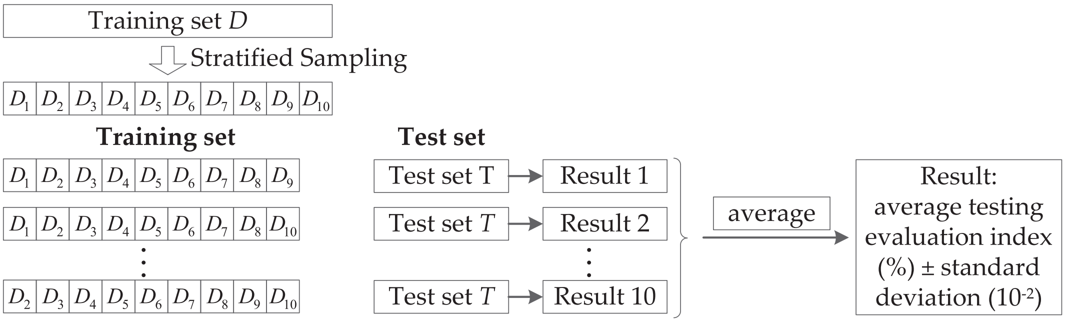
Training set D
Stratified Sampling
D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9 D_{10}
Training set
Test set
D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9
D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_{10}
D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9 D_{10}
Test set T
Test set T
Test set T
Result 1
Result 2
Result 10
average
Result: average testing evaluation index (%) ± standard deviation (10^{-2})

Fig. 10. The process of validation method.

Table 4
Details of the DCNN model used in experiments.

Layer	Parameters
Convolution (C1)	KW = 512; KH = 1; KC = 2; KN = 16; Strides = 2; B = 0; Padding: Yes
Pooling (S1)	S = 2
Convolution (C2)	KW = 65; KH = 1; KC = 16; KN = 32; Strides = 1; B = 0; Padding: Yes
Pooling (S2)	S = 2
Convolution (C3)	KW = 3; KH = 1; KC = 32; KN = 64; Strides = 1; B = 0; Padding: Yes
Pooling (S3)	S = 2
Convolution (C4)	KW = 3; KH = 1; KC = 64; KN = 64; Strides = 1; B = 0; Padding: Yes
Pooling (S4)	S = 2
Convolution (C5)	KW = 3; KH = 1; KC = 64; KN = 64; Strides = 1; B = 0; Padding: Yes
Pooling (S5)	S = 2
Convolution (C6)	KW = 3; KH = 1; KC = 64; KN = 64; Strides = 1; B = 0; Padding: Yes
Pooling (S6)	S = 2
Fully-connected (FC)	Nodes: 100, activation = ReLU
Dropout	Rate: 0.5
Softmax	Nodes: 9, activation = Softmax

KW = kernel width; KH = kernel height; KC = kernel channel; KN = number of kernels in the convolutional layer; B = bias; S = sub-sampling rate.

vation layer are not listed, and the activation layer is sandwiched between the BN layer and the pooling layer. The size of the first convolutional kernel is 512 × 1, the second kernel is 65 × 1 and the rest are 3 × 1, which makes the network deeper to improve the expressivity of the features. The details of every layer can be found in Section 2.1. In addition, the kernel function of SVM is Gaussian radial basis function (RBF). The BPNN model contains three hidden layers with the architecture of 200–100-50, one input layer with the number of nodes according to the sample shape and one output layer with 9 nodes. ReLU is selected as the activation function. The models are all implemented in Python on the software Spyder and an E5-2650 processor at 2.3 GHz and a GPU of NVIDIA Quadro K4200.

4.3. Experimental results

4.3.1. Evaluation metrics [19]

To facilitate the comparison of the performance of the methods with different input or models, four result measurements are calculated to quantify the performance, namely the mean accuracy, recall, precision and F1-score. The formulas of the four measurements can be seen in Eqs. (8)–(11), with $|TP|$ being the number of samples classified correctly into class y; $|TN|$, the number of samples not in y classified into classes except y; $|FP|$, the number of samples classified incorrectly into y; and $|FN|$, the number of samples in y classified into other classes. Recall and precision are two metrics widely used in the field of information retrieval and statistical classification to evaluate the quality of results. Generally speaking, if the condition monitoring system is able to trigger an alarm when the classifier detects a fault, faults could be more easily alerted even if there are some false alarms. However, too many false alarms mean more operational cost due to the unnecessary downtime. In other words, if more alarms are triggered, more faults would be recognized by the maintenance personnel, and recall is high in this situation. But there would be more false alarms, which will result in lower precision. On the other hand, if there are no false alarms and alarms are only triggered when the real fault is detected, but some faults are missed, then precision will be high, but recall is low. Note that it is more dangerous when faults are missed. There is basically a complementary relationship between accuracy and recall, and a good classifier will maximizes both error measurements so that an alarm is triggered when a failure does occur, no misses are missed, and no false alarms occur. The F1-score is an evaluation index that combines the two indicators and is used to comprehensively reflect the overall indicator. Also the accuracy is an evaluation of the correct rate of the classifier as a whole, which is often used to evaluate the performance of different models. In addition, Macro-average is chosen to get the final results [33].

$$Acc = \frac{|TP| + |TN|}{|TP| + |FP| + |FN| + |TN|} \tag{8}$$

$$Precision = \frac{|TP|}{|TP| + |FP|} \tag{9}$$

$$Recall = \frac{|TP|}{|TP| + |FN|} \tag{10}$$

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{11}$$

4.3.2. Results and discussion

t-SNE (t-distributed stochastic neighbor embedding) [34] presented by Laurens van der Maaten and Geoffrey Hinton in 2008 is a nonlinear dimension reduction algorithm to visualize high-dimensional data, which is applied to investigate the feature learning process. Taking one trial in experiments with the proposed method for example, the visualization of the feature learned from the test dataset layer by layer is shown in Fig. 11, where the labels 1–9 correspond to the 9 health states of the planetary gearbox presented in Table 3. The result demonstrates the effectiveness of the proposed method with the powerful ability in extracting discriminant features from the raw vibration data.

By observing Fig. 11, some interesting phenomena can be got. Firstly, the extracted features become more distinguishable as the network layers become deeper. In early layers, the features extracted are relatively shallow, and the features of each category are not divisible, which is reflected in the Fig. 11 (b) with a messy distribution of the data points of each category. As the layers goes deeper, the data points of the same category begin to aggregate, and the data points of different categories are gradually separated. In the last layer of the DCNN, the data points of different conditions

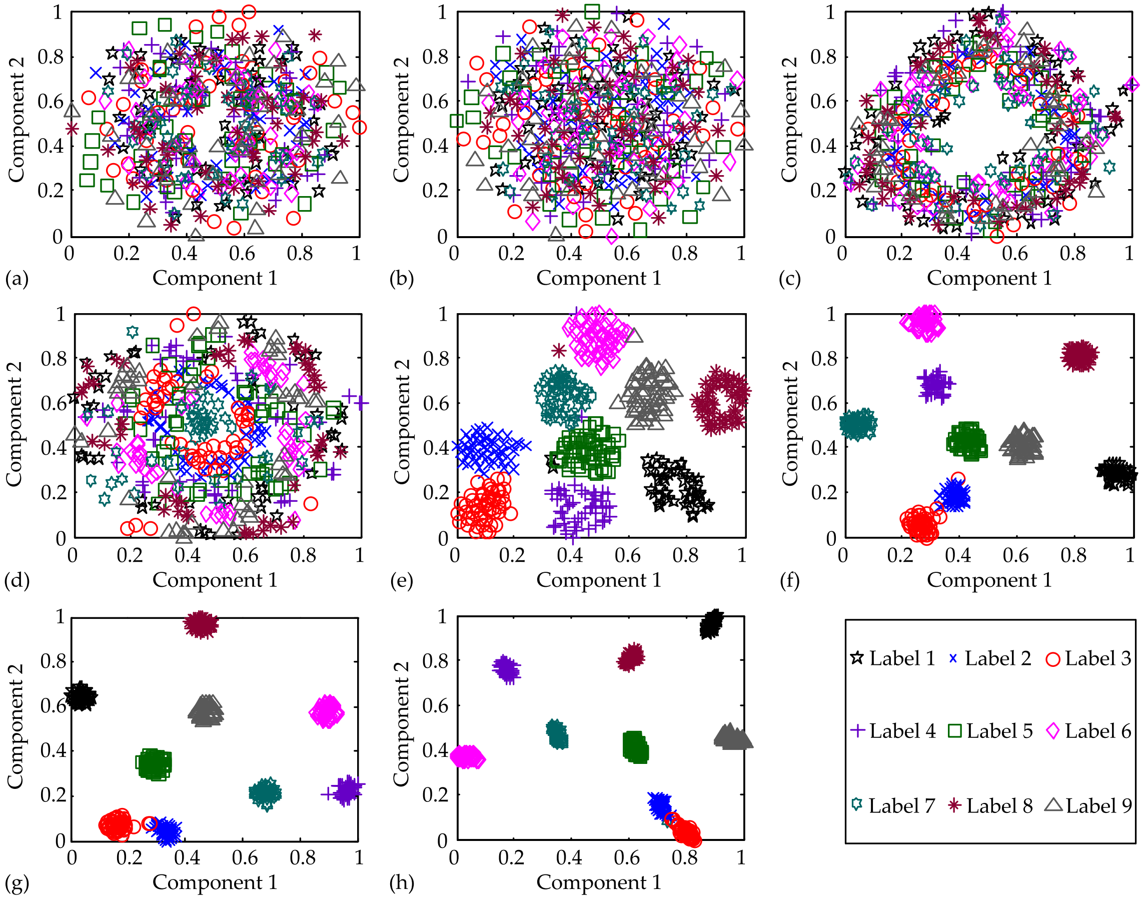

Fig. 11. Visualization of the representations of the test samples learned from the raw data, each convolutional layers and the last fully-connected layer via t-SNE. (a): the raw data; (b)-(g): six convolutional layers from shallow to deep; (f): the fully-connected layer.

Table 5
Performance results of the different diagnosis models with three datasets.

Data processing	Dataset	No. of test samples	Metric	DCNN	SVM	BPNN
No processing	A	900	Accuracy	89.08 ± 9.65	31.28 ± 1.01	44.02 ± 7.41
			Recall	88.88 ± 9.63	31.28 ± 1.01	25.59 ± 13.25
			Precision	91.81 ± 5.82	41.51 ± 1.51	39.67 ± 11.82
			F1-score	90.00 ± 9.56	32.21 ± 1.05	29.55 ± 13.31
Data stitching	B	450	Accuracy	97.38 ± 3.29	56.71 ± 0.89	57.77 ± 2.02
			Recall	97.31 ± 3.34	56.71 ± 0.89	41.18 ± 7.08
			Precision	97.81 ± 2.57	56.13 ± 1.13	49.82 ± 5.01
			F1-score	97.34 ± 3.37	55.85 ± 1.04	44.13 ± 6.44
Channel overlay	C	450	Accuracy	99.22 ± 1.06		
			Recall	99.20 ± 1.12		
			Precision	99.32 ± 0.85		
			F1-score	99.22 ± 1.06		

The format of the result is: average testing evaluation index (%) ± standard deviation (10^{-2}).

reach a better distribution, and there is less intermixing between them with obvious boundary. While in the fully-connected layer, the features are further integrated, and it can be seen from the Fig. 11(h) that the distinguishability between the categories is very good with the data points in each category densely distributed. Secondly, as shown in Fig. 11(h), the feature distribution of Condition (2) and Condition (3) corresponding to the label 2 and the label 3 has a small overlap that cannot be divided, which suggests that the proposed method might not perform so well in slight and severe worn tooth identification. In order to eliminate the influence of random error, the experiment is carried out several times, and the features of the fully-connected layer are visually analyzed. Then we find that the diagnosis method proposed in the paper may not perform very well in discriminating the degree of the failure, but it has no problem identifying the condition type. What is more, the variation of the feature distribution shows the strong nonlinear mapping ability of the DCNN.

The results of the experiments are displayed in Table 5. Among them, the proposed method takes Dataset C by channel overlay as input and uses DCNN for feature extraction and fault diagnosis. It should be noted that the samples in Dataset C cannot reshape into the format suitable for SVM and BPNN. The diagnosis results show that the performance of the proposed method is superior to other methods in health state identification of the planetary gearbox. As can be seen from Table 5, the four error measurements with the DCNN based model are over 89%, yet the results of SVM and BPNN are below 60%, which illustrates that the DCNN based diagnosis model outperforms the SVM and the BPNN based models when the input is raw data and demonstrate the superiority of the deep architecture in learning nonlinear fault features from the raw data automatically.

What is more, data fusion could improve the performance of the diagnosis method. In this study, three datasets contain the same amount of information, but the data from different sensors in Dataset A is independent and that in Dataset B and C works together for the final results. Take DCNN based diagnosis model as an example, it can obviously be seen from the Table 5 that the diagnosis methods with Dataset B and C obtain better results, over 97%, than the method with Dataset A. It proves that the fusion of vibration signals in two directions can improve the performance for fault diagnosis. In addition, the information on run time of different datasets based on DCNN is shown in Fig. 12, where 30 epochs were conducted in one trial of training. And an epoch refers to the process that all data have been sent to the network to perform a forward and back propagation. As can be seen from the figure, the method proposed in this paper is more efficiency than other methods.

Based on the DCNN model, the result of Dataset C is better than Dataset B with higher value and smaller fluctuations of the evaluation metrics. Data stitching connects the signal segments from different sensors end to end, which mainly increases the amount

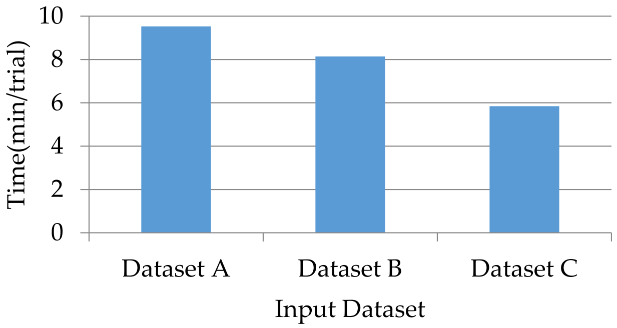

Fig. 12. Run time of different datasets based on DCNN.

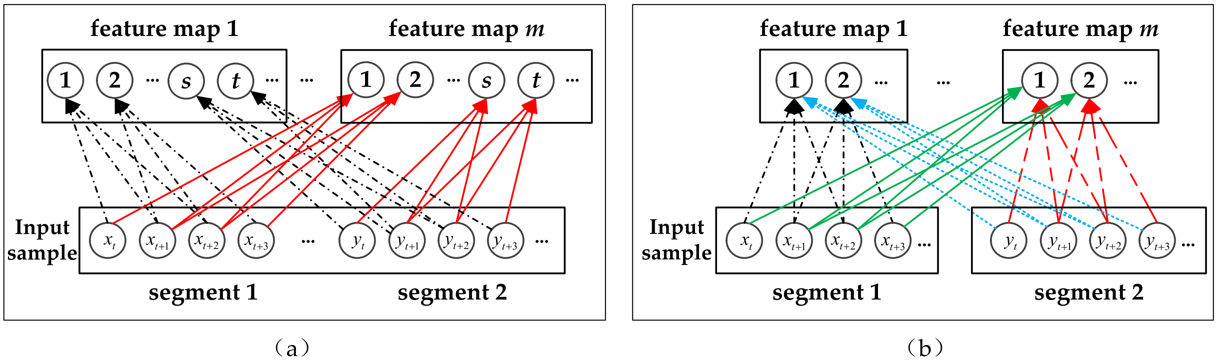

Fig. 13. The implementation principle of convolution. (a): Data stitching; (b): Channel Overlay.

of information of the sample and does not adequately utilize the relevance of data at the same time. In contrast, the proposed method in the paper not only increases the amount of information of the sample, but also preserves the information correspondence of the different signals at the same time. Specifically speaking, it is assumed that at time t, the signals of the horizontal and the vertical vibration sensors are $x(t)$ and $y(t)$ respectively, which determine the vibration information at the moment of the detected object together. In the method proposed in this paper, these two signals are like the channels of a RGB image, which coincide at the same pixel position, and constitute the information of this position. However, the relationship information of $x(t)$ and $y(t)$ is not preserved by data stitching, which leads to a slightly worse result than the proposed method in this paper. On the other hand, we try to explain the reason of the result from the implementation principle of convolution. As shown in Fig. 13, the neuron node in each feature map is the feature extracted from the input sample. In the data stitching method, the features in the feature map are almost extracted from the information from the same data segment. For example, the feature ① in feature map 1 is extracted from x_t , x_{t+1} and x_{t+2} of segment 1. While the data points x_t , y_t of two segments from the sensor signals in two direction act on the same neuron node in the same output feature map, which preserves the time correlation between the signal sequences, as shown in Fig. 13(b). And the learned convolution kernel parameter is equivalent to the weight distribution for data fusion of each data point, which is the essence of data adaptive fusion with the DCNN model.

5. Conclusions

This paper presents a DCNN based data fusion method of two-direction vibration data for health state identification of planetary gearboxes. Two vibration signals emitted by the same gearbox in horizontal and vertical directions are used to cancel out the noise and give a more trustworthy diagnosis result. The proposed method preserves the time correlation between the signals from different sensors with the preprocessing method named channel overlay. In addition, the proposed method can integrate information adaptively and extract features automatically for the final fault diagnosis decision with fewer requirements of the expert diagnosis experience and the signal processing techniques, as DCNN is an end-to-end machine learning system and it could learn transformations on the raw data that result in better representation of the data for the eventual classification task in the output layer. The effectiveness of the proposed method is validated through the data collected from the planetary gearbox test rig. Experiments using the SVM and the BPNN based model in different fusion methods are carried out to demonstrate the superiority of the proposed method. Under the same data processing, the classification accuracy of DCNN based method could achieve 90% even without data fusion, far greater than the result of SVM and BPNN. The contrastive results show that the DCNN can extract features with good distinction from the raw data than the traditional models. Compared with data stitching for DCNN, the proposed method yields an increase in evaluation metrics of approximately 2 percent, with more stable results and less cost in time. However, the method requires teaching for particular diagnostic object, which requires as much data as possible in multiple states.

In future, work will focus on more mechanical objects to test the DCNN based fusion method in feature learning and explain the superiority of the method in more detail in theory. Furthermore, as there are still some misclassifications possible, the fusion of more vibration signals will be considered. In addition, some denosing preprocessing will be combined to improve the performance of the proposed method.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant No.51475463, 51775550), Provincial Natural Sciences Foundation of Hunan (Grant No. 2018JJ3604). All the authors contributed equally to this work. Thanks also go to the anonymous reviewers and the editor for their suggestions and comments, which have helped improving the quality of this paper.

References

[1] Y. Lei, J. Lin, M.J. Zuo, Z. He, Condition monitoring and fault diagnosis of planetary gearboxes: a review, Measurement 48 (2014) 292–305.

[2] M. Khazaee, H. Ahmadi, M. Omid, A. Banakar, A. Moosavian, Feature-level fusion based on wavelet transform and artificial neural network for fault diagnosis of planetary gearbox using acoustic and vibration signals, Insight-Non-Destructive Test. Condition Monit. 55 (2013) 323–330.

[3] Y. Lei, Z. He, J. Lin, D. Han, D. Kong, Research advances of fault diagnosis technique for planetary gearboxes, Jixie Gongcheng Xuebao (Chin. J. Mech. Eng.) 47 (2011) 59–67.

[4] L. Jing, T. Wang, M. Zhao, P. Wang, An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox, Sensors 17 (2017) 414.

[5] M. Safizadeh, S. Latifi, Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell, Inf. Fusion 18 (2014) 1–8.

[6] L. Xuejun, Y. Dalian, G. Dengta, J. Lingli, Fault diagnosis method based on multi-sensors installed on the base and KPCA, Chin. J. Sci. Instrum. 7 (2011) 017.

[7] L. Jiang, Y. Liu, X. Li, A. Chen, Gear fault diagnosis based on SVM and multi-sensor information fusion, J. Central South Univ. (Sci. Technol.) 41 (2010) 2184–2188.

[8] T.P. Banerjee, S. Das, Multi-sensor data fusion using support vector machine for motor fault detection, Inf. Sci. 217 (2012) 96–107.

[9] L. Jing, M. Zhao, P. Li, X. Xu, A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement 111 (2017) 1–10.

[10] R. Socher, B. Huval, B. Bath, C.D. Manning, A.Y. Ng, Convolutional-recursive deep learning for 3d object classification, Adv. Neural Inf. Process. Syst. (2012) 656–664.

[11] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, https://doi.org/10.1038/nature14539.

[12] L. Deng, J. Li, J.-T. Huang, K. Yao, D. Yu, F. Seide, M.L. Seltzer, G. Zweig, X. He, J.D. Williams, Recent advances in deep learning for speech research at Microsoft, ICASSP (2013) 64.

[13] G. Hinton, L. Deng, D. Yu, G.E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T.N. Sainath, Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, IEEE Signal Process Mag. 29 (2012) 82–97.

[14] J. Tao, Y. Liu, D. Yang, Bearing fault diagnosis based on deep belief network and multisensor information fusion, Shock Vib. 2016 (2016).

[15] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409 1556 (2014).

[16] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[17] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, AAAI (2017) 12.

[18] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (1998) 2278–2324.

[19] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, S. Van Hoecke, Convolutional neural network based fault detection for rotating machinery, J. Sound Vib. 377 (2016) 331–345.

[20] X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement 93 (2016) 490–502.

[21] W. Zhang, G. Peng, C. Li, Rolling Element Bearings Fault Intelligent Diagnosis Based on Convolutional Neural Networks Using Raw Sensing Signal, in: Advances in Intelligent Information Hiding and Multimedia Signal Processing, Springer, 2017, pp. 77–84.

[22] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, Sensors 17 (2017) 425.

[23] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Sig. Process. 100 (2018) 439–453.

[24] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, T. Zhang, Deep model based domain adaptation for fault diagnosis, IEEE Trans. Ind. Electron. 64 (2017) 2296–2305.

[25] H.-K. Peng, R. Marculescu, Multi-scale compositionality: identifying the compositional structures of social dynamics using deep learning, PLoS One 10 (2015) e0118309.

[26] Y. Won, P.D. Gader, P.C. Coffield, Morphological shared-weight networks with applications to automatic target recognition, IEEE Trans. Neural Networks 8 (1997) 1195–1203.

[27] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167, (2015).

[28] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.

[29] Y.-T. Zhou, R. Chellappa, Computation of optical flow using a neural network, IEEE Int. Conf. Neural Networks (1988) 71–78.

[30] C.M. Bishop, Pattern recognition and machine learning (information science and statistics), Springer-verlag New York, Inc, Secaucus, NJ, USA, 2006.

[31] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, (2014).

[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

[33] C. Liu, W. Wang, M. Wang, F. Lv, M. Konan, An efficient instance selection algorithm to reconstruct training set for support vector machine, Knowl.-Based Syst. 116 (1) (2017) 65.

[34] L.V.D. Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.