

Article

A Novel Dual-Channel Hybrid Attention Model for Wind Turbine Misalignment Fault Diagnosis

Tong Tong ¹, Xiang Liu ², Jia Zhang ², Dian Long ^{3,*}, Teng Fan ³ and Xiangyang Zheng ³

¹ China Huaneng Clean Energy Research Institute Co., Ltd., Beijing 100031, China; t_tong@qny.chng.com.cn

² Huaneng Qinghai Power Generation Co., Ltd., Xining 810008, China; ttongde@aliyun.com (X.L.); wangtangdi@aliyun.com (J.Z.)

³ School of Mechanical and Electronic Control Engineering, Beijing Jiaotong University, Beijing 100044, China; 24126031@bjtu.edu.cn (T.F.); 23121280@bjtu.edu.cn (X.Z.)

* Correspondence: 23126120@bjtu.edu.cn; Tel.: +86-17773262390

Abstract: Aiming at the problems of inaccurate feature extraction, slow convergence, and low diagnostic accuracy of wind turbine misalignment fault diagnosis under complex working conditions, this paper proposes an innovative diagnostic method based on two channels of U-Net and ResNet50. The model innovatively introduces the multi-head attention mechanism (MHA) in the jump connection of the U-Net architecture to form hybrid U-Net and optimizes the feature fusion process with dynamically learnable weights, which significantly enhances the ability to capture local details and key fault features. In the ResNet50 branch, deep global features are fully mined for extraction. To further achieve the co-optimization of global and local information, a shared hybrid expert attention (SHEA) module is proposed. This module achieves efficient integration of features by adaptively fusing the multi-scale local features output from the hybrid U-Net decoder with the deep global features extracted from the ResNet50 backbone network through a dynamic weighting and expert selection mechanism. The multi-scale features optimized by the SHEA module are fed into the classifier for fault type determination. The experimental results show that the method demonstrates excellent convergence speed and 99.64% classification accuracy under complex working conditions, providing an effective solution for the intelligent diagnosis of wind turbine misalignment faults.



Academic Editor: Hui Ma

Received: 27 March 2025

Revised: 22 April 2025

Accepted: 27 April 2025

Published: 29 April 2025

Citation: Tong, T.; Liu, X.; Zhang, J.; Long, D.; Fan, T.; Zheng, X. A Novel Dual-Channel Hybrid Attention Model for Wind Turbine Misalignment Fault Diagnosis. *Machines* **2025**, *13*, 368. <https://doi.org/10.3390/machines13050368>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: misalignment faults; multi-head attention mechanism; wind turbines; U-Net; ResNet50; shared hybrid expert attention module

1. Introduction

Since the beginning of the 21st century, issues such as global energy security, ecological environmental protection, and climate change have become increasingly severe. Wind energy, as a rapidly developing renewable energy technology in recent decades, is gradually demonstrating its immense potential and value. However, with the annual increase in wind turbine installations and their long-term operation, the number of faulty units has risen significantly, leading to a growing number of accidents and associated losses. The structure of a wind turbine is shown in Figure 1, including blades, gearboxes, couplings, generators, and so on. Common failures of wind turbines mainly include blade damage, gearbox wear, drive train misalignment failure, and generator failure. Among these, drivetrain misalignment refers to the misalignment of the rotor shaft system composed of the gearbox and generator, which are connected by a coupling. During the actual operation of wind turbines, factors such as installation errors, load-induced deformation,

and uneven settling of the turbine's foundation can lead to misalignment. Misalignment faults inevitably lead to turbine vibrations, compromising the reliability of the transmission and power generation systems. As a result, monitoring and diagnosing misalignment faults in transmission systems have become critical areas of research in the development of wind power technology.

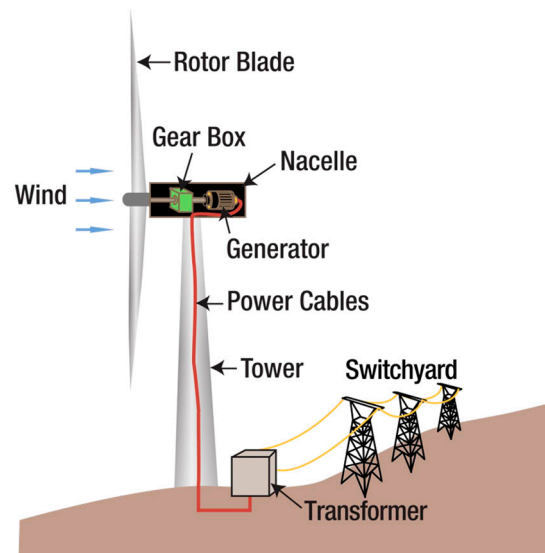


Figure 1. The structure of a wind turbine.

Fault diagnosis methods can be mainly divided into signal analysis methods, machine learning methods, and deep learning methods.

Signal analysis methods mainly include time-domain analysis methods, frequency-domain analysis methods, and time-frequency analysis methods. For example, Bajric et al. proposed a method for diagnosing wind turbine high-speed shaft gear faults based on discrete wavelet transform and time-synchronous averaging [1]. Yi et al. proposed an adaptive fault diagnosis method for rotating machinery based on the incomplete S-transform with maximum kurtosis [2]. Zhao et al. developed a scaled chirp transform method based on frequency-chirp synchronous compression for health monitoring and fault diagnosis of planetary gearboxes and bearings [3]. However, signal analysis methods predominantly depend on manual prior knowledge to extract frequencies and interpret spectra from the collected signals on-site, which imposes stringent requirements on the expertise of the personnel.

With the development of computer technology, machine learning-based fault diagnosis methods have become increasingly popular. For instance, Long et al. proposed a fault diagnosis method for wind turbines based on a cloud bat algorithm-kernel extreme learning machine [4]. Pu et al. developed a feature enhancement mapping method to minimize the intra-class distance of deep features in triaxial vibration signals and fed the fused triaxial features into an echo state network for fault classification [5]. Tang et al. proposed a novel fault diagnosis method for wind turbine transmission systems based on manifold learning and Shannon wavelet support vector machines [6]. Although these methods have achieved notable success, they are constrained by their inability to perform automatic feature extraction, and their classification algorithms exhibit limited learning capabilities, rendering them inadequate for addressing the demands of fault diagnosis in large-scale data scenarios.

In recent years, deep learning has been extensively applied in fault diagnosis owing to its robust capabilities in feature learning, multi-modal data fusion, and large-scale data

processing. Numerous neural network architectures have been developed to effectively handle complex monitoring data; for example, Zhao et al. proposed a Multi-scale Dynamic Graph Mutual Information Network (MDGMIN) for health monitoring of planetary bearings under unbalanced data [7]. Gao et al. introduced an interpretable wavelet basis unit convolutional network for mechanical fault diagnosis [8]. Zhao et al. developed a hierarchical health monitoring model called the Adaptive Threshold and Coordinate Attention Tree Heuristic Network (ATCATN) for monitoring the health of aero-engine bearings under strong background noise [9]. Zhao et al. developed a Multi-Perceptual Graph Convolutional Tree Embedded Network (MPGCTN) [10]. Xu et al. proposed a cross-modal fusion convolutional neural network for mechanical fault diagnosis [11]. Li et al. introduced a method for searching globally optimal hyperparameters for LSTMs using an additive attention serpentine optimization algorithm [12]. Mohammad-Alikhani et al. proposed a long short-term memory-regulated deep residual network for data-driven fault diagnosis in motors [13]. Tao et al. developed a deep neural network algorithm framework based on stacked autoencoders and Softmax regression for bearing fault diagnosis [14]. Zhao et al. proposed a hybrid deep autoencoder network for bearing fault detection and diagnosis [15]. Wang et al. introduced a fault diagnosis method based on a rolling bearing residual shrinkage network [16]. Wang et al. developed a method integrating an improved deep residual network and wavelet transform to address gearbox health detection problems [17]. Shi et al. proposed an intelligent fault diagnosis algorithm for rolling bearings based on a residual dilated pyramid network and a fully convolutional denoising autoencoder [18]. These research methods are able to automatically learn deep discriminative features from raw data but often exhibit slow convergence and low accuracy when faced with signals containing non-smooth and transient abrupt changes.

Since the vibration signals reflecting wind turbine misalignment faults are non-stationary due to the fluctuation of wind speed, in order to increase the convergence speed and further improve the diagnostic accuracy, this paper proposes a wind turbine misalignment fault diagnosis method based on U-Net and ResNet50 feature extractions.

The main innovations of this paper are as follows:

1. Putting forward a dual-branch feature fusion structure to improve model characterization capability.

The dual-branch feature fusion of U-Net and ResNet50 realizes the synergy and complementarity between local detail features and deep global features. The design of this structure enables the model to capture both local key features and global semantic information of wind turbine misalignment faults, which significantly improves the characterization ability and diagnosis accuracy of fault patterns.

2. Introducing the multi-head attention (MHA) mechanism to optimize the U-Net jump connection.

In the U-Net branch, multi-scale feature extraction and the gradual recovery of detail information are realized through the encoding–decoding structure, and the multi-head attention (MHA) mechanism is an attention mechanism that captures the relationship between different subspace features in an input sequence by computing multiple attention heads in parallel, which is innovatively introduced in the jump connection part to enhance the cross-layer feature interaction capability. Meanwhile, the dynamically learnable weight coefficients further optimize the feature fusion effect of the jump connection, which is called hybrid U-Net in this paper, so that the model can focus more on the key fault feature region, thus significantly improving the ability to capture local details and important features.

3. Proposing a shared hybrid expert attention (SHEA) module for efficient integration of global and local features.

In order to further optimize the feature expression ability of the two feature extraction paths, this paper proposes a SHEA module, which dynamically selects and weights the multi-scale local features output from the hybrid U-Net decoder and the deep global features extracted from the ResNet50 backbone network to achieve efficient fusion of multi-scale and multi-dimensional feature maps, which significantly improves the diagnostic accuracy and generalization ability of the model.

The experimental results show that the method in this paper exhibits an excellent fault recognition rate and a fast convergence speed under the complex working conditions of wind turbines.

The rest of the paper is organized as follows. Section 2 describes the proposed diagnostic method in detail. Section 3 presents the experimental design and results analysis. Section 4 summarizes the main conclusions of this study.

2. Methodology

2.1. Architecture Design

This paper proposes a fault diagnosis method for wind turbine misalignment based on dual-channel hybrid attention of U-Net and ResNet50, and its structure is shown in Figure 2.

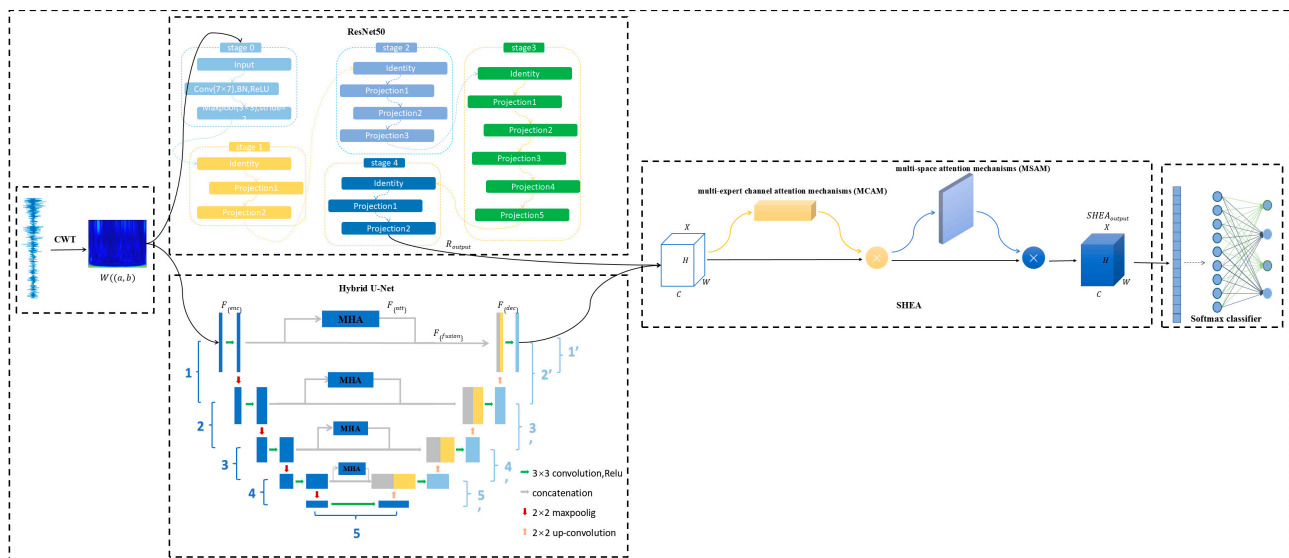


Figure 2. The structure of the dual-channel hybrid attention model.

First, for the time-varying and non-stationary characteristics of wind turbine vibration signals, continuous wavelet transform (CWT) is used to perform multi-scale time-frequency analysis to generate a two-dimensional time-frequency image $W(a, b)$, which provides rich time-frequency information for the feature learning of the subsequent deep learning model.

Then, $W(a, b)$ is fed in parallel to hybrid U-Net and ResNet50 for feature extraction. In the hybrid U-Net branch, multi-scale feature extraction with the gradual recovery of detailed information is realized through the encoding–decoding structure, and the MHA mechanism is introduced in the jump-joining part. In the ResNet50 branch, the residual module is used to realize deep feature mining, which can effectively alleviate the problem of gradient disappearance and improve the model’s ability to learn complex failure modes.

In order to further optimize the feature expression ability of the two feature extraction paths, this paper proposes the SHEA module to weigh the hybrid U-Net decoder output and ResNet50 output to obtain SHEA output.

Finally, the multi-scale and multi-dimensional feature maps optimized by the SHEA module are fused, and the intelligent identification of wind turbine misalignment faults is completed by the classifier.

2.2. Continuous Wavelet Transform

Traditional signal processing methods mainly focus on the spectral characteristics of the signal but often ignore the pattern of change in the signal over time. Time-frequency imaging, as an image representation method that integrates time and frequency information, can capture the dynamic features in the vibration signal more intuitively and improve the accuracy and efficiency of fault diagnosis. Common methods for converting vibration signals into time-frequency images include short-time Fourier transform, discrete wavelet transform, CWT, and Hilbert–Huang transform. Among them, CWT can automatically adjust the time resolution according to the frequency characteristics of the signal, provide fine detail features and extensive generalized information, make the time-frequency changes of the signal more intuitive, and make it easy to find abnormal or fault characteristics. The significant advantages shown in signal processing and analysis make CWT widely used. In this study, CWT is used to convert the original vibration signals of wind turbines into time-frequency images. For a given signal, the definition of CWT is expressed as shown in Equation (1) as follows:

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \cdot \psi^*\left(\frac{t-b}{a}\right) dt \quad (1)$$

where $x(t)$ represents the original signal, $\psi^*(t)$ is the complex conjugate of the wavelet function, and a and b are the scale parameter and translation parameter, respectively. The scale parameter a adjusts the oscillation frequency of the wavelet function by stretching or compressing it, while the translation parameter b shifts the position of the time window [19].

2.3. Hybrid U-Net

The U-Net network is a classical encoder–decoder architecture widely used in image segmentation and feature extraction tasks [20]. Its core idea is to realize multilevel feature extraction and reconstruction of images through the combination of contraction path (encoder) and expansion path (decoder). In order to further enhance the feature extraction capability of U-Net for time-frequency images, in this paper, the feature maps of the corresponding layers in the encoder are spliced with the feature maps of the decoder through jump connections after the up-sampling operation at each layer.

In the jump connection, the feature map of the encoder $F_{\{enc\}}$ is divided into two parts. One part is directly passed to the decoder, and the other part undergoes multi-level multi-space feature extraction through the MHA to obtain $F_{\{att\}}$. The encoder feature map and decoder feature map are merged to obtain $F_{\{fusion\}}$ through a weighted fusion strategy.

The MHA maps the input features into multiple subspaces and performs independent attention calculations so that the model can effectively capture the feature dependencies in multiple time scales and frequency scales and realize the global modeling and accurate discrimination of fault features, and its structure is shown in Figure 3. Each head calculates its attentional output through Equation (2) as follows:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where d_k denotes the dimension size of each head, Q , K , and V represent the query matrix, key matrix, and value matrix in the multi-head attention mechanism, respectively, and T

stands for the transpose of the matrix. The outputs of multiple heads are then projected through the linear transformation layer after the splicing operation to form the final multi-head attention output, as shown in Equation (3) as follows:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_n)W^O \quad (3)$$

where W^O is the output weight matrix.

The MHA weighted hopping connection enhances the model's ability of long-distance-dependent modeling, multi-scale feature fusion, and dynamic feature weighting to improve the feature selection ability and reduce the information loss to enhance the feature reconstruction ability.

This enables the decoding part $F_{\{dec\}}$ to utilize the global and local information more comprehensively, thus improving the accuracy of feature reconstruction, as shown in Equations (4)–(6).

$$F_{\{att\}} = \text{MHA}(F_{\{enc\}}) \quad (4)$$

$$F_{\{fusion\}} = \alpha \cdot F_{\{att\}} + (1 - \alpha) \cdot F_{\{enc\}} \quad (5)$$

$$F_{\{dec\}} = \text{Decoder}(F_{\{fusion\}}) \quad (6)$$

where α is a learnable parameter that adaptively adjusts the contribution of different features to enhance the feature reconstruction in the decoding stage.

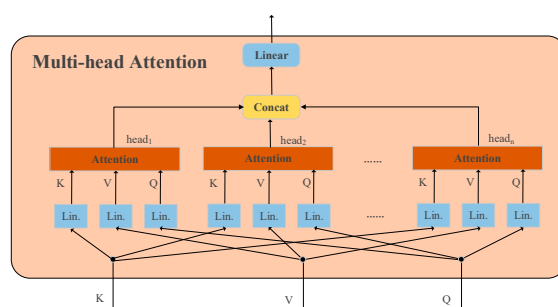


Figure 3. The structure of the multi-attention mechanism.

Where Q , K , and V represent the query matrix, key matrix, and value matrix in the multi-head attention mechanism, respectively.

2.4. ResNet50

To extract deeper and more abstract features that are beneficial for classification, it is essential to increase the depth of the network. However, as the number of network layers increases, the number of parameters grows significantly, and issues such as vanishing or exploding gradients may arise, making the network challenging to train. To address the gradient vanishing problem associated with increasing network depth, He et al. proposed ResNet, a residual learning framework based on convolutional neural networks [21]. In the residual learning framework, network layers do not simply learn a direct nonlinear mapping from input to output. Instead, they decompose this complex mapping into two components: a shortcut path and a residual path. The shortcut path directly transmits the input data to the output of the residual block, while the residual path consists of multiple convolutional and activation layers. At the output of the residual block, the original input information is combined with the information processed by the convolutional layers through an addition operation, producing the final output. The residual block, the core component of ResNet, typically comprises multiple convolutional layers, batch normalization, ReLU activation functions, and a shortcut connection. Residual blocks can

be categorized into two types: identity mapping and projection mapping, as illustrated in Figure 4. Figure 4a illustrates the structure of identity mapping, where the input is x , F represents the residual path in the identity mapping, and the output of the identity mapping is expressed by Equation (7). Figure 4b shows the structure of projection mapping, where the input is x , F represents the residual path in the projection mapping, and H represents the shortcut path in the projection mapping. The output of the projection mapping is expressed by Equation (8).

$$Y_1 = \text{ReLU}(F(x) + x) \quad (7)$$

$$Y_2 = \text{ReLU}(F(x) + H(x)) \quad (8)$$

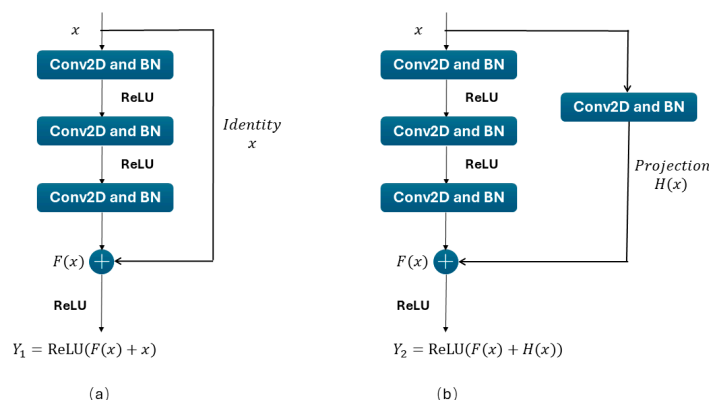


Figure 4. (a) Identity mapping; (b) projection mapping.

Multiple identity mapping blocks and projection mapping blocks are stacked following the initial convolutional layer to construct ResNet50 [22].

$W(a, b)$ goes through ResNet50 to obtain R_{output} , as shown in Equation (9).

$$R_{output} = \text{ResNet50}(W(a, b)) \quad (9)$$

2.5. The Shared Hybrid Expert Attention

In order to emphasize the importance of features and to improve the model's expressiveness by enhancing key feature regions while suppressing irrelevant features, we propose a SHEA module for the feature maps generated by hybrid U-Net and ResNet50, which is derived from the improved convolutional block attention mechanism [23]. SHEA utilizes two mutually independent attention mechanisms: multi-expert channel attention mechanisms (MCAMs) and multi-space attention mechanisms (MSAMs). MCAMs and MSAMs generate the corresponding weight maps to be fused with the input feature maps in an element-by-element multiplication manner. This process is an adaptive refinement of features and is shown in Figure 5, respectively.

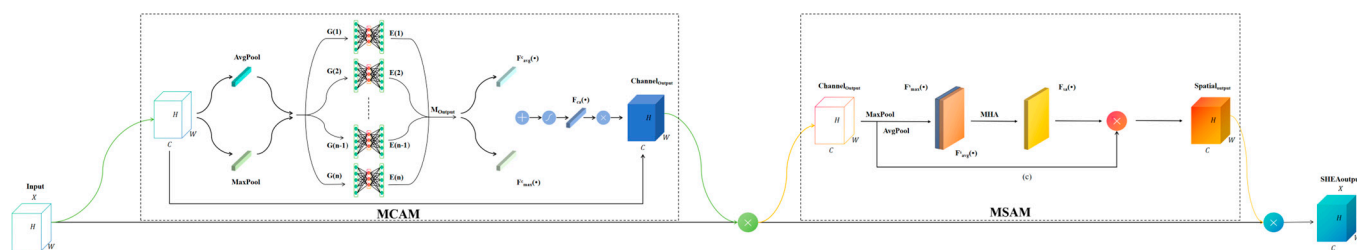


Figure 5. SHEA.

The MCAM enhances the adaptive capability by introducing the dynamic selection mechanism of mixing of experts (MoE) in the channel attention module (CAM), which is a model that can dynamically select and enhance the key channels according to the diversity of the input features, thus improving the diversity of the feature expression and generalization capability. MoE consists of an expert network and a gated network. The gated network calculates the importance of the coefficients of different experts under the current task based on the specific features of the input samples and decides the contribution of each expert in the final output. The expert network, on the other hand, consists of a number of relatively structurally independent submodels, each of which is designed to deal with a particular subspace of the input data or to model specific types of features. In the whole MoE architecture, the computational process can be formalized as Equation (10) as follows:

$$y = \sum_{i=1}^n G(x)_i \cdot Expert_i(x) \quad (10)$$

where x denotes the input sample, n denotes the total number of expert networks, and $Expert_i(x)$ denotes the result of the i th expert network for input x . $G(x)_i$ denotes the weight score computed by the gating network for the i th expert, which satisfies that the sum of the weights of all experts is 1.

The gating network is controlled according to the input x . The weights are dynamically adjusted according to the feature distribution of $G(x)$, thus providing flexible control over the combination of experts. Through the above-weighted fusion mechanism, MoE realizes sparse computation in the inference stage while maintaining a large model capacity, which effectively improves computational efficiency. In addition, the functional specificity among different experts enhances the model's expressiveness and generalization performance for diverse inputs.

Specifically, the MCAM first performs spatial aggregation on the feature map using average and max pooling. Then, the two generated vectors are passed through shared fully connected layers to obtain two sets of channel weights, $F_{avg}^c(\cdot)$ and $F_{max}^c(\cdot)$, which are fused into $F_{ca}(\cdot)$. Finally, $F_{ca}(\cdot)$ is used to weight the input feature map, selectively enhancing or suppressing the features, as shown in Equations (11)–(14).

$$F_{avg}^s(\cdot) = \sum_{i=1}^n G(F_{in})_i \cdot Expert_i(F_{in}) \quad (11)$$

$$F_{max}^c(\cdot) = \sum_{i=1}^n G(F_{in})_i \cdot Expert_i(F_{in}) \quad (12)$$

$$F_{ca}(\cdot) = \text{Sigmoid}\left(F_{avg}^c(\cdot) + F_{max}^c(\cdot)\right) \quad (13)$$

$$Channel_{Output} = F_{ca}(Input) \quad (14)$$

In addition, the MSAM first averages and maximally pools $Channel_{Output}$ to generate two single-channel feature maps, $F_{avg}^s(\cdot)$ and $F_{max}^s(\cdot)$. They are then fused along the channel dimensions to generate the spatial attention weight map, $F_{sa}(\cdot)$, by the MHA. Finally, $F_{sa}(\cdot)$ is used in $Channel_{Output}$ to refine important feature regions and suppress redundant information [24], as shown in Equations (15)–(18).

$$F_{avg}^s(\cdot) = \text{AvgPool}(Channel_{Output}) \quad (15)$$

$$F_{max}^s(\cdot) = \text{MaxPool}(Channel_{Output}) \quad (16)$$

$$F_{sa}(\cdot) = \text{MHA}\left(\text{Concat}\left(F_{avg}^s(\cdot), F_{max}^s(\cdot)\right)\right) \quad (17)$$

$$Spatial_{Output} = F_{sa}(Channel_{Output}) \quad (18)$$

$F_{\{dec\}}$ and R_{output} are SHAE-processed to obtain $SHEA_{output}$, as shown in Equation (19).

$$SHEA_{output} = SHEA(F_{\{dec\}}) + SHEA(R_{output}) \quad (19)$$

3. Experiment

3.1. Data Acquisition

The method proposed in this paper is applied to diagnose misalignment faults in the drivetrain system of wind turbines. Experimental research is conducted using a 1.5 kW misalignment test bench, which includes a frequency converter, generator, coupler, gearbox, and driving motor. The driving motor outputs torque, which is transmitted through a reduction ratio of 1:50, a speed increase ratio of 40:1, and another speed increase ratio of 1.5:1, resulting in an overall transmission ratio of 1.2:1. Displacement sensors and angle sensors are installed on the experimental platform, and the position of the generator is adjusted to simulate parallel misalignment, angular misalignment, and combined misalignment (parallel and angular misalignment). Vibration signals are collected using acceleration sensors. The experimental platform is depicted in Figure 6. Four conditions are considered during the experiment: normal, parallel misalignment, angular misalignment, and combined misalignment, as illustrated in Figure 7. In this study, the vibration dataset comprised 496 signal samples for each condition under multiple speeds, with each signal sample containing 512 data points. The vibration signals are processed using CWT to generate time-frequency images with dimensions of $256 \times 256 \times 3$, resulting in a total of 1984 image samples. The speed settings of the driving motor during the experiment are detailed in Table 1.

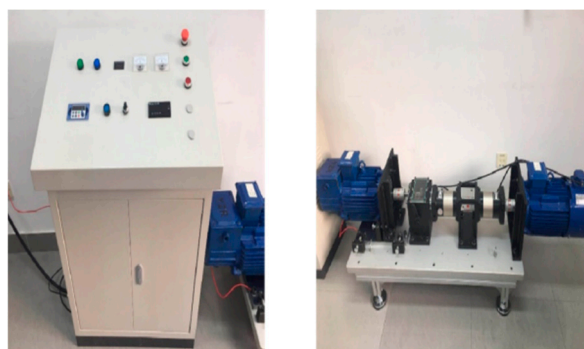


Figure 6. Wind turbine drive system failure test bench.

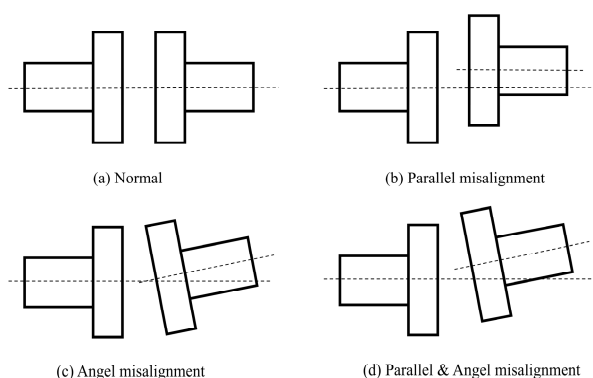


Figure 7. Four types of coupling misalignment faults. The dash lines represent the two drive shafts, the gearbox output shaft and the generator drive shaft.

Table 1. Dataset with different speeds.

Fault	Motor Speed (r/min)
Normal	100/107/200/300/500/700/720
Parallel misalignment	100/200/300/400/500
Angle misalignment	200/300/400/500/720
Parallel and angle misalignment	100/200/300/400/500/600

3.2. Experimental Results

The dual-channel hybrid attention model proposed in this paper is used to diagnose the wind turbine misalignment fault and compare it with the existing deep learning methods. Accuracy, precision, recall, and F1 score, as defined by Equations (20)–(23), are used as evaluation metrics to assess the models' performance. The results are presented in Table 2. The hardware environment of this study is RTX 4090 based on PyTorch 2.1.0 and CUDA 12.1 framework. All the experiments used uniform hyperparameter settings, including training epochs of 100, a batch size of 32, a learning rate of 0.0001, and ReLU as the activation function. Regarding the model architecture, the U-Net network contains five up- and down-sampling operations. ResNet uses the classical [3, 4, 6, 3] configuration, i.e., each layer contains three, four, six, and three Bottleneck modules, respectively. The multi-head attention mechanism (MHA) employs four attention heads to enhance the feature extraction capability. As shown in Table 2, the two-channel hybrid attention model significantly outperforms the other methods in all metrics, accuracy, precision, recall, and F1 score, demonstrating excellent performance improvement. In addition, even though the model has the most trainable parameters, its inference time does not increase significantly and is comparable to other methods. The method proposed in this paper achieves better performance while ensuring efficient inference, fully demonstrating its potential and value in practical applications.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (20)$$

$$Precision = \frac{TP}{TP + FP} \quad (21)$$

$$Recall = \frac{TP}{TP + FN} \quad (22)$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (23)$$

Here, TP , FP , FN , and TN represent the number of true positive, false positive, false negative, and true negative results, respectively [25].

Table 2. Comparison of the results of the different methods.

Model Evaluate	Accuracy	Precision	Recall	F1 Score	Inference Speed (Seconds)	Trainable Parameters (Million)
VGG	97.17%	96.16%	97.14%	97.13%	5.01	165.73
U-Net	99.28%	99.30%	99.29%	99.28%	3.06	31.18
ResNet	98.21%	98.27%	98.21%	98.20%	2.45	23.51
Mobilenet	97.14%	97.16%	97.14%	97.13%	2.03	16.37
MLP-Mixer	98.21%	98.21%	98.21%	98.21%	4.78	76.21
ConvNeXt	96.78%	96.98%	96.78%	96.75%	2.75	28.63
Present Results	99.64%	99.65%	99.64%	99.64%	5.23	325.31

In order to compare the actual performance of different models more clearly, this paper analyzes the training loss curves in Figure 8. It can be observed that during the training process, the dual-channel hybrid attention model always maintains a lower loss value and possesses a faster convergence speed, showing excellent feature reconstruction ability and high-precision classification performance, which fully verifies the effectiveness of the model in feature learning and pattern recognition.

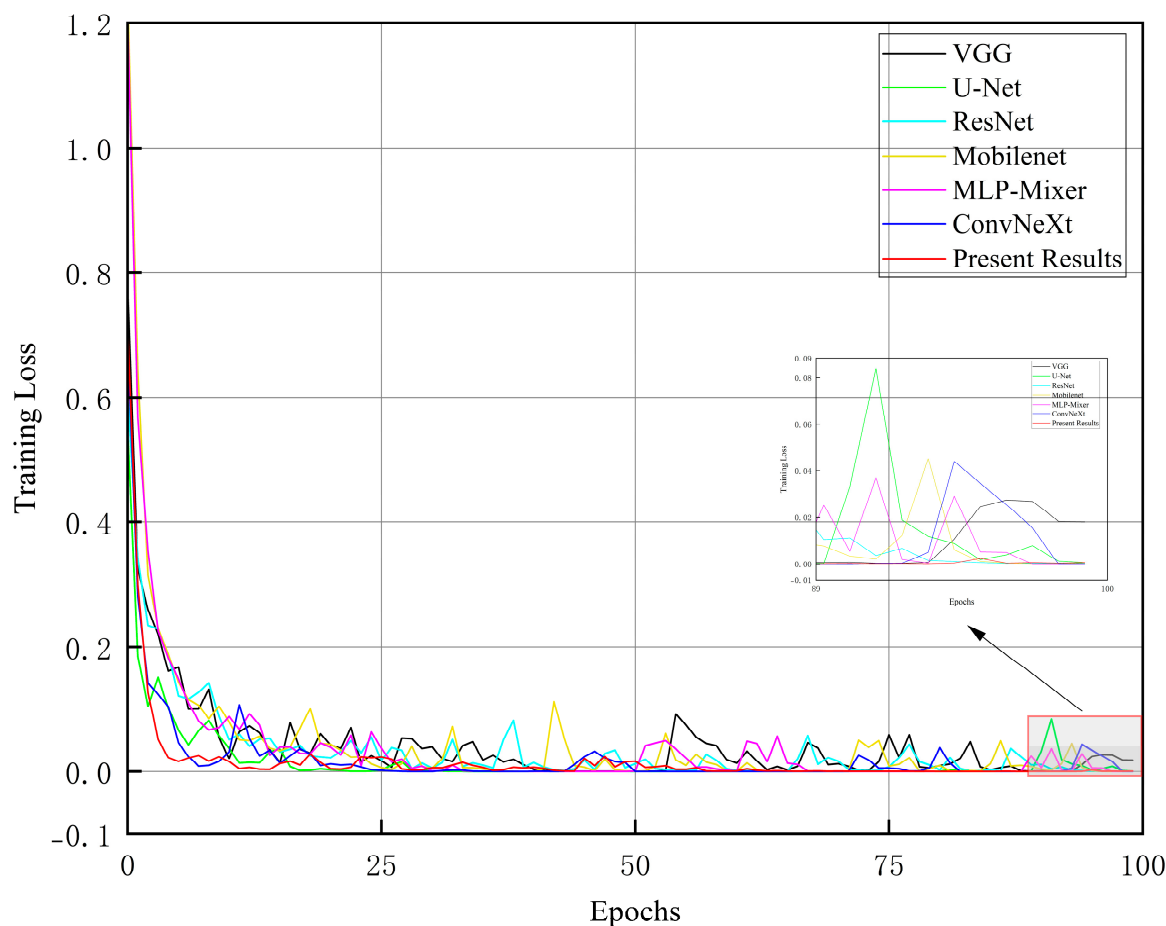


Figure 8. The training loss curves.

Figure 9 shows the visualized scatter plots of different methods in the feature space. It is obvious from the figure that after deep feature extraction by the dual-channel hybrid attention model, the samples of different categories form a clear clustering distribution in the high-dimensional feature space, and each category shows obvious boundary separation. This phenomenon fully demonstrates that the model can effectively extract key features and improve the recognition ability of complex patterns, further proving its advantages in feature learning and spatial distribution optimization.

Combining the quantitative evaluation data in Table 2, the training loss curve in Figure 8, and the feature space visualization results in Figure 9, the excellent performance of the dual-channel hybrid attention model in the task of wind turbine misalignment fault diagnosis can be fully reflected.

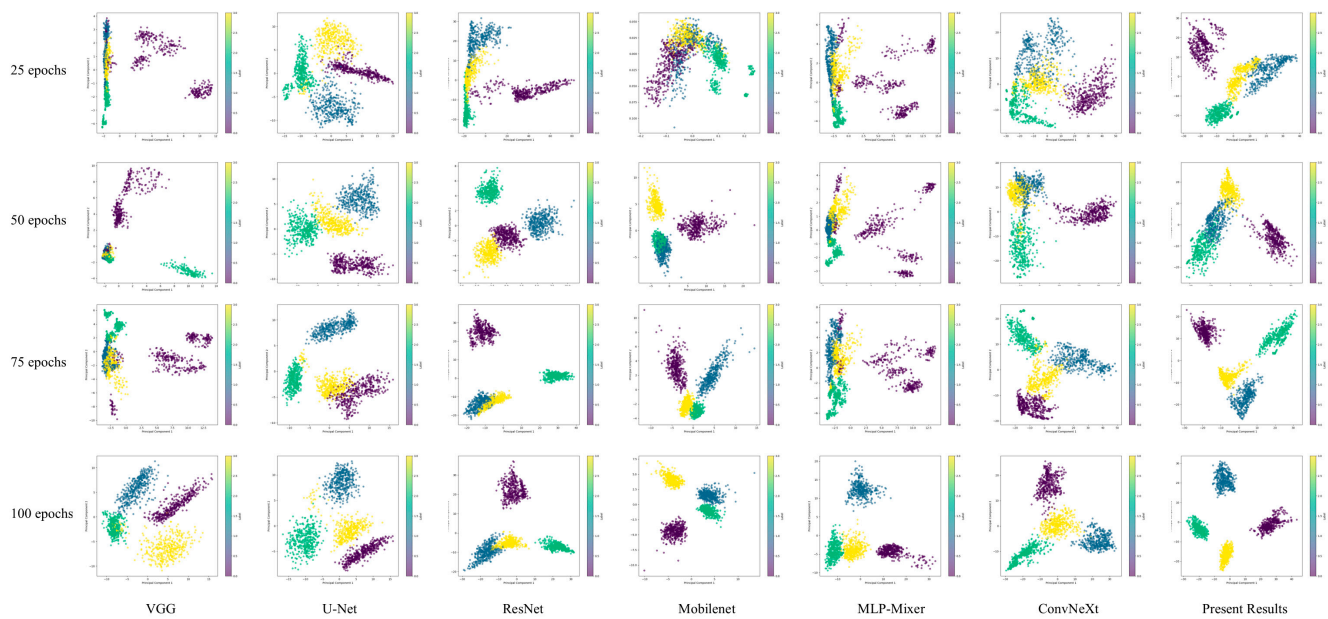


Figure 9. The feature space visualization results.

In summary, the dual-channel hybrid attention model proposed in this paper achieves efficient, accurate, and stable fault diagnosis under multiple operating conditions by virtue of the innovative introduction of the MHA into the hybrid U-Net hopping connection and the enhancement of the two feature extraction paths through the SHEA module.

3.3. Ablation Experiments

To evaluate the specific contribution of SHEA and hybrid U-Net to the model performance, we conducted systematic ablation experiments. The experiments use the same dataset, and the evaluation results are shown in Table 3. The training loss curves in Figure 10 show that the dual-channel hybrid attention model significantly outperforms the comparison model in terms of both convergence speed and final loss values. Further feature visualization analysis in Figure 10 shows that SHEA and hybrid U-Net are able to extract more discriminative deep features effectively. Combining the experimental data in Table 3, the training loss curves in Figure 11, and the feature visualization results in Figure 10, it can be clearly concluded that SHEA and hybrid U-Net are the key components to enhance the classification performance and diagnostic accuracy of the dual-channel hybrid attention model and play an irreplaceable role in the model.

Table 3. The results of the ablation experiments.

Model Evaluate	Accuracy	Precision	Recall	F1Score
U-ResNet	91.79%	91.22%	91.79%	91.44%
Hybrid U-ResNet	96.79%	96.79%	96.78%	96.78%
Present Results	99.64%	99.65%	99.64%	99.64%

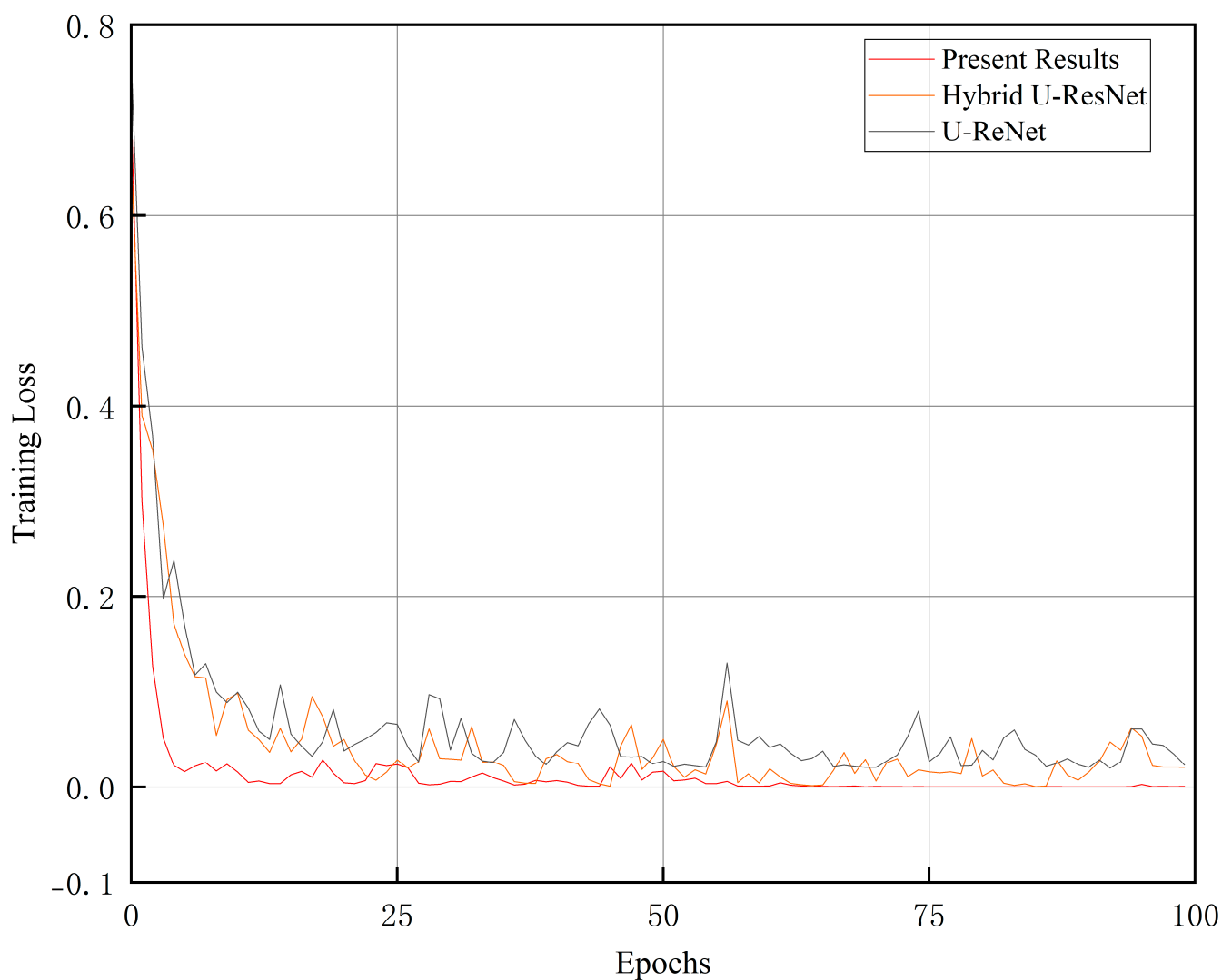


Figure 10. Ablation experiment training loss curves.

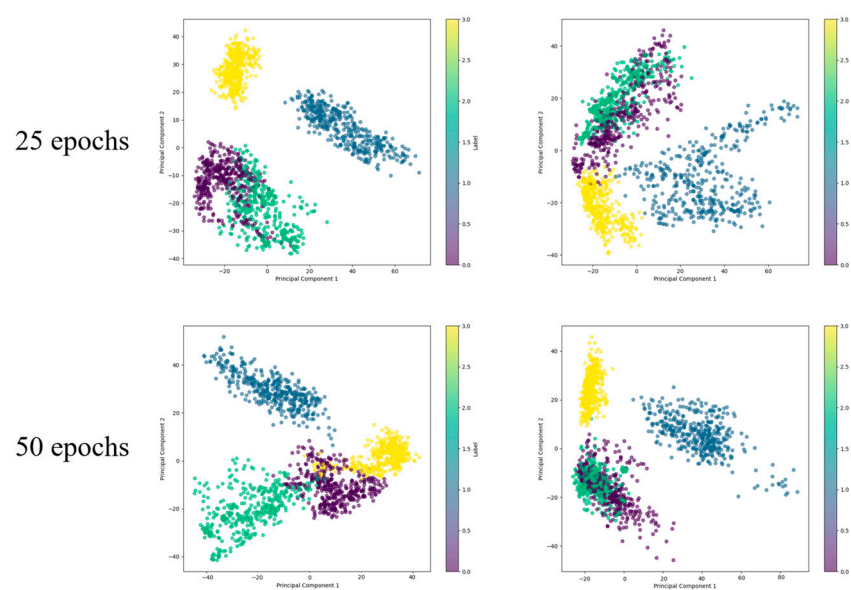


Figure 11. Cont.

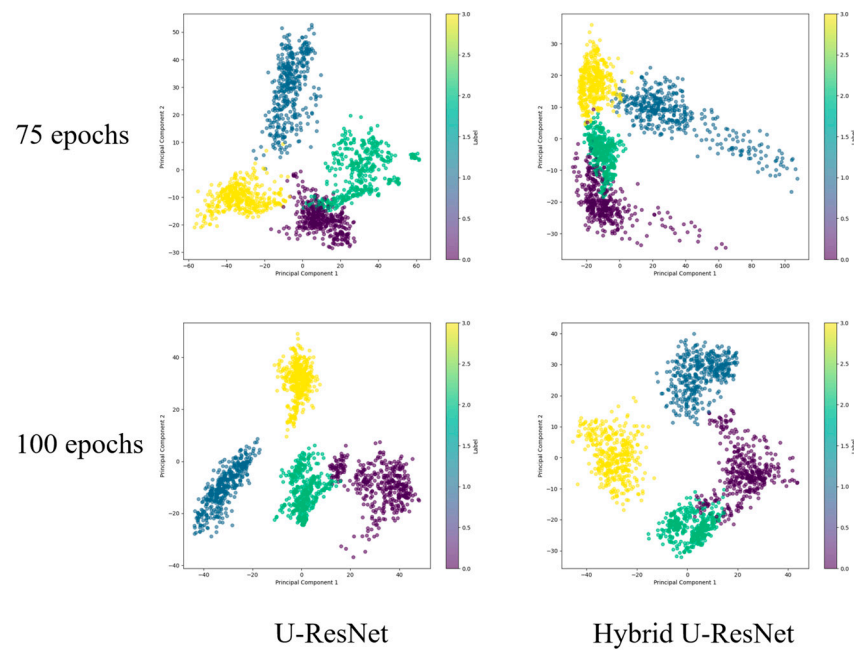


Figure 11. Results of spatial visualization of ablation experiment features.

4. Conclusions

The dual-channel hybrid attention model proposed in this paper significantly improves the performance of wind turbine misalignment fault diagnosis by innovatively introducing the MHA in the jump connection of U-Net and combining the SHEA module to enhance the two feature extraction paths. The model realizes multi-scale feature extraction and detail recovery through the encoding–decoding structure in the hybrid U-Net branch using the MHA to enhance cross-layer feature interactions and the ability to focus on critical regions. In the ResNet50 branch, deep global features are mined, and the SHEA module dynamically weights and selects experts between multi-scale local features and global features to realize efficient fusion of global and local information. Therefore, facing the challenges of signal non-stationarity and transient mutation characteristics in the fault diagnosis of wind turbine drive train misalignment under complex working conditions, this method is able to realize efficient, accurate, and stable fault diagnosis, which provides reliable technical support for the intelligent monitoring of industrial equipment.

In the future, research can focus on optimizing the model’s computational efficiency and lightweight design, enhancing its generalization ability across different wind turbine types and noisy environments, and integrating multi-modal data for more comprehensive fault diagnosis. Additionally, exploring unsupervised or semi-supervised learning methods to reduce reliance on labeled data, improving model interpretability, and extending the framework to fault prognosis and predictive maintenance will further advance its applicability in real-world industrial scenarios.

Author Contributions: Formal analysis, T.T.; methodology, X.L.; software, J.Z.; supervision, T.F.; validation, D.L.; writing—review and editing, T.T. and D.L.; data curation, X.L. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was funded by the Project Supported by China Huaneng Group (HNKJ24-H95).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: Authors Xiang Liu, Jia Zhang were employed by Huaneng Qinghai Power Generation Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest..

References

1. Bajric, R.; Zuber, N.; Skrimpas, G.A.; Mijatovic, N. Feature Extraction Using Discrete Wavelet Transform for Gear Fault Diagnosis of Wind Turbine Gearbox. *Shock Vib.* **2016**, *2016*, 10. [\[CrossRef\]](#)
2. Yi, J.L.; Tan, H.B.; Yan, J.; Chen, X. Adaptive rotating machinery fault diagnosis method using MKIST. *Meas. Sci. Technol.* **2024**, *35*, 045010. [\[CrossRef\]](#)
3. Zhao, D.Z.; Wang, H.H.; Cui, L.L. Frequency-chirprate synchrosqueezing-based scaling chirplet transform for wind turbine nonstationary fault feature time-frequency representation. *Mech. Syst. Signal Process.* **2024**, *209*, 111112. [\[CrossRef\]](#)
4. Long, X.F.; Yang, P.; Guo, H.X.; Zhao, Z.L.; Wu, X.W. A CBA-KELM-Based Recognition Method for Fault Diagnosis of Wind Turbines with Time-Domain Analysis and Multisensor Data Fusion. *Shock Vib.* **2019**, *2019*, 14. [\[CrossRef\]](#)
5. Pu, Z.Q.; Li, C.; Zhang, S.H.; Bai, Y. Fault Diagnosis for Wind Turbine Gearboxes by Using Deep Enhanced Fusion Network. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 11. [\[CrossRef\]](#)
6. Tang, B.P.; Song, T.; Li, F.; Deng, L. Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine. *Renew. Energy* **2014**, *62*, 1–9. [\[CrossRef\]](#)
7. Cai, W.B.; Zhao, D.Z.; Wang, T.Y. Multi-scale dynamic graph mutual information network for planet bearing health monitoring under imbalanced data. *Adv. Eng. Inform.* **2025**, *64*, 103096. [\[CrossRef\]](#)
8. Gao, S.; Zhang, Z.J.; Zhang, X.; Li, H. WBUN: An interpretable convolutional neural network with wavelet basis unit embedded for fault diagnosis. *Meas. Sci. Technol.* **2024**, *35*, 086125. [\[CrossRef\]](#)
9. Zhao, D.Z.; Cai, W.B.; Cui, L.L. Adaptive thresholding and coordinate attention-based tree-inspired network for aero-engine bearing health monitoring under strong noise. *Adv. Eng. Inform.* **2024**, *61*, 102559. [\[CrossRef\]](#)
10. Zhao, D.Z.; Cai, W.B.; Cui, L.L. Multi-perception graph convolutional tree-embedded network for aero-engine bearing health monitoring with unbalanced data. *Reliab. Eng. Syst. Saf.* **2025**, *257*, 110888. [\[CrossRef\]](#)
11. Xu, Y.D.; Feng, K.; Yan, X.A.; Sheng, X.; Sun, B.B.; Liu, Z.; Yan, R.Q. Cross-Modal Fusion Convolutional Neural Networks with Online Soft-Label Training Strategy for Mechanical Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2024**, *20*, 73–84. [\[CrossRef\]](#)
12. Li, R.B.; Yu, P.; Cao, J. Rolling bearing fault diagnosis method based on SOA-BiLSTM. In Proceedings of the 7th International Conference on Electronic Information Technology and Computer Engineering (EITCE), Xiamen, China, 20–22 October 2023; Association for Computing Machinery: New York, NY, USA, 2024; pp. 41–45.
13. Mohammad-Alikhani, A.; Nahid-Mobarakeh, B.; Hsieh, M.F. One-Dimensional LSTM-Regulated Deep Residual Network for Data-Driven Fault Detection in Electric Machines. *IEEE Trans. Ind. Electron.* **2024**, *71*, 3083–3092. [\[CrossRef\]](#)
14. Tao, S.Q.; Zhang, T.; Yang, J.; Wang, X.Q.; Lu, W.N. Bearing Fault Diagnosis Method Based on Stacked Autoencoder and Softmax Regression. In Proceedings of the 34th Chinese Control Conference (CCC), Hangzhou, China, 28–30 July 2015; pp. 6331–6335.
15. Zhao, Y.Y.; Hao, H.J.; Chen, Y.; Zhang, Y. Novelty Detection and Fault Diagnosis Method for Bearing Faults Based on the Hybrid Deep Autoencoder Network. *Electronics* **2023**, *12*, 2826. [\[CrossRef\]](#)
16. Wang, L.J.; Zou, T.X.; Cai, K.L.; Liu, Y. Rolling bearing fault diagnosis method based on improved residual shrinkage network. *J. Braz. Soc. Mech. Sci. Eng.* **2024**, *46*, 172. [\[CrossRef\]](#)
17. Wang, S.Y.; Tian, J.Y.; Liang, P.F.; Xu, X.F.; Yu, Z.Z.; Liu, S.Y.; Zhang, D.L. Single and simultaneous fault diagnosis of gearbox via wavelet transform and improved deep residual network under imbalanced data. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108146. [\[CrossRef\]](#)
18. Shi, H.M.; Chen, J.C.; Si, J.; Zheng, C.C. Fault Diagnosis of Rolling Bearings Based on a Residual Dilated Pyramid Network and Full Convolutional Denoising Autoencoder. *Sensors* **2020**, *20*, 5734. [\[CrossRef\]](#)
19. Dinç, E.; Üstündag, Ö.; Tilkan, G.Y.; Türkmen, B.; Özdemir, N. Continuous wavelet transform methods for the simultaneous determinations and dissolution profiles of valsartan and hydrochlorothiazide in tablets. *Braz. J. Pharm. Sci.* **2017**, *53*. [\[CrossRef\]](#)
20. Ghosh, S.; Chaki, A.; Santosh, K.C. Improved U-Net architecture with VGG-16 for brain tumor segmentation. *Phys. Eng. Sci. Med.* **2021**, *44*, 703–712. [\[CrossRef\]](#)
21. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 770–778.
22. Wen, L.; Li, X.Y.; Gao, L. A transfer convolutional neural network for fault diagnosis based on ResNet-50. *Neural Comput. Appl.* **2020**, *32*, 6111–6124. [\[CrossRef\]](#)
23. Xu, Q.H.; Jiang, H.; Zhang, X.F.; Li, J.; Chen, L. Multiscale Convolutional Neural Network Based on Channel Space Attention for Gearbox Compound Fault Diagnosis. *Sensors* **2023**, *23*, 3827. [\[CrossRef\]](#)

24. Xu, T.Y.; Qi, X.Y.; Lin, S.; Zhang, Y.H.; Ge, Y.H.; Li, Z.L.; Dong, J.; Yang, X. A Neural Network Structure with Attention Mechanism and Additional Feature Fusion Layer for Tomato Flowering Phase Detection in Pollination Robots. *Machines* **2022**, *10*, 1076. [\[CrossRef\]](#)
25. Tao, W.J.; Li, X.W.; Liu, J.L.; Li, Z. Multi-scale attention network (MSAN) for track circuits fault diagnosis. *Sci. Rep.* **2024**, *14*, 8886. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.