electronics

MDPI

Article

Bearing Fault Vibration Signal Denoising Based on Adaptive Denoising Autoencoder

Haifei Lu, Kedong Zhou * and Lei He

School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; lhf_0513@njust.edu.cn (H.L.); hlhlhxh@njust.edu.cn (L.H.)
* Correspondence: zkd81151@njust.edu.cn

Abstract: Vibration signal analysis is regarded as a fundamental approach in diagnosing faults in rolling bearings, and recent advancements have shown notable progress in this domain. However, the presence of substantial background noise often results in the masking of these fault signals, posing a significant challenge for researchers. In response, an adaptive denoising autoencoder (ADAE) approach is proposed in this paper. The data representations are learned by the encoder through convolutional layers, while the data reconstruction is performed by the decoder using deconvolutional layers. Both the encoder and decoder incorporate adaptive shrinkage units to simulate denoising functions, effectively removing interfering information while preserving sensitive fault features. Additionally, dropout regularization is applied to sparsify the network and prevent overfitting, thereby enhancing the overall expressive power of the model. To further enhance ADAE's noise resistance, shortcut connections are added. Evaluation using publicly available datasets under scenarios with known and unknown noise demonstrates that ADAE effectively enhances the signal-to-noise ratio in strongly noisy backgrounds, facilitating accurate diagnosis of faults in rolling bearings.

Keywords: fault diagnosis; autoencoder; vibration signal denoising; convolution and deconvolution

check for updates

Citation: Lu, H.; Zhou, K.; He, L. Bearing Fault Vibration Signal Denoising Based on Adaptive Denoising Autoencoder. *Electronics* **2024**, *13*, 2403. https://doi.org/10.3390/electronics13122403

Academic Editors: Mário P. Véstias and Rui Policarpo Duarte

Received: 14 May 2024
Revised: 14 June 2024
Accepted: 16 June 2024
Published: 19 June 2024

Copyright: © 2024 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/).

1. Introduction

Rolling bearings are crucial components in rotating machinery, significantly influencing equipment reliability and stability [1]. However, their prolonged exposure to complex working environments makes them susceptible to malfunctions, which can negatively impact machine performance and pose safety risks [2,3]. Effective fault diagnosis techniques for rolling bearings are therefore essential [4,5].

Vibration signal analysis is pivotal in diagnosing rolling bearing faults due to the plethora of fault-related features inherent in such signals [6,7]. However, signals collected from rolling bearings in real-world industrial settings often suffer from significant noise interference, presenting substantial challenges [8]. Over the years, considerable attention has been devoted to reducing noise interference and extracting meaningful fault features [9–11]. Various signal denoising techniques have been explored, including empirical mode decomposition (EMD) [12,13], variational mode decomposition (VMD) [14,15], wavelet analysis (WT) [16–18], and threshold denoising [19,20]. Despite their usefulness, these methods often lack adaptability, particularly in noisy environments.

EMD and its variants, such as Ensemble EMD (EEMD), have been widely used for non-stationary signal analysis. EMD decomposes a signal into intrinsic mode functions (IMFs) but can suffer from mode mixing issues. VMD offers an alternative by decomposing the signal into modes with specific sparsity properties in the frequency domain, yet it requires predefinition of mode numbers and can be sensitive to noise.

WT provides a multi-resolution analysis of signals and has been effective in fault diagnosis. It offers a time-frequency representation, making it suitable for transient feature extraction. However, the choice of mother wavelet and decomposition level significantly influences its performance.

Threshold denoising techniques, often combined with WT, apply hard or soft thresholds to the decomposed signal coefficients to reduce noise. Although these methods can enhance signal clarity, they are typically heuristic and may not adapt well to varying noise levels.

Recent advancements in deep learning, particularly convolutional neural networks (CNNs), have significantly enhanced fault diagnosis capabilities by automating feature extraction from raw data [21–23]. CNNs excel in learning hierarchical representations, which are vital for distinguishing fault features from noise [24–26]. Studies have demonstrated that CNN-based models outperform traditional methods in both accuracy and robustness under varying noise conditions [27–29]. These models can capture complex patterns and interactions within the data that are often missed by traditional techniques.

Building upon these advancements, we propose an adaptive denoising autoencoder (ADAE) framework. The ADAE framework introduces a novel adaptive shrinkage unit (ASU) that employs local attention mechanisms to dynamically adjust shrinkage coefficients. This approach effectively removes noise while preserving critical fault features. Additionally, the integration of dropout regularization within the encoder enhances the network's resilience by simulating real-world noise conditions during training. This dual strategy of local attention and dropout regularization not only improves denoising performance but also ensures the model's robustness and adaptability across different industrial applications.

The ADAE framework holds broad applicability in industries reliant on precise machinery health monitoring, including the aerospace, manufacturing, and energy sectors. By effectively mitigating noise interference in vibration signals, ADAE paves the way for early detection of equipment anomalies and predictive maintenance strategies, thereby reducing unscheduled downtime and maintenance costs.

In summary, our ADAE framework addresses the limitations of traditional denoising techniques and leverages the strengths of CNNs to provide a more adaptable and accurate solution for vibration signal analysis in noisy environments. This work represents a significant step forward in the development of robust diagnostic tools for industrial applications.

2. Theoretical Foundations

2.1. Problem Statement

In rolling bearings, vibration signals, denoted as s, are a combination of the original signal u and a noise signal n. In real-world scenarios, these signals are often contaminated with noise from various sources, typically modeled as additive white Gaussian noise (AWGN). The goal is to remove the noise n from the observed vibration signal s to recover the original fault pulse signal u. Since obtaining a pure signal is challenging, we use noisy signals as the training set, adding Gaussian white noise to simulate noise and train the model. This approach allows the model to learn the characteristics of the noise within the contaminated signals and extract the clean signal.

For the training process, a training set $M = \{(s_i, u_i)\}_{i=1}^{N} = S \times U$ is utilized, where S and U represent matrices of noise signals and clean signals, respectively. The network parameters are trained to minimize the loss error L between the network input s_i and output $\hat{s_i}$.

2.2. Basic Components of CNN

Convolutional neural networks (CNNs) are comprised of various fundamental components that play essential roles in processing and extracting features from input data. These components include convolution (Conv), deconvolution (Deconv), nonlinear activation functions (ReLU), batch normalization (BN), and fully connected (FC) layers, among others [30,31].

2.3. Dropout: Addressing Overfitting in Deep Learning

Dropout [32] is a widely adopted technique in deep learning models, renowned for its efficacy in mitigating overfitting. By randomly dropping out neurons during training,

Dropout encourages the network to learn more robust and diverse representations of the data. This prevents the model from relying too heavily on specific features or training samples, thereby improving its generalization performance on unseen data. Dropout is straightforward to implement and does not require significant modifications to the network architecture, making it a popular choice among researchers and practitioners in the field of deep learning. Its ability to enhance model performance while maintaining simplicity has contributed to its widespread adoption across various applications.

2.4. Theoretical Basis of Autoencoders

Autoencoders [33] represent a class of neural networks designed to replicate input data with precision, aiming to distill essential features while discarding redundant information. Comprising input, hidden, and output layers, autoencoders are structured to encode input data into a lower-dimensional latent space and subsequently decode it back to its original form [34]. Positioned between the input and output layers, the hidden layer in autoencoders acts as a bottleneck, deliberately restricting the network's capacity to learn a compressed representation of the input data.

This component transforms the input data x into a latent representation h by means of nonlinear transformations, facilitated by parameters such as the activation function φ, weight matrix W and bias vector b.

$$h = f(x) = \varphi(Wx + b), \tag{1}$$

Subsequently, the decoder network reconstructs the latent representation h back into the original input $\widetilde{x}$. This process involves another set of nonlinear transformations using σ, W' and b' for the activation function, weight matrix, and bias vector of the decoder, respectively.

$$\widetilde{x} = g(h) = \sigma\left(W'h + b'\right), \tag{2}$$

The performance of an autoencoder is evaluated based on its ability to reconstruct the input faithfully. The reconstruction error [35], denoted by $L\left(x, \widetilde{x}\right)$, measures the discrepancy between the original input x and its reconstruction $\widetilde{x}$. The optimization process revolves around fine-tuning the network parameters, to minimize the reconstruction error observed across the training samples. This objective underscores the endeavor to glean insightful representations of the input data, capable of faithfully reconstructing the original inputs with precision and fidelity.

3. The Proposed Method

3.1. Shortcut Connection

Challenges related to gradients, such as gradient vanishing and gradient explosion, are prevalent in the training of deep neural networks and often lead to convergence issues or sluggish training. To ensure that gradients converge meaningfully towards detection rather than diverging, this study adopts several strategies: the use of adaptive learning rate scheduling via the Adam algorithm, which adjusts the learning rate dynamically during training; proper weight initialization techniques; regularization methods; dropout [32]; and skip connections [36]. Skip connections enable the network to bypass one or more layers, directly adding the input from a previous layer to the output of a subsequent layer, thereby facilitating smoother gradient flow and enhancing training stability. This facilitates:

1. Gradient Propagation: Providing a direct path for gradients to propagate back to earlier layers helps mitigate the vanishing gradient problem, thus facilitating smoother training.
2. Information Flow: Faster information flow within the network is enabled, aiding in quicker and more efficient learning of effective feature representations.

3. Feature Reuse: Accessing original inputs or feature maps from preceding layers directly prevents information loss, crucial for tasks such as image super-resolution and segmentation, where preserving detailed information is vital.

In summary, the incorporation of shortcut connection enhances training efficiency, stability, and generalization ability, leading to improved performance across various deep learning tasks.

3.2. Adaptive Shrinkage Unit (ASU)

Traditional denoising techniques such as filter design, wavelet thresholding, and sparse representation are effective in retaining useful information while removing noise from signals. However, these methods often demand substantial domain expertise. Deep learning approaches such as DRSN [37] combine soft thresholding with deep learning to automatically learn channel thresholds. Nevertheless, manually defining threshold functions and their varying impacts present challenges in adaptively selecting appropriate functions for specific problems.

The ASU employs a local attention mechanism to train shrinkage coefficients, which are then used to attenuate noise in the input signal. This approach allows the ASU to adaptively determine the optimal shrinkage values for different parts of the signal, thus effectively removing noise while preserving essential fault-related features.

Training of Shrinkage Coefficients: The shrinkage coefficients are trained using convolutional and deconvolutional operations, as depicted in Figure 1. Specifically, the ASU consists of a convolutional layer followed by a sigmoid activation function to constrain the shrinkage coefficients within the [0, 1] range. The sigmoid function ensures that the coefficients scale the input signal appropriately, reducing the noise components while retaining significant fault information.

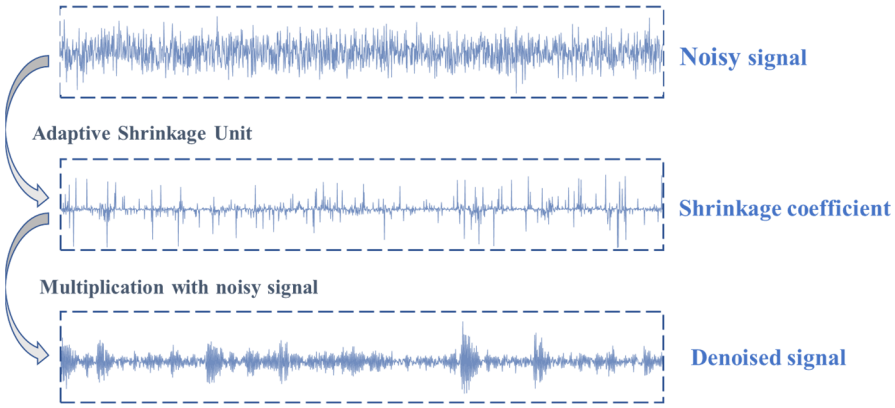


Figure 1. Principle of ASU.

Application to Noisy Signals: Once the shrinkage coefficients are learned, they are applied to the noisy input signal. For a given noisy signal xxx, the ASU computes the shrinkage coefficients α as follows: $\alpha = \sigma(Wx + b)$ where W and b are the weights and biases learned during training, and σ represents the sigmoid activation function. The denoised signal $\hat{x}$ is then obtained by element-multiplication of the shrinkage coefficients with the noisy signal:$\hat{x} = Ax$. This process effectively filters out the noise components, as the shrinkage coefficients adapt to the noise level in different segments of the signal.

Significance in the Denoising Process: The ASU enhances the network's generalization ability and overall performance by reducing the reliance on manually defined threshold functions and improving adaptability to different datasets and noise levels. By incorporating the ASU within both the encoder and decoder of our adaptive denoising autoencoder (ADAE) framework, we ensure that noise is effectively attenuated during both feature extraction and signal reconstruction phases, resulting in more accurate fault diagnosis.

3.3. Architecture of the Proposed Method

Based on the fundamental denoise autoencoder architecture, our work introduces an innovative vibration signal denoising framework, the adaptive denoising autoencoder (ADAE), illustrated in Figure 2. The proposed ADAE framework consists of several key modules:

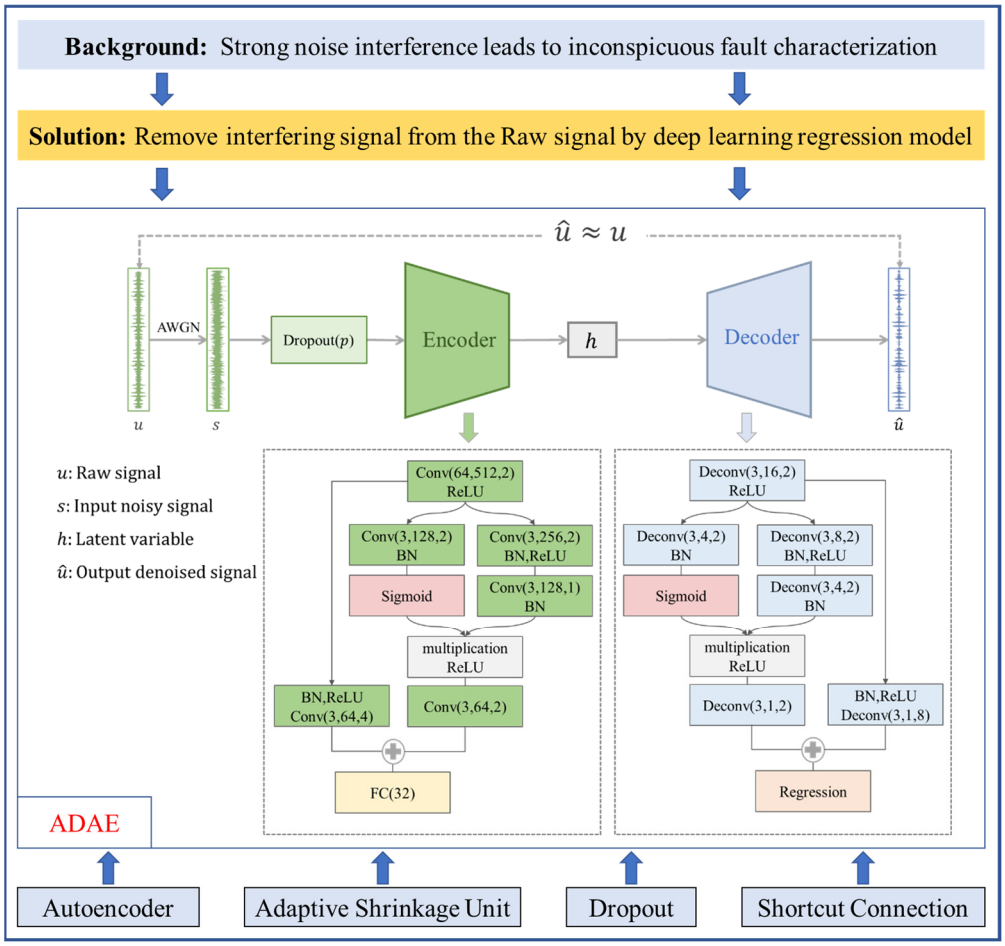


Figure 2. Illustrative overview of the proposed model's network structure.

Encoder: Utilizes convolutional layers to learn data representations, focusing on compressing temporal information.

Decoder: Employs deconvolutional layers for data reconstruction, ensuring the recovery of signal details.

Adaptive Shrinkage Unit (ASU): Trains shrinkage coefficients through local attention mechanisms, effectively removing noise while preserving fault features.

Dropout Regularization: Introduced after the initial convolutional layer to mimic real-world noise conditions, enhancing the network's resilience and adaptability.

Shortcut Connection: Enhances training efficiency, stability, and generalization ability, leading to improved performance across various deep learning tasks.

4. Experimental Validation

The denoising performance of ADAE is rigorously evaluated on the CWRU dataset across different noise levels to ascertain its effectiveness in real-world scenarios. Our experiments were conducted using Matlab-2022a in windows environment.

4.1. Experimental Setup and Data Description

The CWRU dataset is a commonly utilized third-party dataset that offers a reliable method for assessing the effectiveness of existing algorithms. This dataset includes drive-end samples captured at a sampling frequency of 12 kHz, containing signals related to four different health conditions: Normal, rolling element fault, inner race fault, and outer race fault. Each health condition consists of four different operating loads, and within each fault mode, three different fault sizes are examined. Therefore, there are a total of ten categories of bearing data, each corresponding to four operating loads.

To ensure data consistency and comparability, all collected data undergoes z-score standardization, mitigating the potential impact of sensor placement. Gaussian white noise (AWGN) is subsequently added to the data based on signal-to-noise ratio (*SNR*) for performance validation of ADAE and comparison with other mainstream algorithms. The dataset is split into training and testing samples at a ratio of 4:1. In our denoising process, the presence of AWGN necessitates an assessment of the denoising algorithm's effectiveness. We employ the signal-to-noise ratio (*SNR*) as a fundamental metric, defined as:

$$SNR = 10log\frac{||u||_2^2}{||s-u||_2^2}, \tag{3}$$

We aim to eliminate AWGN noise, mimicking real industrial scenarios. To gauge the algorithm's performance comprehensively, we utilize three evaluation metrics [11]: *SNR* improvement and root mean square error (*RMSE*) defined as:

$$SNR_{imp} = 10log\frac{||s-u||_2^2}{||\hat{u}-u||_2^2}, \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}||\hat{u}-u||_2^2}, \tag{5}$$

Here u, $\hat{u}$ and s have the same meanings as before. A higher SNR_{imp} and lower $RMSE$ indicate stronger denoising capabilities of the algorithm.

4.2. Hyperparameter Optimization and Ablation Study

In this section, critical structures and parameters are optimized and tested using the CWRU dataset. All experiments are conducted under a noise level of SNR = -6 dB. We employ a sliding window approach to construct the sample set, with each sample containing 2048 data points, and no overlap between adjacent samples. For each fault category and each load, 50 samples are generated, resulting in a total of 2000 samples. This experimental design ensures a comprehensive evaluation and comparison of model performance under different conditions.

4.2.1. Impact of Dropout Probability on Denoising Performance

This section delves into the significance of dropout probability (p) in dictating the denoising efficiency of the adversarial domain adaptation encoder (ADAE) model, specifically when the signal-to-noise ratio (*SNR*) is set at -6 dB. An exploratory endeavor was undertaken, where p was systematically varied from 0 to 0.6, yielding a comprehensive dataset of denoising outcomes summarized in Table 1.

Table 1. Denoising performance of ADAE with different dropout probabilities at -6 dB.

p	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**Random**
SNR_{imf}	10.27 ± 1.81	11.62 ± 1.98	11.10 ± 1.91	10.67 ± 1.78	9.76 ± 1.51	9.04 ± 1.31	8.21 ± 0.92	11.26 ± 1.99
RMSE	0.64 ± 0.11	0.54 ± 0.13	0.57 ± 0.13	0.60 ± 0.12	0.66 ± 0.12	0.71 ± 0.11	0.78 ± 0.08	0.56 ± 0.12

The findings highlight a peak in denoising performance within the p range of 0.1 to 0.3 across all evaluation metrics. A marked deterioration occurs beyond $p = 0.3$, suggesting that excessive dropout disrupts the signal excessively, hindering effective learning. Conversely, lower p values maintain a balance between regularization and information retention, fostering an environment conducive to noise reduction.

In anticipation of variable noise profiles in practical applications, we introduced a novel strategy: The randomization of p within the interval [0.1, 0.3]. This decision was meticulously considered to imbue the ADAE model with heightened versatility. By dynamically selecting p during training epochs, the model becomes resilient to fluctuations in noise intensity, thereby ensuring consistent denoising performance across a broader spectrum of SNR conditions. Specifically, this stochastic element emulates real-world variability, forcing the model to adapt and learn from a more diverse set of signal-disruption patterns, which, in turn, bolsters its generalization capabilities.

In summary, the analysis underscores the critical role of dropout probability in optimizing the denoising performance of the ADAE model, with randomized p values proving instrumental in enhancing adaptability and efficacy across varying noise levels.

4.2.2. Impact of Initial Convolutional Kernel Width

In delving into the intricacies of the neural network architecture, the initial choice of the convolutional kernel width significantly influences the model's feature extraction efficacy from raw inputs. This section provides a comprehensive elaboration on this parameter, addressing the oversight in previous explanations and enhancing the holistic understanding of the architecture.

Intuitive Understanding of Kernel Width: Analogous to the human visual system, the convolutional kernel operates as a feature perception mechanism, scanning over the data. The width of the kernel determines the "field of view" for feature detection, affecting the balance between capturing local details and integrating broader contextual information. Wider kernels are proficient at incorporating extensive contextual cues, whereas narrower kernels excel in discerning fine-grained, localized features.

Experimental Framework and Insights: Our study intentionally selected kernel widths as powers of 2 (4, 8, 16,. . ., 1024) for systematic comparison and computational efficiency. Under a fixed dropout rate ($p = 0.2$), we assessed the denoising performance of the adversarial domain adaptation encoder (ADAE) across these varying widths, presenting the findings in Table 2 through SNR improvement (SNR_{imf}) and root mean square error (RMSE) metrics under -6 dB SNR conditions.

Table 2. Denoising performance of ADAE with different kernel widths at -6 dB.

Kernel Width	**4**	**8**	**16**	**32**	**64**	**128**	**256**	**512**	**1024**
SNR_{imf}	7.80	8.15	10.01	9.62	10.84	11.54	9.80	10.57	10.4
$RMSE$	0.81	0.78	0.64	0.67	0.58	0.54	0.65	0.60	0.62

Core Findings: The analysis revealed that a kernel width of 128 optimized the ADAE's denoising capability, striking a balance between detailed local feature extraction and the integration of global structural information. While narrower and wider kernels exhibit unique advantages, they do not match the comprehensive denoising efficiency achieved with a medium-sized kernel width.

Conclusion and Optimization Guidance: This section's expanded explanation not only elucidates the direct impact of kernel width on model performance but also offers guidance for future researchers in adjusting this parameter for different tasks and datasets. The careful calibration of the initial convolutional kernel width is thus emphasized as a pivotal design consideration in deep learning models, enriching the understanding of its role in achieving optimal performance.

4.2.3. Ablation Study of Shortcut Connections

In this ablation study, the integration of shortcut connections at various stages within the ADAE model was meticulously investigated, targeting three strategic locations: Position I (connecting the input to the output of the encoder), Position II (linking the input of the encoder to the output of the decoder), and Position III (establishing a direct path from the input to the output of the decoder). The assessment encompassed eight distinct configurations, the outcomes of which are summarized in Table 3.

Table 3. Denoising performance of ADAE with shortcut connections at −6 dB.

Position	None	I	II	III	I & II	I & III	II & III	I & II & III
SNR_{imf}	9.34	11.13	11.25	19.26	11.64	17.00	15.54	16.27
$RMSE$	0.68	0.56	0.56	0.22	0.60	0.28	0.34	0.32

The remarkable performance of configurations featuring solely a Position III shortcut stands out, surpassing all other arrangements by a substantial margin. This superiority can be attributed to the direct transmission path from the input to the decoder output, which bypasses the potential accumulation of error in intermediate layers. By maintaining a purer representation of the input signal, Position III shortcuts facilitate more effective feature reconstruction, resulting in significantly reduced noise levels, as reflected in the lowest $RMSE$ and highest SNR_{imf}.

Combinations inclusive of Position III shortcuts consistently showed favorable performance, validating the strategic advantage of this connection scheme. Conversely, Position I and Position II shortcuts alone or in certain combinations provided less pronounced improvements. Consequently, our optimized ADAE model, depicted in Figure 3, exclusively integrates Position III shortcuts to harness its exceptional denoising capabilities. For completeness, the finalized network specifications are detailed in Table 4.

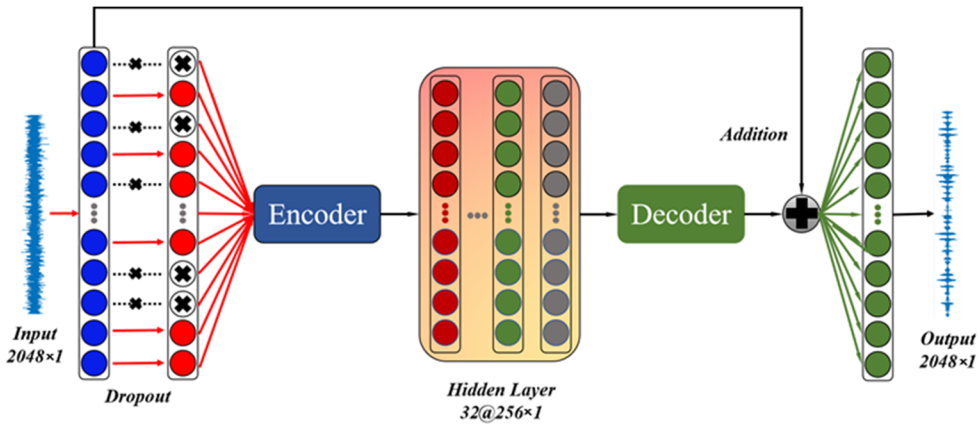


Figure 3. ADAE optimized structure.

4.2.4. Impact of Leaky ReLU Leakiness on Denoising Performance

Convolution and deconvolution operations are inherently linear processes. However, the relationship between signals and noise is complex, requiring nonlinear transformations for effective denoising. While rectified linear unit (ReLU) activation functions are commonly used in computer vision tasks, they tend to disregard negative values during signal denoising. Negative values are pivotal in denoising tasks, especially as signal means are normalized to zero, rendering negative values significant. ReLU simply nullifies negative values, resulting in considerable information loss.

To address this limitation, Leaky ReLU activation functions are introduced, which scale negative values by a leakage factor (α). In this section, we replace all ReLU activations

with Leaky ReLU and explore the impact of different α values on the denoising performance of the adversarial domain adaptation encoder (ADAE).

Table 4. Optimized ADAE network parameters.

Layer	Parameter's Description	Output Size
Input	-	1×2048
conv(128,512,2)	(Kernel width, Kernel number, Stride)	512×1024
conv(3,256,2)	(Kernel width, Kernel number, Stride)	256×512
conv(3,128,1)	(Kernel width, Kernel number, Stride)	128×512
conv(3,128,2)	(Kernel width, Kernel number, Stride)	128×512
conv(3,64,2)	(Kernel width, Kernel number, Stride)	64×256
fc(32)	(Output dimension)	
deconv(3,16,2)	(Kernel width, Kernel number, Stride)	16×512
deconv(3,8,2)	(Kernel width, Kernel number, Stride)	8×1024
deconv(3,4,1)	(Kernel width, Kernel number, Stride)	4×1024
deconv(3,4,2)	(Kernel width, Kernel number, Stride)	4×1024
deconv(3,1,2)	(Kernel width, Kernel number, Stride)	1×2048
regression	-	1×2048

As depicted in Figure 4, denoising performance initially increases with α, reaching a peak before diminishing returns are observed. This trend suggests that increasing the leakage factor accentuates the importance of negative values, thereby enhancing denoising effectiveness. ADAE achieves maximal denoising efficacy when α is set to 0.3. Beyond this value, denoising capability starts to deteriorate, indicating a delicate balance between capturing negative values and nonlinear learning capabilities.

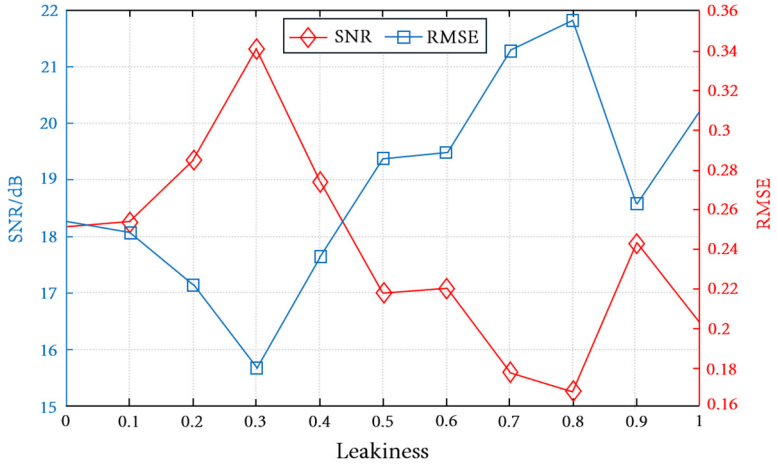


Figure 4. Denoising performance of ADAE with different leakiness under SNR = −6 dB.

In summary, a judicious selection of the leakage factor (α) is crucial to achieving optimal denoising performance in ADAE. The findings suggest that $\alpha = 0.3$ offers the best balance between capturing negative values and nonlinear learning capabilities, thereby maximizing denoising effectiveness. This optimal configuration is adopted for subsequent experiments to ensure robust denoising performance.

4.3. Comparative Analysis

4.3.1. Known Noise Intensity

To underscore the superiority of the proposed method, we conducted a comparative evaluation with several state-of-the-art and traditional denoising techniques, including WT [18], EMD [13], JL-CNN [4], SEAEFD [38], and NL-FCNN [11]. The denoising performance of ADAE was assessed under varying SNRs of −6 dB, −3 dB, and 0 dB, with the results summarized in Table 5.

Table 5. Denoising results of different methods under constant noise intensity.

SNR	Metric	WT	EMD	JL-CNN	SEAEFD	NL-FCNN	Proposed
−6 dB	SNR_{imf}	7.32 ± 0.84	7.86 ± 0.81	13.32 ± 1.48	11.47 ± 1.15	10.22 ± 0.91	22.14 ± 2.61
	$RMSE$	0.91 ± 0.17	0.88 ± 0.15	0.59 ± 0.17	0.64 ± 0.22	0.60 ± 0.32	0.16 ± 0.05
−3 dB	SNR_{imf}	4.45 ± 0.98	5.14 ± 1.04	9.48 ± 0.83	7.77 ± 0.71	8.60 ± 0.96	16.70 ± 2.37
	$RMSE$	0.88 ± 0.18	0.81 ± 0.16	0.47 ± 0.13	0.52 ± 0.16	0.50 ± 0.30	0.21 ± 0.06
0 dB	SNR_{imf}	1.84 ± 1.32	3.01 ± 1.26	5.10 ± 0.63	4.25 ± 0.94	7.20 ± 0.90	15.66 ± 2.60
	$RMSE$	0.84 ± 0.18	0.74 ± 0.15	0.40 ± 0.09	0.52 ± 0.31	0.40 ± 0.29	0.17 ± 0.05

ADAE demonstrates significantly superior denoising performance compared to other methods across all noise intensities, with SNR_{imf} nearly doubling or more in comparison. Even when noise intensity is low, ADAE achieves an SNR_{imf} greater than 15, showcasing its robustness and ability to learn noise characteristics while preserving original fault information.

To provide a more intuitive understanding of ADAE's denoising efficacy, we visualize both temporal and frequency waveforms of signals under the three noise intensities. Figure 5 depicts the denoising results of ADAE under SNR = −6 dB, when noise intensity is known.

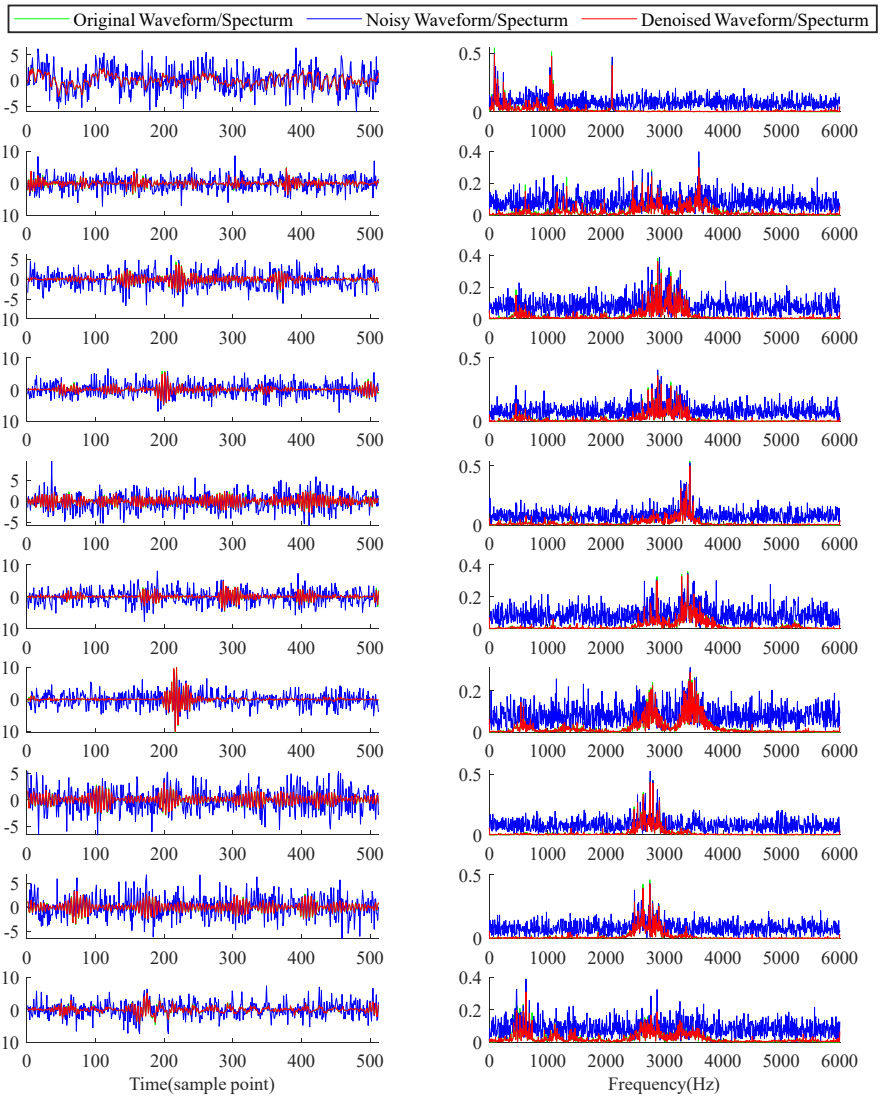


Figure 5. Denoising results of ADAE with SNR = −6 dB when noise intensity is known.

These visualizations underscore ADAE's remarkable denoising capabilities, effectively removing noise while preserving fault information across various fault modes and noise intensities. The distinct denoising effects observed in the frequency spectra demonstrate ADAE's ability to remove irrelevant frequency components, further validating its effectiveness even under challenging noise conditions.

4.3.2. Unknown Noise Intensity

To further evaluate the denoising performance of ADAE in real-world scenarios where noise levels are unpredictable, we conducted experiments with randomly varied SNR ranging from 0 dB to −6 dB. The results, compared with other denoising methods, are presented in Table 6.

Table 6. Denoising results of different methods under random noise intensity.

SNR	Metric	WT	EMD	JL-CNN	SEAEFD	NL-FCNN	Proposed
−6 dB	SNR_{imf}	5.12 ± 0.79	6.24 ± 0.95	14.88 ± 2.14	14.57 ± 1.69	12.85 ± 1.41	24.02 ± 2.20
	$RMSE$	0.93 ± 0.28	0.90 ± 0.21	0.33 ± 0.11	0.51 ± 0.20	0.57 ± 0.25	0.13 ± 0.04
−3 dB	SNR_{imf}	2.88 ± 0.75	4.44 ± 1.62	12.15 ± 0.71	10.47 ± 0.75	7.58 ± 0.88	22.39 ± 2.45
	$RMSE$	0.84 ± 0.35	0.48 ± 0.22	0.39 ± 0.17	0.52 ± 0.16	0.68 ± 0.33	0.16 ± 0.05
0 dB	SNR_{imf}	1.11 ± 0.58	1.23 ± 0.54	8.47 ± 0.55	6.78 ± 0.62	5.74 ± 0.85	22.07 ± 3.02
	$RMSE$	0.56 ± 0.18	0.45 ± 0.12	0.42 ± 0.24	0.47 ± 0.26	0.54 ± 0.35	0.17 ± 0.06

In all tested scenarios, ADAE consistently showcases superior denoising capabilities. When trained on a dataset with randomly varied SNR, ADAE exhibits even better performance compared to the previous scenario with constant noise intensity. Particularly notable is the significant enhancement in the SNR_{imf} metric under 0 dB and −3 dB conditions, where ADAE achieves values surpassing 22, representing an improvement of 7 compared to the previous results. This substantial improvement underscores ADAE's adaptability and effectiveness in handling varying noise levels.

Furthermore, relative to other denoising methods, ADAE consistently outperforms them across all tested scenarios, highlighting its unparalleled advantages in signal denoising. This reaffirms ADAE's robustness and generality, making it a highly reliable solution for denoising tasks in diverse environments.

To provide further insights into ADAE's denoising efficacy, temporal and frequency waveforms of test samples under an SNR of −6 dB are presented in Figure 6. These samples encompass signals from ten bearing health states under 1 hp load. As observed in the figures, ADAE successfully preserves temporal fault features while effectively removing irrelevant frequency components. This capability underscores ADAE's role as an efficient preprocessing method for rolling bearing fault diagnosis, particularly in environments characterized by strong noise backgrounds.

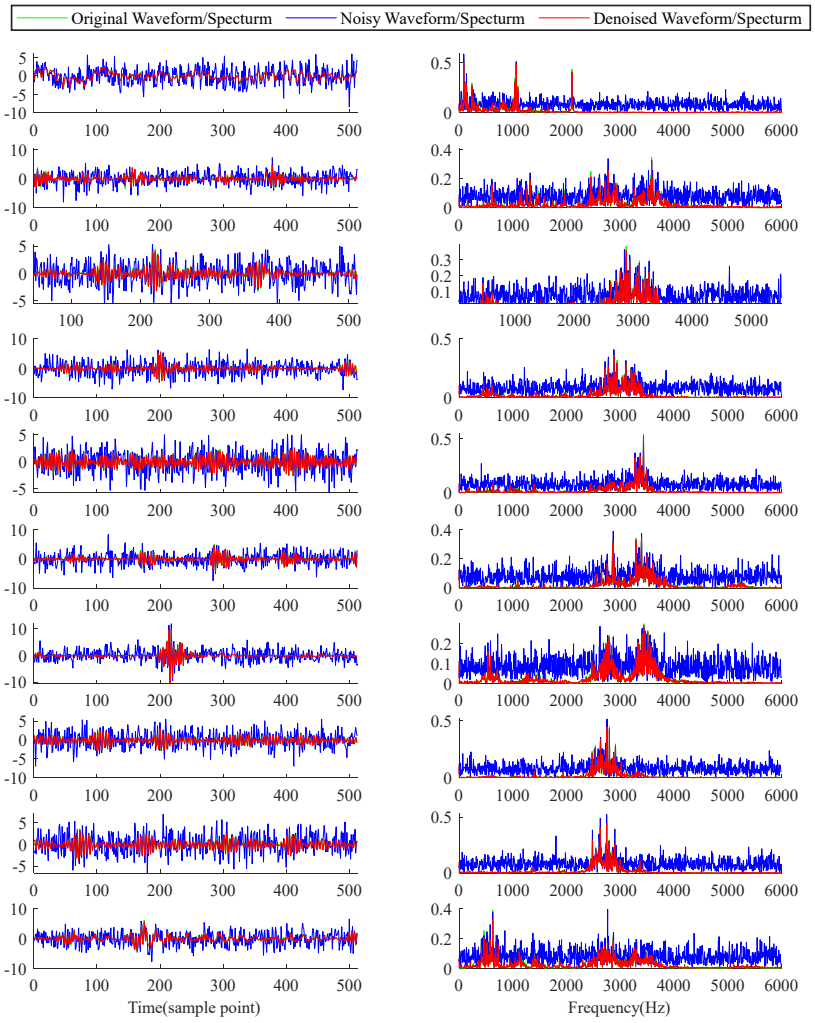


Figure 6. Denoising results of ADAE with SNR = −6 dB when noise intensity is unknown.

5. Conclusions

This study pioneers the adaptive denoising autoencoder (ADAE), an innovative framework designed to mitigate noise in vibration signals with unprecedented effectiveness. By introducing an adaptive shrinkage unit with local attention mechanisms within its architectural core, ADAE innovatively discriminates between noise and fault signatures, preserving vital diagnostic information while suppressing irrelevant noise. The strategic integration of a dynamic dropout strategy further augments its adaptability, endowing the model with versatility across diverse noise profiles.

Our research breaks new ground by exhaustively examining the influence of Leaky-ReLU activation functions and convolutional kernel dimensions on denoising efficacy, contributing fresh insights to the design of neural network architectures for signal processing. Moreover, the exploration of shortcut connections validates their utility in enhancing feature preservation amidst noise suppression.

The empirical validations affirm ADAE's dominance over conventional denoising techniques, evidencing exceptional competence in isolating and removing noise in complex, real-world conditions. Its proficiency in both time and frequency domains, particularly in low SNR environments, underscores its potential as a transformative tool in the realm of fault diagnosis, especially critical for bearings operating under heavy noise interference.

In conclusion, the proposed ADAE framework stands as a groundbreaking advancement in the realm of signal denoising, offering unparalleled performance in challenging noise scenarios and charting a new course for enhancing diagnostic accuracy under severe

noise interference, thereby solidifying its value as a pivotal preprocessing technology for robust machinery health monitoring.

Author Contributions: Conceptualization, H.L. and K.Z.; methodology, K.Z. and L.H.; writing—original draft preparation, H.L.; project administration, K.Z.; funding acquisition, K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Joint Funds of Equipment Advance Research of China, grant number 6141B02040207.

Data Availability Statement: The data presented in this study are available are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Singh, G.K.; Ahmed Saleh Al Kazzaz, S.A. Induction machine drive condition monitoring and diagnostic research—A survey. *Electr. Power Syst. Res.* **2003**, *64*, 145–158. [CrossRef]
2. Wang, Q.; Xu, F.Y.; Ma, T.C. Wavelet packet decomposition with motif patterns for rolling bearing fault diagnosis under variable working loads. *J. Vib. Control* **2024**. [CrossRef]
3. Ju, Y.M.; Tian, X.; Liu, H.J.; Ma, L.F. Fault detection of networked dynamical systems: A survey of trends and techniques. *Int. J. Syst. Sci.* **2021**, *52*, 3390–3409. [CrossRef]
4. Wang, H.; Liu, Z.L.; Peng, D.D.; Cheng, Z. Attention-guided joint learning CNN with noise robustness for bearing fault diagnosis and vibration signal denoising. *ISA Trans.* **2022**, *128*, 470–484. [CrossRef] [PubMed]
5. AU Khoukhi, A.; Khalid, M.H. Hybrid computing techniques for fault detection and isolation, a review. *Comput. Electr. Eng.* **2015**, *43*, 17–32. [CrossRef]
6. Amezquita-Sanchez, J.P.; Adeli, H. Signal Processing Techniques for Vibration-Based Health Monitoring of Smart Structures. *Arch. Comput. Methods Eng.* **2016**, *23*, 1–15. [CrossRef]
7. Zhang, C.W.; Mousavi, A.A.; Masri, S.F.; Gholipour, G.; Yan, K.; Li, X. Vibration feature extraction using signal processing techniques for structural health monitoring: A review. *Mech. Syst. Signal Process.* **2022**, *177*, 109175. [CrossRef]
8. Du, W.L.; Yang, L.; Wang, H.; Gong, X.; Zhang, L.; Li, C.; Ji, L. LN-MRSCAE: A novel deep learning based denoising method for mechanical vibration signals. *J. Vib. Control* **2024**, *30*, 459–471. [CrossRef]
9. Xu, J.; Zhang, H.; Sun, C.K.; Shi, Y.H.; Shi, G.C. Tensor-Based Denoising on Multi-dimensional Diagnostic Signals of Rolling Bearing. *J. Vib. Eng. Technol.* **2023**, *12*, 1263–1275. [CrossRef]
10. Wang, R.; Ding, X.; He, D.; Li, Q.; Li, X.; Tang, J.; Huang, W. Shift-Invariant Sparse Filtering for Bearing Weak Fault Signal Denoising. *IEEE Sens. J.* **2023**, *23*, 26096–26106. [CrossRef]
11. Han, H.R.; Wang, H.; Liu, Z.L.; Wang, J.Y. Intelligent vibration signal denoising method based on non-local fully convolutional neural network for rolling bearings. *ISA Trans.* **2022**, *122*, 13–23. [CrossRef]
12. Lei, Y.G.; Lin, J.; He, Z.J.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126. [CrossRef]
13. Mohguen, W.; Bekka, R.E. EMD-based denoising by customized thresholding. In Proceedings of the International Conference on Control, Automation and Diagnosis (ICCAD), Hammamet, Tunisia, 19–21 January 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 19–23.
14. Liu, W.; Liu, Y.; Li, S.X.; Chen, Y.K. A Review of Variational Mode Decomposition in Seismic Data Analysis. *Surv. Geophys.* **2023**, *44*, 323–355. [CrossRef]
15. Yang, J.; Zhou, C.; Li, X.; Pan, A.; Yang, T. A Fault Feature Extraction Method Based on Improved VMD Multi-Scale Dispersion Entropy and TVD-CYCBD. *Entropy* **2023**, *25*, 277. [CrossRef] [PubMed]
16. Cheng, Z.Q. Extraction and diagnosis of rolling bearing fault signals based on improved wavelet transform. *J. Meas. Eng.* **2023**, *11*, 420–436. [CrossRef]
17. Xi, C.; Gao, Z. Fault Diagnosis of Rolling Bearings Based on WPE by Wavelet Decomposition and ELM. *Entropy* **2022**, *24*, 1423. [CrossRef] [PubMed]
18. Al-Raheem, K.F.; Roy, A.; Ramachandran, K.P.; Harrison, D.K.; Grainger, S. Rolling element bearing faults diagnosis based on autocorrelation of optimized: Wavelet de-noising technique. *Int. J. Adv. Manuf. Technol.* **2009**, *40*, 393–402. [CrossRef]
19. Wang, J.X.; Tang, X.B. Wavelet Denoise Method Applied in Load Spectrum Analysis of Engineering Vehicles. *Adv. Mater. Res.* **2010**, *108–111*, 1320–1325. [CrossRef]
20. Fu, S.; Wu, Y.; Wang, R.; Mao, M. A Bearing Fault Diagnosis Method Based on Wavelet Denoising and Machine Learning. *Appl. Sci.* **2023**, *13*, 5936. [CrossRef]
21. Zhang, X.; Li, J.; Wu, W.; Dong, F.; Wan, S. Multi-Fault Classification and Diagnosis of Rolling Bearing Based on Improved Convolution Neural Network. *Entropy* **2023**, *25*, 737. [CrossRef]

22. Wang, Q.; Xu, F.Y. A novel rolling bearing fault diagnosis method based on Adaptive Denoising Convolutional Neural Network under noise background. *Measurement* **2023**, *218*, 113209. [CrossRef]
23. Zhou, H.; Liu, R.D.; Li, Y.X.; Wang, J.C.; Xie, S.C. A rolling bearing fault diagnosis method based on a convolutional neural network with frequency attention mechanism. *Struct. Health Monit. -Int. J.* **2023**. [CrossRef]
24. Zhang, Q.; Deng, L.F. An Intelligent Fault Diagnosis Method of Rolling Bearings Based on Short-Time Fourier Transform and Convolutional Neural Network. *J. Fail. Anal. Prev.* **2023**, *23*, 795–811. [CrossRef]
25. Liu, X.L.; Lu, J.N.; Li, Z. Multiscale Fusion Attention Convolutional Neural Network for Fault Diagnosis of Aero-Engine Rolling Bearing. *IEEE Sens. J.* **2023**, *23*, 19918–19934. [CrossRef]
26. 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), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.
27. Wang, J.; Tavakoli, H.R.; Laaksonen, J. Fixation Prediction in Videos Using Unsupervised Hierarchical Features. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 2225–2232.
28. Gu, X.; Tian, Y.; Li, C.; Wei, Y.; Li, D. Improved SE-ResNet Acoustic–Vibration Fusion for Rolling Bearing Composite Fault Diagnosis. *Appl. Sci.* **2024**, *14*, 2182. [CrossRef]
29. Zhou, J.; Yang, X.; Li, J. Deep Residual Network Combined with Transfer Learning Based Fault Diagnosis for Rolling Bearing. *Appl. Sci.* **2022**, *12*, 7810. [CrossRef]
30. Bhatt, D.; Patel, C.; Talsania, H.; Patel, J.; Vaghela, R.; Pandya, S.; Ghayvat, H. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. *Electronics* **2021**, *10*, 2470. [CrossRef]
31. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Chen, T. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]
32. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
33. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [CrossRef]
34. Cui, M.L.; Wang, Y.Q.; Lin, X.S.; Zhong, M.Y. Fault Diagnosis of Rolling Bearings Based on an Improved Stack Autoencoder and Support Vector Machine. *IEEE Sens. J.* **2021**, *21*, 4927–4937. [CrossRef]
35. Che, C.C.; Wang, H.W.; Fu, Q.; Ni, X.M. Intelligent fault prediction of rolling bearing based on gate recurrent unit and hybrid autoencoder. *Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci.* **2021**, *235*, 1106–1114. [CrossRef]
36. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Identity Mappings in Deep Residual Networks. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 630–645.
37. Zhao, M.H.; Zhong, S.S.; Fu, X.Y.; Tang, B.P.; Pecht, M. Deep Residual Shrinkage Networks for Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4681–4690. [CrossRef]
38. Wu, H.; Li, J.M.; Zhang, Q.Y.; Tao, J.X.; Meng, Z. Intelligent fault diagnosis of rolling bearings under varying operating conditions based on domain-adversarial neural network and attention mechanism. *ISA Trans.* **2022**, *130*, 477–489. [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.