machines

MDPI

Article

Towards a Fault Diagnosis Method for Rolling Bearings with Time-Frequency Region-Based Convolutional Neural Network

Jiahui Tang [1], Jimei Wu [1,2,*], Bingbing Hu [2] and Jiajuan Qing [1]

[1] School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an 710048, China

[2] Faculty of Printing, Packing and Digital Media Engineering, Xi'an University of Technology, Xi'an 710054, China

* Correspondence: wujimei@xaut.edu.cn; Tel.: +86-1535-362-1328

Abstract: An artificial-intelligence (AI)-based method for fault diagnosis is a strong candidate for industrial applications in the health management of rolling bearings. However, traditional fault diagnosis methods fail to improve the detection accuracy because they only extract a single feature and have limitations in feature representation. In addition, advanced object detection frameworks such as region-based convolutional neural networks have not yet been applied in fault diagnosis. To this end, a fault diagnosis model using a Time-Frequency Region-Based Convolutional Neural Network (TF-RCNN) is proposed in this paper. This method was mainly adopted to extract multiple regions that can characterize fault features from the Time-Frequency Representation (TFR). Specifically, an attention module was introduced so the model could focus on representative features. The existing classification strategy was also enhanced to perform multiple types of fault classification. Finally, an end-to-end rolling bearing fault diagnosis framework based on the TF-RCNN was developed with the aforementioned improvements. The effectiveness of this method was proven experimentally on artificial faults and real faults. The superiority of the proposed method is demonstrated using a comparison with the typical object detection method and an advanced fault diagnosis method.

Keywords: fault diagnosis; TF-RCNN; time-frequency representation; rolling bearing

check for updates

Citation: Tang, J.; Wu, J.; Hu, B.; Qing, J. Towards a Fault Diagnosis Method for Rolling Bearings with Time-Frequency Region-Based Convolutional Neural Network. *Machines* **2022**, *10*, 1145. https://doi.org/10.3390/machines10121145

Academic Editors: Wanqing Song and Francesco Villecco

Received: 26 October 2022
Accepted: 23 November 2022
Published: 1 Decemeber 2022

1. Introduction

Bearings are among the most significant components in rotating machinery. The leading causes of various bearing faults are harsh working environments, heavy loads, and high-speed operation conditions [1,2]. When untimely, the diagnosis and repair of bearing faults can lead to serious production accidents and economic losses. Hence, bearing health management is crucial to increase the service life of rotating machinery, which can also reduce the maintenance cost [3–5].

At present, there are many state-of-the-art and effective fault diagnosis methods for bearings. These methods can be roughly divided into the following two categories: signal-processing-technology-based methods and artificial-intelligence-based methods [6,7]. In general, fault diagnosis methods based on signal processing technology mainly identify faults by extracting the signal components related to the fault features or by using classical techniques to model the signals in the time or frequency. Wang et al. [8] replaced the framing and feature operators with traditional envelopes, and the modulation of the rotation frequency component was eliminated, which resulted in the improvement of the recognition effects. A sparse signal reconstruction approach combining time-frequency manifolds was proposed by He et al. [9], and the results revealed that this approach can enhance the fault features of rolling bearings. De Moura et al. [10] proposed a signal preprocessing technique by combining rescaled range analysis with a detrended fluctuation analysis and applied this technique to the fault monitoring and diagnosis of rolling bearings. Ai et al. [11] introduced a method named the n-dimensional characteristic parameters' distance; using this method, the bearing fault diagnosis was achieved by incorporating four information entropies. An obvious drawback of the method is the limited range of application. This method fails to accurately extract fault feature components under a working environment with a heavy background and other interferences because the vibration signals of the non-detected components can easily mask the fault features. In the context of high-speed and intelligent mechanical equipment, the magnitude of the vibration signals generated by the equipment increases, and the efficiency of traditional fault diagnosis methods often does not meet the expectations.

On the other hand, artificial-intelligence-based algorithms have recently received considerable attention for fault diagnosis. Machine learning algorithms were initially employed to learn the relationship between the mechanical equipment and health status from the generated signals [12]. Chen et al. [13] proposed a method called Adaboost Support Vector Machine (Adaboost-SVM) to address the early fault diagnosis of rolling bearings. Zhang et al. [14] developed a method combining K-nearest-neighbors (k-NN) and an optimal multi-kernel local Fisher discriminant analysis, and this approach improved the recognition of the bearings' condition via the most characteristic feature space. Wan et al. [15] presented a fault diagnosis method based on an improved random forest, and the experimental results indicated that this method had a faster training speed and better fault diagnosis efficiency. However, the limited feature representation ability of machine learning algorithms means that they fail to learn complex nonlinear relationships, which results in a lower diagnostic accuracy.

With the development of deep learning, which has gradually replaced traditional machine learning methods, the mainstream algorithms for intelligent fault diagnosis use methods with more nonlinear transformation layers [16]. Wen et al. [17] developed a hierarchical-convolutional-neural-network (HCNN)-based fault diagnosis method and presented the corresponding model-training method. Li et al. [18] investigated a fault diagnosis approach based on a deep convolutional neural network (DCNN) and performed the diagnosis of newly added faults without exchanging the data through federated learning. A principal component analysis (PCA) was employed to perform dimensionality reduction for the bearing vibration signal of a deep belief network (DBN), and this network realized fault classification and diagnosis using the fault features of the preprocessed signal in [19]. The original feature set was constructed with the dual-tree complex wavelet packet (DTCWPT) by Shao et al. [20], and they also built an adaptive deep belief network (DBN) to achieve bearing fault diagnosis. Guo et al. [21] put forward a fault diagnosis method by combining the DBN and a double-sparse dictionary model, and this approach greatly reduced the training time. Song et al. [22] considered the long short-term memory (LSTM) network as the fundamental architecture for transfer learning, and this structure was employed for bearing fault diagnosis in the presence of a working environment with a changeable load. Liu et al. [23] introduced a fault diagnosis model based on a low-delay, lightweight recurrent neural network (RNN) with less memory used for computation.

In addition, considering the complex functioning of rotating machinery, the appearance of background noise, and the defects of measurement technology in industrial production, it is essential to introduce advanced signal feature extraction technology for extract features from the raw data and then input them into the deep learning model. Xu et al [24]. utilized time-frequency features to train a convolutional neural network (CNN) and measured the Euclidean distance between the semantic features and signal features for compound fault diagnosis. Shi et al. [25] generated a high-quality TFR with a redundant second-generation wavelet transform and employed an adversarial network to perform a single-component health assessment and identify multicomponent fault locations. Ma et al. [26] developed a transfer learning model based on Alexnet, and this model extracts the features of the TFR to achieve the classification of the bearing fault. These investigations of advanced methods proved that replacing the time domain signals with two-dimensional (2D) images as the input can improve the accuracy of intelligent diagnosis. The 2D image used in fault diagnosis is generally obtained by the time-frequency transformation of the vibration signals, and the energy intensity distribution diagram of the signal at different periods and frequencies can be obtained from the image.

The limitations of the above studies can be summarized as follows: (1) The existing researches all use the TFR as the input. All components in the same sample have a relationship with the fault category, which lacks the fault signal intensity distribution and also increases the importance of the noise intensity distribution. (2) The diagnosis mode of the one-to-one relationship between samples and categories restricted the accuracy of fault diagnosis methods. An increasing number of scholars pay considerable attention to further improving diagnosis accuracy. (3) The advanced object detection theory can promote the industrial application of intelligent fault diagnosis in the future. Theories that can accurately map the fault features related to fault conditions from TFR provide insights into fault diagnosis. Nevertheless, there is still a lack of relevant research on the accurate application of object detection theory.

Based on the previous studies, object detection is introduced into the fault diagnosis in this paper, and the fault features are identified from the input signals to accurately establish the corresponding relationship between the fault features and the fault categories. Faster RCNN [27,28] is a widely followed object detection algorithm, and has a substantial value in many fields. In this paper, two attention modules are introduced into the object detector based on the original algorithms, which improve the sensitivity of the backbone network to features and obtain more reliable fault features from the TFR. The classification strategy in the original model is then improved to accurately identify the fault category. Finally, the experiment is tested on the bearing dataset, and the results show that the proposed method can accurately establish the relationship between features and fault categories. The main contributions of the present study can be described as presented below.

(1) The object detection theory is applied to build a rolling bearing fault diagnosis method based on TF-RCNN.

(2) The attention module is introduced to allow the model to focus on representative features, thus improving the sensitivity of the model to fault features.

(3) A new classification strategy is designed to enable the model to adapt to multi-class fault diagnosis, and the model is tested with artificial faults and realistic faults.

The remaining sections are structured as follows. Section 2 provides a brief overview of time-frequency processing and Faster RCNN. The TF-RCNN methodology is described in Section 3. We propose a novel diagnosis procedure in Section 4. In the next section, two case studies are analyzed. We close the paper with a conclusion in Section 5.

2. Preliminaries

2.1. Time-Frequency Processing with Wavelet Transform

The time-frequency analysis used in this study is the currently dominant continuous wavelet transform (CWT) method [29]. CWT can display the time-frequency characteristics of the vibration signal from the generated 2-D image, and the CWT is considered to be one of the most appropriate algorithms for investigating non-linear and non-stationary signals. In particular, this method carries out the correlation coefficients by performing convolved operations between the mother wavelet and the original signal. Afterwards, the original signal is transformed into a time-frequency domain image based on the correlation coefficients, from which information about the fault frequency can be obtained [30,31].

Mathematically, a signal is assumed to be $\psi(t) \in L^2(R)$, and its Fourier transform can be obtained with the following equation:

$$\int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty \tag{1}$$

where $\hat{\psi}(\omega)$ is the Fourier transform result of $\psi(t)$, and ω is the frequency. A family of wavelet sequences can be achieved by shifting and stretching the mother wavelet.

$$\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right)a, b \in R, a > 0 \tag{2}$$

where $\psi_{a,b}(t)$ is the analytic wavelet, a and b are the scale factor and shift factor, respectively.

The scale factor is applied to stretch or expand the mother wavelet to change its shape. The shift factor controls the shift of the mother wavelet of the original signal. The dynamic frequency characteristics of the original signal can be achieved by adjusting these two factors.

A CWT of an arbitrary finite energy signal $[x(t)$ can be described as:

$$W(a,b)=\langle x(t), \psi_{a,b}(t)\rangle = |a|^{-1/2}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{3}$$

where $W(a,b)$ represents the wavelet transform coefficients, and * denotes the conjugate. As can be observed from Equation (4), the CWT maps a 1-D signal into a 2-D time-frequency space. For the inspection of bearing faults using wavelet transform, the filter generates the frequency band correlated with the fault. Next, the TFR is created after obtaining the envelope signal.

The selection of an appropriate mother wavelet is essential for CWT. The commonly employed mother wavelet functions contain the Coiflet, Symlet, Haar and Morlet wavelets. Within the category of CWT, the Morlet wavelet has a relatively lower error rate for extracting impulse information in vibration signals. In addition, the time domain waveform of the Morlet wavelet [32,33] matches the impact characteristics of the bearing fault and achieves its matching relationship with the fault signal in terms of time and frequency resolution.

The mother wavelet of the Morlet can be determined as follows:

$$\psi(t) = c \cdot e^{-\sigma^2 t^2 + j2\pi f_c t} \tag{4}$$

where c is the normalization constant, σ is the shape factor and f_c is the central frequency.

The procedure for sample conversion in this paper is shown in Figure 1. Firstly, a truncated window with 1000 dimensions slides along the original vibration signal, and each move of the window generates one sample of 1000 points. After that, these samples are converted to a TFR utilizing CWT. Finally, the process is repeated until enough training samples are generated.

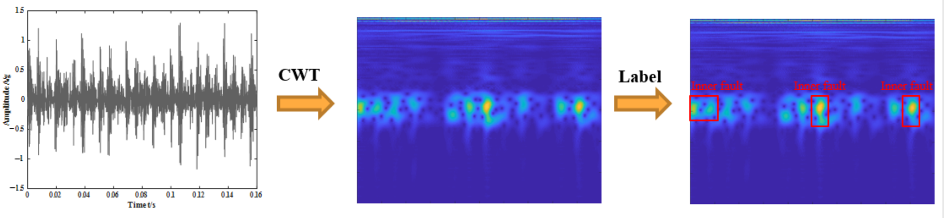


Figure 1. The process of sample transforms.

2.2. *The Framework of Faster RCNN*

As one representative of the two-stage algorithms, Faster RCNN is an improvement of the traditional RCNN network model that introduces a new region proposal network (RPN) for providing proposal boxes. This algorithm has a wide range of applications in many fields such as autonomous driving, monitoring security, drone scene analysis and face recognition. A basic Faster RCNN architecture is shown in Figure 2. It mainly consists of the following four components: feature extraction, the region proposal network (RPN), region pooling and status classification.

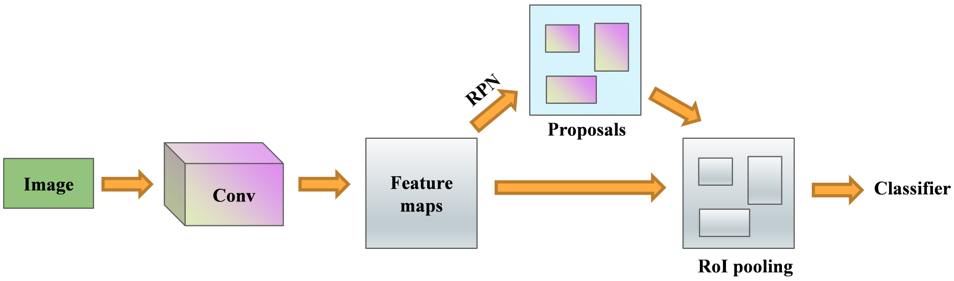


Figure 2. The basic structure of Faster RCNN.

2.2.1. Feature Extraction

In the block, the backbone network is applied to extract feature maps from the input images, which are shared in the subsequent RPN and RCNN. The common backbone networks include VGG, ResNet, etc. To increase the sensitivity of backbone networks to fault features, the attention module was introduced to ResNet in this paper. More details are shown in Section 3.1.

2.2.2. Region Proposal Network

It is worth noting that the aim of using RPN is to obtain the best box containing the target object, but the inevitable overlapping anchor boxes tend to degrade the performance of subsequent models. Therefore, if there are multiple anchor boxes with overlapping borders, the Non-maximum suppression (NMS) algorithm is utilized to fine-tune the position and scale of these boxes, and the box with the highest target confidence (or scores) is ultimately preserved. The core of the NMS algorithm is the filtering of boxes based on their scores, and this process considers boxes with low scores based on the intersection-over-union (IoU) condition. The IoU can be expressed as follows:

$$IoU = \frac{area(A_1) \cap area(A_2)}{area(A_1) \cup area(A_2)} \tag{5}$$

In general, if $IoU > 0.5$, the boxes are kept by default and vice versa.

2.2.3. RoIs Pooling

RoIs pooling can be interpreted as a simplified form of spatial pyramid pooling. Its principle is to map the RoIs output from the previous process to the corresponding position in the feature map. In addition, the RoIs are obtained by performing offset correction and the selection of anchor boxes with different sizes and ratios. It is therefore noted that their dimensions are different. Nevertheless, subsequent fully connected networks require a fixed input dimension, so that the role of RoIs pooling is to transform the different dimensional RoIs to meet the input requirements.

2.2.4. RCNN Fully Connected Network

The fully-connected layer is fed features with the same dimensionality delivered by the RoIs pooling layer. Moreover, classification and bounding box regression calculation are based on these features. In this procedure, the loss between the predicted box and the ground truth is calculated, and the loss is back-propagated to optimize the parameters.

3. Time–Frequency RCNN

The proposed TF-RCNN improves the precision of traditional object detection. Meanwhile, it can be better applied to rolling bearing fault diagnosis, for which specific explanations are described below.

3.1. Backbone Network with Attention

Traditional backbone networks are dominated by multiple variants of CNN, which extract features by mixing information across channels and space. To help the backbone network capture the connections between different objects, and to improve the sensitivity and relevance of the model to different pixels in the feature map, the backbone network introduces the attention mechanism. In detail, an improved backbone is built using the channel and spatial attention module to emphasize representative features along those two principal dimensions [34].

3.1.1. Attention Module

The channel and spatial attention modules are adopted to enhance the feature sensitivity of the model. The channel attention module (as shown in Figure 3) uses max pooling and average pooling to aggregate different spatial information from a feature map; then, the information is fed into a shared network of multi-layer perceptron (MLP) to generate a channel attention map. Finally, the output vectors are merged based on element-wise summation. The channel attention module can be described as follows:

$$F_{out}{=}sigm(MLP(Avgpool(F_{in})){+}MLP(Maxpool(F_{in}))) \tag{6}$$

in which $sigm$ denotes the sigmoid function, and F_{in} , and F_{out} represent the input feature and output feature, respectively.

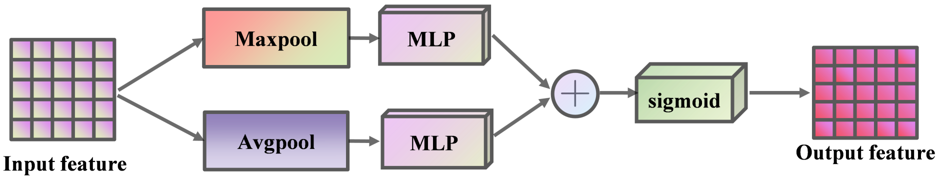


Figure 3. Structure of channel attention module.

For the spatial attention module (as shown in Figure 4), the meaningful features are initially extracted by using max pooling and average pooling along the channel axis. The convolution layer accepts the concatenation of both types of features and transforms them into a spatial feature map. Significantly, the feature map encodes information on the position of emphasis or suppression. This process can be achieved as follows:

$$F_{out} = sigm(Conv_{7\times7}([Avgpool(F_{in})Maxpool(F_{in})])) \tag{7}$$

in which $Conv_{7\times7}$ denotes a convolution operation with a kernel size of 7×7.

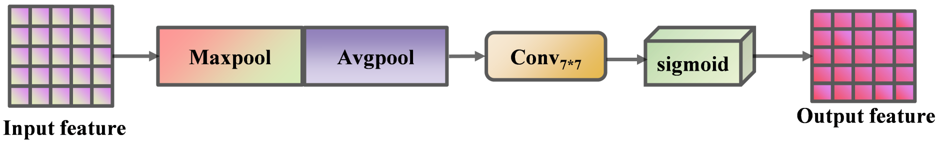


Figure 4. Structure of spatial attention module.

3.1.2. Architecture of the Attention ResNet

Although the TFR of the vibration signal can indicate local features of the bearing fault, the TFRs of different fault types are sometimes not significantly different. This phenomenon explains the inability of traditional backbone networks to learn inaccurate features. For this purpose, channel and spatial attention modules are introduced in the backbone network of the traditional Faster RCNN in this paper, thereby helping the network to emphasize effective features from these two main dimensions and achieve an effective feature flow.

The ResNet is selected as the backbone network for feature extraction. ResNet has received considerable attention in recent research on object detection due to identity shortcuts. The residual building unit (RBU) is the basic structure of ResNet. In detail, a standard RBU contains two convolutional layers, two activation layers, two batch normalization layers, and an identity shortcut. The RBU structure with two attention modules is shown in Figure 5.

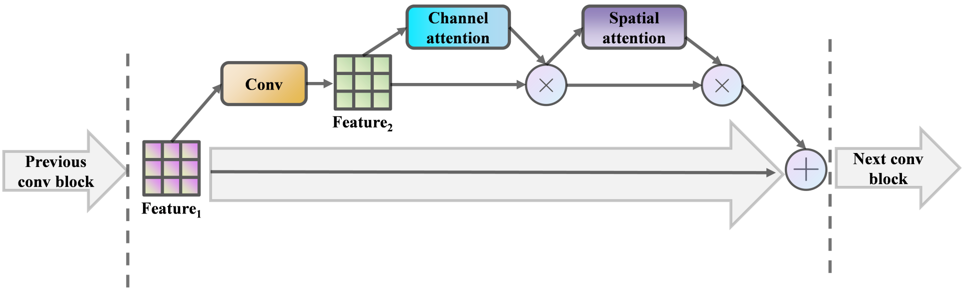


Figure 5. Attention module integrated with RBU in ResNet.

3.2. Classification Strategy

The classification block of the standard Faster RCNN is fed with RoIs generated from region pooling and their corresponding feature maps. Next, the category probabilities are calculated using *softmax*, and the final prediction position for each RoI is calculated with a fully connected network. To improve the applicability of the TF-RCNN in fault diagnosis, this paper further improves on the standard faster RCNN classification strategy

To identify the fault category of the TFR of the bearing fault, the proposed classification strategy integrates all RoIs of the whole feature map and sorts the categories probability of each corresponding RoI. Finally, the class with the highest probability among all RoIs is the output.

3.3. Overview of Time-Frequency RCNN

The overall architecture of the proposed TF-RCNN is shown in Figure 6. It is revealed from the figure that this method is divided into the following four parts: feature extraction, RPN part, RoI pooling, and classification.

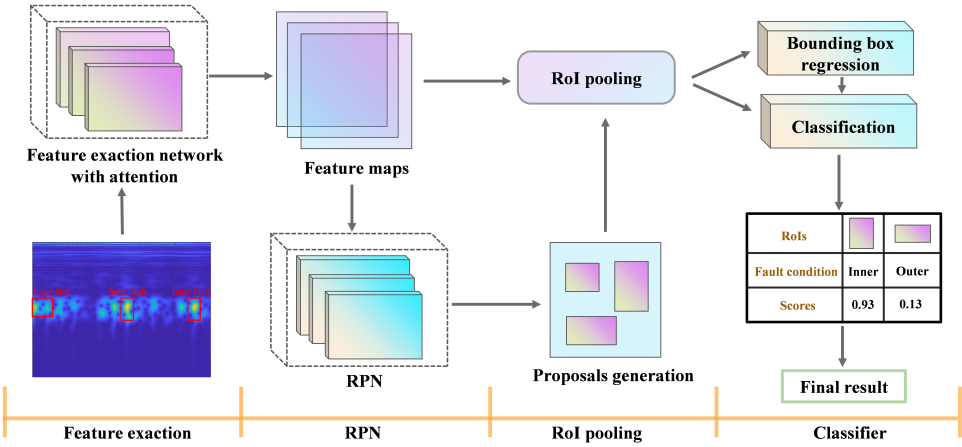


Figure 6. The network architecture of TF-RCNN.

4. Fault Diagnosis Framework Based on TF-RCNN

We propose a novel approach named TF-RCNN. This approach enhances previous methods by introducing object detection theories in fault diagnosis. Figure 7 shows the proposed framework for bearing fault diagnosis. The general diagnosis procedures are outlined below.

(1) The vibration signal of the rolling bearing is collected by the signal acquisition system.

(2) The vibration signal is transformed into a TFR via CWT. Then, these images are labelled with the corresponding working conditions.

(3) The proposed TF-RCNN is initialized, e.g., learning rate, batch size, iterations number, etc.

(4) The model is fully trained using randomly selected training samples, and the optimal hyperparameters and model structure are determined by cross-validation experiments.

(5) The trained TF-RCNN predictor is adopted for rolling bearing fault detection.

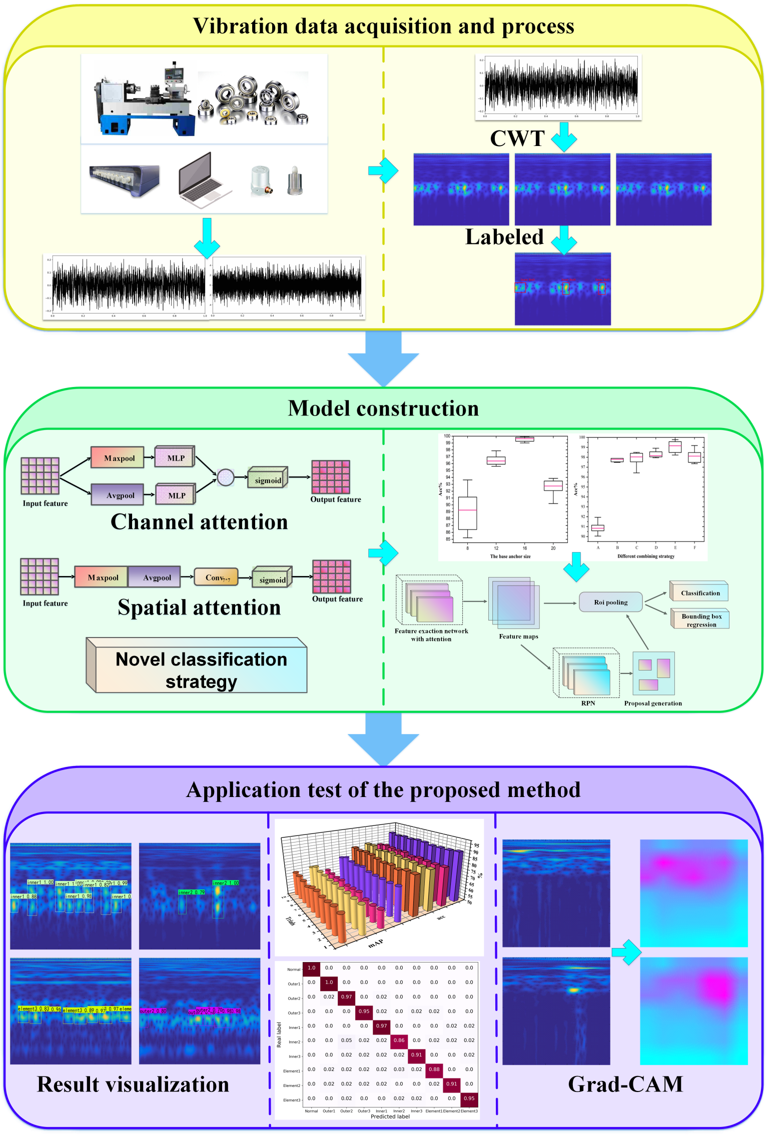


Figure 7. The framework of the proposed method.

5. Case Analysis

To assess the effectiveness of the proposed method, two cases are performed in this section. Case 1 investigates artificial damages, and Case 2 analyzes realistic damages. The present method is written in Python 3.7 by Pytorch, and the system configuration is Intel Core i7-9700 3.00 GHz CPU; 8 GB RAM; 128 GB SSD.

5.1. Case 1: Diagnosis for Artificial Damages

5.1.1. Dataset Description

Here, the bearing fault dataset is provided by the Case Western Reserve university (CWRU) [35]. Figure 8 displays the bearing test bench at CWRU. The fault of the drive-end motor bearing is produced by electric spark cutting (EDM) technology. There are four bearing working conditions, namely inner race fault, outer race fault, ball fault and normal. Each fault contains three diameters: 0.18, 0.36 and 0.54 mm. The motor operates under four loads (0–3 horsepower), and the sampling frequency is 12 kHz. Details about the dataset are listed in Table 1.

After truncating, the collected signals are converted into a time-frequency diagram as mentioned in Section 2.1. These TFRs are the input of the TF-RCNN fault diagnosis model.

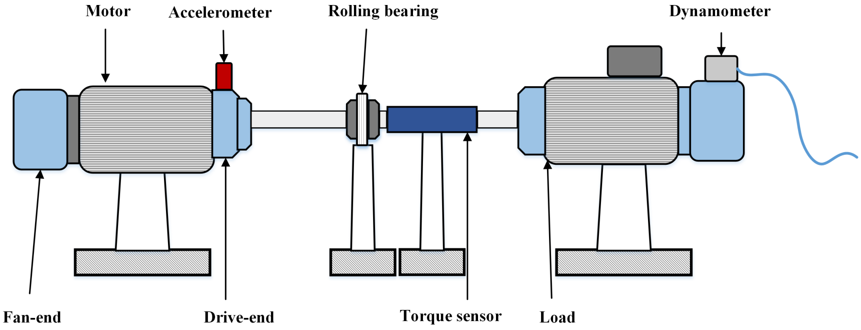


Figure 8. The platform for bearing fault diagnosis experiments of CWRU.

Table 1. Distribution of rolling bearing dataset.

Bearing Working Condition	Fault Diameter	The Number of Samples
Normal	—	200
	0.1778	200
Inner race fault	0.3556	200
	0.5334	200
	0.1778	200
Outer race fault	0.3556	200
	0.5334	200
	0.1778	200
Ball fault	0.3556	200
	0.5334	200

5.1.2. Model Selection

In this paper, all parameters (learning rate, input size, dropout value, base anchor size, NMS threshold and attention combining strategy) are determined by performing a trial-and-error analysis. To facilitate accurate analysis, these parameters are cross-validated ten times, and evaluated according to the average diagnostic accuracy. From the analysis result, the base anchor size and attention combining strategy have a considerable impact on results.

As illustrated in Figure 9a, too small an initial base anchor size enables the model to only extract local features, which leads to an inferior identification ability and poorer network performance. However, too large an initial base anchor size enables the model to extract irrelevant features, which results in lower accuracy. Figure 9b presents the combining strategy of different attention module, where A: ResNet(base), B: ResNet + channel, C: ResNet+spatial, D: ResNet+spatial+channel, E: ResNet+channel+spatial, F: ResNet+channel & spatial(parallel). Both attention modules enhance the performance of the ResNet model. The channel-first approach (E) has better performance than the space-first approach. It also can be concluded that inappropriate parameter settings cause poorer network performance as well as a more dispersed error distribution. Finally, Table 2 demonstrates the optimal parameters of this model.

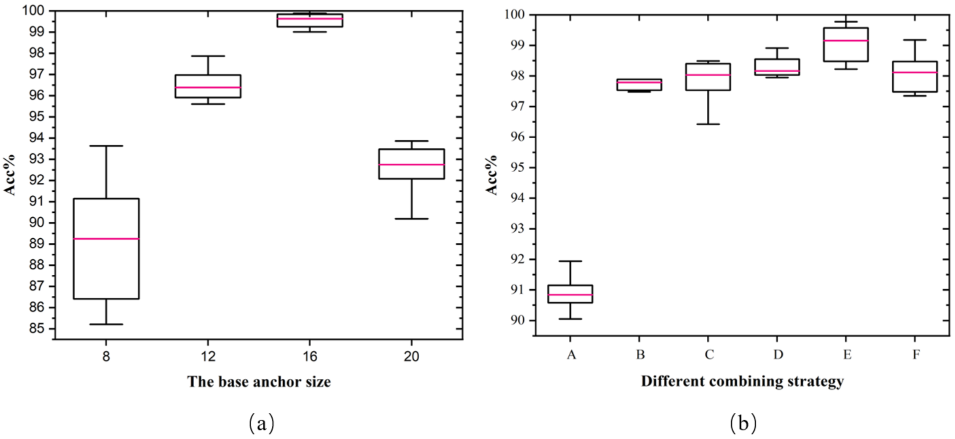


Figure 9. The relationship between diagnosis accuracy and (**a**) base anchor size (**b**) attention combining strategy.

Table 2. Experimental parameters of faults diagnosis for rolling bearing.

Parameter Description	Value
Input size	[600, 600]
Epochs	100
Learning rate	4×10^{-4}
Batch size	4
Optimizer	Adam
Attention combining strategy	channel + spatial
The base anchor size	16
Dropout	0.4
NMS threshold	0.5

5.1.3. Result Analysis

To demonstrate the advantages of the proposed method, the standard Faster RCNN, Single Shot MultiBox Detector (SSD), and You only look once (YOLO) models are taken as comparison methods. All initial configurations are consistent for these models to ensure the reliability of this experiment. Feature extraction networks with two depths, ResNet50 and ResNet101, are adopted to verify the stability of model performance. In this paper, all classifiers are based on the proposed classification strategy.

The evaluation metrics of the model are the mean average precision (mAP), diagnosis accuracy and standard deviation, respectively. The mAP can be defined as follows:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{8}$$

Precision is the fraction of positive predictions that actually belong to the positive class.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{9}$$

Recall is the fraction of positive examples in the dataset that are predicted as positive.

Average precision is the precision averaged across all recall values between 0 and 1. This indicator is obtained by calculating the area under the curve (AUC) of the precision versus recall curve. The different detectors in this paper are compared with mAP.

The comparison results between the present method and other target detection methods are shown in Table 3. These obtained results are the average results of ten repetitions of an experiment. It is apparent from the table that the proposed method with different backbone depths has a larger mAP (89.09%, 91.13%) than the others. This result proves the attention module introduced in feature extraction can be regarded as a novel design paradigm for the backbone network. More importantly, the proposed method achieves the highest accuracy (99.74%, 99.61%), which is remarkably higher than other methods. It is because the proposed method has an excellent feature representation capability. The performance of other methods is slightly disappointing in comparison, since their basic convolution and multiplication operations lack the grasp of most representative features. The proposed method also has a smaller diagnostic variance, which results in higher accuracy and better stability.

Table 3. Performance of comparison methods based on object detection.

Approach	Backbone	mAP	Diagnosis Accuracy	Standard Deviation
Standard Faster RCNN	ResNet50	86.34%	90.13%	0.61
	ResNet101	88.62%	91.06%	0.54
SSD	ResNet50	87.81%	92.52%	0.60
	ResNet101	88.58%	94.71%	0.44
Yolo	ResNet50	85.35%	91.55%	0.30
	ResNet101	85.79%	93.24%	0.37
Proposed method	ResNet50 with attention	89.09%	99.74%	0.36
	ResNet101 with attention	91.13%	99.61%	0.34

To test the effectiveness of the present method in fault diagnosis, advanced networks such as CNN, deep auto-encoder (DAE), and DBN are selected for comparison. Table 4 presents the model structure and hyperparameter settings. It is worth noting that the input of these networks is the TFR of the vibration signal after CWT processing.

Table 4. Structure and hyperparameter settings.

Model	Structure	Hyperparameter

CNN	ResNet50
	Dense (10)
		Dropout = 0.5
		Batch size = 32
		Optimizer = Adam

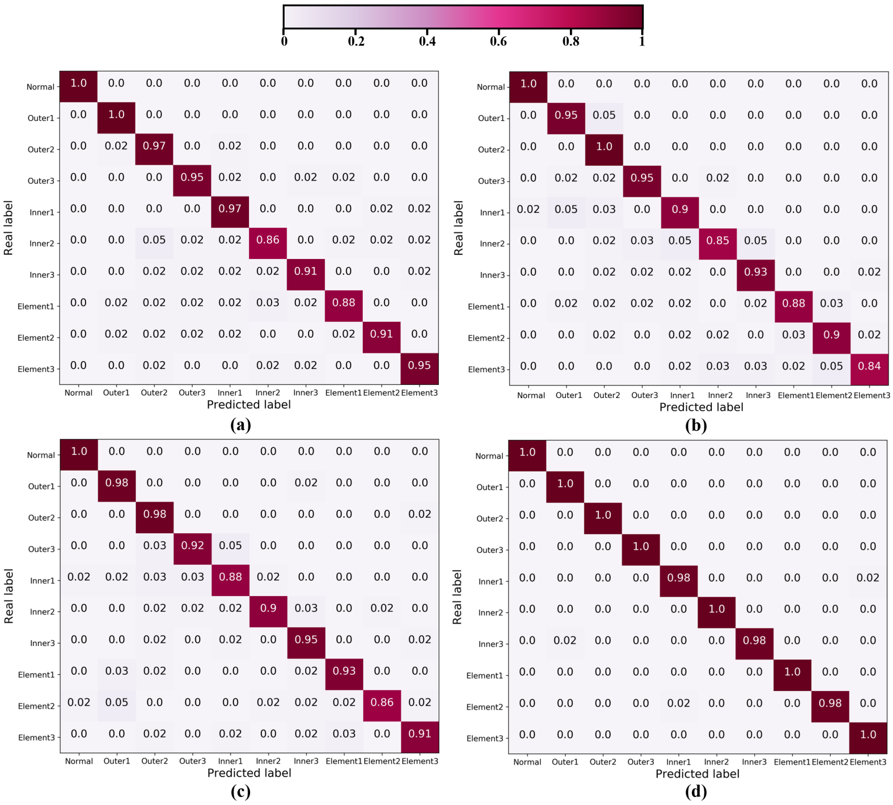
DBN	three RBMs
	Structure [1000, 800, 500, 100, 10]
	Softmax
		Learning rate = 0.001
		Dropout = 0.5
		Batch size = 32
		Optimizer = Adam

DAE	three hidden layers.
	structure [1000, 800, 500, 100, 10]
	Softmax
		Learning rate = 0.003
		Dropout = 0.4
		Batch size = 32
		Optimizer = Adam

All experiments have been repeated in 10 trials, and the results are drawn in Table 5. From the table, one can see that the proposed method has higher accuracy and better stability than other methods. Specifically, the fault type is determined with the multiple RoIs in the proposed method, thereby leading to a better diagnosis performance. The Figure 10 shows the confusion matrix of all methods in this paper. The x-axis represents the predicted labels for different fault working conditions, and the y-axis represents the real labels, it can be noted that all methods achieve better diagnostic accuracy for normal and outer race faults. The misclassification of other comparison methods is mainly concentrated in the inner race and ball faults. However, the proposed method can achieve the classification of all fault types with a high accuracy (mostly 100%).

Table 5. Test performance of different methods.

Approach		Diagnosis Accuracy	Standard Deviation
CNN		93.78%	0.63
DBN		91.89%	0.57
DAE		94.17%	0.55
Proposed method	ResNet50 with attention	99.74%	0.36
	ResNet101 with attention	99.64%	0.34



(a)

(b)

(c)

(d)

Figure 10. Confusion matrix of the fault detection result. (**a**) CNN; (**b**) DBN; (**c**) DAE; (**d**) Proposed method (ResNet50 with attention).

To clearly demonstrate the feature extraction ability of the proposed method, some examples of the test results are shown in Figure 11. It can be seen from the figure that this method can accurately mark many different types of RoIs. These RoIs not only extract the frequency-dependent features in the time-frequency diagram, but also present the temporal information of different fault features via the position coordinates. This allows the subsequent classification model to identify the fault type of the original time-frequency diagram accurately and quickly.

Based on these results, one can conclude that: (1) The proposed method can screen out multiple RoIs from the original sample to determine the fault category, which can lead to a more accurate diagnosis than other fault diagnosis methods. Moreover, this method can extract more effective features in the context of limited information in practical engineering. (2) Compared with other object detection methods in fault diagnosis, the attention module introduced in this method enhances the grasp of representative features, thereby reducing the occurrence of model misdiagnosis. (3) It is vital for fault diagnosis methods to strike a balance between computation time and diagnostic accuracy. The proposed method requires more training time when training complex models; however, it is believed that this problem can be addressed with the future improvement of hardware.

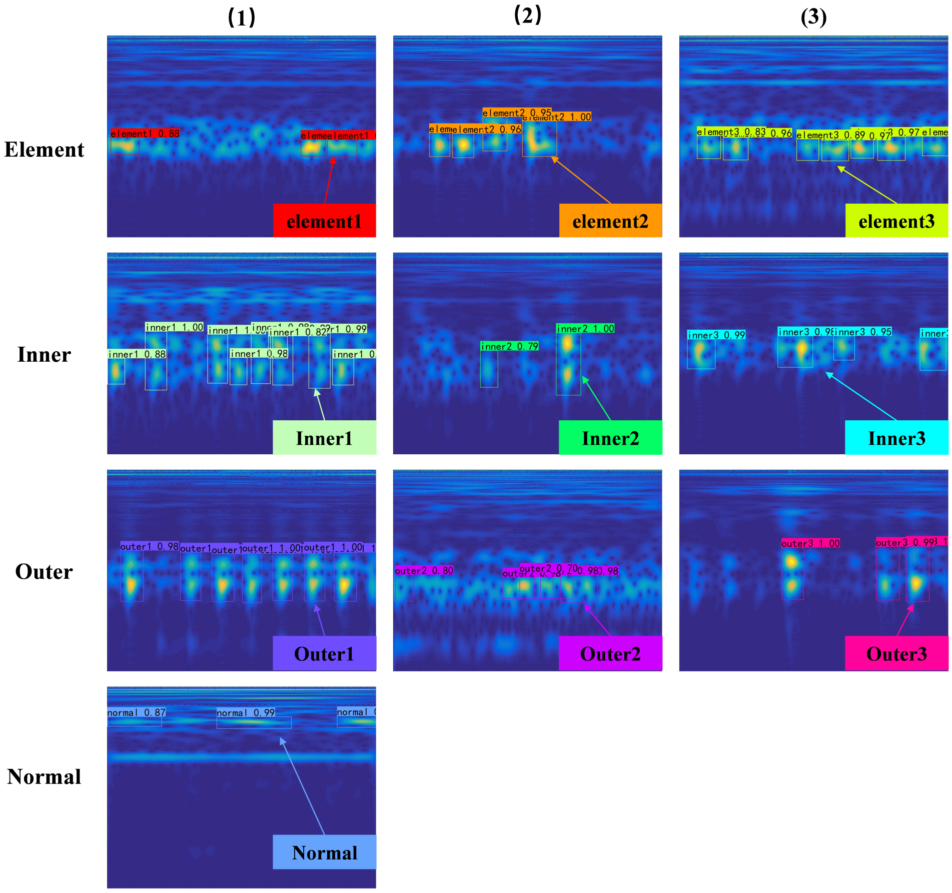


Figure 11. Data samples and their corresponding detection results in artificial damages (ResNet50 with attention).

5.2. *Case 2: Diagnosis for Realistic Damages*

In this section, the dataset comes from the chair of design and drive technology, Paderborn university [36], which is adopted to test the effectiveness of the proposed method. Additionally, rolling bearings with realistic damages that are obtained from an accelerated lifetime test are selected to verify the diagnosis accuracy of this method towards different fault types.

5.2.1. Dataset Description

The tested bearings operate under radial loads applied by spring screws, and this radial force is commonly greater than in usual bearing applications. It aims at accelerating the formation of fatigue damages, but not exceeding the static load capacity of the bearing. Figure 12 presents the structure of the test bench. The bearing in the experiment contains the following two operating states: normal and realistic damages. These realistic damages are caused by fatigue pitting and plastic deformation. Realistic bearing damages in the shape of plastic or pitting deformation are shown in Figure 13. More details about the working conditions are provided in Table 6. The dataset is processed in the same way as in Section 5.1. For further details, refer to Table 7.

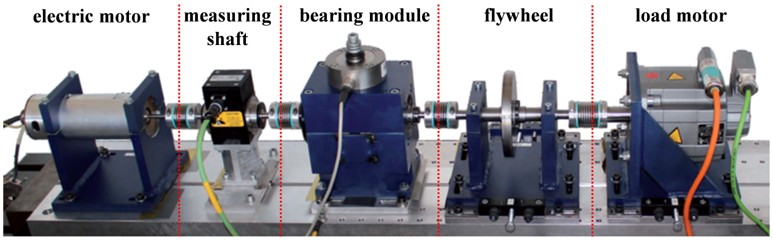


Figure 12. Modular test rig of Paderborn.

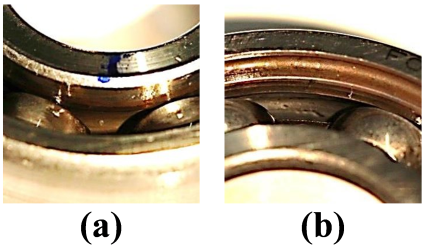


Figure 13. The realistic bearing damages. (**a**) Indentation at the outer ring. (**b**) pitting at the inner ring.

Table 6. Test bench parameters.

Parameter Description	Value
Bearing type	6203
Rotational speed	1500 rpm
Load torque	0.7 Nm
Radial force	1000 N
Sampling time	4 s
Sampling rate	64 kHz

Table 7. The rolling bearing operation conditions of all kinds. (* A: single point without repetitive damage. B: single point damage with random distribution).

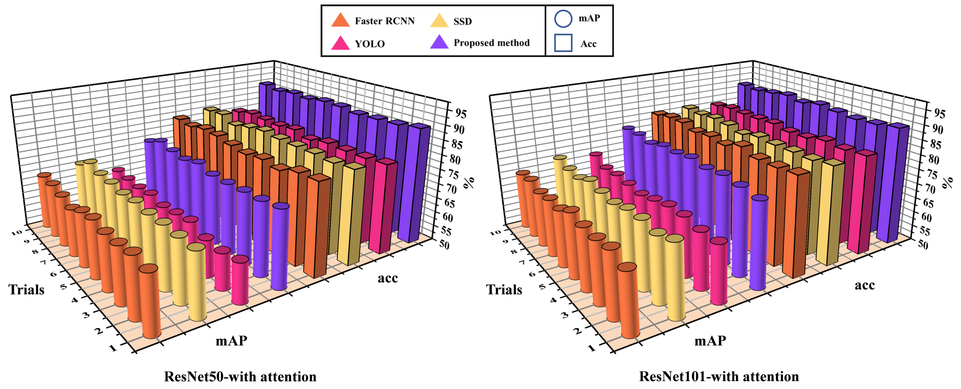
Data Index	Damage Mode	Component	Fault Diameter (mm)	Characteristic of Damage *	Size of Training Testing Samples
Inner1	Fatigue: Pitting	Inner race	6	A	180/50
Inner2	Fatigue: Pitting	Inner race	1	B	180/50
Inner3	Fatigue: Pitting	Inner race	2.5	A	180/50
Outer1	Plastic deform: Indentations	Outer race	<1	A	180/50
Outer2	Fatigue. Pitting	Outer race	2&3	B	180/50
Normal	—	—	—	—	180/50

5.2.2. Result Analysis

The comparative experiments are set up in the same way as in Section 5.1; ten experiments are performed to analyze the robustness of the proposed method in practical applications. The experimental results are illustrated in Table 8, and Figure 14 shows their detailed results in each trial. It can be observed that mAP and diagnostic performance decreased compared to the previous section. This is because bearing damage is usually a gradual process in practical industrial applications, and the vibration signal characteristics in different periods of the fault are quite different. Further, it can also be observed that the proposed method achieves the highest mAP (75.09%, 79.34%) in backbone networks with different depths, and the highest diagnostic accuracy (89.01%, 89.31%) in actual fault diagnosis. These results prove that the proposed method has better stability and a more efficient realization of fault diagnosis ability.

Table 8. Performance of comparison methods based on object detection.

Approach	Backbone	mAP	Diagnosis Accuracy	Standard Deviation
Standard Faster RCNN	ResNet50	68.77%	80.51%	0.79
	ResNet101	70.26%	83.19%	0.61
SSD	ResNet50	71.65%	81.99%	0.68
	ResNet101	72.18%	82.70%	0.50
Yolo	ResNet50	64.05%	80.36%	0.49
	ResNet101	69.44%	82.94%	0.43
Proposed method	ResNet50 with attention	75.09%	89.01%	0.44
	ResNet101 with attention	79.34%	89.31%	0.51



ResNet50-with attention

ResNet101-with attention

Figure 14. Comparison results of the 10 trials.

The comparison results between the proposed method and other methods are analyzed in Table 9. Although the accuracy of the proposed method in diagnosing artificial faults decreases, it still has advantages compared with other methods. The average accuracy can reach 89.01% and 89.31%, which is more stable than the DAE algorithm, but weaker than CNN and DBN.

Table 9. Test performance of different methods.

Approach		Diagnosis Accuracy	Standard Deviation
CNN		85.98%	0.20
DBN		82.38%	0.33
DAE		82.69%	0.78
Proposed method	ResNet50 with attention	89.01%	0.44
	ResNet101 with attention	89.31%	0.51

Table 10 shows the diagnostic accuracy of the proposed method, standard Faster RCNN and CNN for different fault types. It is apparent from the table that the misclassification mainly exists in the outer race faults. Surprisingly, the diagnostic accuracy of the proposed method is greater than other methods in each fault category, and it has better robustness. Figure 15 demonstrates the diagnosis results of the proposed method. As shown, the proposed method not only extracts RoIs that can characterize the fault, but also extracts some misidentified RoIs, which has a slight effect on the results.

Table 10. Diagnosis accuracy of each condition.

Approach	Inner1	Inner2	Inner3	Outer1	Outer2	Normal
CNN	84.68%	82.53%	88.47%	88.22%	76.35%	90.87%
Standard Faster RCNN	81.41%	82.66%	80.71%	77.47%	73.41%	88.63%
Proposed method	90.98%	90.11%	90.25%	89.67%	79.55%	93.49%

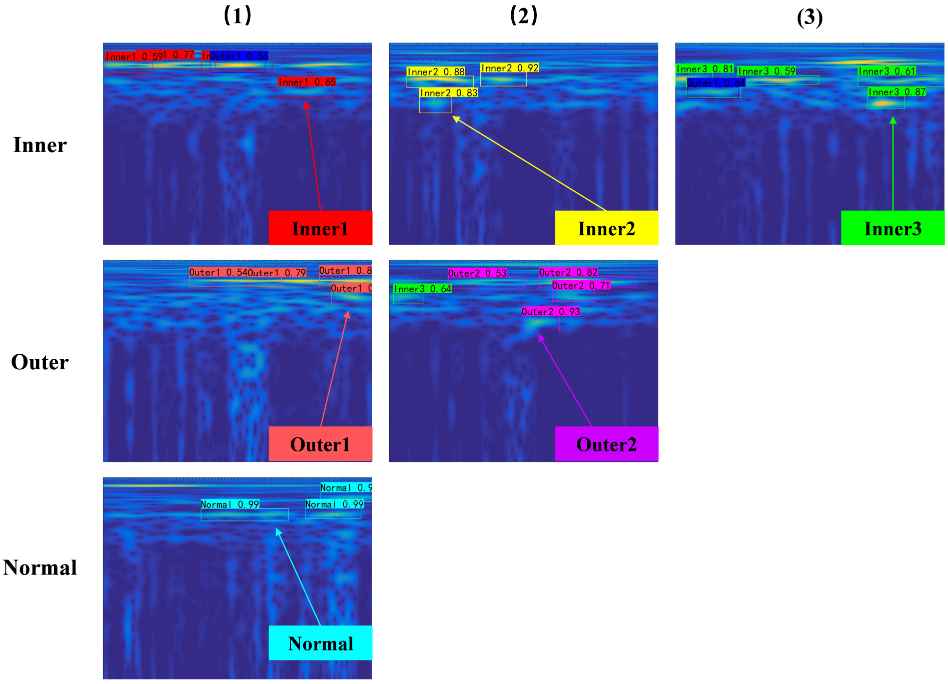


Figure 15. Data samples and their corresponding detection results in realistic damages (ResNet50 with attention).

5.2.3. The Contribution of Attention

For qualitative analysis, the part that the attention module introduced in this paper plays in fault feature extraction is investigated. Here, gradient-weighted class activation mapping (Grad-CAM) is applied to all methods in this paper. Grad-CAM is an excellent visualization model, which utilizes gradients to generate a rough localization map to describe the importance of the spatial locations in convolutional layers. By observing the important regions identified by the feature extraction, we try to explain the process of network learning features.

In Figure 16, the visualization results of different categories of original images, ResNet50, and ResNet50 with attention are demonstrated. The color distribution represents the distribution of gradients, and the softmax scores for the condition class are also presented in the figure. Compared with ResNet50, the proposed method more accurately covers the part of the input image that contains the fault frequency, which enables the subsequent model to utilize information in these areas. This method can reduce the learning of irrelevant components and improve the subsequent model accuracy.

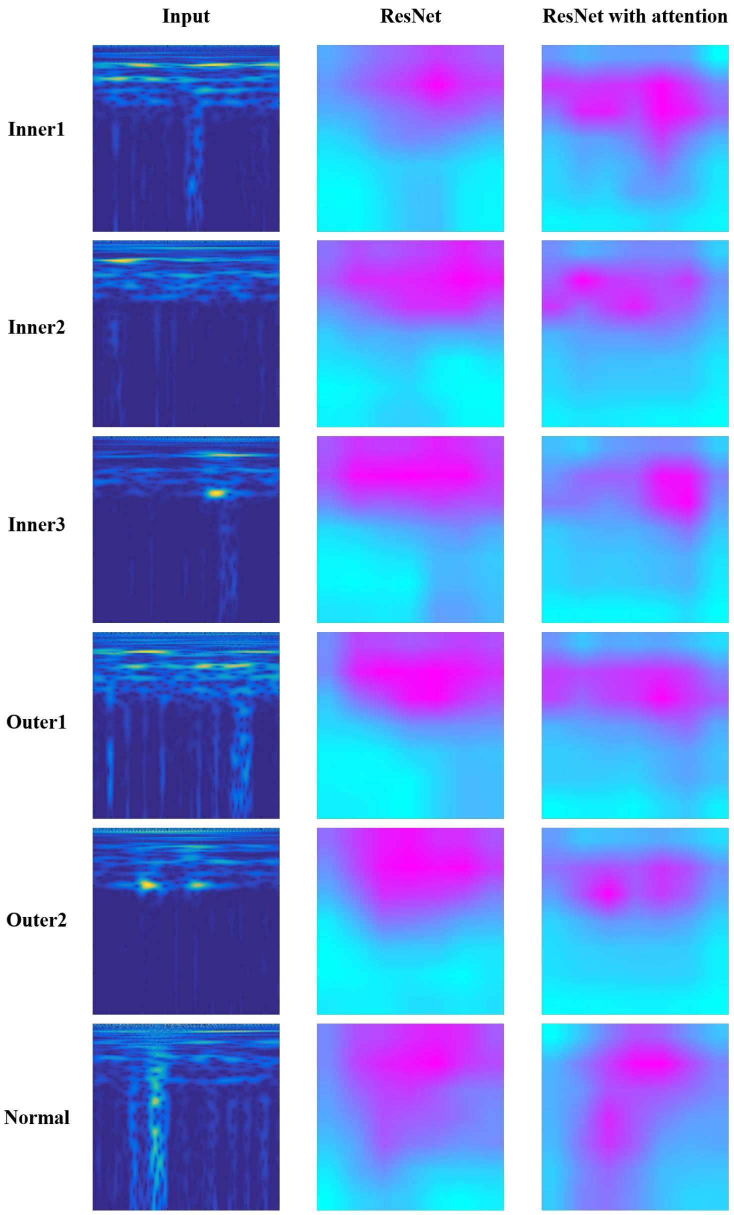


Figure 16. Grad-CAM visualization results of proposed method (ResNet50).

6. Concluding Remarks

The main contribution of this paper is to develop a fault diagnosis method based on object detection. It is inspired by the massive success of object detection architecture in the detection task. In this regard, a fault diagnosis method of rolling bearing based on TF-RCNN is proposed. The CWT is employed to process the original vibration signals, and then obtain the corresponding TFR and label them. Finally, these TFRs are fed into the model for feature extraction, RoIs proposal and fault identification.

The proposed method in this paper has the following characteristics: (1) TF-RCNN removes the one-to-one relationship between samples and categories in traditional fault diagnosis methods, and adopts multiple RoIs to identify the fault type of samples. (2) The attention module is introduced to improve the feature extraction ability, thus meaning that the model can accurately focus on the representative feature region of the input image.

Furthermore, artificial damages and realistic bearing damages are adopted to verify the effectiveness of the proposed method. Meanwhile, the results reveal that the present method has the following advantages compared with other advanced methods: (1) Compared with other object detection methods, the proposed method can pay more attention to the effective information of the original sample, thereby improving the diagnosis accuracy and efficiency. (2) Compared with other outstanding diagnosis methods based on deep learning, the proposed method achieves more accurate fault diagnosis. (3) The proposed method has excellent diagnostic performance for both artificial damages and complex realistic damages.

Considering the better application performance of the proposed method, our future interest is to realize the compound fault diagnosis of rolling bearings and to address the problems of accuracy in compound fault diagnosis.

Author Contributions: Conceptualization, J.T.; methodology, J.T. and J.W.; validation, B.H.; data curation, J.T.; writing—original draft preparation, J.T.; writing—review and editing, J.W. and B.H.; supervision, J.Q.; funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research benefited from a grant from the National Natural Science Foundation of China (No. 52075435), the Natural Science Basic Research Program Key Project of Shaanxi Province (No. 2022JZ-30), the Natural Science Special Project of Education Department of Shaanxi Provincial Government (No. 21JK0805), and the Doctoral Dissertation Innovation Fund of Xi'an University of Technology (Grant No. 252072105).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587–106626. [CrossRef]
2. Sharma, A.; Upadhyay, N.; Kankar, P.K.; Amarnath, M. Nonlinear dynamic investigations on rolling element bearings: A review. *Adv. Mech. Eng.* **2018**, *10*, 1–15. [CrossRef]
3. Zhao, H.M.; Liu, H.D.; Jin, Y.; Dang, X.J.; Deng, W. Feature extraction for data-driven remaining useful life prediction of rolling bearings. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [CrossRef]
4. Tang, J.; Wu, J.; Hu, B.; Liu, J. Towards a fault diagnosis method for rolling bearing with bi-directional deep belief network. *Appl. Acoust.* **2022**, *192*, 108727. [CrossRef]
5. Liang, H.P.; Zhao, X.Q. Rolling bearing fault diagnosis based on one-dimensional dilated convolution network with residual connection. *IEEE Access* **2021**, *9*, 31078–31091. [CrossRef]
6. Yi, C.; Lv, Y.; Dang, Z.; Xiao, H.; Yu, X. Quaternion singular spectrum analysis using convex optimization and its application to fault diagnosis of rolling bearing. *Measurement* **2017**, *103*, 321–332. [CrossRef]
7. Lai, Z.H.; Wang, S.B.; Zhang, G.Q.; Zhang, C.L.; Zhang, J.W. Rolling bearing fault diagnosis based on adaptive multiparameter-adjusting bistable stochastic resonance. *Shock Vib.* **2020**, *2020*, 6096024. [CrossRef]
8. Wang, G.; Zhao, B.; Xiang, L.; Li, W.; Zhu, C. Information interval spectrum: A novel methodology for rolling-element bearing diagnosis. *Measurement* **2021**, *183*, 109899–109916. [CrossRef]
9. He, Q.; Song, H.; Ding, X. Sparse signal reconstruction based on time-frequency manifold for rolling element bearing fault signature enhancement. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 482–491. [CrossRef]
10. de Moura, E.P.; Souto, C.R.; Silva, A.A.; Irmão, M.A.S. Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by rs and df analyses. *Mech. Syst. Signal Process.* **2011**, *25*, 1765–1772. [CrossRef]
11. Ai, Y.-T.; Guan, J.-Y.; Fei, C.-W.; Tian, J.; Zhang, F.-L. Fusion information entropy method of rolling bearing fault diagnosis based on n-dimensional characteristic parameter distance. *Mech. Syst. Signal Process.* **2017**, *88*, 123–136. [CrossRef]
12. Moura, M.D.; Zio, E.; Lins, I.D.; Droguett, E. Failure and reliability prediction by support vector machines regression of time series data. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 1527–1534. [CrossRef]
13. Chen, F.; Cheng, M.; Tang, B.; Chen, B.; Xiao, W. Pattern recognition of a sensitive feature set based on the orthogonal neighborhood preserving embedding and adaboost svm algorithm for rolling bearing early fault diagnosis. *Meas. Sci. Technol.* **2020**, *31*, 105007. [CrossRef]
14. Zhang, Q.; Li, H.; Zhang, X.; Wang, H. Optimal multi-kernel local fisher discriminant analysis for feature dimensionality reduction and fault diagnosis. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2021**, *235*, 1041–1056. [CrossRef]
15. Wan, L.; Gong, K.; Zhang, G.; Yuan, X.; Li, C.; Deng, X. An efficient rolling bearing fault diagnosis method based on spark and improved random forest algorithm. *IEEE Access* **2021**, *9*, 37866–37882. [CrossRef]

16. Jia, F.; Lei, Y.G.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72–73*, 303–315. [CrossRef]
17. Wen, L.; Li, X.; Gao, L. A new two-level hierarchical diagnosis network based on convolutional neural network. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 330–338. [CrossRef]
18. Li, Z.; Li, Z.; Li, Y.; Tao, J.; Mao, Q.; Zhang, X. An intelligent diagnosis method for machine fault based on federated learning. *Appl. Sci.* **2021**, *11*, 12117. [CrossRef]
19. Zhu, J.; Hu, T.Z.; Jiang, B.; Yang, X. Intelligent bearing fault diagnosis using pca-dbn framework. *Neural Comput. Appl.* **2020**, *32*, 10773–10781. [CrossRef]
20. Shao, H.; Jiang, H.; Wang, F.; Wang, Y. Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet. *ISA Trans.* **2017**, *69*, 187–201. [CrossRef]
21. Guo, J.; Zheng, P. A method of rolling bearing fault diagnose based on double sparse dictionary and deep belief network. *IEEE Access* **2020**, *8*, 116239–116253. [CrossRef]
22. Song, X.; Zhu, D.; Liang, P.; An, L. A new bearing fault diagnosis method using elastic net transfer learning and lstm. *J. Intell. Fuzzy Syst.* **2021**, *40*, 12361–12369. [CrossRef]
23. Liu, W.; Guo, P.; Ye, L. A low-delay lightweight recurrent neural network (llrnn) for rotating machinery fault diagnosis. *Sensors* **2019**, *19*, 3109. [CrossRef] [PubMed]
24. Xu, J.; Zhou, L.; Zhao, W.; Fan, Y.; Ding, X.; Yuan, X. Zero-shot learning for compound fault diagnosis of bearings. *Expert Syst. Appl.* **2022**, *190*, 16197. [CrossRef]
25. Shi, Z.; Chen, J.; Zi, Y.; Zhou, Z. A novel multitask adversarial network via redundant lifting for multicomponent intelligent fault detection under sharp speed variation. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [CrossRef]
26. Ma, P.; Zhang, H.; Fan, W.; Wang, C.; Wen, G.; Zhang, X. A novel bearing fault diagnosis method based on 2d image representation and transfer learning-convolutional neural network. *Meas. Sci. Technol.* **2019**, *30*, 055402. [CrossRef]
27. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
28. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 142–158. [CrossRef] [PubMed]
29. Konar, P.; Chattopadhyay, P. Bearing fault detection of induction motor using wavelet and support vector machines (svms). *Appl. Soft Comput.* **2011**, *11*, 4203–4211. [CrossRef]
30. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [CrossRef]
31. Kankar, P.K.; Sharma, S.C.; Harsha, S.P. Fault diagnosis of ball bearings using continuous wavelet transform. *Appl. Soft Comput.* **2011**, *11*, 2300–2312. [CrossRef]
32. Su, W.; Wang, F.; Zhu, H.; Zhang, Z.; Guo, Z. Rolling element bearing faults diagnosis based on optimal morlet wavelet filter and autocorrelation enhancement. *Mech. Syst. Signal Process.* **2010**, *24*, 1458–1472. [CrossRef]
33. Tang, B.; Liu, W.; Song, T. Wind turbine fault diagnosis based on morlet wavelet transformation and wigner-ville distribution. *Renew. Energy* **2010**, *35*, 2862–2866. [CrossRef]
34. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In *Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 3–19.
35. Case Western Reserve University. Available online: http://www.eecs.case.edu/laboratory/bearing/download.htm (accessed on 1 July 2022).
36. Lessmeier, C.; Kimotho, J.K.; Zimmer, D.; Sextro, W. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In Proceedings of the European Conference of the Prognostics and health Management, Bilbao, Spain, 5–8 July 2016; pp. 5–8.