

Article

Gearbox Fault Diagnosis Based on Gramian Angular Field and CSKD-ResNeXt

Yanlin Liu, Shuihai Dou, Yanping Du and Zhaohua Wang *

Department of Mechanical and Electrical Engineering, Beijing Institute of Graphic Communication, No. 1, Xinghua Street, Daxing District, Beijing 102699, China; wn25816482021@163.com (Y.L.); doushuihai@bigc.edu.cn (S.D.); duyanning@bigc.edu.cn (Y.D.)

* Correspondence: 18515525865@163.com; Tel.: +86-10-185-1552-5865

Abstract: For most rotating mechanical transmission systems, condition monitoring and fault diagnosis of the gearbox are of great significance to avoid accidents and maintain stability in operation. To strengthen the comprehensiveness of feature extraction and improve the utilization rate of fault signals to accurately identify the different operating states of a gearbox, a gearbox fault diagnosis model combining Gramian angular field (GAF) and CSKD-ResNeXt (channel shuffle and kernel decomposed ResNeXt) was proposed. The original one-dimensional vibration signal of the gearbox was converted into a two-dimensional image by GAF transformation, and the image was used as the input of the subsequent diagnosis network. To solve the problem of channel independence and incomplete information caused by group convolution, the idea of channel shuffle is introduced to enable the branches of the group convolution part to establish information exchange. In addition, to improve the semantic expression ability of the model, the convolutional kernel of the network backbone is split and replaced. The model is verified under the different working conditions of the gearbox and compared with other methods. The experimental results show that the diagnostic accuracy of the model is up to 99.75%, and the precise identification of gearbox faults is realized.

Keywords: fault diagnosis; gearbox; GAF; ResNeXt; vibration signal



Citation: Liu, Y.; Dou, S.; Du, Y.; Wang, Z. Gearbox Fault Diagnosis Based on Gramian Angular Field and CSKD-ResNeXt. *Electronics* **2023**, *12*, 2475. <https://doi.org/10.3390/electronics12112475>

Academic Editors: Chiara Boccaletti, Séjir Khojet El Khil and Monia Ben Khader Bouzid

Received: 23 April 2023

Revised: 21 May 2023

Accepted: 23 May 2023

Published: 31 May 2023



Copyright: © 2023 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

1.1. Motivation

Rotary machinery is mainly used to drive mechanical equipment and plays a crucial role in the mechanical equipment. Therefore, the reliability and safety requirements for rotating machinery are extremely high [1]. At the same time, with the rapid development of intelligent manufacturing, mechanical equipment tends to have high precision and high reliability. The failure of a component often causes a chain reaction, leading to a severe accident, which greatly increases the economic cost of equipment operation [2–4]. The gear is the one of three basic components of rotating machinery. Gear failure accounts for a large proportion of mechanical failures [5]. Therefore, it is very crucial to accurately identify gearbox state and diagnose and predict gearbox fault [6,7].

1.2. Analysis of Related Works

At present, fault diagnosis methods are mainly divided into three types, which include model-based fault diagnosis methods, signal-processing-based fault diagnosis methods, and data-driven fault diagnosis methods. In the data-driven fault diagnosis methods, the methods can be further divided into two types, namely, traditional machine learning fault diagnosis methods and deep learning fault diagnosis methods.

Model-based fault diagnosis methods utilize the correlation between gearbox fault features and physical model, analyze its fault mechanism to build and optimize the model, and realize real-time fault diagnosis and prediction [8]. However, it is difficult to establish

an accurate gearbox model in practical application, which greatly limits the application of model diagnosis methods.

The fault diagnosis methods based on signal processing determine the effective indicators of diagnosis by analyzing the correlation between signals and faults. Fault diagnosis is achieved by constructing fault features through the dimensional and dimensionless indicators of signals [9,10]. However, the working conditions of the gearbox are complex and changeable, and the selected features are difficult to use in different conditions. Therefore, mining the commonness of fault data in massive data is an effective means of fault diagnosis.

In recent years, due to the large increase in training resources and the rapid development of computing power, data-driven fault diagnosis methods have gradually attracted more attention [11,12]. The development of machine learning algorithms provides a new path for gearbox fault diagnosis. According to the signal processing technology, the signal is analyzed and the feature vector that can effectively express the fault is constructed. Then, a machine learning algorithm is adopted for intelligent fault diagnosis, such as support vector machine (SVM) [13], KNN [14], random forest [15], etc.

However, in traditional machine learning algorithms, the screening and extraction of fault features still rely on manual operation, which brings uncertainty to fault diagnosis and fails to achieve the purpose of real intelligent diagnosis. The deep learning method with a powerful feature-learning ability can realize automatic feature extraction and fault classification, so it is widely used in the fault diagnosis field [16].

The input of the diagnosis model based on deep learning includes two fault sample types, which are the one-dimensional (1D) vibration signal and two-dimensional (2D) image [17]. The former directly extracts fault features from 1D vibration signals for diagnosis, while the latter combines signal processing technology to convert vibration signals into 2D images. Many studies have utilized signal preprocessing technology to improve sample quality in the conversion process. It has abundant data and strong computing power in the current fault diagnosis field. The fault feature extraction method, which inputs fault image samples into a deep learning model, is a necessary choice for accurate fault identification.

The analysis of the above fault diagnosis methods is shown in Table 1.

Table 1. Analysis of fault diagnosis methods.

Fault Diagnosis Methods	Diagnostic Limitations	Related Researches
Model-based fault diagnosis method	Fault mechanism and physical models are combined to analyze the nature of the fault but are more applicable to systems that can be modeled accurately.	Saxena A et al. (2016) [18] Sanchez H et al. (2015) [19] Sun et al. (2020) [20]
Signal processing-based fault diagnosis method	It does not need to rely on a large amount of data and also has better performance for signals with low SNR. However, the signal processing method is localized, and different research objects usually correspond to different fault diagnosis indexes.	Shanbr S et al. (2018) [21] Wang et al. (2017) [22] Lv et al. (2014) [23] Misael Lopez-Ramirez et al. (2016) [24] Tang et al. (2021) [25]
Traditional machine learning-based fault diagnosis method	Machine learning algorithms inject intelligence into the field of fault diagnosis, but the feature extraction process and classification task are two independent subjects. How to extract the optimal features is still a problem that many researchers are paying attention to.	Zeng et al. (2020) [26] Wang et al. (2021) [27] Toma R N et al. (2021) [28] Pang et al. (2021) [29]
Deep-learning-based fault diagnosis method	It has low computational complexity and is suitable for real-time and low-cost applications, but the applicability of one-dimensional signals and most network structures is poor. The internal setup of the model is the problem facing to improve the applicability of one-dimensional diagnostic model.	Wang et al. (2019) [30] Yu et al. (2020) [31] Zhou et al. (2020) [32] Yang et al. (2022) [33]
	One-dimensional signal as input	
	The signal is converted into a two-dimensional image as input	Xu et al. (2023) [34] Wang et al. (2017) [35] Zhang et al. (2020) [36] Huang et al. (2023) [37]

Based on current research, common deep learning models include long short-term memory (LSTM) [38], convolutional neural network (CNN) [39], recurrent neural networks (RNN) [40], artificial neural network (ANN) [41], etc. Because of its powerful feature extraction and classification ability in the face of complex data, CNN has been widely used in the field of fault identification [42]. However, gradient dispersion/explosion will occur in some networks when the depth of the network is increased, such as AlexNet [43] and VGG [44]. The proposal of BatchNorm can alleviate the gradient problem to a certain extent [45], but there is network degradation. The problem of network degradation was solved by ResNet [46], as proposed by He et al. in 2017, but it increased the difficulty of network design and the cost of calculation. ResNeXt (suggesting the next dimension) [47] adopted the residual module and added the ideas of group convolution and stack to reduce the number of hyperparameters and calculation cost on the basis of ensuring accuracy.

ResNeXt has been applied in various recognition and classification tasks because of its strong comprehensive performance. Gao et al. [48] used ResNeXt50 to identify individual underwater fish. Zhang et al. [49] used ResNeXt-50 as a backbone network to detect an abnormal object in X-ray images. Wang et al. [50] identified the degree of maize disease occurrence by ResNeXt101. Fang et al. [51] realized accurate recognition of dynamic gesture by using ResNeXt. All the above studies gave full play to ResNeXt's excellent image recognition ability and achieved ideal experimental results. Therefore, it is effective to apply ResNeXt to gearbox fault identification and classification.

1.3. Contributions

Since ResNeXt has an excellent learning ability in "vision", it is adopted in this paper as a diagnosis model for gearbox faults, which can make up for the gaps in its application in the field of fault diagnosis. In view of this, a diagnostic method based on GAF (Gramian angle field) [52] and CSKD-ResNeXt is proposed in this paper. Using GAF, a signal conversion method that can preserve the correlation between signal and time and effectively express valid fault information, one-dimensional gear box vibration data are converted into two-dimensional images as the input of ResNeXt. To enable the diagnosis model to learn the gearbox state information more comprehensively, the structure of ResNeXt was optimized to improve the gearbox fault feature extraction ability. The main contributions of this paper are as follows:

- i. In this paper, a Gramian image is used as the sample diagram of model input. After comparing the performance of GADF (Gramian angular difference field) and GASF (Gramian angular summation field), one with good effect is selected to process one-dimensional vibration signals, and the output two-dimensional sample image is used to express time-dependent signal characteristics.
- ii. The 7×7 convolutional kernel in the backbone of the ResNeXt model was decomposed into three 3×3 convolutional kernels, which reduced the feature extraction ambiguity caused by a large convolutional kernel and improved model semantic capability. After receiving vibration signals, the convolution kernel can extract more accurate and detailed feature information and improve the diagnostic accuracy.
- iii. For the purpose of feature communication, channel shuffle is added to the group convolution part to break the isolation between channels and exchange data. The data flow in the model is enriched to obtain a more competitive feature-mining capability. In addition, the process of fault identification and classification is demonstrated by using t-SNE visual dimension reduction.

The remaining sections are outlined as follows. Section 2 provides the methods of GAF and CSKD-ResNeXt. The introduction and partitioning of the dataset is illustrated in Section 3. Experimental results are shown in Section 4. Finally, concluding remarks are given in Section 5.

2. Methods

In this section, the preliminary knowledge of GAF and ResNeXt and their significance are first introduced, and then, the problems existing when ResNeXt is used as a fault diagnosis model are analyzed. Finally, channel shuffle and kernel decomposed are introduced to establish CSKD-ResNeXt.

2.1. The GAF

The Gramian angle field uses a one-dimensional time series in the Cartesian coordinate system for numerical scaling [53], and the matrix based on polar coordinates encodes the time series into images to maintain the correlation between signals and time, and then uses trigonometric functions to generate a GAF matrix and convert it into two-dimensional images [54]. Suppose that the original time series has n values, $X = \{x_1, x_2, x_3, \dots, x_n\}$, and the sequence is normalized to between $[-1, 1]$ and $[0, 1]$, denoted as $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \dots, \tilde{x}_n\}$, and \tilde{x}_i is the value of the normalized time series. Map data to polar coordinates as \tilde{x}_{-1}^i and \tilde{x}_0^i :

$$\tilde{x}_{-1}^i = \frac{(x_i - \max X + (x_i - \min X))}{\max X - \min X} \quad (1)$$

$$\tilde{x}_0^i = \frac{x_i - \min X}{\max X - \min X} \quad (2)$$

The time series is represented in polar coordinates; in Equation (2), \tilde{x}_i is mapped to angle ϕ_i , and time stamp t_i is mapped to radius r_i .

$$\begin{cases} \phi_i = \arccos \tilde{x}_i, -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r_i = \frac{t_i}{N}, t_i \in N \end{cases} \quad (3)$$

where X is the time stamp, and the interval $[0, 1]$ is divided into N equal parts, so that the span of polar coordinate system is regularized. The encoding mapping of Equation (3) has two important properties. First, the transformation is bijective because $\cos(\phi_i)$ is monotonically decreasing at $\phi_i \in [0, \pi]$; there is a unique corresponding value in the polar coordinate system when given a time series, and its inverse mapping is unique. Second, the transformation preserves the time information, and the time value can be determined by the radius coordinates.

The correlation between each time point is defined using trigonometric difference or trigonometric sum:

$$G = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} = \tilde{X}' \cdot \tilde{X} - \sqrt{I - \tilde{X}^2} \cdot \sqrt{I - \tilde{X}^2} \quad (4)$$

$$G = \begin{bmatrix} \sin(\phi_1 - \phi_1) & \cdots & \sin(\phi_1 - \phi_n) \\ \sin(\phi_2 - \phi_1) & \cdots & \sin(\phi_2 - \phi_n) \\ \vdots & \ddots & \vdots \\ \sin(\phi_n - \phi_1) & \cdots & \sin(\phi_n - \phi_n) \end{bmatrix} = \sqrt{I - \tilde{X}^2}' \cdot \tilde{X} - \tilde{X}' \cdot \sqrt{I - \tilde{X}^2} \quad (5)$$

where $\phi_i (i = 1, 2, \dots, n)$ is the angle of the i th time point in polar coordinates, and I is the unit row vector. In the formula, the inner product is redefined, $\langle x, y \rangle = x \cdot y - \sqrt{1 - x^2} \cdot \sqrt{1 - y^2}$ and $\langle x, y \rangle = \sqrt{1 - x^2} \cdot y - x \cdot \sqrt{1 - y^2}$, and a penalty term is added to reduce the interference of noise.

Figure 1 shows the conversion process of vibration signals into images through GAF: (a) represents the time series containing 1000 vibration signal points, (b) represents the representation of vibration signal mapped in polar coordinates through Equation (3), and (c) represents the two-dimensional image of the final GAF transformation.

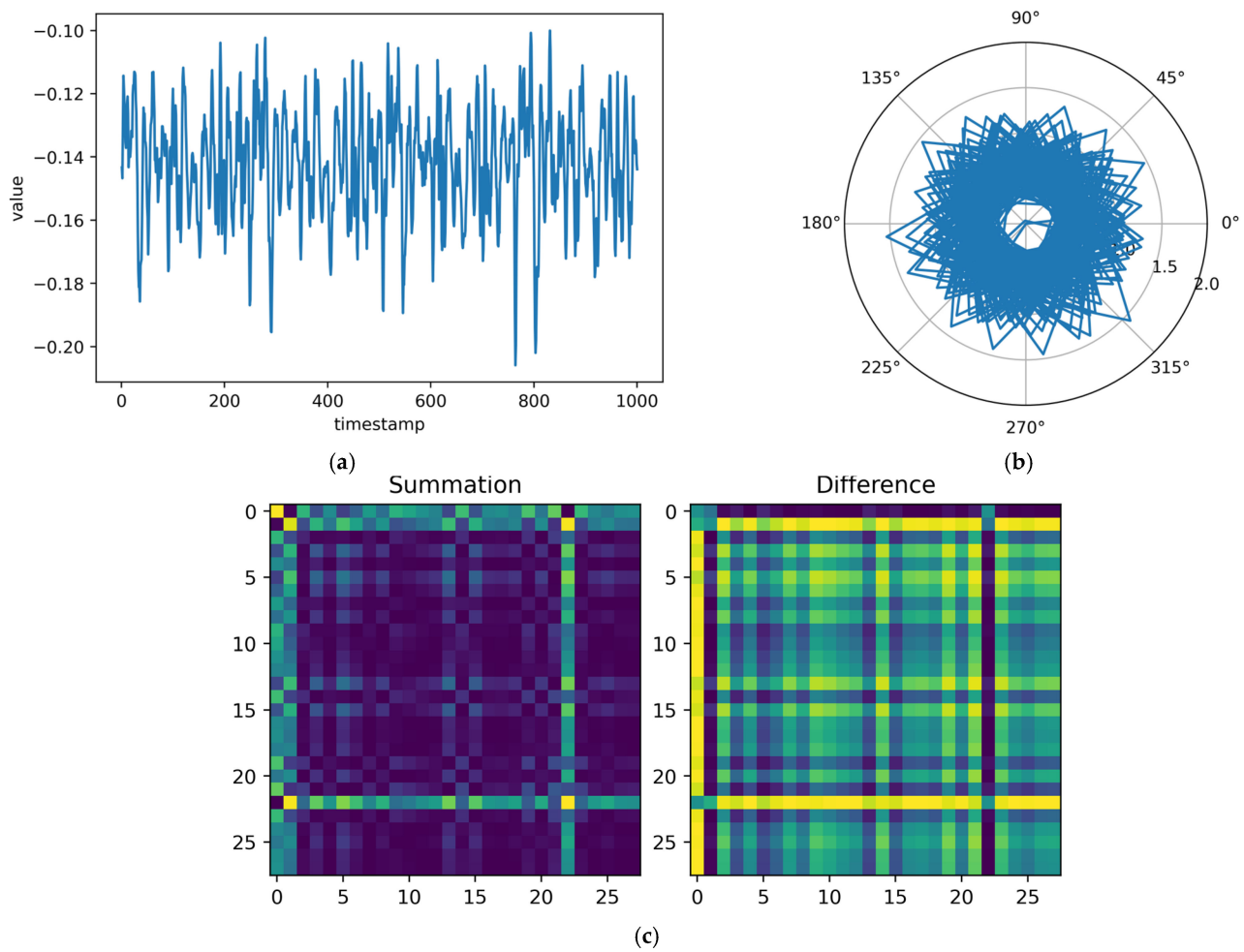


Figure 1. Conversion process of GAF: (a) original signal, (b) polar coordinates, and (c) Gramian angular field.

The advantage of the Gramian angle field in converting time-series data into image data is that it cannot only retain the complete information of the signal but also maintain the dependence of the signal on time. Then, the advantages of ResNeXt in image classification and recognition will be further made full use of for state recognition.

2.2. ResNeXt

On the basis of the residual structure, ResNeXt proposes a new dimension of cardinality and uses group convolution [55] to replace the three-layer convolution structure of ResNet, which not only improves the accuracy of the neural network but also reduces the parameter complexity so that ResNeXt performs better in neural network models with the same complexity. In addition, based on the ResNet structure of ResNeXt, the idea of parallel topology is introduced to increase the cardinality to 32, as shown in Figure 2. The residual part of ResNeXt is composed of grouping convolution, which makes ResNeXt more accurate and more efficient than ResNet.

The increase in cardinality means that the ResNeXt structure contains more parallel topologies, which can be seen in Equation (6):

$$out(x) = \sum_{i=1}^n (\omega * C(x)) \quad (6)$$

where ω is the weight of different topologies, $C(x)$ is the output value of the flat same topology, and n is the number of identical branches that a module has.

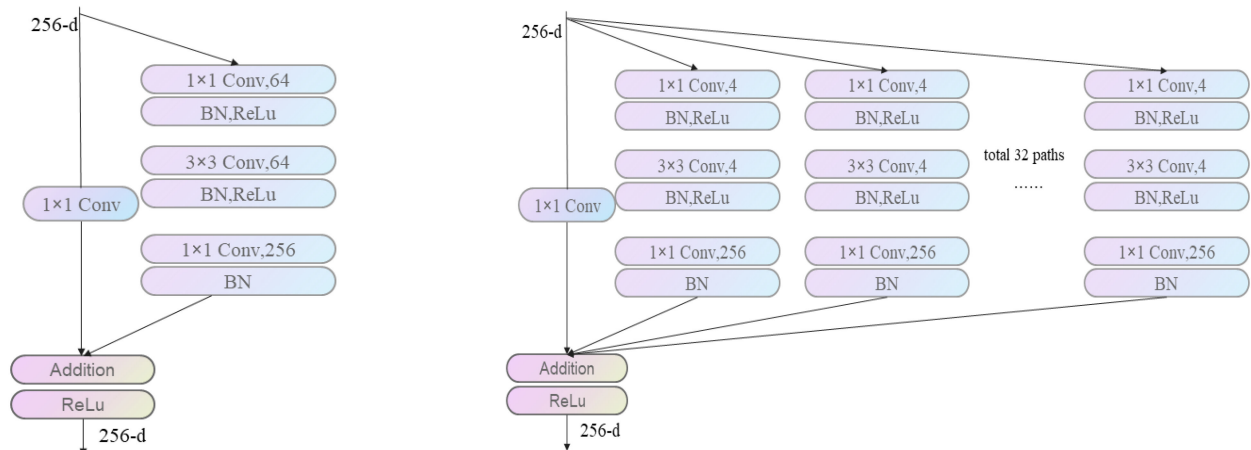


Figure 2. Residual block structure of ResNet and ResNeXt.

Although the introduction of cardinality improves the computational efficiency and identification accuracy of ResNeXt, it also brings the problem of channel independence. The independence of channels leads to the output only being derived from a small part of the input channels. As a result, there is no information flow between channels, and the generated features lack representativeness and weaken the model generalization ability. In addition, the first convolution layer in the ResNeXt backbone is the first place for feature extraction after sample input. Whether the extraction of sample information is comprehensive and accurate has a great influence on the subsequent processing. The large convolution kernel has less nonlinear ability than the small convolution kernel. In the case of the same inductive field, multiple small convolution layers have more nonlinear functions, which can make the decision function more deterministic and play the role of implicit regularization. In view of the above problems, this paper makes the following improvements:

(1) Channel Shuffle

ShuffleNet [56], proposed in 2017, solved the problem of feature graph communication between different groups caused by channel sparse connectivity, such as group convolution. Different from the dense pointwise convolution (which requires considerable complexity) adopted by Xception, MobileNet, and other networks, channel shuffle has no expensive calculation cost or high complexity and can make the input and output channels completely related. Therefore, channel shuffle is used in this paper to solve the grouping convolution problem of ResNeXt to provide help for information flow between channels. The main steps are as follows:

- i. Reshape: the input layer is assumed to be divided into g groups, and the total number of channels is $g \times n$. The input channel dimension is reshaped into two dimensions (g, n) , which represent the number of convolution groups and the number of channels contained in each convolution group.
- ii. Transpose: transpose two extended dimensions into (n, g) .
- iii. Flatten: the transposed channel flatten is reshaped into dimension $g \times n$, and channel shuffle can be finished.

After channel mixing, the feature graphs received by the subsequent group convolution layer from the previous layers are mixed fully correlated.

(2) Kernel Decomposed

By observing the structure of ResNeXt, it can be seen that the first layer in the input backbone consists of a 7×7 convolutional kernel, whose receptive fields are the same as those of three 3×3 convolutional kernels. The computational cost backbone of the convolution layer is the square of the width of the convolution kernel or the height of the convolution kernel, so the computational amount of one 7×7 convolution kernel is equal

to 5~6 times that of three 3×3 convolution kernels. Meanwhile, adding an activation function between the additional network layers can increase the nonlinear representation ability of the network when the receptive field is the same. On the premise that the details of the convolutional layer are not lost, the number of network model parameters is reduced, and the mining depth and feature precision of the model are improved.

2.3. Establishing the CSKD-ResNext Network

In this paper, feature graphs after group convolution are first “reorganized”, i.e., “uniformly disrupted”, to ensure that information can flow between different groups. Second, the 7×7 convolution kernel in the input backbone is replaced by three 3×3 convolution kernels, whose stride is 1; the output channel size is 64, and batch normalization is adopted. Meanwhile, the LeakyReLU function is adopted as the activation function in the convolution layer to solve the problems of neuron “extinction” and gradient disappearance caused by the Relu function. Set the output dimension of the full connection layer to 5, which corresponds to 5 different states of the gearbox. Table 2 below shows the details of the CSKD-ResNeXt network structure.

Table 2. Detailed structure of CSKD-ResNeXt.

Layer	Type	Output	Parameter
Conv1	Convolution	$64 \times 112 \times 112$	Three 3×3 Conv, stride = 1
Pool	MaxPooling	$64 \times 56 \times 56$	3×3 , Maxpool, stride = 2
Bottleneck1	Convolution	$256 \times 56 \times 56$	$\left(\begin{array}{c} 1 \times 1 \text{Conv} \\ 3 \times 3 \text{Conv, Group} = 32 \\ 1 \times 1 \text{Conv} \end{array} \right) \times 3$, stride = 1
Bottleneck2	Convolution	$512 \times 28 \times 28$	$\left(\begin{array}{c} 1 \times 1 \text{Conv} \\ 3 \times 3 \text{Conv, Group} = 32 \\ 1 \times 1 \text{Conv} \end{array} \right) \times 4$, stride = 1
Bottleneck3	Convolution	$1024 \times 14 \times 14$	$\left(\begin{array}{c} 1 \times 1 \text{Conv} \\ 3 \times 3 \text{Conv, Group} = 32 \\ 1 \times 1 \text{Conv} \end{array} \right) \times 6$, stride = 1
Bottleneck4	Convolution	$2048 \times 7 \times 7$	$\left(\begin{array}{c} 1 \times 1 \text{Conv} \\ 3 \times 3 \text{Conv, Group} = 32 \\ 1 \times 1 \text{Conv} \end{array} \right) \times 3$, stride = 1
Pool	MaxPooling	$2048 \times 1 \times 1$	Adaptive Average Pool
FC	Fully-connected	$2048 \times 1 \times 1$	Fc, Softmax

Based on the above, a gearbox fault diagnosis method based on GAF and CSKD-ResNeXt is proposed in this paper, as shown in Figure 3. The time series is converted into GASF and GADF images, which are used as the input of the subsequent convolutional network. On the basis of ResNeXt-50, the backbone convolutional kernel is split, and the channel shuffle is introduced to obtain CSKD-ResNeXt, which is used to extract features from the input sample graph. After three 3×3 convolutional kernels and four convolutional layers composed of different block numbers, deep learning and feature mining of samples in different states are carried out. The final predicted fault category is output through a softmax classifier after global pooling. The performance of the model was evaluated by the accuracy and loss of the test set. The t-sne scatter plot and confusion matrix were used to visually display the diagnostic process and results.

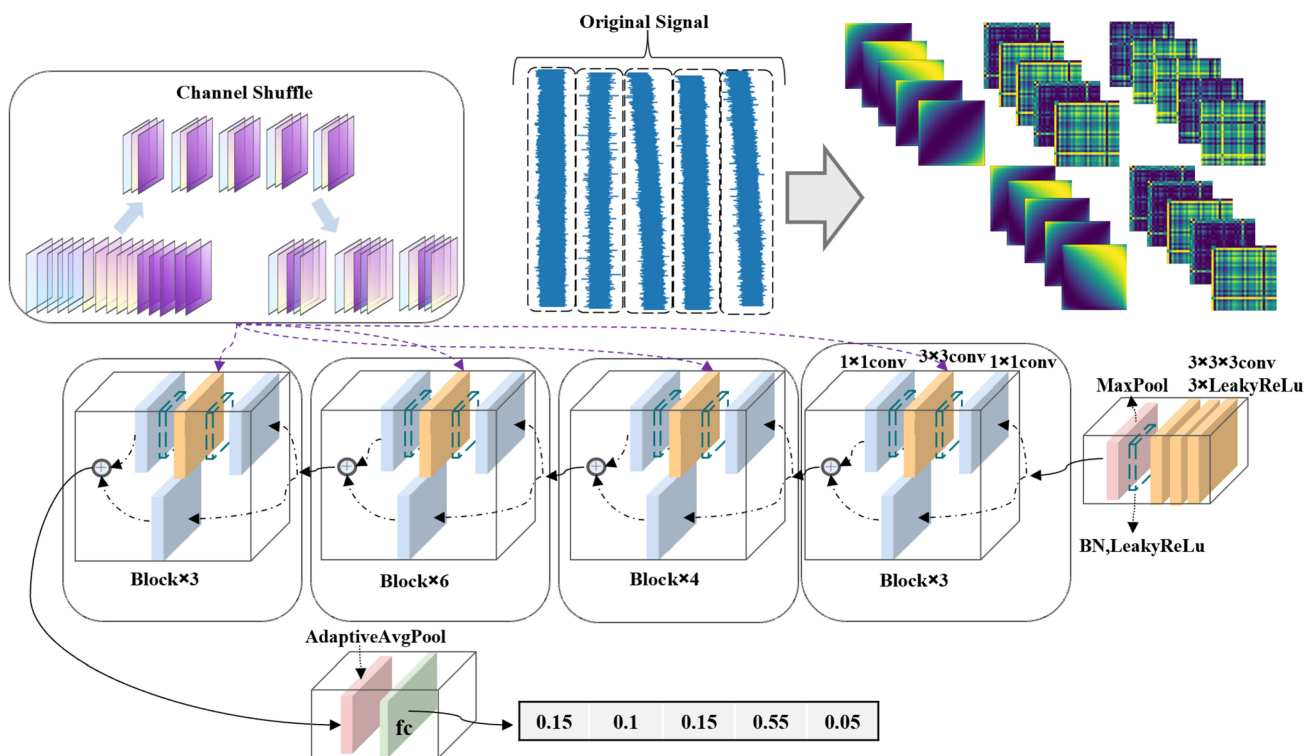


Figure 3. GAF-ResNeXt.

3. Data Description

In this section, the sources, types, working conditions, and other information of the datasets are introduced; the division of training sets and test sets and the number of sample sets are also shown in detail, and the configuration of the experimental platform and the common parameters of the operating framework are explained.

3.1. Datasets

The gearbox dataset in this paper comes from the experimental setup for a gearbox of Southeast University in China [57], as shown in Figure 4. The dataset includes 20 Hz–0 V and 30 Hz–2 V load conditions. The gearbox state has four fault states and one health state. Each state signal includes the vibration signal of the motor, motor torque, planetary gearbox in x, y, and z directions, and parallel gearbox in x, y, and z directions. The data types are shown in Table 3.

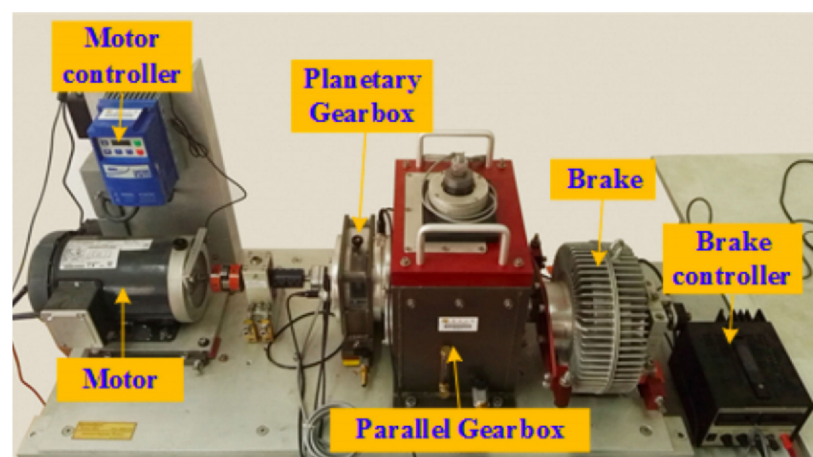


Figure 4. Experimental setup for gearbox.

Table 3. Partitioning of gearbox datasets.

Operating Condition		20 Hz–0 V		30 Hz–2 V	
Dataset Type		Training	Validation	Training	Validation
Health	normal state	666	166	666	166
Chipped	crack occurs in the feet	666	166	666	166
Miss	missing one of feet in the gear	666	166	666	166
Root	crack occurs in root of the gear feet	666	166	666	166
Surface	wear occurs in the surface of gear	666	166	666	166
Total	8320	3330	830	3330	830

Each of the 10,000 data points in the dataset was truncated to generate sample graphs with a size of 224×224 RGB three-channel. Under the working condition of 20 Hz–0 V, each fault has eight columns of vibration data corresponding to eight positions or directions of the gearbox test stand. Each column contains 1.04 million vibration data points and generates 104 images, so each failure type has 832 images. The operating data of 30 Hz–2 V is the same. The gearbox datasets under two working conditions contained a total of 8320 sample graphs, which were divided into training sets and test sets in a ratio of 4:1. Each fault type was composed of 1332 training samples and 332 test samples.

3.2. Experimental Platform Setting

The gearbox fault diagnosis model runs on the Pytorch framework, and the experimental platform is configured as follows: 64-bit Windows 10 operating system, Intel (R) Xeon (R) Gold 6330@ 2.00 GHz (CPU), RTX 3090 (24 GB) (GPU), and code written in the Python 3.8 environment. The adaptive moment estimation (Adam) algorithm [58] is used to update the network training parameters. The initial learning rate is 0.001. Use ReduceLROnPlateau to update the learning rate to achieve the self-attenuation process. It takes the accuracy of the test set as the adjustment indicator, and patience in ReduceLROnPlateau was set as 4 according to the results after repeated experiments. The model loss is calculated using cross entropy, and the dropout in the model is set to 0.2.

4. Analysis of Model Results

In this section, performance verification and comparison experiments are performed on the proposed model with accuracy, loss, and other indicators, including comparison between GADF and GASF, comparison between GAF and STFT and CWT, comparison between CSKD-ResNeXt and classical networks, and visualization of the fault classification identification process of key convolution layers.

4.1. Model Verification

To compare the effectiveness of the GADF and GASF methods, the images generated by GADF and GASF are respectively input into CSKD-ResNeXt, as shown in Table 4. In this paper, image samples of GADF are selected as the input of subsequent models.

Table 4. Comparison of GADF and GASF effects.

	20 Hz–0 V		30 Hz–2 V	
	Accuracy	Loss	Accuracy	Loss
GADF	0.998	0.016	0.993	0.013
GASF	0.984	0.021	0.980	0.024

The accuracy and loss on the test set of the fault diagnosis model proposed in this paper under two working conditions are shown in Figure 5. It can be seen that the model does not converge, and the accuracy fluctuates in the early stage of training. After about 40 epochs, the accuracy fluctuates between 95% and 100%; after 60 epochs, the accuracy of the two conditions tends to be stable and converges to 99.75% and 99.27%, respectively, while the loss gradually approaches zero. The accuracy and loss trends of the two conditions are not too different, which proves that the model has a certain generalization ability.

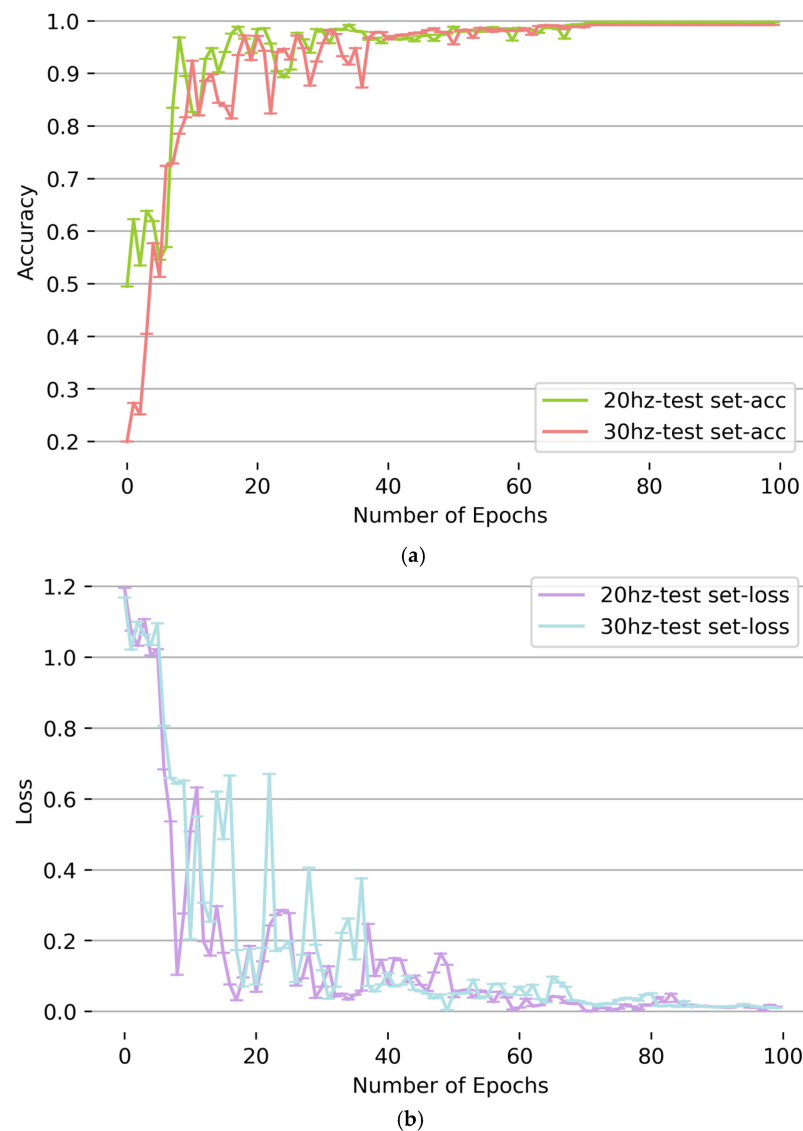


Figure 5. Accuracy and loss under two working conditions: (a) accuracy under two working conditions and (b) loss under two working conditions.

After testing, under the working condition of 20 Hz–0 V, 828 of the 830 gearbox image samples were correctly classified, and one sample of miss and root, respectively, was misjudged as normal. Under the working condition of 30 Hz–2 V, a total of six samples were misjudged, among which miss had the largest number of misjudged samples, four of which were misjudged as root. In addition, all health and surface samples were judged correctly, and the two confusion matrices are shown in Figure 6. In general, CSKD-ResNeXt can avoid state confusion and can identify different faults well.

To prove the superiority of GAF, the following fault diagnosis methods are selected to compare with the method in this paper. ① STFT+ CSKD-ResNeXt: a one-dimensional time series is converted into two-dimensional time-frequency image by STFT, and then, the two-dimensional image is used as the input of CSKD-ResNeXt. ② CWT+ CSKD-ResNeXt: a one-dimensional time series is processed into a two-dimensional time-frequency graph by continuous wavelet transform, and then, the time-frequency graph is used as the input of CSKD-ResNeXt. The fault diagnosis results of different methods are shown in Figure 7. The method in this paper (GAF+ CSKD-ResNeXt) has the best performance, while the accuracy of the STFT+ CSKD-ResNeXt method is the lowest, which is 94.06% and 95.85%, respectively, under two working conditions. The accuracy of the CWT+ CSKD-ResNeXt method can reach 96.75% and 97.69%. Therefore, GAF is used to process time-series signals

with higher accuracy, which further indicates that the two-dimensional image transformed by GAF can retain the relevant information between the original time-series data better.

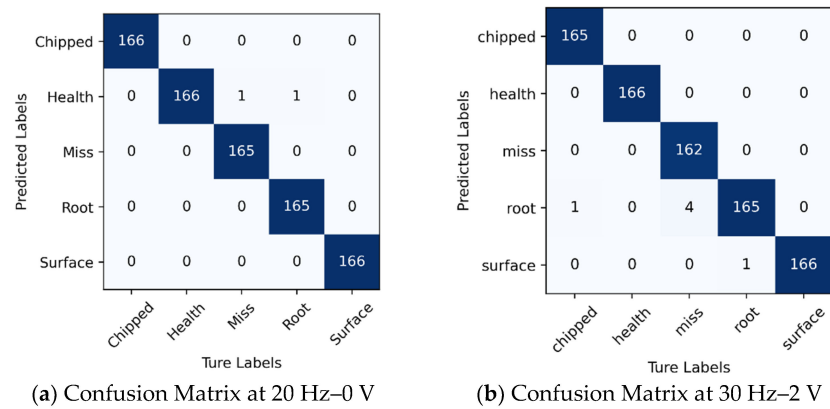


Figure 6. Confusion matrix.

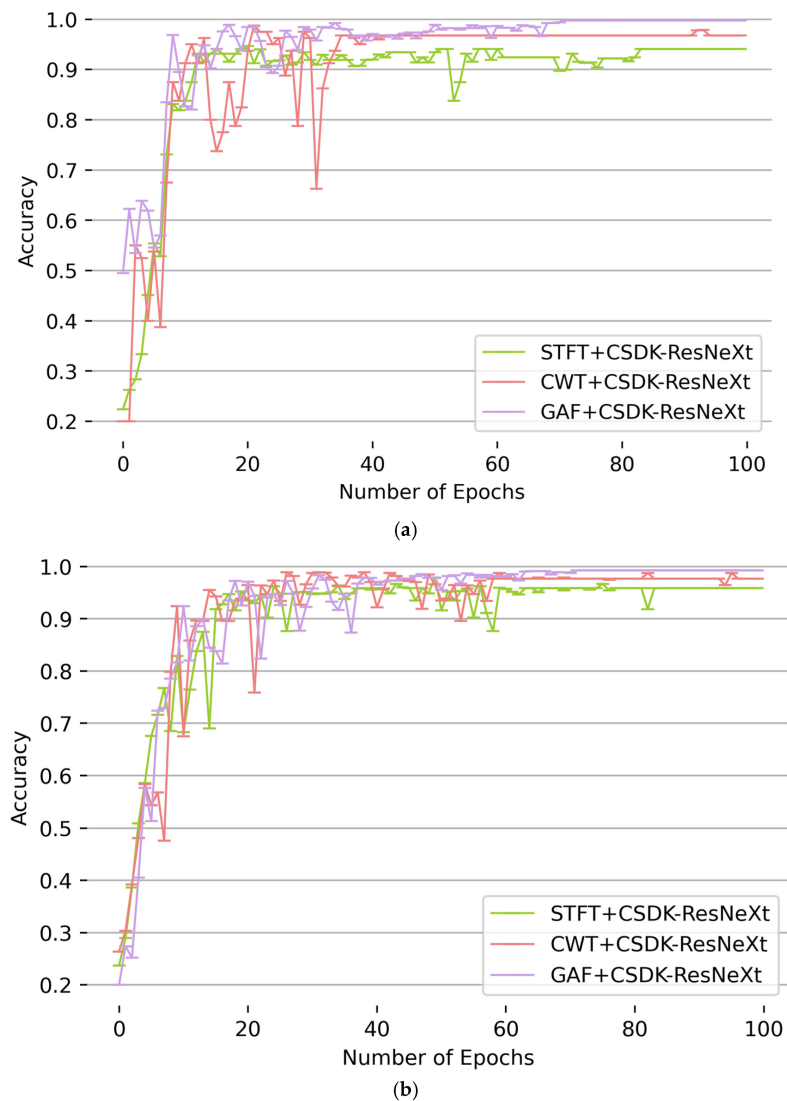


Figure 7. Effects of different methods: (a) effects of different methods at 20 Hz-0 V and (b) effects of different methods at 30 Hz-2 V.

To intuitively show the influence brought by channel shuffle and convolution kernel splitting in CSKD-ResNeXt, ablation experiments were set to show and compare the improvement degree of modified parts, as shown in Table 5. ResNeXt represents the network without channel shuffle and convolutional kernel splitting; 7×7 -ResNeXt represents the network with convolutional kernel splitting but without channel shuffle; CSKD-ResNeXt represents the network with both operations, namely, the network proposed in this paper. Other settings are consistent, such as LeakyReLU, initial learning rate, ReduceLROnPlateau (including patience), Adam, etc.

Table 5. Ablation experiment.

	Accuracy		Loss		Precision		Recall		F1	
	20 Hz	30 Hz	20 Hz	30 Hz	20 Hz	30 Hz	20 Hz	30 Hz	20 Hz	30 Hz
ResNeXt	0.943	0.945	0.162	0.144	0.946	0.944	0.943	0.945	0.945	0.945
7×7 ResNeXt	0.972	0.963	0.024	0.034	0.972	0.964	0.972	0.963	0.972	0.963
CSKD-ResNeXt	0.998	0.993	0.016	0.013	0.998	0.993	0.998	0.993	0.998	0.993

Convolution kernel splitting can improve the comprehensiveness and delicacy of feature mining, and channel shuffle can make up for the defect of a group convolution's independent information channel. Both of these make CSKD-ResNeXt improve the breadth and depth of feature mining, so as to extract fault features more fully.

4.2. t-SNE Visualization

The popular t-SNE [59] is used to make the output of the representative stage in the model be low-dimensional mapped and visualized. Five colors represent five gearbox states. The visualized result of dimensionality reduction is shown in Figure 8. With the deepening of layers, data points in the same state gradually gather, while data points in different states gradually separate.

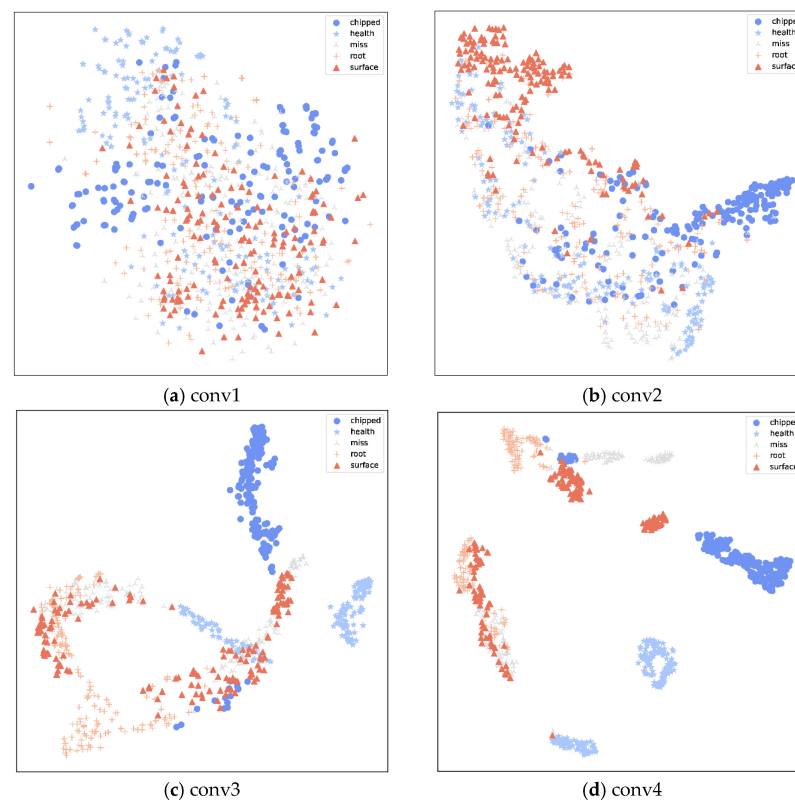


Figure 8. Cont.

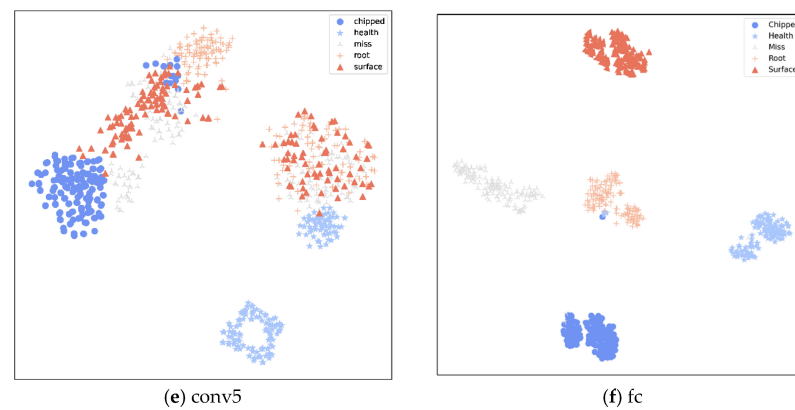


Figure 8. t-SNE dimension reduction results.

4.3. Contrast of Classical Model

To further verify the recognition performance of the proposed GAF+ CSKD-ResNeXt model and other fault diagnosis models, classical convolutional neural networks AlexNet, ResNet50, and DenseNet were selected for comparison experiments. The softmax classifier was used for all comparison models. The comparison results are shown in Figure 9. The accuracy of the GAF+ CSKD-ResNeXt model is higher than that of other models under two working conditions, and the convergence and stability of the GAF+ CSKD-ResNeXt model are better than those of other models.

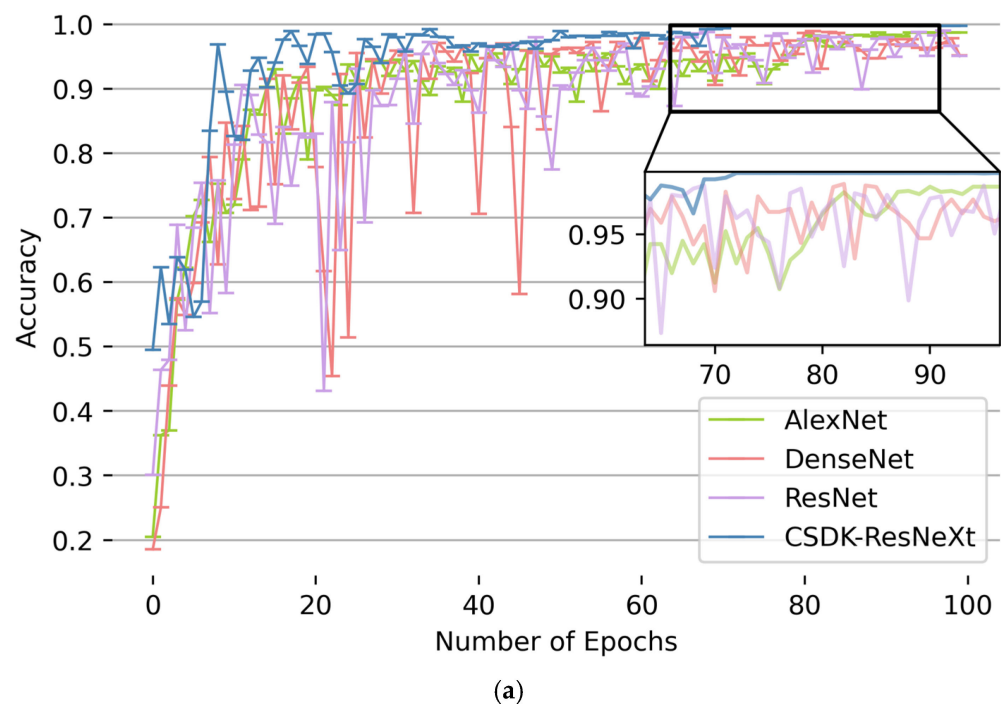


Figure 9. Cont.

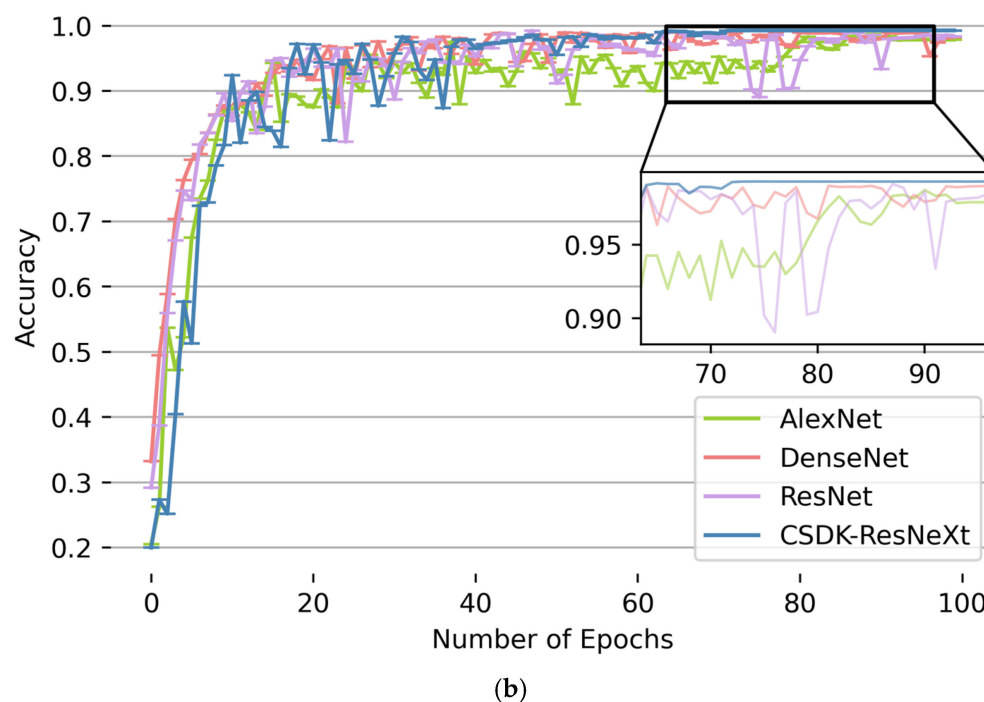


Figure 9. Effects of different methods: (a) effects of different models at 20 Hz—0 V and (b) effects of different models at 30 Hz—2 V.

5. Conclusions

In this paper, a gearbox fault detection method combining GAF and improved ResNeXt is proposed. The performance of the proposed method is verified in two operating conditions of the gearbox, and the accuracy of fault identification can reach 99.75%. The comparison experiment results between GAF and the time-frequency conversion method show that GAF has better ability to express different state features. After comparison between GASF and GADF, GADF is selected to output two-dimensional images according to the accuracy and loss performance. The ablation experiment shows that the modified ResNeXt model can promote the information exchange in the network and improve the feature-mining ability. The experimental results of comparison between the proposed model and other classical network models show that the GAF-CSKD-ResNeXt method has higher recognition accuracy and faster convergence speed and can effectively classify gearbox faults. Our future work will further improve the feature expression ability of vibration signals, reduce the workload of the feature selection process, and pay more attention to the interpretability of the feature selection process to further build a model with stronger generalization ability, higher stability, and better interpretation. Further consideration will be given to the fault diagnosis of the gearbox under the condition of load, voltage, speed, and other conditions changing at the same time.

Author Contributions: Y.L.: writing—original draft preparation; S.D.: writing—review and editing; Y.D. and Z.W.: supervision and revision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the R&D Program of the Beijing Municipal Education Commission. Project name: Research on Dynamic Characteristics and Optimal Design of Complex Transmission System for High-end-printing-equipment to Meet High-quality Printing Needs, grant number: KZ202210015019.

Data Availability Statement: The public datasets used in this experiment can be obtained here: <https://gitee.com/zhengkun110/Mechanical-datasets/tree/master/gearbox/gearset> (accessed on 15 March 2023).

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

References

- Wang, R.; Zhan, X.; Bai, H.; Dong, E.; Cheng, Z.; Jia, X. A Review of Fault Diagnosis Methods for Rotating Machinery Using Infrared Thermography. *Micromachines* **2022**, *13*, 1644. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lu, L.; He, Y.; Ruan, Y.; Yuan, W. An optimized stacked diagnosis structure for fault diagnosis of wind turbine planetary gearbox. *Meas. Sci. Technol.* **2021**, *32*, 75102. [\[CrossRef\]](#)
- Guo, Y.J.; Jiang, S.F.; Yang, Y.D.; Jin, X.H.; Wei, Y.D. Gearbox Fault Diagnosis Based on Improved Variational Mode Extraction. *Sensors* **2022**, *22*, 1779. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sohaib, M.; Munir, S.; Islam MM, M.; Shin, J.; Tariq, F.; Rashid SM, M.; Kim, J. Gearbox fault diagnosis using improved feature representation and multitask learning. *Front. Energy Res.* **2022**, *10*, 998760. [\[CrossRef\]](#)
- Schoen, R.R.; Habetler, T.G.; Kamran, F.; Bartfield, R.G. Motor bearing damage detection using stator current monitoring. *IEEE Trans. Ind. Appl.* **1995**, *31*, 1274–1279. [\[CrossRef\]](#)
- Zhong, R.Y.; Xu, X.; Klotz, E.; Newman, S.T. Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering* **2017**, *3*, 616–630. [\[CrossRef\]](#)
- Lee, J.; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L.; Siegel, D. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mech. Syst. Signal Process.* **2014**, *42*, 314–334. [\[CrossRef\]](#)
- Alvarez, E.J.; Ribaric, A.P. An improved-accuracy method for fatigue load analysis of wind turbine gearbox based on SCADA. *Renew. Energy* **2018**, *115*, 391–399. [\[CrossRef\]](#)
- Li, H.; Feng, G.; Zhen, D.; Gu, F.; Ball, A.D. A normalized frequency-domain energy operator for broken rotor bar fault diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3500110. [\[CrossRef\]](#)
- Li, C.; Zhang, S.; Qin, Y.; Estupinan, E. A systematic review of deep transfer learning for machinery fault diagnosis. *Neurocomputing* **2020**, *407*, 121–135. [\[CrossRef\]](#)
- Contin, A.; D’Orlando, S.; Fenu, G.; Menis, R.; Milo, S.; Parisini, T. Experiments on actuator fault diagnosis: The case of a nonlinearly controlled AC motor. In Proceedings of the European Control Conference (ECC), Porto, Portugal, 4–7 September 2001; pp. 2747–2752. [\[CrossRef\]](#)
- Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
- You, D.; Gao, X.; Katayama, S. Multisensor Fusion System for Monitoring High-Power Disk Laser Welding Using Support Vector Machine. *IEEE Trans. Ind. Inform.* **2014**, *10*, 1285–1295. [\[CrossRef\]](#)
- Tian, J.; Morillo, C.; Azarian, M.H.; Pecht, M. Motor Bearing Fault Detection Using Spectral Kurtosis-Based Feature Extraction Coupled With K-Nearest Neighbor Distance Analysis. *IEEE Trans. Ind. Electron.* **2016**, *63*, 1793–1803. [\[CrossRef\]](#)
- Shevchik, S.A.; Saeidi, F.; Meylan, B.; Wasmer, K. Prediction of Failure in Lubricated Surfaces Using Acoustic Time–Frequency Features and Random Forest Algorithm. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1541–1553. [\[CrossRef\]](#)
- Lei, Y.; Karimi, H.R.; Chen, X. A novel self-supervised deep LSTM network for industrial temperature prediction in aluminum processes application. *Neurocomputing* **2022**, *502*, 177–185. [\[CrossRef\]](#)
- Arel, I.; Rose, D.C.; Karnowski, T.P. Deep Machine Learning—A New Frontier in Artificial Intelligence Research [Research Frontier]. *IEEE Comput. Intell. Mag.* **2010**, *5*, 13–18. [\[CrossRef\]](#)
- Saxena, A.; Parey, A.; Chouksey, M. Time varying mesh stiffness calculation of spur gear pair considering sliding friction and spalling defects. *Eng. Fail. Anal.* **2016**, *70*, 200–211. [\[CrossRef\]](#)
- Sanchez, H.; Escobet, T.; Puig, V.; Odgaard, P.F. Fault Diagnosis of an Advanced Wind Turbine Benchmark Using Interval-Based ARRs and Observers. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3783–3793. [\[CrossRef\]](#)
- Sun, R.; Yang, Z.; Yang, L.; Qiao, B.; Chen, X.; Gryllias, K. Planetary gearbox spectral modeling based on the hybrid method of dynamics and LSTM. *Mech. Syst. Signal Process.* **2020**, *138*, 106611. [\[CrossRef\]](#)
- Shanbr, S.; Elasha, F.; Elforjani, M.; Teixeira, J. Detection of natural crack in wind turbine gearbox. *Renew. Energy* **2018**, *118*, 172–179. [\[CrossRef\]](#)
- Wang, J.; Peng, Y.; Qiao, W.; Hudgins, J.L. Bearing Fault Diagnosis of Direct-Drive Wind Turbines Using Multiscale Filtering Spectrum. *IEEE Trans. Ind. Appl.* **2017**, *53*, 3029–3038. [\[CrossRef\]](#)
- Lv, Y.; Guan, N.; Liu, J.; Cai, T. The fault diagnosis of rolling bearing in gearbox of wind turbines based on second generation wavelet. In Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, Lanzhou, China, 13–16 July 2014; pp. 43–49. [\[CrossRef\]](#)
- Lopez-Ramirez, M.; Romero-Troncoso, R.J.; Morinigo-Sotelo, D.; Duque-Perez, O.; Ledesma-Carrillo, L.M.; Camarena-Martinez, D.; Garcia-Perez, A. Detection and diagnosis of lubrication and faults in bearing on induction motors through STFT. In Proceedings of the International Conference on Electronics, Communications and Computers (CONIELECOMP), Cholula, Mexico, 24–26 February 2016; pp. 13–18. [\[CrossRef\]](#)
- Tang, G.; Wang, Y.; Huang, Y.; Wang, H. Multiple time-frequency curve classification for tacho-less and resampling-less compound bearing fault detection under time-varying speed conditions. *IEEE Sens. J.* **2021**, *21*, 5091–5101. [\[CrossRef\]](#)
- Zeng XJ Yang, M.; Bo, Y.F. Gearbox oil temperature anomaly detection for wind turbine based on sparse Bayesian probability estimation. *Int. J. Electr. Power Energy Syst.* **2020**, *123*, 106233. [\[CrossRef\]](#)
- Wang, Z.Y.; Li, G.S.; Yao, L.G.; Qi, X.L.; Zhang, J. Data-driven fault diagnosis for wind turbines using modified multiscale fluctuation dispersion entropy and cosine pairwise-constrained supervised manifold mapping. *Knowl.-Based Syst.* **2021**, *228*, 107276. [\[CrossRef\]](#)

28. Toma, R.N.; Kim, J.M. Bearing Fault Classification of Induction Motors Using Discrete Wavelet Transform and Ensemble Machine Learning Algorithms. *Appl. Sci.* **2020**, *10*, 5251. [\[CrossRef\]](#)
29. Pang, J.S.; Chen, Y.M.; He, S.Z.; Qiu, H.H.; Wu, C.L.; Mao, L.B. Classification of Friction and Wear State of Wind Turbine Gearboxes Using Decision Tree and Random Forest Algorithms. *J. Tribol. -Trans. ASME* **2021**, *143*, 91702. [\[CrossRef\]](#)
30. Wang, H.; Liu, Z.; Peng, D.; Qin, Y. Understanding and Learning Discriminant Features based on Multiattention 1DCNN for Wheelset Bearing Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 5735–5745. [\[CrossRef\]](#)
31. Yu, J.; Zhou, X. One-Dimensional Residual Convolutional Autoencoder Based Feature Learning for Gearbox Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6347–6358. [\[CrossRef\]](#)
32. Xingkang, Z.; Jianbo, Y. Gearbox Fault Diagnosis Based on One-dimension Residual Convolutional Auto-encoder. *J. Mech. Eng.* **2020**, *56*, 96–108. [\[CrossRef\]](#)
33. Yang, S.; Liu, L.; Zhou, J.; Zhao, Y.; Hua, G.; Sun, H.; Zheng, N. Robust and Efficient Star Identification Algorithm based on 1-D Convolutional Neural Network. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, *58*, 4156–4167. [\[CrossRef\]](#)
34. Xu, H.; Cai, C.Z.; Chi, Y.L.; Zhang, N. Fault diagnosis of gearbox based on adaptive wavelet de-noising and convolution neural network. *Adv. Mech. Eng.* **2023**, *15*, 16878132231157186. [\[CrossRef\]](#)
35. Wang, L.-H.; Zhao, X.-P.; Wu, J.-X.; Xie, Y.-Y.; Zhang, Y.-H. Motor Fault Diagnosis Based on Short-time Fourier Transform and Convolutional Neural Network. *Chin. J. Mech. Eng.* **2017**, *30*, 1357–1368. [\[CrossRef\]](#)
36. Zhang, Y.; Xing, K.; Bai, R.; Sun, D.; Meng, Z. An enhanced convolutional neural network for bearing fault diagnosis based on time-frequency image. *Measurement* **2020**, *157*, 107667. [\[CrossRef\]](#)
37. Huang, D.; Zhang, W.A.; Guo, F.; Liu, W.; Shi, X. Wavelet Packet Decomposition-Based Multiscale CNN for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Trans. Cybern.* **2023**, *53*, 443–453. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Liu, H.; Mi, X.; Li, Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. *Energy Convers. Manag.* **2018**, *159*, 54–64. [\[CrossRef\]](#)
39. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [\[CrossRef\]](#)
40. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [\[CrossRef\]](#)
41. Babu, T.N.; Ali PS, N.; Prabha, D.R.; Mohammed, V.N.; Wahab, R.S.; Vijayalakshmi, S. Fault Diagnosis in Bevel Gearbox Using Coiflet Wavelet and Fault Classification Based on ANN Including DNN. *Arab. J. Sci. Eng.* **2022**, *47*, 15823–15849. [\[CrossRef\]](#)
42. Zhang, K.; Chen, J.; Zhang, T.; Zhou, Z. A Compact Convolutional Neural Network Augmented with Multiscale Feature Extraction of Acquired Monitoring Data for Mechanical Intelligent Fault Diagnosis. *J. Manuf. Syst.* **2020**, *55*, 273–284. [\[CrossRef\]](#)
43. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
44. Gu, F.C. Application of the convolutional neural network in partial discharge spectrum recognition of power apparatus. *IET Sci. Meas. Technol.* **2023**, 1–10. [\[CrossRef\]](#)
45. O'Shea, T.J.; Roy, T.; Clancy, T.C. Over-the-Air Deep Learning Based Radio Signal Classification. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 168–179. [\[CrossRef\]](#)
46. Liu, J.; Wang, Y.C.; Siong, T.C.; Li, X.J.; Zhao, L.P.; Wei, F.R. On the combination of adaptive neuro-fuzzy inference system and deep residual network for improving detection rates on intrusion detection. *PLoS ONE* **2022**, *17*, e0278819. [\[CrossRef\]](#)
47. Xie, S.N.; Girshick, R.; Dollar, P.; Tu, Z.W.; He, K.M. Aggregated residual transformations for deep neural networks. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [\[CrossRef\]](#)
48. Gao, C.; Wu, J.; Yu, H.; Yin, J.; Guo, S. FIRN: A Novel Fish Individual Recognition Method with Accurate Detection and Attention Mechanism. *Electronics* **2022**, *11*, 3459. [\[CrossRef\]](#)
49. Zhang, Y.T.; Zhuo, L.; Ma, C.J.; Zhang, Y. Abnormal Object Detection in X-ray Images with Self-normalizing Channel Attention and Efficient Data Augmentation; In Proceedings of the International Workshop on Advanced Imaging Technology (IWAIT), Hong Kong, China, 4–6 January 2022. [\[CrossRef\]](#)
50. Wang, G.W.; Wang, J.X.; Yu, H.Y.; Sui, Y.Y. Research on Identification of Corn Disease Occurrence Degree Based on Improved ResNeXt Network. *Int. J. Pattern Recognit. Artif. Intell.* **2022**, *36*, 2250005. [\[CrossRef\]](#)
51. Fang, J.; Xu, C.; Wang, C.; Li, H. Dynamic Gesture Recognition Based On Multimodal Fusion Model. In Proceedings of the 2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS), London, UK, 20–22 December 2021; pp. 172–177. [\[CrossRef\]](#)
52. Zhou, Y.J.; Long, X.Y.; Sun, M.W.; Chen, Z.Q. Bearing fault diagnosis based on Gramian angular field and DenseNet. *Math. Biosci. Eng.* **2022**, *19*, 14086–14101. [\[CrossRef\]](#)
53. Xi, C.P.; Liu, R.Q. Detection of Small Floating Target on Sea Surface Based on Gramian Angular Field and Improved EfficientNet. *Remote Sens.* **2022**, *14*, 4364. [\[CrossRef\]](#)
54. Xue, Y.M.; Huang, W.M.; Yang, C. Hyperspectral image classification based on gramian angular fields encoding. In Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE), Halifax, NS, Canada, 18–20 September 2022. [\[CrossRef\]](#)

55. Dong, S.J.; Li, Y.; Zhu, P.; Pei, X.W.; Pan, X.J.; Xu, X.Y.; Liu, L.H.; Xing, B.; Hu, X.L. Rolling bearing performance degradation assessment based on singular value decomposition-sliding window linear regression and improved deep learning network in noisy environment. *Meas. Sci. Technol.* **2022**, *33*, 045015. [[CrossRef](#)]
56. Zhang, X.; Zhou, X.Y.; Lin, M.X.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices; In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018. [[CrossRef](#)]
57. Shao, S.; Mcaleer, S.; Yan, R.; Baldi, P. Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2446–2455. [[CrossRef](#)]
58. Diederik, K.; Ba, J.L. ADAM: A method for stochastic optimization. *AIP Conf. Proc.* **2014**, *1631*, 58–62. [[CrossRef](#)]
59. Lu, W.P.; Yan, X.F. Variable-weighted FDA combined with t-SNE and multiple extreme learning machines for visual industrial process monitoring. *ISA Trans.* **2022**, *122*, 163–171. [[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.