Applied Mathematics, Modeling and Computer Simulation C. Chen (Ed.) © 2022 The authors and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/ATDE220088

# Fault Diagnosis for Aerostat Model Based on Multiscale Feature Fusion Deep Residual Networks

Pingan WANG<sup>a,1</sup>, Haibin SHI<sup>b</sup>, Shouwu HOU<sup>a</sup>, Liangzhi FAN<sup>c</sup> and Haiping ZHU<sup>b</sup> <sup>a</sup>NO.38 Research Institute of CETC, China <sup>b</sup>School of Mechanical Science and Engineering, Huazhong University of Science and Technology, China

<sup>c</sup> School of Mechanical engineering and automation, Wuhan Textile University, China

Abstract. Deep learning has been widely used in fault diagnosis, especially the convolutional neural network (CNN). However, traditional CNN models usually use single-scale kernels to extract features, which ignores the multiscale features of input data. This article develops a novel residual neural network named multiscale feature fusion deep residual networks for fault diagnosis of aerostat. The designed multiscale feature fusion block (MFF Block) realizes automatic extraction, fusion and compression of multiscale features. The series connection of multiple MFF Blocks makes the proposed model able to extract deeper and wider features from raw signal segments. Then a multiscale pooling layer is developed to extract the most effective features for accurate fault diagnosis. The proposed model is evaluated on an exclusive aerostat strain signal dataset. The comparison results illustrate that the proposed model achieves superior diagnostic performance than several popular models.

Keywords. Deep learning; fault diagnosis; feature fusion; residual learning; convolutional neural network.

# 1. Introduction

In practical engineering applications, the failure of equipment is difficult to be found in time, which is a serious threat to the safety of industrial production [1]. Because of the complex work environment and long-running time, unexpected faults occur frequently, which might cause a tremendous amount of loss and great maintenance cost. Thus, it is necessary to implement the in-time and precise fault diagnosis for industrial equipment.

In traditional fault diagnosis, the consensus is to get valuable information through signal processing methods such as filtering and demodulation. Marco *et al.* [2] applied the short-time Fourier transform (STFT) in the feature extraction for damage detection and use the sum of STFT coefficients as the damage index. However, such methods have quite high requirements for professional knowledge and signal processing experience.

<sup>&</sup>lt;sup>1</sup> Corresponding Author, Pingan WANG, NO.38 Research Institute of CETC, China; E-mail: 9764777@qq.com.

With the rapid exaltation of computer performance and the booming of industrial big data, data-driven fault diagnostic methods have been widely studied [3]. Among these methods, machine learning prevails because it is based on sensor data rather than particular signal processing models or signal patterns. For example, Yuwono *et al.* [4] combined particle swarm clustering algorithm and hidden Markov model, and achieved remarkable diagnostic performance on rolling bearings. However, machine learning based models rely heavily on manual feature extraction methods.

Deep learning [5] has become popular because of its ability in mining and capturing deep level information from high dimensional data, which can get rid of the manual feature extraction. Deep learning based methods such as deep neural networks (DNN) [6], convolutional neural networks (CNN) [7-8] and deep auto-encoder [9] have shown their prominent feature extracting capabilities. As deep learning attracts more attention, a lot of excellent models have been applied in fault diagnosis. For example, Zhu *et al.* [10] proposed a model named stacked pruning sparse denoising autoencoder, which can reduce information loss during training iterations and achieve higher diagnostic accuracy. However, as networks get deeper, the degradation problem [11] occurs. If the deep networks are just simply stacked layer by layer, the training loss will stop declining because the identity mapping is hard to be represented by traditional deep networks.

To address the aforementioned problems, a novel multiscale feature fusion deep neural networks (MFF-DRN) is proposed for the fault diagnosis. In the proposed model, the designed multiscale feature fusion block (MFF Block) realizes automatic extraction, fusion and compression of multiscale features. Then, the series connection of multiple MFF Blocks makes the proposed model able to extract deeper and wider features from raw signal segments. Finally, the multiscale pooling layer extracts the most effective features for accurate fault diagnosis.

The main contributions of this paper are concluded as follows.

(1)An end-to-end model derived from CNN is proposed for industrial equipment fault diagnosis, which could efficiently extract features from one-dimensional raw signal segments without manually extracting features or signal processing.

(2)A novel multiscale feature fusion block (MFF Block) is designed, which is capable of automatically extracting, fusing and compressing multiscale features. This structure realizes efficient multiscale feature extraction with fewer filter channels. Additionally, a new multiscale pooling method is proposed to broaden the model's receptive field. Meanwhile, residual learning is adopted to heighten the model's learning ability.

(3)The proposed MFF-DRN is evaluated on an aerostat strain signal dataset. A series of multifaceted comparisons with some popular models powerfully manifest the superiority of MFF-DRN.

The remainder of this paper is organized as follows. Section 2 introduces the proposed model in detail. Section 3 illustrates the datasets and experimental validations. Finally, section 4 concludes this study.

## 2.1. Residual Learning Module

The one-dimensional convolutional neural network (1-D CNN) is derived from feedforward networks. The strategies of local connection and weight sharing can greatly reduce the parameters of CNN models. A classical CNN mainly consists of convolutional layers and pooling layers.

To fetch information from complex data, convolutional neural networks are becoming deeper and deeper. However, there lies a degradation problem in a deep neural network: As the network becomes deeper, its training accuracy saturates or even declines [11]. To address the aforementioned problem, residual learning was proposed by *He et al.* [11], which can be defined as follows:

$$H(X) = F(X) + X \tag{1}$$

Where H represents the output function, F represents the residual function and X denotes the input data.



Figure 1. Residual Learning Block

Figure 2. Structure of Multiscale Pooling Block

As shown in Figure 1, a residual block (ResBlock) that mainly consists of two convolutional blocks (ConvBlocks) is built for feature extraction. The ConvBlock consists of a convolutional layer, a batch normalization layer (BatchNorm) and a ReLU activation layer. The batch normalization layer can address the internal covariate shift problem during training iterations by normalizing the input data [12].

#### 2.2. Proposed MFF-DRN Architecture

The MFF-DRN takes signal fragments as input, which consists of an initial ConvBlock, three MFF Blocks, a multiscale pooling layer formed of three multiscale pooling blocks (MSPs) and a classification layer using softmax function.

Each MFF Block consists of three ConvBlocks, three ResBlocks mentioned in Figure 1 and a bottleneck layer whose kernel size is  $1 \times 1$  ( $1 \times 1$  Conv). The only difference among these ConvBlocks is that their kernel sizes are different, and so are ResBlocks. To get multiscale features, feature maps that output from different Resblocks are connected through the concatenation operation. The important function of  $1 \times 1$  Conv is compressing the feature maps, which can reduce feature map channels without information loss. These components make the MFF Block able to extract and fuse

multiscale features so can improve the mapping ability and classification accuracy of MFF-DRN. Besides, channels of feature maps generated in the proposed MFF-DRN are much fewer than in other CNN based models during training, because the application of MFF Blocks reduces redundant features.

Then a multiscale pooling layer is constructed to extract the most valuable information from features output by the MFF Block. As shown in Figure 2, each MSP in the multiscale pooling layer consists of a bottleneck layer and an average pooling layer. In each MSP, the channels of bottleneck layers and pooling kernel size are set to be different to extract valuable features at different scales.

# 3. Model Evaluation

To evaluate the effectiveness and the universality of the MFF-DRN, a case on strain signal dataset gathered from the surface of an airship model is studied.

Besides accuracy, the macro F1-score F is used as the other evaluation index to better evaluate the model performance. Whose definition is as follows:

$$p_{i} = \frac{TP_{i}}{TP_{i} + FP_{i}}$$

$$r_{i} = \frac{TP_{i}}{TP_{i} + FN_{i}}$$

$$f_{i} = 2 \cdot \frac{p_{i} \cdot r_{i}}{p_{i} + r_{i}}$$

$$F = \frac{1}{n} \cdot \sum_{i=1}^{n} f_{i}$$
(2)

Where  $TP_i$  denotes the number of samples that are correctly classified as class *i*,  $FP_i$  denotes the number of samples that are falsely classified as class *i* and  $FN_i$  represents the number of samples that actually belong to class *i* but are classified as others.  $p_i$ ,  $r_i$ ,  $f_i$  denotes the precision, recall and micro F1-score of class *i*, respectively. Therefore, macro F1-score *F* is a comprehensive index for classification.

All models in the case studies are implemented by Python 3.7 with Pytorch 1.7 and run on Windows 10 with a NVIDIA GTX 1070 GPU.

#### 3.1. Data Description

In the case, strain signals collected from the surface of the airship model are used to verify the validity and generalization of the model. The dataset is derived from the vibration and displacement signals of the airship surface excited by sound, which are collected by strain sensor. The experimental device is shown in Figure 3, The surface of the airship model is made of flexible composite material, which is composed of a load-bearing layer and a sealing layer.



Figure 3. Overview of strain signal acquisition devices for airship model

During each signal acquisition, an air compressor was used to maintain the airship pressure at 5kPa to simulate the actual working condition of the airship. A speaker close to the airship surface was used to play 200Hz fixed frequency sound signals to stimulate the vibration of the model. The speaker power is 20W. The carrier frequency of the sound is 44kHz, and the sensor acquisition frequency is 2kHz. As shown in Figure 4, the 5mm transverse slit defect (F1) and cross defect (F2) are constructed at the mark in the figure, which are too small to be seen with the eye. The distance from the sensor to the defect is within 10cm. The signal acquisition experiments in various states have been repeated 10 times, the signal collection duration is 50s in the healthy state and 20s in the defective state, so as to construct an unbalanced airship surface strain dataset.



Figure 4. Relative location of defects and sensors

The entire preprocessing procedure is illustrated in Figure 5. The signals are split into isometric signal segments to get more samples for model training. To avoid noise and disturbance at the beginning and end of the raw signals, the first one-tenth and last one-tenth of the original signal are cut off. The length of signal segments is set to 5120, and there is no overlap between adjacent signal segments.



Figure 5. Data Preprocess Schematic Diagram

Then, the min-max normalization process is applied on each signal segment to accelerate the model training, which is defined as follows:

$$\tilde{X} = \frac{X - \min(X)}{\max(X) - \min(X)}$$
(3)

Where X denotes the signal segment, max(X) denotes the biggest value in X and min(X) denotes the smallest value in X.

Finally, a dataset including 10,944 signal segments is constructed for fault diagnosis. With the samples shuffled, 80% of these segments are selected for training set and the remaining 20% are selected for testing set.

# 3.2. Results Comparison and Analysis

In this subsection, the DNN, CNN and DRN-3 (a deep residual network whose kernel size is 3), DRN-7 and DRN-11 are used to be compared with the MFF-DRN. Settings for the training of all models are as follows: the number of epochs is set to 40, the learning rate is 0.001; the Adam [13] optimization algorithm is used to update models' parameters; the batch size is set to 16; the L2 Regularization is used in convolutional layers to prevent overfitting [5], its weight is set to 0.0001. All models are tested on the same testing set and the testing set will be re-selected after an experiment. The results are presented in Figure 6 and Table 1.



Figure 6. Diagnostic Accuracy Comparison of Five Independent Experiments

Model	Max Acc	Min Acc	Mean Acc	SD	Mean F
DNN	80.27	77.66	79.01	1.08	73.89
CNN	85.29	83.69	84.58	0.64	81.22
DRN-3	83.01	79.12	80.91	1.75	74.31
DRN-7	90.09	86.80	88.45	1.28	85.46
DRN-11	91.05	88.35	89.50	0.98	84.37
MFF-DRN	92.05	90.27	91.17	0.63	88.13

Table 1. Experimental results of MFF-DRN and other models

In Figure 6, The accuracy of MFF-DRN was the highest in each experiment, and its accuracy was the most stable among all the models. TABLE 1 shows the numerical results. The mean accuracy and mean macro F1-score of DRN-7 and DRN-11 are higher than DRN-3. The standard deviation of CNN is much smaller than other DRNs. The performance of MFF-DRN is the best for all indicators. The experimental results strongly prove the superiority of MFF-DRN in the classification of surface defects of the aerostat.

# 4. Conclusion

A novel MFF-DRN model is presented, which directly extracts features from signal segments. Firstly, the convolutional and residual learning blocks of different scales are developed. Then, the channel concatenation operation is used to connect different scale features and the application of bottleneck layers are used to fuse and compress features. Additionally, a multiscale pooling layer composed of bottleneck layers and average pooling layers is developed to further compress features. Deep learning tricks such as the batch normalization method, the Adam optimization algorithm and the L2 regularization are used to improve the diagnostic performance. The aerostat strain signal dataset is used for model validation. Popular models including DNN, CNN and DRNs are applied for comparison. The MFF-DRN outperforms these models on the dataset, which shows the excellent fault classification ability of MFF-DRN on the unbalanced dataset that based on real defects.

In the future, the denoising components will be developed and embedded in the model for fault diagnosis on heavily noisy signals. Furthermore, model distillation and pruning operation will be studied to simplify the structure and accelerate model learning.

## 5. Acknowledgments

The authors gratefully acknowledge the funding support by the National Natural Science Foundation of China (Grant No. 52075202).

## References

- R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," Mechanical Systems and Signal Processing, vol. 108, pp. 33-47, 2018.
- [2] M. Cocconcelli, R. Zimroz, R. Rubini, and W. Bartelmus, "STFT based approach for ball bearing fault detection in a varying speed motor," in Condition Monitoring of Machinery in Non-Stationary Operations: Springer, 2012, pp. 41-50.
- [3] K. Tidriri, N. Chatti, S. Verron, and T. Tiplica, "Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges," Annual Reviews in Control, vol. 42, pp. 63-81, 2016.

- [4] M. Yuwono, Y. Qin, J. Zhou, Y. Guo, B. G. Celler, and S. W. Su, "Automatic bearing fault diagnosis using particle swarm clustering and Hidden Markov Model," Engineering Applications of Artificial Intelligence, vol. 47, pp. 88-100, 2016.
- [5] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning (no. 2). MIT press Cambridge, 2016.
- [6] G. E. Dahl, Y. Dong, D. Li, and A. Acero, "Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 1, pp. 30-42, 2012.
- [7] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks," IEEE/ASME Transactions on Mechatronics, vol. 23, no. 1, pp. 101-110, 2018.
- [8] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks," IEEE Transactions on Industrial Electronics, vol. 63, no. 11, pp. 7067-7075, 2016.
- [9] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," Mechanical Systems and Signal Processing, vol. 95, pp. 187-204, 2017.
- [10] H. Zhu, J. Cheng, C. Zhang, J. Wu, and X. Shao, "Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings," Applied Soft Computing, vol. 88, 2020.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
- [12] S. Ioffe and C. J. a. p. a. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015.
- [13] D. P. Kingma and J. J. a. p. a. Ba, "Adam: A method for stochastic optimization," 2014.