Application of Local Histogram Clipping Equalization Image Enhancement in Bearing Fault Diagnosis

Zhiqiang Zeng1, Rui Zhang1, Wenan Cai2 and Yanfeng Li1

1School of mechanical engineering, North University Of China, Taiyuan, CO 030051 China
2School of Mechanical Engineering, Jinzhong University, Taiyuan, 030619, China

Corresponding author: Yanfeng Li (e-mail: yanfengli0002@163.com).

This work was partially supported by Natural Science Foundation of Shanxi Province (201801D221339)

ABSTRACT Aiming at the problem that the time-frequency image of bearing fault characteristics is relatively weak and difficult to identify. This paper presents a time-frequency analysis method of Local maximum synchrosqueezing transform based on image enhancement. Firstly, the instantaneous frequency of the collected vibration signal is obtained through local maximum synchrosqueezing transformation. Secondly, a local histogram cropping equalization image enhancement algorithm is proposed, which is used to obtain time-frequency images with clearer textures. Then, in order to extract fault features from the enhanced instantaneous frequency (IF) image, A new neural network is proposed. The network consists of Multi-size convolution kernel module, Dual-channel pooling layer and Cross Stage Partial Network (MDCNet). Finally, the fault signal was collected on the bearing fault test bench for prediction, and the accuracy rate reached 99.7%. And compared with AlexNet, VGG-16, Resnet and other methods. The results show that the method can meet the needs of actual engineering.

INDEX TERMS fault diagnosis, image enhancement, the instantaneous frequency

I. INTRODUCTION

In the process of modern industrial manufacturing, the bearing is a key component of mechanical equipment. The stability of mechanical equipment determines the reliability of products[1]. Whether micro faults can be diagnosed in time is the key to ensure safe operation. However, bearings often fail at low speed and heavy load, and their closed operating environment and environmental noise make it difficult to find faults in time. Vibration signal analysis is a common fault diagnosis method without disassembling machine parts. Data-driven bearing fault diagnosis has become the mainstream of research and attracted the wide attention of the academic community[2-5]. Bearing fault identification as an important part of fault diagnosis, is used to determine the necessity of replacing the bearing without destroying the equipment. It provides good reference information for equipment reliability. This is of great significance for equipment maintenance.

Extracting fault characteristics from time domain signals is a key method for bearing fault diagnosis. Time-frequency domain analysis can reflect the time-domain information of signal and frequency-domain information in a coordinate system. That is to say, it can reflect the law that the signal changes with time in frequency. Signals gather near their instantaneous frequency (IF) and different fault types are represented by different IF images. Therefore, to accurately identify fault types, clear time-frequency images are required.

Nowadays, many researches use time-frequency analysis method to transform vibration signals into time-frequency images and then classify bearing faults. Classic short-time Fourier transform (STFT) causes the problem of low time-frequency resolution. Time-frequency reassigned (RM) improves the readability of signals by reclassifying them into IF trajectories in the direction of time and frequency by collecting the energy of the signal. Synchrosqueezing transformation (SST) compresses the TF coefficients near mid-frequency in the direction of frequency, greatly improving the time-frequency (TF) resolution[6]. SST method has been applied in many other fields, such as earthquake prediction [7], gravitational wave, bearing intelligent fault diagnosis [8], etc. However, there are also some shortcomings in SST method. When processing FM signal with SST method, TF generated by SST means that energy will diverge seriously, resulting in low resolution.

Many studies have shown that time-frequency energy divergence can lead to inaccurate identification of bearing faults and different improvement methods have been proposed. The second-order SST [9, 10] and the high-order SST [11, 12] greatly improve the time-frequency aggregation,
but they will double the amount of calculation and are susceptible to noise interference. Yu [13] proposed Synchronous Extraction Transformation (SET), which significantly improved the resolution of TF results. However, this method also has certain shortcomings. This method cannot achieve perfect signal reconstruction when processing FM signal. Zhu et al. [14] proposed synchronous extraction of chirplet transform to improve the accuracy of intermediate frequency estimation. The purpose of the Synchrosqueezing matching pursuit algorithm is to enhance the energy aggregation in the TF plane [15]. This method compresses the TF distribution to the vicinity of the center time of the selected wavelet through a two-dimensional Gaussian function, which can improve the time-frequency aggregation.

Yu et al. [16] proposed Multiple Synchrosqueezing Transform (MSST), which realizes synchrosqueezing through frequency reassignment. This method redistributes TF coefficients through multiple iterative operations to suppress energy dispersion, so as to achieve more time-frequency aggregation on the TF plane. Fourier et al. [17] and He et al. [18] respectively propose time-frequency analysis method for TF coefficient compression based on time reallocation. Liang et al. [19] proposed a Kaiser window S transform (KST), which further enhances time-frequency aggregation by adaptively adjusting the Kaiser window.

Deep learning has been widely used in fault diagnosis. Convolutional neural network (CNN) is a popular method in bearing fault diagnosis. In order to extract fault information from time-domain signals, Sun et al. [20] uses multi-synchronous compression transformation to obtain time-frequency images and then extract features. Finally, the trained linear support vector machine is used to classify and diagnose faults of test samples. Xin Y et al. [21] calculates time-frequency characteristics by STFT and takes pseudocolor mapping as a new recognition object. A sparse self-coding and linear decoding method is proposed to extract these time-frequency characteristics. Liu Q et al. [22] designed a LeNet-5 network to realize 10 classifications of bearings through synchrosqueezing wavelet transform (CWT). Wen [23] transformed the time-domain signal into two-dimensional gray image, and then realized four fault classification of rolling bearing using LeNet-5. G. Sun et al. [24] obtains the time-frequency image with high resolution through the Second-order Time-Reassigned Multi-Synchrosqueezing Transform (STMSST) and then realizes the bearing classification by combining the uniform small batch training method with CNN. The above method is convenient and has been proved to be effective, but there are still some shortcomings: (1) In IF images, time-frequency ridges may be ignored because they are small (2) With the increase of network depth in fault diagnosis process, gradient disappearance and over-fitting problems occur, resulting in low accuracy and weak generalization ability.

In this paper, combining local maximum synchrosqueezing transform (LMSST) and local histogram clipping equalization image enhancement algorithm, the IF image obtained by the LMSST is image enhanced, and then the image closing operation, brightness adjustment and other operations are performed to obtain clarity Time-frequency image. The accuracy of classification depends on the quality of the transformed image, as well as the feature extraction algorithm and classifier. This paper proposes MDCNet, which can accurately identify the fault state of the bearing. The main contributions of this paper are as follows: (1) A local histogram cropping and equalization image enhancement algorithm is proposed. By extracting the IF image of the vibration signal, the original signal is converted into an image, and the image is enhanced by the local histogram cropping equalization image enhancement algorithm. Get the IF image with clear texture. (2) Propose an MDCNet model for fault diagnosis, through multi-size convolution kernel module, cross-stage partial modules, improve classification accuracy, and reduce memory consumption; (3) Build an experimental platform to test bearing faults. The fault signal was collected on the stage for prediction, and the accuracy rate reached 99.7%.

The rest of this article is organized as follows. The section II is the Local Maximum Synchrosqueezing Transformation based on image enhancement. The section III gives the proposed MDCNet model. The section IV presents the test results of the method on two different data sets. Section V shows the conclusion of the proposed method.

II. Local Maximum Synchrosqueezing Transformation for Image Enhancement

A. Synchrosqueezing Transform

This paper starts with the framework of STFT [25]. The signal \( s(t) \) expression is

\[
s(t) = \sum_{k=1}^{K} A_k(t) e^{j \phi_k(t)}
\]  

Equation (1) is the exponential form of multi-component signal, in which, \( K \) is the number of signal modes, \( A_k(t) \) is the phase of component signal, \( \phi_k(t) \) represents the amplitude of component signal and \( \phi'_k(t) = \frac{d}{dt} \phi_k(t) \) represents instantaneous frequency (IF).

Equation (2) is the result of Taylor expansion of the above signals at \( u = t \). Ignoring the higher-order term, \( A_k(u) = A_k(t) \), \( \phi_k(u) = \phi_k(t) + \phi'_k(u-t) \), and the signal can be expressed as:

\[
s(u) = \sum_{k=1}^{K} A_k(t) e^{j \left[ \phi_k(t) + \phi'_k(u-t) \right]}
\]  

In order to realize the hypothesis of stationary signal, the STFT divides the signal into small intervals according to a fixed length. Then, the Fourier transform of the signal in each interval is allocated to the central time sample. The signal \( s(t) \in L^2(\mathbb{R}) \) uses a real window function at \( t \in [-\Delta t, \Delta t] \), and its STFT is expressed as:
The SST method with signal phase information is reassignment method (RM) method reassigning of time-frequency samples to the ideal IF trajectory. The resolution of the TF plane is improved by accurately reallocating samples in both time and frequency directions.

The SST method with signal phase information is introduced to overcome the deficiency of redistribution method in reconstructing the original signal. The SST expression is

$$\text{SST}(t, \eta) = \int_{-\infty}^{\infty} S^{\alpha \alpha}(t, \omega) \delta(\omega - \phi_1(t)) d\omega$$

(6)

Among them, $\phi_1(t, \omega)$ stands for IF, which is equivalent to $\hat{\omega}(t, \omega)$, which is obtained by entering the formula:

$$\frac{\partial}{\partial t} S^{\alpha \alpha}(t, \omega) = \frac{\partial}{\partial t} \left\{ \sum_{k=1}^{K} A_k(t) e^{i \phi_k(t) \omega} \right\} = i \phi_k(t) S^{\alpha \alpha}(t, \omega)$$

(7)

For each $(t, \omega)$, where $S^{\alpha \alpha}(t, \omega) \neq 0$, IF is expressed as

$$\phi_1(t) = \omega_k(t, \omega)$$

$$\frac{\partial}{\partial \omega} S^{\alpha \alpha}(t, \omega) = \frac{i \omega_k(t)}{s_k(t)} S^{\alpha \alpha}(t, \omega)$$

$$\frac{\partial}{\partial \omega} S^{\alpha \alpha}(t, \omega) = \frac{i \omega_k(t)}{s_k(t)} S^{\alpha \alpha}(t, \omega)$$

(8)

The actual calculation uses the real component $\omega_k(t, \omega)$, so IF [23] is

$$\omega_k(t, \omega) = \omega - \operatorname{Im} \left[ \frac{S^{\alpha \alpha}(t, \omega)}{S^{\alpha \alpha}(t, \omega)} \right]$$

(9)

The component of the original signal in expression (1) is defined as

$$s_k(t) = \frac{1}{2 \pi \delta(0)} \int_{-\infty}^{\infty} \sin \left( \frac{\omega - \omega_k(t, \omega)}{\omega - \omega_k(t, \omega)} \right)$$

(10)

Where $\delta$ is the SST reconstruction bandwidth. Because SST has good performance in signal reconstruction, it can be used in modal decomposition, signal noise reduction and other applications.

### B. Local maximum synchrosqueezing transform

The SST method has superior performance in signal reconstruction. RM method can maintain the time-frequency aggregation of signals. In order to take into account these two advantages at the same time, the coefficients of STFT method distributed on the time-frequency TF plane need to be reassigned to each if along the frequency direction. This is the advantage of ideal time-frequency analysis. It has high-resolution representation on TF plane and the ability of complete signal reconstruction. The LMSST method uses RM to gather the scattered energy and redistribute the coefficients to the real if trajectory [26].

Assuming that two arbitrary modes of the vibration signal are far apart in the frequency domain, for each $t \in \mathbb{R}$ and $k \in \{1, \ldots, K\}$, there is $\phi_{k+1}(t) - \phi_k(t) \geq 4 \Delta \omega$, and the redistribution is as follows:

$$\omega_k(t, \omega) = \begin{cases} \arg \max_{\omega} \left| S^{\alpha \alpha}(t, \omega) \right|, & \omega \in [-\Delta \omega, \omega + \Delta \omega] \\ 0, & \text{otherwise} \end{cases}$$

(11)

The reassignment operator defined by expression (11) reallocates the coefficients in the TF plane to new positions, where $\left| S^{\alpha \alpha}(t, \omega) \right|$ and $\omega \in [-\Delta \omega, \omega + \Delta \omega]$ have their maximum values. Due to $\hat{\omega}(\omega) \leq \hat{\omega}(0)$, the reassignment operator in expression (11) is defined as

$$\phi_k(t, \omega) = \begin{cases} \phi_{k+1}(t), & \text{if } \omega \in [-\Delta \omega, \omega + \Delta \omega] \\ 0, & \text{otherwise} \end{cases}$$

(12)

Therefore, the expression of LMSST is:

$$\text{LMSST}(t, v) = \int_{-\infty}^{\infty} S^{\alpha \alpha}(t, \omega) \delta(v - \omega_k(t, \omega)) d\omega$$

(13)
C. Local histogram clipping equalization image enhancement algorithm based on LMSST

For some weak fault signals, the time-frequency images through LMSST still contain little information and are difficult to be identified. Therefore, an improved local histogram clipping equalization (LHCE) image enhancement algorithm is proposed. This method is based on Contrast Limited Adaptive Histogram Equalization (CLAHE), which replaces histogram equalization with automatic contrast and automatic color levels. The main steps of LHCE algorithm are shown in Figure 1.

(1) The image is divided into sub-images $X_i$ by selecting the appropriate number of horizontal $i$ and vertical grids $j$.

(2) Histogram information of each sub-image R, G, B channel and overall image brightness are calculated and represented as Hist (B), Hist (G), Hist (R), Histgram (L), respectively.

(3) The gray histogram of the sub-image and the histogram information of the whole image are fused. Adaptation is the fusing factor, and its expression is as follows:

\[
\text{HistB'}[X_i] = (\text{HistB}[X_i] \ast \text{Adaptation} + (100 - \text{Adaptation}) \ast \text{HistgramB}[X_i]) / 100
\]

(4) The fusion results are fused with Histgram (L) again, and the fusion process is as follows:

\[
\text{HistB''}[X_i] = (\text{HistB'}[X_i] \ast \text{Correction} + (100 - \text{Correction}) \ast \text{HistgramB'}[X_i]) / 100
\]

(5) Cut the histogram according to CALHE, and then equalize the cut histogram to get a mapping table for each block.

(6) In order to avoid large singularities or noises in the data of the new mapping table, the new mapping table is smoothed by Gauss and a smoother mapping table is obtained.

(7) Finally, bilinear interpolation is used to interpolate the mapping table between each sub-block to obtain new pixel values.

After contrast enhancement, the image should be post-processed: Since there are discontinuous line segments in the IF of the time-frequency image, the closed operation is performed on the reconstructed image. Connect the broken time-frequency ridges inside the image. Compared with histogram equalization algorithm, this method can effectively avoid the problem of excessive brightness after image enhancement. The enhanced image quality is higher and more in line with the observation habit of human eyes.

III. MDCNet

In order to improve the ability of neural network to identify and extract weak faults, this paper proposes an MDC-Net network. The network consists of multi-scale convolution kernel module, dual-channel pooling layer and Cross Stage Partial network. Figure 2 shows the schematic image of the MDCNet network proposed in this paper.

Firstly, the use of multi-scale convolution kernel module in the training process is conducive to extract more information. Secondly, the two pooling methods are combined to form a double channel pool layer. Finally, the use of cross stage partial (CSP) network block can greatly reduce the storage cost, and the use of this structure can enhance the learning ability of neural network and improve the training accuracy of network. Here are more details about the network.

A. Multi-dimensional convolution kernel module

Szegedy et al. [27] proposed an Inception v1 structure, which stacks the convolution and pooling operations commonly used in CNN together, and finally merges the results of each path operation on the channel by the Concat function. The 5×5 size convolution kernel can cover most of the input of the receiver layer, and the structure performs a pooling operation along with convolution, which helps to reduce the network space size and overfitting.

In this paper, a multi-scale convolution kernel module is proposed, in which the commonly used convolutional kernel (1×1, 3×3, 5×5) and pooling operation (3×3) are stacked together, and uses the Add function to connect the outputs of each channel. The amount of feature information under each dimension described by the Add function will increase. It does not increase the dimension describing the characteristics of the image itself. This operation is conducive to improving the classification accuracy of the image.

![FIGURE 1. flowchart of LHCE method](image-url)
**B. Dual-channel pooling layer**

There are two types of pooling layers in the residual network, the maximum pooling layer and the average pooling layer. The purpose of maximum pooling is to take the maximum activated in the receptive field as the final pooling output. Average pooling is achieved by using the average activated in the receptive field as the final pooling output. In order to combine the advantages of these two methods, a two-channel pooling layer structure is proposed in this paper. The output characteristics of the previous step are divided into two parts, which are pooled by maximum and average respectively. Finally, the output results of the two channels are combined into the next CPS block by the add function.

**C. Cross-stage partial network**

In this paper, the Residual block in the original network was improved by CSP block instead of using the Residual block in ResNet18 network as the feature extraction network. The improved CPSNet basic module is shown in Fig. 3:

The main idea of CSPNet module is to divide feature map into two parts. In order to extract feature information, one part is convoluted while the other part and the results of the previous part are spliced directly. This operation can greatly reduce the memory consumption of the computer and improve the training accuracy[28].

The output of k layer in the neural network can be expressed as:

\[ y = G(x_k) = x_k = G(x_{k-1}, G(x_{k-2}, G(x_{k-3}, \ldots, G(x_2, x_1)) \)  (16)

Where G represents the combination of convolution operation and nonlinear activation operation. Then ResNet can be expressed as:
Among them, \( P \) represents the residual layer, including a combination of convolution operation and nonlinear activation operation. The residual layer can minimize the path length that the gradient flows through, making it more effective back propagation. However, this connection will also add a lot of redundant information. For example, the information of the \( k \)-th layer needs to be passed to the \( k-1, k-2, \ldots, 1 \) layer. The specific operation of the CPS block is as follows. The basic feature layer in the previous stage is divided into two parts \( x_0' \) and \( x_0'' \) through the channel \( x_0 = [x_0', x_0''] \), and the \( x_0' \) part is subjected to operations such as convolution by the CPS block, and then connected with the output to generate the output. The gradients of the CPS block are integrated separately, and the feature maps that do not pass through the CPS block are independent of \( x_0'' \). There will not be the same part on both sides of the gradient information used to update the weight.

The operation flow of CPS block is as follows: Through the strategy of channel division and merge, the multi-stage thought can alleviate the gradient duplication caused by repeated feature map; Convolutional Input Output (CIO) is a method used to evaluate the DRAM IO state. Suppose the feature map size in Residual block is \( m \times n \times q \), the growth rate is \( k \), with a total of \( P \) layers. Then the CIO of the Residual block is \( q \times p + \frac{(p^2 + p) \times k}{2} \), and the CPS block is

\[
x_k = P_1 x_{k-1} + x_{k-1} = P_1 x_{k-1} + P_1 x_{k-2} + \ldots + P_1 x_0 + x_0 \quad (17)
\]
While \( k \) and \( p \) are usually far smaller than \( q \), the CPS block can save about half of the memory.

Combined with MDCNet, convert the original signal into an image, then enhance the IF image, and finally train the image data set. The flow chart is shown in Figure 4.

Step 1: collect the signal of rolling bearing in complex working environment with acceleration sensor.

Step 2: use LMSST method to convert time domain signal into IF image.

Step 3: enhance the image through the LHCE method to obtain a clear IF image.

Step 4: normalize the data set and divide it into training set and test set. Save the MDCNet model parameters after training.

Step 5: input the test set into MDCNet to get the fault diagnosis results.

IV. Experiments and Results

To verify the validity of the proposed method, the method is tested on the bearing data set from Case Western Reserve University (CWRU) and on the bearing failure test bench. The network and comparison methods proposed in this paper are written in Matlab 2020, and run on a computer with cpu i7-11800H, RAM 32.00GB, RTX2060GPU, and the operating system is 64-bit win10. In the experiment, the images enhanced by LHCE and directly generated are compared and analyzed, and the texture features of the images are compared. Finally, the MDCNet is compared with the traditional network method to prove the feasibility of this method.

First, the theoretical performance of the model is analyzed by 3 metrics. (1) Time complexity: refers to the number of floating-point operations that occur during a complete forward propagation of a single sample of input. (2) Spatial complexity: This refers to the total number of memory swaps that occur during one complete forward propagation of a single input sample. (3) Computational intensity of the model: The computational intensity of the model can be obtained by dividing the time complexity by the space complexity, which indicates the number of floating-point operations per Byte of memory used in the computation of this model. The unit is FLOPs/Byte, and it can be seen that the higher the computational intensity of the model, the more efficient its memory usage.

From Table I, it can be seen that the computation of the MDCNet proposed in this paper is 4.6 GFLOPS, and the space complexity is 107 M, which has a small space occupation. It can be calculated that the computational intensity is 42.9. In comparison with other methods, the proposed method in this paper has the highest computational intensity and the most efficient use of memory.

### Table I

| Method   | Time complexity /GFLOPS | Spatial complexity /M | Computational intensity FLOPs/Byte |
|----------|-------------------------|-----------------------|-----------------------------------|
| MDCNet   | 4.6                     | 107                   | 42.9                              |
| AlexNet  | 0.7                     | 295                   | 2.37                              |
| VGG-16   | 16                      | 590                   | 27.1                              |
| Resnet18 | 1.8                     | 72.5                  | 24.8                              |
| Resnet50 | 3.8                     | 211                   | 18.0                              |
| Resnet152| 11                      | 472                   | 23.3                              |
| GoogleNet| 2                       | 57.8                  | 34.6                              |

Case1

The experimental data set is the CWRU bearing data set. Fig. 5 is a picture of the experimental device of the experimental platform. The platform is composed of a motor with a speed of 1797r/min, a torque sensor and an accelerometer. The system contains two test bearings, located at the motor drive end and the motor fan end. The experiment mainly studies the bearing model of 6203-2RS JME SKF.

![FIGURE 5. Diagram of CWRU experimental setup](image)

In this paper, 10 states of bearings were selected: normal state and 9 fault states, which were samples with different fault diameters of the rolling body, inner ring and outer ring at 0.007, 0.014 and 0.021. Each sample speed was 1797rpm/min, and the sampling frequency was 12k. LMSST method was used to transform the signal into IF graph, with each sample image number of 2000 and a total of 20000 images. Table II lists more details about the CWRU data set.

### TABLE II

| Fault location | Fault diameter (inches) | Speed/(rpm) | Sampling frequency | label |
|----------------|-------------------------|-------------|--------------------|-------|
| Normal         | 0.007                   | 1797        | 12K                | N     |
| Ball           | 0.014                   |             |                    | BF_1  |
| fault          | 0.021                   |             |                    | BF_3  |
| Inner          | 0.007                   | 1797        | 12K                | IRF_1 |
| race           | 0.014                   |             |                    | IRF_2 |
| fault          | 0.021                   |             |                    | IRF_3 |
| Outer          | 0.007                   |             |                    | ORF_1 |
| race           | 0.014                   |             |                    | ORF_2 |
| fault          | 0.021                   |             |                    | ORF_3 |
The IF image is shown in Fig. 6, Fig. 7. Fig. 6 is the IF image generated directly without image enhancement, and Fig. 7 is the IF image after LHCH enhancement. It can be seen from the figure that the time-frequency ridges in Figure 6 are easily mixed with the background, resulting in fewer time-frequency ridges that can be seen in the figure. Especially in the picture (a) in Fig. 6, that is, the IF picture in the normal state can hardly see useful information. After the image is enhanced, obviously 3 time-frequency ridges can be seen. In (d) in Fig. 7, time-frequency ridges can be clearly seen in Fig. 7 (d). In order to explore the performance of LHCE proposed in Section 2, this paper uses information entropy (IE) of the image as an objective quantitative indicator. Information entropy refers to the amount of information covered by the image. The larger the value, the richer the image details. The clearer the hierarchy and structure, the higher the image quality, the formula is as follows:

\[
E = -\sum_{i=0}^{2^R-1} P(R) \log_2 P(R)
\]  

(18)

In expression (18): \( R \) is the gray level of the image pixel, usually \( R = 2^8 - 1 \), \( P(R) \) is the probability of the image appearing at that point when the gray value \( R \) of a certain point in the image. Select the instantaneous frequency image of normal state, rolling body fault, inner ring fault and outer ring fault. LHCE method was compared with histogram equalization (HE), single-scale Retinex (SSR) algorithm, and Contrast Limited Adaptive Histogram Equalization (CLAHE). The specific comparison test is shown as follows:

**FIGURE 6.** IF image directly generated in case 1 (a)N (b)IRF (c)ORF (d)BF

**FIGURE 7.** IF image enhanced by LHCE in case 1 (a)N (b)IRF (c)ORF (d)BF
This paper uses IE to evaluate the quality of different IF images. The higher the IE value, the more fault information contained in picture. Fig. 8 shows the IF image results of different image enhancement methods for four fault signals. The IE values are written in the table below. As shown in Fig. 7, the IE of the LHCE method achieves the maximum value in the three data sets of N, IRF, BF. Only the data set of ORF lags behind the CLAHE method and is ahead of other image enhancement algorithms. The image information entropies enhanced by HE, CLAHE, SSR and LHCE methods are 1.06, 3.19, 2.34 and 3.37 times of the original image respectively. Therefore, the LHCE displays more texture features in the IF image to show subtle differences between different fault categories.

In order to classify different types of faults, this paper imports the original IF image and the LHCE image-enhanced data set into the Resnet network for classification. Each data set contains 10 categories and 20000 pictures. The classification results of the two data sets are as follows Table III. The method of directly generating pictures is compared with the LHCE image enhancement method. It can be seen that the accuracy of the LHCE image enhancement method is higher than that of the directly generating picture method. The average accuracy, minimum accuracy and standard deviation of image classification generated by LHCE are: 98.45%, 97.74%, 0.4425. It can be seen from the results that these models can basically meet the classification requirements.

![Graph of Information Entropy](image)

**FIGURE 8.** Comparison of LHCE and other image enhancement methods in case 1

| NO | original | LHCE |
|----|----------|------|
| 1  | 97.25    | 98.65|
| 2  | 98.17    | 98.36|
| 3  | 97.55    | 97.74|
| 4  | 98.43    | 98.91|
| 5  | 98.29    | 98.59|
| Mean | 97.938 | 98.45|
| Std  | 0.4568  | 0.4425|

In order to further improve the classification accuracy, this paper proposes MDCNet network and LHCE to obtain a high-resolution time-frequency image size of 512x1024. In order to meet the input size of the proposed MDCNet model, the image is randomly cropped to the size of MDCNet model training. Then, the training image is input into MDCNet for training. Finally, all test images were input into the trained MDCNet model for fault diagnosis to obtain diagnostic accuracy.

![Confusion Matrix](image)

**FIGURE 9.** Confusion matrix obtained from MDCNet network

With the LHCE image enhancement method, it can be seen that the fault diagnosis results of the proposed model are very close to the actual results.
In order to predict the classification performance of the proposed CNN model, other classification methods were chosen to compare the prediction accuracy in this case. The original vibration signal is converted into an IF image and then classified using different methods. There are support vector machines (SVM) based on radial basis function (RBF), K nearest neighbors (KNN), AlexNet network, VGG-16 network, and Resnet network. The average predicted accuracy is the final measure of the comparison. The comparison results are shown in Table IV. The method of this study is effective, with an accuracy rate of 99.9%. The prediction results of KNN, AlexNet, VGG-16, RBF-SVM, and Resnet are respectively 95.70%, 98.51%, 97.64%, 93.46%, and 98.88%. The comparison can be obtained. The traditional network does not have the high classification accuracy of the MDCNet network. The classifier did not identify some weak fault features, which showed the superiority of the proposed MDCNet method.

Table IV

| Method   | Mean Accuracy (%) |
|----------|-------------------|
| MDCNet   | 99.90             |
| KNN      | 95.70             |
| AlexNet  | 98.51             |
| VGG-16   | 97.64             |
| RBF-SVM  | 93.46             |
| Resnet   | 98.88             |

Case 2

The bearing failure test bench is shown in Fig. 11. The platform is composed of motors, supporting bearing seats, vibration sensors, hydraulic resistors, couplings and other mechanisms. The simulated faults of 1mm of bearing inner ring, outer ring, and rolling elements are respectively processed by EDM as shown in Fig. 12. The data set is the vibration signal collected at a sampling frequency of 12.8 kHz and a motor speed of 1500 r/min. The test mainly selected the vibration signals of the rolling bearing under four health conditions, namely normal conditions, inner ring failure, outer ring failure and rolling element failure, for testing and verification. The details of the data set are shown in Table V:

Table V

| Fault location | Motor Load (kg) | Speed (/rpm) | Sampling frequency (Hz) | label |
|----------------|-----------------|--------------|-------------------------|-------|
| Normal         | 500             | 1500         | 12800                   | 1     |
| Ball fault     | 500             | 1500         | 12800                   | 2     |
| Inner race fault | 500             | 1500         | 12800                   | 3     |
| Outer race fault | 500             | 1500         | 12800                   | 4     |
The result of image enhancement is shown in Fig. 13. Fig. 14. Fig. 13 directly generated IF without image enhancement. Fig. 14 is the IF image after LHCE enhancement. The picture in Fig. 14 contains more texture details, especially some faint details, have no obvious features in Fig. 13, and can be clearly identified after image enhancement. Select the instantaneous frequency maps of normal state, rolling element failure, inner ring failure, and outer ring failure, respectively, with HE, SSR and CLAHE. As a comparison, the specific comparison test is as follows.

Fig. 15 shows the IF image results of different image enhancement methods for four fault signals. The IE values are written in the table below.
As shown in Fig. 15, the IE of the LHCE method achieves the maximum value in the three data sets of N, IRF, ORF. Only when the BF data set lags behind the CLAHE method and is ahead of other image enhancement algorithms. The image information entropies enhanced by HE, CLAHE, SSR and LHCE methods are 1.04, 2.84, 1.15 and 3.83 times of the original image respectively. Therefore, The LHCE displays more texture features in the IF image to show subtle differences between different fault categories.

![Graph showing information entropies](image)

**FIGURE 15.** comparison between LHCE and other image enhancement method

The IF image generated directly and the data set enhanced by LHCE image are input into Resnet network for classification. Each data set contains 4 categories and 8000 pictures. The classification results of the two data sets are as follows. The method of directly generating image by case 2 is compared with that of LHCE image enhancement. The results of table VI show that: The accuracy of LHCE image enhancement method is higher than that of direct image generation method. It shows that image enhancement is beneficial to classification. The average accuracy and standard deviation of image classification generated by LHCE are 99.19% and 0.46. It can be seen from the results that these models can basically meet the classification requirements.

| NO | 1     | 2     | 3     | 4     | 5     | Mean | Std |
|----|-------|-------|-------|-------|-------|------|-----|
| original | 98.56 | 98.47 | 97.89 | 98.43 | 98.29 | 98.33 | 0.26 |
| LHCE     | 98.65 | 99.36 | 98.74 | 99.61 | 99.59 | 99.19 | 0.46 |

In order to verify the performance of the extracted method in this paper, the image after image enhancement is randomly clipped to become the size required by CNN model, and the training image is input into CNN for training to obtain the diagnostic accuracy. Fig. 16 is a confusion matrix for MDCNet network classification. The classification accuracy can reach 99.7%. Only in the Ball fault category, the fault is misclassified as a normal signal. It is proven that the fault diagnosis results of the proposed model are quite close to the actual results.

Fig. 17 shows the visualization of the output layer of the MDCNet network. The similar samples gather without any intersection at a distance from each other, which means that the feature fault feature classification is more effective and meets the conditions for bearing fault classification. The practicability and accuracy of bearing fault classification are proved.

![Confusion Matrix](image)

**FIGURE 16.** Confusion Matrix for Case 2 Data Set
In order to further verify the classification performance of the proposed MDCNet, this paper chooses to perform a comparative experiment with other classifiers. The comparison results are shown in Table V. The forecasting results of KNN, AlexNet, VGG-16, RBF-SVM and ResNet are 93.99%, 96.25%, 94.87%, 96.47% and 98.25% respectively. Compared with other methods, the method studied in this paper has better effect and the accuracy rate reaches 99.7%. The validity of the proposed fault diagnosis method is demonstrated.

Table VII. Comparison of different classification methods

| Method   | Mean Accuracy (%) |
|----------|-------------------|
| MDCNet   | 99.70             |
| KNN      | 93.99             |
| AlexNet  | 96.25             |
| VGG-16   | 94.87             |
| RBF-SVM  | 96.47             |
| ResNet   | 98.25             |

V. Conclusion

The proposed fault diagnosis method includes two steps: time-frequency image enhancement and fault classification. The IF image is extracted from the time domain signal and input into the MDCNet network through image enhancement to test the bearings with different types of faults. The main contribution of this paper is to propose the LHCE image enhancement algorithm to enhance texture features of time-frequency images. Compared with image enhancement methods, this method has a higher information entropy value and can clearly display weak features of bearing faults. Then, the MDCNet is designed to help better predict the accuracy of the model and reduce memory consumption in the calculation process. The application and effectiveness of the method are verified by different data sets, and the test accuracy is 99.9% and 99.7% respectively. In addition, compared with other existing diagnostic methods, this method has obvious advantages in accuracy.

Future work: Further work is planned to understand the application of the proposed method to acoustic and gear fault signal analysis of fault diagnosis.

REFERENCES

[1] R. Bai, Q. Xu, Z. Meng, L. Cao, K. Xing, and F. Fan, “Rolling bearing fault diagnosis based on multi-channel convolution neural network and multi-scale clipping fusion data augmentation,” Measurement, vol. 184, p. 109885, 2021.
[2] Z. Ye and J. Yu, “AKRNet: A novel convolutional neural network with attentive kernel residual learning for feature learning of gearbox vibration signals,” Neurocomputing, vol. 447, pp. 23-37, 2021.
[3] Z. Wang, J. Zhou, W. Du, Y. Lei, and J. Wang, “Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution,” Mech. Syst. Signal. Pr., vol. 162, p. 108018, 2022.
[4] D. He, H. Cao, S. Wang, and X. Chen, “Time-reassigned synchrosqueezing transform: The algorithm and its applications in mechanical signal processing,” Mech. Syst. Signal. Pr., vol. 117, pp. 255-279, 2019.
[5] W. Zhao et al., “Multiscale inverted residual convolutional neural network for intelligent diagnosis of bearings under variable load condition,” Measurement, vol. 188, p. 110511, 2022.
[6] I. Daubechies, J. Lu, and H.-T. Wu, “Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool,” Appl. Comput. Harmon. A., vol. 30, no. 2, pp. 243-261, 2011.
[7] A. Mahdavi, A. Roshandel Kahoo, M. Radad, and M. Soleimani Monfared, “Application of the local maximum synchrosqueezing transform for seismic data,” Digit. Signal Process., vol. 110, 2021.
[8] Z. Wang, W. Zhao, W. Du, N. Li, and J. Wang, “Data-driven fault diagnosis method based on the conversion of erosion operation signals into images and convolutional neural network,” Process Saf. Environ., vol. 149, pp. 591-601, 2021.
[9] T. Oberlin, S. Meignen, and V. Perrier, “Second-order synchrosqueezing transform or invertible reassignment? Towards ideal time-frequency representations,” IEEE T. Signal Proces., vol. 63, no. 5, pp. 1335-1344, 2015.
[10] R. Behera, S. Meignen, and T. Oberlin, “Theoretical analysis of the second-order synchrosqueezing transform,” Appl. Comput. Harmon. A., vol. 45, no. 2, pp. 379-404, 2018.
[11] D. H. Pham and S. Meignen, “High-Order Synchrosqueezing Transform for Multicomponent Signals Analysis—With an Application to Gravitational-Wave Signal,” IEEE T. Signal Proces., vol. 65, no. 12, pp. 3168-3178, 2017.
[12] W. Liu, W. Chen, and Z. Zhang, “A Novel Fault Diagnosis Approach for Rolling Bearing Based on High-Order Synchrosqueezing Transform and Determined Fluctuation Analysis,” IEEE Access, vol. 8, pp. 12533-12541, 2020.
[13] Y. Gang, M. Yu, and C. Xu, “Synchroextracting Transform,” IEEE T. Ind. Electron., vol. 64, no. 10, pp. 8042-8054, 2017.
[14] X. Zhu, Z. Zhang, J. Gao, B. Li, and G. Wen, “Synchroextracting chirplet transform for accurate IF estimation and perfect signal reconstruction,” Digit. Signal Process., 2019.
[15] L. Xu, X. Yin, Z. Zong, and K. Li, “Synchrosqueezing Matching Pursuit Time-Frequency Analysis,” IEEE Geosci. Remote S., vol. PP, no. 99, pp. 1-5.
[16] G. Yu, Z. Wang, and P. Zhao, “Multisynchrosqueezing Transform,” IEEE T. Ind. Electron., vol. 66, no. 7, pp. 5441-5455, 2018.
[17] D. Fourer and F. Auger, “Second-order Time-Reassigned Synchrosqueezing Transform: Application to Draupner Wave Analysis,” pp. 1-5, 2019.
[18] Z. He, X. Tu, W. Bao, Y. Hu, and F. Li, “Gaussian-modulated linear group delay model: Application to second-order time-reassigned synchrosqueezing transform,” Signal Process., vol. 167, no. Feb., pp. 107275.1-107275.7, 2020.
Z.Zeng et al.: Application Of Local Histogram Clipping Equalization Image Enhancement In Bearing Fault Diagnosis

[19] C. Liang et al., "A Kaiser window-based S-transform for time-frequency analysis of power quality signals," IEEE T. Ind. Inform., vol. 18, no. 2, pp. 965-975, 2021.

[20] G. Sun, Y. Gao, K. Lin, and Y. Hu, "Fine-Grained Fault Diagnosis Method of Rolling Bearing Combining Multisynchrosqueezing Transform and Sparse Feature Coding Based on Dictionary Learning," Shock Vib., vol. 2019, pp. 1-13, 2019.

[21] Y. Xin, S. Li, C. Cheng, and J. Wang, "An intelligent fault diagnosis method of rotating machinery based on deep neural networks and time-frequency analysis," J. Vibroeng., vol. 20, no. 6, pp. 2321-2335, 2018.

[22] Q. Liu and C. Huang, "A Fault Diagnosis Method Based on Transfer Convolutional Neural Networks," IEEE Access, vol. 7, pp. 171423-171430, 2019.

[23] Long et al., "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," IEEE T. Ind. Electron., vol. 65, no. 7, pp. 5990-5998, 2017.

[24] G. Sun, Y. Gao, Y. Xu, and W. Feng, "Data-Driven Fault Diagnosis Method Based on Second-Order Time-Reassigned Multisynchrosqueezing Transform and Evenly Mini-Batch Training," IEEE Access, vol. 8, pp. 120859-120869, 2020.

[25] G. Yu, Z. Wang, P. Zhao, and Z. Li, "Local maximum synchrosqueezing transform: An energy-concentrated time-frequency analysis tool," Mech. Syst. Signal Pr., vol. 117, pp. 537-552, 2019.

[26] Y. He, Z. Jiang, M. Hu, and Y. Li, "Local Maximum Synchrosqueezing Chirplet Transform: An Effective Tool for Strongly Nonstationary Signals of Gas Turbine," IEEE T. Instrum. Meas., vol. 70, pp. 1-14, 2021.

[27] Szegedy, C. Wei Liu, Yangqing Jia, Sermanet, P. Reed, S.Anguelov, D. Rabinovich, A. (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2015.7298594

[28] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN, 2020 IEEE," 2020: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1571-1580.