Fault Detection of Planetary Gearboxes Based on Deep Convolutional Neural Network

Zhe Cheng, Niaoqing Hu, Jiageng Chen
Laboratory of Science and Technology on Integrated Logistics Support
National University of Defense Technology
Changsha, China
chengzhe@nudt.edu.cn

Ming Gao*, Qifeng Zhu
College of Intelligence Science and Technology
National University of Defense Technology
Changsha, China
gaoming@ nudt.edu.cn

***Abstract*—Due to the reliance on the expert experience and the signal processing approaches, traditional diagnosis methods lead to uncertainty in feature extraction and fault detection results. deep learning is a great method to overcome the shortcomings of traditional fault diagnosis. For the other side, the accelerometers in a single direction are not suitable enough to position-shift damages and the vibration data is generally nonstationary and noisy, which impacts the accuracy of fault detection. Therefore, as the reason that different measurement locations provide different sensitivity degree or complete data of the damages, this research presents a method based on deep convolutional neural network (DCNN) of vibration signal for early fault detection. The accuracy of this approach is validated based on the sensor data sets collected from a experimental rig. The results show that the DCNN based fault detection method presented in this paper could obtain promising identification results.**

Keywords- DCNN; Feature extraction; Fault detection; Planetary gearbox

I. INTRODUCTION

As well known, helicopter is a type of universal aircraft and widely used in lots of areas. Planetary gearbox is a critical component of the main transmission system of helicopter. Once there is a fault in the planetary gearbox, the safety of the helicopter will be threatened seriously. For this reason, it is one of the important enabling techniques in CBM area to monitor and detect the early faults in the planetary gearbox.

Deep Convolutional Neural Network (DCNN) has been applied in two-dimensional data processing areas successfully, such as Image data analysis and processing, and now DCNN based approaches are researched widely in condition monitoring and fault diagnosis of rotary machinery. Nowadays, lots of new architectures of DCNN have been developed after the first structure presented by LeCun [1-4]. Some structures of DCNN have been applied to the fault diagnosis of mechanical systems, such as bearing fault detection[5-9]. Taking the planetary gearbox of the main transmission system in helicopter as the object, a fault feature extraction and early detection method based on DCNN is researched in this research. Finally, the faults seeded experiments are carried out to validate the methods presented in this research.

National Natural Science Foundation of China (Grant No. 51775550)
Provincial Natural Sciences Foundation of Hunan (Grant No.2018JJ3604).

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

II. BASIC THEORY OF DEEP CONVOLUTIONAL NEURAL NETWORK

A. Architecture

As a typical feed-forward neural network, DCNN is a type of filter which could extract the essential features from the input data. The operation of convolution and pooling are carried out on the input data layer by layer via the filter, and then the topological features hidden in the input data are extracted layer by layer. As the layer of the neural network gets deeper, the structure features extracted gradually become abstractive. Finally, the feature representation of the input data is obtained, which is independent of the translation, rotation, and scaling of the input data. The main characteristic of DCNN is that the sparse sampling, the weight sharing and the sub-sampling in time domain or space domain are combined in one neural network. Sparse connections establish a non-fully connected spatial relationship between layers through a topology structure to reduce the number of training parameters; Weight sharing could avoid the over fitting of the algorithms effectively; and sub-sampling takes the advantage of the features such as the locality contained in the input data itself, to reduce the data dimensions, to optimize the network structure, and to ensure the displacement invariance. As the result of that, DCNN is very suitable for the data processing and learning of the big data.

DCNN is a type of multilayer perception, which is composed of input layer, convolutional layer, pooling layer, fully-connected layer and output layer, as shown in Figure 1. The convolutional layer and the sub-sampling layer are two typical network structures which are used for feature extraction. In other words, the high-level features of the input data will obtained via the feature extraction structure of the DCNN.

(1) Convolutional layer

The core component of the convolutional neural network is the convolutional layer, which has two characteristics as local connection and weight sharing. The convolutional layer consists of a convolution kernel that mapping the network node matrix of the previous layer into a unit node matrix of the next layer of neural networks. The unit node matrix is a node matrix with a length and a width of 1, and with an unlimited depth. The convolutional layer may contain more than one type of convolution kernel filter. In the process of feature extraction, the number of convolution kernels and the convolution kernel form could be adjusted as needed. Each convolution kernel is corresponding to a feature map, and the number of the feature maps obtained is the depth of the convolution layer.

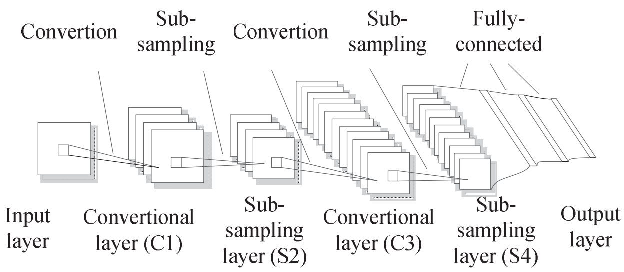


Figure 1. A typical structure of DCNN.

The forward propagation calculation method of the convolution kernel is similar to the full connection layer. Supposing that the current layer network is the lth layer, w_{ij}^{l} is used to represent the weight of the jth node of the output unit node matrix corresponding to the ith node of the input convolutional kernel filter. b_j^l is used to represent the bias parameter corresponding to the jth output node. Then, the value of the jth node of the unit matrix x_j^l is expressed as:

$$x_j^l = \sum_{i \in M_j} x_i^{l-1} \times w_{ij}^l + b_j^l \tag{1}$$

where M_j is the set of the input feature maps. The convolution kernel is used to traverse the feature map of the previous layer, after that the convolution operation is performed, and then all the local features in the feature map of the previous layer is extracted by convolution operation. The convolution kernel filter is moved in each step, and a value can be calculated. By stitching these values into a new matrix, the forward propagation of the convolutional layer is operated. After the convolution operation, the obtained result is usually nonlinearly processed. The activation functions commonly used are Sigmoid function and Rectified Linear Units (Relu) function. There are two key problems which should be solved in the convolution operation: convolution edge processing method and convolution step size. In the design process of the deep convolutional neural networks, the convolution step size and edge processing method could be selected flexibly. The specific convolution effect varies from target to target.

(2) Sub-sampling layer

The feature map obtained in the general convolution operation has a large dimension, which is not conducive to the subsequent operations. Therefore, after the convolution layer, there is usually a sub-sampling layer which also called as pooling layer. The sub-sampling process can be seen as the process of feature dimensionality reduction, thereby reducing the training data in the network. At the same time, the sub-sampling process reduces the complete learning of the features appropriately, which can prevent over-fitting and speed up the learning operation of the model. Not only that, the sub-sampling operation could ensure the invariance of the feature position. The calculation method of this neurons could be expressed as:

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

where $\mathrm{down}(\bullet)$ is sub-sampling function, β is network multiplicative bias.

However, it should be noted that the convolution layer is not followed by a sub-sampling layer all the time. There are two common methods for sub-sampling, one is the maximum pooling method, and the other is the mean pooling method. The essence of the pooling process is the expression of the local features. After the pooling operation, the number of feature maps does not change. Considering the loss of information, the size of pooling matrix should not be too large. In addition to reducing the dimension, the role of the sub-sampling layer also achieves the aim of quadratic feature extraction. In the deep convolutional neural network, the convolutional layer is followed by the structure of the sub-sampling layer, which is equivalent to two operation process of feature extractions.

B. Training

For the purpose of making the DCNN have the ability of learning, it is necessary to train the network. No need to find the exact mathematical expression between the output and the input, the DCNN could be trained with the labeled samples to obtain the mapping relationship between the output and the input. The training process of DCNN consists of two steps: forward propagation and back propagation. Forward propagation is to get the network output with the input of samples. For back propagation process, it is to calculate the error between the network output and the target output firstly, and then propagate the error value back to obtain the error of each layer, then adjust the network parameters by the random gradient descent method until the network converges or reaches the specified termination condition.

The cost function of the model based on the DCNN is the cross entropy between the target output value and the estimated output value, which is calculated as

$$L = -\sum_x p(x) \log q(x) \tag{3}$$

where $q(x)$ is the estimated output value of the probability distribution for the softmax layer, and is the target output class value of the probability distribution.

For the purpose of optimizing the DCNN parameters, a stochastic optimization algorithm named AdaME[10] is used, to minimize the function of cost for the model based on DCNN presented in this research. Compared to the traditional similar algorithms, this stochastic optimization algorithm has a better performance in the learning rate and efficiency.

III. Fault Detection Approach based on DCNN

A early fault detection approach based on DCNN is researched for the planetary gearbox of main transmission system in helicopter. Dual channel sensor data sets are used as input samples of DCNN directly, and one channel data is from the horizontal accelerometer installed on the top of the gearbox, the other channel data is from the vertical accelerometer installed on the left side of the gearbox. For the aim of improve the accuracy of fault detection, the data from two channels are fused to enhance the condition information of the planetary gearbox by the DCNN. Without considering preprocessing of the raw data in other traditional method, this approach extracts the fault features directly from the raw sensor data intelligently.

As shown in Figure2, the flow chart of the proposed approach is depicted layer by layer. There are 4 steps listed as follows: 1) dual channel data synchronous acquisition, that means the vibration data are acquired from two channel sensors by the data collection system; 2) dual channel data processing and fusion, in this step the two channel raw data are separated and overlapped to enrich the condition information of each samples, as shown in Figure 3. What should be noted is that each data fragment is applied to develop a sample used as the input of the DCNN in Step 2). The number of the data segments could be calculated by the Equation (4):

$$n=(l-N)/m \tag{4}$$

where N denotes the sample length; m denotes the data shift length. 3) multiple feature extraction, this operation is carried out based on the convolutional layers and the pooling layers of the DCNN, which has multiple processing repeatedly. The input samples are used to train and adjust the structure of the DCNN; 4)fault detection and recognition, and this is the final step of the proposed method, a fully connected layer and a softmax layer is included in this step. Additionally, the experimental data sets with some typical faults seeded is applied to validate the accuracy of the presented fault detection approach based on the DCNN.

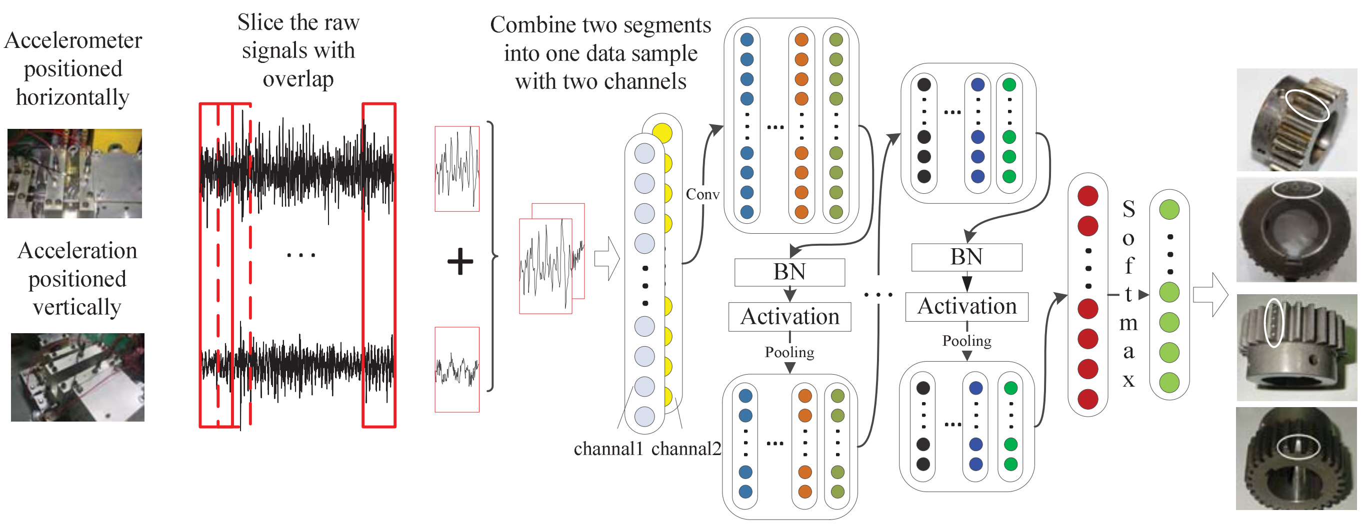


Figure 2. The flow chart of the proposed approach.

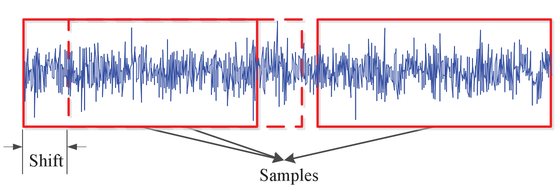


Figure 3. The overlapping method of the raw data.

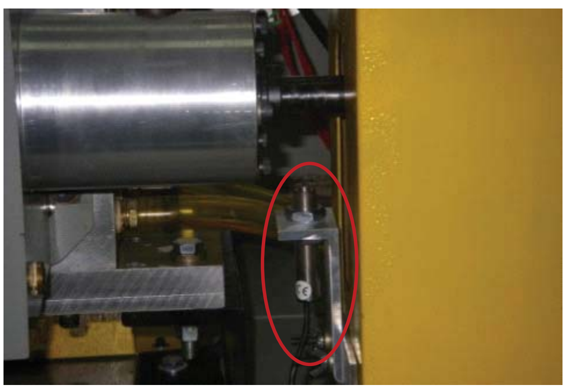
IV. Validation based on the Experimetal Data

A. *Illustration of the Experiments*

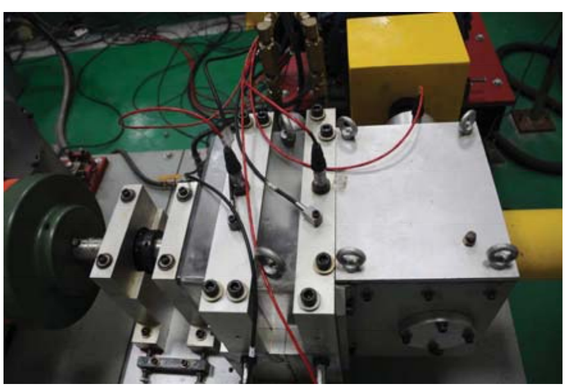
A set of experiments were conducted to validate the feature presented above on the test rig (shown in Figure 4(a)), which simulated the planetary gearbox of a rotary aircraft. The transmission system of the test rig is composed of one bevel gearbox, a two-stage deceleration planetary gearbox and one acceleration gearbox. The parameters of the bevel gearbox and the planetary gearbox are listed in Table 1. One optical speed sensor and five accelerators are installed on the test rig, as shown in Figure 4(b) and Figure 4(c).

As shown in Figure 4 (b), the optical tachometer was mounted on the casing of the first coupling which connected the output shaft of the drive motor and the input shaft of the spur-bevel gearbox, and this sensor was used to measure the speed of the input shaft of the test rig. The vibration signal generated by the planetary gearbox was picked up by four accelerometers bolted on the planetary gearbox casing referring to Figure 4 (c), and the electrical signal was transferred to the data acquisition system, which has a fore-charge-amplifier. The sampling frequency f_s for all the signal including vibration signal and tachometer signal is 10 kHz. The data sets had been acquired in all experiments corresponding to a time length of 10 s, so as to ensure all the interesting frequency components are included in the test data. A number of faults seeded experiments with different fault levels have been conducted on the test rig, and some fault seeded parts could be seen in Figure 4 (d).

(a) The configuration of the test rig.



(b) The optical tachometer in the experiments.



(c) The sensors mounted on the test rig.

(d) Some parts with faults seeded

Figure 4. The test rig, sensors and faults seeded parts in the experiments..

TABLE I. THE PARAMETERS OF THE PLANETARY GEARBOX.

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

Each type of the data sets collected from the two-sensor-channel is separated, and each segment is one sample, and the number of each sample is 1024. Then, the one-dimensional raw signal in time domain is processed to obtain a 32×32 two-dimensional raw signal map, as shown in Figure 5.

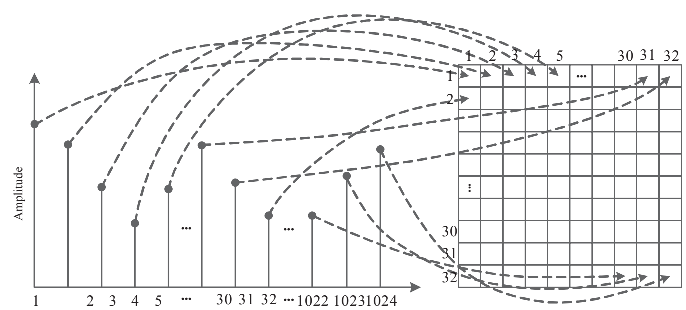


Figure 5. Diagram of 1D signal transforming to 2D signal map.

Finally, 8000 data samples were obtained, in which there are 2000 samples for each class. The number of the training samples was 7920, and the number of the test samples were 80 samples. The data sets and their tag numbers are shown in Table 2.

TABLE II. THE DATA SETS AND THEIR LABEL NUMBERS.

Class	Number of training samples	Number of testing samples	Sample labels
Health	1980	20	1
Sun gear with tooth chipped	1980	20	2
Sun gear with tooth pitting	1980	20	3
Sun gear with tooth crack	1980	20	4

B. Validation Results

The DCNN used in this research has a structure as 6C-2S-12C-2S-6C-2S, which consists of 3 convolutional layers and 3 sub-sampling layers. The number of nodes in the convolutional layer is 6 , 12 and 6 respectively, and the size of the convolution kernel is 5×5. The convolutional layer and the convolution kernel size are determined by the size of the raw signal matrix and the selected neural network structure. The data set under each load condition is diagnosed by a convolutional neural network, and the obtained results are as follows.

The experimental data sets as illustrated in Table 2 were used to train and test the DCNN model presented above. The validation results is shown in Figure 6 and Table 3. In the testing of 80, 77 samples are recognized correctly, and the detection accuracy is 96.25%. The confusion matrix of fault mode could be seen in Table 3, from which it could be found that the number and the rate of false alarm are 0, and the number of missed alarm is 1, and the rate of missed alarm is 1.25%.

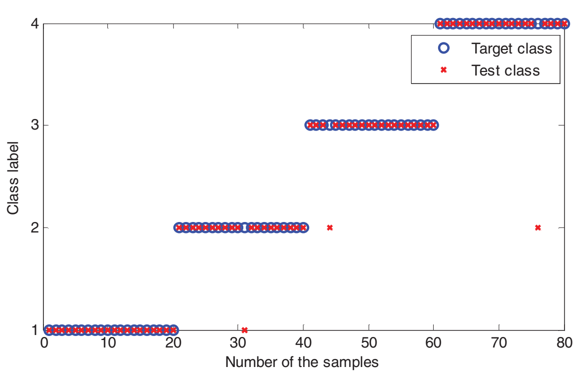


Figure 6. Detection results of DCNN.

TABLE III. THE CONFUSION MATRIX OF VALIDATION RESULTS.

Labels	1	2	3	4
1	20	0	0	0
2	1	19	0	0
3	0	1	19	0
4	0	1	0	19

V. CONCLUSIONS

A novel approach based on DCNN is presented to detect the early fault of planetary gearbox in the main transmission system of helicopter. Two channel sensor data sets are utilized as the input samples in this approach, and this operation reduce the effect of environment noise and the position shift of the damage in the moving part of planetary gearbox. The proposed method has a intelligent process in feature extraction and fault detection, and a promising results are provided based on the validation with the fault seeded experimental data sets. More structures of DCNN and the fusion with other DNN will be explored to detect, diagnose and predict the early faults in the rotary machinery in the future.

REFERENCES

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

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

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

[4] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, 86 (1998) 2278-2324.

[5] 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”, Journal of Sound and Vibration, 2016, 377, pp.331-345.

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

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

[8] 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, 2017, 17, pp. 425.

[9] 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”, Mechanical Systems and Signal Processing, 2018, 100, pp. 439-453.

[10] D.P. Kingma, J. Ba, Adam, “A method for stochastic optimization”, arXiv preprint arXiv,2014, 1412.6980.