Research on Multi-scale Convolutional Neural Networks for Gearbox Fault Diagnosis

ChunZhi Wu, FuZhou Feng, ShouJun Wu, An Wan
PengCheng Jiang
Department of Vehicle Engineering Academy of Army Armored Forces
Beijing, China
e-mail addresses: Chunzhi.wu@qq.com

Chuang Ding
No.32382 PLA
China

Abstract—**For traditional fault diagnosis methods, it is usually necessary to manually extract features and then classify them by pattern recognition. Convolutional neural networks (CNN) can self-learn features from raw data and classify faults. The convolutional neural network is sensitive to the interval points of the training set signals, and the increased interval points will reduce the accuracy of the model for fault identification. Therefore, we propose a multi-scale batch normalized convolutional neural network model (MSBNCNN), with different convolution steps at each scale. The model improves the training speed while reducing the sensitivity to the training set signal interval points, reducing the number of training sets, and making the model diagnosis fault more stable.**

Keywords-component; Deep learning; Multi-scale; Batch normalization; Gearbox; Fault diagnosis

I. INTRODUCTION

Gearboxes are the key components of mechanical transmission systems which are widely used in wind turbines, armored vehicle transmissions and so on. Once the gearbox fails, it may cause production stagnation and even lead to casualties [1]. Therefore, real-time condition monitoring and fault diagnosis of the gearbox is crucial [2].

As an important branch of machine learning, deep learning has been widely used in various fields in recent years with the innovation of computing equipment [3]. The Convolutional Neural Network (CNN) is a typical deep learning method. The biggest advantage is that it automatically learns the features from the original image without having to manually select it. It optimizes the convolution kernel parameters through training to achieve image classification and identification [4][5].

Some scholars have applied CNN to the fault diagnosis of gearboxes and achieved certain results. Anssens et al. [6] and Lu et al. [7] perform a Fourier transform on the acquired vibration signal and use the spectral signal as an input to the CNN network to achieve rotational mechanical condition monitoring. Chen et al. [8] used the characteristic statistical graph of the vibration signal as the input of the CNN to realize the classification diagnosis of the fault. Jing et al. [7] directly used the frequency data of the vibration signal as the input of the CNN and compared it with the three types of data from the original data, the spectrum and the characteristics of the combined time-frequency data. The results show that CNN can adaptively learn features from frequency data and has higher diagnostic accuracy than other comparison methods. Ince et al. [9] used raw vibration data as input to the CNN to detect motor faults. Chen et al. [10] proposed a layered learning rate adaptive deep convolutional neural network based on improved algorithm for diagnosing bearing faults and determining their severity.

One of the major drawbacks of deep learning is that it requires manual selection of parameters, such as learning rate parameters, initialization weight attenuation coefficients, and dropout ratios. Batch normalization for neural network models was proposed in 2015 [11]. The main effect is to increase the learning rate that can be set to remove or drastically reduce dropout without affecting accuracy. One of the great advantages of deep learning is that it can achieve end-to-end fault diagnosis. Therefore, the input of the model is the original vibration signal, and the fault can be diagnosed without other processing. However, when constructing the model training set, there is a certain requirement on the interval points of the signal, that is, the signal interval points cannot be too long. In order to make the training set cover more lengths of signals, more training sets are needed in the case of smaller intervals, which leads to longer training time of the model. Therefore, this paper introduces the concept of multi-scale on the basis of CNN, and proposes a multi-scale batch normalized convolutional neural network algorithm, which can reduce the sensitivity of the training set signal interval and greatly reduce the training time of the model. In this study, we used the fault data of a certain type of tank gearbox as the analysis object.

This paper is organized as follows. Section 2 introduces the convolutional neural network and the improved multi-scale batch normalized convolutional neural network model. Section 3 introduces the fault data of a certain type of tank transmission used in this study. Section 4 compares the proposed method with other methods. The final summary and outlook is in the Section 5.

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

II. MULTISCALE NORMALIZED CONVOLUTIONAL NEURAL NETWORK

A. Convolutional neural network

The CNN network structure consists of an input layer, a hidden layer, a fully connected layer, and an output layer, as shown in Figure 1. The convolutional layer and the pooled layer form a hidden layer. Convolutional layer uses local link and weight sharing. Its mathematical expression can be expressed as follows:

$$z^{l+1} = \boldsymbol{w}^{l} + x^{l} + \boldsymbol{b}^{l} \tag{1}$$

x^{l} is the output of the previous layer and the output is x^{l+1}.

$$x^{l+1} = f\left(z^{l}\right) \tag{2}$$

$\boldsymbol{w}^{l}$ is the convolution kernel of the l layer; $\boldsymbol{b}^{l}$ is the bias; f is the activation function. If the convolution layer is directly behind the classifier, it will become over-fitting because the input dimension is too high. Therefore, the pooling layer needs to be added for dimensionality reduction.

Define the pooling function as $\downarrow Subconv$, then express it as follows:

$$X_{k}^{l+1} = f\left(\boldsymbol{w}^{l+1} \downarrow Subconv(R_{k}) + \boldsymbol{b}^{l+1}\right) \tag{3}$$

$\boldsymbol{w}^{l+1}$ is the weight; $\boldsymbol{b}^{l+1}$ is the offset of the function. The role of the convolutional layer is to take a deeper analysis of each small part of the neural network and abstract it to obtain higher features. The pooling layer is primarily intended to significantly reduce the spatial dimensions of the input convolutional layer and to control overfitting. After passing through the two convolutional layers and the pooling layer, a fully connected layer is connected. Similar to the classical neural networks, fully connected layers can be applied to different classification models.

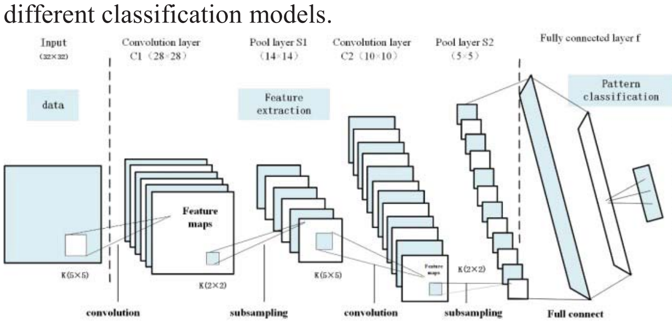


Figure 1. Typical CNN structure

When dealing with a one-dimensional signal such as a vibration signal, changing the convolutional form can result in a one-dimensional CNN model, as shown in Figure 2.

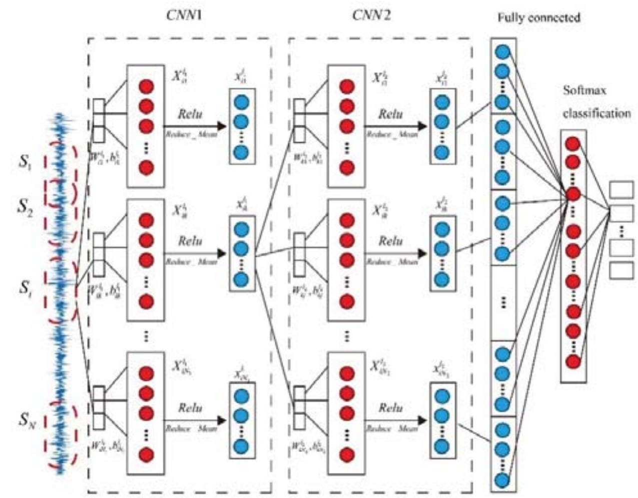


Figure 2. 1-DCNN network structure

B. Proposed method

In order to improve the dimension of feature extraction, the multi-scale concept is introduced into the one-dimensional convolutional neural network model, and a multi-scale convolutional neural network model is proposed, as shown in Figure 3. Among them, $Convd1$, $Convd2$, and $Convd3$ respectively represent the moving steps of the convolution kernel as 1, 2, and 3. Different moving steps can make different scale feature quantities after convolution.

A Batch Normalization (BN) operation is added to the last fully connected portion of each CNN layer group. The BN layer can improve the training speed and accuracy, and ensure the stability of the data after the three sets of vector stitching, effectively avoiding the gradient disappearance and gradient explosion phenomenon.

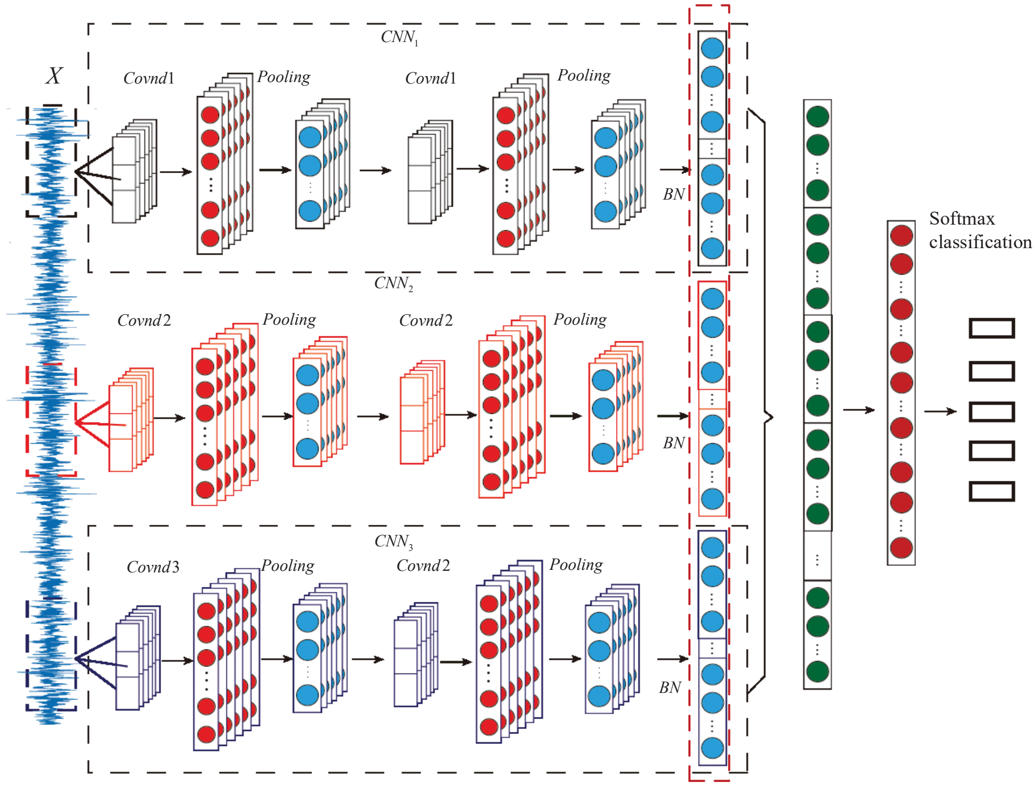
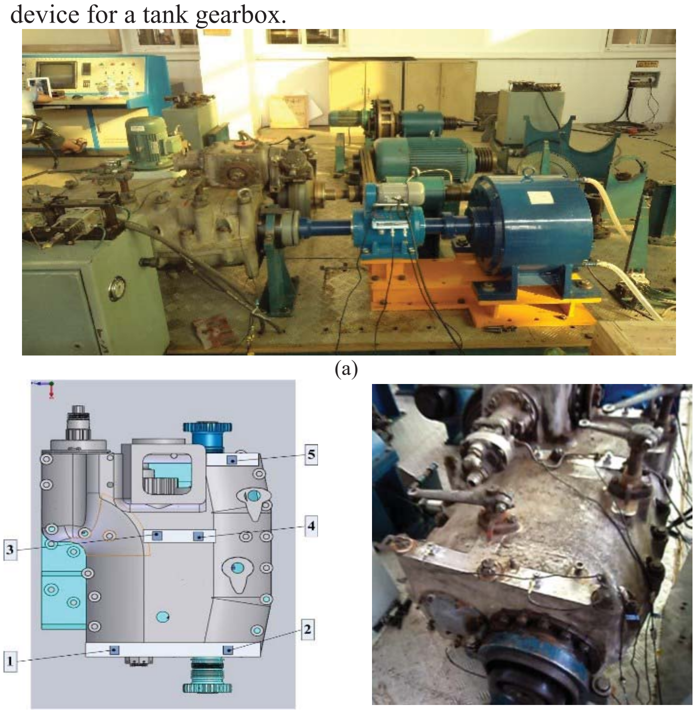


Figure 3. Multi-scale BNCNN structure

III. EXPERIMENT SETUP

Figure 4 (a) is a composite fault simulation experimental device for a tank gearbox.



(a)

(b) (c)

Figure 4. (a) Overview of the experimental platform (b) Schematic diagram of test points (c) Actual location of test points

It consists of a speed control console, a three-phase asynchronous motor, a test platform and an electromagnetic dynamometer. The gearbox gear is an involute spur gear. The experimental device system is powered by a three-phase asynchronous motor, and the output speed of the motor can be adjusted by a speed console. The motor transmits the speed to the input shaft of the gearbox through the belt drive, and the drive shaft of the gearbox is transmitted to the spindle through the intermediate shaft to output the speed. The spindle is connected to an electromagnetic dynamometer, and the electromagnetic dynamometer acts as a load here. Figure 4(b) and Figure 4(c) show the position of the measuring point of the box. The third-order vibration data of measuring point 3 is selected as the experimental data. The measuring condition is the speed of 1000r/min, the load is 150N·m, and the sampling frequency is 20kHz.

The experiment set up four kinds of faults: gear broken gear, gear wear, bearing outer ring crack, bearing rolling body wear and so on. Figure 5 shows the time domain diagram of the vibration signal under several fault conditions and normal operating conditions. As can be seen from the figure, in addition to the gear breakage fault is more obvious, the other states are difficult to distinguish directly from the time domain map.

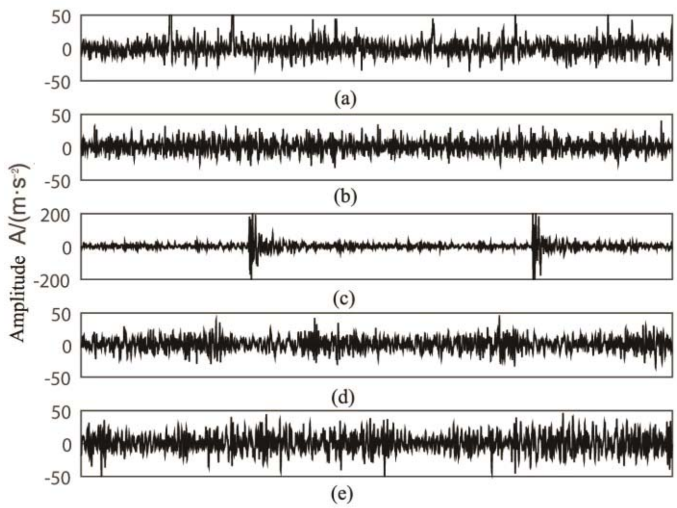


Figure 5. Time domain and frequency domain signal diagrams for different faults (a) normal (b) broken teeth (c) bearing outer ring crack (d) bearing rolling element wear (e) gear wear

When the model is training, the sample signal length is required to be consistent. In this model, the length of the signal intercepted in each segment is 2400 points, which basically contains a full cycle of signals. The interval points of the signal represent the phase information of the sample. The fewer the number of interval points, the more phase information the sample contains. However, in order to ensure the data length covered by the sample and reduce the number of interval points, the corresponding number of samples is increased. The figure below shows how to sample the sample by setting the number of intervals for one test data.

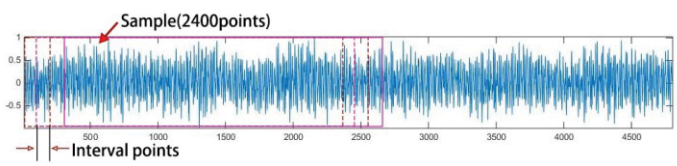


Figure 6. Schematic diagram of the interval points of the sample

IV. Discussion and Analysis

A. *Related parameter settings*

By adjusting the parameters, the number of parameters of the two models before the full connection is similar, which makes the comparison fairer. The specific parameter settings are shown in TABEL I. MSBNCNN is composed of three sets of two-layer CNN structures, so there is less than a single CNN model in the second-layer convolution kernel setting of each group, so that the parameters of the fully-connected layer are consistent.

TABLE I. Modal decomposition algorithm parameter selection and model parameter setting

1DCNN and BNCNN Structural parameters
First layer convolution kernel size:21 No.1 kernels:6 First layer convolution kernel size:11 No.2 kernels:14 Number of three sets of stitching layer features:8960 Learning rate:5e-4
MSBNCNN Structural parameters
First layer convolution kernel size:21 No.1 kernels:6 First layer convolution kernel size:11 No.2 kernels:10 Number of three sets of stitching layer features:8800 Learning rate:5e-4

B. *Case 1: The impact of BN on training*

When the number of set interval points is 30, in order to ensure the coverage length of the training set and the test set, the number of training set groups is 1200 groups and the number of test set groups is 600 groups in each fault state. Taking 2400 points as a whole cycle, the training set and test set cover a total of 23 full cycles. Figure 7 shows the change in 1DCNN test accuracy for 200 training sessions.

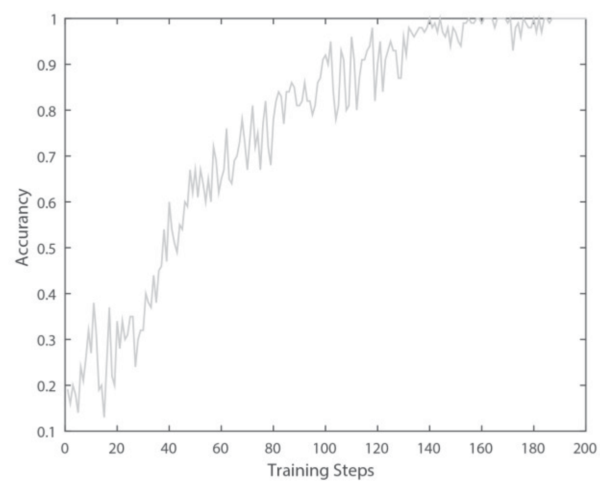


Figure 7. 1DCNN accuracy curve

The solid line portion in Figure 7 is an accuracy curve after smoothing by Smooth method. It can be seen from the figure that after 200 trainings, the accuracy rate is over 99%, but the model training process has large fluctuations and more training times.

After adding the BN module to the 1DCNN, the accuracy of the training with the MSBNCNN model is shown in Figure 8. It can be seen from the figure that after the batch normalization is added, the convergence speed of the model increases without changing the learning rate. In particular, the MSBNCNN model achieved 100% accuracy in about 20 times.

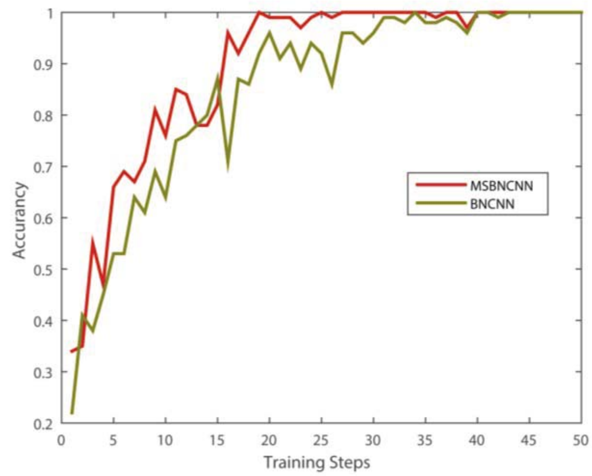


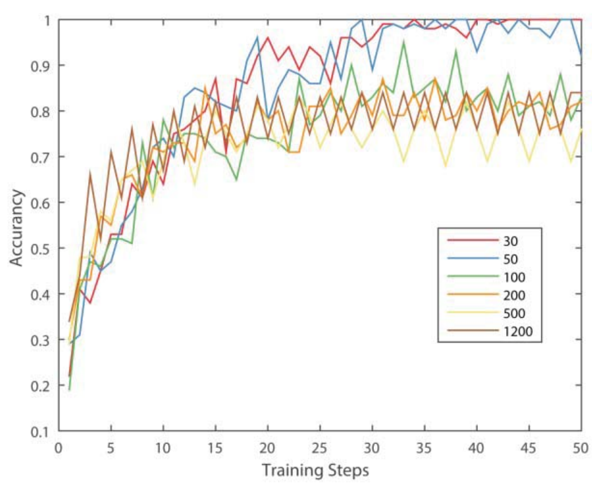
Figure 8. BNCNN and MSBNCNN accuracy curves

C. *Case 2: The effect of different interval points on accuracy*

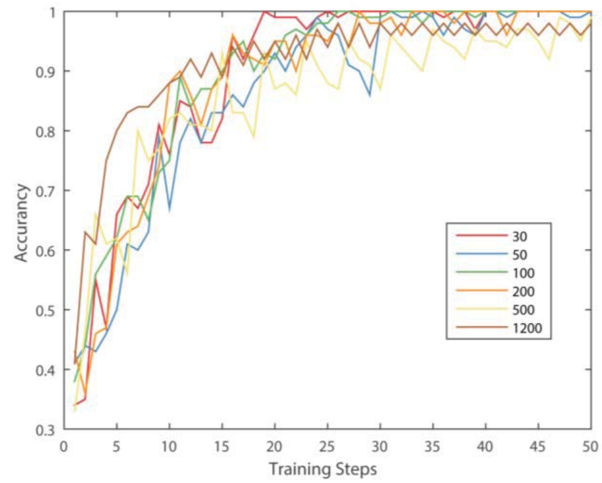
This experiment sets different sample intervals and compares the changes in model training accuracy. The table for setting the relationship between different sample intervals and the number of training sets is as shown in TABEL II, while ensuring that the total length of the entire cycle is constant.

TABLE II. SAMPLE INTERVAL POINTS AND TRAINING SETS NUMBER

Sample interval points	Number of training sets
30	1200
50	720
100	360
200	180
500	72
1200	30



(a)



(b)

Figure 9. BNCNN and MSBNCNN accuracy curves with different interval points (a) BNCNN (b)MSBNCNN

The data set input model of different sample intervals is trained. The training frequency and accuracy curve of BNCNN model are shown in Figure 9(a). When the sample interval exceeds 50 points, the accuracy begins to drop to about 80%. The MSBNCNN accuracy curve is shown in Figure 9(b). The accuracy of the model diagnosis has little effect on the sample interval and remains above 95%.

V. CONCLUTION AND FUTURE WORK

In this paper, a multi-scale batch normalized convolutional neural network fault diagnosis model is proposed. The vibration signal is used as input, and the typical fault diagnosis of the gearbox is completed end-to-end. The experiment compares the impact of increasing the batch standardization module and setting different sampling intervals on the diagnostic accuracy. Although 1DCNN can identify faults well through vibration signals, it has higher requirements on the number of sample interval points in the training set. Small intervals can achieve higher accuracy, but the number of training sets required is large. The more the number of interval points, the lower the recognition accuracy. The multi-scale batch normalized convolutional neural network model proposed in this paper is not affected by the number of sample interval points, and can achieve high fault diagnosis accuracy under different interval points, so the model is robust.

REFERENCES

[1] Li C , Sanchez, René-Vinicio, Zurita G , et al. "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals." Mechanical Systems and Signal Processing , 76 (2016) 283–293.

[2] S. Polak, "Gearbox and gear system problems", in: Gears Conf. London, GEAR TECH., Imeche seminar, 1999, pp. 65–72.

[3] G.E. Hinton, R.R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", Science 313 (5786) (2006) 504-507.

[4] R. Girshick, J. Donahue, T. Darrell, J. Malik., "Rich feature hierarchies for accurate object detection and semantic segmentation", in:

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[5] T. Jonathan, A. Jain, R. Goroshin, Y. Lecun, “Efficient object localization using Convolutional Networks.” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2015.pp. 648-656.

[6] O. Janssensa1, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, et al. “Convolutional Neural Network Based Fault Detection for Rotating Machinery.” Journal of Sound and Vibration (2016), 377:331-345.

[7] 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 Journal of the International Measurement Confederation 111(2017):1-10.

[8] Z. Chen, L. Chuan, S. René-Vinicio, “Gearbox Fault Identification and Classification with Convolutional Neural Networks.” Shock and Vibration 2015(2015):1-10.

[9] I. Turker , S. Kiranyaz, L. Eren, M. Askar, “Real-Time Motor Fault Detection by 1D Convolutional Neural Networks.” IEEE Transactions on Industrial Electronics, 2016, 63(11):7067-7075.

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

[11] I. Sergey, C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift.” International Conference on International Conference on Machine Learning JMLR.org, 2015.