Convolutional Neural Networks for Fault Diagnosis Using Rotating Speed Normalized Vibration

Dongdong Wei[1], KeSheng Wang[1(✉)], Stephan Heyns[2], and Ming J. Zuo[3]

[1] Equipment Reliability, Prognostics and Health Management Lab, School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China
keshengwang@uestc.edu.cn

[2] Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa

[3] Department of Mechanical Engineering, University of Alberta, Edmonton, Canada

Abstract. Fault diagnosis is vital for the health management of rotating machinery. The non-stationary working conditions is one of the major challenges in this field. The key is to extract working-condition-invariant but fault-discriminative features. Traditional methods use expert knowledge on the machines and signal processing to extract fault features from vibration signals manually. This paper regards this issue as a domain adaption problem and utilizes deep learning technique to learn fault discriminative features automatically. We teach deep Convolutional Neural Networks to pronounce diagnostic results from raw vibration data and propose a Rotating Speed Normalization method to improve the domain adaption ability of the neural network models. A case study of rotor crack diagnosis under non-stationary and ever-changing rotating speeds is presented. Using 95600 signal segments, we compare the diagnostic performance of ours and reported Convolutional Neural Network models. The results show that our model gives solid diagnostic accuracy from non-stationary vibration signals, and the proposed Rotating Speed Normalization method can successfully boost the performance of all investigated CNN models.

Keywords: Fault diagnosis · Rotating machine · Deep learning · Domain adaption · Convolutional Neural Networks

1 Introduction

Fault diagnosis for rotating machinery is very important in many industries such as the automobile, mining, and aerospace. It aims to reduce maintenance costs and avoid casualties. Many studies, e.g. [1] and [13], took data driven approaches in which diagnostic models are built upon historical data. However, traditional

A. Fernandez Del Rincon et al. (Eds.): CMMNO 2018, ACM 15, pp. 67–76, 2019.
https://doi.org/10.1007/978-3-030-11220-2_8

data driven methods often use expert knowledge and human labours to extract potential fault indicative features from raw sensory data. A new way is to extract features with deep learning models automatically [2,8].

Convolutional Neural Network (CNN) have been proven to be one of the most powerful deep learning models for many pattern recognition tasks, such as image classification [5] and music tagging [11]. In the past three years, CNNs have been applied to solve fault classification tasks for rotating machines. An overview is given in Table 1.

Table 1. CNN architectures and data sets in surveyed papers

Research item	#Layers	Max kernel	#Training
Chen et al. [3]	1	5×5	7200
Guo et al. [4]	3	5×5	2000
Lu et al. [14]	2	5×5	N/A
Li et al. [12]	3	5×5	10000
Jing et al. [9]	3	65×1	2620
Janssens et al. [7]	1	64×2	19200
Jing et al. [10]	1	$64, 32 \times 1$	360, 298
Zhang et al. [23]	5	64×1	19800
Zhang et al. [22]	6	64×1	19800
Xia et al. [19]	2	$17 \times 3, 2$	2520, 4200
Ince et al. [6]	3	9×1	468
This paper	9	3×1	40000

In Ref. [3,4,12,14], input signals are organized as 2D arrays to adapt CNNs designed for image classification. Other researches in Table 1 used one-dimensional CNNs and take 1D signals or frequency spectrum as inputs. This one-dimensional form is a more natural for fault diagnosis.

State-of-the-art CNN models [5] for image recognition go to 152 layers while fault diagnostic CNN models reviewed in Table 1 are comparatively shallow. In general, deeper models are more likely to overfit and demand larger data sets for training. However, neurons in deeper layers of CNNs receive information from larger signals segments [17], which can be beneficial for fault diagnosis. Although wide convolutional kernels used in [7,9,10] can enlarge the receptive field of neuron in shallow layers, CNNs are usually designed as deep structures to learn more complicated representations. Zhang et al. [22,23] used wide kernels only in the first layer and deepened the structure to 6 layers. Deeper structures with only small kernels for fault diagnosis have not been investigated according to our literature survey.

Despite the fact that deep learning has surprised us with high fault classification accuracy, the variable working conditions of machines still cause trouble. To collect the data of all possible working conditions for model training maybe prohibitive. This causes distributional differences in source and target data domains where we draw training and testing samples from. Cross domain learning problems are thus introduced [16]. Taking rotating speed as an example, the vibration behavior of a machine would change as the rotating speed varies. Training data collected in a limited range of rotating speed will be different to testing data from other ranges. In such a case, diagnostic models trained on the source domain data may not be well generalized for the target test data.

To perform domain adaption in fault diagnosis, Ref. [15,20,21] introduced statistical methods to force models to learn work-condition-invariant but fault-discriminative features. The above studies presented solid cross domain learning results for PHM09 Challenge data (different rotating speeds [15,20]) and CWRU Bearing data (different loads [15,21]).

However, two problems remain unsolved. One is that aforementioned methods require target domain data for training. In many industry situations, target domain data only show up during the running stage when the model training is already completed. The other one is that current studies only consider some fixed working conditions. In reality, the operational conditions, such as rotating speed, of a machine may be ever-changing and non-stationary. This is to say that the source and target domains may contain data collected under two ranges of working conditions instead of different but fixed ones.

Using raw vibration data and rotating speed information, in this paper, we adopt deep Convolutional Neural Networks (CNNs) to perform fault diagnosis. A new domain adaption method, rotating speed normalization (RSN), for vibration signals of rotating machines is proposed to tackle the cross domain learning problem caused by the fluctuation of rotating speed. The diagnostic performance and the effectiveness of the proposed domain adaption method are validated through a rotor crack diagnosis case study. The main contributions of this paper are summarized as follows:

1. Diagnosis performance of several different CNN structures, including a proposed 9 layer one, are studied and compared;
2. Experimental data are collected under large speed fluctuations instead of fixed ones, and the total amount of samples is 95600;
3. Considering the effect of centrifugal force, domain adaption for rotating speed fluctuation is implemented without target domain data.

In the following parts of this paper, Sect. 2 describes our CNN design, Sect. 3 explains the proposed RSN method. Section 4 presents the conducted case study. Section 5 concludes the paper.

2 CNN Architecture

Typically, CNN models are hierarchically arranged with convolutional layers, pooling layers, and fully connected layers. In a convolutional layer, convolutional kernels (also called filters) are trained to capture the local information of its input using convolutional operations. Pooling layers are designed to reduce the number of captured features and fully connected layers can be regarded as a conventional classifier that takes learned features as inputs. The output size of the last fully connected layer corresponds to the number of classes (health conditions) of a certain task. Instead of revisiting detailed foundations and training procedures for CNNs, paper [23] is recommended for further reading.

The CNN architecture used in [6] is depicted on the left side of Fig. 1. The WDCNN architecture used in [23] is drawn in the middle. Our proposed architecture is shown on the right hand side of Fig. 1. The black boxes represent inputs or outputs while the coloured boxes are different layers in CNNs. Taking the design of vgg-net in [17], we construct a 9 layer CNN (DCNN9) with only small kernels. The three CNN architectures in Fig. 1 are compared using the same data set. It should be noticed that the output size of the last fully connected layer will be set equal to the number of classes of a task.

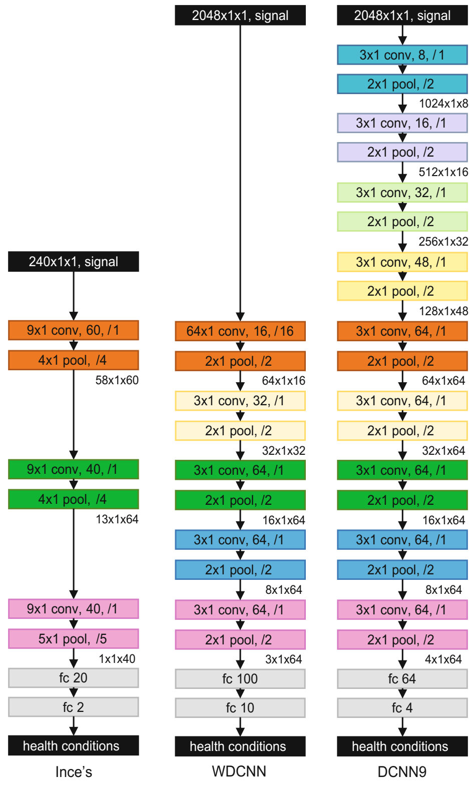

Fig. 1. Network architectures for fault diagnosis. Left: Ince's model [6] with 3 convolutional layers. Middle: WDCNN model [23] with 1 wide kernel layer and 4 small kernel layers. Right: Proposed deep CNN with 9 layers of small kernels. E.g., '64×1 conv, 16, /16' denotes a convolutional layer with 16 kernels, size 64×1, stride 16; a box marked '2×1 pool, /2' means a pooling layer with a size of 2×1, stride 2; 'fc 100' is a fully connected layer with 100 output dimensions. The numbers between boxes are the output sizes of the previous layers.

In our DCNN9, the kernel sizes (F) are 3, the stride (S) and the padding size (P) are set to be 1. According to the formula $S_{out} = (S_{in} - F + 2P)/S + 1$, the size of input ($S_{in}$) and output ($S_{out}$) of a convolutional layer will be the same. Pooling layers are used to reduce the dimension of features. The choice of the number of kernels and layer output size is adopted from the design of WDCNN.

3 Rotating Speed Normalization

To address the cross domain learning problem caused by rotating speed fluctuation, we propose an instance reweighing domain adaption technique for rotating machines. Rotating speed is utilized to preprocess the vibration signals in both training and testing stages. No testing data is required during the training process.

In [18], Stander et al. presented a load demodulation normalization (LDN) for vibration signal to get a more diverged hand-craft feature vector when the load is fluctuating. Original signals are divided by a filtered modulation signal before feature extraction. Similarly, considering the effect of the centrifugal force, we propose a simple rotating speed normalization (RSN) for displacement vibration signals. It is known that centrifugal force $F = mw^2r$ (where m stands for unbalance mass, w is angular speed, and r is the distance of unbalance mass to the rotation centre) of a unbalance mass is proportional to the square of its rotating speed. This indicates that the amplitude of measured vibration should be proportional to the square of the rotating speed. Without rigorous mathematics but simply based on empirical observation, we divide the measured vibration by the square of the rotating speed (w^2) before training and testing the CNN models. The square of the global mean rotating speed is used to control the scale of the vibration amplitude. The RSN process can be written as

$$X_{rsn} = \overline{R_{tr}}^2 X_{raw} \otimes [R]^{-2}$$

where R_{tr} is the rotating speed matrix of training samples, X_{raw} and X_{rsn} are raw and RSN vibration, and R is the rotating speed matrix corresponding to the vibration matrix. We also have $X_{rsn}, X_{raw}, R \in \Re^{n\times l}$, where n stands for the number of samples, and l for the length of each signal sample. The overline above R_{tr} gives its average value, $\otimes$ denotes element-wise matrix multiplication, and $[\cdot]^k$ calculates element-wise k power. To be clear, $\overline{R_{tr}}^2$ can also be any other suitable scalar that preserves the numerical stability of X_{rsn}.

4 Experimental

In this section, an experimental demonstration to the CNN based diagnosis and the validation of our RSN is given. All the data are collected from a MFS-MG test rig of SpectraQuest Inc., shown in Fig. 2. The CNNs codes are implemented with Pytorch and we run them on a PC with Intel i7-7700k CPU and a single Nvidia GTX-1080 GPU.

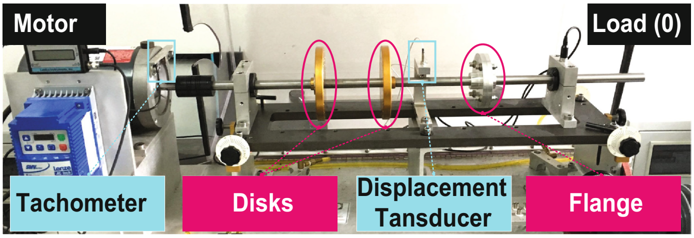

Fig. 2. Experimental test rig with a crack shaft study kit. The vibration signal is obtained by a displacement transducer and the rotating speed is measured via a tachometer (one pulse per revolution).

4.1 Data Collection

This experiment use the flange in Fig. 2 with removable bolts to simulate rotor crack conditions. Data sets SF01 and SF03 are collected respectively when the drive speed curve is (a) and (c) in Fig. 3. The rotating speeds are measured using a tachometer and the vibrations are collected by a displacement transducer marked by blue square frames in Fig. 2. Under each speed curve, four health conditions are considered, including a healthy case and cracked cases with three different levels simulated by removing 1, 2 or 3 bolts on the flange. For each condition, 2390 s of vibration signal and tachometer signal are collected with a sampling frequency of 10240 Hz. Then the vibration signals are sliced as input data of CNNs sized at $2048 \times 1 \times 1$. No data augmentation is used in this experiment. In total, 95600 data samples are measured and each data set has 47800 samples. For both data sets (SF01 and SF03), their samples are then randomly split into training and testing sets with sizes of 40000 and 7800.

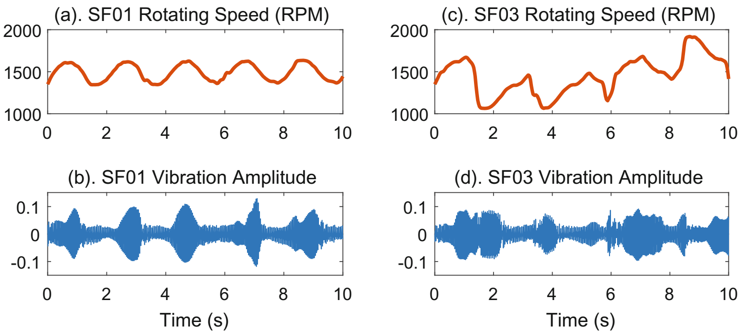

Fig. 3. Examples of collected vibrations and rotating speed curves. (a) and (c): Rotating speed curve and vibration signal of data set SF01. (b) and (d): Rotating speed curve and vibration signal of data set SF03. SF is the abbreviation for "Speed Fluctuation".

4.2 CNN Parameters

Determining hyper-parameters for CNN trainings is an important issue. In this work, ReLU activations are used after every pooling layer and in the fully connected layers. Batch normalization is applied for the input layer and every activation layers. Cross entropy is used as loss function and Adam algorithm with default $\beta = (0.9, 0.999)$ is implemented to train the models. An L2 penalty is appended to the loss function for weight decay and the decay factor is 1e−4. We set the initial learning rate at 1e−4 and then halve it every 5 epoch. Batchsize is 200 and epoch number goes to 20. The number of output classes is 4 for the 4 health conditions in our experiment.

4.3 Results and Discussions

To justify the choice of small kernels and demonstrate the effectiveness of RSN on domain adaption, Table 2 gives the test performance of three CNN models (Ince's model [6], Zhang's WDCNN [23] and our proposed DCNN9) with raw signals and the test performance after utilizing the proposed RSN method. For each model, the whole training and testing procedure is run 20 times with all 40000 training samples and 7800 test samples. Average and variance of testing accuracies are written as average ± variance. Generally, WDCNN (6 layers) and DCNN9 (9 layers) both achieved solid diagnostic performances while Ince's model (3 layers) did not. Testing accuracy approaches 100% when the testing data are drawn from the same data set as training data, but it could drop to 66.71% for DCNN9 under a cross domain learning scenario. In the SF01→SF03 scenario, RSN successfully boosted the domain adaption ability of three CNNs by 11.89% (Ince's, 62.74 to 74.63), 16.21% (WDCNN, 67.10 to 83.31), and 18.45% (DCNN9, 66.71 to 85.16). There are two interesting differences between the results of SF01→SF03 and SF03→SF01. All implemented models showed a better transfer ability when they are trained by SF03 whose rotating speed fluctuating range is wider. A wider range of working conditions produces vibration data featured with richer information of the machine, thus the models can learn more from such vibrations. Besides, in SF01→SF03, the performance can be largely improved by RSN. In SF03→SF01, although RSN fails to boost the mean value, it managed to suppress the variance of the accuracies by 2.52%, 2.8% and 4.47% w.r.t. Ince's model, WDCNN and DCNN9.

The number of training samples versus testing errors of DCNN9 is plotted in Fig. 4. Generally, error rates show downward trends as the number of training samples increases. The solid lines obtained with RSN are lower than corresponding dash lines from raw vibrations. RSN shows its advantage when the number of training samples is relatively small. This is a good evidence that RSN can be used as a regularization of data to guard against the over-fitting phenomenon for our CNN model. Except for SF01→SF03, error rates output from raw and RSN data are approached as the number of training samples increases. SF01→SF03 is a scenario where the models are tested with data that are collected from unseen rotating speed on training stage. A model's diagnostic accuracy can be

Table 2. Performance comparison of models

Method	SF01→SF01	SF01→SF03	SF03→SF01	SF03→SF03
Ince's [6]	90.85 ± 1.13	62.74 ± 1.67	72.35 ± 6.80	77.63 ± 1.14
Zhang's [23]	**99.99** ± **0.00**	67.10 ± 7.01	**91.30** ± **7.78**	**99.91** ± **0.00**
Proposed	99.98 ± 0.00	66.71 ± 5.02	89.76 ± 7.89	99.84 ± 0.01
RSN-Ince's	93.94 ± 0.89	74.63 ± 0.97	79.20 ± 4.28	86.22 ± 0.97
RSN-Zhang's	**99.99** ± **0.00**	83.31 ± 3.72	91.01 ± 4.98	99.89 ± 0.00
RSN-proposed	**99.99** ± **0.00**	**85.16** ± **3.38**	90.22 ± 3.42	99.84 ± 0.00

boosted by collecting more training samples but would still drop dramatically when rotating speed range expands. RSN as a data preprocessing method is effective when dealing with the cross domain learning problem brought along by the rotating speed variation.

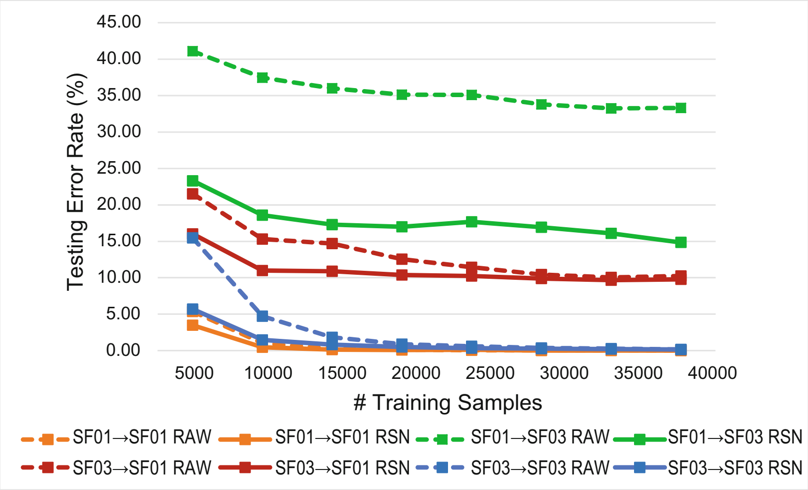

Fig. 4. Testing error trends of DCNN9 for 8 different scenarios as the number of training samples increases.

5 Conclusion

CNN is a powerful deep learning tool that can be utilized to solved many pattern recognition problems including fault diagnosis. However, CNNs have trouble identifying vibration signals collected under unseen working conditions. The proposed RSN is a good compensation for this issue. The main findings of this study can be concluded as follows:

1. The health condition of a rotating machine is successfully diagnosed by trained CNNs using a 0.2 s vibration data sequence even if the rotating speed is fluctuating;
2. It is shown that a wider range of dynamic response of a machine helps CNNs models learn better diagnostic ability;
3. Generalization ability of CNNs can be boosted by feeding them with more data. But a knowledge based signal processing can be very effective when adapting unseen working conditions.

Further investigations to RSN are to carried out to adapt contacted accelerometers that are commonly used for fault diagnosis. Other effective forms of combination of expert diagnosis knowledge and deep learning could be introduced. We also believe that there exists intelligent models who are capable of finding the correlation between vibration data and rotating speed by themselves.

Acknowledgment. This research is supported by National Natural Science Foundation of China (51305067), National Key Research and Development Program of China (2017YF- C0108401, 2016YFB1200401), Natural Sciences and Engineering Research Council of Canada (Grant #RGPIN-2015-04897), and Future Energy Systems research under Canada First Research Excellent Fund (CFREF) (FES-T14-P02). The authors would also like to acknowledge the help of Mr. Peng Chen and Mr. Mian Zhang.

References

1. Chen P, Wang K, Feng K (2016) Application of order-tracking holospectrum to cracked rotor fault diagnostics under nonstationary conditions. In: 2016 Prognostics and system health management conference (PHM-Chengdu)
2. Chen Y, Zaki MJ (2017) KATE: K-competitive autoencoder for text. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, KDD 2017. ACM, pp 85–94
3. Chen Z, Li C, Sanchez RV (2015) Gearbox fault identification and classification with convolutional neural networks. Shock and Vibration
4. Guo X, Chen L, Shen C (2016) Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement 93:490–502
5. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
6. Ince T, Kiranyaz S, Eren L, Askar M, Gabbouj M (2016) Real-time motor fault detection by 1-D convolutional neural networks. IEEE Trans Ind Electron 63(11):7067–7075
7. Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccuer M, Verstockt S, Van de Walle R, Van Hoecke S (2016) Convolutional neural network based fault detection for rotating machinery. J Sound Vib 377:331–345
8. Jia F, Lei Y, Lin J, Zhou X, Lu N (2016) Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Sig Process 72:303–315
9. Jing L, Wang T, Zhao M, Wang P (2017) An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox. Sensors 17(2):414

10. 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
11. Lee J, Nam J (2017) Multi-level and multi-scale feature aggregation using pre-trained convolutional neural networks for music auto-tagging. IEEE Sig Process Lett 24(8):1208–1212
12. Li S, Liu G, Tang X, Lu J, Hu J (2017) An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis. Sensors 17(8):1729
13. Liu Z, Jin Y, Zuo MJ, Feng Z (2017) Time-frequency representation based on robust local mean decomposition for multicomponent AM-FM signal analysis. Mech Syst Sig Process 95:468–487
14. Lu C, Wang Z, Zhou B (2017) Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Adv Eng Informat 32:139–151
15. Lu W, Liang B, Cheng Y, Meng D, Yang J, Zhang T (2017) Deep model based domain adaptation for fault diagnosis. IEEE Trans Ind Electron 64(3):2296–2305
16. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
17. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556
18. Stander CJ, Heyns PS, Schoombie W (2002) Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions. Mech Syst Sig Process 16(6):1005–1024
19. Xia M, Li T, Xu L, Liu L, de Silva CW (2017) Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks. IEEE/ASME Trans Mechatron PP(99):101–110
20. Xie J, Zhang L, Duan L, Wang J (2016) On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on Transfer Component Analysis. In: 2016 IEEE international conference on prognostics and health management (ICPHM)
21. Zhang R, Tao H, Wu L, Guan Y (2017) Transfer learning with neural networks for bearing fault diagnosis in changing working conditions. IEEE Access 5:14347–14357
22. Zhang W, Li C, Peng G, Chen Y, Zhang Z (2018) A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. Mech Syst Sig Process 100:439–453
23. Zhang W, Peng G, Li C, Chen Y, Zhang Z (2017) A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors 17(2):425