Intelligent Bearing Fault Diagnosis Based on Scaled Ramanujan Filter Banks in Noisy Environments

Ruixian Li, Li Zhuang, Yongxiang Li, and Changqing Shen

***Abstract*—Bearing fault diagnosis plays an essential role in the maintenance of rotating machines in industries. Challenges posed in developing an effective and robust bearing fault diagnosis method include the essential complexity of vibration data and the external interference caused by the data collection. This study develops an intelligent data-driven method for bearing fault diagnosis in noisy environments, consisting of the feature transformation of vibration data and fault recognition based on transformed features. Firstly, an extension of the Ramanujan filter banks (RFB) method, Scaled-RFB, is introduced to suppress the noises and convert original time series vibration data into representative RGB images. Next, a strip convolutional neural network (Strip-CNN) is developed with strip convolution to recognize the health condition of bearings based on the obtained RGB images. Two vibration datasets collected from Soochow University and a public data source are utilized to validate the effectiveness and robustness of the proposed method individually. Six levels of Gaussian noises are separately added into the two datasets to further demonstrate the performance of the proposed method in noisy environments. Compared with six benchmarking methods, the proposed method can achieve the best performance on bearing fault diagnosis in most scenarios and shows promising performance on datasets with a higher noise level. When the signal-noise ratio (SNR) is -10 dB, the** average $Precision$**,** $Recall$**, and** $F1$ **scores of the proposed method on both datasets are at least** 51.79%**,** 52.49%**, and** 52.47% **higher than benchmarking methods, respectively.**

***Index Terms*—Bearing fault diagnosis, Ramanujan filter banks, Deep neural network, Strip convolution, Noisy vibration data**

I. Introduction

BEARINGS are critical components of rotating machines that have been widely deployed in industries [1]. The health of bearings is crucial to ensure the reliability of mechanical systems [2]. Due to long-term usage and harsh working environments, the failure of bearings increases [3]. The occurrence of bearing failures can reduce productivity, leading to stochastic machine breakdown, and even threaten the safety of crews [4]. Nowadays, vibration data, which indicates the condition of bearings, can be effectively collected via acceleration sensors. Therefore, it is meaningful to investigate advanced methods to diagnose bearing faults by using the vibration data.

Traditional bearing fault diagnosis relies on primary signal processing methods, including analyses of statistical time-domain features (i.e., the root mean square value) and frequency-domain features (i.e., the spectral envelope) individually [5]. The performance of such basic signal processing methods is bound to the complexity of collected data. With the rapid development of information technology, machine learning based methods have been applied to bearing fault diagnosis to achieve better performance. The procedure of bearing fault diagnosis via machine learning based methods consists of two steps, feature transformation and fault recognition [6]. At the first step, feature transformation is performed by using conventional signal processing methods to extract representative features in the time and frequency domains from raw signals simultaneously, such as fast Fourier transform (FFT) [7], short-time Fourier transform (STFT) [8], empirical mode decomposition (EMD) [9], and Hilbert-Huang transform (HHT) [10]. Based on transformed features, fault recognition is further applied to evaluate the health condition of bearings through effective methods, such as support vector machine (SVM) [10], k-nearest neighbor (kNN) [11], and artificial neural network (ANN) [12]. Numerous studies addressing bearing fault diagnosis by using machine learning based methods have been reported. Samanta et al. [13] utilized statistical time-domain features including mean, skewness, derivative, and so on to characterize the health condition of bearings and applied ANN and SVM to recognize bearing faults individually. Deng et al. [14] proposed two SVM based methods with extracted time and frequency features for the real-time sensor fault detection. Lei et al. [15] applied the EMD and wavelet packet transform (WPT) to representing characteristics of slight rub faults on heavy oil catalytic cracking units and identified the types of faults via a developed ANN automatically. Haddad et al. [7] introduced a fault diagnosis method by considering a combination of FFT and STFT as well as linear discriminant analysis (LDA) to achieve satisfactory recognition results. Although reported machine learning based methods can be applied to bearing fault diagnosis, their performance drops significantly due to increased external noises and interference caused by data collection. Moreover, it is computationally expensive to realize bearing fault diagnosis via the reported methods due to the massive datasets.

Recent deep learning methods have demonstrated the capability of better addressing bearing fault diagnosis with improved recognition accuracy. Convolutional neural networks (CNNs) [16], a milestone network architecture targeting com-

This work was supported in part by the National Natural Science Foundation of China under Grant 72101147 and in part by Shanghai Pujiang Talent Program under Grant 21PJ1405500.

Ruixian Li is with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai, China, e-mail: liruixian@sjtu.edu.cn.

Li Zhuang *(Corresponding author)* is with the School of Data Science, City University of Hong Kong, China, e-mail: lizhuanglily@gmail.com.

Yongxiang Li *(Corresponding author)* is with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai, China, e-mail: yongxiangli@sjtu.edu.cn.

Changqing Shen is with the School of Rail Transportation, Soochow University, Suzhou, China, e-mail: cqshen@suda.edu.cn.

puter vision based tasks, and the extensions of CNN have been widely applied in intelligent bearing fault diagnosis[17]. Quite a few pioneer studies using CNNs directly for diagnosing bearing faults have been presented. Ince et al. [18] developed a fast and accurate fault-detection system based on 1-D CNN to classify input signals acquired from the motor current. Sun et al. [19] presented a discriminative convolutional method to extract features from raw vibration data of induction motors, and applied SVM to fault diagnosis based on extracted features. Wang et al. [20] applied a combination of a CNN and a hidden Markov model to identify bearing faults.

However, diagnosing bearing faults by applying deep learning methods directly to raw signals is time-consuming, computationally expensive, and has a higher probability of encountering false diagnoses [21]. Therefore, more advanced studies addressing complicated bearing fault diagnosis via deep learning methods have been reported, which consisted of the feature transformation for raw vibration data first and next to the decision-making [22]. Dong et al. [23] developed a framework for identifying bearing faults using the combination of CNN and deep belief network (DBN) based on transformed features via STFT. Zhu et al. [24] applied STFT to feature transform of raw signals and proposed a novel capsule network with an inception block to recognize the category of faults. Ding et al. [25] presented a CNN based spindle bearing fault diagnosis method using wavelet packet energy images. Xu et al. [26] applied continuous wavelet transform (CWT) to transform raw signals to gray-scaled images, and next developed a LeNet-5 based model to extract multi-level features automatically. Based on extracted features, an ensemble of random forest classifiers was utilized to recognize bearing faults. Jiang et al. [27] introduced a spectral kurtosis based filtering approach to denoise for raw data first. The denoised data was transformed into images and classified via a developed CNN. Chen et al. [28] investigated the second-order cyclostationary behavior of vibration data, and conducted the fault recognition via a CNN model. It is observable that bearing fault recognition using vibration data has shifted machine learning based methods to methods developed by the interaction of feature transformation and CNNs. However, previously reported studies mainly focus on bearing fault diagnosis based on vibration data with noise-free conditions. Thus, it is valuable to explore more advanced deep learning methods for bearing fault diagnosis in noisy environments.

Based on deficiencies of the aforementioned methods for bearing fault diagnosis in strong noisy environments, this research develops an intelligent data-driven method with noise resistance capability. The developed method consists of two components, Scaled-RFB for feature transformation and Strip-CNN for fault recognition. The Scaled-RFB is extended from the Ramanujan filter banks (RFB) [29], [30] by generalizing the period interval, which targets extracting the hidden period of signals in noisy environments and has offered promising performance on medical datasets [31], [32]. On top of the characteristics of the RFB, the Scaled-RFB can significantly improve the identifiability of transformed features without increasing computational complexity. Next, a strip-convolutional neural network (Strip-CNN) is developed by applying the strip convolution to automatically recognize the health condition of bearings based on the transformed features via the Scaled-RFB. Six state-of-art benchmarking algorithms are compared in comparative experiments to validate the effectiveness of the proposed method. The novelty of this paper is to operate the Scaled-RFB as a feature for bearing fault recognition, and the main contributions of this study are summarized below:

(1) An extension of the RFB, named Scaled-RFB, is introduced to transform the hidden period and the temporal information of vibration data into RGB images in noisy environments.

(2) An intelligent data-driven method based on a novel CNN architecture with strip convolution is developed, making full use of the information transformed by the Scaled-RFB.

(3) When the signal-noise ratio (SNR) is -10 dB, the average $Precision$, $Recall$, and $F1$ scores of the proposed method on both datasets are at least 51.79%, 52.49%, and 52.47% higher than benchmarking methods, respectively.

The remaining parts of this paper are organized as follows. Section II develops the proposed data-driven method, including Scaled-RFB and Strip-CNN. Comprehensive case studies are conducted, and their computational results are analyzed in Section III. Finally, a conclusion of this study is provided in Section IV.

II. The Data-Driven Bearing Fault Diagnosis Method

The proposed data-driven method for bearing fault diagnosis consists of two components, the Scaled-RFB based feature transformation and the Strip-CNN based fault recognition. A schematic diagram of the proposed method is illustrated in Fig. 1.

To facilitate feature transformation, the sequence of the raw vibration data is first divided into several individual samples and further passed to Scaled-RFB to obtain a set of RFB features. RFB features can be visualized as RGB images and next fed into the Strip-CNN for fault recognition. Two convolutional modules with the same architecture employed in the developed Strip-CNN are utilized to extract high-level features from the set of RGB image inputs. The extracted features are further fused for fault recognition. Thus, different types of bearing faults can be diagnosed via the developed method.

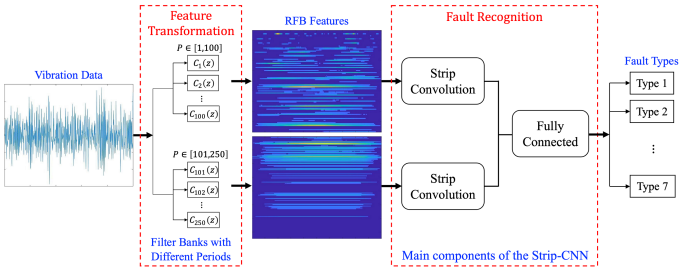


Fig. 1. Illustration of the intelligent data-driven bearing fault diagnosis method

A. The Scaled-RFB based Feature Transformation

Based on period estimation using high-dimensional dictionary representations of signals, the RFB [29], [30] is introduced to track periodicities of time series data. Moreover, it has been demonstrated that the RFB has outstanding performance on time series data analytics in noisy environments by comparing it with traditional period estimation techniques in signal processing, such as STFT [30]. Thus, the RFB has the potential to characterize the hidden periodicity of bearing faults and is considered a fundamental element for feature transformation in this study.

A discrete time series data $x\left(n\right)$ with length N is considered a periodic signal with period P, if P is the smallest positive integer as described in

$$x\left(n+P\right)=x\left(n\right), \tag{1}$$

where n represents the time point with an integer value.

As bearing components strike the fault at regular intervals, vibration data can be considered the periodic signal $x\left(n\right)$. To extract the periodicity of $x\left(n\right)$, the RFB is developed based on the pth Ramanujan sum [33] described in

$$c_p\left(n\right)=\sum_{k=1}^{p}e^{-j\frac{2\pi kn}{p}}=\sum_{k=1}^{p}W_p^{kn}\iff\gcd\left(k,p\right)=1, \tag{2}$$

where $\gcd$ represents the greatest common divisor. The e is the base of the natural logarithm, and the basis function $e^{-j\frac{2\pi}{p}}$ is denoted as W_p for simplification. The value of $c_p\left(n\right)$ equals the sum of basis functions W_p^{kn} that the ks are coprime to p.

Based on the Ramanujan sum, Ramanujan subspace representation (RSR) [34], [35] is introduced to represent $x(n)$ with multiple periodicities. The RSR removes redundant basis functions used in discrete Fourier transform (DFT) and only includes a minimal set of basis functions W_p^{kn}, where k and p are coprime. Thus, the signal $x\left(n\right)$ can be formulated as

$$\begin{aligned}x\left(n\right)&=\sum_{p=1}^{P_{max}}\sum_{k=0}^{p-1}a_{pk}W_p^{kn}&&\text{(DFT)}\\&=\sum_{p=1}^{P_{max}}\sum_{k=1}^{p}b_{pk}W_p^{kn}\tau_{\gcd(k,p)=1}&&\text{(RSR)},\end{aligned} \tag{3}$$

where

$$\tau_{\gcd(k,p)=1}=\begin{cases}1 & \gcd\left(k,p\right)=1\\0 & \text{otherwise.}\end{cases}$$

P_{max} indicates the maximum value of the period p. a_{pk} and b_{pk} are corresponding coefficients. W_p^{kn} denotes the basis function that has the periodicity with the constraint $1\leq p\leq P_{max}<N$.

As investigated in [36], a dictionary method provides a high dimensional representation for periodic signals. Based on RSR, a Ramanujan dictionary method is presented to represent periodic components of $x\left(n\right)$ via

$$\boldsymbol{x}=\boldsymbol{A}_{N\times\varphi(P_{max})}\boldsymbol{s}, \tag{4}$$

where $\boldsymbol{x}=\left[\begin{array}{cccc}x\left(0\right), & x\left(1\right), & \cdots, & x\left(N-1\right)\end{array}\right]^T$ consisting of specific samples of signal $x(n)$. $\boldsymbol{s}\in\mathbb{C}^{\varphi(P_{max})}$ denotes the coefficients b_{pk} with the number of $\varphi\left(P_{max}\right)$. Here $\varphi\left(\cdot\right)$ denotes the Euler totient function, and $\varphi\left(P_{max}\right)$ means the number of integers in $[1,P_{max}]$ coprime to P_{max}. $\boldsymbol{A}$ indicates a Ramanujan dictionary which consists of a set of basis functions W_p^{kn} in columns. Since the Ramanujan dictionary $\boldsymbol{A}$ and its coefficients matrix $\boldsymbol{s}$ are unique to $x(n)$, the hidden period of $x\left(n\right)$ can be identified from the index ℓ of its non-zero components s_ℓ based on the confirmed $\boldsymbol{s}$.

To facilitate the calculation of $\boldsymbol{s}$ in the Ramanujan dictionary in Eq. 4, the RFB is proposed based on the filter banks theory [37]. The RFB consists of P_{max} filters and the frequency response of the pth filter is described in

$$\begin{aligned}C_p\left(e^{j\omega}\right)&=\mathcal{F}\left[c_p\left(n\right)\right]\\&=2\pi\sum_{k=1}^{p}\delta\left(\omega-\tfrac{2\pi k}{p}\right)\tau_{\gcd(k,p)=1},\end{aligned} \tag{5}$$

where $\mathcal{F}\left[\cdot\right]$ denotes the Fourier transform and $\delta\left(\cdot\right)$ is the delta function. The range of ω is $[0,2\pi]$.

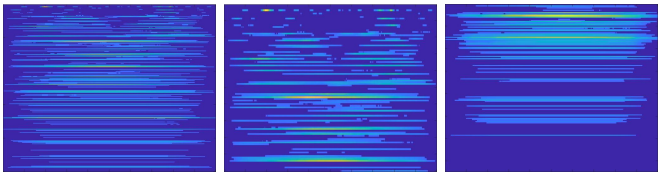
Since $c_p\left(n\right)$ is periodic, The response of $C_p\left(e^{j\omega}\right)$ is a line spectrum. According to the definition of $c_p\left(n\right)$, the line spectrum of $C_p\left(e^{j\omega}\right)$ is non-zero if the frequency equals $\frac{2\pi k_i}{p}$ and $\gcd\left(k_i,p\right)=1$. In other words, the output of $C_p\left(e^{j\omega}\right)$ will be non-zero if and only if the decomposition of $x(n)$ to the RSR has a component W_p. Thus, it is a valid estimation of the period of $x\left(n\right)$ by considering the least common multiple of indices of the Ramanujan filters. Due to the finite length N of $x(n)$, the filters in Eq. 5 will be represented with finite impulse response (FIR) filters as described in

$$C_p^{(l)}\left(z\right)=\sum_{n=0}^{pl-1}c_p\left(n\right)z^{-n}, \tag{6}$$

where each filter $c_p\left(n\right)$ is truncated to l consecutive periods. The time duration of the pth filter is equal to pl samples, which offer localization information.

The original format of the RFB is a matrix with the size of $P_{max}\times N$. To illustrate the transformed feature via the RFB in a more intuitive way, the output matrix of the RFB is depicted as an RGB image in this study. Since high-frequency components are more significant in bearing faults, it is reasonable to focus on components of signals with a small-scale period. To extract more high-frequency components and reduce computational complexity simultaneously, the first half of the periodic components is considered based on the characteristics of vibration data in this study. Therefore, P_{max} in Eq. 4 is set to 250.

Fig. 2(a) shows an example of the transformed feature via the RFB. The vertical and horizontal axes of such images indicate the period p and the time, respectively. Moreover, the colors of the horizontal line in the RGB image represent the intensity of output power of the pth filter. Thus, the transformed feature of raw vibration data is capable of compressing the original matrix and reducing the computation time in further fault recognition.



(a) Original RFB : $P \in [1, 250]$ (b) Scaled-RFB (Part 1): $P \in [1, 100]$ (c) Scaled-RFB (Part 2): $P \in [101, 250]$

Fig. 2. Transformed features via the original RFB and the Scaled-RFB

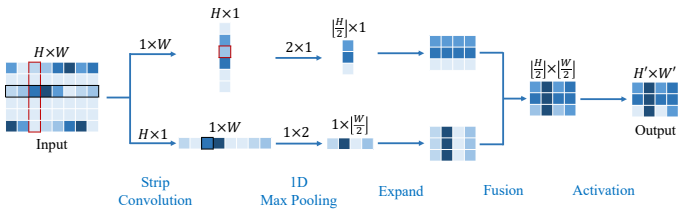


Fig. 3. Strip convolution

To strengthen high-frequency components of bearing faults in transformed features, a Scaled-RFB is proposed in this study. The Scaled-RFB can further enhance the feature transformation of bearing faults based on the RFB. In the Scaled-RFB, a new parameter P_{min} is introduced to represent the minimum value of the period, which targets refining the feature transformation performed via the RFB. A refined $x'(n)$ can be considered as a subset of $x(n)$ and is calculated via

$$x'(n) = \sum_{p=P_{min}}^{P_{max}} \sum_{k=1}^{p} b_{pk} W_p^{kn} \tau_{\gcd(k,p)=1} \quad . \tag{7}$$

A set of period components with the value of P, $P \in [P_{min}, P_{max}]$, will be extracted via the Scaled-RFB. The Scaled-RFB provides an opportunity to refine the transformed feature by using several RFB images with a set of different ranges of p. Based on the relatively short range of p, more informative features of bearing faults can be described via the Scaled-RFB. Moreover, RGB images transformed via the Scaled-RFB can be further processed separately, facilitating the parallelism of feature transformation.

In this study, two sub-intervals, including $1 \leq p_1 \leq 100$ and $101 \leq p_2 \leq 250$ are further set in the Scaled-RFB. The transformed features of bearing faults based on the Scaled-RFB are depicted as two RGB images as shown in Fig. 2(b) and (c). Fig. 2(b) and (c) provide more sparse information than Fig. 2(a), which makes adjacent periodic features easier to distinguish instead of being mixed (especially the yellow areas in the figures). This characteristic also facilitates the subsequent analysis using machine learning algorithms. Meanwhile, the information in each sub-interval of Scaled-RFB is processed separately and does not interfere with each other. In this way, it is more robust when the signal contains abnormal period components (this can explain why Fig. 2(b) and (c) look brighter). Thus, a shorter period range in the Scaled-RFB can provide profound information with the same size as the RGB image by comparing it with the RFB.

B. The Strip-CNN for Fault Recognition

Numerous state-of-art convolutional neural networks have been proposed and widely applied to computer vision based tasks, such as image identification [38], [16]. In reported studies, square convolution is commonly applied to extract left-right and up-down features simultaneously. However, RGB images obtained via the Scaled-RFB show more significant relationships along the horizontal of the image. To adapt to the characteristic of transformed features via the Scaled-RFB, a novel CNN architecture with strip convolution is developed for fault identification recognition.

1) Strip Convolution: A convolution layer includes several convolutional kernels named filters. The convolution is conducted by applying a filter times the corresponding region of the input iteratively. A traditional convolution kernel has a square size, such as 5×5, which targets extracting left-right and up-down information. However, strong dependencies along the horizontal of an RGB image transformed via the Scaled-RFB are shown in Fig. 2(b) and (c). They indicate that more attention should be paid to the left-right neighborhoods to extract the temporal information.

Suppose the two-dimensional input is $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$, where C denotes the number of channels, H and W are the spatial height and width, respectively. Inspired by strip pooling [39], strip shape kernels are deployed in the convolution of the CNN to effectively extract hidden information along the horizontal and vertical dimensions of the input. Such convolution is named strip convolution. The reason for considering the vertical dimension here is that the saliency of the feature is relative, and considering these two dimensions together can help us better build long-range dependencies on one of them.

Mathematically, use $\mathbf{k}$ to denote the strip kernel. And the size of horizontal and vertical strip kernels are denoted as S_h and S_v, and their total number are denoted as C_h and C_v, respectively. To describe the strip convolution more clearly, only the simplest case is considered here, that is, $S_h = 1 \times W$, $S_v = H \times 1$, and the stride is equal to the kernel size. As shown in Fig. 3, the input is fed into two parallel pathways, each of them contains a horizontal or vertical strip convolutional layer, which gives $\mathbf{y}^h \in \mathbb{R}^{C_h \times H}$ and $\mathbf{y}^v \in \mathbb{R}^{C_v \times W}$:

$$\begin{cases} y^h_{c_h,i} &= \sum_j \mathbf{X}_{i,j} \circ \mathbf{k}_{c_h,j} \\ y^v_{c_v,j} &= \sum_i \mathbf{X}_{i,j} \circ \mathbf{k}_{c_v,i}, \end{cases} \tag{8}$$

where $1 \leq c_h \leq C_h$, $1 \leq c_v \leq C_v$, $1 \leq i \leq H$, $1 \leq j \leq W$, and $\circ$ denotes the element-wise product. The strip convolutional layer is followed by a one-dimension (1D) max pooling layer with kernel size 2:

$$\begin{cases} \gamma^h_{c_h,i'} &= \max\left(y^h_{c_h,2i'-1}, y^h_{c_h,2i'}\right) \\ \gamma^v_{c_v,j'} &= \max\left(y^v_{c_v,2j'-1}, y^v_{c_v,2j'}\right), \end{cases} \tag{9}$$

where $1 \leq i' \leq \lfloor \frac{H}{2} \rfloor = H'$, $1 \leq j' \leq \lfloor \frac{W}{2} \rfloor = W'$, and $\lfloor \cdot \rfloor$ means round-down operation. If $C_h = C_v = C'$, then γ^h and γ^v will be combined to obtain $\gamma \in \mathbb{R}^{C' \times H' \times W'}$ with more useful global priors according to

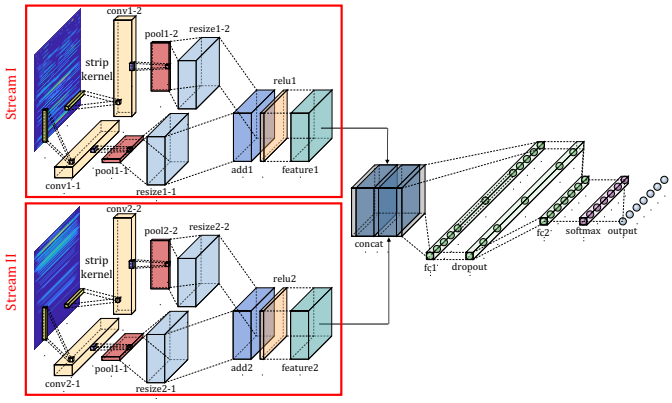


Fig. 4. Structure of the proposed Strip-CNN

$$\gamma_{c',i',j'} = \gamma^{h}_{c',i'} + \gamma^{v}_{c',j'}, \tag{10}$$

where $1 \leq c' \leq C'$, Finally, the output $\mathbf{z} \in \mathbb{R}^{C' \times H' \times W'}$ is computed as $\mathbf{z} = \sigma(\mathbf{y})$, where $\sigma(\cdot)$ is the Rectified Linear Unit (ReLU) activation function.

The kernel size S_h and S_v could be reduced proportionally to retain more information, then the convolution operation will obtain the two-dimension (2D) feature map. In this case, the 1D max pooling is still applied, so the output size is always $\lfloor \frac{H}{2} \rfloor \times \lfloor \frac{W}{2} \rfloor = H' \times W'$ for the input with size $H \times W$.

2) The Architecture of Strip-CNN: Based on strip convolution, an architecture described in Fig. 4 is developed for fault recognition. Since two RGB images are generated from one vibration data sample via the Scaled-RFB with period intervals of $[1, 100]$ and $[101, 250]$, two identical modules named Stream I and Stream II are developed in the Strip-CNN to process such inputs. Stream I and Stream II are constructed with two blocks. Each block comprises a horizontal or vertical strip convolutional layer, a 1D max pooling layer, and a resize layer. Then the outputs of two blocks are added together and through a ReLU function. After that, the outputs of Stream I and Stream II are further concatenated and fed into two fully connected layers, with a dropout layer in the middle of them. Finally, the output will input a softmax function to obtain the recognition results. The architecture of the Strip-CNN is summarized in Table I.

C. Performance Assessment

To validate the performance of the developed method on bearing fault diagnosis, a comprehensive metric, the $F1$ score, is utilized based on two evaluation metrics, including $Precision$ and $Recall$. The calculation of these three scores are described in

$$\begin{cases} Precision & = \frac{TP}{TP+FP} \\ Recall & = \frac{TP}{TP+FN} \\ F1 & = \frac{2 \times precision \times recall}{precision + recall}, \end{cases} \tag{11}$$

where TP denotes the true positive in recognizing the type of bearing faults and FP denotes the false recognition of such faults. FN represents the number of false negatives. For each type, its scores are calculated according to Eq.11, and then the average score of all types are calculated to get the final score. In this case, the value of $F1$ in the following may not equal the harmonic average of the $Precision$ and $Recall$.

III. Case Studies

In this section, a comprehensive case study is conducted to validate the performance of the proposed data-driven method on bearing fault diagnosis. Two datasets, a set of vibration data provided by our academic partner in Soochow University (SCU) and a set of open-source vibration data collected from the Case Western Reserve University (CWRU), are utilized in this study. Six benchmarking algorithms combine feature transformation via STFT and fault recognition using state-of-art CNNs are applied in comparative experiments.

A. Data Description and Preprocessing

1) SCU Dataset: One set of vibration data used in this study is collected from an experimental platform in SCU [40], as shown in Fig. 5. The platform consists of a drive motor, a bolt-and-nut loading system, a healthy bearing, a testing bearing, and a vibration acceleration sensor. Meanwhile, an adjustable mechanical loading system is deployed along the radial direction of the motor shaft. Although the running state of bearings under different working conditions can be simulated via such a system, we only consider the working condition with zero loads. To collect vibration data under the aforementioned working condition, the sampling frequency and the rotating speed are set to 10k Hz and 896.1 rpm, respectively. Meanwhile, the type of test bearing is 6205-2RS SKF.

Based on the experimental platform, a wire-electrode machine is deployed to generate defects with widths of 0.2 and 0.6 mm at three locations of test bearings, including inner race (IR), outer race (OR), and ball (BA). A description of the health condition of bearings in the SCU dataset is summarized in Table II, where the normal (No) bearing with no fault is also included.

2) CWRU Dataset: Another vibration data is the CWRU dataset, a publicly available dataset used to validate the proposed algorithms on bearing fault diagnosis. CWRU dataset is

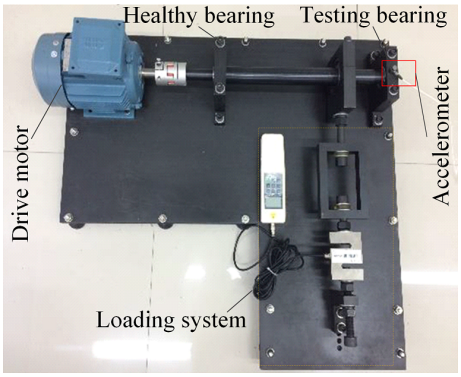


Fig. 5. Experimental platform in SCU

TABLE I
THE ARCHITECTURE OF STRIP-CNN

Layer		Type	Filters	Size / Stride	Input	Output
Stream I	1-1-1	Convolution	5	1×226 / 1×226	$537 \times 679 \times 3$	$537 \times 3 \times 5$
	2-1-1	Max Pooling		2×1 / 2×1	$537 \times 3 \times 5$	$268 \times 3 \times 5$
	3-1-1	Resize			$268 \times 3 \times 5$	$268 \times 339 \times 5$
	1-1-2	Convolution	5	179×1 / 179×1	$537 \times 679 \times 3$	$3 \times 679 \times 5$
	2-1-2	Max Pooling		1×2 / 1×2	$3 \times 679 \times 5$	$3 \times 339 \times 5$
	3-1-2	Resize			$3 \times 339 \times 5$	$268 \times 339 \times 5$
	4-1	Addition			$268 \times 339 \times 5$,$268 \times 339 \times 5$	$268 \times 339 \times 5$
	5-1	ReLU			$268 \times 339 \times 5$	$268 \times 339 \times 5$
Stream II	1-2-1	Convolution	5	1×226 / 1×226	$537 \times 679 \times 3$	$537 \times 3 \times 5$
	2-2-1	Max Pooling		2×1 / 22×1	$537 \times 3 \times 5$	$268 \times 3 \times 5$
	3-2-1	Resize			$268 \times 3 \times 5$	$268 \times 339 \times 5$
	1-2-2	Convolution	5	179×1 / 179×1	$537 \times 679 \times 3$	$3 \times 679 \times 5$
	2-2-2	Max Pooling		1×2 / 1×2	$3 \times 679 \times 5$	$3 \times 339 \times 5$
	3-2-2	Resize			$3 \times 339 \times 5$	$268 \times 339 \times 5$
	4-2	Addition			$268 \times 339 \times 5$,$268 \times 339 \times 5$	$268 \times 339 \times 5$
	5-2	ReLU			$268 \times 339 \times 5$	$268 \times 339 \times 5$
Both	6	Concatenation			$268 \times 339 \times 5$,$268 \times 339 \times 5$	$268 \times 339 \times 10$
	7	Fully Connected			$268 \times 339 \times 10$	$1 \times 1 \times 120$
	8	Dropout			$1 \times 1 \times 120$	$1 \times 1 \times 120$
	9	Fully Connected			$1 \times 1 \times 120$	$1 \times 1 \times 7$

TABLE II
SUMMARY OF BEARING FAULTS IN SCU DATASET

Health Types	Fault Width (mm)	Location of Bearing Faults
1	0.2	IR
2	0.2	IR
3	0.2	BA
4	0.6	BA
5	0.6	OR
6	0.6	OR
7	N.A.	No

TABLE III
DESCRIPTION OF SELECTED SUB-DATASETS IN CWRU DATASET

Health Types	Dataset number	Fault width (mm)	Location of Bearing Faults	Locations of Bearings
1	3001	0.72	IR	DE
2	118	0.18	BA	DE
3	197	0.36	OR centered	DE
4	290	0.36	BA	FE
5	298	0.18	OR orthogonal	FE
6	302	0.18	OR opposite	FE
7	97	N.A.	Normal	N.A.

collected by a motor drive system mounted with accelerometers that are deployed at drive end (DE), fan end (FE), and base plate (BP) with two sampling rates and four loads [41]. The test bearing in such a system is the same as that in the SCU system. In addition, the 6205-2RSJEM SKY bearing and the 6203-2RSJEM SKY bearing are deployed in DE and FE separately. Various bearing faults on different locations, including inner race (IR), outer race (OR), and ball (BA), can be collected through an electrical discharge process with different severities, including 0.18, 0.36, 0.54, and 0.72 mm. To validate the performance of the proposed method, a challenging sub-dataset recommended in [42] is utilized in this study. The selected sub-dataset sampled by the accelerometer deployed at DE with the frequency of 12k and zero loads are tagged with six types of health conditions of bearings. The detailed information of the CWRU sub-dataset is described in Table III.

3) Data Preprocessing: Since raw vibration data collected by the accelerometers is a sequence of long-term time series data, we first partition such raw data in the SCU and CWRU datasets into individual samples with a fixed length separately. The total number of data samples in the SCU dataset and CWRU dataset are both three hundred. All data samples in the SCU dataset are further divided into three sub-datasets, the training, validation, and test datasets with the ratio of $6:2:2$. And no overlap is allowed. Three sub-datasets in the CWRU dataset are divided with the same mechanism in the SCU dataset.

To consider more complex scenarios in practice, such as data in noisy environments, Gaussian noises are added into data samples in SCU and CWRU dataset separately to synthesize another six new groups of datasets with noise. Levels of noise are set with different values in such six datasets. A smaller SNR indicates a stronger noisy environment. Although noises will inevitably be included in both datasets in the experimental environments, the SNR of the collected signal is difficult to measure and control. Therefore, this paper has to consider the manually added Gaussian noises and does not consider the inherent noise of the original signal. In this way, the SNR used in this paper is slightly larger than its actual value, which will make the fault diagnosis more challenging than it should be. Values of the SNR and the corresponding energy of six groups of synthesized noisy datasets are $[10, 3, 0, -3, -7, -10]$ dB and $[0.1, 0.5, 1, 2, 5, 10]$ times of the original signal respectively, where the energy of signal $x(n)$ is defined as $E = \sum_{n=1}^{N} |x(n)|^2$.

B. Experiment Settings

The loss function of the proposed Strip-CNN is cross-entropy. To optimize the parameters of the Strip-CNN, stochastic gradient descent with momentum (SGDM) is utilized. The initial learning rate and the momentum are set to 0.000001 and 0.9, respectively. The learning rate is reduced by half every five epochs to prevent overfitting. The total number of

TABLE IV
RESULTS OF BEARING FAULT DIAGNOSIS VIA THE PROPOSED METHOD ON BOTH DATASETS

	SCU dataset							CWRU dataset						
Noise Level	Original	I	II	III	IV	V	VI	Original	I	II	III	IV	V	VI
SNR	N.A.	10	3	0	-3	-7	-10	N.A.	10	3	0	-3	-7	-10
Relative Energy of Noise	N.A.	0.1	0.5	1	2	5	10	N.A.	0.1	0.5	1	2	5	10
$Precision$	0.9882	0.9954	0.9789	0.9713	0.9264	0.8016	0.5525	1	1	0.9887	0.9953	0.9867	0.9236	0.7585
$Recall$	0.9881	0.9952	0.9786	0.9714	0.9262	0.8071	0.5524	1	1	0.9881	0.9952	0.9857	0.9214	0.7619
$F1$	0.9880	0.9953	0.9786	0.9712	0.9257	0.8027	0.5516	1	1	0.9881	0.9952	0.9857	0.9213	0.7580
Testing Time (s)	11.15	10.32	10.29	10.34	10.43	10.62	10.60	11.33	10.02	10.13	10.15	10.01	10.41	10.51

epoch for training the Strip-CNN is 200. The performance of the Strip-CNN is validated per two epochs. Meanwhile, the early stopping mechanism is applied to avoid overfitting in the training procedure, which will be triggered if the performance of the Strip-CNN on the validation dataset has not been improved after five epochs. In addition, a random seed is set to 42.75 in all experiments to ensure the same initialization condition.

The case study is conducted based on a computer with two Inter Xeon E5-2690 v2 CPUs and 128G memory, as well as an NVIDIA GeForce RTX 2080 Ti GPU with 11G memory. The implementation of our proposed method is based on the Windows 10 Professional operating system and the Matlab 2020b platform.

C. Results of Bearing Fault Diagnosis

The proposed method is applied to test datasets in the SCU dataset and CWRU dataset separately. The performance of bearing fault diagnosis via the proposed method is summarized in Table IV. The relative energy of noise indicates the ratio of the energy of noise to the energy of the original signal.

According to Table IV, the proposed method achieves a promising performance on the SCU dataset. Although the performance of bearing fault diagnosis on the SCU dataset decreases due to the increase of the noise level, the proposed method still demonstrates its ability to overcome challenges caused by such noises. It is observable that the performance on the SCU dataset is reduced by less than 7%, while the relative energy of noise is increased by 20 times by comparing noise level I with noise level IV. If the noise level continuously increases until the corresponding energy equals five times compared to the original data, the proposed method can still achieve relatively good performance with all three scores are around 0.8. Meanwhile, the proposed method offers the opportunity to obtain good performance on the challenging sub-dataset of CWRU with different levels of Gaussian noises. In contrast, three envelope-based methods introduced in the [42] fail to diagnose the bearing faults.

Furthermore, the proposed method is capable of obtaining a better performance using datasets with Gaussian noises. The reason is that the training procedure can bring some random errors. In addition, Gaussian noises may also provide extra energy to the practical components in the original vibration data. Subsequent such noises can be suppressed via feature transformation by using the Scaled-RFB. More significant characteristics of bearing faults can be shown in transformed features. Therefore, the proposed method can effectively transform original vibration data into RGB images and achieve satisfactory recognition results.

D. Ablation Studies

In this section, the ablation studies for detailed component analysis of the proposed Strip-CNN are conducted on both datasets. As shown in Table V, methods A1-A3 investigate the needed number of strip convolutional layers to balance the performance and the runtime cost of the proposed approach. Since the output size will be reduced to half of the input size after the strip convolutional layer, the channel number in the next layer will be twice the previous layer. Meanwhile, methods A4-A6 explore the impact of kernel size on the results. Then, to show the advantages of the proposed Strip-CNN over traditional 2D CNN, method A7 replaces the strip convolutional layer in A1 with an ordinary 2D convolutional layer. Note that all methods are trained with the same setting, and the Testing time refers to the time required for the method to process all four hundred and twenty samples in the test dataset.

Table VI depicts the methods A1 and A6 have the best performance on the SCU dataset, but the efficiency of method A7 is superb. However, the performance of method A5 is also competing on the CWRU dataset. To show the comprehensive performance of these methods, the average of the scores under different noise levels are calculated and shown in Table VII. Comparing method A1 with A7, the former performs better when measured by three scores. Although method A1

TABLE V
THE DESCRIPTION OF ABLATION STUDIES

Notation	Description			
	Layer Type	Layer Number	Kernel Size(s)	$\frac{\text{Input Size}}{\text{Kernel Size}}$
A1	Strip Convolution	1	21×1, 1×27	5^2
A2	Strip Convolution	2	21×1, 1×27	5^2
A3	Strip Convolution	3	21×1, 1×27	5^2
A4	Strip Convolution	1	537×1, 1×679	1^2
A5	Strip Convolution	1	59×1 , 1×75	3^2
A6	Strip Convolution	1	10×1 , 1×13	7^2
A7	2D convolution	1	5×5	N.A.

TABLE VI
RESULTS OF ABLATION STUDIES ON THE BOTH DATASETS

	Noise Level	SCU dataset							CWRU dataset						
		Original	I	II	III	IV	V	VI	Original	I	II	III	IV	V	VI
Precision	A1	0.9882	**0.9954**	**0.9789**	**0.9713**	**0.9264**	**0.8016**	0.5525	**1**	**1**	0.9887	**0.9953**	0.9867	**0.9236**	0.7585
	A2	0.9905	0.9736	0.9596	0.9471	0.9203	0.7505	0.5147	**1**	**1**	0.9886	0.9907	0.9795	0.8943	0.7558
	A3	0.9737	0.9696	0.9573	0.9272	0.8324	0.6732	0.4718	**1**	**1**	0.9887	0.9929	0.9689	0.8560	0.7082
	A4	0.9786	0.9638	0.9594	0.9560	0.8591	0.5925	0.3087	**1**	0.9977	0.9860	0.9881	0.9698	0.8120	0.5760
	A5	0.9858	0.9859	0.9668	0.9520	0.9130	0.7287	0.4825	**1**	**1**	**0.9930**	**0.9953**	0.9722	0.9105	0.7628
	A6	**0.9930**	0.9884	0.9744	0.9616	0.9199	0.7923	**0.5615**	**1**	**1**	**0.9930**	**0.9953**	**0.9977**	0.9168	**0.7825**
	A7	0.9725	0.9680	0.9624	0.9328	0.8617	0.7124	0.4967	**1**	0.9977	0.9884	0.9787	0.9697	0.8403	0.7294
Recall	A1	0.9881	**0.9952**	**0.9786**	**0.9714**	**0.9262**	**0.8071**	0.5524	**1**	**1**	0.9881	**0.9952**	0.9857	**0.9214**	0.7619
	A2	0.9905	0.9738	0.9595	0.9476	0.9214	0.7595	0.5167	**1**	**1**	0.9881	0.9905	0.9786	0.8857	0.7524
	A3	0.9738	0.9690	0.9571	0.9286	0.8381	0.6857	0.4786	**1**	**1**	0.9881	0.9929	0.9643	0.8524	0.7095
	A4	0.9786	0.9643	0.9595	0.9548	0.8619	0.6071	0.3214	**1**	0.9976	0.9857	0.9881	0.9690	0.8119	0.5738
	A5	0.9857	0.9857	0.9667	0.9524	0.9143	0.7429	0.5000	**1**	**1**	**0.9929**	**0.9952**	0.9667	0.9095	0.7619
	A6	**0.9929**	0.9881	0.9738	0.9619	0.9190	0.8024	**0.5690**	**1**	**1**	**0.9929**	**0.9952**	**0.9976**	0.9143	**0.7833**
	A7	0.9714	0.9667	0.9619	0.9333	0.8619	0.7143	0.4976	**1**	0.9976	0.9881	0.9786	0.9690	0.8357	0.7310
$F1$	A1	0.9880	**0.9953**	**0.9786**	**0.9712**	**0.9257**	**0.8027**	0.5516	**1**	**1**	0.9881	**0.9952**	0.9857	**0.9213**	0.7580
	A2	0.9905	0.9735	0.9590	0.9468	0.9205	0.7468	0.5133	**1**	**1**	0.9881	0.9905	0.9786	0.8834	0.7479
	A3	0.9736	0.9690	0.9571	0.9266	0.8335	0.6777	0.4746	**1**	**1**	0.9881	0.9929	0.9639	0.8492	0.7057
	A4	0.9783	0.9637	0.9593	0.9549	0.8590	0.5964	0.3137	**1**	0.9976	0.9857	0.9881	0.9691	0.8101	0.5724
	A5	0.9857	0.9857	0.9664	0.9521	0.9134	0.7316	0.4887	**1**	**1**	**0.9929**	**0.9952**	0.9666	0.9090	0.7589
	A6	**0.9929**	0.9881	0.9738	0.9616	0.9185	0.7916	**0.5631**	**1**	**1**	**0.9929**	**0.9952**	**0.9976**	0.9130	**0.7816**
	A7	0.9712	0.9658	0.9615	0.9325	0.8608	0.7088	0.4947	**1**	0.9976	0.9881	0.9786	0.9690	0.8364	0.7295
Testing Time (s)	A1	11.15	10.32	10.29	10.34	10.43	10.62	10.60	11.33	10.02	10.13	10.15	10.01	10.41	10.51
	A2	14.09	10.66	12.80	15.07	11.80	11.90	10.85	16.38	10.68	11.31	10.91	11.59	13.02	11.85
	A3	11.63	14.72	13.32	11.52	10.71	15.33	12.47	11.41	15.91	10.96	10.47	13.56	14.79	16.44
	A4	11.43	10.39	10.26	10.34	10.39	10.44	10.60	11.15	10.18	13.69	10.20	11.90	13.60	11.08
	A5	11.30	10.38	10.52	10.34	10.38	10.43	10.52	11.15	10.13	10.02	10.14	10.33	10.45	10.61
	A6	11.38	10.39	10.27	10.19	10.28	10.45	10.58	11.09	10.14	10.24	10.20	10.23	10.55	10.67
	A7	**10.69**	**9.73**	**9.762**	**9.82**	**9.65**	**9.88**	**9.95**	**10.45**	**9.61**	**9.66**	**9.46**	**9.69**	**9.75**	**9.95**

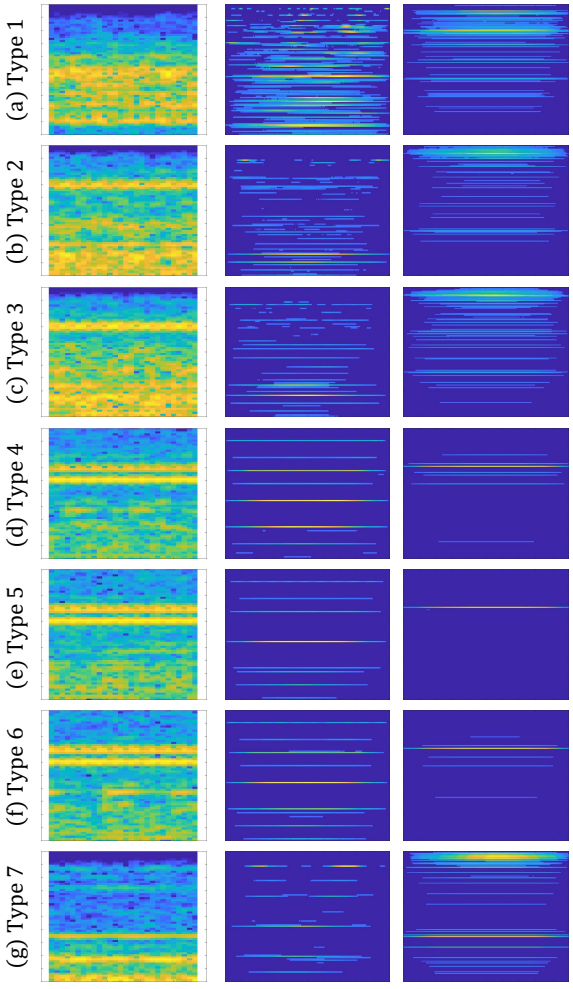


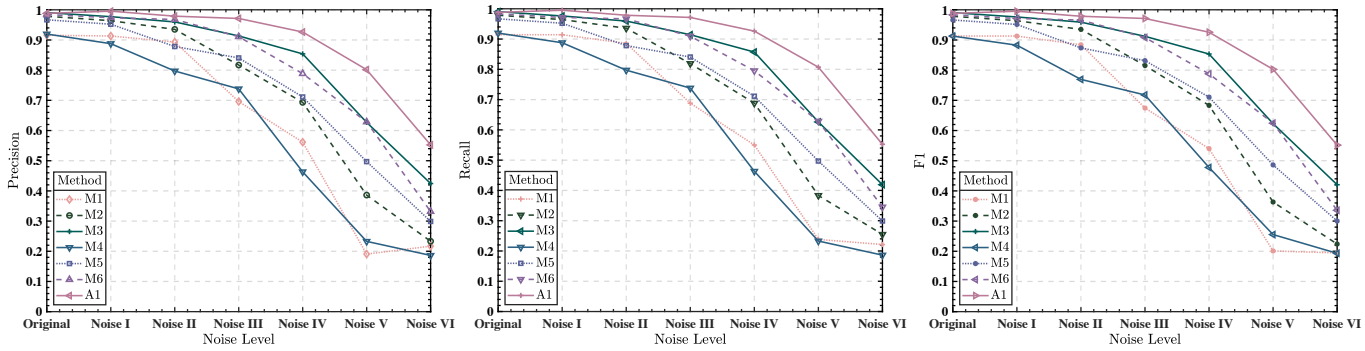
Fig. 6. Example of STFT (the left one) and Scaled-RFB feature images

has some shortcomings in efficiency, it brings more score improvements at the expense of time.

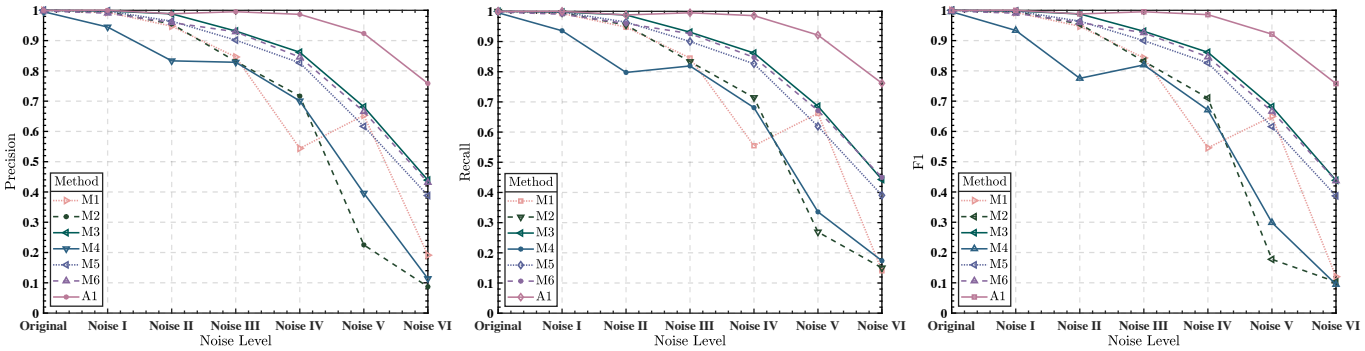
TABLE VII
THE AVERAGE RESULTS OF DIFFERENT NOISE LEVELS ON BOTH DATASETS

		SCU dataset		CWRU dataset	
		Score Value	Relative Value	Score Value	Relative Value
Precision	A1	0.8878	100%	0.9504	100%
	A2	0.8652	97.45%	0.9441	99.34%
	A3	0.8293	93.41%	0.9307	97.93%
	A4	0.8026	90.40%	0.9042	95.14%
	A5	0.8592	96.78%	0.9477	99.72%
	A6	0.8844	99.62%	0.9550	100.48%
	A7	0.8438	95.04%	0.9292	97.77%
Recall	A1	0.8884	100%	0.9503	100%
	A2	0.8670	97.59%	0.9422	99.15%
	A3	0.8330	93.76%	0.9296	97.82%
	A4	0.8068	90.81%	0.9037	95.10%
	A5	0.8640	97.25%	0.9466	99.61%
	A6	0.8867	99.81%	0.9548	100.47%
	A7	0.8439	94.99%	0.9286	97.72%
$F1$	A1	0.8876	100%	0.9498	100%
	A2	0.8643	97.37%	0.9412	99.09%
	A3	0.8303	93.54%	0.9285	97.76%
	A4	0.8036	90.54%	0.9033	95.10%
	A5	0.8605	96.95%	0.9461	99.61%
	A6	0.8842	99.62%	0.9543	100.47%
	A7	0.8422	94.89%	0.9285	97.76%
Testing Time (s)	A1	10.54	100%	10.37	100%
	A2	12.45	118.12%	12.25	118.13%
	A3	12.81	121.54%	13.36	128.83%
	A4	10.55	100.09%	11.69	112.73%
	A5	10.55	100.09%	10.40	100.29%
	A6	10.51	99.72%	10.45	100.77%
	A7	9.93	94.21%	9.80	94.50%

Meanwhile, method A1 surpasses A2-3 both in effectiveness



(a) Comparative results on the SCU dataset



(b) Comparative results on the CWRU dataset

Fig. 7. Comparative results on both datasets

and efficiency, this may be because the shallow information captured from the scaled-RFB features is already enough for diagnosis, and the deeper information captured by more strip convolution layers will aggravate overfitting. Comparing method A1 with A4-5, it seems that the scores will increase as the size of the kernel sizes decrease. However, the results of method A6 indicates that smaller kernel sizes may not lead to better results. Therefore, method A1 has better overall performance.

E. Comparative Experiments

To compare the ability of STFT and the Scaled-RFB in feature transformation, a set of transformed features based on examples of original vibration data in the CWRU dataset is shown in Fig. 6. According to Fig. 6, it is observable that the feature transformed via the Scaled-RFB depicts the more precise period information of the vibration data by comparing features transformed via STFT. As shown in features of type 2 and type 3 faults, the characteristic of the transformed feature via STFT (yellow parts in the left one of Fig. 6(b) and (c)) are almost the same, which indicates there is no significant difference between such two types of faults. However, the transformed feature via the Scaled-RFB for the same vibration data shows more information by comparing the middle one of Fig. 6 (b) and (c). A strong small period component (the upper part of the figure) appears intermittently in the middle one of Fig. 6(b), but not in the middle one of Fig. 6(c). At the same time, the two large period feature components appear in different locations (the lower area of the figure) in the two figures.

Moreover, six state-of-art deep learning algorithms with the feature transformation via STFT are considered benchmarking algorithms in this study, including the AlexNet [43], GoogLeNet [44], ResNet-18 [45], SqueezeNet [46], ShuffleNet [47], and EfficientNet [48]. Method M1-6 represents these benchmarks in turn for simplification. The AlexNet is a deep and wide CNN model and outperforms all traditional machine learning approaches. The GoogleNet is developed with a novel inception module based on the conventional architecture of CNN, and the module deployed in the GoogleNet allows a deeper network without an uncontrolled blow-up in the computational complexity. The deep residual network (ResNet) is introduced to address the degradation issue caused by the increase in the number of layers in the network. To consider the complexity of vibration data and computation source, the ResNet-18 with a relatively shallow structure is considered. The SqueezeNet is a minor CNN architecture that requires fewer parameters while maintains a competitive accuracy. It is mainly composed of fire modules constructed via that are squeeze convolution layers with only 1×1 filters. The outputs of fire modules are passed into an expand layer with a mix of 1×1 and 3×3 convolution filters. The ShuffleNet is an efficient model proposed for the application in resource-constrained scenarios. A residual module including a depth-wise convolution unit is introduced in the ShuffleNet, and a channel shuffle unit. Based on the observation that a better performance can be obtained by balancing network

TABLE VIII
COMPARATIVE RESULTS ON THE BOTH DATASETS

		SCU dataset							CWRU dataset						
	Noise Level	Original	I	II	III	IV	V	VI	Original	I	II	III	IV	V	VI
$Precision$	M1	0.9157	0.9132	0.8932	0.6960	0.5615	0.1907	0.2169	**1**	0.9932	0.9476	0.8459	0.5440	0.6532	0.1912
	M2	0.9785	0.9639	0.9351	0.8169	0.6932	0.3861	0.2332	**1**	0.9954	0.9548	0.8314	0.7167	0.2251	0.0871
	M3	**0.9906**	0.9772	0.9598	0.9135	0.8539	0.6270	0.4245	**1**	0.9977	0.9884	0.9315	0.8615	0.6812	0.4392
	M4	0.9197	0.8878	0.7972	0.7380	0.4632	0.2328	0.1863	0.9954	0.9448	0.8331	0.8284	0.7006	0.3959	0.1137
	M5	0.9669	0.9522	0.8784	0.8406	0.7112	0.4971	0.2989	**1**	0.9977	0.9641	0.9016	0.8266	0.6168	0.3882
	M6	0.9841	0.9738	0.9679	0.9109	0.7885	0.6286	0.3317	0.9977	0.9906	0.9597	0.9290	0.8454	0.6656	0.4315
	A1	0.9882	**0.9954**	**0.9789**	**0.9713**	**0.9264**	**0.8016**	**0.5525**	**1**	**1**	**0.9887**	**0.9953**	**0.9867**	**0.9236**	**0.7585**
$Recall$	M1	0.9143	0.9143	0.8857	0.6881	0.5500	0.2381	0.2214	**1**	0.9929	0.9476	0.8452	0.5548	0.6619	0.1405
	M2	0.9786	0.9643	0.9357	0.8190	0.6881	0.3833	0.2548	**1**	0.9952	0.9548	0.8333	0.7143	0.2690	0.1500
	M3	**0.9905**	0.9762	0.9595	0.9143	0.8571	0.6262	0.4190	**1**	0.9976	**0.9881**	0.9310	0.8619	0.6857	0.4429
	M4	0.9197	0.8878	0.7972	0.7380	0.4632	0.2328	0.1863	0.9952	0.9357	0.7976	0.8190	0.6810	0.3357	0.1738
	M5	0.9669	0.9522	0.8784	0.8406	0.7112	0.4971	0.2989	**1**	0.9976	0.9643	0.9000	0.8262	0.6190	0.3905
	M6	0.983	0.9714	0.9667	0.9095	0.7952	0.6286	0.3452	0.9976	0.9905	0.9595	0.9262	0.8476	0.6714	0.4500
	A1	0.9881	**0.9952**	**0.9786**	**0.9714**	**0.9262**	**0.8071**	**0.5524**	**1**	**1**	**0.9881**	**0.9952**	**0.9857**	**0.9214**	**0.7619**
$F1$	M1	0.9132	0.9129	0.8841	0.6746	0.5402	0.2010	0.1946	**1**	0.9929	0.9474	0.8427	0.5459	0.6493	0.1201
	M2	0.9784	0.9638	0.9350	0.8154	0.6833	0.3632	0.2239	**1**	0.9952	0.9548	0.8317	0.7104	0.1775	0.1030
	M3	**0.9905**	0.9762	0.9591	0.9121	0.8535	0.6228	0.4205	**1**	0.9976	**0.9881**	0.9307	0.8614	0.6819	0.4384
	M4	0.9129	0.8827	0.7694	0.7178	0.4776	0.2553	0.1932	0.9952	0.9336	0.7754	0.8196	0.6709	0.2990	0.0947
	M5	0.9662	0.9514	0.8737	0.8315	0.7107	0.4860	0.3001	**1**	0.9976	0.9642	0.8998	0.8262	0.6159	0.3872
	M6	0.9833	0.9711	0.9659	0.9081	0.7878	0.6245	0.3357	0.9976	0.9905	0.9595	0.9260	0.8459	0.6665	0.4367
	A1	0.9880	**0.9953**	**0.9786**	**0.9712**	**0.9257**	**0.8027**	**0.5516**	**1**	**1**	**0.9881**	**0.9952**	**0.9857**	**0.9213**	**0.7580**
Testing Time (s)	M1	**4.56**	**4.10**	**3.91**	**3.95**	**3.99**	**3.95**	**3.94**	**4.12**	**4.02**	**3.99**	**4.00**	**4.03**	**4.04**	**4.02**
	M2	9.67	9.74	9.36	9.43	9.43	9.41	9.50	9.64	9.66	9.67	9.63	9.70	9.58	9.60
	M3	5.66	6.07	5.10	5.10	5.12	5.05	5.10	5.16	5.05	5.10	5.10	5.06	5.08	5.05
	M4	6.33	6.16	6.52	6.58	6.17	6.11	6.11	6.13	6.08	6.05	6.05	6.03	60.8	6.08
	M5	6.23	5.21	5.57	5.64	5.66	5.53	5.60	5.53	5.33	5.29	5.33	5.44	5.35	5.29
	M6	22.73	20.22	20.02	20.09	20.16	20.09	20.08	22.19	20.24	20.29	20.17	20.01	19.99	20.12
	A1	11.15	10.32	10.29	10.34	10.43	10.62	10.60	11.33	10.02	10.13	10.15	10.01	10.41	10.51

depth, width, and resolution, the EfficientNet family has been proposed to minimize the main weaknesses in the existing CNN technologies. The baseline network EfficientNet-B0 already outperforms ResNet-50, which is designed by neural architecture search.

All benchmarking algorithms are initialized with weights pre-trained on the ImageNet, a public image dataset for visual-based studies [49]. The initial learning rate used to train benchmarking algorithms are set to 0.001, and the rest hyperparameters are kept the same as those used to train the Strip-CNN.

Results of benchmark algorithms on bearing fault diagnosis based on SCU and CWRU datasets are summarized in Table VIII. Table VIII indicates that in a total of fourteen experiments, method A1 (Strip-CNN) achieves the highest $Precision$, $Recall$ and $F1$ scores 13 times. Besides, the method M3 (ResNet-18) performs best in the other case. Method A1 lags behind M3 by about 0.2% when there is no noise on the SCU dataset, but this disadvantage will be reversed when the noise occurs. At noise level VI, the average $Precision$, $Recall$, and $F1$ scores of the proposed method on both datasets are 51.79%, 52.49%, and 52.47% higher than M3, respectively. Although the testing time consumed by method M3 is only about half of A1, method A1 still has a relatively high time efficiency considering that the time is the total time for processing four hundred and twenty samples.

Moreover, Fig. 7 describes the overall performance of three scores via different methods on different levels of noise for two datasets.

According to Fig. 7(a), it is observable that when measuring the relative merits of methods by three scores, almost identical conclusions could be made. The scores of Methods M1 (AlexNet) and M4 (SqueezeNet) are a bit low when there is no noise, but others are very similar. As noise levels increase, scores of different methods decline in similar patterns, but the gap widens. When the noise level exceeds II, comparative results demonstrate the superiority of method A1 over M1-6, which makes the proposed method more suitable for bearing fault diagnosis in noisy environments. It is worth noting that although method M6 (EfficientNet) outperforms M3 on the ImageNet dataset, it performs worse on the SCU dataset. This phenomenon attests to the need to design networks specifically for bearing fault diagnosis in noisy environments.

Fig. 7(b) further validate the performance of the proposed method on the challenging CWRU sub-dataset. The relative advantages of the benchmark methods are a little different-between the two datasets, but method A1 is still in the first echelon. As the noise level increases, the scores of method A1 decrease significantly more slowly than others, so its advantages increase with the noise level. Especially at noise level III and above, where more noise than valid data can be found, the results indicate that the proposed method is more robust to noise and performs better at challenging scenarios.

To further demonstrate the capability of the proposed method in bearing fault diagnosis in noisy environments, t-SNE [50] is utilized to visualize features extracted via different methods. Based on t-SNE, features from the same cluster are shown with the same color, and better feature extraction can be validated if there are well-separated clusters. Fig. 8 and Fig. 9 show cluster results of original features of raw data and high-

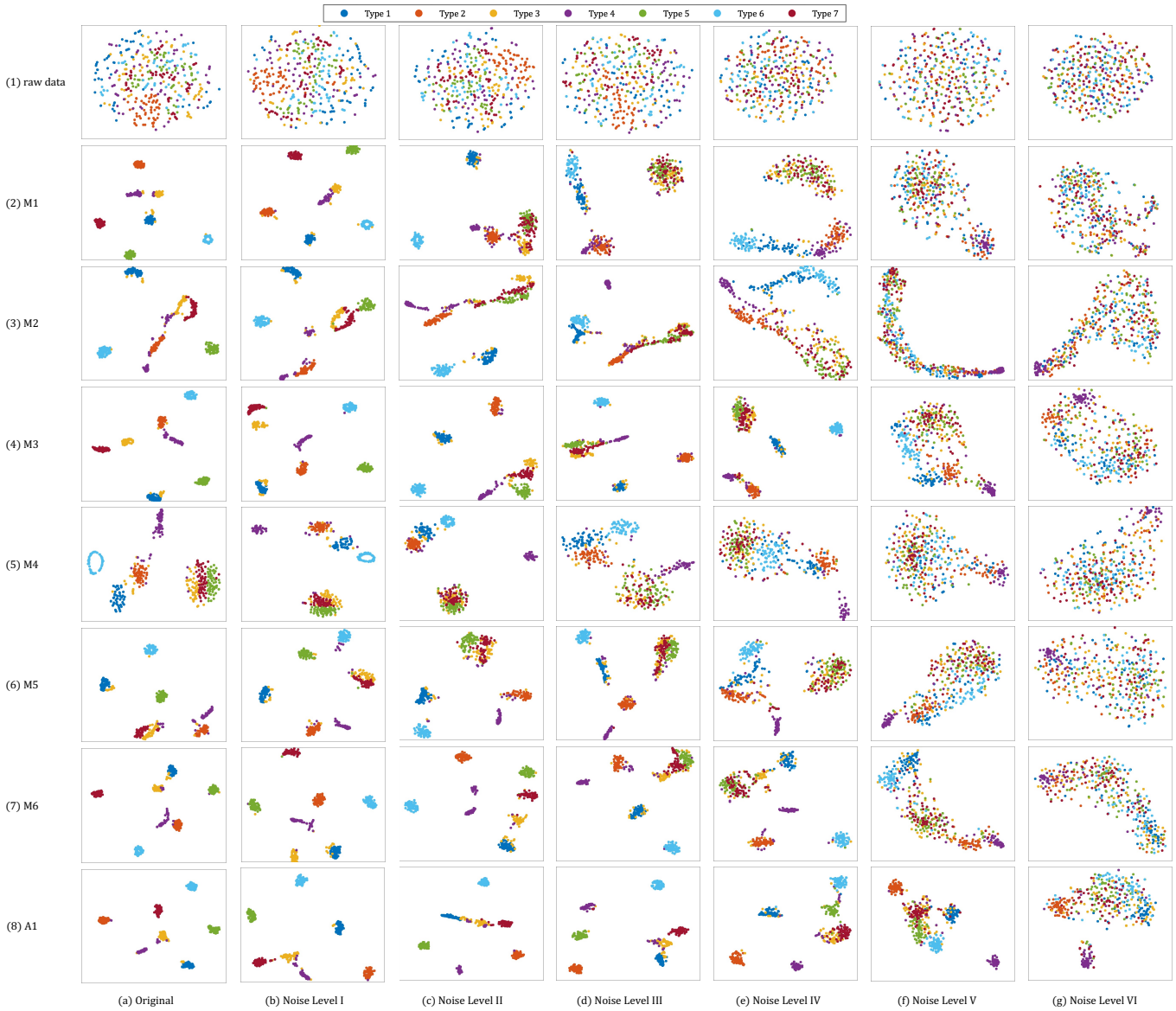


Fig. 8. Visualization of primal features and high-level features learned by models on the SCU dataset

level features obtained via different methods, respectively.

According to Fig. 8(a), it indicates that the original data of the SCU dataset is complicated, and bearing fault type 3 is the most challenging fault to recognize. All methods fail to recognize fault type 3 from fault type 1 perfectly. Fig. 8(a) also shows that method M4 may misclassify fault types 3 from fault types 7. As the noise increases, method M1 and M5 (ShuffleNet) will make the same mistake at noise level II as shown in Fig. 8(c). Fig. 8(e) demonstrates that method M2 (GoogLeNet) is almost eliminated at noise level IV. Finally, when the noise reaches level VI, as shown in Fig. 8(g), method A1 can still separate most type 4 samples from others.

The situation on the CWRU dataset is almost similar. As shown in Fig. 9(a)-(e), samples from fault types 1 and 7 can be easily distinguished from others in the raw data when noise is level IV and below. At the same time, Fig. 9(b) indicates that the most challenging task on the CWRU dataset is to separate samples marked as fault types 5 and 6, which the method M6 fails to fulfill. According to Fig. 9(c), sample features of fault types 5 and 6 extracted by method M1-2 and M4-5 are mixed at noise level II. Fig. 9(d) shows that method M3 is unable to separate these two kinds of samples when the noise reaches level III, while method A1 can still give a clear classification result. Finally, as shown in Fig. 9(g), only method A1 has noticeable clustering features at noise level VI, while other methods almost have no clusters.

IV. Conclusions

A data-driven method for automatically diagnose bearing faults in noisy environments based on vibration signals was developed in this study. The proposed method was developed with two stages, feature transformation and fault recognition. At the first stage, the Scaled-RFB was introduced to transform raw vibration data into RGB images in various noisy environments. Next, the Strip-CNN was developed to recognize

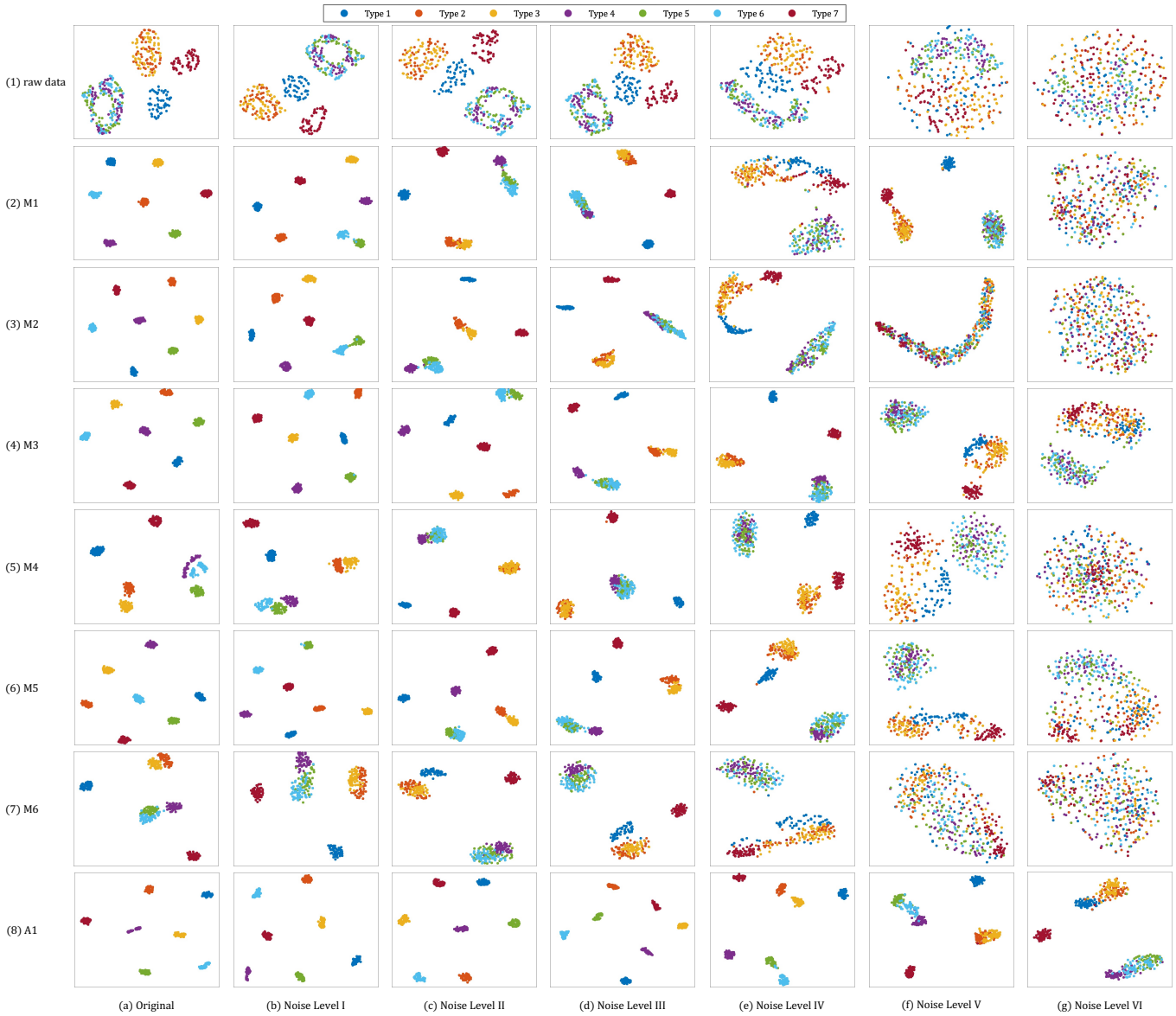


Fig. 9. Visualization of primal features and high-level features learned by models on the CWRU dataset

bearing faults based on transformed features via the Scaled-RFB. In comparative experiments, the proposed method was compared with the combination of six state-of-art benchmark algorithms with the STFT feature. All methods are validated with two datasets at seven different levels of Gaussian noise. According to comparative results, the proposed method can offer more accurate and robust diagnosis results than benchmark algorithms. Thus, we can conclude that the proposed method is more suitable for bearing fault diagnosis, especially in noisy environments.

In order to better remove Gaussian noise, this paper ignores the modulation effect of non-Gaussian noises on the signal, such as randomly distributed large impulses, so further works are needed to exploit the performance of the Scaled-RFB on signals containing heavy non-Gaussian noises. At the same time, Strip-CNN is designed based on characteristics of the scaled-RFB features under constant working conditions, so our future work will also explore its fault recognition capability under variable working conditions.

REFERENCES

[1] S. Gao, Q. Wang, and Y. Zhang, "Rolling bearing fault diagnosis based on CEEMDAN and refined composite multiscale fuzzy entropy," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–8, 2021.

[2] Y. Li, H. Zhao, W. Fan, and C. Shen, "Extended noise resistant correlation method for period estimation of pseudoperiodic signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[3] B. Li, B. Tang, L. Deng, and M. Zhao, "Self-attention ConvLSTM and its application in RUL prediction of rolling bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[4] R. Guo, Y. Wang, H. Zhang, and G. Zhang, "Remaining useful life prediction for rolling bearings using EMD-RISI-LSTM," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[5] Y. Hao, L. Song, L. Cui, and H. Wang, "A three-dimensional geometric features-based SCA algorithm for compound faults diagnosis," *Measurement*, vol. 134, pp. 480–491, 2019.

[6] T. Pan, J. Chen, Z. Zhou, C. Wang, and S. He, "A novel deep learning network via multiscale inner product with locally connected feature extraction for intelligent fault detection," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5119–5128, 2019.

[7] R. Z. Haddad, C. A. Lopez, J. Pons-Llinares, J. Antonino-Daviu, and E. G. Strangas, "Outer race bearing fault detection in induction machines using stator current signals," in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, 2015, Conference Proceedings, pp. 801–808.

[8] R. Liu, B. Yang, X. Zhang, S. Wang, and X. Chen, "Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 75, pp. 345–370, 2016.

[9] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets-Empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, pp. 696–711, 2012.

[10] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on hilbert-huang transform, support vector machine, and regression," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 1, pp. 52–62, 2015.

[11] M. Amarnath, S. Arunav, H. Kumar, V. Sugumaran, and G. S. Raghvendra, "Fault diagnosis of helical gear box using large margin k-nearest neighbors classifier using sound signals," *arXiv preprint arXiv:1508.04734*, 2015.

[12] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. Massi Pavan, "A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks," *Renewable Energy*, vol. 90, pp. 501–512, 2016.

[13] B. Samanta and C. Nataraj, "Use of particle swarm optimization for machinery fault detection," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 2, pp. 308–316, 2009.

[14] F. Deng, S. Guo, R. Zhou, and J. Chen, "Sensor multifault diagnosis with improved support vector machines," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 2, pp. 1053–1063, 2017.

[15] Y. Lei, Z. He, and Y. Zi, "Application of an intelligent classification method to mechanical fault diagnosis," *Expert Systems with Applications*, vol. 36, no. 6, pp. 9941–9948, 2009.

[16] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.

[17] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.

[18] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067–7075, 2016.

[19] W. Sun, R. Zhao, R. Yan, S. Shao, and X. Chen, "Convolutional discriminative feature learning for induction motor fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1350–1359, 2017.

[20] S. Wang, J. Xiang, Y. Zhong, and Y. Zhou, "Convolutional neural network-based hidden Markov models for rolling element bearing fault identification," *Knowledge-Based Systems*, vol. 144, pp. 65–76, 2018.

[21] X.-W. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE access : practical innovations, open solutions*, vol. 2, pp. 514–525, 2014.

[22] H. Shao, H. Jiang, X. Li, and T. Liang, "Rolling bearing fault detection using continuous deep belief network with locally linear embedding," *Computers in Industry*, vol. 96, pp. 27–39, 2018.

[23] S. Dong, Z. Zhang, G. Wen, S. Dong, Z. Zhang, and G. Wen, "Design and application of unsupervised convolutional neural networks integrated with deep belief networks for mechanical fault diagnosis," in *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, 2017, Conference Proceedings, pp. 1–7.

[24] Z. Zhu, G. Peng, Y. Chen, and H. Gao, "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis," *Neurocomputing*, vol. 323, pp. 62–75, 2019.

[25] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 1926–1935, 2017.

[26] G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, p. 1088, 2019.

[27] Q. Jiang, F. Chang, and B. Sheng, "Bearing fault classification based on convolutional neural network in noise environment," *IEEE access : practical innovations, open solutions*, vol. 7, pp. 69 795–69 807, 2019.

[28] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on Cyclic Spectral Coherence and Convolutional Neural Networks," *Mechanical Systems and Signal Processing*, vol. 140, p. 106683, 2020.

[29] S. V. Tenneti and P. Vaidyanathan, "Ramanujan filter banks for estimation and tracking of periodicities," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, Conference Proceedings, pp. 3851–3855.

[30] P. Vaidyanathan and S. Tenneti, "Properties of Ramanujan filter banks," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, Conference Proceedings, pp. 2816–2820.

[31] S. V. Tenneti and P. Vaidyanathan, "Detecting tandem repeats in DNA using Ramanujan filter bank," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2016, Conference Proceedings, pp. 21–24.

[32] ——, "Detection of protein repeats using the ramanujan filter bank," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, Conference Proceedings, pp. 343–348.

[33] S. Ramanujan, "On certain trigonometrical sums and their applications in the theory of numbers," *Trans. Cambridge Philos. Soc*, vol. 22, no. 13, pp. 259–276, 1918.

[34] P. Vaidyanathan, "Ramanujan sums in the context of signal processing-Part I: Fundamentals," *IEEE transactions on signal processing*, vol. 62, no. 16, pp. 4145–4157, 2014.

[35] ——, "Ramanujan sums in the context of signal processing-Part II: FIR representations and applications," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4158–4172, 2014.

[36] P. Vaidyanathan and P. Pal, "The Farey-dictionary for sparse representation of periodic signals," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, Conference Proceedings, pp. 360–364.

[37] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Prentice-hall, 1995.

[38] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[39] Q. Hou, L. Zhang, M.-M. Cheng, and J. Feng, "Strip pooling: Rethinking spatial pooling for scene parsing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, Conference Proceedings, pp. 4003–4012.

[40] C. Shen, X. Wang, D. Wang, Y. Li, J. Zhu, and M. Gong, "Dynamic joint distribution alignment network for bearing fault diagnosis under variable working conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[41] "Case western reserve university bearing data center website."

[42] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64-65, pp. 100–131, 2015.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, Conference Proceedings, pp. 1–9.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, Conference Proceedings, pp. 770–778.

[46] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and¡ 0.5 MB model size," *arXiv preprint arXiv:1602.07360*, 2016.

[47] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2018, Conference Proceedings, pp. 6848–6856.

[48] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, Jun. 2019, pp. 6105–6114.

[49] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein, "Imagenet large

scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[50] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," in *International Conference on Machine Learning 2014*. PMLR, 2014, Conference Proceedings, pp. 647–655.

Ruixian Li received a B.S. degree in the Department of Industrial Engineering and Management from Shanghai Jiao Tong University, Shanghai, China in 2019, where he is currently working toward his Master degree. His research interests include signal processing and fault diagnosis.

Li Zhuang received her Ph.D. in Data Science from City University of Hong Kong, Hong Kong, China, in 2020. She is currently a postdoctoral fellow at the Department of Architecture and Civil Engineering, City University of Hong Kong. Her research interests focus on computer vision, machine learning, computational intelligence, and related applications in intelligent transportation, renewable energy, and the industry.

Yongxiang Li received his Ph.D. degree in data science from City University of Hong Kong in 2019. He was an Engineer at Kuang Chi Institute of Advanced Technology and a Research Assistant at City University of Hong Kong. Currently, he is an Assistant Professor in the Department of Industrial Engineering and Management and in the Chinese Institute for Quality Research at Shanghai Jiao Tong University, China. His research interests comprise both applied and theoretical aspects of data science integrated with domain knowledge, including computer experiment, quality, reliability, signal processing and statistical learning.

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 signal processing and fault diagnosis.