2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)
June 28-30, 2023. Seattle, Washington, USA

A Robust Wavelet-integrated Residual Network for Fault Diagnosis of Machines with Adversarial Training

Xiwei Li, Yaguo Lei*, Xiang Li, Bin Yang

Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System
Xi'an Jiaotong University
Xi'an 710049, China
E-mail: yaguolei@mail.xjtu.edu.cn

***Abstract*—In engineering applications, the performance of intelligent fault diagnosis models can be significantly impacted by noise interference in the signals. This paper aims to address this issue by analyzing the influence of noise interference on diagnosis models, focusing on the network structure and training methods. Based on the analysis findings, a wavelet-integrated residual network (WResNet) is proposed to improve the noise-robustness. WResNet integrates discrete wavelet transformation (DWT) into the residual network architecture to mitigate potential problems related to frequency aliasing caused by traditional down-sampling operations. By incorporating DWT, WResNet could reduce the impact of noise interference. In addition, a gradient-based adversarial training method is adopted for optimizing the loss function of WResNet. By minimizing the maximal risk for label-preserving fluctuations of input signals, adversarial training is able to enhance the stability of WResNet. The effectiveness of WResNet is validated by using the monitoring data from a motor with different signal-noise-ratio. The results show that compared with ResNet and the method that using wavelet transformation as a pre-processing step, WResNet is able to achieve higher diagnosis accuracy while owning better noise-robustness.**

***Keywords*—Intelligent fault diagnosis of machines, noise interference, discrete wavelet transformation, adversarial learning.**

I. INTRODUCTION

Currently, deep neural networks have been widely studied in the field of machine fault diagnosis since they are able to extract health information of the equipment from the monitoring signals directly [1, 2]. However, in real-world engineering scenarios, signals are often susceptible to noise disturbances, leading to performance degradation in diagnosis models [3-5]. Upon analyzing existing intelligent diagnosis models, we have identified certain deficiencies in their network structures and training processes that may contribute to poor noise robustness.

1) In terms of network structure, most existing intelligent diagnosis models lack the ability to effectively counteract noise interference during down-sampling operations. These operations, such as max pooling, average pooling, are commonly used to reduce the feature dimensions. However, these operations overlook the sampling theorem, leading to a phenomenon known as frequency aliasing [6]. This means that after these operations, the original signal's high-frequency components can interfere with the low-frequency band, further exacerbating the noise interference. To elaborate, let's consider a diagnosis model using max-pooling as the down-sampling operation. If no anti-aliasing filter is applied before the max-pooling operation, the noise components in the high-frequency band will pass through and contaminate the low-frequency band. As the signal progresses through multiple layers, the noise accumulates, degrading the accuracy of the diagnosis model.

2) In terms of network training, one crucial aspect that is often overlooked in most existing diagnosis models is the flatness property of the loss function during the training process. These models primarily focus on minimizing the loss function values on the training samples, aiming to achieve optimal performance. However, this myopic approach comes with a potential risk of overfitting. An overfitted diagnosis model is usually characterized by a non-flat loss function, even though it may reach minima on the training samples. As a consequence, even slight fluctuations in the input signals can cause drastic variations in the loss function values. When the signals are affected by noise, the diagnosis results of the model become unstable. Therefore, it can be argued that the poor noise robustness of a diagnosis model is partly attributable to the lack of flatness in its loss function [7, 8].

To address the aforementioned challenges, a novel wavelet-integrated residual network namely WResNet is proposed. WResNet tackles these issues by incorporating the discrete wavelet transformation (DWT) as a down-sampling operation and leveraging adversarial training to enhance noise-robustness. There are multiple wavelet-integrated residual blocks (WResBlocks) in the WResNet. Within each WResBlock, the features undergo a decomposition using DWT, effectively separating them into the low-frequency band and the high-frequency band without experiencing frequency aliasing [9]. Subsequently, the features in different frequency bands are processed independently within the WResBlock. This approach ensures that the down sampling operation maintains the integrity of the signal and minimizes noise interference. To optimize the loss function of WResNet, a gradient-based adversarial training strategy is employed. During training, adversarial noise is deliberately introduced into the input signals to challenge the diagnosis model [10]. This adversarial noise acts as a form of attack, forcing the model to become more robust against noise interference.

The rest of the paper is organized as follows. Section II introduced the proposed WResNet in detail. In Section III, the

This work was supported in part by National Key R&D Program of China (2022YFB3402100), and the National Science Fund for Distinguished Young Scholars of China under Grant 52025056.

noise-robustness of the WResNet is demonstrated through the noise simulation experiment on a dataset collected from a motor test rig. Finally, conclusions are drawn in Section IV.

II. Proposed Method

The proposed WResNet architecture primarily consists of several stacked WResBlocks, along with a classification sub-network. The overall structure of WResNet is visually represented in Fig. 1.

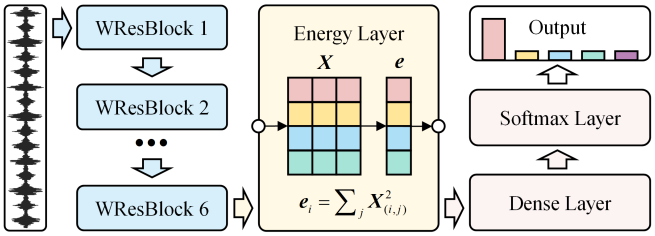


Fig. 1 Structure of WResNet.

A. Wavelet Integrated Residual block

To mitigate the impact of noise on the neural network, wavelet transformation is utilized as the down-sampling operation. In particular, DWT is integrated into the network architecture. Based on this, a WResBlock is constructed, as shown in Fig. 2.

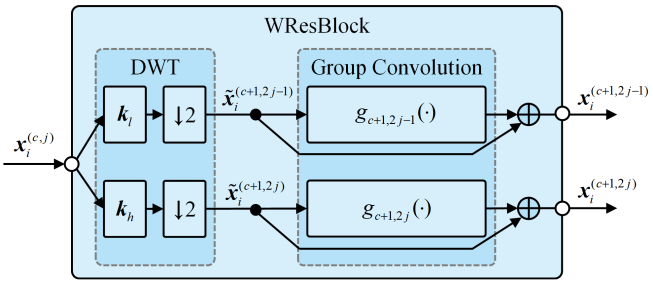


Fig. 2 Structure of WResBlock.

Stacked WResBlocks can be used to decompose a signal hierarchically like wavelet package transformation (WPT). For a given signal $\boldsymbol{x}_i \in \mathbb{R}^L$, define its decomposed components via c WResBlocks are denoted by $\boldsymbol{X}_i^c = [\boldsymbol{x}_i^{(c,1)}, \cdots, \boldsymbol{x}_i^{(c,2^{\wedge}c)}] \in \mathbb{R}^{[L/(2^{\wedge}c)] \times (2^{\wedge}c)}$. The calculation process of a WResBlock is formulated as follows.

A WResBlock halves the length of input features and doubles the number of input channels. There are several layers in a WResBlock. At first, each channel of the input is decomposed by DWT without information loss, i.e., given a low-pass filter $\boldsymbol{k}_l$ and a high-pass filter $\boldsymbol{k}_h$ of an orthogonal wavelet, $\boldsymbol{x}_i^{(c,j)}$ can be decomposed into

$$\begin{cases} \tilde{\boldsymbol{x}}_i^{(c+1,2j-1)} = (\boldsymbol{x}_i^{(c,j)} * \boldsymbol{k}_l) \downarrow 2 \\ \tilde{\boldsymbol{x}}_i^{(c+1,2j)} = (\boldsymbol{x}_i^{(c,j)} * \boldsymbol{k}_h) \downarrow 2 \end{cases}, \quad j = [1, 2, \cdots, 2^c] \tag{1}$$

where $\tilde{\boldsymbol{x}}_i^{(c+1,2j-1)}$ and $\tilde{\boldsymbol{x}}_i^{(c+1,2j)}$ are the low-frequency component and high-frequency component of $\boldsymbol{x}_i^{(c,j)}$ respectively. Symbol $\downarrow$ represents a down-sampling operation. After that, information enhancement and noise suppression are applied to $\tilde{\boldsymbol{x}}_i^{(c+1,2j-1)}$ and $\tilde{\boldsymbol{x}}_i^{(c+1,2j)}$ by $g_{c+1,2j-1}(\cdot)$ and $g_{c+1,2j}(\cdot)$ respectively. In this step, group convolution is adopted to carry out the independent transformation of different components synchronously. Borrowed the structure of Residual Network, identity mapping is also used in a WResBlock, which allows the network to converge better. The final output of a WResBlock with the input $\boldsymbol{x}_i^{(c,j)}$ is

$$\begin{cases} \boldsymbol{x}_i^{(c+1,2j-1)} = \tilde{\boldsymbol{x}}_i^{(c+1,2j-1)} + g_{c+1,2j-1}(\tilde{\boldsymbol{x}}_i^{(c+1,2j-1)}) \\ \boldsymbol{x}_i^{(c+1,2j)} = \tilde{\boldsymbol{x}}_i^{(c+1,2j)} + g_{c+1,2j}(\tilde{\boldsymbol{x}}_i^{(c+1,2j)}) \end{cases} \tag{2}$$

It should be noted that the proposed WResNet is different from the method that using DWT as a preprocessing step of training signals before a neural network [11, 12]. In the WResNet, multiresolution analysis is embedded into the network. The features in the high-frequency band and low-frequency band produced by each WResBlock are further transformed respectively. Therefore, the noise component in a certain frequency band would not influence the others. Especially, a WResBlock will shrink to DWT when the output of the residual branch equals 0, i.e.

$$\begin{cases} \boldsymbol{x}_i^{(c+1,2j-1)} = \tilde{\boldsymbol{x}}_i^{(c+1,2j-1)} \\ \boldsymbol{x}_i^{(c+1,2j)} = \tilde{\boldsymbol{x}}_i^{(c+1,2j)} \end{cases} \tag{3}$$

B. Classification Sub-network

The classification sub-network aims to identify the health state of $\boldsymbol{x}_i$ according to the features $\boldsymbol{X}_i^C$ produced by C stacked WResBlocks. In WPT, the energy of each frequency band can be used to reflect the health state. Since the proposed WResBlock can be regarded as an extension of DWT, we consider to use energies as features that can indicate the health state. The energy features of $\boldsymbol{x}_i$ are calculated by

$$\boldsymbol{e}_i = \left[\sum_{j=1}^{j=L/(2^C)} \boldsymbol{X}_{i,(1,j)}^C, \cdots, \sum_{j=1}^{j=L/(2^C)} \boldsymbol{X}_{i,(2^C,j)}^C \right]^{\mathrm{T}} \tag{4}$$

Then, after a dense layer and a softmax layer, the probability of prediction is generated by

$$\boldsymbol{p}_i = \mathrm{softmax}(\boldsymbol{w} \cdot \boldsymbol{e}_i + \boldsymbol{b}) \tag{5}$$

where $\boldsymbol{w}$ and $\boldsymbol{b}$ are the trainable parameters of the dense layer. Finally, the predicted label of $\boldsymbol{x}_i$ is obtained by

$$\hat{y}_i = \mathrm{argmax}(\boldsymbol{p}_i) \tag{6}$$

C. Adversarial Training

Adversarial training is adopted to improve the robustness of the diagnosis model by adding adversarial perturbations to the original signals during the training process. Adversarial

training seeks to minimize the adversarial loss of the input signals. The illustration of the adversarial training process is shown in Fig. 3.

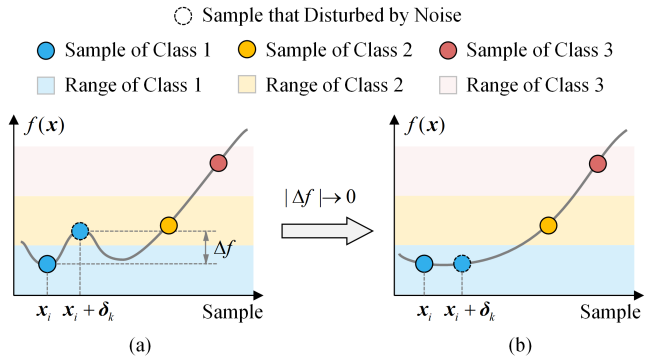


Fig. 3 Illustration of adversarial training process. (a) gradients of the loss function around the sample of class 1 change dramatically. (b) gradients of the loss function around the sample of class 1 become close to 0 after adversarial training.

Denote the whole WResNet as $f_{\theta}(\cdot)$. For a given perturbation $\boldsymbol{\delta}$ to the signal $\boldsymbol{x}_i$, the prediction result becomes

$$\hat{y}_i = \text{argmax}(f_{\theta}(\boldsymbol{x}_i + \boldsymbol{\delta})). \tag{7}$$

It is expected that the prediction of the model should not change when the perturbation $\boldsymbol{\delta}$ is added, as shown in Fig. 3 (b). Then the loss function of adversarial training can be expressed as

$$\min_{\theta} \sum_{i=1}^{N} \left[\max_{\|\boldsymbol{\delta}\| \le \varepsilon} \mathcal{L}\left(f_{\theta}(\boldsymbol{x}_i + \boldsymbol{\delta}), y_i\right) \right] / N \tag{8}$$

where N is the number of samples in one batch, y_i is the label of $\boldsymbol{x}_i$, and $\mathcal{L}$ is the cross-entropy loss function as follows.

$$\mathcal{L} = -\text{I}(y_i) \cdot \ln\left(f_{\theta}(\boldsymbol{x}_i + \boldsymbol{\delta})\right) \tag{9}$$

where $\text{I}(y) = [t_1, \cdots, t_Y]$, $t_i = 1$ when $i = y$ otherwise $t_i = 0$. Assume that the loss function is locally linear when $\|\boldsymbol{\delta}\| \le \varepsilon$, then $\boldsymbol{\delta}$ can be updated in each iteration by

$$\boldsymbol{\delta}_{k+1} = \Gamma_{\|\boldsymbol{\delta}\| \le \varepsilon}\left(\boldsymbol{\delta}_k + \lambda \times \frac{g(\boldsymbol{\delta}_k)}{\|g(\boldsymbol{\delta}_k)\|}\right) \tag{10}$$

where

$$g(\boldsymbol{\delta}_k) = \nabla_{\boldsymbol{\delta}} \sum_{i=1}^{N} \mathcal{L}\left(f_{\theta}(\boldsymbol{x}_i + \boldsymbol{\delta}), y_i\right) \tag{11}$$

is the gradient of the prediction loss with respect to $\boldsymbol{\delta}$, λ is a hypermeter that controls the change rate of $\boldsymbol{\delta}$, and $\Gamma_{\|\boldsymbol{\delta}\| \le \varepsilon}(\cdot)$ constraints the norm of $\boldsymbol{\delta}$ within ε. To get high robustness of the diagnosis model, gradient accumulation is performed in K iteration steps, and $\boldsymbol{\theta}$ of the model is updated only once for every K steps. Through gradient accumulation, the training process is equivalent to training the model by a larger virtual batch consisting of $\{\boldsymbol{x}_i + \boldsymbol{\delta}_k \mid i = 1, \cdots, N, k = 0, \cdots, K\}$. By employing adversarial training, the diagnosis model becomes less susceptible to interference caused by noise. The adversarial training process for WResNet is illustrated in Algorithm 1.

Algorithm 1 Pseudocode of Adversarial Training.

Initialization: training parameters $\boldsymbol{\theta}$; hyperparameters $\varepsilon, \lambda, \mu$;

for epoch $n = 1$ to N_{ep} **do**
  for mini-batch $i = 1$ to N **do**
    $\boldsymbol{\delta}_0 \leftarrow U(-\varepsilon, \varepsilon) / \sqrt{N_{\delta}}$
    $\boldsymbol{g}_{\theta,0} \leftarrow 0$
    for $k = 1$ to K **do**
      Accumulate gradient of $\boldsymbol{\theta}$
      $\boldsymbol{g}_{\theta,k} \leftarrow \boldsymbol{g}_{\theta,k-1} + \frac{1}{K} \nabla_{\theta} \sum_{i=1}^{N} \mathcal{L}\left(f_{\theta}(\boldsymbol{x}_i + \boldsymbol{\delta}_{k-1}), y_i\right)$
      Update $\boldsymbol{\delta}$ via gradient ascent
      $\boldsymbol{g}_{\delta} \leftarrow \nabla_{\delta} \sum_{i=1}^{N} \mathcal{L}\left(f_{\theta}(\boldsymbol{x}_i + \boldsymbol{\delta}_{k-1}), y_i\right)$
      $\boldsymbol{\delta}_k \leftarrow \Gamma_{\|\boldsymbol{\delta}\| \le \varepsilon}\left(\boldsymbol{\delta}_k + \lambda \times \frac{\boldsymbol{g}_{\delta}}{\|\boldsymbol{g}_{\delta}\|}\right)$
    end for
    $\boldsymbol{\theta} = \boldsymbol{\theta} - \mu \times \boldsymbol{g}_{\theta,K}$
  end for
end for
Return parameters $\boldsymbol{\theta}$

Compared to training the model by simply adding random noise to the original data, adversarial training can be seen as a form of hard example mining within the collection of noise-disturbed samples. Adversarial training focuses on identifying and utilizing the noise-disturbed samples that are particularly challenging to classify.

III. Experimental Investigation

A. Dataset Description

The motor fault simulation test rig is utilized to simulate various health states of the motor. This test rig consists primarily of two main components: the test motor and a brake system. Fig. 4 provides a visual representation of the test rig setup.

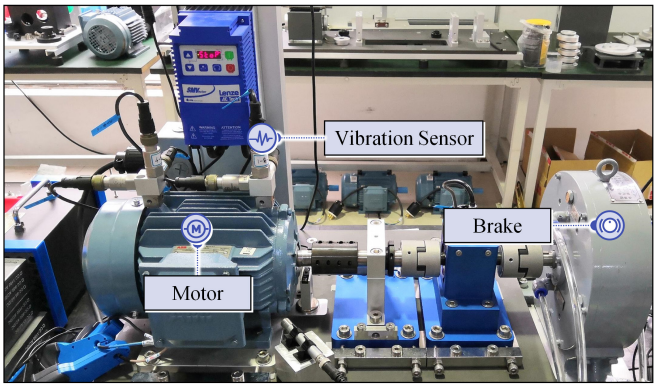


Fig. 4 Test rig for motor fault simulation.

In the experiment, a vibration sensor installed on the drive-end of the motor was used for data acquisition at a sampling frequency of 12.8 kHz. The data under 6 different health states were collected respectively, including the normal condition (NC), the bearing inner race fault (BIF), the bearing outer race fault (BOF), the bearing roller fault (BRF), the shaft misalignment (SM) and the shaft bending (SB). Finally, 540 samples were obtained under each health state, with each sample consisting of 1200 data points. The detail of the dataset is shown in Table 1.

TABLE I. DETAILS OF THE DATASET

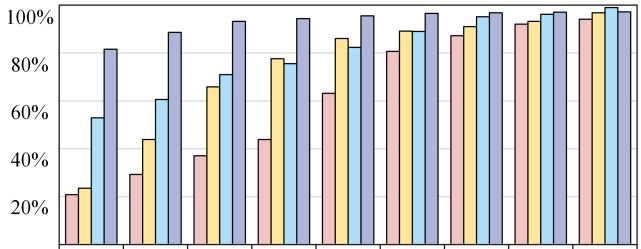
Health State	Rotation Speed (Hz)	Load (N)	Health State	Rotation Speed (Hz)	Load (N)
NC	15	0	BRF	15	0
BOF	25	10	SM	25	10
BIF	35		SB	35	
	45			45	

B. *Diagnosis Results*

To verify the effectiveness of the proposed WResNet, 20% of the samples are selected randomly for training the model and the rest were used for evaluating the model. Within the WResBlock, the Haar wavelet was implemented for the DWT operation due to its desirable properties of having the fewest coefficients and good orthogonality.

The proposed method is compared with some other diagnosis methods. For the sake of fair comparison, the number of decomposition levels for all methods is 6. Among these methods, Method 1 (M1) first decomposes the original signal by WPT to the 3rd level, then uses the features to obtain the final diagnosis through 3 convolution layers and a dense layer. Method 2 (M2) is a ResNet with 6 residual blocks. Method 3 (M3) has the same network structure as the proposed WResNet (M4), but it does not apply adversarial training during the training process.

To validate the noise-robustness of these methods, white Gaussian noise with different power is added to the original test samples. The signal-noise-ratio (SNR) of test samples is set to -4 dB, -2 dB, 0 dB, 2 dB, 4 dB, 6 dB, 8 dB, and 10 dB respectively. We define SNR = +∞ dB means no noise is added to the test samples. The mean diagnosis results under different SNR settings with ten trails are recorded in Fig. 5.



SNR	-4	-2	0	2	4	6	8	10	+∞
M1	20.83	29.31	37.11	43.81	63.18	80.69	87.11	91.98	94.03
M2	23.46	43.78	65.82	77.61	85.96	89.02	90.94	93.17	96.71
M3	52.91	60.59	70.93	75.49	82.25	88.95	95.12	96.16	98.87
M4	81.50	88.61	93.15	94.31	95.53	96.55	96.78	96.99	97.18

Fig. 5 Diagnosis accuracies of different methods.

It can be seen from Fig. 5 that Method 3 achieves the average diagnosis accuracy of 98.87% without adversarial training when no noise is added to the test samples. In this situation, Method 3 outperforms all other methods. Method 1 and Method 2 exhibit accuracies of 94.03% and 96.71% respectively when no noise is added, which are not significantly different from WResNet. However, when Gaussian noise is introduced to the test samples, the accuracies of Method 1 and Method 2 experience a dramatic decrease. Particularly, at an SNR of -4 dB, their accuracies drop below 25%. This indicates that noise interference significantly affects the diagnosis results of Method 1 and Method 2. In contrast, Method 3 demonstrates better noise robustness. Even when the SNR of the test samples is higher than 0 dB, the accuracy of Method 3 remains above 70%. These results suggest that the integration of DWT and independent feature extraction can attenuate the influence of noise. When WResNet is optimized using adversarial training, its noise robustness is significantly enhanced. Although it appears that the adversarial training causes WResNet to achieve slightly lower accuracy than Method 3 on the raw test samples, the accuracy of WResNet remains above 90% when the SNR of the test samples is higher than 0 dB. Even at an SNR of -4 dB, WResNet maintains an accuracy of over 81%. These findings indicate that adversarial training effectively stabilizes the performance of the diagnosis model and improves its noise robustness.

C. *Visualization*

To better show the noise-robustness of different methods, the features of signals with different SNR values are projected onto a two-dimensional space using the uniform manifold approximation and projection (UMAP) algorithm [13]. In Fig. 6, the features extracted from the proposed WResNet are compared with the features extracted from Method 2.

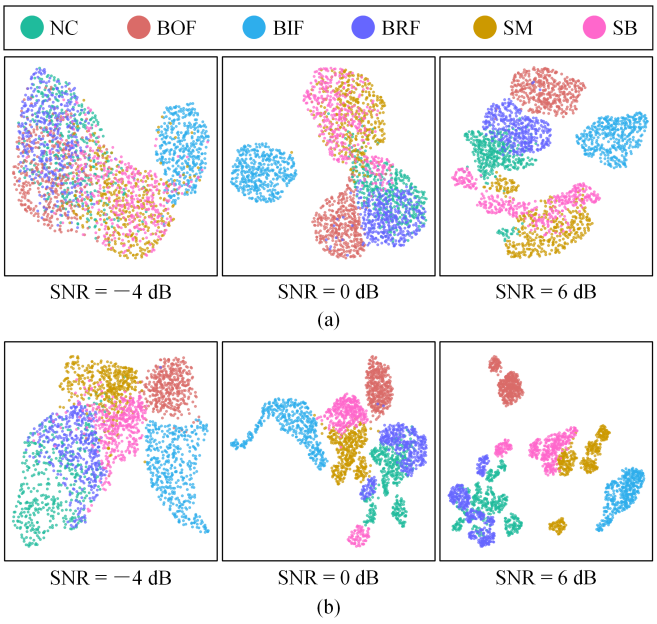


Fig. 6 Visualization of the features from (a) Method 2. (b) WResNet.

From Fig. 6(a), it is evident that Method 2 experiences a decline in diagnosis performance due to noise interference. As

the SNR of the test samples decreases, the boundaries between different health states become blurred. Particularly, at an SNR of -4 dB, the points representing the various health states are heavily mixed, except for those representing BIF. This indicates that Method 2 has lost its diagnostic capability under these conditions. In contrast, the proposed WResNet exhibits a more stable classification surface as the SNR of the test samples changes, as depicted in Fig. 6(b). Although the range of points representing the same health state widens with decreasing SNR, the clusters of points corresponding to different health states remain distinct, albeit being in close proximity to each other. This highlights the superior performance of WResNet compared to Method 2 in terms of preserving the separability of health states even under noisy conditions.

The excellent noise robustness exhibited by WResNet can be attributed to two key factors. Firstly, the integration of DWT into the convolutional neural network enhances the performance of the down-sampling process. This integration resolves the issue where high-frequency components affect the low-frequency components in methods such as max-pooling and other conventional down-sampling techniques. By employing DWT, WResNet effectively avoids this problem, resulting in improved noise resilience. Secondly, through adversarial training, WResNet is trained to produce stable outputs in the vicinity of the training samples, even when faced with noise interference. This approach enhances the model's ability to maintain consistent and accurate predictions despite the presence of noise. As a result, WResNet demonstrates enhanced noise robustness due to the stabilizing effect of adversarial training.

IV. Conclusion

This paper proposes a wavelet-integrated residual network named WResNet for machine fault diagnosis. WResNet is robust to noise interference. In WResNet, DWT is integrated to improve the performance of the down-sampling operation, so the potential risk of frequency aliasing caused by down-sampling can be eliminated. Additionally, WResNet adopts an adversarial training approach by using a gradient-based method. The objective of adversarial training is to enhance the model's ability to maintain stable output even when the signal is disturbed by noise.

WResNet is validated by the data from a test motor with a wide range of SNR. The results demonstrate that compared with some existing methods, WResNet is more robust to noise interference. This superior performance can be attributed to the integrated DWT operation and the way of adversarial training.

Furthermore, considering that the structure of WResNet can be seen as an extension of the WPT to some extent, it is plausible that WResNet may possess certain interpretable and valuable attributes. Future research endeavors will focus on exploring these attributes in greater detail.

Acknowledgment

This work was supported in part by National Key R&D Program of China (2022YFB3402100), and the National Science Fund for Distinguished Young Scholars of China under Grant 52025056.

References

[1] 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, Apr, 2020.

[2] S. Xing, Y. Lei, B. Yang, and N. Lu, "Adaptive Knowledge Transfer by Continual Weighted Updating of Filter Kernels for Few-shot Fault Diagnosis of Machines," *IEEE Trans. Ind. Electron.*, pp. 1-1, 2021.

[3] H. Shi and Y. Shang, "Initial Fault Diagnosis of Rolling Bearing Based on Second-Order Cyclic Autocorrelation and DCAE Combined With Transfer Learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-18, 2022.

[4] P. Peng, L. Ke, and J. Wang, "Fault Diagnosis of RV Reducer with Noise Interference," *Journal of Mechanical Engineering,* vol. 56, no. 1, pp. 30-36, 2020, 2020.

[5] S. Li, Z. Kang, and J. Tao, "Gear fault diagnosis based on information fusion and stacked de-noising auto-encoder," *Journal of Vibration and Shock,* vol. 38, no. 05, pp. 216-221, 2019.

[6] Q. Li, L. Shen, S. Guo, Z. Lai, and Ieee, "Wavelet Integrated CNNs for Noise-Robust Image Classification," *2020 Ieee/Cvf Conference on Computer Vision and Pattern Recognition*, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7243-7252, 2020.

[7] A. Shafahi, M. Najibi, A. Ghiasi, Z. Xu, J. Dickerson, C. Studer, L.S. Davis, G. Taylor, and T. Goldstein, "Adversarial training for free!," *arXiv preprint arXiv:1904.12843*, 2019.

[8] C. Zhu, Y. Cheng, Z. Gan, S. Sun, T. Goldstein, and J. Liu, "Freelb: Enhanced adversarial training for natural language understanding," *arXiv preprint arXiv:1909.11764*, 2019.

[9] T. Guo, T. Zhang, E. Lim, M. López-Benítez, F. Ma and L. Yu, "A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities," in *IEEE Access*, vol. 10, pp. 58869-58903, 2022.

[10] W. Zhao, S. Alwidian, and Q. H. Mahmoud, "Adversarial Training Methods for Deep Learning: A Systematic Review," *Algorithms*, vol. 15, no. 8, p. 283, Aug. 2022.

[11] N. Gharesi, M.M. Arefi, R. Razavi-Far, J. Zarei, and S. Yin, "A neuro-wavelet based approach for diagnosing bearing defects," *Advanced Engineering Informatics,* vol. 46, Oct, 2020.

[12] R. Chen, X. Huang, L. Yang, X. Xu, X. Zhang, and Y. Zhang, "Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform," *Computers in Industry,* vol. 106, pp. 48-59, Apr, 2019.

[13] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," *arXiv pre-print server*, 2020-09-18, 2020.