

Fault diagnosis of a rotor-bearing system under variable rotating speeds using two-stage parameter transfer and infrared thermal images

Haidong Shao, Wei Li, Min Xia, Yu Zhang, Changqing Shen, Darren Williams, Andrew Kennedy, Clarence W. de Silva, *Fellow, IEEE*

Abstract—Current fault diagnosis methods for rotor-bearing system are mostly based on analyzing the vibration signals collected at steady rotating speeds. In those methods, the data collected under one operating condition cannot be accurately used for diagnosis under a different condition. Moreover, in vibration monitoring, installing the necessary sensors will affect the equipment structure and hence the vibration response itself. The present paper proposes a new method based on two-stage parameter transfer and infrared thermal images for fault diagnosis of rotor-bearing system under variable rotating speeds. The method of parameter transfer enables the use of data (or parameters) acquired under one operating condition (called the source domain) to be extended for use in a different operating condition (called the target domain). First, scaled exponential linear unit (SELU) and modified stochastic gradient descent (MSGD) are used to construct an enhanced convolutional neural network (ECNN). Second, a stacked convolutional auto-encoder (CAE) trained based on unlabeled source-domain thermal images is employed to initialize a source-domain ECNN. Third, model parameters from the pre-trained source-domain ECNN are transferred to the target-domain ECNN to adapt to the characteristics of the target domain. The collected thermal images for a rotor-bearing system under variable speeds are used to test the transfer diagnosis performance of the proposed method. The experimental results demonstrate the performance improvement and the advantages of the proposed method.

Index Terms—Rotor-bearing system, Fault diagnosis, Variable rotating speeds, Two-stage parameter transfer, Infrared thermal images.

This work was partially supported by the National Key Research and Development Program of China under Grant No. 2020YFB1712100, the National Natural Science Foundation of China under Grant No. 51905160, the Natural Science Fund for Excellent Young Scholars of Hunan Province under Grant No. 2021JJ20017, and the EU Horizon 2020 under Grant No. 101007005. (*Corresponding author: Min Xia.*)

H. Shao and W. Li are with the College of Mechanical and Vehicle Engineering, Hunan University, China (e-mail: hdshao@hnu.edu.cn; liwei2020@hnu.edu.cn)

M. Xia is with the Department of Engineering, Lancaster University, United Kingdom (e-mail: m.xia3@lancaster.ac.uk).

Y. Zhang is with the Department of Aeronautical and Automotive Engineering, Loughborough University, United Kingdom (e-mail: Y.Zhang@lboro.ac.uk)

C. Shen is with the School of Rail Transportation, Soochow University, China (email: cqshen@suda.edu.cn)

D. Williams and A. Kennedy are with the Department of Engineering, Lancaster University, United Kingdom (e-mail: d.williams8@lancaster.ac.uk; a.kennedy3@lancaster.ac.uk).

C. W. de Silva is with the Department of Mechanical Engineering, The University of British Columbia (e-mail: desilva@mech.ubc.ca).

I. INTRODUCTION

Rotating machinery has extensive applications in modern manufacturing industry [1]. Rotor-bearing systems are a crucial component in rotating machinery. Accurate and automated fault diagnosis of rotor-bearing systems can improve the maintainability and safety of rotating machinery [2-4].

For decades, researchers have made valuable achievements in intelligent fault diagnosis using machine learning techniques such as shallow network models [5-7]. In the past several years, growing attention has been paid to the application of deep learning-based diagnosis techniques [8-12], which can eliminate the dependence on manual feature engineering, thus facilitating the automation of diagnosis. In addition to containing rich dynamic diagnostic information, vibration signals can be acquired by accelerometers with low hardware cost and good sensibility. Oh *et al.* [13] applied deep belief network (DBN) to diagnose various faults of a rotor system using vibration-imaging. Saufi *et al.* [14] constructed stacked auto-encoder (SAE) for the fault identification of a gearbox based on multi-sensor vibration signals. Wen *et al.* [15] developed a two-level hierarchical convolutional neural network (CNN) for rotating machinery fault classification by analyzing the time-frequency images converted from the vibration signals, whose state-of-the-art feasibility was verified by three cases. Although most of the existing diagnosis methods for rotating machinery depend on analyzing the vibration signals, there are two major problems. First, the vibration sensors need to be firmly fixed at a key location of the equipment, which may affect the equipment structure and hence the vibration response itself during operation [16, 17]. Second, due to the coupled vibration of multiple components and complicated transmission paths, the collected vibration signals are usually corrupted by noise disturbances [18]. Infrared thermal images are increasingly used as sensory data for condition monitoring of electromechanical equipment [19-21]. Unlike vibration-based monitoring, the use of infrared thermal images has the advantages of being contactless, convenient, high precision, and wide coverage [18]. It has been shown that the temperature characteristics are stable to the wear and tear, type of fault, the level of severity, and the rotating speeds of bearings [22]. When the rotating machinery operates with localized damage, it would excite the higher temperature generated by the friction [23]. Therefore the thermograms will contain different characteristics of various faults, which could

be used to diagnose the faults of rotor-bearings. Besides, the installation of the necessary equipment is simple and straightforward regardless of the structural complexity of the monitored machine [23]. For image-based fault diagnosis, CNN is considered the most suitable deep learning model due to its specific structure. Recently, some scholars have started to explore the application of CNN and thermal images in intelligent fault diagnosis of rotating machinery. In 2018, Janssens *et al.* [24] proposed a new CNN for machine health monitoring using thermal images. In 2019, Jia *et al.* [18] combined thermal imaging and CNN for fault identification of rotor-bearing systems. In 2020, Li *et al.* [17] employed CNN for fault diagnosis of rotating machinery by analyzing thermal images. In 2020, He *et al.* [25] constructed enhanced CNN after unsupervised initialization to classify fault types of bearing and shaft with small labeled thermal images. CNN-based approaches with thermal images have provided a new way for fault diagnosis of rotor-bearing systems. However, the existing techniques mainly consider diagnosis under a fixed rotating speed, i.e., the training and testing samples are from the same rotating speed. In practice, the rotating speed of a rotor-bearing system will typically vary, which will inevitably lead to characteristic differences, especially the statistical distributions, among the measured samples [26]. In other words, the CNN-based diagnosis model trained with the thermal images collected under one specific rotating speed usually cannot represent other working conditions.

Parameter transfer is currently the most successful application of transfer learning theory [27, 28], which concerns transferring knowledge that is learned under one condition, for use in a different condition. In the present context of fault diagnosis of machinery, it provides the capability of initializing the target-domain model (which is the actual application domain) by using rich knowledge of model parameters gained from a pre-trained source-domain model [29] (which is the domain where the training data is collected). Thanks to the concept and principles of parameter transfer, a deep learning model trained by using source-domain samples has good potential to represent a target-domain, even if the data from the two situations have different characteristics and statistical distributions. Recently, parameter transfer has been applied to improve the generalization abilities of CNNs for fault diagnosis of rotating machinery. In 2018, Cao *et al.* [30] combined a well pre-trained CNN and the approach of parameter transfer for gearbox fault classification. In 2019, Shao *et al.* [29] introduced the approach of parameter transfer in conjunction with properly pre-trained CNN to machinery fault diagnosis under variable rotating speeds. In 2019, Zhong *et al.* [31] proposed a new fault identification method for gas turbine using parameter transfer and CNN. In 2020, He *et al.* [28] constructed ensemble transfer CNNs based on multi-channel signals for fault diagnosis of planetary gearbox. In 2020, Wen *et al.* [32] developed transfer CNN using ResNet-50 to diagnose faults of rotating machinery, whose accuracy and efficiency are better than the existing methods. Although these methods can largely overcome the differences in the statistical distribution between the data samples in the source domain and the target domain, three problems remain to be addressed. First, these methods still use vibration analysis, which has the shortcomings mentioned above. Second, these methods only pay attention to parameter

transfer between the same-type models, which cannot make full use of the characteristics of various models. Third, even though these methods need only small sets of target-domain samples, adequate and labeled source-domain samples are still necessary to realize a properly trained CNN that would provide good performance in the target domain. On one hand, collecting large quantities of labeled source-domain samples will significantly increase the development cost. On the other hand, pre-training a good CNN from scratch, based on large-scale samples, is difficult and time-consuming since a large number of model parameters need to be adjusted [29]. Therefore, reducing the dependence on labeled source-domain samples while enhancing the adaptability of the deep learning model is to be investigated for effective fault diagnosis in rotating machinery.

In this paper, a new method based on two-stage parameter transfer and thermal images is proposed for the fault diagnosis of rotor-bearing systems. The thermal images of a rotor-bearing system, collected under variable rotating speeds, are used to evaluate the performance of the developed method. The experimental result demonstrates the advantages of the developed two-stage parameter transfer strategy. The main contributions of the paper are as follows.

- 1) A new framework for the fault diagnosis of a rotor-bearing system under variable rotating speeds is developed, in which the concept of two-stage parameter transfer is used to facilitate the use of data from a source domain to train a CNN for use in the target domain, and infrared thermal images are utilized to characterize various fault states.
- 2) Scaled exponential linear unit (SELU) and modified stochastic gradient descent (MSGD) are used to construct an enhanced convolutional neural network (ECNN), which is shown to improve performance.
- 3) A stacked CAE trained with unlabeled source-domain thermal images is employed to initialize the source-domain ECNN, reducing the dependence on labeled information.
- 4) Model parameters of the pre-trained source-domain ECNN are further transferred to the target-domain ECNN to adapt to the characteristics of the target-domain thermal images.

The organization of the rest of the paper is as follows. In Section II, a brief review of classical CNN is given. Section III gives details of the proposed method and its implementation. An experimental study, evaluation of the results, and discussion are presented in Section IV. Section V concludes this paper and highlights possible future work.

II. CONCEPTS OF CLASSICAL CNN

The most common and classical CNN model that is applicable in the present problem is LeNet-5, which can use two-dimensional (2D) grayscale images as the input data. As shown in Fig. 1, LeNet-5 usually comprises an input layer, convolutional layer 1, pooling layer 2, convolutional layer 3, pooling layer 4, and a fully connected layer 5.

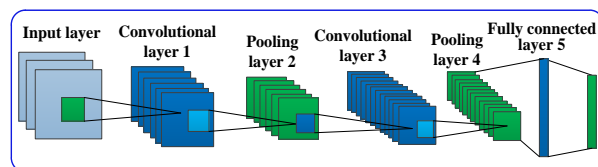


Fig. 1. Structure of a classical CNN (LeNet-5) for image processing.

The calculation principle of convolution layer is given by,

$$h_{r,s}^{k,d} = a\left(\sum_{k=1}^K \sum_{r,s=1}^{N_W} \sum_{i,j=1}^{N_W} W_{i,j}^{k,d} * x_{r+i-1,s+j-1}^{k,d-1} + b^{k,d}\right) \quad (1)$$

With the rectified linear unit (ReLU) activation function,

$$a(x) = \max(0, x) \quad (2)$$

where, $h_{r,s}^{k,d}$ is the element located at the r^{th} row and the s^{th} column of the k^{th} output feature map in the d^{th} layer, $W_{i,j}^{k,d}$ is the k^{th} weight with size of $N_W \times N_W$, $x_{r+i-1,s+j-1}^{k,d-1}$ is the k^{th} input feature map, $*$ is 2D convolution operation, and $b^{k,d}$ is bias.

After each convolutional layer, a max pooling layer is added for reducing the feature dimension, expressed as

$$f_{r,s}^{k,d} = \underset{(r,s) \in R_p}{\text{maxpl}}(h_{r,s}^{k,d}) \quad (3)$$

in which, $\text{maxpl}(\cdot)$ denotes the max pooling operation, $f_{r,s}^{k,d}$ is the corresponding element after max pooling, and R_p is the pooling region with each size of $N_p \times N_p$. High-level features can be acquired through successive convolution and pooling layers, which are then used as the input of a fully connected layer and a softmax layer, to generate the classification output.

III. THE PROPOSED METHOD

A. ECNN construction

In the present paper, to enhance the capability of the classical CNN, the improvements in activation function and training algorithm are combined by constructing the ECNN.

Exponential linear unit (ELU) can effectively overcome the neuron death problem of ReLU, and has reported slightly better performance compared with ReLU in several benchmark data sets. As a further improved version of ELU, scaled ELU (SELU) [33] can make the neural network have self-normalizing ability, defined as

$$a(x) = \begin{cases} \lambda x & , \text{ if } x > 0 \\ \lambda \alpha (e^x - 1) & , \text{ if } x \leq 0 \end{cases} \quad (4)$$

where, λ and α are two control coefficients of SELU.

In the stochastic gradient descent (SGD) algorithm, the learning rate stays the same during the training process, which is inadequate to effectively adjust all the weights and biases at different layers. Although basic adaptive learning rate strategy has been proved to be feasible in many cases, it cannot synchronously and timely capture the change of the MSE function. Based on the moving average gradient and square gradient, Adam optimizer has the advantages of high efficiency and smooth gradient, which has become a priority choice of neural network training. However, Adam optimizer has three parameters that need to be specified [34]. Here, to update these weights and biases more easily and efficiently, a modified SGD (MSGD) algorithm with tracking learning rate and momentum is used, specifically,

$$\theta_{t+1} = \theta_t - \eta_t (\partial E_t / \partial \theta_t) + \mu_t (\theta_t - \theta_{t-1}) \quad (5)$$

$$\eta_{t+1} = (1 + \gamma) \cdot \eta_t \quad (6)$$

$$\gamma = \tau \cdot (1 - e^{(E_{t+1} - E_t) / E_t}) \quad (7)$$

where t is the iteration number, θ_t is the trained parameter set at the t^{th} iteration, E_t , g_t , η_t are the MSE function, gradient, and learning rate at the t^{th} iteration, respectively, γ is a self-adjusting coefficient that depends on the tracking factor τ ,

μ_t is the momentum. Generally, at the beginning of training, it is unlikely to fall into the local optimum, so the momentum value does not need to be very large to reach a certain steady state. Later, when looking for the optimal solution, it needs to jump out of the local optimum, so it needs a larger momentum value. Following this idea and according to some published researches and their open codes [35], specify the initial momentum at the beginning (current iteration < a boundary iteration number) of training is suggested close to 0.5, and the final momentum is suggested close to 0.95. In the MSGD, the learning rate has a direct and synchronous relationship with the change of the MSE at different iterations. So far, the ECNN based on SELU and MSGD has been constructed. Next, the two-stage parameter transfer strategy will be introduced.

B. The first transfer stage: parameter transfer from stacked CAE to source-domain ECNN

The effective training of ECNN cannot avoid the use of large-scale labeled samples. Unsupervised initializations can not only help reduce the dependence on labeled samples, but also avoid local minima [36]. As integration of auto-encoder (AE) and CNN, convolutional AE (CAE) simultaneously has unsupervised learning ability and accommodates the 2D image structures [37]. In addition, the weights and the biases of CAE are both shared among all the feature maps, providing a strong foundation for parameter transfer from CAE to CNN.

The encoder of CAE is a convolution unit consisting of a convolutional layer and a max pooling layer. The k^{th} feature map of a 2D grayscale image \mathbf{x} with no labeled information is,

$$\mathbf{h}^k = g(\mathbf{x} * \mathbf{W}^k + \mathbf{b}^k) \quad (8)$$

where, \mathbf{h}^k is the k^{th} hidden feature map, $g(\cdot)$ is nonlinear activation function, \mathbf{W}^k is the k^{th} weight matrix with the same size as CNN, and \mathbf{b}^k is the corresponding bias vector. Then, a max pooling layer should also follow the convolutional layer.

As an inverse process, the decoder has a deconvolutional layer and a max unpooling layer. The deconvolutional layer is to recover the grayscale image from the hidden feature maps after max pooling, and the reconstruction can be calculated as,

$$\mathbf{y} = \sigma\left(\sum_{k=1}^K \mathbf{h}^k * \hat{\mathbf{W}}^k + \mathbf{c}\right) \quad (9)$$

where, K is the number of hidden feature maps, $\hat{\mathbf{W}}^k$ is the transpose of \mathbf{W}^k , \mathbf{c} is the bias, and \mathbf{y} is the reconstruction of \mathbf{x} with the same size by adding a max unpooling layer after the deconvolutional layer. SGD is used to update the weights and biases to minimize the following function,

$$E_{CAE} = \frac{1}{2S} \sum_{i=1}^S (\mathbf{x}^{(i)} - \mathbf{y}^{(i)})^2 \quad (10)$$

in which, S denotes the total number of 2D grayscale images (unlabeled), and $\mathbf{x}^{(i)}$ denotes the i^{th} input image.

To further learn the high-level compressed features of the grayscale images, this paper constructs a stacked CAE using two CAEs, which is similar to the stacked SAE. Specifically, the stacked CAE in this paper contains two convolutional layers, two pooling layers, a fully connected layer, two unpooling layers, and two deconvolutional layers. The first CAE is trained to learn the first level features of the grayscale images in an unsupervised way. After that, the second CAE continues to mine the second level features based on the first level features. To achieve a small reconstruction error, two deconvolutional layers and two unpooling layers are combined to recover the grayscale images from the two-level features.

Finally, the constructed stacked CAE can be used to acquire high-level features of the unlabeled source-domain grayscale images and provide good trained weights and biases to initialize the source-domain ECNN with the same structure, which no longer requires many labeled samples. The parameter transfer from stacked CAE to source-domain ECNN is shown in Fig. 2.

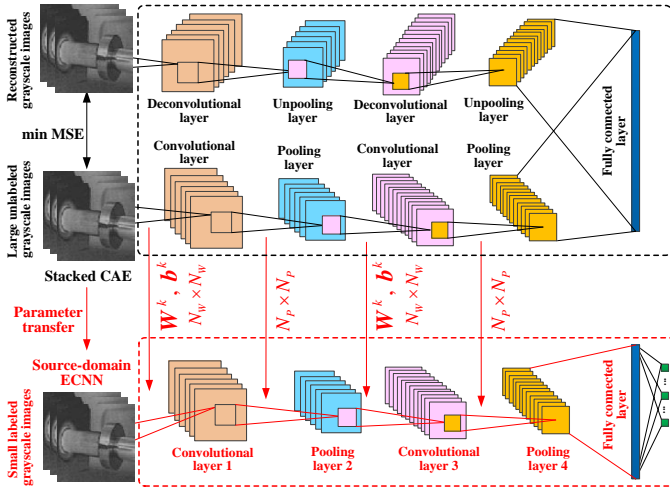


Fig. 2. Parameter transfer from the stacked CAE to source-domain ECNN.

C. The second transfer stage: parameter transfer from source-domain ECNN to target-domain ECNN

Fig. 3 shows the second parameter transfer process from the source-domain ECNN to target-domain ECNN, summarized as follows. (1) Build a target-domain ECNN with the same structure and hyperparameters as the pre-trained source-domain ECNN initialized by a trained stacked CAE; (2) Transfer all the pre-trained weights and biases to the target-domain ECNN from the source-domain ECNN; and (3) Fine-tune the target-domain ECNN with limited labeled target-domain grayscale images. In detail, no weights are fixed while fine-tuning, that is, all of the weights are first pre-trained, then transferred, and finally fine-tuned.

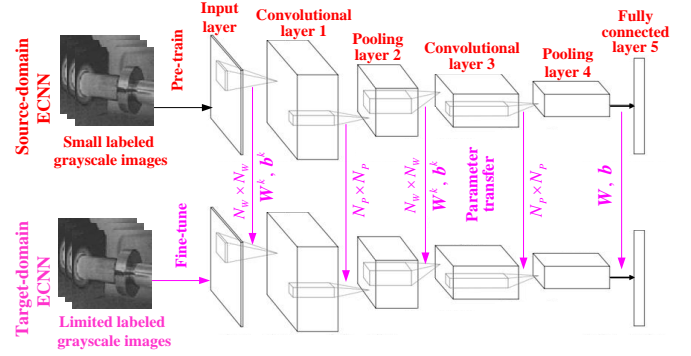


Fig. 3. Parameter transfer from source-domain ECNN to target-domain ECNN.

D. Procedures of the proposed method

The flowchart of the proposed method is shown in Fig. 4. The main steps are as follows.

Step 1: Infrared thermal images of the rotor-bearing system under variable rotating speeds are collected and converted to grayscale images. Then, these grayscale images are divided into the source domain and the target domain, both of which include a small number of labeled samples.

Step 2: SELU activation function and MSGD algorithm are used to construct the ECNN.

Step 3: Some unlabeled source-domain samples are used to train a stacked CAE.

Step 4: Transfer all the trained weights and biases of the stacked CAE to initialize the source-domain ECNN with the same network architecture, which is then pre-trained using a small number of labeled source-domain samples.

Step 5: Transfer all the pre-trained weights and biases of the source-domain ECNN to initialize the target-domain ECNN with the same network architecture and hyperparameters.

Step 6: Fine-tune the target-domain ECNN using a small number of labeled target-domain samples to further update all the weights and the biases.

Step 7: Use unlabeled target-domain samples to test the transfer diagnosis performance of the proposed method.

IV. CASE STUDY

A. Introduction to the infrared thermal images

The fault diagnosis test bench for the experimental study is shown in Fig. 5, including a motor, tested bearing unit, rotating shaft, and a transmission belt. An infrared thermal imaging camera (FLIR Ax5: 320*256 pixels) is used for monitoring the bearing and the shaft [38].

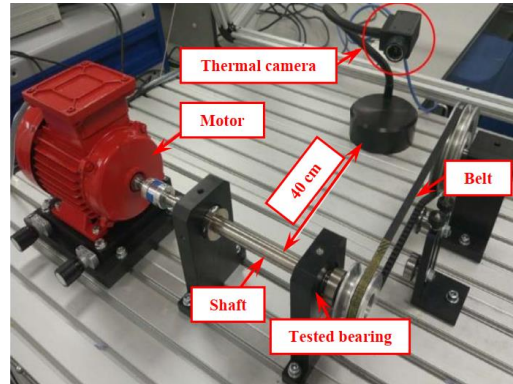


Fig. 5. Fault simulation test bench of the rotor-bearing system.

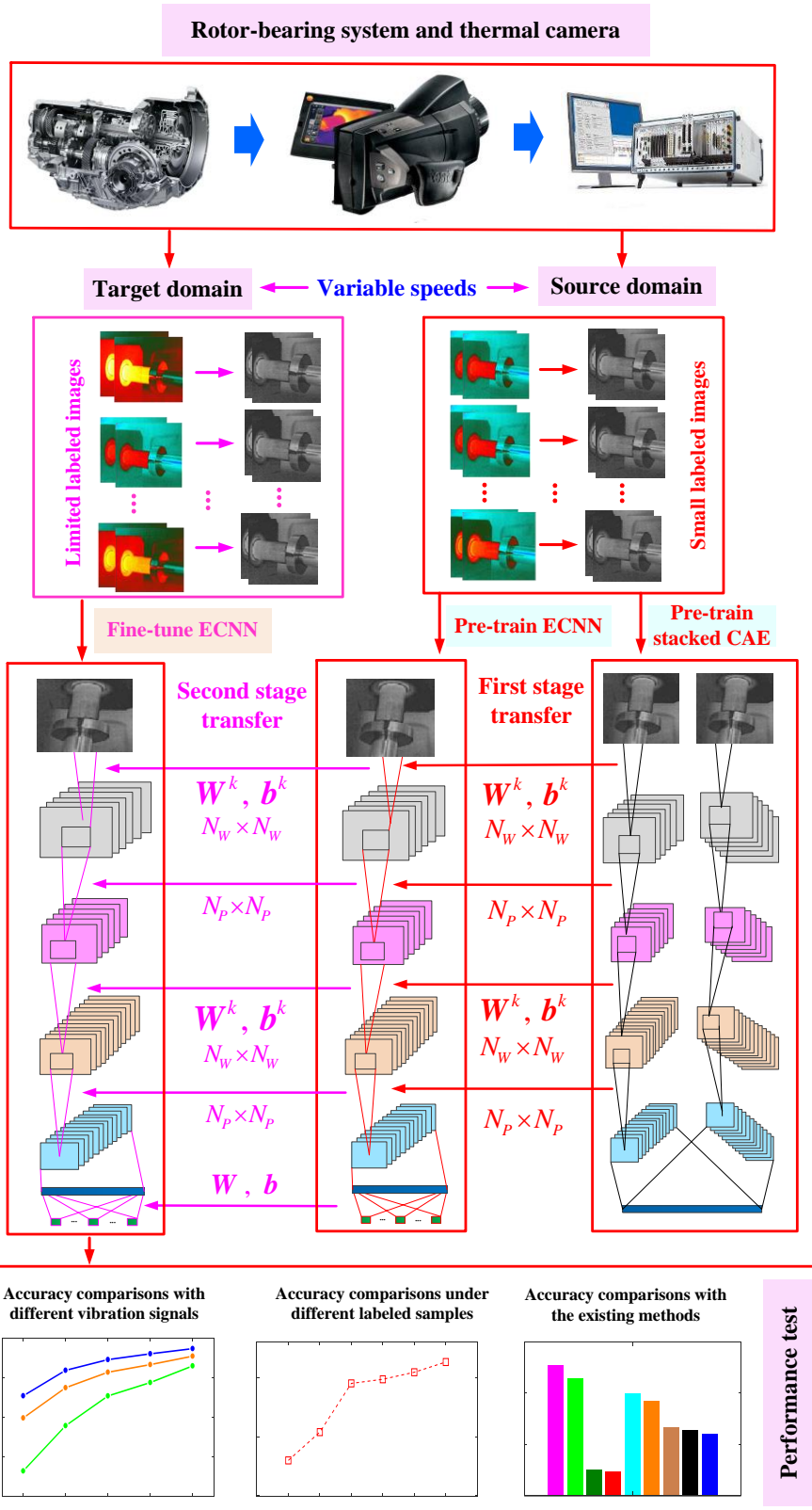


Fig. 4. The overall system framework of the proposed method.

TABLE II
SOURCE-DOMAIN AND TARGET-DOMAIN DATASETS

The grayscale images	Rotating speeds	Extra noise	Sizes of the training samples for each health state		Sizes of the testing samples for each health state
			Unlabeled samples	Labeled samples	
Source domain	2000 rpm	No	70 for training stacked CAE	30 for pre-training source-domain ECNN	50 for testing source-domain ECNN
Target domain	3000 rpm	Yes	-----	5 for fine-tuning target-domain ECNN	100 for testing target-domain ECNN

As listed in **Table I**, eight health conditions are selected for administering to the rotor-bearing system. The unbalanced fault of the rotating shaft is generated by installing three mass blocks on the same side of a disk. Limited by the existing experimental device, the infrared thermal images collected under 2000 rpm and 3000 rpm are treated as the source domain and the target domain, respectively. As the aim of this paper is to achieve fault transfer diagnosis with small amounts of both labeled source-domain samples and target-domain samples, only 30 source-domain samples and 5 target-domain samples have labeled information for each condition, as shown in **Table II**.

TABLE I

DETAILS OF THE EIGHT CONDITIONS OF THE ROTOR-BEARING SYSTEM

Health conditions of the rotor-bearing system	Labels of the conditions
Normal bearing and normal rotating shaft	1 (Condition 1)
Outer race fault and normal rotating shaft	2 (Condition 2)
Inner race fault and normal rotating shaft	3 (Condition 3)
Ball fault and normal rotating shaft	4 (Condition 4)
Normal bearing and unbalanced rotating shaft	5 (Condition 5)
Outer race fault and unbalanced rotating shaft	6 (Condition 6)
Inner race fault and unbalanced rotating shaft	7 (Condition 7)
Ball fault and unbalanced rotating shaft	8 (Condition 8)

Since the infrared thermal images are often sensitive to the environment, the region of interest should be firstly selected to increase the analysis efficiency and suppress the background interference. Specifically, the region of interest is manually determined in this study, which mainly consists of the tested bearing, disk, rotating shaft and housing. To further test the robustness of the proposed method, extra Gaussian noise with zero-mean and variance of 0.005 is added infrared thermal images in the target domain to increase the distribution differences with the source domain. **Fig. 6** and **Fig. 7** successively show the infrared thermal images and the corresponding grayscale images of the four health conditions (Conditions 3, 4, 7, and 8) in the source domain and the target domain, based on the selected region of interest.

Here, based on the image level, an evaluation index called structural similarity (SSIM) is used to measure the similarity of two images from the brightness, contrast and structure, and the defined formula can be found in Ref. [39]. Take the random 4 images of source-domain Condition 3 and the random 4 images of target-domain Condition 3 with no noise as examples, the 16 SSIM values are calculated in **Table III**. It can be seen that although the SSIM values between different pairs show slight

diversities due to the random mechanism of the algorithm, most of them are larger than 0.58. For comparison, SSIM values between the random 4 images of source-domain Condition 3 and the random 4 images of target-domain Condition 3 with Gaussian noise are also calculated. We can find from **Table IV** that most of the SSIM values are smaller than 0.29. Similar situations also appear in the other 7 conditions. Therefore, adding Gaussian noise can indeed increase the differences between the two domains.

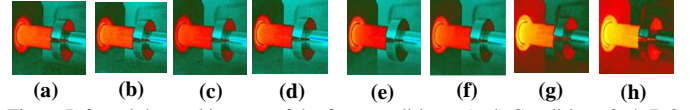


Fig. 6. Infrared thermal images of the four conditions: (a-d) Conditions 3, 4, 7, 8 from the source domain; (e-h) Conditions 3, 4, 7, 8 from the target domain.

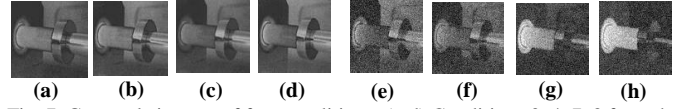


Fig. 7. Grayscale images of four conditions: (a-d) Conditions 3, 4, 7, 8 from the source domain; (e-h) Conditions 3, 4, 7, 8 (with noise) from the target domain.

TABLE III

SSIM VALUES BETWEEN THE RANDOM 4 IMAGES OF SOURCE-DOMAIN CONDITION 3 AND THE RANDOM 4 IMAGES OF TARGET-DOMAIN CONDITION 3 WITH NO EXTRA GAUSSIAN NOISE

Random 4 images of source-domain Condition 3	Random 4 images of target-domain Condition 3 with no extra Gaussian noise			
	Image 1	Image 2	Image 3	Image 4
Image 1	0.5906	0.5889	0.5899	0.5830
Image 2	0.5967	0.5882	0.5895	0.5869
Image 3	0.6015	0.5937	0.5916	0.5934
Image 4	0.6038	0.5966	0.5915	0.5916

TABLE IV

SSIM VALUES BETWEEN THE RANDOM 4 IMAGES OF SOURCE-DOMAIN CONDITION 3 AND THE RANDOM 4 IMAGES OF TARGET-DOMAIN CONDITION 3 WITH EXTRA GAUSSIAN NOISE

Random 4 images of source-domain Condition 3	Random 4 images of target-domain Condition 3 with extra Gaussian noise			
	Image 1	Image 2	Image 3	Image 4
Image 1	0.2884	0.2849	0.2816	0.2808
Image 2	0.2880	0.2861	0.2815	0.2821
Image 3	0.2859	0.2834	0.2771	0.2822
Image 4	0.2826	0.2819	0.2763	0.2774

TABLE V
AVERAGE FAULT TRANSFER DIAGNOSIS RESULTS OF DIFFERENT METHODS

Different diagnosis methods	Numbers of labeled source-domain / target-domain samples	Parameter transfer strategies	Average results
Method 1 (CAE & ECNN: SELU & MSGD, Proposed)	30 / 5	Two-stage parameter transfer	94.54% (7563/8000)
Method 2 (CAE & CNN: SELU & SGD)	30 / 5	Two-stage parameter transfer	88.91% (7113/8000)
Method 3 (CAE & CNN: SELU & Adam)	30 / 5	Two-stage parameter transfer	92.43% (7394/8000)
Method 4 (CAE & CNN: ReLU & MSGD)	30 / 5	Two-stage parameter transfer	89.56% (7165/8000)
Method 5 (ECNN: SELU & MSGD)	30 / 5	One-stage parameter transfer	74.03% (5922/8000)
Method 6 (CNN: ReLU & MSGD)	30 / 5	One-stage parameter transfer	71.53% (5722/8000)
Method 7 (ECNN: SELU & MSGD)	60 / 5	One-stage parameter transfer	87.72% (7018/8000)
Method 8 (CNN: ReLU & MSGD)	60 / 5	One-stage parameter transfer	84.46% (6757/8000)
Method 9 (ECNN: SELU & MSGD)	90 / 5	One-stage parameter transfer	92.15% (7372/8000)

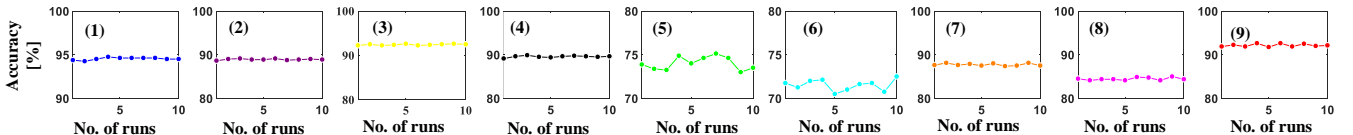


Fig. 8. Detailed fault transfer diagnosis results of different methods for the ten runs: (1)-(9) mean Method 1 to Method 9.

B. Superiority of the two-stage parameter transfer idea

The performance of the two-stage parameter transfer and ECNN are both tested to demonstrate their superiorities. Two transfer strategies, two activation functions, three optimizers are considered for comparison. All of the methods aim to classify the target-domain testing samples using the diagnosis models pre-trained with the source-domain samples.

Fig. 8 shows the detailed fault transfer diagnosis results of each method for the ten independent runs to avoid the contingency of the results, i.e., each method runs for ten repeated times using the randomly selected samples, and the statistical results are presented in **Table V**. The average transfer diagnosis accuracy rate of the proposed method is about 95% with a total correct diagnosis number of 7563, outperforming the comparative methods. Specifically, by comparing Methods 1, 2, and 3, the improvements of average accuracies using MSGD are about 6% and 2%, respectively. The improvement of average accuracy based on SELU is about 5%. The average diagnosis results of Methods 5, 7, and 9 indicate that the training of the source-domain ECNN depends on the labeled samples using one-stage parameter transfer. It is believed that the transfer diagnosis accuracy can be further improved by adding more labeled samples, which however will increase the cost at the same time. The positive role of the first stage parameter transfer is demonstrated by comparing Methods 1 and 5.

For the first run, the specific fault transfer diagnosis accuracy of the proposed method is 94.25% (754/800). The multi-class confusion matrix is given in **Fig. 9**, and the F-scores are calculated and presented in **Fig. 10**. The vertical coordinate and horizontal coordinate in **Fig. 9** refer to the actual and predicted labels, respectively. The matrix elements located on the main diagonal refer to the diagnostic accuracies of the corresponding condition, and the other elements refer to the misdiagnosis rates. It can be found that Conditions 1, 4, and 5 all have 100% accuracy. Condition 1 (normal bearing and normal shaft) is intuitively far away from other conditions that results in a high diagnosis accuracy. The misclassification happens between Conditions 2, 3, and among Conditions 6, 7, 8. The main reason is that Conditions 2 and 3 have similar characteristics of temperature distributions, so as for Conditions 6, 7, 8 (with unbalanced shaft). The factor of unbalanced shaft brings the group of Conditions (5, 6, 7, 8) distanced from the other conditions with normal shaft. Condition 5 with normal bearing is easier to be separated from Conditions 6, 7, and 8.

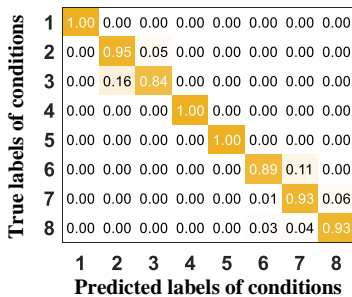


Fig. 9. Confusion matrix of the proposed method for the first run.

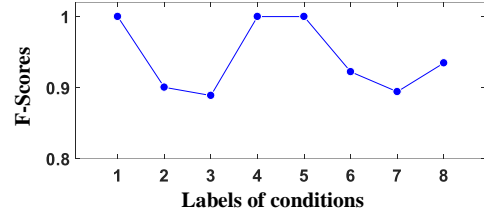


Fig. 10. F-Scores of the proposed method for the first run.

After acquiring a series of infrared thermal images, their regions of interest need to be first decided and converted into grayscale images, to meet the requirement of the inputs for stacked CAE and ECNN. The conversion process from infrared thermal images with a size of 320×256 to grayscale images with a size of 88×88 can be implemented by successively using three functions in MATLAB 2016b with names of ‘rgb2gray’, ‘imcrop’ and ‘imresize’. For each run, the architecture settings of the stacked CAE and ECNN are given in **Table VI**. The hyper-parameters are chosen as follows. (1) Stacked CAE: the learning rate = 0.1, and iteration = 100; (2) ECNN: the initial learning rate = 0.0008, iteration = 400, boundary iteration = 40, two control coefficients of SELU $\lambda = 1.05$, $\alpha = 1.67$, dropout rate = 0.4, and tracking factor $\tau = 6.5$. The model architectures and hyper-parameters are selected according to the suggested value from related papers and a grid search of the hyper-parameters.

Generally, small size of training samples and large iterations will easily cause the overfitting problem. This paper aims to solve the fault transfer diagnosis problem with only a small amount of labeled source-domain and target-domain samples at the same time. Thus, several techniques are combined to avoid overfitting, including dropout, SELU activation function, MSGD algorithm, early stopping, and small initial learning rate. Thus, despite the maximal iteration number is set as 400, the training process in many experiments converged within 200 iterations with early stop if the validation error started to increase. The software environment and hardware configuration are as follows: Windows 10 64-bit operating system, MATLAB 2016b, Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz, and 16 GB RAM. During the ten independent runs, the average computing time of the proposed method is about 148s, mainly consists of the training of stacked CAE and source-domain ECNN. The total number of unlabeled samples for training the stacked CAE is 70*8, and the total number of labeled samples for training the source-domain ECNN is 30*8. Each sample refers to a grayscale image with a size of 88*88 converted from the infrared thermal image.

TABLE VI
MODEL ARCHITECTURES OF THE STACKED CAE AND ECNNs

Settings of model architectures	Details
Weight matrices of stacked CAE and ECNNs (Layer 1)	(5x5) x6
Activation functions of stacked CAE and ECNNs (Layer 1)	ReLU / SELU
Max pooling of stacked CAE and ECNNs (Layer 2)	(2x2) x6
Weight matrices of stacked CAE and ECNNs (Layer 3)	(5x5) x12
Activation functions of stacked CAE and ECNNs (Layer 3)	ReLU / SELU
Max pooling of stacked CAE and ECNNs (Layer 4)	(2x2) x12

C. Advantage of monitoring through thermal images

Two types of monitored signals are compared in this subsection, which are thermal images and vibration signals. An acceleration sensor is fixed near the tested bearing in the

support structure, and the sampling frequency is 32768 Hz. As before, the collected vibration data at the speeds 2000 rpm and 3000 rpm is treated as the source domain and the target domain, respectively. Here, each data sample is a signal segment consisting of 1024 sequence points. Before feeding into the stacked CAE and ECNN, they are translated into time-frequency images. **Fig. 11** and **Fig. 12** show the data samples based on the original time-domain vibration signals and time-frequency images of the four conditions, respectively. To obtain the 2D time-frequency images, continuous wavelet transform is used with the complex Morlet wavelet. The bandwidth parameter and wavelet center frequency are both set as 3. Different total scales have been tried and 256 is selected as it produces the highest accuracy. Therefore, the frequency resolution is 64Hz ($32768/2*256$). In **Fig. 12**, the horizontal axis refers to time with the range of $[0, 1024/32768]$ s, and the vertical axis refers to the frequency with the range of $[0, 32768/2]$ Hz. Note that, here, all of the samples from the target domain are not mixed with extra Gaussian noise.

The procedures of fault transfer diagnosis using the original vibration signals and time-frequency images are the same as those used for the infrared thermal images. After utilizing 70 labeled source-domain samples to train the stacked CAE and 5 labeled target-domain samples to fine-tune the source-domain ECNN, the average diagnosis accuracies of the ten repeated runs for three types of inputs are shown in **Fig. 13**. It can be seen that (1) the accuracies with infrared thermal images (with extra noise) are always higher than those with the other two inputs (with no extra noise) under different numbers of labeled source-domain samples, which are used to pre-train the source-domain ECNN; (2) the average accuracies always increase as the number of labeled samples is increased for the three types of inputs. Specifically, with the proposed method, 20 labeled samples have achieved an accuracy level of 90% and 30 labeled samples have increased that level to 95%.

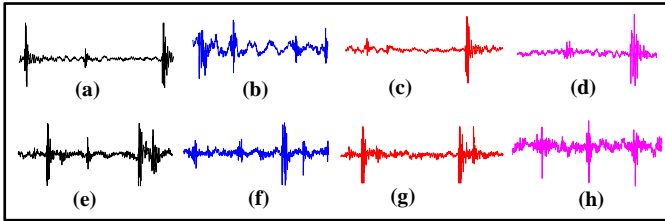


Fig. 11. Original vibration signals of the four conditions: (a-d) Conditions 3, 4, 7, 8 from the source domain; (e-h) Conditions 3, 4, 7, 8 from the target domain.

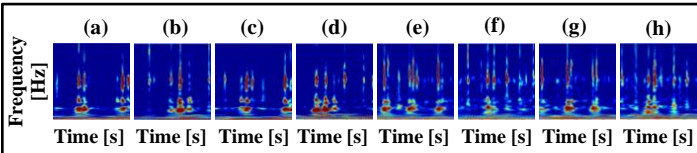


Fig. 12. Time-frequency images of the four conditions: (a-d) Conditions 3, 4, 7, 8 from the source domain; (e-h) Conditions 3, 4, 7, 8 from the target domain.

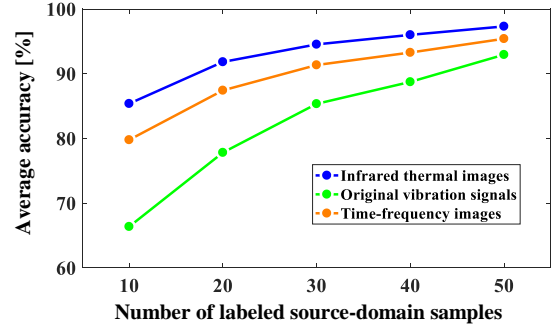


Fig. 13. Average transfer diagnosis accuracies under three types of inputs with different numbers of labeled source-domain samples.

D. Discussions on the numbers of labeled target-domain samples for fine-tuning

The use of labeled target-domain samples aims at fine-tuning the pre-trained source-domain ECNN to adapt to the characteristics of the target-domain that the fault diagnosis is to apply. Due to the notable difficulties in obtaining labeled target domain samples, the analysis of the number of labeled target domain samples needed has practical significance.

As listed in **Table VII**, six cases with a different number of the labeled target-domain samples are considered for the proposed method, and the number of the target-domain samples used for testing is set at 100 for all the cases. **Fig. 14** shows the average transfer diagnosis accuracies of the six cases. It is seen that the diagnosis accuracy shows a steady upward trend with the increase of the number of labeled target-domain samples used for fine-tuning. Specifically, with only 2 labeled samples, the diagnosis accuracy reaches 88%. The accuracy climbs to 95% with 5 labeled samples. Thus, with a limited number of labeled samples from the target domain, satisfactory transfer diagnosis results can be realized through the proposed method.

TABLE VII
SIX SITUATIONS OF THE LABELED TARGET-DOMAIN SAMPLES

Different cases	Numbers of the target-domain samples for each condition	
	Labeled for fine-tuning ECNN	Unlabeled for testing ECNN
Case 1	2	100
Case 2	3	100
Case 3	5	100
Case 4	6	100
Case 5	10	100
Case 6	15	100

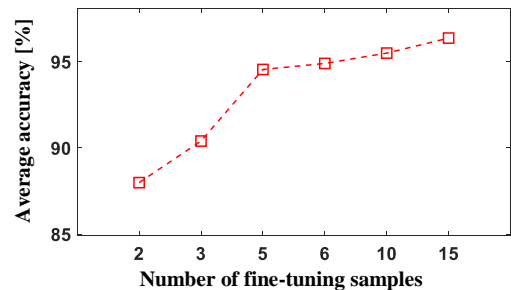


Fig. 14. Average transfer diagnosis accuracy with different numbers of labeled target-domain samples for fine-tuning.

V. CONCLUSIONS

Aiming to boost the diagnosis capability and adaptability of the approach for diagnosing faults in a rotor-bearing system

under different rotating speeds, a new technique based on two-stage parameter transfer and infrared thermal images was presented. The thermal images were collected to reflect the various health states of a rotor-bearing system. Thermal images were chosen due to the advantages of this method of sensing such as non-contact, convenient installation, wide coverage, and high precision. SELU and MSGD were used to construct an ECNN to improve the fault diagnosis performance. Then, a stacked CAE trained with unlabeled source-domain thermal images was used to initialize the source-domain ECNN. Finally, weights and biases of the pre-trained source-domain ECNN were transferred to the target-domain ECNN, to adapt to the characteristics of the target domain.

The transfer diagnosis performance of the proposed method was tested using thermal images collected from a rotor-bearing system at different rotating speeds. The experimental results confirmed the advantages of the two-stage parameter transfer strategy and the use of thermal images, compared with other state-of-art methods. Nevertheless, the proposed method has the following drawbacks, including relatively high hardware cost and difficulties in detecting sudden faults, selecting representative interest areas, and determining suitable hyper-parameters. Future work will mainly consider combining different transfer learning strategies and fault transfer diagnosis using multi-sensor fusion, to further enhance the accuracy and the generalization ability.

VI. REFERENCES

- [1] Z. Hu, Y. Wang, M. Ge and J. Li, "Data-Driven Fault Diagnosis Method Based on Compressed Sensing and Improved Multiscale Network", *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3216-3225, Apr. 2020.
- [2] D. Egorov, I. Petrov, J. Pyrhonen, J. Link and R. Stern, "Hysteresis Loss in Ferrite Permanent Magnets in Rotating Electrical Machinery", *IEEE Trans. Ind. Electron.*, vol. 65, no. 12, pp. 9280-9290, Dec. 2018.
- [3] M. Hamadache, D. Lee, K. Veluvolu, "Rotor Speed-Based Bearing Fault Diagnosis (RSB-BFD) Under Variable Speed and Constant Load", *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6486-6495, Oct. 2015.
- [4] S. Ma and F. Chu, "Ensemble deep learning-based fault diagnosis of rotor bearing systems", *Comput. Ind.*, vol. 105, pp. 143-152, Feb. 2019.
- [5] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li and A. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap", *Mech. Syst. Signal Process.*, vol. 138, 106587, Apr. 2020.
- [6] S. Pandarakone, Y. Mizuno and H. Nakamura, "Evaluating the Progression and Orientation of Scratches on Outer-Raceway Bearing Using a Pattern Recognition Method", *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 1307-1314, Feb. 2019.
- [7] C. Yang, B. Gunay, Z. Shi and W. Shen, "Machine Learning-Based Prognostics for Central Heating and Cooling Plant Equipment Health Monitoring", *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 1, pp. 346-355, Jan. 2021.
- [8] M. Zhao, M. Kang, B. Tang, and M. Pecht, "Deep Residual Networks With Dynamically Weighted Wavelet Coefficients for Fault Diagnosis of Planetary Gearboxes," *IEEE T. Ind. Electron.*, vol. 65, no.5, pp. 4290-4300, May. 2018.
- [9] 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.
- [10] W. Du, H. Liu, J. Xu, H. Zhao and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network", *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7319-7327, Oct. 2020.
- [11] M. Xia, T. Li, T. Shu, J. Wan, C. W. de Silva and Z. Wang, "A Two-Stage Approach for the Remaining Useful Life Prediction of Bearings Using Deep Neural Networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3703-3711, Jun. 2019.
- [12] 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 Trans. Mech.*, vol. 23, no. 1, pp. 101-110, Feb. 2018.
- [13] 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.
- [14] S. Saufi, Z. Ahmad, M. Leong and M. Lim, "Gearbox fault diagnosis using a deep learning model with limited data sample", *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6263-6271, Oct. 2020.
- [15] L. Wen, X. Li and L. Gao, "A New Two-Level Hierarchical Diagnosis Network Based on Convolutional Neural Network", *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 330-338, Feb. 2020.
- [16] 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.
- [17] Y. Li, X. Du, F. Wang, X. Wang and H. Yu, "Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging", *Chinese J. Aeronaut.*, vol. 33, no. 2, pp. 427- 438, Feb. 2020.
- [18] 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.
- [19] A. Mohammad-Alikhani, M. Rahnama and A. Vahedi, "Neighbors Class Solidarity Feature Selection for Fault Diagnosis of Brushless Generator Using Thermal Imaging", *IEEE Trans. Instrum. Meas.*, vol. 69, no.9, pp. 6221-6227, Sep. 2020.
- [20] 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.
- [21] D. Pal, R. Meyur, S. Menon, M. Reddy and D. Mohanta, "Real-time condition monitoring of substation equipment using thermal cameras", *IET Gener. Transm. Dis.*, vol. 12, no. 4, pp. 895-902, Jan. 2018.
- [22] A. Choudhary, D. Goyal and S. S. Letha, "Infrared Thermography-Based Fault Diagnosis of Induction Motor Bearings Using Machine Learning," *IEEE Sens J.*, vol. 21, no. 2, pp. 1727-1734, Jan. 2021.
- [23] Y. Li, X. Wang, S. Si and X. Du, "A New Intelligent Fault Diagnosis Method of Rotating Machinery under Varying-Speed Conditions Using Infrared Thermography", *Complexity*, vol. 5, pp. 1-12, Aug. 2019.
- [24] O. Janssens, R. Walle, M. Loccupier 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.
- [25] Z. He, H. Shao, X. Zhong, Y. Yang and J. Cheng, "An intelligent fault diagnosis method for rotor-bearing system using small labeled infrared thermal images and enhanced CNN transferred from CAE", *Adv. Eng. Inform.*, vol. 46, 101150, Oct. 2020.
- [26] H. Shao, M. Xia, G. Han, Y. Zhang and J. Wan, "Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer CNN and thermal images," *IEEE T. Ind. Inform.*, 2020, DOI 10.1109/TII.2020.3005965.
- [27] X. Li, H. Jiang, R. Wang and M. Niu, "Rolling bearing fault diagnosis using optimal ensemble deep transfer network", *Knowl-Based Syst.*, vol. 213, 106695, Oct. 2021.
- [28] Z. He, H. Shao, X. Zhong and X. Zhao, "Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions", *Knowl-Based Syst.*, vol. 207, 106396, Nov. 2020.
- [29] 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.
- [30] P. Cao, S. Zhang and J. Tang, "Pre-Processing-Free Gear Fault Diagnosis Using Small Datasets with Deep Convolutional Neural Network-Based Transfer Learning", *IEEE Access*, vol. 6, pp. 26241-26253, May. 2018.
- [31] S. Zhong, S. Fu and L. Lin, "A novel gas turbine fault diagnosis method based on transfer learning with CNN", *Measurement*, vol. 137, pp. 435-453, Apr. 2019.
- [32] L. Wen, X. Li and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50", *Neural Comput. Appl.*, vol. 32, pp. 6111-6124, May. 2020.
- [33] G. Klambauer, T. Unterthiner, A. Mayr and S. Hochreiter, "Self-normalizing neural networks", *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, Long Beach, USA, 971-980, 2017.
- [34] K.K. Chandriah, R.V. Naraganahalli, "RNN / LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand

forecasting”, *Multimed. Tools. Appl.*, vol. 80, pp. 26145-26159, Apr. 2021.

- [35] Deep Learning (Neural Networks), <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/deep-learning.html>
- [36] D. Erhan, Y. Bengio, A. Courville, P.A. Manzagol, P. Vincent and S. Bengio, “Why does unsupervised pre-training help deep learning?”, *J. Mach. Learn. Res.*, vol. 11, no. 3, pp. 625-660, Feb. 2010.
- [37] W. Luo, J. Li, J. Yang, W. Xu and J. Zhang, “Convolutional Sparse Autoencoders for Image Classification”, *IEEE Trans. Neur. Net. Lear.*, vol. 29, no. 7, pp. 3289-3294, July 2018.
- [38] 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.
- [39] L. Galteri, L. Seidenari, M. Bertini and A. D. Bimbo, “Deep Universal Generative Adversarial Compression Artifact Removal,” *IEEE Trans. Multimedia*, vol. 21, no. 8, pp. 2131-2145, Aug. 2019.



Haidong Shao received the B.S. degree in Electrical Engineering and Automation and the Ph.D. degree in Vehicle Operation Engineering from Northwestern Polytechnical University, Xi'an, China, in 2013 and 2018, respectively. He is currently an Assistant Professor in the College of Mechanical and Vehicle Engineering, Hunan University, Changsha, China. From 2019 to 2021, he was a Postdoctoral Fellow with Luleå University of Technology, Luleå, Sweden. His current research interests include fault diagnosis, intelligent maintenance and information fusion.



Wei Li received the B.S. degree in Mechanical Design, Manufacturing and Automation from Huazhong Agricultural University, Wuhan, P. R. China in 2020. He is now studying for the M.S. degree in Mechanical Engineering at College of Mechanical and Vehicle Engineering, Hunan University, Changsha, P. R. China. His research interests are deep learning theories and their applications in rotating machinery fault diagnosis.



Min Xia (S'11-M'17) is currently a lecturer in the Department of Engineering at the Lancaster University, UK. He received B.S. degree from Southeast University, China (2009); M.S. degree from the University of Science and Technology of China, China (2012); and Ph.D. degree from the University of British Columbia, Canada (2017). His research interests include smart manufacturing, machine diagnostics and prognostics, deep neural networks, wireless sensor network and sensor fusion.



Yu (Eve) Zhang is a Senior Lecturer in Digital Engineering in the Department of Aeronautical and Automotive Engineering at Loughborough University, UK. She obtained her BEng degree in engineering mechanics from the School of Aerospace Engineering and Applied Mechanics at Tongji University, China in 2004, and her PhD degree in nano-mechanics from the Department of Civil Engineering at the University of Nottingham in 2011. Her current main research

interests include condition monitoring, fault diagnosis and the development of data analysis and machine learning algorithms, especially for industrial applications.



Changqing Shen received the B.S. and Ph.D. degrees in Instrument Science and Technology from the University of Science and Technology of China in 2009 and 2014, respectively. He also obtained the Ph.D. degree in Systems Engineering and Engineering Management from the City University of Hong Kong in 2014. He is currently an Associate Professor with the School of Rail Transportation, Soochow University, China. His research interests include machinery data mining, and intelligent fault diagnosis.



Darren Willimas is the Welding Systems Integration Team Leader at TWI, and the Professor and Director of the Joining 4.0 Innovation Centre at Lancaster University. He received Doctor of Engineering in System Engineering with Management and Accounting from University of Bath in 2013. He is a Chartered Engineer, and Fellow of the Institution of Engineering and Technology. His research interests include digital manufacturing, Industry

4.0, welding and joining technologies



Andrew Kennedy is the Professor for Advanced Manufacturing in the Department of Engineering at the Lancaster University, UK. He received his undergraduate degree from Imperial College London, UK (1989) and a Ph.D. degree from the University of Cambridge, UK (1993). His research interests include process optimization, additive manufacturing and advanced materials processing.



Clarence W. de Silva (F'98) is a Fellow of: IEEE, ASME, Canadian Academy of Engineering, and Royal Society of Canada. He received Ph.D. degrees from Massachusetts Institute of Technology (1978); and University of Cambridge, U.K. (1998); and honorary D.Eng. degree from University of Waterloo, Canada (2008), and an earned higher doctorate (ScD) from University of Cambridge (2020). He has been a Professor of Mechanical Engineering and Senior Canada Research Chair and NSERC-BC Packers Chair in Industrial Automation, at the University of British Columbia, Vancouver, Canada since 1988.