sensors

MDPI

Article

Optimization of Gearbox Fault Detection Method Based on Deep Residual Neural Network Algorithm

Zhaohua Wang, Yingxue Tao, Yanping Du *, Shuihai Dou and Huijuan Bai

Department of Mechanical and Electrical Engineering, Beijing Institute of Graphic Communication, No. 1, Xinghua Street, Beijing 102600, China; wzh@bigc.edu.cn (Z.W.); taoyingxue111@126.com (Y.T.); doushuihai@bigc.edu.cn (S.D.); bhj07@bigc.edu.cn (H.B.)

* Correspondence: duyanping@bigc.edu.cn; Tel.: +86-158-0165-7157

Abstract: Because of its long running time, complex working environment, and for other reasons, a gear is prone to failure, and early failure is difficult to detect by direct observation; therefore, fault diagnosis of gears is very necessary. Neural network algorithms have been widely used to realize gear fault diagnosis, but the structure of the neural network model is complicated, the training time is long and the model is not easy to converge. To solve the above problems and combine the advantages of the ResNeXt50 model in the extraction of image features, this paper proposes a gearbox fault detection method that integrates the convolutional block attention module (CBAM). Firstly, the CBAM is embedded in the ResNeXt50 network to enhance the extraction of image channels and spatial features. Secondly, the different time–frequency analysis method was compared and analyzed, and the method with the better effect was selected to convert the one-dimensional vibration signal in the open data set of the gearbox into a two-dimensional image, eliminating the influence of the redundant background noise, and took it as the input of the model for training. Finally, the accuracy and the average training time of the model were obtained by entering the test set into the model, and the results were compared with four other classical convolutional neural network models. The results show that the proposed method performs well both in fault identification accuracy and average training time under two working conditions, and it also provides some references for existing gear failure diagnosis research.

Keywords: gearbox; fault detection; ResNeXt50 model; CBAM; time–frequency analysis method

check for updates

Citation: Wang, Z.; Tao, Y.; Du, Y.; Dou, S.; Bai, H. Optimization of Gearbox Fault Detection Method Based on Deep Residual Neural Network Algorithm. *Sensors* **2023**, *23*, 7573. https://doi.org/10.3390/s23177573

Academic Editors: Fanglai Zhu and Iñigo Barandiaran

Received: 9 July 2023
Revised: 27 August 2023
Accepted: 30 August 2023
Published: 31 August 2023

1. Introduction

The purpose of gearbox fault detection is to diagnose and prevent various abnormal states and fault states in a timely and correct manner, eliminate the impact of faults on equipment operation, improve the reliability, safety and effectiveness of equipment operation, and reduce the fault loss to the lowest level [1]. At present, gearbox fault detection can be roughly divided into three stages. The first stage is to rely on the experience and knowledge of professional and technical personnel to judge their working status; The second stage is fault detection technology with fault mechanism and signal processing as the core. However, the fault diagnosis methods in the first and second stages mostly rely on the expertise and experience of technicians. With the rapid development of pattern recognition and other disciplines, the third stage of intelligent fault detection with an artificial intelligence algorithm as the core has gradually developed, and the research into intelligent fault detection based on artificial intelligence algorithms has become increasingly in-depth.

At present, the third stage of gearbox intelligent fault detection methods with artificial intelligence as the core can be divided into two categories: the traditional machine learning-based method and the method based on deep learning [2,3]. The intelligent fault detection method based on traditional machine learning includes three processes: data acquisition, artificial feature extraction, and pattern recognition [4]. However, traditional machine learning methods generally do not have a deep architecture and are relatively closed themselves; therefore, the upper limit of the algorithm performance is low. Compared with traditional machine learning methods, the fault detection model based on deep learning is a typical end-to-end model with superior nonlinear mapping fitting ability, which can automatically extract data features in the optimization process. It makes up for the limitations of manual feature extraction in traditional machine learning methods and reduces the dependence on mechanism research and subjective experience [5–7]. To some extent, it improves the accuracy and intelligence of the diagnosis a–nd is more conducive to realization of the diagnosis of mechanical equipment faults on an automatic and large scale [8].

Today, in the field of troubleshooting, the most commonly used deep learning methods are deep belief networks (DBNs), stack auto-encoders (SAEs), recurrent neural networks (RNNs), generative adversarial networks (GANs) and convolutional neural networks (CNNs). Deep learning models are also being applied in the field of fault diagnosis. Xu et al. [9] used DBNs to extract the signal from the bearings and performed a rolling bearing fault detection. Shang et al. [10] proposed a rolling bearing fault detection model based on DBNs, and experiments proved that the model could accurately identify all kinds of faults and had good fault detection ability. Liu et al. [11] have proposed an effective deep learning approach called stacked auto-encoders (SAEs). This method extracted significant features directly from the frequency domain signal and eliminated the use of manual features to solve the problem of gearbox fault detection. Dai et al. [12] combined a sparse auto-encoder and a denoising auto-encoder, proposed the stacked sparse denoising auto-encoder diagnosis model, and applied the model to rolling bearing fault detection. Liu et al. [13] combined recurrent neural networks and autoencoders to realize the intelligent fault detection of rolling bearings. An et al. [14] used recurrent neural networks to realize the intelligent fault detection of bearings under time-varying working conditions. Wang et al. [15] used GAN in the fault diagnosis of rotating machinery in nuclear power plants to improve the generalization ability of the model when the data are inconsistent.

Compared with other deep learning models, the unique convolutional layer and pooled layer algorithm design of the CNN model can effectively reduce the complexity of the model, reduce the number of training parameters, and make the model easy to train and optimize. Common convolutional neural network models include the visual geometry group (VGG), GoogleNet, ResNet and ResNeXt. In order to improve the recognition accuracy of the model, the traditional convolutional neural network needs to deepen or widen the network, but with the deepening of the network layers, there will be gradient explosion or gradient disappearance, and network degradation will occur. As a variant of ResNet, the ResNeXt model adopts the idea of an inception network [16] and uses group convolution instead of traditional convolution to widen the network and reduce the number of training parameters of the model [17]. Because of its strong feature recognition ability, the ResNeXt model is widely used in the field of image recognition, such as: Yao Xiao et al. [18] proposed an automatic insect recognition system based on SE-ResNeXt, which realized the visual display of insect recognition results and the digital storage of insect data. Wang Guowei et al. [17] proposed an improved ResNeXt model to identify the occurrence degree of maize diseases. Pant Gaurav et al. [19] used an improved ResNeXt model to solve the problem of identification and classification of algae species. The above studies have proved the performance of the ResNeXt model in the field of image recognition, so this paper adopts the ResNeXt model as the gearbox fault identification and classification model.

However, when there is too much input, the ResNeXt model also becomes more complex and the training time is extended. In order to alleviate this problem, attention mechanisms are applied to neural network algorithms. Zheng et al. [20] proposed a multi-scale residual aggregation network by adding attention mechanisms to the network. Jiang et al. [21] applied the attention mechanism to predict the life of rolling bearings. Wu et al. [22] and Wu et al. [23] applied the attention mechanism to cross-domain fault detection, enabling neural network models to learn and update fault-related features in the training process, and improving the influence of the source domain with a stronger correlation. All the above studies show that the introduction of attention mechanisms into neural network models can effectively focus on significant information while ignoring redundant features, improving the feature extraction ability of neural network models for fault information, and thus reducing the amount of information processed by models and reducing computing resources.

In view of this, in this paper the ResNeXt model with strong feature extraction ability is selected as the basic model. The CBAM is embedded after the first convolutional layer and before the last convolutional layer of the model to highlight key information of different fault types and suppress useless information, and a CBAM-ResNeXt50 model is built. By comparing CWT and STFT, the effective method is selected to convert the vibration signal into time-frequency images as input to the CBM-ResNeXt50 model to reduce the influence of redundant background noise and improve the feature extraction ability of the model on the time-frequency images. The CBAM-ResNeXt50 model is compared with the original ResNeXt50 model and four other classical convolutional neural network models. The effectiveness of adding the CBAM and the robustness of the CBAM-ResNeXt50 model are verified by comparing the average training time and the accuracy of the test set of each model.

The organization structure was as follows:

i. In Section 2, the theory of two time-frequency analysis methods, the basic principle of the CBAM, and the basic structure of the ResNeXt50 model are introduced.
ii. In Section 3, the CBAM-ResNeXt50 model integrated into the CBAM module is established and the flow of the algorithm is detailed.
iii. In Section 4, the proposed method is experimentally studied and verified using the open gearbox data set from Southeast University. The confusion matrix is obtained by comparing with four other classical convolutional neural networks. T-distributed stochastic neighbor embedding (t-SNE) is used to simplify the classification results into a two-dimensional plane and visualize them in the form of scatter plots. In addition, the anti-confusion capability of the proposed method is verified by the above methods.
iv. The conclusions of this study and future research directions are presented in Section 5.

2. Basic Theory

The main work of this paper is to add the CBAM block to the ResNeXt50 model to build the CBAM-RESNEXT50 model and to make the model pay more attention to the fault information. Vibration signals are converted into two-dimensional images by CWT and STFT, and the images are divided into a training set and a test set to train and test the constructed model. Compared to the original ResNeXt50 model and the four other classical convolutional neural network models, the performance of the proposed method is judged.

The second section is divided into three parts according to the main work, and respectively introduces the basic theories involved in the research, including the basic knowledge of CWT and STFT, the basic theory of attention mechanism, and the basic model structure of ResNeXt.

2.1. Time–Frequency Analysis Method

In the acquisition of vibration signals, there is a lot of noise interference, and it is difficult to train the network to judge the fault type directly according to the obtained signals. It is necessary to use the method of time–frequency analysis and processing to make the characteristics of the fault prominent. Time–frequency analysis methods include short-time Fourier transform (STFT) and continuous wavelet transform (CWT) [24].

In 1946, Dennis Gaor proposed the STFT algorithm [25]. The basic idea is: In the framework of the traditional Fourier transform, the nonstationary signal is regarded as the superposition of a series of short-time stationary signals. The short-time is realized by adding windows in the time domain, after which it is transformed into a local spectrum in a very small period time near the time t. The whole time domain is covered by translation parameters, that is, the local spectrum of any position is obtained by moving in the whole time domain with the change in time t, and then the time–frequency–energy analysis is carried out [26].

For a given non-stationary signal, the STFT of the signal $s(t) \in L^2(R)$ is defined as:

$$\mathrm{STFT_S}(t,\omega) = \int_{-\infty}^{+\infty} s(\tau)h(\tau - t)e^{-i\omega t}d\tau \tag{1}$$

where: $h(t)$ is the window function.

It can be seen from Equation (1) that, in the STFT, the translation parameter t is first used to segment signal $s(t)$ through window function $h(t)$, namely:

$$s_t = s(\tau)h(\tau - t) \tag{2}$$

The signal at time τ is obtained near time t, and the Fourier transform is then applied to that signal $s_t(\tau)$.

The time resolution and frequency resolution of STFT are contradictory. The higher the time resolution is, the lower is the frequency resolution, and vice versa, which causes STFT to have certain limitations when analyzing high frequency signals. In addition, the analysis results of STFT are affected by the window function. Different window functions will produce different analysis results, so it is necessary to choose the appropriate window function according to the specific situation.

CWT [27] is derived from the decomposition and reconstruction of the input signal using a finite length or fast attenuation "mother wavelet", and the mother wavelet fits the input signal through scale contraction and wavelet translation. The original signal in the time–domain $y(x)$ is decomposed into a series of wavelet transform coefficients $\lambda(k,x)$ after inner product operation by CWT through a specific mother wavelet $\varphi(x)$, so as to construct a localized time–frequency signal with good time–domain and frequency domain.

For any square integrable function $y(x) \in L^2(R)$, its CWT is defined as [28]:

$$\lambda(k,x) = \int y(x)\varphi^*_{k,\tau}(x)dx \tag{3}$$

where, $\varphi^*_{k,\tau}(x)$ is the conjugate operation of the wavelet basis function $\varphi_{k,\tau}(x)$.

A series of wavelet basis functions can be obtained by stretching and time-delaying the transformation of the mother wavelet $\varphi(x)$. If we allow $\varphi(x) \in L^2(R)$, its process can be expressed as:

$$\varphi_{k,\tau} = \frac{1}{\sqrt{k}}\varphi\left(\frac{x-\tau}{k}\right),\ k,\tau \in R, k > 0 \tag{4}$$

where k is the scale parameter, when $k > 1$, the stretching of $\varphi(x)$ is realized, and when $0 < k < 1$, the compression of $\varphi(x)$ is realized. τ is the translation parameter, which specifies the position of the translation of the wavelet function along the time axis. k cannot be negative, so the normalization unified treatment becomes positive, and $|k|^{-1/2}$ realizes the normalization processing of the signal capability.

Compared with the STFT, CWT has a different size time–frequency window at different times and frequencies, which can achieve higher frequency resolution in the low frequency region. However, the time–frequency window of CWT is not completely adaptive; it also needs to choose a suitable mother wavelet.

2.2. Attention Mechanism

An attention mechanism is a way to realize adaptive attention in neural networks. Generally speaking, it makes the network pay more attention to the effective unit and suppresses the invalid unit in the feature extraction process [29]. CBAM [30] is a lightweight attention module, which is composed of a channel attention module and a spatial attention module. It can perform the attention mechanism along the channel and space, infer the weight coefficient, and then multiply the feature map [31]. Its structure is shown in Figure 1:

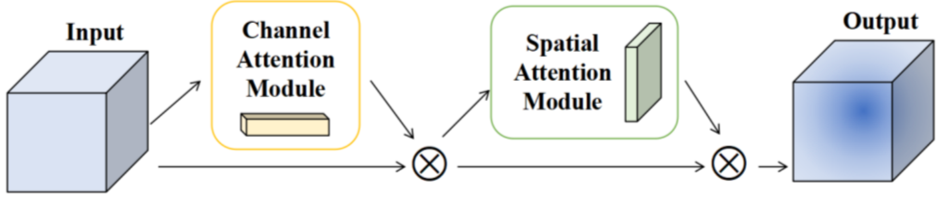


Figure 1. The structure of CBAM.

Channel attention module. Channel attention is the feature on which channel it is meaningful to focus, as shown in Figure 2.

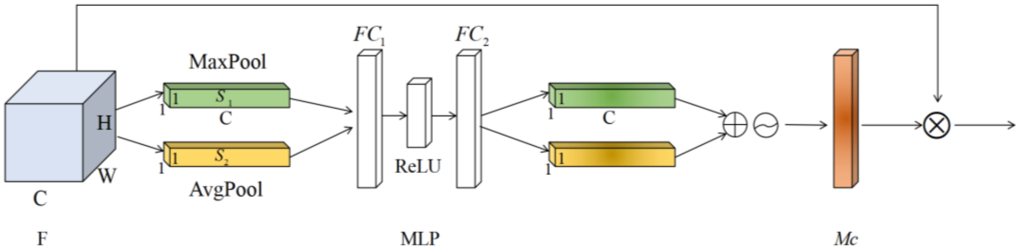


Figure 2. Channel Attention Module.

Suppose that the input feature map (C, W, H) is denoted as F, where W is the width of the input feature map, H is the height of the input feature map, and C is the number of channels in the input feature map. The global spatial information of F is compressed through the maximum pooling layer and the average pooling layer to generate two feature maps s_1 and s_2, both of which are $1 \times 1 \times C$. s_1 and s_2 obtained two one-dimensional feature maps by sharing a multi-layer perception (MLP) consisting of the fully connected layer FC_1, FC_2 and ReLU functions. After summing the two one-dimensional feature graphs according to channels, the Sigmoid function is used to normalize them, and the weighted statistic value Mc of each channel size $1 \times 1 \times C$ is obtained. The process can be expressed by Formula (5):

$$\begin{aligned} Mc(F) &= \sigma\{MLP[Avgpool(F)] + MLP[Maxpool(F)]\} \\ &= \left\{ MLP\left[\frac{1}{H \times W} \sum_{x_0=1}^{H} \sum_{y_0=1}^{W} f_{k_0}(x_0, y_0) \right] \right\} + MLP\left[max f_{k_0}(x_0,y_0)\right] \end{aligned} \tag{5}$$

where, σ represents the Sigmoid function, and $f_{k_0}(x_0, y_0)$ represents the pixel value of the point whose coordinate is x_0, y_0 in the k_0 channel of the input feature graph F.

Spatial attention mechanism. After the output of channel attention, the spatial attention module is introduced to focus on which part of the space features are meaningful, as shown in Figure 3.

The spatial attention mechanism takes the feature map F'' output by the channel attention module as the input of this module and carries out global maximum pooling and global average pooling, respectively, on the channel dimension to obtain two channel descriptions P_1 and P_2. The two descriptions are concatenated together according to the channels to obtain the feature description P_3. The convolutional layer is used to encode and fuse the information of different positions in P_3, and the spatial weighted information Ms is obtained. Finally, we multiply the weight coefficient and feature F' to obtain the new feature after scaling. The process can be expressed by Formula (6):

$$Ms(F) = \sigma\left\{ f^{7\times7}\left[Avgpool(F'); Maxpool(F')\right] \right\} \tag{6}$$

where: $f^{7\times7}$ represents a convolution layer with a convolution kernel of size 7 × 7.

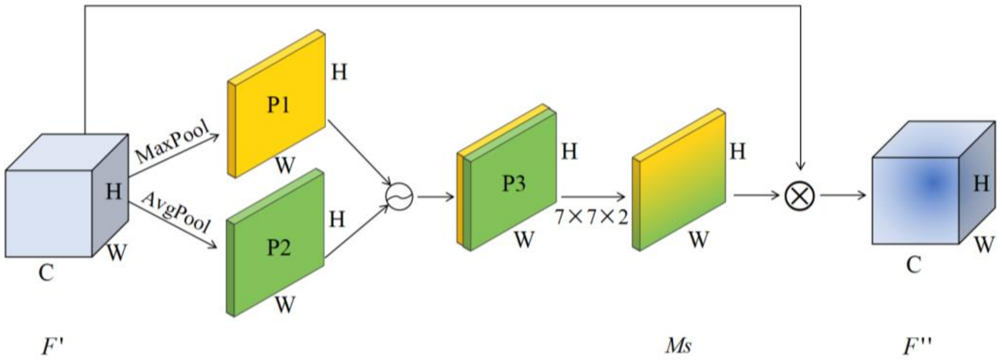


Figure 3. Spatial Attention Module.

CBAM attention mechanism [32]. Multiply the input feature map F with the corresponding elements of the two matrices of the weight value Mc of each channel, carry out the feature recalibration of F, and obtain the feature mapping F' which can effectively reflect the feature key channel information. On the basis of channel weighting, the series spatial attention mechanism is used to carry out adaptive weighting on the spatial feature information. F' is taken as the input to the spatial attention module and multiplied by the corresponding elements of the two matrices of the spatial weight coefficient Ms to obtain the significant feature graph F' containing the channel location information and spatial location information. The network can pay more attention to the strong input features of the gearbox two-dimensional image, and improve its spatial feature selection ability. Its process can be expressed by the formula:

$$\begin{aligned} F' &= Mc \otimes F \\ F'' &= Ms \otimes F' \end{aligned} \tag{7}$$

2.3. ResNeXt Model

Residual Block. By calculating residuals through identity mapping, part of the upper-layer information can be directly transmitted to the last layer, which is conducive to the back propagation of the gradient.

If we suppose that the input sample of part of the neural network is x, $F(x)$ represents the residual error, and the two together form the expected output $H(x)$. As shown in Formula (8), $H(x) = x$ forms the identity mapping when $F(x) = 0$.

$$H(x) = F(x) + x \to F(x) = H(x) - x \tag{8}$$

The Bottleneck Structure. The bottleneck has two structures. The two structures correspond to two situations: Different input and output channel counts (Bottleneck1), and identical input and output channel counts (Bottleneck2).

Bottleneck1 has four mutable parameters C, W, C_1, and S. Compared with Bottleneck2, Bottleneck1 has one more right-side convolution layer, which is designated as a function $H(x)$. Bottleneck1 corresponds to the different input x and output $F(x)$ channel numbers, and it is the added convolutional layer that changes x into $H(x)$, which matches the input and output dimensional differences ($H(x)$ and $F(x)$ have the same number of channels).

Bottleneck2 has two mutable parameters C and W. Assume that the shape of the input x is (C, W, W), and the three convolution blocks (along with the related BN and ReLU) to the left of Bottleneck2 are the function $F(x)$. After adding the two together ($F(x) + x$) and passing through a ReLU function, the output of Bottleneck2 is obtained. The shape of the output remains (C, W, W), namely, Bottleneck2 corresponds to situations where the number of input and output channels is identical.

ResNeXt Block. The ResNeXt network architecture uses the stack of VGG networks and the idea of partition and integration inception, using a block with the same structure for the stack [33]. The essence of ResNeXt is grouping convolution, controlling the number of groups through cardinality [34]. The ResNeXt block divides the input channels into groups, then each group performs a convolution calculation with its own convolution kernel, and finally merges the output of all groups by a 1 × 1 convolution. Because the ResNeXt block has the same topology, it can be used directly to replace the remaining blocks in the regular residual network and stack them to construct the desired structure.

3. Gearbox Fault Detection Method Based on the CBAM-ResNeXt50 Model

The data set in this paper was a labeled data set with four fault categories and one normal category, each of which was distinguished according to the different characteristics of the fault. The flow of fault diagnosis method based on the CBM-ResNeXt50 model is shown in Figure 4. Firstly, one-dimensional vibration signals of two working conditions in the data set were converted into different two-dimensional images as sample data by CWT and STFT, and the sample data were divided into the training sets and test sets according to 4:1 by labeling. The training set was used as input to the CBAM-ResNeXt50 model and features were extracted. The extracted features were then input into the CBAM to make the model pay more attention to the faulty areas in the time–frequency graph. In addition, the CBAM information output was input into the global average pooling layer and the full connection layer. Furthermore, the fault categories of time–frequency images were output and the trained CBAM-ResNeXt50 model was saved. In the classification module shown in Figure 4, the output result is the probability of five failure types, including "Chipped" 0.03, "Health" 0.8, "Miss" 0.06, "Root" 0.04, and "Surface" 0.07. Because the probability of Health is the highest, this gear is judged to be "Health". Finally, the test sets under the two conditions were input into the trained network model, respectively, for testing to obtain the accuracy of the test set. The classification scatter diagram obtained by the t-SNE technique and the confusion matrix diagram were used to judge the merits and demerits of this model.

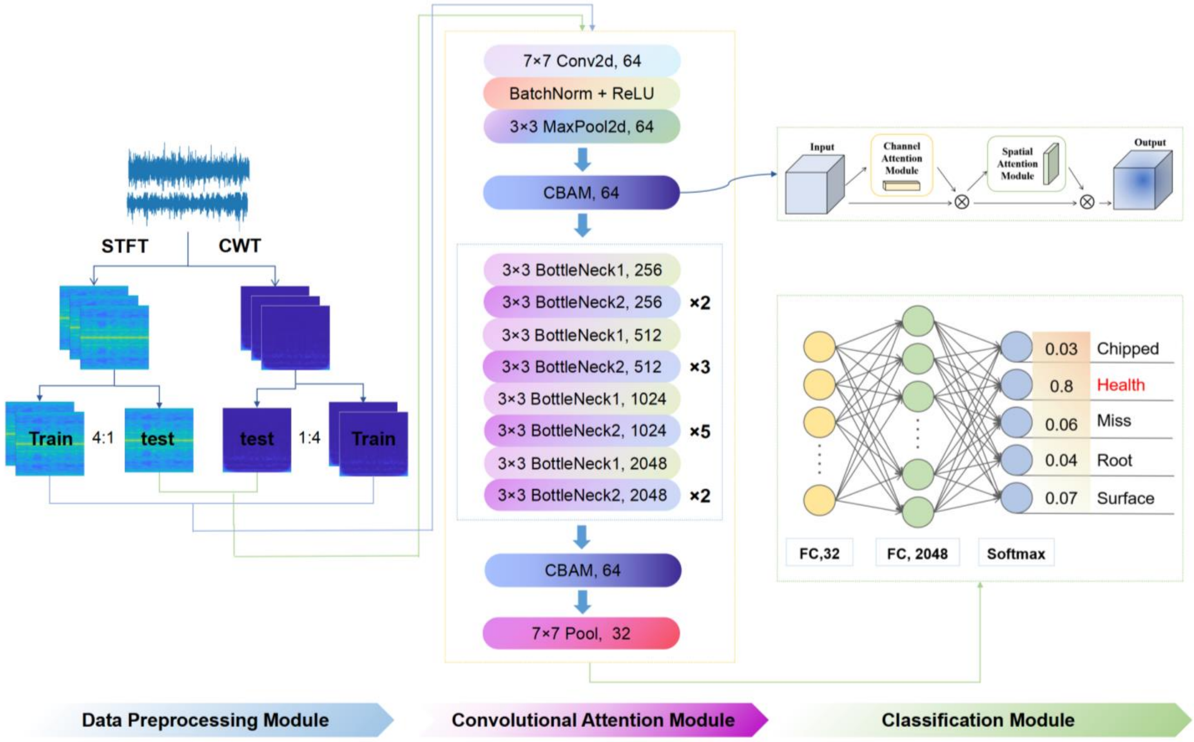


Figure 4. Algorithm flow chart.

4. Experimental Research

4.1. Experimental Data and Setting of Parameters

The experimental data are obtained from the drivetrain dynamic simulator (DDS) of Southeast University [35]. The platform is shown in Figure 5 and mainly consists of six sections: motor, motor controller, planetary gearbox, parallel gearbox, brake, and brake controller. The whole experiment was carried out under two working conditions according to different settings of speed and load. The speed was set at 20 Hz and 30 Hz, and the load was set at 0 V and 2 V while the signals of four kinds of gear fault state and the signal of one kind of gear health state were collected. 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 highest frequency of the collected signal was 2000, and the sampling frequency was 5120 Hz, namely, 5120 sampling points within 1 s. Four types of gear failure data include chipped, miss, root, and surface. The detailed data set type division is shown in Table 1.

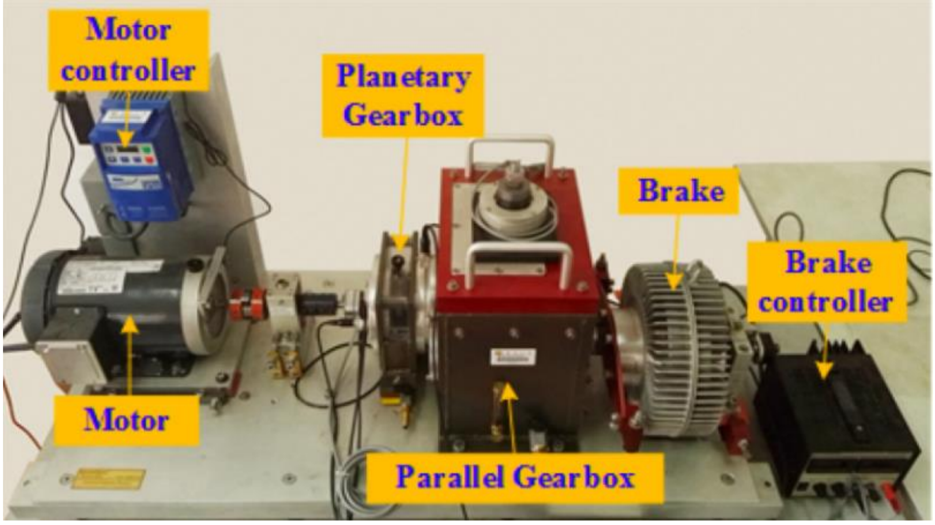


Figure 5. Experimental setup for gearbox.

Table 1. Data set division of gearbox of DDS experimental platform.

	Type	Health	Chipped	Miss	Root	Surface	Total	Working Conditions
Dataset 1	Training Validation	640 160	640 160	640 160	640 160	640 160	4000	20 Hz–0 V
Dataset 2	Training Validation	640 160	640 160	640 160	640 160	640 160	4000	30 Hz–2 V

ResNeXt model is widely used in the field of image recognition due to its strong feature recognition ability. Therefore, this paper truncated all 3000 data points in the data set, and sample graphs with a size of 224 × 224 RGB three-channel through STFT and CWT were generated, respectively, to enhance the feature extraction capability of the ResNeXt model for the image. Finally, the sum of 4000 two-dimensional images of different fault categories under two working conditions was obtained, part of which is shown in Figures 6 and 7. It can be seen from Figures 6 and 7, after STFT and CWT transformation, that the signal used was stable, and that images of various fault types have different characteristics.

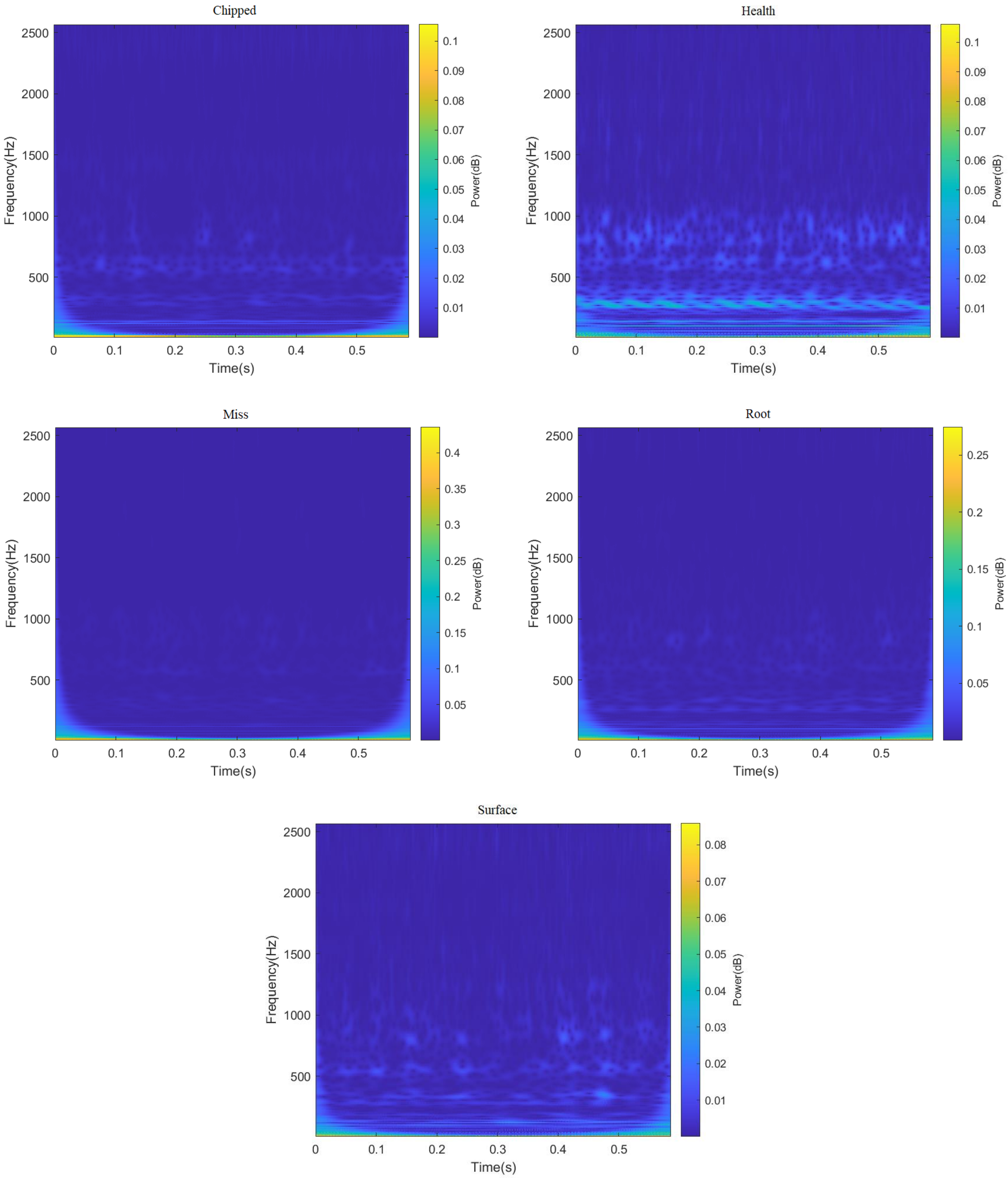


Figure 6. Time–frequency decomposition diagram of gear vibration signals with different fault types after passing through CWT.

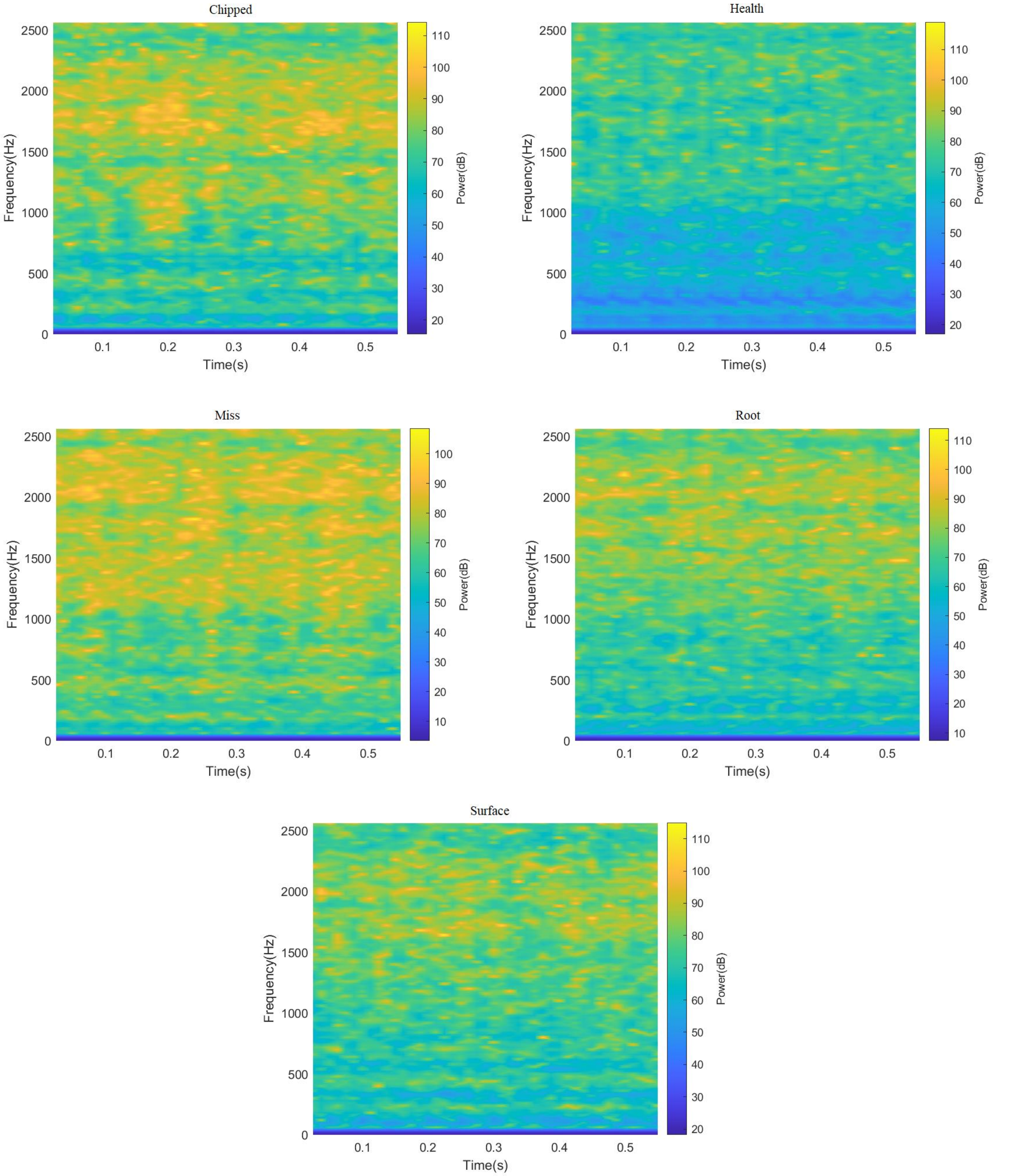


Figure 7. Spectrogram of gear vibration signals with different fault types after passing through STFT.

In STFT transformation, the select window type was the Hanning window with size 256 and moving step size 64. Because a complex morlet wavele adopted a Gaussian function with minimum time–frequency window area and had good localization performance in the time–frequency domain and good symmetry, a complex morlet wavele was selected as the parent wavelet in this paper. The mathematical expression of the complex morlet wavele is:

$$\varphi(t) = \frac{1}{\sqrt{\pi f_b}} e^{2i\pi f_c t} e^{(\frac{-t^2}{f_b})} \tag{9}$$

where f_b is the bandwidth of the wavelet, f_c is the center frequency of the wavelet *and* $2\pi f_c \geq 5$.

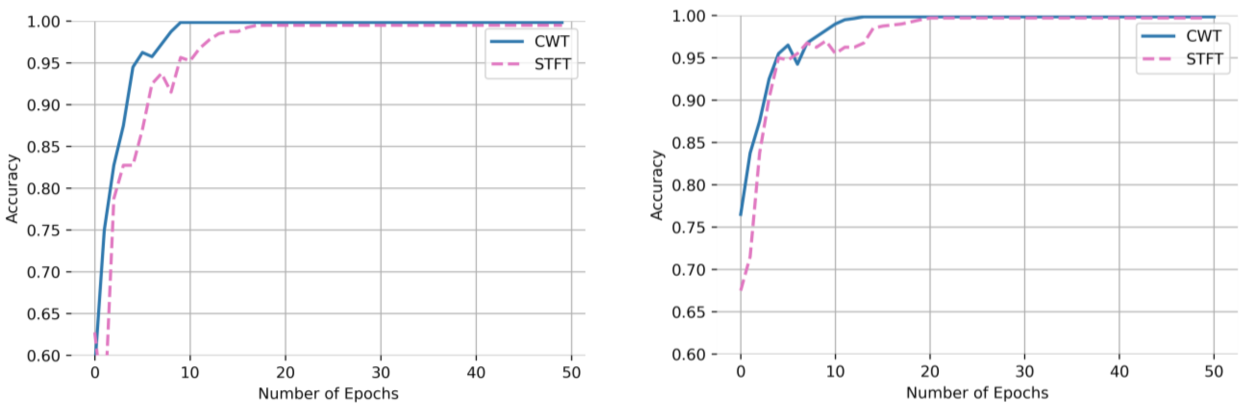
All experiments in this paper were performed on the same platform. The experimental operating environment was ubuntu20.04 OS, the GPU was RTX 3080(10 GB), and the CPU was 12-core Intel(R) Xeon(R) Platinum 8255C. During the experiment, the parameters were adjusted and used as shown in Table 2. The network training epoch was set to 50. In the training process, the initial value of the learning rate was set to 0.001, and the automatic adjustment strategy was adopted to realize the automatic adjustment of the learning rate. The dropout value was set to 0.2. All models were trained using adaptive moment estimation (Adam).

Table 2. Parameter selection.

Hyper Parameter	Values
Epoch	50
Batch_size	32
Learning rate	0.001
Learning rate decay	0.1
Dropout rate	0.2
Adam	-

4.2. Comparison of Time–Frequency Analysis Method

In this paper, two-dimensional images of CWT and STFT would be used as the input of the CBAM-ResNeXt50 model, respectively. Figure 8 is a comparison of the accuracy and convergence rate of the model test set obtained. As can be seen from Figure 8, the accuracy and convergence speed of the test set after CWT were better than the STFT, which is more suitable for model training. Therefore, this article used it as a time–frequency analysis method to carry out the time–frequency transformation of one-dimensional time series fault signals of the gearbox.



(**a**) (**b**)

Figure 8. Comparison of CWT and STFT under two working conditions (**a**) Working Condition 1: 20 Hz–0 V, (**b**) Working Condition 2: 30 Hz–2 V.

4.3. Model Training and Result Analysis

After constructing the network, the classification prediction results were obtained through 50 training epochs. The training convergence curves of the CBAM-ResNeXt50 model under two working conditions are shown in Figure 9. As can be seen from Figure 9, the error in the network decreased gradually with the increase in the epoch. After 10 epochs, the network converged completely. At this time, the network accuracy rate reached 99.95% and 99.875%, and the loss reached 0.0015 and 0.0135, respectively. It can be seen from the above that the CBAM-ResNeXt50 model had a good performance in fault classification.

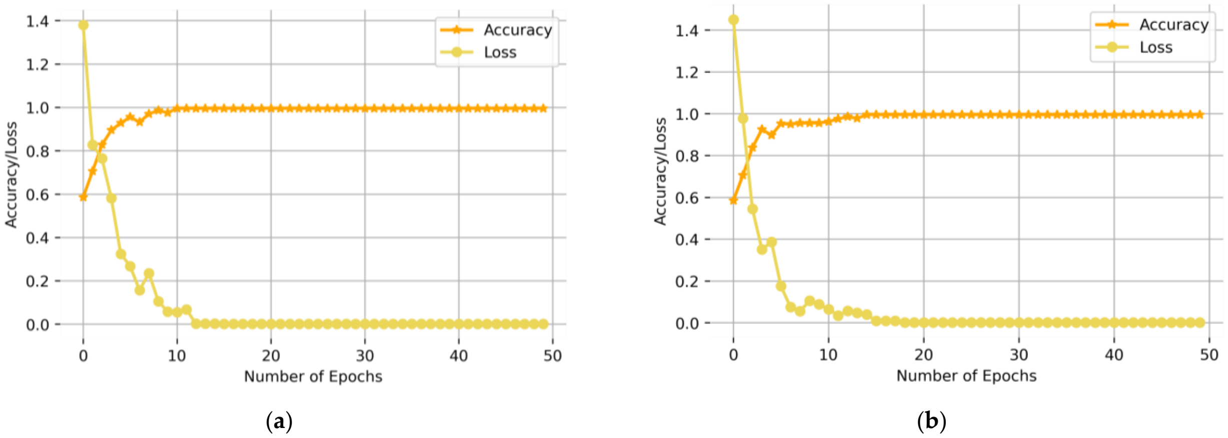


Figure 9. Accuracy and loss rate convergence curves of CBAM-ResNeXt50 under working conditions (**a**) Working condition 1: 20 Hz–0 V, (**b**) Working condition 2: 30 Hz–2 V.

For complex tasks with a limited sample size but requiring the use of a strong model, the model is prone to overfitting phenomena with a small loss on the training set and a large loss on the test set. Figure 10 shows the loss rates of the training set and the test set of the CBAM model under two working conditions. As can be seen from Figure 10, the loss rates of the model test set under the two working conditions are very small, and there is no overfitting problem.

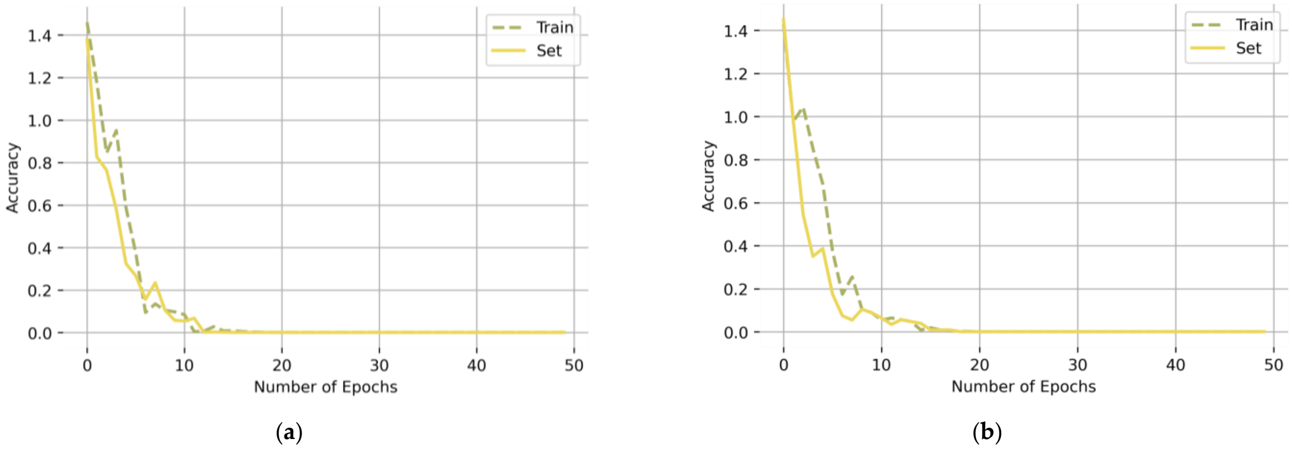


Figure 10. Accuracy of the training and test sets under working conditions (**a**) Working Condition 1: 20 Hz–0 V, (**b**) Working condition 2: 30 Hz–2 V.

To demonstrate the effectiveness of CBAM in the fusion model in this paper, comparative experiments were conducted on whether to use the CBATM module. Among them, CBAM-ResNeXt50 represented the ResNeXt50 model integrated into CBAM, and ResNeXt50 represented the original model not integrated into CBAM. At the same time, the performance of classical convolutional neural networks such as the DenseNet121, the ResNet50, and the AlexNet was compared.

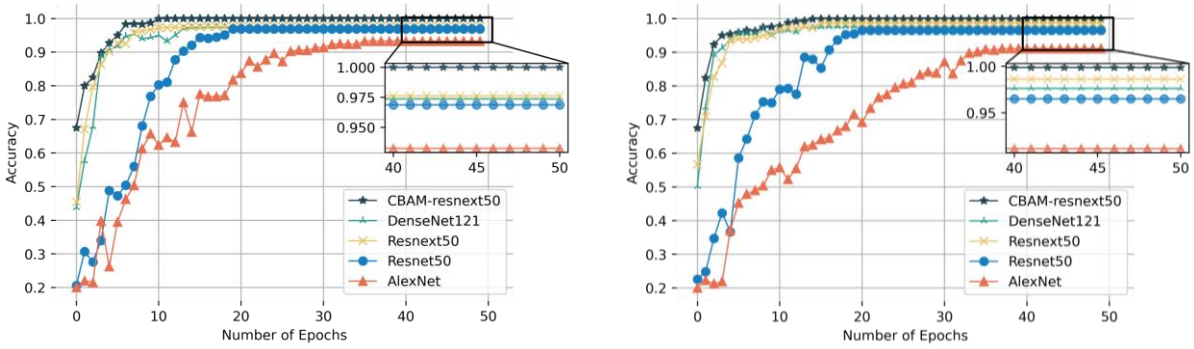
Table 3 shows the results of the training time and fault recognition accuracy of each model. It can be seen from Table 3 that the accuracy of the CBAM-ResNeXt50 model reached

99.95% and 99.875%, respectively, under two working conditions. The addition of CBAM significantly improved the accuracy of the ResNeXt50 network in identifying different faults. The experiment proved the effectiveness of CBAM for gearbox fault identification.

Table 3. Training results of five models under two working conditions.

Working Conditions	Neural Network Model	Accuracy (%)
Condition 1 20 Hz–0 V	CBAM-ResNeXt50	99.95%
	ResNeXt50	97.625%
	ResNet50	96.875%
	DenseNet121	97.25%
	AlexNet	93.25%
Condition 2 30 Hz–2 V	CBAM-ResNeXt50	99.875%
	ResNeXt50	98.75%
	ResNet50	96.5%
	DenseNet121	97.625%
	AlexNet	91.125%

Figure 11a,b shows the accuracy curves of five model test sets. Under the two working conditions, with an increasing number of epochs, the increasing trend of test accuracy of different models was different. After about 10 epochs, the CBAM-ResNeXt50 model began to converge, the DenseNet121 and ResNeXt50 models began to converge after small oscillations, while the ResNet50 and AlexNet models still did not converge. Comparing several models, we see that CBAM could improve the diagnostic accuracy and convergence speed of the model.



(**a**) (**b**)

Figure 11. Training convergence curves of five models under working conditions (**a**) Working Condition 1: 20 Hz–0 V, (**b**) Working condition 2: 30 Hz–2 V.

Figure 12 compared the average training time of five neural network models under two working conditions. The number of training epochs was 50, and the sample sizes of the training and test data sets were 3200 and 800. As can be seen in Figure 12, the average running time of the CBAM-ResNeXt50 model was shorter than those of the other three models due to the integration of CBAM. This indicated that the addition of CBAM can shorten the time of the model training on the premise of improving the accuracy of the model training. However, the AlexNet model had a single structure and only used a few convolution layer parameters, so the shortest average training time was reasonable. Although the AlexNet model took the shortest time, its model training accuracy was low and could not converge quickly.

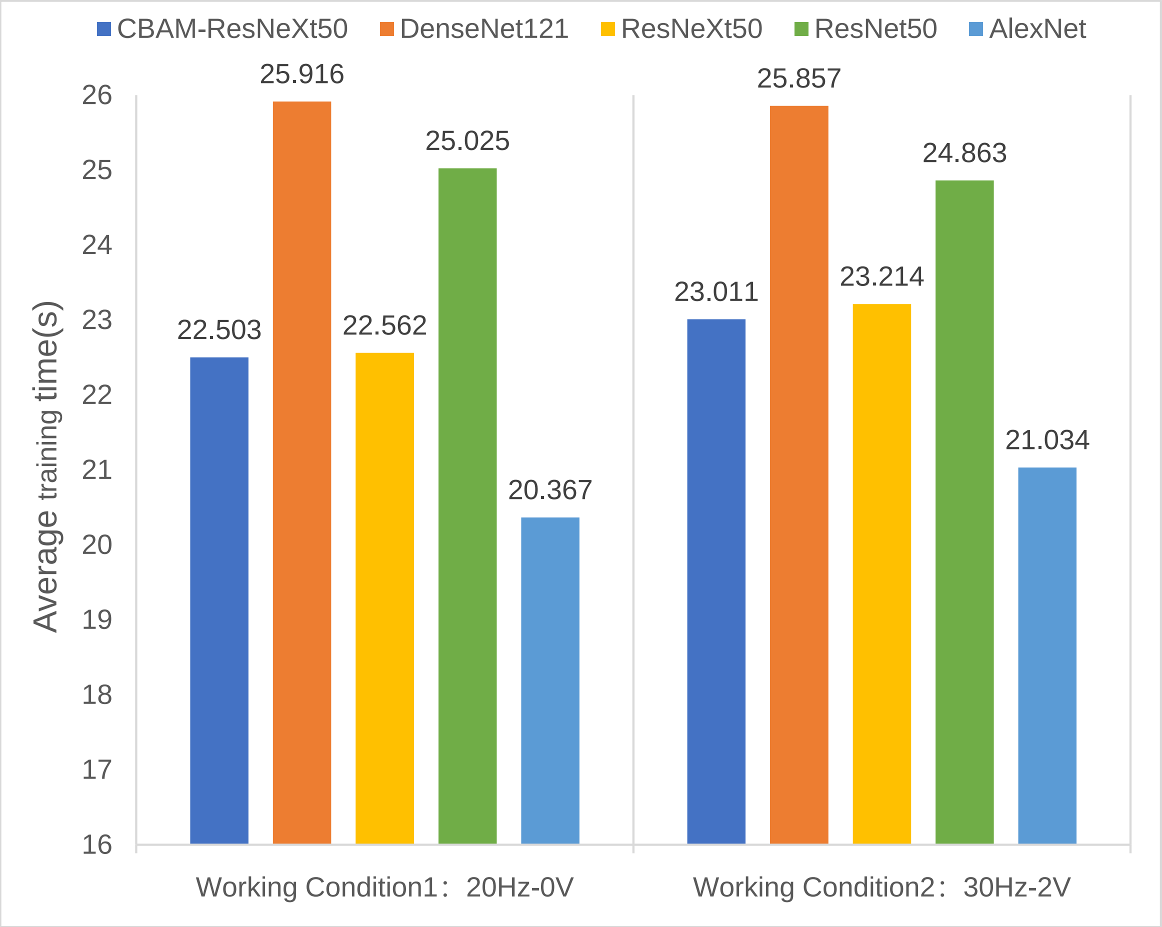
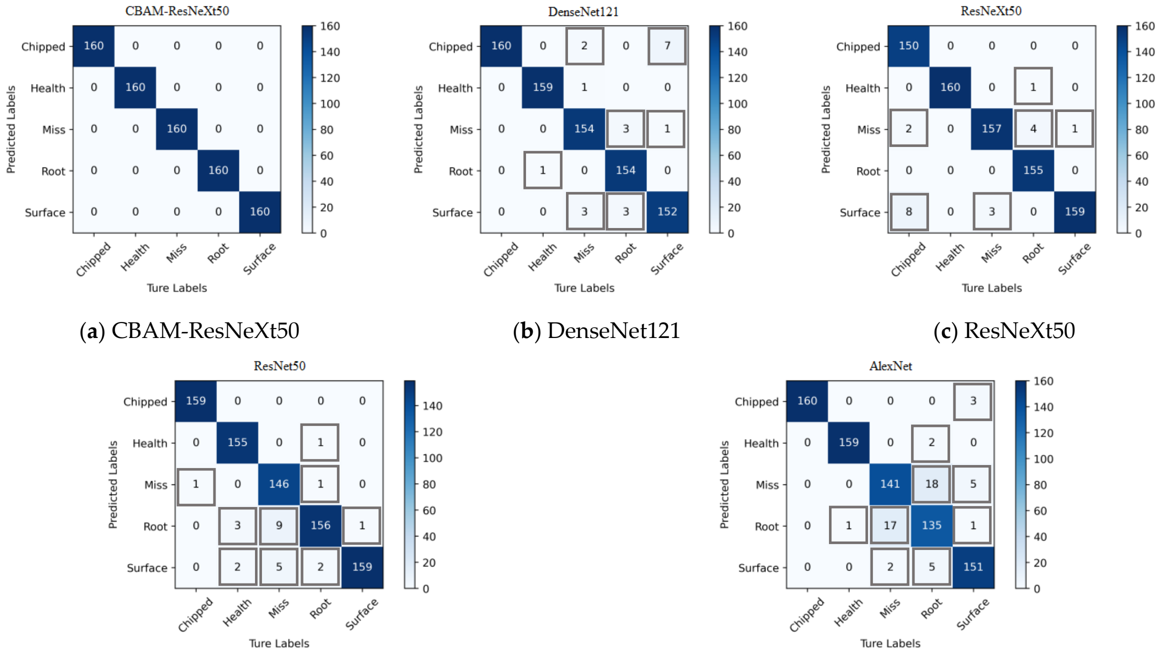


Figure 12. Comparison of the average training time of five models under two working conditions.

The two models were verified by using the two-dimensional images obtained by CWT. Figures 13 and 14 show the confusion matrix obtained by fault classification by the CBAM-ResNeXt50 model and the ResNeXt50 model under two working conditions, respectively. The abscissa represented the real fault and the ordinate represented the predicted result, which was, respectively, chipped, health, miss, root, and surface. The depth of the color represents the number of identified samples. The darker the color is, the greater is the number of correctly identified samples, and the greater is the accuracy.

In this paper, the average precision (AP) and average recall (AR) of five fault types were used to analyze the confusion matrix obtained. The precision represented the ratio of correct samples to the total samples in the prediction results. The recall rate represented the ratio of the number of predicted samples to the total number of actual samples [36].

The gearbox data set published by Southeast University was used to conduct experiments on the CBAM-ResNeXt50 model, the ResNeXt50 model, and three other convolutional network models. As can be seen from Table 4, the DenseNet121 and ResNeXt50 model both had an average precision and recall rate of around 97%, the average precision and recall rate of the ResNet50 model were both around 96%, and the average precision and recall rate of the AlexNet model were both about 93%. Based on the original ResNeXt50 model, CBAM was integrated into the method proposed in this paper, which optimized the average precision rate and the average recall rate of gear fault type detection in the model.



CBAM-ResNeXt50

	Chipped	Health	Miss	Root	Surface
Chipped	160	0	0	0	0
Health	0	160	0	0	0
Miss	0	0	160	0	0
Root	0	0	0	160	0
Surface	0	0	0	0	160

DenseNet121

	Chipped	Health	Miss	Root	Surface
Chipped	160	0	2	0	7
Health	0	159	1	0	0
Miss	0	0	154	3	1
Root	0	1	0	154	0
Surface	0	0	3	3	152

ResNeXt50

	Chipped	Health	Miss	Root	Surface
Chipped	150	0	0	0	0
Health	0	160	0	1	0
Miss	2	0	157	4	1
Root	0	0	0	155	0
Surface	8	0	3	0	159

(**a**) CBAM-ResNeXt50 (**b**) DenseNet121 (**c**) ResNeXt50

ResNet50

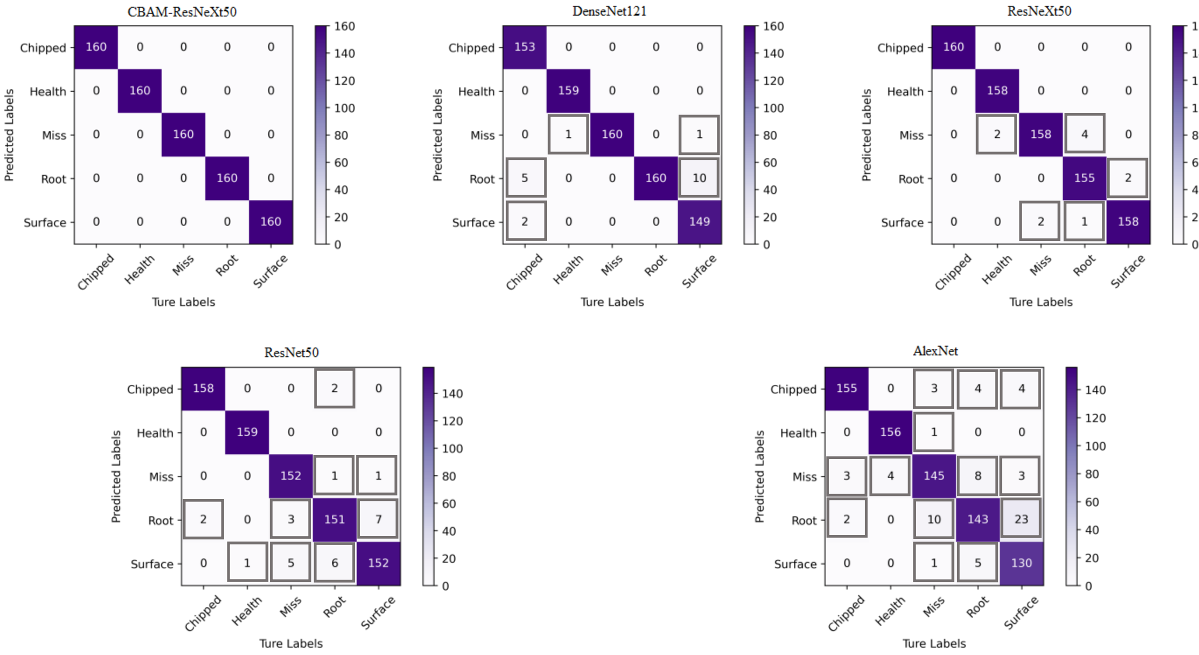
	Chipped	Health	Miss	Root	Surface
Chipped	159	0	0	0	0
Health	0	155	0	1	0
Miss	1	0	146	1	0
Root	0	3	9	156	1
Surface	0	2	5	2	159

AlexNet

	Chipped	Health	Miss	Root	Surface
Chipped	160	0	0	0	3
Health	0	159	0	2	0
Miss	0	0	141	18	5
Root	0	1	17	135	1
Surface	0	0	2	5	151

(**d**) ResNet50 (**e**) AlexNet

Figure 13. Confusion matrix for fault detection of five models under working condition 1.



CBAM-ResNeXt50

	Chipped	Health	Miss	Root	Surface
Chipped	160	0	0	0	0
Health	0	160	0	0	0
Miss	0	0	160	0	0
Root	0	0	0	160	0
Surface	0	0	0	0	160

DenseNet121

	Chipped	Health	Miss	Root	Surface
Chipped	153	0	0	0	0
Health	0	159	0	0	0
Miss	0	1	160	0	1
Root	5	0	0	160	10
Surface	2	0	0	0	149

ResNeXt50

	Chipped	Health	Miss	Root	Surface
Chipped	160	0	0	0	0
Health	0	158	0	0	0
Miss	0	2	158	4	0
Root	0	0	0	155	2
Surface	0	0	2	1	158

ResNet50

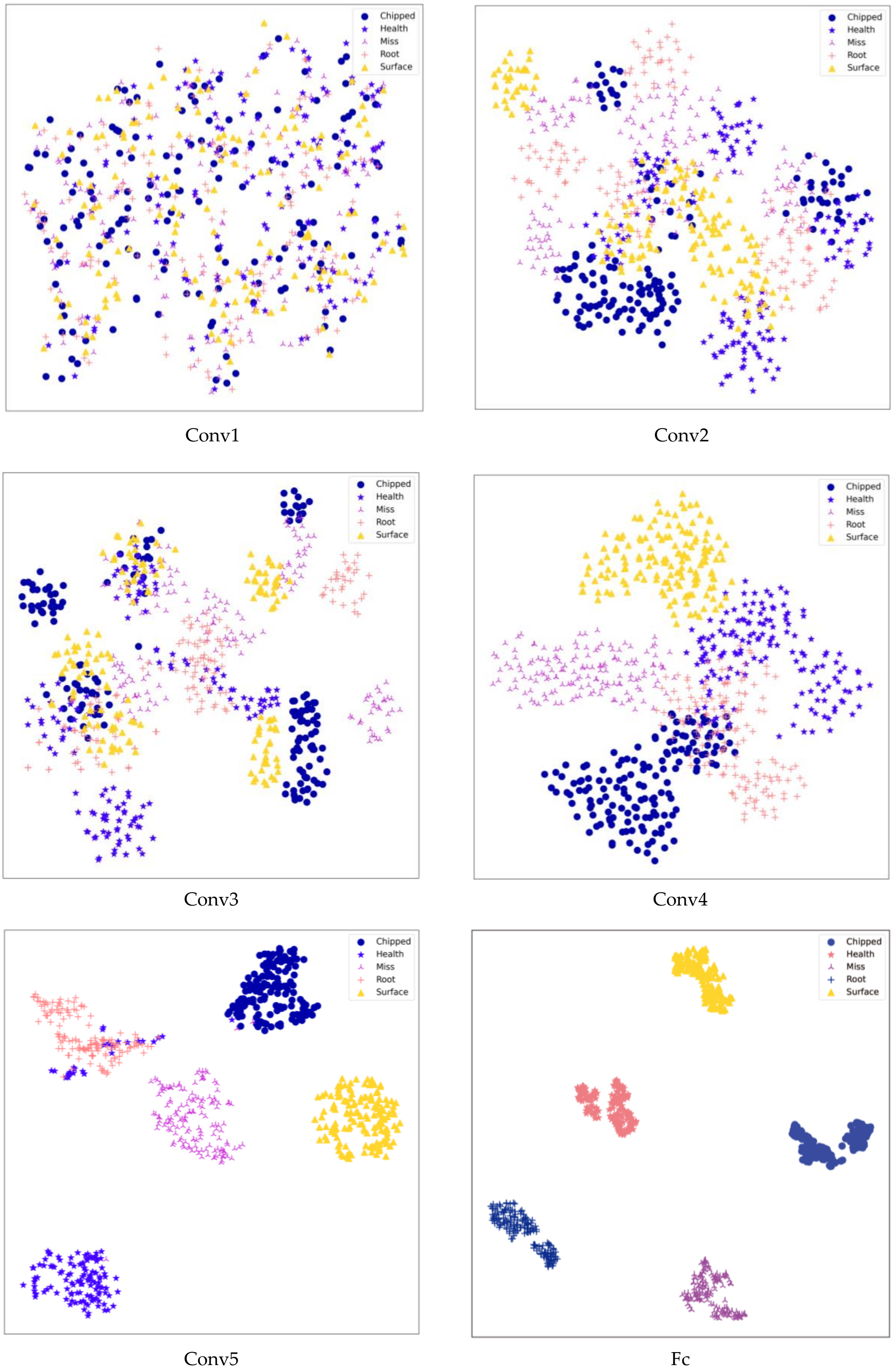
	Chipped	Health	Miss	Root	Surface
Chipped	158	0	0	2	0
Health	0	159	0	0	0
Miss	0	0	152	1	1
Root	2	0	3	151	7
Surface	0	1	5	6	152

AlexNet

	Chipped	Health	Miss	Root	Surface
Chipped	155	0	3	4	4
Health	0	156	1	0	0
Miss	3	4	145	8	3
Root	2	0	10	143	23
Surface	0	0	1	5	130

Figure 14. Confusion matrix for fault detection of five models under working condition 2.

Using the t-SNE algorithm, the prediction of the CBAM-ResNeXt50 model on the test set could be displayed visually, so that the multidimensional output prediction data could be displayed in the two-dimensional space. t-SNE technology was used to reduce the feature dimension of trained data in data set 1. The result is shown in Figure 15. Dimensionality reduction visualization showed that at the beginning of classification, different fault data were mixed and difficult to distinguish. After training, the division between classes gradually became clear and there were five clear distributions in the entire connection layer.



Conv1

Conv2

Conv3

Conv4

Conv5

Fc

Figure 15. Dimension reduction results of each layer of the CBAM-ResNeXt50 model.

Table 4. Evaluation indexes of confusion matrices of five models under working condition 1.

Neural Network Model	Average Precision		Average Recall	
	Working Condition 1	Working Condition 2	Working Condition 1	Working Condition 2
CBAM-ResNeXt50	1.0	1.0	1.0	1.0
DenseNet121	0.9741	0.9778	0.9738	0.9776
ResNeXt50	0.9747	0.9864	0.9738	0.9863
ResNet50	0.9699	0.9655	0.9687	0.9650
AlexNet	0.9322	0.9152	0.9325	0.9113

5. Conclusions

This paper presented a fault detection method for the gearbox based on the CBAM-ResNeXt50 model. Compared with other existing convolutional neural network models, a CBAM attention mechanism was added to the ResNeXt basic model in this paper, which made the model pay more attention to fault information during training. We used the gearbox data set published by Southeast University to verify the model proposed in this paper. The results show that under the two working conditions, the accuracy of the test set was 100% and 99.875%, and the average training time was 22.619 s and 23.083 s. We comprehensively compared this result with four models, including DenseNet121, ResNeXt50, ResNet50, and AlexNet, and found that the performance of the proposed method in this paper was more stable. The model parameters and experimental results provided in this paper also provide some data reference for existing fault diagnosis research. However, the existing gear fault data are few, difficult to collect, and the model training requires a large amount of data, so in the case of a small amount of data, using neural network model training and achieving good stability is one of the future research directions.

Author Contributions: Z.W.: writing—original draft preparation; Y.T.: Software and Validation; Y.D.: writing—review and editing; S.D. and H.B.: 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. This research was supported by the Project of Construction and Support for high-level Innovative Teams of Beijing Municipal Institutions, grant number: BPHR20220107.

Data Availability Statement: The public datasets used in this experiment can be obtained at: 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

1. Li, C.; Zhang, S.; Qin, Y.; Estupinan, E. A systematic review of deep transfer learning for machinery fault detection. *Neurocomputing* **2020**, *407*, 121–135. [CrossRef]
2. Jiao, J.; Zhao, M.; Lin, J.; Liang, K. A comprehensive review on convolutional neural network in machine fault detection. *Neurocomputing* **2020**, *417*, 36–63. [CrossRef]
3. Lei, Y.; Lin, J.; Zuo, M.; He, Z. Condition monitoring and fault detection of planetary gearboxes: A review. *Measurement* **2014**, *48*, 292–305. [CrossRef]
4. Sun, T.; Yu, G.; Gao, M.; Zhao, L.; Bai, C.; Yang, W. Fault detection Methods Based on Machine Learning and its Applications for Wind Turbines: A Review. *IEEE Access* **2021**, *9*, 147481–147511. [CrossRef]
5. Yang, J.; Guo, Y.; Zhao, W. Long short-term memory neural network based fault detection and isolation for electro-mechanical actuators. *Neurocomputing* **2019**, *360*, 85–96. [CrossRef]
6. Ding, Y.; Ma, L.; Ma, J.; Suo, M.; Tao, L.; Cheng, Y.; Lu, C. Intelligent fault detection for rotating machinery using deep Q-network based health state classification: A deep reinforcement learning approach. *Adv. Eng. Inform.* **2019**, *42*, 100977. [CrossRef]

7. Wang, Y.; Wu, D.; Yuan, X. LDA-based deep transfer learning for fault detection in industrial chemical processes. *Comput. Chem. Eng.* **2020**, *140*, 106964. [CrossRef]
8. Chen, Z.; Mauricio, A.; Li, W.; Gryllias, K. A deep learning method for bearing fault detection based on Cyclic Spectral Coherence and Convolutional Neural Networks. *Mech. Syst. Signal Process* **2020**, *140*, 106683. [CrossRef]
9. Xu, F.; Fang, Y.; Wang, D.; Liang, J.; Tsui, K. Combining DBN and FCM for fault detection of Roller Element Bearings without Using Data Labels. *Shock Vib.* **2018**, *2018*, 3059230. [CrossRef]
10. Shang, Z.; Liao, X.; Geng, R.; Gao, M.; Liu, X. Fault detection method of rolling bearing based on deep belief network. *J. Mech. Sci. Technol.* **2018**, *32*, 5139–5145. [CrossRef]
11. Liu, G.; Bao, H.; Han, B. A Stacked Autoencoder-Based Deep Neural Network for Achieving Gearbox fault detection. *Math. Probl. Eng.* **2018**, *2018*, 5105709. [CrossRef]
12. Dai, J.; Tang, J.; Shao, F.; Huang, S.; Wang, Y. Fault detection of Rolling Bearing Based on Multiscale Intrinsic Mode Function Permutation Entropy and a Stacked Sparse Denoising Autoencoder. *Appl. Sci.* **2019**, *9*, 2743. [CrossRef]
13. Liu, H.; Zhou, J.; Zheng, Y.; Jiang, W.; Zhang, Y. Fault detection of rolling bearings with recurrent neural network-based autoencoders. *ISA Trans* **2018**, *77*, 167–178. [CrossRef] [PubMed]
14. An, Z.; Li, S.; Wang, J.; Jiang, X. A novel bearing intelligent fault detection framework under time-varying working conditions using recurrent neural network. *ISA Trans* **2020**, *100*, 155–170. [CrossRef]
15. Wang, Z.C.; Xia, H.; Yin, W.Z.; Yang, B. An improved generative adversarial network for fault diagnosis of rotating machine in nuclear power plant. *Ann. Nucl. Energy* **2023**, 180. [CrossRef]
16. Liu, Y.; Dou, S.; Du, Y.; Wang, Z. Gearbox fault detection Based on Gramian Angular Field and CSKD-ResNeXt. *Electronics* **2023**, *12*, 2475. [CrossRef]
17. Wang, G.; Wang, J.; Yu, H.; Sui, Y. Research on Identification of Corn Disease Occurrence Degree Based on Improved ResNeXt Network. *Int. J. Pattern Recognit. Artif. Intell.* **2022**, *36*, 2250005. [CrossRef]
18. Xiao, Y.; Zhou, A.; Zhou, L.; Zhao, Y. Automatic insect identification system based on SE-ResNeXt. *Int. J. Syst. Control Commun.* **2023**, *14*, 81–98. [CrossRef]
19. Pant, G.; Yadav, D.P.; Gaur, A. ResNeXt convolution neural network topology-based deep learning model for identification and classification of Pediastrum. *Algal Res.* **2020**, *48*, 101932. [CrossRef]
20. Zheng, Q.; Chen, Z.; Liu, H.; Lu, Y.; Li, J.; Liu, T. MSRANet: Learning discriminative embeddings for speaker verification via channel and spatial attention mechanism in alterable scenarios. *Expert Syst. Appl.* **2023**, *217*, 119511. [CrossRef]
21. Jiang, C.; Liu, X.; Liu, Y.; Xie, M.; Liang, C.; Wang, Q. A Method for Predicting the Remaining Life of Rolling Bearings Based on Multi-Scale Feature Extraction and Attention Mechanism. *Electronics* **2022**, *11*, 3616. [CrossRef]
22. Wu, Z.; Jiang, H.; Zhu, H.; Wang, X. A knowledge dynamic matching unit-guided multi-source domain adaptation network with attention mechanism for rolling bearing fault detection. *Mech. Syst. Signal Process.* **2023**, *189*, 110098. [CrossRef]
23. Wu, H.; Li, J.; Zhang, Q.; Tao, J.; Meng, Z. Intelligent fault detection of rolling bearings under varying operating conditions based on domain-adversarial neural network and attention mechanism. *ISA Trans.* **2022**, *130*, 477–489. [CrossRef] [PubMed]
24. Zhang, Q.; Deng, L. An Intelligent fault detection 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, Q.; Xu, S.Q.; Cao, X.Y. Research on residual life prediction of rolling bearings based on STFT-CNN. *Testing Const.* **2022**. [CrossRef]
26. Zhou, S.; Xiao, M.; Bartos, P.; Filip, M.; Geng, G. Remaining Useful Life Prediction and fault detection of Rolling Bearings Based on Short-Time Fourier Transform and Convolutional Neural Network. *Shock Vib.* **2020**, *2020*, 8857307. [CrossRef]
27. Aguiar-Conraria, L.; Soares, M. *The Continuous Wavelet Transform: A Primer*; NIPE-Universidade do Minho: Braga, Portugal, 2011.
28. Chen, S.L.; Li, J.L. The contract finite element analysis for the small module gear of the car door lock devices. *J. Mech. Transm.* **2012**, *36*, 77–80.
29. Li, Z.; Li, B.; Ni, H.; Ren, F.; Lv, S.; Kang, X. An Effective Surface Defect Classification Method Based on RepVGG with CBAM Attention Mechanism (RepVGG-CBAM) for Aluminum Profiles. *Metals* **2022**, *12*, 1809. [CrossRef]
30. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19. [CrossRef]
31. Cui, W.; Meng, G.; Gou, T.; Wang, A.; Xiao, R.; Zhang, X. Intelligent Rolling Bearing fault detection Method Using Symmetrized Dot Pattern Images and CBAM-DRN. *Sensors* **2022**, *22*, 9954. [CrossRef]
32. Chen, C.; Wu, B.; Zhang, H. An Image Recognition Technology Based on Deformable and CBAM Convolution Resnet50. *IAENG Int. J. Comput. Sci.* **2023**, *50*, IJCS_50_1_30.
33. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5987–5995. [CrossRef]
34. Liang, Z.; Tao, M.; Wang, L.; Su, J.; Yang, X. Automatic Modulation Recognition Based on Adaptive Attention Mechanism and ResNeXt WSL Model. *IEEE Commun. Lett.* **2021**, *25*, 2953–2957. [CrossRef]

35. Shao, S.; Stephen, M.; Yan, R.; Pierre, B. Highly accurate machine fault detection using deep transfer learning. *IEEE Trans. Ind. Informatics* **2018**, *15*, 2446–2455. [CrossRef]
36. Fränti, P.; Mariescu-Istodor, R. Soft precision and recall. *Pattern Recognit. Lett.* **2023**, *167*, 115–121. [CrossRef]

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.