Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer CNN and thermal images

***Abstract*—The existing intelligent fault diagnosis methods of rotor-bearing system mainly focus on vibration analysis under steady operation, which has low adaptability to new scenes. In this paper, a new framework for rotor-bearing system fault diagnosis under varying working conditions is proposed by using a modified convolutional neural network (CNN) with transfer learning. First, infrared thermal images are collected and used to characterize the health condition of rotor-bearing system. Second, modified CNN is developed by introducing stochastic pooling and Leaky rectified linear unit to overcome the training problems in classical CNN. Finally, parameter transfer is used to enable the source modified CNN to adapt to the target domain, which solves the problem of limited available training data in the target domain. The proposed method is applied to analyze thermal images of rotor-bearing system collected under different working conditions. The results show that the proposed method outperforms other cutting edge methods in fault diagnosis of rotor-bearing system.**

***Index Terms*—Rotor-bearing system, Intelligent fault diagnosis, Modified CNN, Thermal images, Parameter transfer.**

I. Introduction

Rotating machinery plays an increasing role in electric power, manufacturing, transportation, and other industries [1, 2]. As a safety-critical part of rotating machinery, various faults of rotor-bearing system during its service life may cause severe security accidents [3, 4]. Due to automatic detecting capability, more and more intelligent fault diagnosis techniques have been proposed and applied to health monitoring of rotor-bearing system and other industrial equipment [5, 6].

A large number of studies have claimed that the diagnosis results of different shallow learning models are largely affected by the effectiveness of extracted features [7-9]. Recently, more and more attention has been paid to deep learning-based approaches with automatic feature learning capability, such as deep belief network (DBN), stacked auto-encoder (SAE), convolutional neural network (CNN), long short-term memory (LSTM), etc. Oh *et al.* [10] classified different faults of rotor system using DBN and vibration-imaging. Ma *et al.* [11] employed enhanced DBN for fault classification of gear and bearing by fusing vibration and acoustic signals. Abid *et al.* [12] constructed SAE to diagnose bearing faults using the extracted multidomain features of vibration signals. Saufi *et al.* [13] designed SAE for fault recognition of gearbox using vibration signals of multi-sensors. Pan *et al.* [14] modified the standard CNN to identify faults of motor bearing based on noisy vibration data Jia *et al.* [15] presented new CNN for fault detection of planetary gearbox through analysis of transverse vibration and torsional vibration. Shao *et al.* [2] combined DBN and CNN for locomotive bearing fault diagnosis based on the raw vibration analysis. Through literature review, it can be found that most of the existing methods focus on vibration analysis of the rotor-bearing. However, vibration analysis has the problem of affecting the equipment structures and the difficulty in installing sensors [16]. Besides, the processing of vibration signals is very complicated due to the long transfer path of the signal, changeable working conditions, and strong noise in real applications [1, 16].

Infrared thermography (IRT) has provided an advanced tool for equipment condition monitoring in recent years [17-19]. Compared with vibration monitoring, IRT holds unique superiorities such as non-contact, simple installation, high precision, high sensitivity, etc [16]. Considering the special abilities of processing two-dimensional (2D) images, CNNs have been gradually used for fault diagnosis of industrial equipment using the infrared thermal images in the past few years. In 2018 and 2019, Janssens *et al.* utilized deep CNNs to analyze the thermal images for machine fault classification and oil-level prediction [20, 21]. In 2019, Jia *et al.* proposed a new fault diagnosis method of rotor-bearing system based on CNN and thermal images [16]. In 2019, Nasiri *et al.* constructed CNN for intelligent fault identification of cooling radiator using thermal images [22].

However, the current CNN-based diagnosis approaches with thermal images can only deal with the same working condition that is rarely the case in real applications. Also, these methods are all based on the availability of a large number of training samples which are difficult and expensive to acquire. In engineering practice, the operating conditions of equipment frequently change, which will result in different distributions of the collected samples [23]. In addition, it is challenging to train an excellent CNN from scratch only using a few samples [24]. Thus, how to enable the CNNs trained with limited thermal images to achieve satisfactory fault diagnosis accuracy of rotor-bearing system under varying working conditions has become an urgent task.

Transfer learning is considered to have great potential to complete different but similar tasks from the source domain to the target domain [25, 26]. Parameter transfer, the most widely applied transfer learning technique, aims to provide valuable parameter knowledge for the target model from a well pre-trained model (source model) [24, 27]. With well-located

initial parameters and a small number of target samples, the target model can be quickly fine-tuned to solve the target task. Since 2017, the transfer diagnosis performance of CNNs integrated with parameter transfer has been demonstrated by a few case studies [24, 26]. Thus, CNN and parameter transfer can be investigated for fault diagnosis of rotor-bearing system under different working conditions with limited samples.

In this paper, a modified transfer CNN driven by thermal images is proposed to diagnose faults of rotor-bearing system under varying working conditions. The results confirm a better fault diagnosis performance of the proposed method compared with the existing methods. The main contributions of this article are as follows:

1) A new framework for rotor-bearing system fault diagnosis under varying working conditions is proposed by using CNN with transfer learning. Infrared thermal images are collected and used to characterize the health condition of rotor-bearing system.
2) A novel modified CNN model is developed by introducing stochastic pooling and Leaky rectified linear unit. It overcomes the training problems in classical CNN.
3) Parameter transfer is used to enable the source modified CNN to adapt to the target domain to solve the problem of limited available training data in the target domain.

The remaining of this article is mainly described as follows. Section II reviews the brief theory of classical CNN. The proposed method is introduced in detail in Section III. Section IV presents the case study. Finally, the conclusion summaries the paper and proposed future work in Section V.

II. The Classical CNN Theory

Among different types of CNN models, the LeNet-5 is the most classical with less trainable parameters, more mature theory and higher computational efficiency. Besides, LeNet-5 is specifically designed to process 2D grayscale images, which mainly consists of an input layer, two convolutional layers, two pooling layers and a fully connected layer, shown in **Fig. 1**.

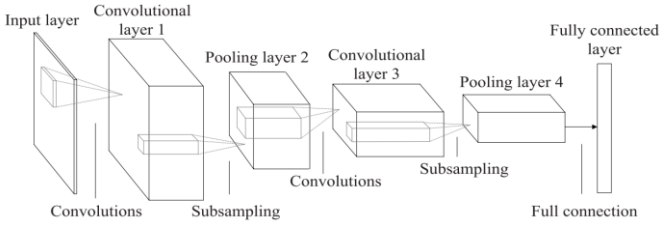


Fig. 1. Model architecture of the classical CNN.

For the convolutional layer in the l^{th} layer, its output is

$$a_{i,j}^{k,l} = f\left(\sum_{k=1}^{K}\sum_{i,j=1}^{N_W}\sum_{r,s=1}^{N_W} W_{r,s}^{k,l} * x_{i+r-1,j+s-1}^{k,l-1} + b^{k,l}\right) \tag{1}$$

$$f(x) = \max(0,x) \text{ or } 1/(1+e^{-x}) \tag{2}$$

where $a_{i,j}^{k,l}$ means the element in the i^{th} row and the j^{th} column of the k^{th} output feature map in the l^{th} layer, $x_{i+r-1,j+s-1}^{k,l-1}$ and $W_{r,s}^{k,l}$ are the elements of the k^{th} input feature map and convolution kernel located in the corresponding position, respectively, $N_W \times N_W$ is the size of each $W_{r,s}^{k,l}$, $b^{k,l}$ is bias, $*$ is a convolution operation, and $f(\cdot)$ is the activation function.

The subsampling operation is used to obtain compressed feature representation and decrease calculation cost, which is usually carried out by max pooling or average pooling as

$$f_{i,j}^{k,l} = \underset{(i,j)\in R_P}{maxpool}(a_{i,j}^{k,l}) \text{ or } \underset{(i,j)\in R_P}{avepool}(a_{i,j}^{k,l}) \tag{3}$$

where $maxpool(\cdot)$ is max pooling, $avepool(\cdot)$ is average pooling, $f_{i,j}^{k,l}$ is the element of the k^{th} output feature map after pooling, and R_P is the pooling region with each size of $N_P \times N_P$. Transformed by the convolution and pooling layers, deep feature maps are finally fed into a fully connected layer and a softmax layer for classification. Weights and biases are adjusted by stochastic gradient descent (SGD) to minimize the errors between the true and predicted labels.

III. The Proposed Method

A. Modified CNN design

In this paper, to overcome the problems of the classical CNN, modifications of the basic CNN are carried out in two aspects: pooling strategy and activation function.

Max and average operations are the two popular pooling strategies in the basic CNN. Some potential information is ignored using max pooling since only the strongest feature elements are selected. Although average pooling considers all the elements, they are always treated equally, reducing the contributions of those important elements. Besides, the certainties caused by these two pooling strategies are more likely to result in over fitting.

As a new subsampling strategy, stochastic pooling can address the problems of max and average operations [28] by fusing all the feature elements according to their contributions, and the sampled probabilities are

$$p_{i,j} = \frac{a_{i,j}}{\sum_{(i,j)\in R_S} a_{i,j}} \tag{4}$$

where $p_{i,j}$ is the probability of element $a_{i,j}$, and R_S is the pooling region. In the back-propagation phase, the fused element is weighted by the probabilistic form of averaging as

$$f_{i,j} = \sum_{(i,j)\in R_S} (p_{i,j} \cdot a_{i,j}) \tag{5}$$

In addition to pooling strategy, the activation function also has an impact on CNN performance. The gradient vanishing and computational complexity are the main limitations of sigmoid (Sigm). The neurons of rectified linear unit (ReLU) always stop learning when the inputs are negative. Leaky ReLU (LReLU) is an enhanced variant of ReLU, which can address these problems by giving a small positive value α when the inputs are negative, helping neurons continue working [29].

$$f(x) = \begin{cases} x & , \text{if } x>0 \\ \alpha x, & \text{if } x \le 0 \end{cases} \tag{6}$$

By now, the modified CNN has been designed, as shown in **Fig. 2**. In order to improve the training performance, SGD algorithm with learning rate decay and momentum are applied to update model parameters

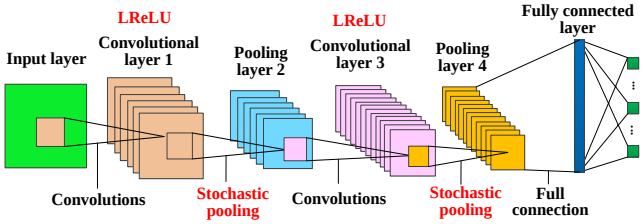


Fig. 2. The modified CNN using stochastic pooling and LReLU.

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \nabla E(\boldsymbol{\theta}_t) + \mu_t(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}) \tag{7}$$

$$\eta_{t+1} = \eta_t / \varepsilon \tag{8}$$

$$\mu_t = \begin{cases} 0.5, & \text{if } t < \text{mom\_epoch} \\ 0.95, & \text{if } t \geq \text{mom\_epoch} \end{cases} \tag{9}$$

where t is the epoch number, $\boldsymbol{\theta}_{t+1}$ is the trained parameters of the modified CNN, $\nabla E(\boldsymbol{\theta}_t)$ is the derivative of cross-entropy error between the true and predicted labels, η_t is learning rate, ε is decay factor, μ_t is momentum, mom_epoch is the boundary epoch number.

B. *Construction of modified transfer CNN*

As shown in **Fig. 3**, by introducing parameter transfer into the modified CNN, a modified transfer CNN can be constructed as follows. (1) Enough source-domain samples are used to pre-train a source modified CNN by minimizing cross-entropy error between the true and predicted labels according to Eqs. (7-9). (2) Prepare a target modified CNN holding the same structure and hyperparameters as the source model. (3) Transfer all the well-learned weights and biases from the source modified CNN to the target. (4) A few target-domain samples are used to fine-tune the target modified CNN. During the fine-tuning stage, all the weights are adjusted. After these steps, the modified transfer CNN can be applied to solve tasks cross domains.

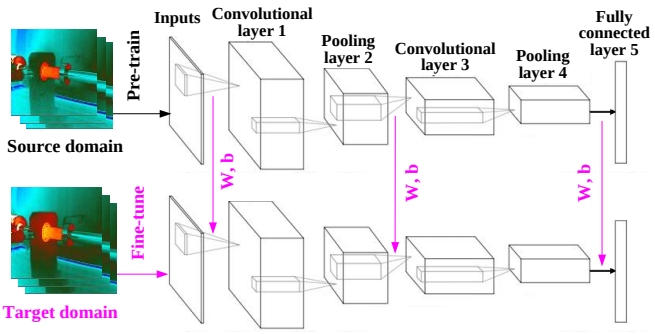


Fig. 3. Construction of modified transfer CNN.

C. *Procedures of the proposed method*

As shown in **Fig. 4**, this paper presents modified transfer CNN and thermal images for fault diagnosis of rotor-bearing system cross working conditions, and the followings are the main procedures:

Step 1: Collect thermal images of rotor-bearing system under different working conditions, and they are first translated into grayscale images and then divided as the source domain and target domain. The source domain contains enough labeled samples while the target domain has a few labeled samples.

Step 2: Stochastic pooling and LReLU are combined to design modified CNN.

Step 3: A source modified CNN model is first well pre-trained using enough samples from the source domain.

Step 4: Transfer all the trained weights and biases provided by the source modified CNN to initialize a target modified CNN with the same structure and hyperparameters.

Step 5: Fine-tune the target modified CNN using small samples from the target domain to further adjust the weights and biases.

Step 6: Test the diagnosis performance of the fine-tuned target modified CNN using the rest samples from the target domain.

IV. Case Validation

A. *Thermal images of rotor-bearing system*

In this paper, the experimental data is from the GUNT PT500 rotating machinery shown in **Fig. 5**, mainly consists of driven motor, rotor shaft, tested bearing, transmission belt, and thermal camera (FLIR A5 with a resolution of 320*156 pixels) [30]. The thermal camera is installed at 40cm away from the fixed rotor shaft.

In the experiments, eight health conditions of rotor-bearing system are created through combining four kinds of bearing conditions and two kinds of rotor conditions, shown in Table I. Three mass blocks (5g) are placed on a side of a disk fixed on the rotating rotor to simulate unbalanced cases, as shown in **Fig. 6(a)**.

As listed in **Table II**, the thermal images collected under steady operation of 2000 rpm and 3000 rpm are employed as the source and target domains, respectively. We acquire a total of 150 source samples and 104 target samples. It should be noted that each target condition only has 2 training (fine-tuning) samples. A raw thermal image of condition 3 with its region of interest is shown in **Fig. 6(b)**. According to the restricted region, the collected thermal image samples of four health conditions (Conditions 3, 4, 7, 8) from the source and target domains are plotted in **Fig. 7**. To meet the 2D input forms of CNN, all of these thermal images are first translated into the grayscale images with sizes of 104*104.

.

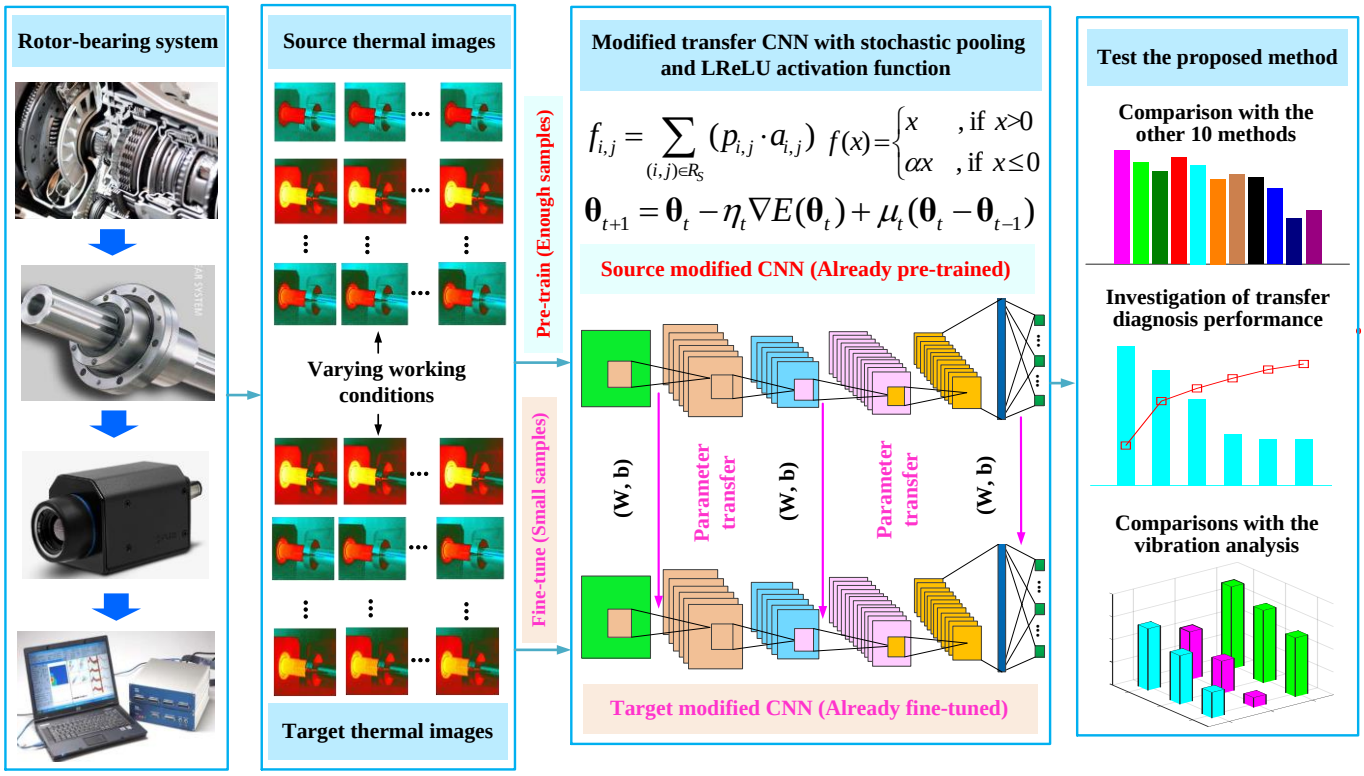


Fig. 4. The overall framework of the proposed method.

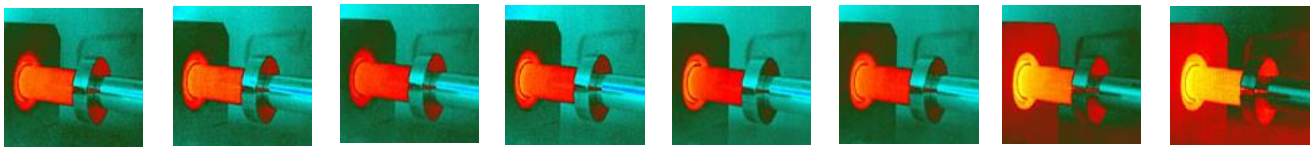


Fig. 7. Collected thermal images of four health conditions: (a-d) Conditions 3, 4, 7, 8 from the source domain; (e-h) Conditions 3, 4, 7, 8 from the target domain.

TABLE III
TRANSFER FAULT DIAGNOSIS RESULTS OF DIFFERENT METHODS

Diagnosis methods	Diagnosis strategies	Average diagnosis results
Method 1 (CNN: LReLU & SP, Proposed method)	Supervised learning and parameter transfer	95.55% (7797/8160)
Method 2 (CNN: LReLU & MP)	Supervised learning and parameter transfer	92.52% (7550/8160)
Method 3 (CNN: LReLU & AP)	Supervised learning and parameter transfer	90.17% (7358/8160)
Method 4 (CNN: ReLU & SP)	Supervised learning and parameter transfer	93.66% (7643/8160)
Method 5 (CNN: ReLU & MP)	Supervised learning and parameter transfer	91.48% (7465/8160)
Method 6 (CNN: ReLU & AP)	Supervised learning and parameter transfer	88.93% (7257/8160)
Method 7 (CNN: Sigm & SP)	Supervised learning and parameter transfer	90.12% (7354/8160)
Method 8 (CNN: Sigm & MP)	Supervised learning and parameter transfer	88.95% (7258/8160)
Method 9 (CNN: Sigm & AP)	Supervised learning and parameter transfer	86.83% (7085/8160)
Method 10 (Basic DBN)	Unsupervised learning and parameter transfer	80.36% (6557/8160)
Method 11 (Basic SAE)	Unsupervised learning and parameter transfer	81.68% (6665/8160)

Remarks: SP: Stochastic pooling; MP: Max pooling; AP: Average pooling.

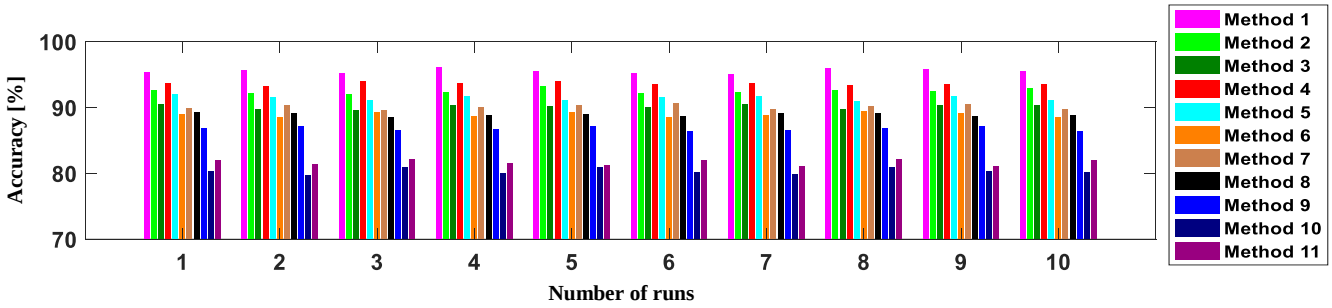


Fig. 8. Detailed transfer diagnosis results of different methods.

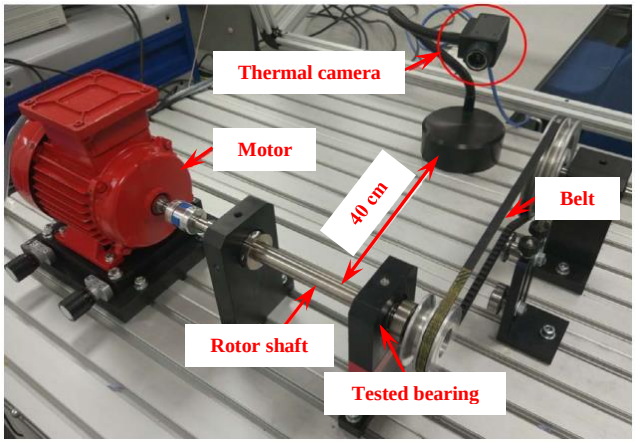
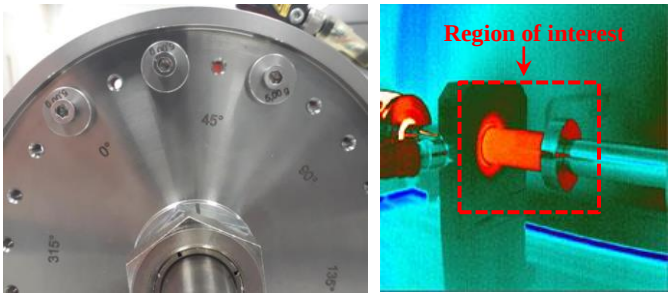


Fig. 5. Diagnosis system of GUNT PT500 rotating machinery.



(a) **(b)**

Fig. 6. (a) The disc fixed on the rotating rotor with three mass blocks and (b) A raw thermal image of condition 3 under steady operation of 2000 rpm.

TABLE I
THE EIGHT HEALTH CONDITIONS OF ROTOR-BEARING SYSTEM

Health conditions of rotor-bearing system	Labels of conditions
Normal bearing and normal rotor	Condition 1 (C1)
Outer race fault of bearing and normal rotor	Condition 2 (C2)
Inner race fault of bearing and normal rotor	Condition 3 (C3)
Ball fault of bearing and normal rotor	Condition 4 (C4)
Normal bearing and unbalanced rotor	Condition 5 (C5)
Outer race fault of bearing and unbalanced rotor	Condition 6 (C6)
Inner race fault of bearing and unbalanced rotor	Condition 7 (C7)
Ball fault of bearing and unbalanced rotor	Condition 8 (C8)

TABLE II
DETAILED INFORMATION OF THE SOURCE AND TARGET DATASETS

Datasets of the thermal images	Rotating speeds	Health conditions	Sizes of the training/ testing samples
Source domain (S)	2000 rpm	Conditions 1 - 8	100*8 / 50*8
Target domain (T)	3000 rpm	Conditions 1 - 8	**2*8** / 102*8

B. Comparisons with other methods

In order to prove the feasibility of the modifications in the proposed method, several currently popular models are utilized for comparisons, including another eight types of CNNs built with different pooling strategies and activation functions, and two unsupervised deep learning models, i.e., DBN and SAE. The diagnosis goals of all the methods are to classify the testing samples in the target domain using the deep models pre-trained by the source-domain samples.

Each method is performed for ten repeated runs so as to avoid contingency and particularity. The transfer fault diagnosis results of various methods are recorded in **Table III**, and the detailed information is available in **Fig. 8**. Through the statistical analysis, the max and mean values of the diagnosis results given by the proposed method are 96.20% (785/816) and 95.55% (7797/8160, 8160=102*8*10), respectively. The average accuracies based on the 10 contrastive methods are 92.52%, 90.17%, 93.66%, 91.48%, 88.93%, 90.12%, 88.95%, 86.83%, 80.36% and 81.68%, respectively, and they are all lower than the proposed method.

From **Fig. 8**, the best diagnosis result of the proposed method occurs in the fourth run with the corresponding confusion matrix shown in **Fig. 9**. In **Fig. 9**, the horizontal X-axis and vertical Y-axis refer to the predicted label and true label, respectively, and the diagonal elements are the accuracies of each state. The precision rate, recall rate, and F-Score of the proposed method for the fourth run are calculated in **Fig. 10** according to the following formulas [2]

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{F-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

in which *TP*, *FP*, and *FN* denote the sizes of the true positive, false positive, and false negative samples, respectively. Every individual condition has a corresponding F-Score, and higher F-Score represents better classification performance. It can be found from **Fig. 9** and **Fig. 10** that except conditions 2 and 3, the F-Score values provided by the proposed method for other conditions are very high.

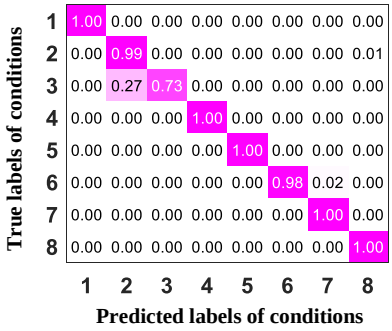


Fig. 9. Confusion matrix of the proposed method for the fourth run.

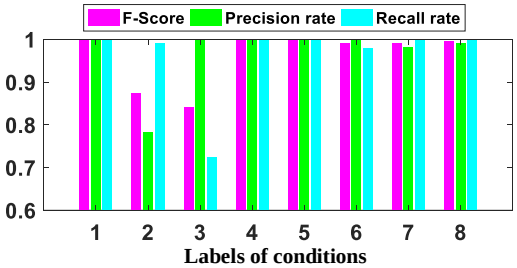


Fig. 10. Precision rate, recall rate, and F-Score of the proposed method for the fourth run.

According to our experience, two convolutional layers and two pooling layers are deep enough in this case study. To strike a good balance between testing accuracy and computing time, the model architectures of the constructed CNN are given in **Table V**. The other hyper-parameters are selected as follows. The initial learning rate is 0.008, initial momentum is 0.5, final momentum is 0.95, epoch number is 500, boundary epoch is 8, decay factor is 5 and the small positive value of LReLU is 0.004, which are mainly determined according to experiences and experimentations. The number of fine-tuning samples from the target domain is set as 2. The relationships between average diagnosis accuracies, standard deviations, and the numbers of fine-tuning samples are investigated in **Fig. 11**. It can be clearly observed that the diagnosis accuracies and standard deviations show respectively steady upward and downward trends as the number of target training samples become larger.

TABLE V
STRUCTURE SETUP OF THE PROPOSED CNN MODEL

Parameters	Descriptions
Input feature map	104x104
Convolution kernels in layer 1	(5x5) x6
Activation function in layer 1	LReLU
Stochastic pooling in layer 2	(2x2) x6
Convolution kernels in layer 3	(5x5) x12
Activation function in layer 3	LReLU
Stochastic pooling in layer 4	(2x2) x12
Softmax layer	8

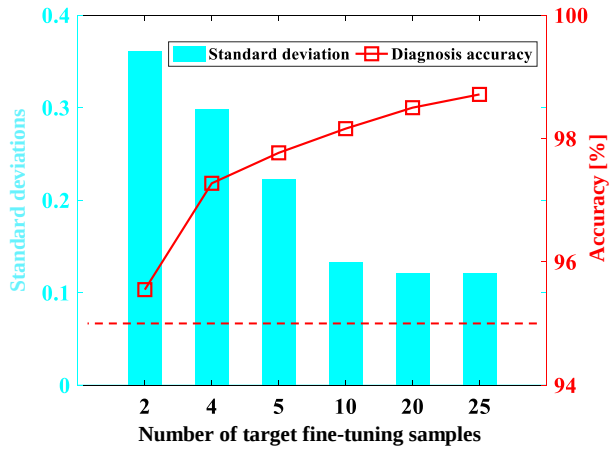


Fig. 11. Transfer diagnosis performance under different fine-tuning samples from target domain.

C. *Superiority of infrared thermal images analysis*

The superiority of thermal images over vibration analysis is investigated in this section. The piezo-electric accelerometer is placed on the side of the support with a sampling frequency of 32768 Hz. Similar to Section A, working conditions of 2000 rpm and 3000 rpm are used as the source and target domains. Data samples of each health condition constructed by the raw vibration signals and time-frequency signals are shown in **Fig. 12** and **Fig. 13**, respectively. Each raw data sample refers to a sequence containing 1024 (32*32) data points without overlap.

The numbers of the source-domain training samples for each health condition are set as 50 (50S), 80 (80S) and 100 (100S), respectively, while the samples sizes in the target domain are the same as Section A. Based on the transfer diagnosis during the ten repeated times, the average diagnosis accuracies of the proposed method for three kinds of inputs are listed in Table VI. **Fig. 14** shows the best accuracies using different inputs during the ten repeated times. From Table VI, two main conclusions are drawn as follows. (1) With the increase of the numbers of the source-domain training samples, all of the accuracies based on different inputs become higher. (2) The accuracies using infrared thermal images (Input 1) are always higher than the other two inputs (Input 2 and Input 3).

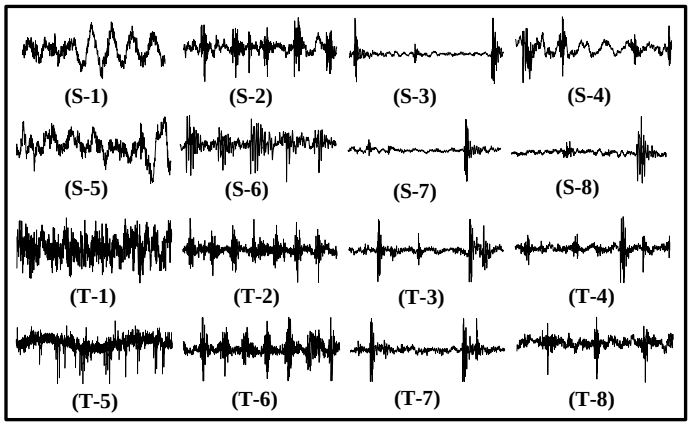


Fig. 12. Data samples of each health condition constructed by raw vibration signals. (T: Target domain, S: Source domain).

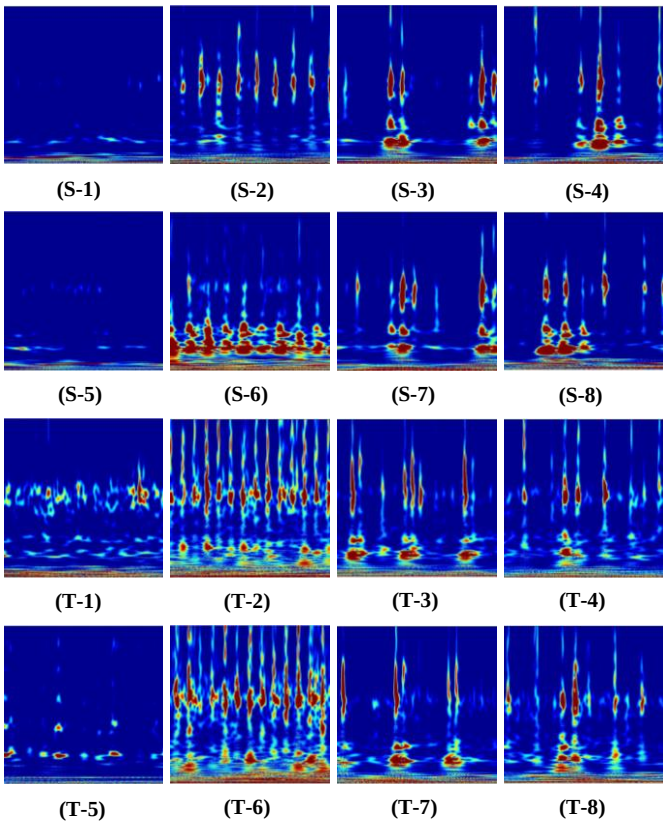


Fig. 13. Data samples of each health condition constructed by time-frequency signals. (T: Target domain, S: Source domain).

TABLE VI
AVERAGE TRANSFER DIAGNOSIS ACCURACIES OF THE PROPOSED METHOD FOR THREE KINDS OF INPUTS

Different inputs of the proposed CNN model	Sizes of the source training samples		
	50 (50S)	75 (75S)	100 (100S)
Infrared thermal images (Input 1)	85.18%	91.53%	**95.55%**
Raw vibration signals (Input 2)	62.11%	72.93%	79.04%
Time-frequency images (Input 3)	70.96%	80.26%	86.18%

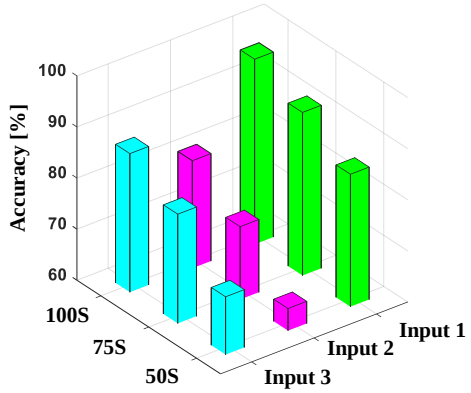


Fig. 14. The best accuracies using different inputs during the ten repeated times

D. Limitations of the proposed method

Despite the proposed method shows better performance for transfer fault diagnosis of rotor-bearing system than the basic CNN and vibration analysis, we have to admit some limitations during the experiments.

(1) Fluctuation problem. From the detailed results shown in **Fig. 8**, it can be found that the proposed method provides different results for each run. Though the average and highest accuracies of the proposed method are both larger than others, the standard deviation is not. The randomness of the stochastic pooling and particularity of the 2 fine-tuned samples randomly selected from the target domain is the main reasons for fluctuation problems. With the increase of the fine-tuned samples from the target domain, the standard deviations will become small to a great degree, as shown in **Fig. 11**.

(2) Hyper-parameter selection. Although the basic CNN is modified by using some skill, more hyper-parameters have to be predetermined simultaneously, including the initial learning rate, initial momentum, final momentum, boundary epoch, decay factor, etc. In this paper, these hyper-parameters are all selected by experimentations. Hyper-parameter selection is an inherent problem in designing different deep learning models, which will more or less be solved with the accumulation of experience and knowledge.

(3) Region of interest for thermal images. From **Fig. 6(b)**, it can be observed that the raw thermal images always contain heavy background noise, i.e., the unrelated information, which will largely affect diagnosis performance. Advanced image segmentation techniques can select high-quality regions of interest to focus on fault characteristics of rotor-bearing system. However, it is carried out manually in this paper.

V. CONCLUSIONS

This article focuses on the transfer diagnosis of rotor-bearing system faults under different working conditions based on advanced monitoring measures. Because of the superiorities of non-contact, simple installation, high precision, and high sensitivity, infrared thermal images are used to characterize the health condition of rotor-bearing system. Stochastic pooling and LReLU are combined to improve the performance of the basic CNN. Parameter transfer is used to provide good initial parameters to enable the designed CNN to adopt new domains with limited available training data.

The proposed method is applied to analyze the thermal images of rotor-bearing system collected under different working conditions. The comparison results confirm the superior performance of the proposed method compared with the existing methods in fault diagnosis of rotor-bearing system. Deep transfer learning has shown the potential to solve practical fault diagnosis tasks with varying working conditions. Future work includes the further improvement of the CNN model, interpretation of the learned features, and fusion of data from multiple sensors to increase the diagnosis accuracy and robustness.

REFERENCES

[1] C. Ding, M. Zhao, J. Lin, J. Jiao and K. Liang, "Sparsity based Algorithm for Condition Assessment of Rotating Machinery Using Internal Encoder Data", *IEEE Trans. Ind. Electron.*, 2019, DOI 10.1109/TIE.2019.2941172.

[2] H. Shao, H. Jiang, H. Zhang and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network", *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727-2736, Mar. 2018.

[3] N. Chandra and A. Sekhar, "Fault detection in rotor bearing systems using time frequency techniques", *Mech. Syst. Signal Process.*, vol. 72-73, pp. 105-133, Mar. 2016.

[4] X. Xia, J. Zhou, J. Xiao and Han Xiao, "A novel identification method of Volterra series in rotor-bearing system for fault diagnosis", *Mech. Syst. Signal Process.*, vol. 66-67, pp. 557-567, Jan. 2016.

[5] M. Zhao, J. Jia and J. Lin, "A Data-Driven Monitoring Scheme for Rotating Machinery Via Self-Comparison Approach", *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2435-2445, Apr. 2019.

[6] L. Wan, G. Han, L. Shu, S. Chan and N. Feng, "PD Source Diagnosis and Localization in Industrial High-Voltage Insulation System via Multimodal Joint Sparse Representation", *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2506-2516, Apr. 2016.

[7] M. Xia, T. Li, T. Shu, J. Wan, C. Silva and Z. Wang, "A Two-Stage Approach for Remaining Useful Life Prediction of Bearings using Deep Neural Networks", *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3703-3711, Jun. 2019..

[8] T. Ko and H. Kim, "Fault Classification in High-Dimensional Complex Processes Using Semi-Supervised Deep Convolutional Generative Models", *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2868-2877, Apr. 2020.

[9] M. Zhao, S. Zhong, X. Fu, B. Tang and M. Pecht, "Deep Residual Shrinkage Networks for Fault Diagnosis", *IEEE Trans. Ind. Informat.*, 2019, DOI 10.1109/TII.2019.2943898.

[10] H. Oh, J. Jung, B. Jeon and B. Youn, "Scalable and Unsupervised Feature Engineering Using Vibration-Imaging and Deep Learning for Rotor System Diagnosis", *IEEE Trans. Ind. Electron.*, vol. 65, no. 4, pp. 3539-3549, Apr. 2018.

[11] M. Ma, C. Sun and X. Chen, "Deep Coupling Autoencoder for Fault Diagnosis With Multimodal Sensory Data", *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1137-1145, Mar. 2020.

[12] A. Abid, M. Khan and M. Khan, "Multidomain Features-Based GA Optimized Artificial Immune System for Bearing Fault Detection", *IEEE Trans. Ind. Syst Man Cy-s.*, vol. 50, no. 1, pp. 348-359, Jan. 2020.

[13] M. Saufi, Z. Ahmad, M. Leong and M. Lim, "Gearbox fault diagnosis using a deep learning model with limited data sample", *IEEE Trans. Ind. Informat.*, 2019, DOI 10.1109/TII.2020.2967822.

[14] J. Pan, Y. Zi, J. Chen, Z. Zhou and Biao Wang, "LiftingNet: A Novel Deep Learning Network With Layerwise Feature Learning From Noisy

Mechanical Data for Fault Classification", *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 4973-4982, Jun. 2018.

[15] 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.

[16] Z. Jia, Z. Liu, C. Vong and M. Pecht, "A Rotating Machinery Fault Diagnosis Method Based on Feature Learning of Thermal Images", *IEEE Access*, vol. 7, pp. 12348-12359, Jan. 2019.

[17] M. Delgado-Prieto, J. Carino-Corrales, J. Saucedo-Dorantes, R. Romero-Troncoso and R. Osornio-Rios, "Thermography-Based Methodology for Multifault Diagnosis on Kinematic Chain", *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5553-5562, Dec. 2018.

[18] V. Tran, B. Yang, F. Gu and A. Ball, "Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis", *Mech. Syst. Signal Process.*, vol. 38, no. 2, pp. 601-614, Jul. 2013.

[19] M. Karakose and O. Yaman, "Complex Fuzzy System Based Predictive Maintenance Approach in Railways", *IEEE Trans. Ind. Informat.*, 2020, DOI 10.1109/TII.2020.2973231.

[20] O. Janssens, R. Walle, M. Loccufier and S. Hoecke, "Deep Learning for Infrared Thermal Image Based Machine Health Monitoring", *IEEE-ASME Trans. Mech.*, vol. 23, no. 1, pp. 151-159, Feb. 2018.

[21] O. Janssens, M. Loccufier and S. Hoecke, "Thermal Imaging and Vibration-Based Multisensor Fault Detection for Rotating Machinery", *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 434-444, Jan. 2019.

[22] A. Nasiria, A. Taheri-Garavandb, M. Omida and G. Carlomagno, "Intelligent fault diagnosis of cooling radiator based on deep learning analysis of infrared thermal images", *Appl. Therm. Eng.*, vol. 163, 114410, Sep. 2019.

[23] B. Yang, Y. Lei, F. Jia and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings", *Mech. Syst. Signal Process.*, vol. 122, pp. 692-706, Mar. 2019.

[24] S. Shao, S. McAleer, R. Yan and P. Baldi, "Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning", *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446-2455, Apr. 2019.

[25] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan and X. Chen, "Deep Transfer Learning Based on Sparse Autoencoder for Remaining Useful Life Prediction of Tool in Manufacturing", *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2416-2425, Apr. 2019.

[26] Z. Chen, K. Gryllias and W. Li, "Intelligent Fault Diagnosis for Rotary Machinery Using Transferable Convolutional Neural Network", *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 339-349, Jan. 2020.

[27] X. Li, H. Jiang, K. Zhao and R. Wang, "A Deep Transfer Nonnegativity-Constraint Sparse Autoencoder for Rolling Bearing Fault Diagnosis With Few Labeled Data", *IEEE Access*, vol. 7, pp. 91216-91224, Jul. 2019.

[28] M. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks", *1st International Conference on Learning Representations (ICLR)*, Scottsdale, Arizona, 1-9, 2013.

[29] A. Maas, A. Hannun and A. Ng, "Rectifier nonlinearities improve neural network acoustic models", *Proceedings of the 30th International Conference on Machine Learning (ICML)*, Atlanta, Georgia, USA, 1-6, 2013.

[30] Z.Q. Huo, Y. Zhang, R. Sath and L. Shu, "Self-adaptive fault diagnosis of roller bearings using infrared thermal images", *IEEE IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, Beijing, China, 6113-6118, 2017.