Multiscale Dense Convolutional Networks for Intelligent Fault Diagnosis of Rolling Bearing

Yadong Xu and Beibei Sun*
School of Mechanical Engineering, Southeast University
Nanjing, Jiangsu Province 211189, China
ydxu@seu.edu.cn, bbsun@seu.edu.cn

***Abstract*—With the continuous development of deep learning, convolutional neural network (CNN) has been widely applied to the field of fault diagnosis. However, most of the previous methods rely on complex signal processing knowledge or feature extraction methods, and thus cannot achieve end-to-end fault diagnosis. In this paper, we put forward a multiscale dense convolutional network (MSDCN) for the fault identification of rolling bearing. First of all, a modified coarse-graining procedure is introduced to incorporate multiscale learning ability into CNN model. Then a novel dense convolutional neural network architecture (DCNN) is designed for the feature extraction of the mechanical vibration signals. Finally, an end-to-end fault diagnosis framework which is based on the improved coarse-grained process and the designed DCNN, is presented. The bearing data collected from the fault simulator is used to verify the effectiveness of the proposed method. Experiments show that the proposed approach outperforms some competitive methods in terms of diagnostic accuracy.**

Keywords: Convolutional neural network (CNN); fault diagnosis; multiscale dense convolutional network (MSDCN); mechanical vibration signals.

I. Introduction

As an important part mechanical equipment, the failure of rolling element bearings during operation will lead to high maintenance costs [1]. Therefore, it is of great significance to use the one-dimensional vibration signal collected by the sensor to carry out real-time status monitoring and fault diagnosis of the rolling element bearings.

In recent years, with the continuous development of machine learning research, data-driven fault diagnosis methods have gradually become mainstream applications in the field of fault diagnosis [2]. Traditional intelligent fault diagnosis methods usually consist of three steps including feature extraction, feature selection and fault classification. Feature extraction is a key step in fault diagnosis [3]. The frequently used methods include wavelet transform (WT), spectral analysis (SA), empirical mode decomposition (EMD), Fourier transform (FFT) and so on [4]. Feature selection can eliminate low-sensitivity, cross-correlation, and useless features from the extracted features, thereby reducing the dimension of features. Popular feature extraction approaches include principal component analysis (PCA) and independent component analysis (ICA) etc. [5]. Finally, fault classification inputs selected features into the fault classifier, and fault classification results can be obtained through iterative training of the classifier. Backpropagation neural networks (BPNN), support vector machine (SVM) and K-nearest neighbor method (KNN) are the most widely used classifiers [6].

The above data-driven methods have largely promoted the development of fault diagnosis. However, they also reveal some shortcomings [7]. On the one hand, feature selection often relies on the experience and expertise of engineers, which brings subjectivity and blindness to diagnostic work. In addition, fault features extracted manually are easily interfered by noise, and features reflecting weak faults are easily concealed by noise. On the other hand, the features extracted by such methods are mainly used for specific fault diagnosis, which leads to their poor generalization ability, making it difficult to apply them in industrial practice.

In 2006, Hinton *et al.* [8] proposed the concept of deep learning (DL). DL applies a deep neural network structure to extract features from input sample data layer by layer, and learns the nonlinear relationship between data and labels. It is able to extract feature information intelligently, which gets rid of the shortcomings of traditional methods that require manual feature extraction and expert experience. Convolutional neural network (CNN) is one of the important branches of deep learning [9]. It has powerful feature extraction capabilities and has been widely used in image recognition and natural language processing [10], [21], [22]. In recent years, some scholars have applied CNN to the field of fault diagnosis.

Zhang *et al.* proposed a domain-adaptive convolutional neural network for the fault identification of rolling bearing operating under different working conditions [11]. A coupled dense connected CNN model for fault recognition of planetary gearbox was put forward by Jiao *et al.* [12]. Peng *et al.* [13] proposed a multi-branch and multi-scale convolutional neural network for the fault identification of

* Corresponding author

2767-6684/21/$31.00 ©2021 IEEE
DOI 10.1109/DSA52907.2021.00020

wheelset bearing. Jiang *et al.* [14] proposed a new multi-scale CNN model for the fault diagnosis of gearbox. These existing CNN models have greatly enriched the means of fault diagnosis.

The existing approaches are expected to extract rich feature information from the raw vibration signals. Nevertheless, mechanical equipment usually runs under various operating conditions [23]. This brings huge challenges to feature extraction and fault diagnosis tasks. This paper focuses on building an end-to-end network framework for intelligent fault diagnosis. The main contributions are given as follows:

[1] A modified coarse-graining procedure is introduced to incorporate multiscale feature extraction capability into the traditional CNN models.

[2] A novel dense connection convolutional neural network (DCNN) structure is designed to extract abundant feature information from mechanical signals.

[3] An end-to-end fault diagnosis framework is proposed based on the designed DCNN model.

[4] The proposed method outperforms some competitive methods in terms of the overall diagnostic accuracy.

The rest of this paper is organized as follows. Section II will elaborate the proposed method. An experiment will be carried out to verify the effectiveness of the proposed approach in Section III. The conclusion will be given in Section IV.

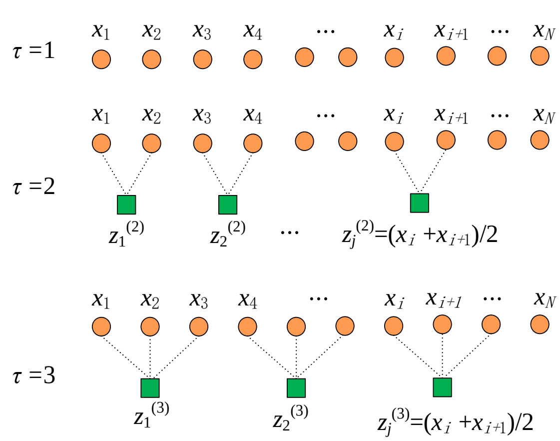

Figure 1. Flowchart of the coarse-graining procedure

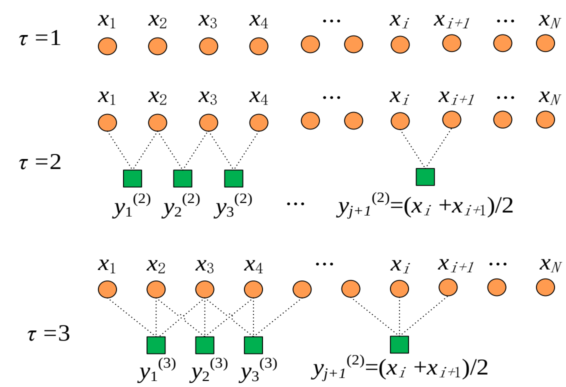

Figure 2. Flowchart of the modified coarse-graining procedure

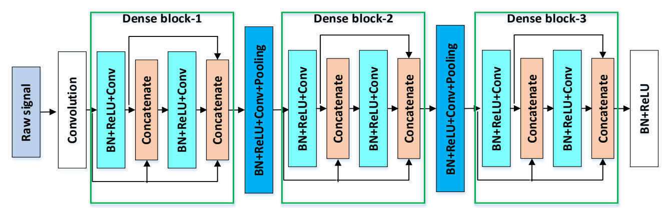

Figure 3. Architecture of dense connection convolutional network proposed in this paper

II. Proposed Method

In this Section, a new CNN model named multiscale dense convolutional network (MSDCN) will be presented in detail.

A. Modified Coarse-graining Procedure

The purpose of introducing the coarse-graining procedure is to incorporate multiscale learning ability into traditional CNN models. The flowchart of the traditional coarse-graining procedure is shown in Fig. 1. Given a time series $x_i,\ 1 \le i \le N$, the coarse-grained time series can be obtained by the following equation:

$$z_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)+1}^{j\tau} x_i,\ 1 \le j \le Round\left(\frac{N}{\tau}\right), \tag{1}$$

where τ denotes the scale factor, and $\tau \in \{1,2,\cdots\}$; $Round(\cdot)$ represents the rounding operation. When τ =1, the obtained coarse-grained time series is actually the raw signal. When τ>1, the coarse-graining operation can be regarded as the combination of average operation and down sampling operation [15].

The traditional coarse-graining procedure is effective for the multiscale analysis of vibration signal. However, it also has some shortcomings [16]. First of all, the traditional coarse-graining procedure does not have continuous shift operation, which leads to the loss of feature information of the raw signal. In addition, the length of the obtained coarse-grained time series is shorter than the original signal, which is not conducive to subsequent analysis. To deal with this problem, a modified coarse-graining procedure is introduced. The flowchart of the modified coarse-graining procedure is displayed in Fig. 2. Given a time series x_i, $1 \le i \le N$, the coarse-grained time series can be obtained by the following equation:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=j}^{j+\tau-1} x_i,\ 1 \le j \le N-\tau+1, \quad (2)$$

where τ denotes the scale factor, and $\tau \in \{1,2,\cdots\}$. The modified coarse-graining procedure can overcome the shortcomings of traditional coarse-graining procedure. In addition, the modified coarse-graining procedure is easy to implement in comparison with some complex multiscale transformation.

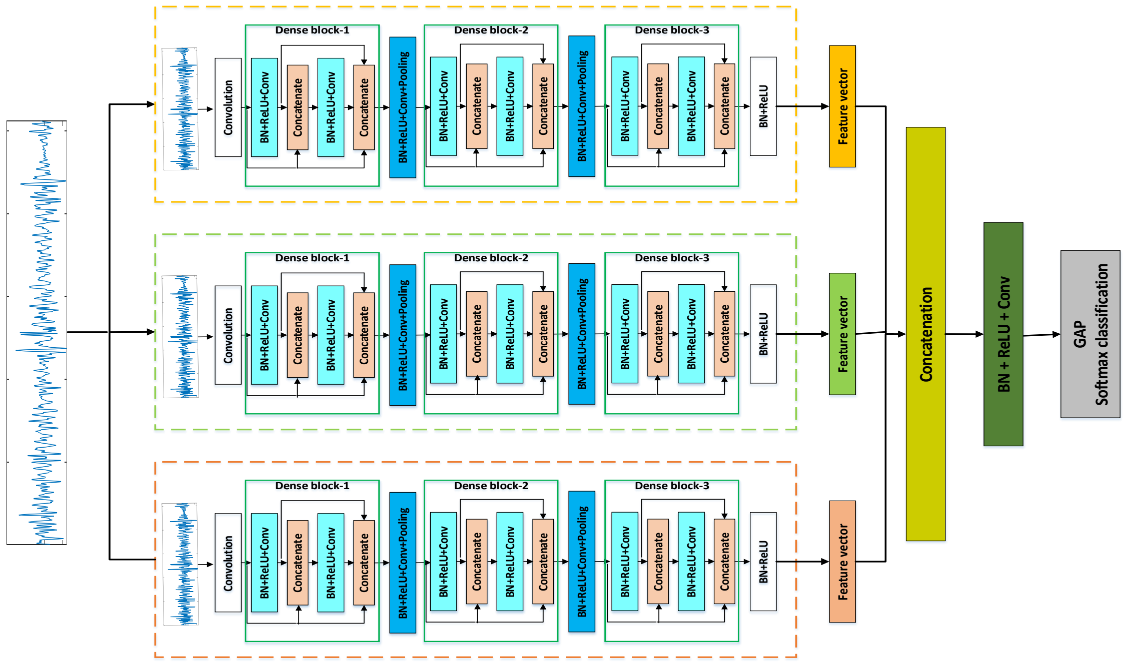

Figure 4. Framework of the proposed approach

TABLE I. DETAILED CONFIGURATION OF THE DESIGNED MSDCN

Lanyers	Output Size	Parameters
Input	512×1	/
Convolution	8-[512×1]	9×1conv, stride 1
Dense Block-1	16-[512×1] 40-[512×1]	9×1conv, stride 1
Transition Block-1	40-[256×1]	1×1conv, stride 1 2×1pool, stride 2
Dense Block-2	56-[256×1] 72-[256×1]	9×1conv, stride 1
Transition Block-2	72-[128×1]	1×1conv, stride 1 2×1pool, stride 2
Dense Block-3	88-[128×1] 104-[128×1]	9×1conv, stride 1
Concatenate	312-[128×1]	/
Convolution	312-[128×1]	1×1conv, stride 1
Global Ave-pool	312 ×1	/
Classification layer	7 ×1	/

B. Dense Connection Learning

Dense connection learning is a recently proposed structure that aims to improve the feature extraction learning ability of traditional CNN [24]. As is described in previous studies, the dense connection learning block is a state-of-the-art learning structure, which is composed of dense layers and transition layers. The detailed description can be found in [12], [17], and [18]. In this paper, we propose a new dense connection convolutional neural network structure (DCNN). As can be seen from Fig. 3, the proposed DCNN includes three dense connection blocks in its architecture. Each dense block has two convolutional layers; the kernels of 9 × 1 with stride 1 are used in all convolutional layers. The growth rate of dense block is set to 16. The designed DCNN structure is expected to extract rich feature information from the input signals.

C. Framework of the Proposed Approach

In this Section, we will put forward a multiscale dense convolutional network (MSDCN) for the fault identification of rolling bearing. The framework of the proposed MSDCN is graphically displayed in Fig. 4. First of all, the modified coarse-graining procedure is carried out to get the multiscale representations of the input signal. Then the obtained coarse-grained signals are fed into three DCNN connected in a parallel manners. The three DCNN subnets share the same architecture as displayed in Table I. Let the extracted feature vectors by the three subnets denoted as C_1, C_2 and C_3, respectively. Then the extracted feature vectors are concatenated:

$$C = [C_1, C_2, C_3], \tag{3}$$

where C denotes the obtained feature vector; [·] represents concatenation operator. Furthermore, a 1×1 convolutional layer is used to reduce the aliasing effect of the feature vector C. Finally, we input the feature C into a global average pool layer and a softmax layer, and then we can get the health condition of the monitored machine. The detailed structure information of proposed MSDCN is shown in Table I. The proposed framework can achieve end-to-end fault classification tasks, which means that no prior knowledge is required in the fault diagnosis process.

TABLE II. EXPERIMENTAL RESULTS OF MSDCN AND DCNN

Model	Accuracy	F_1 score
MSDCN	99.25%	99.25%
DCNN	98.61%	98.60%

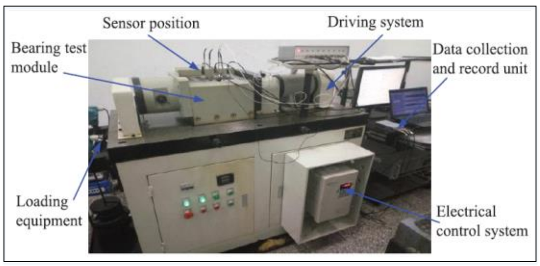

Figure 5. Motor-bearing system used in the experiment.

III. EXPERIMENTAL VERIFICATION

In this Section, we will implement an experiment to validate the effectiveness of MSDCN.

A. Experiment Setting and Data Description

All the algorithms involved in this research are running on the python 3.7 platform. The back propagation algorithm is applied to learn the parameters of the proposed model. We use Adam algorithm with learning rate 0.005 to accelerate the training of MSDCN. The batch size and training epoch are set to 200 and 50, respectively.

As is graphically displayed in Fig. 5, the rolling bearing data is collected from a motor-bearing system. The rolling bearing HRB 6205 is selected as the study object. There are seven health conditions for the bearing, i.e., Healthy status (H), ball fault of the bearing (F1), Inner ring fault of the bearing (F2), Outer ring fault of the bearing (F3), Outer ring and ball failure of the bearing (F4), Inner ring and ball failure of the bearing (F5), Inner outer ring and ball failure of the bearing (F6). For each condition, there are 1300 samples for training procedure and 400 samples for testing procedure. There are 512 points in each sample. Since there are 7 health conditions for the bearing, the fault diagnosis task can be regarded as a 7 classification problems.

B. Effectiveness of the Modified Coarse-graining Procedure

First of all, we will discuss the effectiveness of the modified coarse-graining procedure. The DCNN model is used to compare with the proposed MSDCN. The architecture of the DCNN model is the same as that of MSDCN's subnetwork. The experimental results of the two methods are displayed in Table II. It can be seen that MSDCN outperforms DCNN in terms of accuracy and F_1 score. It is worth pointing out that the accuracy obtained by DCNN is as high as 98.61%, which verifies the effectiveness of the designed DCNN architecture. In addition, the validation cures of the two models are also displayed in Fig. 6. It can be seen that the MSDCN model converges significantly faster than DCNN model, and its accuracy is higher.

TABLE III. EXPERIMENTAL RESULTS OF THE FOUR METHODS

Model	Accuracy	F_1 score
CNN	97.18%	97.17%
WDCNN	97.68%	97.68%
MSCNN	97.36%	97.36%
MSDCN	99.25%	99.25%

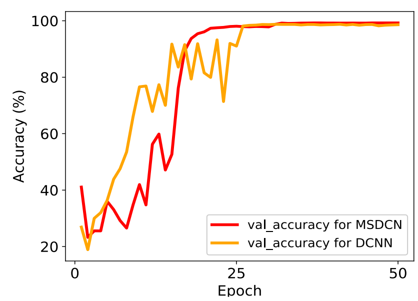

Figure 6. Testing cures of MSDCN and DCNN.

C. Compared with Competitive Methods

The proposed MSDCN are compared with some competitive methods, namely, a five layer CNN in [20], WDCNN [19], and MSCNN [14]. The experimental settings of these methods are same. The experimental results of the four methods are shown in Table III. Compared with CNN, WDCNN and MSCNN, the accuracy of MSDCN is increased by 2.07%, 1.57% and 1.89%, respectively. The confusion matrix of the results is also given in Fig. 7. It can be seen that the proposed method can obtain the highest accuracy rate in each category, except for healthy samples, which implies that the proposed MSDCN has achieved satisfactory results. In addition, the F_1 score of these methods is also listed in Table III. F_1 score is completely consistent with the accuracy result, which further verifies the previous analysis.

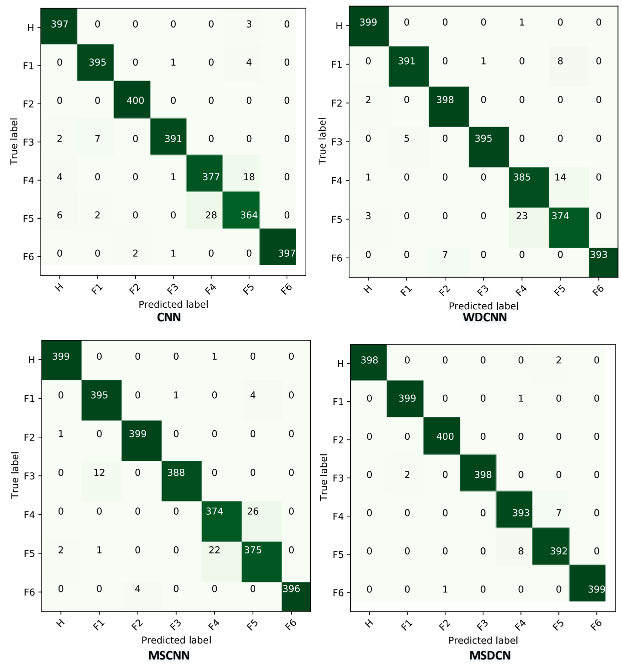

Figure 7. Confusion matrix of the four models.

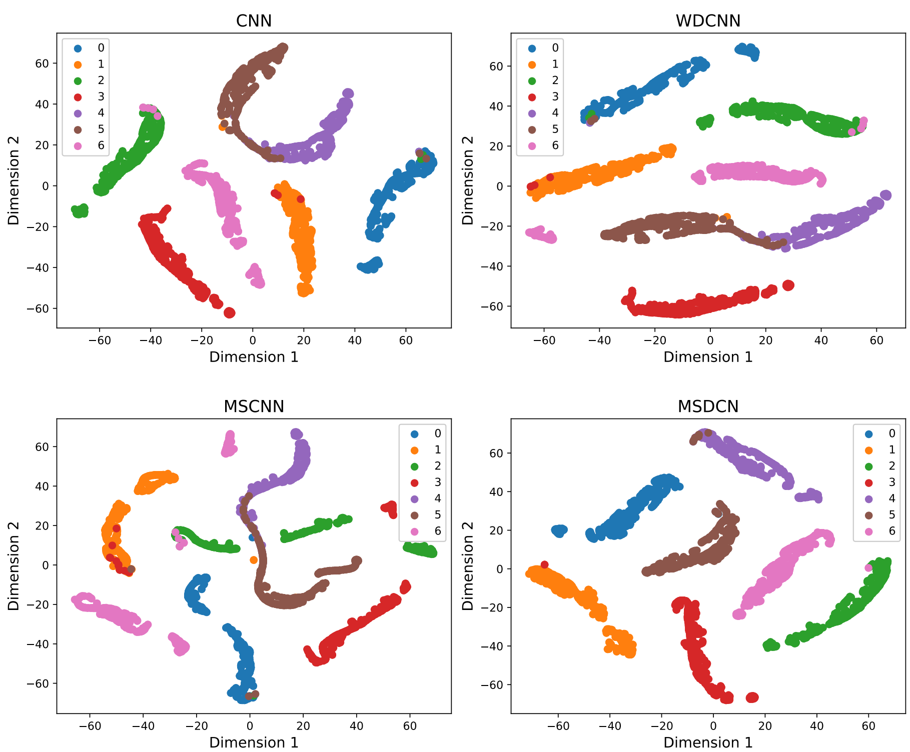

Figure 8. Feature visualization via t-SNE.

In addition to the above analysis on accuracy and F_1 score, we also conduct visualization analysis to give readers a more intuitive understanding of MSDCN. So t-distributed stochastic neighbor embedding (t-SNE) [25] method is introduced to give a more intuitive result. The results by the four methods are shown in Fig. 8. It can be seen that, compared with CNN, WDCNN and MSCNN, the proposed method can better aggregate samples of the same type of label, and at the same time better distinguish samples of different labels, which shows the superiority of MSDCN. In summary, MSDCN can get a more satisfactory result than the other three competitive methods.

IV. Conclusion

In this paper, a novel MSDCN is developed and an end-to-end framework based on MSDCN is presented for the fault identification of rolling element bearing. To begin with, we introduce a modified coarse-graining procedure to incorporate multiscale learning ability into CNN model. Then a novel dense connection architecture (DCNN) is designed for the feature extraction of the mechanical vibration signals. Finally, an end-to-end fault diagnosis framework, which is based on improved coarse-grained process and the designed DCNN, is presented. The bearing data collected by the motor-bearing system is used to verify the effectiveness of the proposed method. Experimental results show that, compared with CNN, WDCNN and MSCNN, the accuracy of MSDCN is increased by 2.07%, 1.57% and 1.89%, respectively.

The proposed framework can achieve end-to-end fault classification tasks, which means that no prior knowledge is required in the fault diagnosis process. In the future, we will implement the proposed MSDCN on more dataset to further validate its effectiveness.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFB2006404), in part by the Fundamental Research Funds for the Central Universities (Grant No. 2242019K3DN05), in part by the cooperative research project of Southeast University and Nanjing Medical University (Grant No. 3202009351). The author would like to appreciate the editors and the anonymous reviewers for their insightful comments.

References

[1] Y. Xu, J. Hu, “Weak fault detection of rolling bearing using a DS-based adaptive spectrum reconstruction method,” *J. Instrum*., vol. 14, no. 03, pp. 03022-03022, 2019.

[2] X. Yan, Y, Liu, Y. Xu and M. Jia, “Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition”, *Energ. Convers & Manage*, vol. 255, pp. 113456, 2020, doi: 10.1016/j.enconman.2020. 113456.

[3] X. Yan, Y, Liu, Y. Xu and M. Jia, “Multichannel fault diagnosis of wind turbine driving system using multivariate singular spectrum decomposition and improved Kolmogorov complexity”, *Renew. Energ*., vol. 170, 2021.

[4] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: A review,” *Mech. Syst. Signal Process*., vol. 108, pp. 33–47, 2018.

[5] D. Peng, H. Wang, Z. Liu, W. Zhang, M. J. Zuo and J. Chen, "Multibranch and Multiscale CNN for Fault Diagnosis of Wheelset Bearings Under Strong Noise and Variable Load Condition," *IEEE Trans. Ind. Informat*., vol. 16, no. 7, pp. 4949-4960, July 2020

[6] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron*., vol. 63, no. 11, pp. 7067-7075, 2016.

[7] Z. Zhao, T. Li, and J. Wu, “Deep Learning Algorithms for Rotating Machinery Intelligent Diagnosis: An Open Source Benchmark Study,” *ISA Transactions*, 2020.

[8] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015

[9] Q. Guo, Y. Li, Y. Song, D. Wang and W. Chen, "Intelligent Fault Diagnosis Method Based on Full 1-D Convolutional Generative Adversarial Network," *IEEE Trans. Ind. Informat*., vol. 16, no. 3, pp. 2044-2053, Mar. 2020.

[10] Y. Xu, C. Yang, B. Sun, X. Yan and M. Chen, “A novel multi-scale fusion framework for detail-preserving low-light image enhancement,” *Information Sciences*, vol. 548, no.12, pp.378-397, 2021.

[11] B. Zhang, W. Li and X. Li, “Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks,” *IEEE Access*, 2018, 6: 66367 — 66384.

[12] J. Jiao, M. Zhao, J. Lin and C. Ding, "Deep Coupled Dense Convolutional Network With Complementary Data for Intelligent Fault Diagnosis," *IEEE Trans. Ind. Electron*., vol. 66, no. 12, pp. 9858-9867, Dec. 2019.

[13] D. Peng, H. Wang, Z. Liu, W. Zhang, M. J. Zuo and J. Chen, "Multibranch and Multiscale CNN for Fault Diagnosis of Wheelset Bearings Under Strong Noise and Variable Load Condition," *IEEE Trans. Ind. Informat*., vol. 16, no. 7, pp. 4949-4960, July 2020.

[14] G. Jiang, H. He, J. Yan and P. Xie, "Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox," *IEEE Trans. Ind. Electron*., vol. 66, no. 4, pp. 3196-3207, Apri. 2019.

[15] Y. Li, G. Li and Y. Yang, “A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy,” *Mech. Syst. Signal Process*, vol.105, pp.319-337 2018.

[16] X. Yan, Y Liu, M. Jia, “Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions,” *Knowl.-Based Syst*. Vol. 193, pp.105484, 2020.

[17] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261-2269.

[18] G. Huang, S. Liu, L. v. d. Maaten and K. Q. Weinberger, "CondenseNet: An Efficient DenseNet Using Learned Group Convolutions," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 2752-2761.

[19] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, “A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals,” *Sensors*, vol. 17, no. 2, pp. 425–446, 2017.

[20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[21] Y. Xu, B, Sun, X. Yan, J. Hu. and M. Chen, “Multi-focus image fusion using learning based matting with sum of the Gaussian-based modified Laplacian, “ *Digital Signal Process*, vol. 106, pp. 202821, 2020

[22] Y. Xu, B, Sun, “Color-compensated multi-scale exposure fusion based on physical features, “*Optik*, vol. 223, pp. 165494, 2020.

[23] M. Zhao and J. Lin, “Health assessment of rotating machinery using a rotary encoder,” IEEE Trans. Ind. Electron., vol. 65, no. 3, pp. 2548–2556, Mar. 2018.

[24] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: A review,” Mech. Syst. Signal Process., vol. 108, pp. 33–47, Aug. 2018.

[25] L. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal Machine Learning Research*, vol. 9, no. 2605, pp. 2579-2605, 2008.