International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 1260-1277

https://doi.org/10.70517/ijhsa464118

An Intelligent Bearing Fault Diagnosis Method Based on SE-ResNet 50 and GAF-MTF Encoding Method

Tianlin Song[1,*], Yuankang Qu[1], Huidong Qu[1], Zihao Tang[1], Ruixian Xue[1], Chuanzhe Ren[1] and Zongheng Ma[1]
[1] School of Mechanical and Electrical Engineering, Shandong Jianzhu University, Jinan, Shandong, 250101, China
Corresponding authors: (e-mail: 13181470546@163.com).

Abstract Current methods for bearing fault diagnosis under complex conditions face limitations in capturing temporal signal correlations, multi-dimensional features, and adaptive feature extraction. This research introduced a sophisticated method that integrates the Gramian Angular Field-Markov Transition Field (GAF-MTF) encoding method with an enhanced Squeeze-and-Excitation ResNet 50 (SE-ResNet 50) model to effectively address the issue at hand. The GAF-MTF method fuses static (GASF), dynamic (GADF), and probabilistic transition (MTF) features to convert 1D timing signals to 2D images, preserving temporal correlations and enhancing fault representation. These images are processed by SE-ResNet 50, which employs channel attention mechanisms to dynamically prioritize critical features and enhance stability. Experiments on the CWRU dataset achieved 99.87% accuracy, with validation on the Jiangnan University dataset yielding 99.05%, demonstrating great generalization ability. Additionally, we utilized t-SNE to reduce feature dimensions and analyzed the role of every residual layer. The framework provides reliable fault diagnosis under variable conditions, with future work targeting computational efficiency and lightweight architectures for broader industrial deployment.

Index Terms fault diagnosis, ResNet, time series classification, SE attention mechanism

I. Introduction

As Industry 4.0 continues to gain momentum and smart manufacturing evolves at breakneck speed, mechanical equipment has emerged as a cornerstone of modern production systems. Among the key elements driving these systems, rolling bearings stand out as indispensable components, crucial for keeping machinery running smoothly and efficiently. A severe failure in rolling bearings can directly lead to the cessation of rotating mechanisms and machine shutdowns, resulting in significant economic losses for factories. In the actual operational process of machinery, real-time monitoring of bearings is necessary to enable timely diagnosis and maintenance upon failure detection, thereby preventing further deterioration of the fault. Therefore, precise, effective, and straightforward bearing fault diagnosis [1] is crucial for enhancing equipment safety and preventing avoidable production losses [2].

The swift progress in artificial intelligence and computing capabilities has greatly accelerated the digital evolution and smart enhancement of fault diagnosis methods. Machine learning and deep learning-based models are capable of automatically analyzing and identifying multiple failure modes via sophisticated training processes [3], [4]. This innovative advancement enables diagnostic systems to flexibly adjust to the intricacies of diverse machinery and shifting environmental factors, significantly boosting their adaptability and overall sophistication [5]. Currently, deep learning-based intelligent fault diagnosis methods are extensively applied in engineering. Compared with traditional methods [6]-[8], these ways show better feature extraction capabilities and diagnostic accuracy.

To acquire details about potential internal malfunctions during the machine's operation, we can only ascertain its internal condition by analyzing relevant external data. According to measurement science literature, the original vibration signal is the most effective and essential tool for diagnosing rolling bearing faults [9], [10]. General intelligent fault diagnosis methods include data collection, feature extraction, and feature classification. The data collection stage underpins the analysis and is crucial for ensuring the accuracy of later steps [11]. Feature extraction derives key features from the source signal, and commonly used feature extraction techniques include Time-Domain Statistical Analysis [12], Wavelet Transform [13], and Fourier Spectrum Analysis [14]. In addition, common classifiers include K-Nearest Neighbor (KNN) [15], Multi-Layer Perceptron (MLP) [16], Support Vector Machine (SVM) [17], Decision Tree (DT) [18], etc. Many researchers have proved these methods to be suitable in the field of bearing fault identification. Lou et al. [19] introduced an approach utilizing the Wavelet Transform for accelerometer signal analysis, yielding feature vectors suitable for training neural fuzzy inference systems. Experimental findings demonstrate that the suggested approach can dependably distinguish distinct faults under diverse load circumstances.

Wang et al. [20] proposed a bearing fault diagnosis method using vibration-acoustic data fusion and a 1D-CNN. By processing and fusing multi-modal sensor signals, the model leverages CNN's feature extraction capability, achieving higher accuracy than single-mode approaches. Currently, the majority of bearing fault diagnosis models are trained using ample labeled data. In order to diagnose bearing faults without sufficient labeled data, Fu et al. [21] developed AC-FEGAN, a bearing fault diagnosis method combining feature-enhanced generative adversarial networks and auxiliary classifiers. The model automatically extracts representative defect features, overcomes shallow CNNs' feature extraction limitations, and integrates a self-attention module for fine feature refinement.

In recent years, researchers have generally found that single-dimensional processing of bearing fault signals often fails to achieve better diagnostic expectations, which may be due to the multi-dimensional features of timing signal. Single-dimensional techniques might segregate time and frequency domain data, but the features extracted cannot fully represent the signal's complete information. Therefore, multi-dimensional feature fusion technology has gradually become the focus of research, which mainly presents two technical paths: The first one is to combine multi-mode signals such as vibration and acoustics [22], and obtain multi-dimensional information by capturing various original signal features. However, in the input of heterogeneous features, channels of different modes may cause feature space mismatch due to differences in distribution (such as dimension and numerical range), forcing the model to consume a large number of parameters in the shallow network for feature alignment instead of focusing on deep semantic extraction. The second is to carry out multiple encoding conversions on the same signal to capture multi-dimensional features. The multi-dimensional features are extracted from the same signal by mathematical encoding, and the underlying signal sources of all feature channels are consistent, and their mathematical representations are intrinsically related, avoiding the problem of multi-modal heterogeneity. Although multi-dimensional features reflect more comprehensive information, which has higher requirements on hardware equipment and increased costs. Benefiting from the development of residual networks, He et al. [23] of Microsoft proposed a residual network in 2016, which is a variant of CNN. By employing identity mapping and skip connections, the depth of the network layer was significantly enhanced, substantially decrease computational volume, and elevate computational efficiency. By using the ResNet structure, the model can extract most features without consuming too much computing resources. Hou et al. [24] integrated Transformer, ResNet, feature extraction, and Transfer Learning to build a model with high noise resistance and superior prediction accuracy compared to traditional DL networks. He et al. [25] introduced a rolling bearing fault diagnosis method via multi-signal fusion and MTF-ResNet. By integrating temporal signal correlations and multi-modal data, the approach extracts and fuses image features to generate comprehensive representations. The model enables compound fault diagnosis under variable conditions, validating ResNet's capability to learn complex features in challenging scenarios.

The above research shows that residual network is a feasible method to solve the bearing fault diagnosis problem. ResNet ensures diagnostic effectiveness irrespective of professional expertise and exhibits resilience to noise in disturbed settings. However, although ResNet achieves deep networks by stacking small convolutional nuclei, its effective perceptual field of view is much smaller than the theoretical value, resulting in weak global context information extraction ability, which limits the model's performance in long-distance dependent tasks. In addition, the processing ability of ResNet for non-stationary signals is limited, and the static residual structure makes it difficult to adjust the feature weight adaptively, so there are inevitable defects in the fault diagnosis classification of ResNet [26]. Consequently, our research endeavors to enhance in two key areas: (1) discovering a more efficient encoding approach, such that the generated 2D images possess comprehensive and non-linear information regarding vibration faults; (2) Seeking to refine ResNet, ceasing to overly depend on local information, boosting its global feature-processing capacity, thereby enabling it to dynamically regulate the weight of each feature and ensuring that the network model exhibits higher accuracy and generalization capabilities.

In this study, a new fault diagnosis method was mentioned. Firstly, the 1D timing signal undergoes GAF-MTF encoding to form a 2D image, which is subsequently fed into the enhanced SE-ResNet model for training. The original vibration signal, encoded through the GAF-MTF method, possesses both static and dynamic features, which can be effectively extracted and trained using the SE-ResNet 50. The model excels on the CWRU bearing data set and shows consistent generalization across additional data sets. The remainder of this paper is structured as follows: The second section introduces the theoretical background. The third section introduces the encoding method and model structure. The fourth section introduces four main experimental processes and results. The fifth section concludes the full text.

II. Theoretical Background

II. A.Gramian Angular Field

Gramian Angular Field is an encoding approach capable of transforming time series into 2D images. The method maps time series into 2D space for angle representation, constructs a 2D matrix based on these angles, and visualizes the 2D matrix to form 2D images. The image's color intensity per distance unit indicates the data's time series correlation. The encoding methodology is detailed as:

(1) For a set of given time series $X = \{x_1, x_2, x_3, \cdots, x_n\}$, where x_i is the i-th sample signal. To maintain data range uniformity, equation (1) is used to normalize the signal data, resulting in x being scaled within the interval [0,1]. The normalized sample signal is represented by $\tilde{x}_i$.

$$\tilde{x}_i = \frac{x_i - \max(X) + x_i - \min(X)}{\max(X) - \min(X)} \tag{1}$$

(2) The normalized x_i is encoded as the included angle cosine value φ, and the time series is encoded as the radius r. Normalized time series are represented in polar coordinates. Thus, the initial time domain signal is transformed into a polar coordinate time series, represented by equation (2).

$$\begin{cases} \varphi = \arccos(\tilde{x}_i), -1 < \tilde{x}_i < 1, \tilde{x}_i \in \tilde{X} \\ r = \dfrac{t_i}{N}, t_i \in N \end{cases} \tag{2}$$

where t_i is the time stamp, N is the constant factor of the generating space of the regularized polar coordinate system. By encoding in this way, the original information is well preserved.

(3) When the normalized time series is converted into polar coordinates, the temporal relationships across various intervals become apparent by examining the cosine of the polar angle and the angular shifts between individual data points. GAF can be divided into GASF and GADF according to different mathematical operations, and their mathematical models can be expressed by equations (3) and (4) respectively.

$$GASF = \left[\cos\left(\varphi_i + \varphi_j\right)\right] = \begin{bmatrix} \cos(\varphi_1 + \varphi_1) & \cdots & \cos(\varphi_1 + \varphi_n) \\ \cos(\varphi_2 + \varphi_1) & \cdots & \cos(\varphi_2 + \varphi_n) \\ \vdots & \ddots & \vdots \\ \cos(\varphi_n + \varphi_1) & \cdots & \cos(\varphi_n + \varphi_n) \end{bmatrix} \tag{3}$$

$$GADF = \left[\sin\left(\varphi_i - \varphi_j\right)\right] = \begin{bmatrix} \sin(\varphi_1 - \varphi_1) & \cdots & \sin(\varphi_1 - \varphi_n) \\ \sin(\varphi_2 - \varphi_1) & \cdots & \sin(\varphi_2 - \varphi_n) \\ \vdots & \ddots & \vdots \\ \sin(\varphi_n - \varphi_1) & \cdots & \sin(\varphi_n - \varphi_n) \end{bmatrix} \tag{4}$$

GASF can extract similar signal features from complex signal sequences and is suitable for signals with strong stationarity. GADF emphasizes signal dynamic changes, making it ideal for analyzing non-stationary signals or those with abrupt variations.

GAF provides a way to maintain a time-dependent, increasing time as positions in the matrix move from the top left to the bottom right. Such a matrix contains temporal dependencies because $G_{(i,j||i-j|=k)}$represents relative dependencies with respect to superpositions or differences in the direction k of time intervals. When $k = 0$, $G_{(i,j)}$ is the element on the main diagonal, which holds the raw time series data. Using this approach, a series of length n is transformed into an $n \times n$ matrix.

II. B.Markov Transition Field

Markov Transition Field represents the probabilistic transitions in a time series, which is mainly based on the first-order Markov chain, aiming at the low time dependence of the Markov transition matrix on the sequence, a method is proposed by adding the position relation of time t. This encoding method uses transition probability matrix to represent the dynamic transition information of time series, and enhances the ability of data visualization while preserving the time and frequency of the sequence. The reliance on time series X is insufficiently robust, and matrix W incurs excessive data loss compared to the original series, necessitating enhancement. The specific encoding process of MTF is as follows:

(1) For the 1D time series $X = \{x_1, x_2, x_3, \cdots, x_n\}$, it is divided into Q quantile units according to the numerical range. Each value in the 1D time series is quantized by quantile $q_j(j \in \{1,2,\ldots,Q\})$. By identifying the quantile, the value $x_i(i \in \{1,2,\ldots,n\})$ in the sequence corresponds to the unique q_i.

(2) The Markov transition matrix W_{QQ} is constructed from $w_{i,j}$ with the dimensions $Q \times Q$. Where $w_{i,j}$ is determined by the probability P of data migration from q_i to q_j, and $w_{i,j}$ can be represented by equation (5).

$$w_{i,j} = P(x_t \in q_i \left| x_{t-1} \in q_j \right.) \tag{5}$$

(3) Probabilities are listed chronologically, forming a Markov transition matrix of a given size, as depicted in equation (6).

$$M = \begin{bmatrix} w_{i,j} \left| z_1 \in q_i, z_1 \in q_j \right. & \cdots & w_{i,j} \left| z_1 \in q_i, z_n \in q_j \right. \\ w_{i,j} \left| z_2 \in q_i, z_1 \in q_j \right. & \cdots & w_{i,j} \left| z_2 \in q_i, z_n \in q_j \right. \\ \vdots & \ddots & \vdots \\ w_{i,j} \left| z_n \in q_i, z_1 \in q_j \right. & \cdots & w_{i,j} \left| z_n \in q_i, z_n \in q_j \right. \end{bmatrix} \tag{6}$$

II. C.ResNet

In the realm of image recognition, it's commonly thought that the more layers a network has, the more adept it is at extracting features from the images being recognized. To achieve superior recognition outcomes, the network frequently grows to tens or even hundreds of layers. However, too many layers often cause problems with disappearing gradients or exploding, causing the network to fail to converge to the desired effect. The key concept of a residual network is to mitigate gradient descent issues by incorporating a residual link module between layers in a standard convolutional neural network.

The residuals module is the core of the ResNet, which consists of multiple convolution layers that extract the features of the residual module input, with batch normalization and activation functions between the convolution layers, and finally the input x is added to the output. Different from the traditional CNN model, ResNet adopts a skip connection. Utilizing a skip connection, the initial input data bypasses several convolutional stages and is promptly forwarded to the following layers. The output of the convolutional stage feeds into the subsequent layers, powered by an activation function, to yield the ultimate result from the residual learning unit. In a ResNet, the identity map is used as part of a skip connection, allowing forward and reverse signals to travel directly between modules. This design allows the network to learn more deeply while maintaining training stability. The use of identity mapping has also been shown to be a key factor in improving network performance [27]. Figure 1 shows the detailed structure of the residual module.

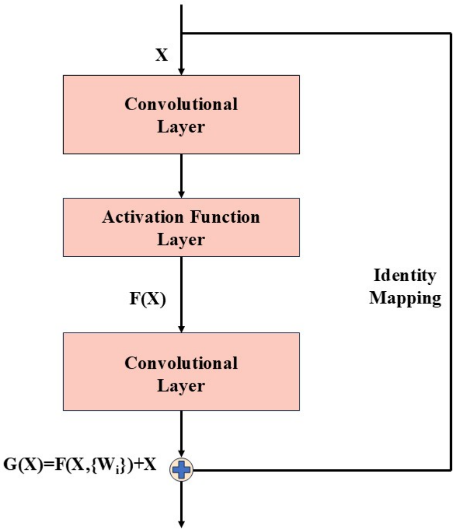


Figure 1: Structure of residual module

As shown in Figure 1, $F(X)$ represents the residual mapping function, W_i denotes the weights obtained after the i-th convolutional layer, and X indicates the input of the residual module. When the input dimension does not match the output dimension, the model applies linear transformation W_S to X via the skip connection. Equation (7) illustrates the resulting output objective function.

$$G(X) = F\left(X, \{W_i\}\right) + W_s X \tag{7}$$

ResNet tackles the issue of gradient vanishing and explosion in deep neural networks without augmenting the layer count, additional learning parameters, or computational complexity through a simple connection method that passes errors back to the previous layer during training.

Bottleneck is an optimized version of a standard residual module that balances model depth and computational efficiency, and is mainly used for deep networks, such as ResNet 50/101. Bottleneck is mainly composed of three parts. 1×1 convolution dimension reduction: compress the number of input channels to reduce the amount of subsequent computation; 3×3 convolutional feature extraction: core space feature extraction is carried out in low-dimensional space to retain local detail information; 1×1 convolution increment: restores the number of channels to the original dimension, ensuring that it matches the dimension of the residual branch. Bottleneck, on the basis of retaining residual connections, reduces noise interference by dimensionality reduction, ensures feature expression ability, and greatly reduces the number of parameters, which improves computation efficiency.

II. D.Squeeze-and-Excitation Module

Squeeze-and-excitation module (SE) is a channel attention mechanism, including three parts: squeeze, excitation and feature scale, which are used to enhance the representation ability of feature maps. It dynamically assigns weights to each channel by learning the dependencies between them, thereby enhancing important features and suppressing unimportant ones. The core idea of this mechanism is ‘information interaction between channels’, which can help models better extract global features. The specific structure of the SE module is shown in Figure 2.

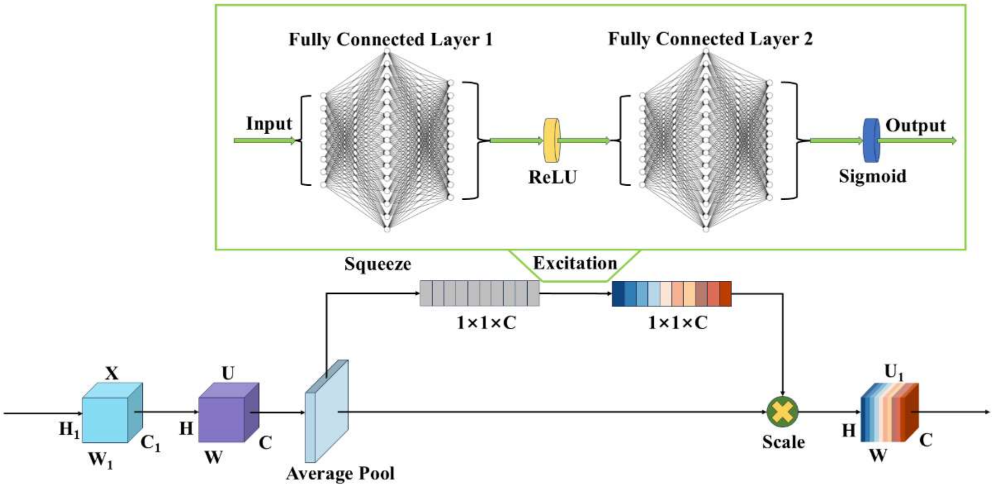


Figure 2: Structure of SE module

X is processed through a convolutional procedure, producing the feature map U, which then becomes the input for the attention module. The initial stage involves a squeeze operation, where global average pooling is applied to U. This operation compresses spatial information across all channels into a scalar value for each channel, which globally represents the channel-wise significance. This process can be formally expressed as equation (8).

$$Z_C = GAP\left(Y_C\right) = \frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H} Y_C\left(i, j\right) \tag{8}$$

In the equation, Z_C represents the output value of the C channels after global average pooling, Y_C represents the C channels of the input feature map, W is the length of the channel, H is the height of the channel. Then comes the excitation part, which includes two operations, reduction and restoration. The output of $1 \times 1 \times C$ reduces the channel to C/r channels after passing through the first fully connected layer to reduce the computational load (r refers to the proportion of reduction), and enhances network nonlinearity via ReLU activation, enabling the system to capture intricate features, then the output enters the second fully connected layer. The squeezed channels are then restored into C. Finally, the Sigmoid function is used to obtain the weight vector S, which represents the importance of each channel. The computational formula is presented in equation (9). The weight vector S is multiplied by the input U channel by channel to complete feature recalibration, and the final output $\widetilde{U} \in R^{H\times W\times C}$ is obtained.

$$S = \sigma\left(W_2 \cdot ReLU\left(W_1 \cdot Z_1\right)\right) \tag{9}$$

where Z_1 is the channel descriptor obtained after global average pooling, $Z_1 \in R^{1\times 1\times C}$, W_1 and W_2 are the parameters of the two fully connected layers respectively.

III. Proposed Method

The proposed method's structure is depicted in Figure 3, which is a fault diagnosis method with ResNet 50 and SE as the core. This section will explain this method in detail.

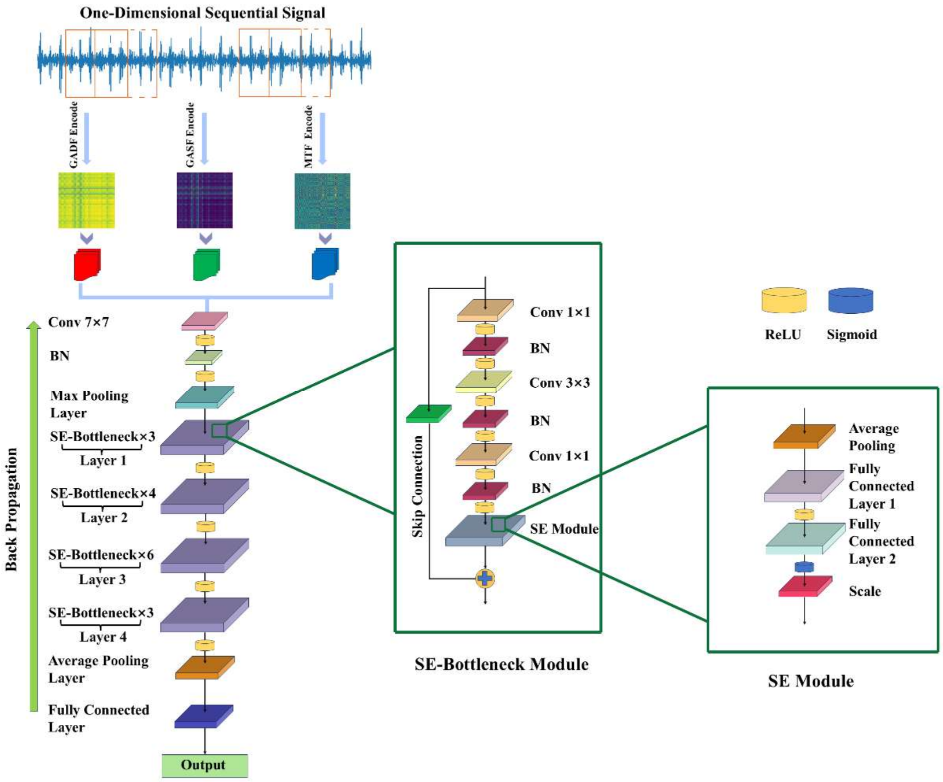


Figure 3: The framework of proposed method

III. A. *Improved Encoding Method*

In this experiment, we anticipate that the encoded 2D image will capture the majority of the fault signal features, thereby enhancing the model's capability to extract fault features. For GADF, GASF and MTF encoding methods to convert time series into 2D images, this study presents a new method to fuse the three encoding results. For a group of partitioned sequence time $X = \{x_1, x_2, x_3, \cdots, x_n\}$, GADF, GASF and MTF encoding are carried out respectively, and the encoded 2D images are normalized to obtain three groups of different encoding results. The three groups of 2D images have different features of timing signals. Both can be extracted and learned by SE-ResNet 50.

The detailed steps of the new method are as follows:

The time series $X = \{x_1, x_2, \cdots, x_n\}$ is encoded by GADF, GASF and MTF respectively, three sets of RGB color images are obtained and normalized in the range of [0,1].

Convert the obtained three 2D images into grayscale images and reduce the number of channels to single channel.

Convert the pixel value of each grayscale image to NumPy array.

Stack three gray arrays along channel dimensions to generate a three-channel array, where three channels (R, G, B) correspond to the gray values of GADF, GASF, and MTF images respectively.

Convert the stacked array to RGB image and perform normalization operation.

The 2D image obtained by this method integrates the signal features captured by the three encoding methods, and will be used as the input of SE-ResNet 50.

III. B. *Model Building*

The SE-ResNet 50 model consists of an independent convolution layer, a maximum pooling layer, a global average pooling layer, a fully connected layer, and four residual layers. Each residual layer is composed of improved SE-Bottleneck, which contains SE module, Bottleneck module, ReLU activation function, and Sigmoid activation function. The function of ReLU is to introduce nonlinear, avoid gradient disappearing and accelerate training speed. The purpose of Sigmoid is to limit the output range so that the output is expressed as a probability. Two activation functions are shown in the Figure 4.

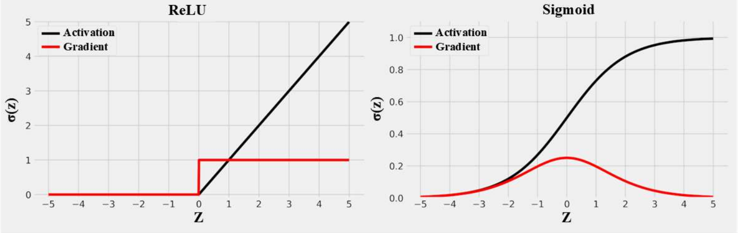


Figure 4: Each activation function and its gradient

The function of the SE module is to dynamically adjust the channel weights of the feature graph, enhance advanced features and suppress unimportant features. After the input image enters the first convolution kernel with the size of 7×7×64 to extract low-level features, the batch normalization(BN) process is carried out.BN employs whitening to transform any neuron's input distribution within a neural network layer into a standard normal distribution. A standard normal distribution has a mean of 0 and a variance of 1. This consistent distribution ensures that covariates remain stable across the network, preventing unwanted shifts. The derivative linked to the activation input value lies well outside the saturation zone, meaning even minor adjustments to the input can lead to significant fluctuations in the loss function. This dynamic allows for faster convergence of the neural network, optimizing its efficiency. The results after BN enter the maximum pooling layer for downsampling to reduce the spatial size of the feature map. Then, the output results enter the first residual layer, which is composed of three SE-Bottleneck with 256 channels. The first residual layer extracts medium-level features, and enhances feature representation through residue learning and SE module. The output result of the first residual layer is fed into next residual layer which is composed of four SE-Bottleneck with 512 channels. The second residual layer is used to extract medium-level and high-level features and downsample the feature graph. The output result of the second residual layer is fed into next residual layer which is composed of six SE-Bottleneck with 1024 channels. The third residual layer is used to extract high-level features and downsample the feature graph. The output result of the third residual layer is fed into the last residual layer which is composed of three SE-Bottleneck with 2048 channels. The fourth residual layer is used to extract the highest-level features and downsample the feature graph. The output of the four residual layers will be compressed by the feature map through the global averaging pooling layer, which is used to obtain the global feature vector. The feature vector will be fed into the fully connected layer, yielding the classification outcome.

To clarify the model's structure and purpose, Table 1 shows the detailed structure of the SE-ResNet 50 model, Table 2 shows the detailed structure of the SE-Bottleneck, and Table 3 shows the detailed structure of the SE module.

Table 1: Layer structure of the SE-ResNet 50 model

Network Layer	Size of Convolution Kernels	Input	Output
Convolution 1	7×7×64; Stride=2; Padding=3	224×224×3	112×112×64
BN1	/	112×112×64	112×112×64
Max Pooling Layer	3×3×64	112×112×64	56×56×64
Residual Layer 1	SE-Bottleneck×3	56×56×64	56×56×256
Residual Layer 2	SE-Bottleneck×4	56×56×256	28×28×512
Residual Layer 3	SE-Bottleneck×6	28×28×512	14×14×1024
Residual Layer 4	SE-Bottleneck×4	14×14×1024	7×7×2048
Average Pooling	/	7×7×2048	1×1×2048
Fully Connected Layer	/	1×1×2048	10

Table 2: Layer structure of the SE-Bottleneck

Network Layer	Size of Convolution Kernels	Input	Output
Convolution 1	1×1×C/4	H×W×C	H×W×C/4
BN1	/	H×W×C/4	H×W×C/4
Convolution 2	3×3×C/4	H×W×C/4	H×W×C/4
BN2	/	H×W×C/4	H×W×C/4
Convolution 3	1×1×C	H×W×C/4	H×W×C
BN3	/	H×W×C	H×W×C

SE Module	/	H×W×C	H×W×C
Skip Connection	/	H×W×C	H×W×C

Table 3: Layer structure of the SE module

Network Layer	Number of Channels	Input	Output	Activate Function
Average Pooling Layer	C	H×W×C	1×1×C	/
Fully Connected Layer	C/r	C	C/r	ReLU
Convolution 2	C	C/r	C	Sigmoid
BN2	C	H×W×C	H×W×C	/

IV. Experiment and Results

IV. A. Encoding Experiment

IV. A. 1) Data Description

To assess the properties of diverse encoding methods across distinct models, the CWRU bearing dataset was employed for experimentation. The experimental data were obtained from the platform shown in Figure 5, and the sampling frequency was 12kHz. In the sampling experiment, the motor load was categorized into four classes, select the class with having a load of 0 HP and an operating speed of 1797 r/min. Taking the deep groove ball bearing SKF-6205 as an example, the outer fault, inner fault, and roller fault were subjected to electrical discharge machining to create single-point faults of 0.007 inches, 0.014 inches, and 0.021 inches, respectively. The three types of faults are shown in Figure 6. This yielded nine unique fault categories. Additionally, normal data samples were also established for comparison.

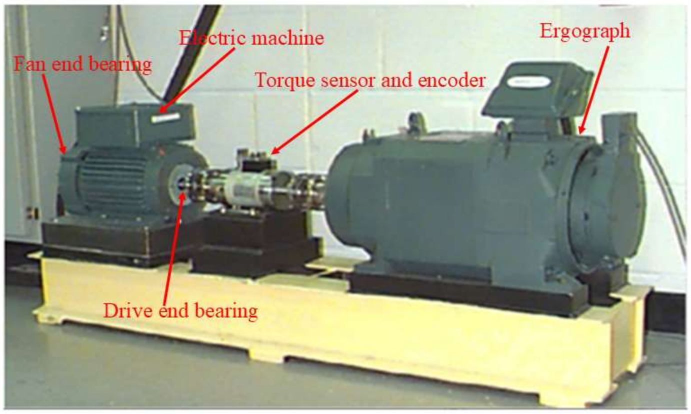
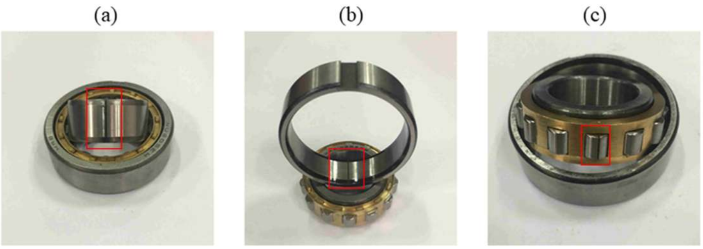


Figure 5: Timing signals collection platform.



(a) Inner fault; (b) Outer fault; (c) Roller fault.

Figure 6: Type of bearing faults

The base accelerometer data (BA), the driver accelerometer data (DE) and the fan accelerometer data (FE) are measured in the experiment. The obtained data are processed and converted into 1D vibration signals. As shown in the Figure 7, the normal and roller fault signals have strong randomness, while the inner and outer signals have certain periodic characteristics, and it is difficult to distinguish the bearing status directly.

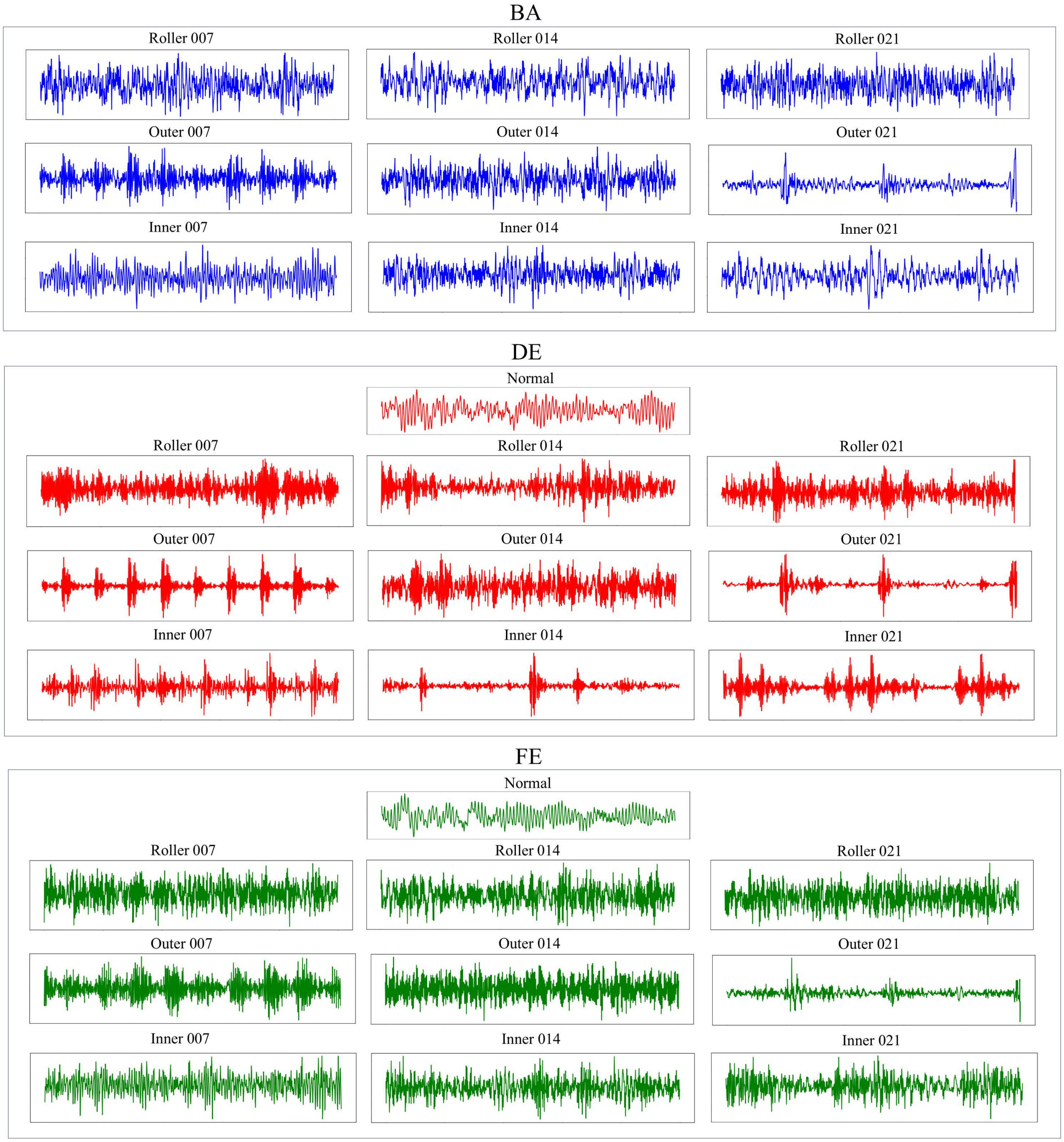


Figure 7: Vibration signal diagram for different bearing faults

1D time series signal is generated by GADF, GASF, MTF encoding method. To minimize model overfitting and expand the dataset's sample diversity, the sliding window captures time series data points and generates 2D images from the corresponding ones. To guarantee that every sample encompasses a single cycle, the size of the sliding window must be ascertained. The count of sampling points during one bearing rotation should be computed as per equation (10).

$$L = \frac{60 \times f}{S} \tag{10}$$

L represents the number of sampling points of the bearing within one cycle, which corresponds to the size of the sliding window. S represents the bearing rotational speed, and f indicates the sampling frequency.

The common input image sizes are 56×56, 128×128, 224×224, and 256×256. Considering that the input image must provide sufficient spatial resolution to capture the key features of the object, but also considering the speed of training, we have pre-selected a 2D image size of 224×224 as the input.

According to equation (10), to ensure each image has sampling points within one cycle, the bearing gathers around 400 data points per revolution. Thus, the minimum sliding window dimension is established at 400. The experimental data is sampled using the smallest possible sliding window. Given the length of the acquired signal, during the preprocessing stage, overlapping sampling with a 50% overlap rate is employed for basic data augmentation. This results in a step size of 112 data points per sample. This approach effectively doubles the number of samples while minimizing feature loss caused by direct truncation of the data.

Using the parameters previously chosen, the count of samples yielded by each fault category is determined, as depicted in equation (11).

$$N = \left| \frac{H - L}{O} \right|_{floor} + 1 \tag{11}$$

where N represents the number of 2D sample images, H represents the length of a 1D time series, L represents the size of the sliding window, O refers to the step size of the sliding window, and $floor$ represents rounding down. According to equation (11), there are 1080 fault instances per type, 2160 normal cases, and a combined total of 11,880 samples. To avoid classification bias potentially caused by particular data segmentation, 80% of the dataset was randomly chosen for training, with the remaining 20% allocated for testing. The resulting dataset is presented in Table 4.

Table 4: Data set partitioning in Encoding experiment

Group Number	Fault Types	Training Set	Testing Set
0	Inner 007	864	216
1	Inner 014	864	216
2	Inner 021	864	216
3	Outer 007	864	216
4	Outer 014	864	216
5	Outer 021	864	216
6	Normal	1728	432
7	Roller 007	864	216
8	Roller 014	864	216
9	Roller 021	864	216
Total	/	9504	2376

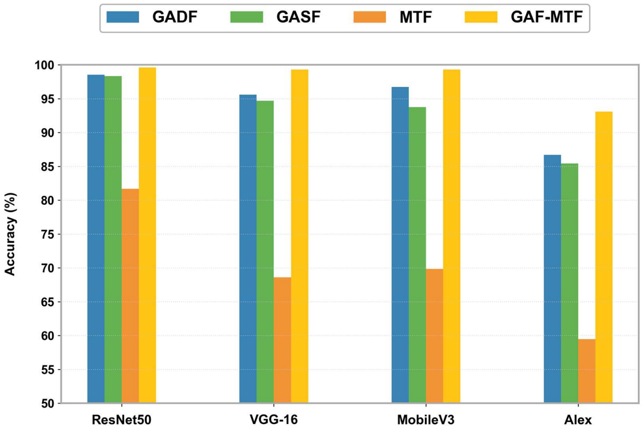
IV. A. 2) Experiment Processing and Result Analyzing

The optimizer employs Adam. Adam is capable of dynamically adjusting the learning rate of training parameters while smoothing parameter updates. The starting learning rate is configured at 0.001. To ensure rapid convergence of weights and prevent entrapment in local optima, a small-batch training method is utilized, with the batch size set to 32. The training loss function is the cross-entropy function, which integrates the softmax function and negative log-likelihood loss. This function accurately measures the variance between the model's predicted outcomes and their actual label distributions. The loss function is represented by equation (12).

$$Loss = -\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{n}\left[y^{(i)} \log \hat{y}^{(i)} + \left(1 - y^{(i)}\right) \log\left(1 - \hat{y}^{(i)}\right) \right] \tag{12}$$

where m is the number of samples, y is the actual sample class, $\hat{y}$ is the predicted sample class, and n is the sample class.

In this experiment, to comprehensively analyze the strengths and weaknesses of various encoding methods, four networks, ResNet 50, VGG-16, Mobile V3 and Alex, were introduced in the experiment and combined with four encoding methods, GADF, GASF, MTF and GAF-MTF, respectively. Through this combination, a total of 16 groups of experiments were designed. To minimize experimental randomness, 10 experiments were conducted in each group, and 100 epochs were carried out in each experiment. The average of the highest accuracy of the testing set in the 10 experiments was taken, and the outcomes were presented in Figure 8.



	GADF	GASF	MTF	GAF-MTF
ResNet50	98.54%	98.32%	81.68%	99.59%
VGG-16	95.57%	94.68%	68.59%	99.31%
MobileV3	96.73%	93.74%	69.82%	99.29%
Alex	86.69%	85.43%	59.47%	93.07%

Figure 8: The test accuracy of different encoding methods in four models

Through the examination of the outcomes, the following findings are derived:

In the single encoding mode, the 2D feature maps encoded by GADF demonstrate superior performance across all models, outperforming both GASF and MTF encoding modes. Notably, the MTF-encoded 2D images exhibit relatively poor performance, achieving accuracy above 80% only in the ResNet 50 model. We speculate that this is due to GADF's ability to more sensitively reflect the dynamic differences of signals at various time points, effectively capturing temporal characteristics with strong time correlation. In contrast, due to the smoothing effect resulting from the superposition of angle combinations, GASF cannot capture high - frequency features in the timing signal. Additionally, MTF relies heavily on probability and statistics for quantile state transitions, making it less sensitive to instantaneous signal changes and thus resulting in suboptimal performance.

Among the four models, ResNet 50 demonstrates the most superior performance. We speculate that this is attributable to the unique architecture of the residual network, which incorporates residual connections and enables gradients to bypass certain layers during backpropagation. This mechanism effectively mitigates the issues of gradient vanishing and degradation commonly encountered in deep networks. Additionally, the presence of skip connections and identity mappings allows ResNet 50 to preserve critical features while filtering out redundant ones, thereby compelling the network to learn residuals rather than complete reconstructions. Such a constraint directs the network to focus on distinctive features rather than redundant global information.

The GAF-MTF encoding method, when contrasted to the single encoding method, the highest diagnostic accuracy was achieved among the four models. Compared with the GADF encoding mode, the GAF-MTF encoding of the ResNet 50, VGG-16, Mobile V3, and Alex network models has respectively improved by 1.05% and 3.74%. 2.56%, 6.38% accuracy. The increase of accuracy is due to the transformation from single dimensional features to multi-dimensional features. The GAF-MTF encoding method captures most features of vibration signals, enabling the model to fully extracts the information of vibration signals. Therefore, the GAF-MTF encoding method has good generalization ability in a variety of networks.

IV. B. Experiment on Input Size

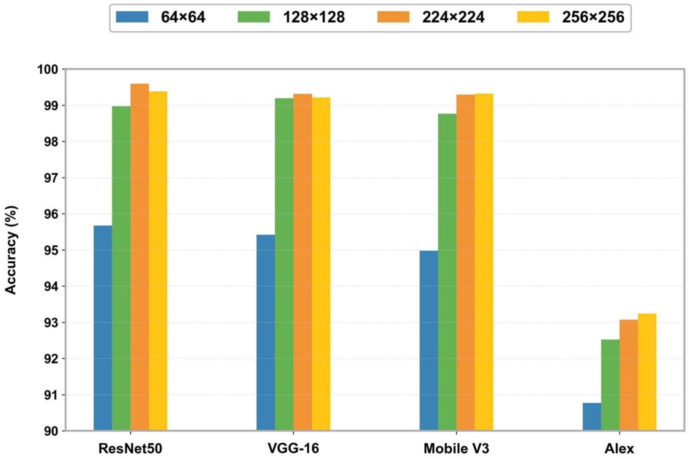
In the encoding experiment, when the input image size is 224×224, the model demonstrates satisfactory performance. However, this does not confirm whether feature maps of other sizes would be better represented within the model. Consequently, in this experiment, the size of the input feature map was systematically adjusted. The feature maps encoded by GAF-MTF were then input into ResNet 50, MobileV3, VGG-16, and Alex with size of 64×64, 128×128, 224×224, and 256×256 for validation. Similarly, to minimize experimental randomness, every experimental group underwent 10 repetitions.

The 1D vibration signal length in the CWRU dataset remains constant. The length of data points of inner, outer and roller fault is 108W (There are 12W of each fault size), and normal state data points total 24W. Due to the particularity of GADF, GASF encoding method, upon applying GAF to a signal with length n, the result is an $n \times n$ matrix. The four sizes are all contained in a sliding window, and overlapping sampling is not designed in this experiment. The GAF-MTF encoded feature images will be split into 80% for training and 20% for testing. The detailed data division is shown in Table 5.

Table 5: Data set partitioning for different input sizes

Group Number	Input Size	Training Set	Testing Set
A	64×64	16500	4125
B	128×128	8250	2062
C	224×224	4714	1178
D	256×256	4125	1031

After obtaining the feature images required for the experiment through GAF-MTF encoding, the feature images of varying sizes were individually input into four network models for training. The average of the highest accuracies achieved across 10 experiments is presented in Figure 9.



	64×64	128×128	224×224	256×256
ResNet50	95.67%	98.97%	99.59%	99.38%
VGG-16	95.42%	99.19%	99.31%	99.21%
Mobile V3	94.98%	98.76%	99.29%	99.32%
Alex	90.77%	92.52%	93.07%	93.24%

Figure 9: The test accuracy of different input sizes in four models

By analyzing the experimental results, the following conclusions are obtained:

At an input resolution of 64×64 pixels, the four models exhibit a notably low accuracy due to the limited feature set present in the input data, which fails to provide adequate learning resources for the model. Furthermore, we observed that while a larger input size generally contains more features, the experimental accuracy does not exhibit a strict linear relationship with the input size. In cases where the input size becomes excessively large, the experimental accuracy does not improve further.

By comparing the accuracy of each group, it can be clearly observed that for ResNet 50, when the input size is 224×224, the testing set accuracy reaches up to 99.59%, indicating that ResNet 50 is more suitable for an input size of 224×224. For Mobile V3, the testing set accuracy achieves 99.32% with an input size of 256×256, suggesting that Mobile V3 is better suited for input sizes of 256×256. Among the four models compared, the testing set accuracy obtained by inputting a feature map of 224×224 size into ResNet 50 is the highest. Consequently, ResNet 50 shows more excellent performance than the other three network models. Additionally, VGG-16 does not exhibit a significant advantage over Mobile V3, while Alex performs the worst.

IV. C. SE Experiment

In Experiments 4.1 and 4.2, despite the excellent performance achieved by ResNet 50 when the GAF-MTF encoded feature map was used as input, the accuracy rate exhibited significant fluctuations during training, which substantially reduced the convergence rate of ResNet 50. After reviewing relevant literature [28]-[30], it was found that this phenomenon is not accidental. However, the specific cause has not been explicitly addressed by researchers. This issue may be attributed to the architecture of ResNet 50, where the skip connections in its structure can lead to gradient superposition effects during backpropagation, potentially causing uneven distribution of gradient magnitudes across different layers. Furthermore, ResNet 50 demonstrates sensitivity to weight initialization, which could also contribute to its instability during training. The testing set accuracy of ResNet 50 in the training process is shown in Figure 10, and the change of loss function is shown in Figure 11.

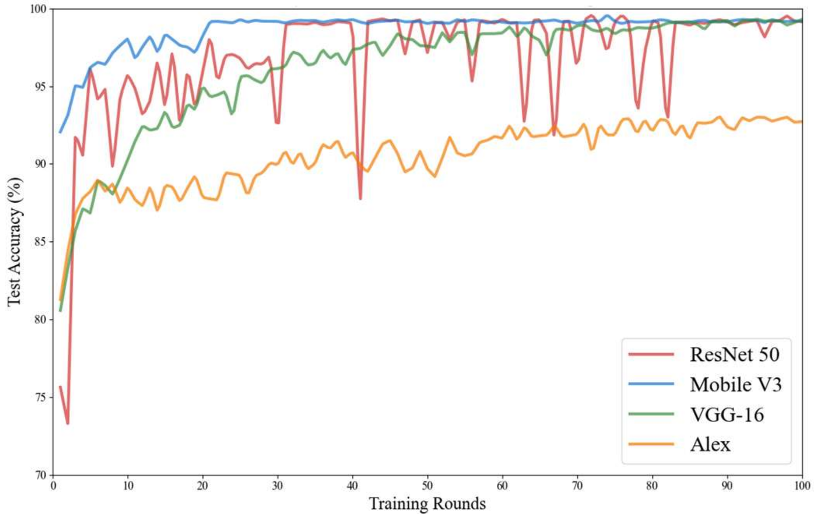


Figure 10: Fluctuations in testing set accuracy during the training process

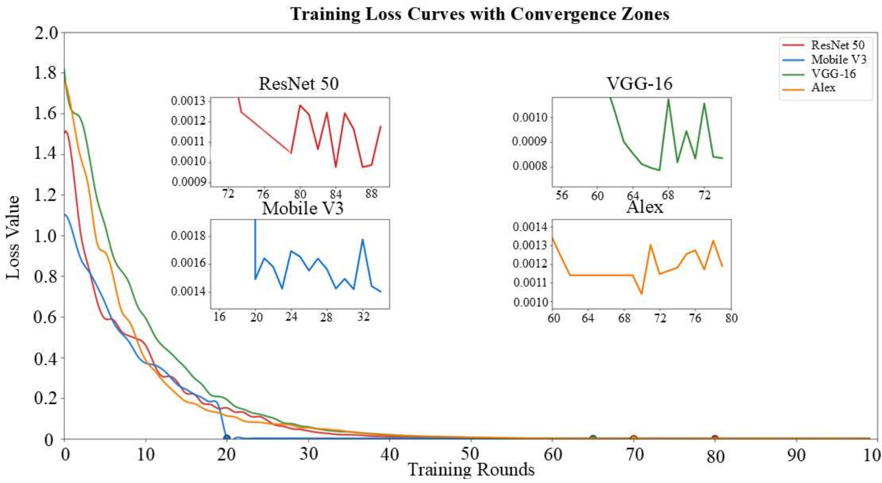


Figure 11: Fluctuations in testing set loss values during the training process

In Figure 10 and Figure 11, the x-axis depicts the number of training rounds, and the y-axis shows the testing set accuracy and loss values respectively. Apart from Alex, the other three network models demonstrate satisfactory performance. Mobile V3 achieves an accuracy of 92% in the first training round and converges earliest, with the testing set accuracy exceeding 99% after 20 training rounds. VGG-16 exhibits an initial training accuracy greater than 80%, though its convergence rate is relatively slow, requiring approximately 65 training rounds to converge. ResNet 50 performs the worst in the first training round, even underperforming compared to Alex, however, it demonstrates strong learning capability, with the testing set accuracy increasing rapidly. Nevertheless, the training process experiences significant fluctuations, indicating instability in its learning behavior. Convergence is achieved after approximately 80 training rounds. Despite the inconsistent learning behavior of ResNet 50, its overall performance surpasses the other three networks after 100 training rounds.

To accelerate the convergence and optimize the performance of the ResNet 50 network, the SE module is incorporated into the design. This addition ensures a stable learning process for ResNet 50 while improving its learning capacity. The SE module comprises two primary stages: squeeze and excitation. During the excitation stage, the hyperparameter r is introduced to reduce the dimensionality of the intermediate layer, thus reducing the quantity of model parameters and computational intricacy. Additionally, the value of r plays a crucial role in influencing channel weights, and variations in the weight vector S significantly impact the final output.

In this experiment, to investigate the impact of SE on ResNet 50 and analyze the differences among various r values, three distinct r values (r=8, r=16, r=32) were chosen for verification. The CWRU bearing dataset, encoded using GAF-MTF, was utilized as the input feature map. The feature map had a size of 224×224. Detailed data parameters are described in Section 4.1. 80% of the input data was assigned to training, with 20% reserved for testing. The outcomes are displayed in Figure 12.

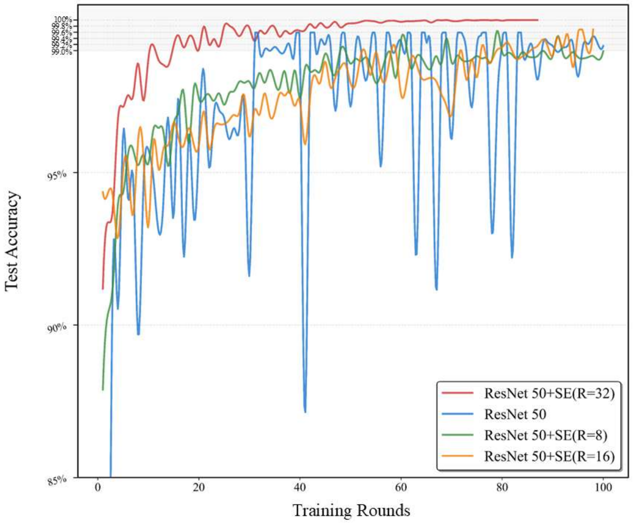


Figure 12: Model test results based on different *r* values

The results indicate a significant enhancement in the model's stability with the inclusion of the SE module. Specifically, when r=8, the accuracy of the initial model training exceeds 85%. For r=16 and r=32, the initial training accuracy surpasses 90%, in comparison to the ResNet 50 model without the SE module, the highest testing set accuracy improves by 0.1% and 0.28%, respectively. Notably, when r=32, the training set accuracy reaches 100% and testing set accuracy reaches 99.87%, with the loss function converging to 0.0005. Additionally, the model effectively learns the data features, enabling it to accurately diagnose faults.

To get a more lucid comprehension of the function of residual layers within the model, we utilized the t-SNE approach to examine the characteristics of each residual layer. Through t-SNE, the feature dimension was shrunk down to 2D, as depicted in Figure 13. t-SNE is a nonlinear method for decreasing the dimensions of data, transforming it from a high-dimensional state into a lower-dimensional representation suitable for visualization and clustering. The t-SNE algorithm focuses on maintaining local neighborhood relationships during dimensionality reduction, ensuring that originally adjacent high-dimensional data points remain clustered in low-dimensional embeddings. Visualization analysis allows researchers to evaluate model classification accuracy, detect class boundary overlaps, and identify clustering errors for targeted optimization.

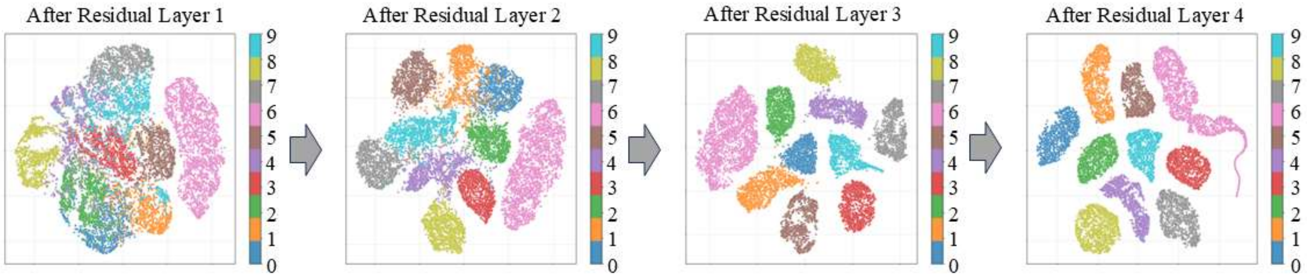


Figure 13: Visualized features through 4 residual layers

As can be seen from Figure 13. After traversing residual layer 1, the fault features remain hard to differentiate, indicating the need to deepen the depth of the residual blocks. After passing through three residual layers, the distinguishability of the fault features has improved significantly, but the clustering effect of the scattered points still needs to be improved. After undergoing treatment with residual layer 4, obvious, discriminative, and stable features were extracted.

To verify the feasibility of the GAF-MTF-SE-ResNet 50 method, the testing set's confusion matrix is depicted in Figure 14. The confusion matrix serves as a graphic representation that juxtaposes the model's forecasted outcomes against the actual classifications across various categories. The numbers along the diagonal indicate the quantity of instances that have been accurately categorized.

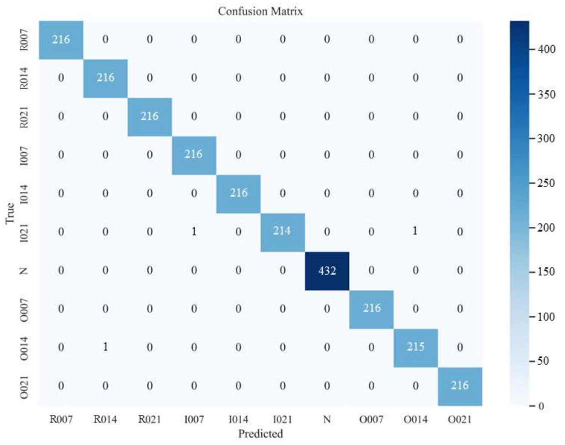


Figure 14: Confusion matrix for the testing set

The testing set's confusion matrix shows that GAF-MTF-encoded features perform robustly in the SE-ResNet 50 framework. Among the 2,376 test samples, only three samples exhibited prediction errors: two errors in the fault prediction for inner 007 and one error in the fault prediction for outer 014. This suggests that the diagnostic capability of the model for these two specific fault types requires further enhancement. The overall prediction accuracy reaches an impressive 99.87%.

IV. D. Performance of the Proposed Method in other Datasets and Comparison with other Models

To assess the generalization ability of the GAF-MTF encoding approach when combined with SE-ResNet 50 on other datasets, the bearing dataset from Jiangnan University was employed for validation. The signal used for encoding is generated by the experimental device for Mitsubishi SB-JR induction motor, rated power 3.7kW, voltage 220 V, quadrupole, rated speed 1800 rpm. Two roller bearings supported the rotor. In the experiment, faults were artificially introduced via wire cutting, including outer fault, inner fault, and roller fault. Specifically, N205 bearings were employed for normal operation, outer fault, and roller fault scenarios, while NU205 bearings (with separable outer rings) were used for inner fault case. Sampling experiments were conducted at three different speeds, with a sampling frequency of 50 kHz and a duration of 20 seconds using an acceleration transducer. The signal collected at 600 rpm was selected for GAF-MTF encoding to generate feature images.

In this experiment, after the timing signal undergoes GAF-MTF encoding, in the size of 224×224, the sliding sample rate is 50%, modify the full connection layer to achieve a four-class classification output. The output is shown in Figure 15.

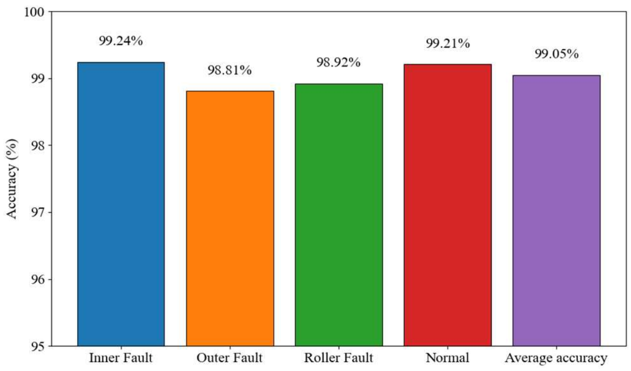


Figure 15: Testing set accuracy based on Jiangnan university dataset

The figure illustrates the diagnostic capability of SE-ResNet 50 after 100 epochs of training. The diagnostic accuracy rates are as follows: 99.24% for inner fault, 98.81% for outer fault, 98.92% for roller fault, and 99.21% for normal bearing conditions. The average accuracy rate is 99.05%. These results indicate that SE-ResNet 50 exhibits robust performance across various datasets and demonstrates strong generalization capabilities.

To further clarify the benefits of the proposed model, GAF-MTF-SE-ResNet 50 was compared with several commonly employed methods, and the comparison is shown in Table 6.

Table 6: Comparison of different bearing fault diagnosis methods

Methods	Categories	Accuracy
MTF-ResNet34 [31]	10	98.52%
ResNet-LSTM [32]	10	98.95%
CNNEPDNN [33]	10	98.10%
Improved 1D LetNet-5 network [34]	10	99.66%
Proposed Method	10	99.87%
	4	99.05%
VI-CNN [35]	4	100%
Spark-IRFA [36]	4	98.12%

While the proposed approach may not be optimal, GAF-MTF transformations have proven effective in capturing bearing fault features. The proposed method considers the performance, classification accuracy, and generalization capability of the model. This demonstrates that the innovative GAF-MTF encoding method can be effectively integrated into the SE-ResNet 50 model, thereby achieving the objective of fault diagnosis across diverse environments.

V. Conclusions

In this study, a bearing fault diagnosis method utilizing GAF-MTF for image encoding was introduced, ResNet 50 as the backbone model, and the SE attention mechanism to enhance the model's overall performance. Specifically, the method employs GADF, GASF, and MTF technique to convert timing signals into 2D image representations. These generated images are then converted into grayscale images, and their grayscale values are utilized to construct new RGB images for training the SE-ResNet 50 model. To investigate the effects of different encoding methods, input image sizes, and parameter variations in the SE module on the model's properties and generalization capability, a sequence of experiments was devised. Additionally, this study applies the t-SNE method to analyze the functions of key layers within the model and visualize the classification process. Finally, the proposed method is evaluated against methods outlined in prior studies. The findings reveal that the proposed method delivers top-notch results in both 10-class and 4-class categorization challenges on different datasets, boasting accuracy rates of 99.87% and 99.05%, respectively. This demonstrates the superior diagnostic performance of the proposed method over current approaches. In this study, the encoding of timing signals into 2D images was improved, providing an effective solution for multi-dimensional feature extraction by the model. In addition, the SE module improves the stability, generalization ability and accuracy of ResNet 50. Notably, the SE module is the sole attention mechanism employed in this work for enhancing ResNet 50. However, due to the increased depth and complexity introduced by the attention mechanism, the SE-ResNet 50 model may experience degraded performance and significantly reduced training speed when handling larger and more complex datasets. Therefore, the selection of an appropriate attention mechanism, the enhancement of model performance, and the achievement of model lightweighting will be key focuses in future research.

Author Contributions

Conceptualization, T.S. and Y.Q.; methodology, T.S. and Z.P.; software, T.S. and Y.Q.; validation, Z.T. and H.Q.; formal analysis, Z.T. and Z.P.; resources, C.R. and Z.M.; data curation, H.Q.; writing—original draft preparation, T.S. and C.R.; writing—review and editing, Z.P. and Z.T.; visualization, T.S.; supervision, Z.M.; project administration, T.S. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Natural Science Foundation of Shandong Province (grant number: ZR2024QE358)

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Case Western Reserve University Bearing Data https://engineering. case.edu/bearingdatacenter. Jiangnan University Bearing Data https://github.com/ClarkGableWang/JNU-Bearing-Dataset.

Acknowledgments

Not applicable.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Zhang, H.; Zhang, C.; Wang, C.; Xie, F. A survey of non-destructive techniques used for inspection of bearing steel balls. Measurement 2020, 159, 107773.

[2] Guo, S.; Yang, T.; Gao, W.; Zhang, C. A Novel Fault Diagnosis Method for Rotating Machinery Based on a Convolutional Neural Network. Sensors 2018, 18, 1429.

[3] Yang, J.; Wang, Z.; Guo, Y.; Gong, T.; Shan, Z. A novel noise-aided fault feature extraction using stochastic resonance in a nonlinear system and its application. IEEE Sens J. 2024, 24, 11856-11866.

[4] Liu, S.; Yin, J.; Hao, M.; Liang, P.; Zhang, Y.; Ai, C.; Jiang, W. Fault diagnosis study of hydraulic pump based on improved symplectic geometry reconstruction data enhancement method. Adv Eng Inform. 2024, 61, 102459.

[5] Zhang, Z.; Wang, J.; Li, S.; Han, B.; Jiang, X. Fast nonlinear blind deconvolution for rotating machinery fault diagnosis. Mech Syst Signal Process. 2023, 187, 109918.

[6] Yang, B.; Lei, Y.; Li, X.; Roberts, C. Deep targeted transfer learning along designable adaptation trajectory for fault diagnosis across different machines. IEEE Trans Ind Electron. 2023, 70, 9463–9473.

[7] Wang, Z.; He, X.; Yang, B.; Li, N. Subdomain adaptation transfer learning network for fault diagnosis of roller bearings. IEEE Trans Ind Electron. 2022, 69, 8430–8439.

[8] Liang, P.; Yu, Z.; Wang, B.; Xu, X.; Tian, J. Fault transfer diagnosis of rolling bearings across multiple working conditions via subdomain adaptation and improved vision transformer network. Adv Eng Inform. 2023, 57, 102075.

[9] Liu, Y.; Zhang, J.; Bi, F.; Lin, J.; Ma, W. A fault diagnosis approach for diesel engine valve train based on improved ITD and SDAG-RVM. Meas. Sci. Technol. 2015, 26, 025003.

[10] Chen, B.; He, Z.; Chen, X.; Cao, H.; Cai, G.; Zi, Y. A demodulating approach based on local mean decomposition and its applications in mechanical fault diagnosis. Meas. Sci. Technol. 2011, 22, 055704.

[11] Wang, Z.; Xu, X.; Zhang, Y.; Wang, Z.; Li, Y.; Liu Z.; Zhang, Y. A Bearing Fault Diagnosis Method Based on a Residual Network and a Gated Recurrent Unit under Time-Varying Working Conditions. Sensors 2023, 23, 6730.

[12] Nayana, B.; Geethanjali, P. Analysis of Statistical Time-Domain Features Effectiveness in Identification of Bearing Faults From Vibration Signal. IEEE Sens J. 2017, 17, 5618-5625.

[13] Tian, C.; Zheng, M.; Zuo, W.; Zhang, B.; Zhang, Y.; Zhang, D. Multi-stage image denoising with the wavelet transform. Pattern Recognit. 2023, 134, 109050.

[14] Rikam, L.; Bitjoka, L.; Nketsa, A. Quaternion Fourier Transform spectral analysis of electrical currents for bearing faults detection and diagnosis. Mech. Syst. Sig. Process. 2022, 168, 108656.

[15] Sharma, A.; Jigyasu, R.; Mathew, L.; Chatterii, S. Bearing fault diagnosis using weighted K-nearest neighbor. 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 11-12 May 2018; pp.1132-1137.

[16] Sinitsin, V.; Ibryaeva, O.; Sakovskaya, V.; Eremeeva, V. Intelligent bearing fault diagnosis method combining mixed input and hybrid CNN-MLP model. Mech. Syst. Sig. Process. 2022, 180, 109454.

[17] Zhou, J.; Xiao, M.; Niu, Y.; Ji, G. Rolling bearing fault diagnosis based on WGWOA-VMD-SVM. Sensors 2022, 22, 6218.

[18] Amarnath, M.; Sugamaran, V.; Kumar, H. Exploiting sound signals for fault diagnosis of bearings using decision tree. Measurement 2013, 46, 1250-1256.

[19] Lou, X.; Loparo, K. Bearing fault diagnosis based on wavelet transform and fuzzy inference. Mech. Syst. Sig. Process. 2004, 18, 1077-1095.

[20] Wang, X.; Mao, D.; Li, X. Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network. Measurement 2021, 173, 108518.

[21] Fu, W.; Jiang, X.; Tan, C.; Li, B.; Chen, B. Rolling bearing fault diagnosis in limited data scenarios using feature enhanced generative adversarial networks. IEEE Sens J. 2022, 22, 8749-8759.

[22] Gunerkar, R.S.; Jalan, A.K. Classification of ball bearing faults using vibro-acoustic sensor data fusion. Exp. Tech. 2019, 43, 635-643.

[23] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016; pp. 770-778.

[24] Hou, S.; Lian, A.; Chu, Y. Bearing fault diagnosis method using the joint feature extraction of Transformer and ResNet. Meas. Sci. Technol. 2023, 34, 075108.

[25] He, K.; Xu, Y.; Wang, Y.; Wang, J.; Xie, T. Intelligent diagnosis of rolling bearings fault based on multisignal fusion and MTF-ResNet. Sensors 2023, 23, 6281.

[26] Rangel, G.; Cuevas-Tello, J.; Nunez-Varela, J.; Puente, C.; Silva-Trujillo, A. A survey on convolutional neural networks and their performance limitations in image recognition tasks. J. Sens. 2024, 1,2797320.

[27] Liang, P.; Wang, W.; Yuan, X.; Liu, S.; Zhang, L.; Cheng, Y. Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment. Eng. Appl. Artif. Intell. 2022, 115, 105269.

[28] Gu, X.; Xie, Y.; Tian, Y.; Liu, T. A lightweight neural network based on GAF and ECA for bearing fault diagnosis. Metals 2023, 13, 822.

[29] Wen, L.; Li, X.; Gao, L. A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural. Comput. Appl. 2020, 32, 6111-6124.

[30] Abid, M.; Haq, I.ul.; Cai, Z.; Zhang, S. Faults identification with deep CNN fine-tuned ResNet50 model for rolling bearings. IET Conference Proceedings CP835. Stevenage, UK: The Institution of Engineering and Technology, 2023, 2023(9): 533-539.

[31] Yan, J.; Kan, J.; Luo, H. Rolling bearing fault diagnosis based on Markov transition field and residual network. Sensors 2022, 22, 3936.

[32] Wang, Y.; Cheng, L. A combination of residual and long–short-term memory networks for bearing fault diagnosis based on time-series model analysis. Meas. Sci. Technol. 2020, 32, 015904.

[33] Li, H.; Huang, J.; Ji, S. Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network. Sensors 2019, 19, 2034.

[34] Wan, L.; Chen, Y.; Li, H.; Li, C. Rolling-element bearing fault diagnosis using improved LeNet-5 network. Sensors 2020, 20, 1693.

[35] Hoang, D.T.; Kang, H.J. Rolling element bearing fault diagnosis using convolutional neural network and vibration image. Cognit. Syst. Res. 2019, 53, 42-50.

[36] Wan, L.J.; Gong, K.; Zhang, G.; Yuan, X.P.; Li, C.Y.; Deng, X.J. An Efficient Rolling Bearing Fault Diagnosis Method Based on Spark and Improved Random Forest Algorithm. IEEE Access 2021, 9, 37866–37882.