Mathematical Biosciences and Engineering

MBE, 19(12): 14086–14101.
DOI: 10.3934/mbe.2022656
Received: 20 August 2022
Revised: 09 September 2022
Accepted: 20 September 2022
Published: 26 September 2022

http://www.aimspress.com/journal/MBE

Research article

Bearing fault diagnosis based on Gramian angular field and DenseNet

Yajing Zhou, Xinyu Long*, Mingwei Sun* and Zengqiang Chen

College of Artificial Intelligence, Nankai University, Tianjin 300350, China

* **Correspondence:** Email: longxinyu@mail.nankai.edu.cn, sunmw@nankai.edu.cn.

Abstract: Rolling bearings are the core components of mechanical and electrical systems. A practical fault diagnosis scheme is the key to ensure operational safety. There are excessive characteristic parameters with remarkable randomness and severe signal coupling in the rolling bearing operation, which makes the fault diagnosis to be challenging. To deal with this problem, the Gramian angular field (GAF) and DenseNet are combined to perform feature extraction and fault diagnosis. The GAF can convert 1-dimensional time series into an image, which can guarantee the completeness of feature information without temporal dependence. The GAF images are then trained by using the DenseNet to generate a data set network. In this process, the transfer learning (TL), which can solve the problem of insufficient samples, is integrated to the DenseNet to enhance its extensibility. The comparative simulations are carried out to illustrate the effectiveness of the proposed method.

Keywords: fault diagnosis; Gramian angular field (GAF); DenseNet; transfer learning (TL)

1. Introduction

With the development of automation, electromechanical systems play an important role in modern industry [1,2]. Steering gear and rolling bearing are the core components of the mechanical and electrical systems. Once there is a fault, it might cause economic costs and even serious accidents [3–5]. Hence, a practical fault diagnosis scheme is a key to improve the reliability and safety. During the past several decades, Mechanical Failure Prevention Group (MFPG) and Mechanical Health Monitor Center (MHMC) have been established for fault diagnosis and prediction. Simultaneously, the mechanism of fault diagnosis is also investigated [6,7]. However, the operation of electromechanical system is a dynamic process, and the existing fault detection methods are mostly static. It is difficult to use the vibrational data directly for the rolling bearing fault diagnosis.

To solve this problem, it is recommended that the features should be extracted from the vibrational data before the fault diagnosis. Ramachandran et al. [8] proposed a proximal support vector machine (PSVM) for fault diagnosis, which uses decision tree to select the best features. In [9], the SVM, naive Bayes and K-nearest neighbor were combined to extract the envelope spectrum of current as the feature. Nishchal et al. [10] employed unsupervised learning sparse auto-encoder (SAE) to extract fault features, and then the SVM is used to diagnose the shafts and valves. In addition, the effectiveness was verified in comparison with the Mahalanobis distance fast classifier. However, modern electromechanical equipment usually has multi-monitoring points, long monitoring time and high sensor sampling frequency, which can collect massive data [11,12]. These factors intensify the difficulty of fault diagnosis. Specifically, the previous feature extraction methods based on expert knowledge are no longer suitable. Regarding this new scenario, a practical method, which can extract features automatically from massive data, is urgently needed.

In the past decade, the deep learning (DL), which has strong feature extraction ability, is developing rapidly [13]. The DL can extract the features from the original signal, which will greatly enhance the accuracy of fault diagnosis and prediction [14]. Ahmed et al. [15] introduced the artificial neural network into expert system, and the fault diagnosis accuracy can be improved. In order to optimize the fault diagnosis performance of sparse automatic encoder, Jayaswal et al. [16] utilized sparse auto-encoder depth network and backpropagation algorithm in the fine-tuning stage (the second stage), which can achieve high classification accuracy even from highly compressed measurements. In [17], a health monitoring method based on convolution network is used for fault diagnosis of bearings, wherein the data is transformed from the time domain to the frequency domain and then the data are fed into a convolution network to carry out training. By doing so, the accuracy can be increased by 6%. Kumbhar et al. [18] combined the adaptive neural fuzzy inference system (ANFIS) and dimensional analysis to conduct the fault diagnosis under different working conditions. Ajagekar et al. [19] applied quantum computing based DL to fault diagnosis, which is suitable for big data problem. In [20], a three-channel data set was established by integrating time/frequency/time-frequency domain information and a new transfer learning (TL) model was provided. Zhao et al. [21] combined dynamic wavelet weighting coefficients (DWWC) with planetary gearbox diagnosis based on the deep residual network, which can dynamically tune the weights and greatly improve the diagnosis performance. Wu et al. [22] integrated empirical wavelet transform, fuzzy entropy and SVM to diagnose the motor bearing fault. Here, the vibrational signal is decomposed into amplitude and frequency modulation components to calculate the fuzzy entropy, then the faults can be identified by using the SVM. Yu et al. [23] utilized denoising autoencoder and elastic network to denoise the signal, and then the sparse exponential discriminant analysis was used to identify the fault neurons. By doing so, the relevant fault variables on each fault neuron can be separated. A deep enhanced fusion network (DEFN) was proposed for the fault diagnosis in [24], wherein three sparse auto-encoders are applied to extract deep features from three-axial vibrational signals, respectively. By using a feature enhancement mapping, the fused three-axis features are then fed into an echo state network for fault classification. Yang et al. [25] proposed a fault diagnosis method that combines hierarchical symbolic analysis with convolutional neural network (CNN). The initial feature extraction and automatic feature learning are implemented by using a simplified network structure. In summary, the above mentioned methods have good performance in fault diagnosis under "Big Data". In practice, the useful measurements are insufficient. In addition, the CNN has remarkable capability in image recognition, classification, target detection and other areas, however, it will encounter difficulties in dealing with 1-dimensional time series.

Motivated by these previous investigations, the bearing fault diagnosis by using the Gramian angular field (GAF) and the DenseNet is proposed in this paper. The GAF can transform the 1-dimensional time series into a 2-dimensional image, which can maintain the time-dependence. Then, the converted 2-dimensional images can train a DenseNet, which can extract the image features to be used in the fault detection. In order to deal with insufficient samples, the TL is combined with the DenseNet to enhance the accuracy and effectiveness of the training model.

The rest of this paper is organized as follows. Section 2 presents the visualization of time series by using the GAF. The DenseNet based fault diagnosis is provided in Section 3. The data sets are preprocessed in Section 4. Section 5 offers the simulation results. The concluding remarks are given in Section 6.

2. Visualization of time series

The GAF, which maintains the time dependency, can transform a 1-dimensional time series into a 2-dimensional image [26]. The polar coordinate is used to represent the time series, and then the sum/difference trigonometric function are calculated. The Gramian angular summation field (GASF) and Gramian angular difference field (GADF) represent the sum and difference of two corners, respectively.

Consider a time series $X = \{x_1, x_2, \cdots, x_n\}$. The normalization can map each element onto [-1, 1] and [0, 1] as

$$\tilde{x}_i = \frac{x_i - x_{min}}{xmin_{max}} \tag{1}$$

$$\widetilde{x_i} = \frac{(x_i - x_{min}) + (x_i - x_{max})}{x_{max} - x_{min}} \tag{2}$$

where x_{min} and x_{max} are the minimum and maximum values of the data sequence, respectively. Time series can be expressed in the polar coordinates by using the angle φ_i and the radius r_i, $(i = 1, 2, \cdots, n)$. φ_i preserves the absoluteness of time relations, r_i is represented by a timestamp t_i, which guarantees the time dependence. They have the forms of

$$\begin{cases} \varphi_i = arccos(\tilde{x}_i), -1 \le \tilde{x}_i \le 1 \\ r_i = \frac{t_i}{n} \end{cases} \tag{2}$$

when the time series is mapped onto [-1, 1], φ_i is mapped onto [0, π]. n is the constant factor of the regularized polar coordinate system, [0, 1] can be divided into n equivalent subintervals. Except the initial point 0, we can obtain n segmentation points, which are associated with $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n\}$.

Due to the monotonicity of the function, each sequence has a unique polar mapping. The GAF matrixes are constructed by using the sum and difference formulae of

$$GASF=\begin{bmatrix} cos(\varphi_1+\varphi_1) & \cdots & cos(\varphi_1+\varphi_n) \\ cos(\varphi_2+\varphi_1) & \cdots & cos(\varphi_2+\varphi_n) \\ \vdots & \vdots & \vdots \\ cos(\varphi_n+\varphi_1) & \cdots & cos(\varphi_n+\varphi_n) \end{bmatrix} = \tilde{X}^T \times \tilde{X} - \sqrt{I-\tilde{X}^2}^T \times \sqrt{I-\tilde{X}^2} \quad (4)$$

$$GADF=\begin{bmatrix} sin(\varphi_1-\varphi_1) & \cdots & sin(\varphi_1-\varphi_n) \\ sin(\varphi_2-\varphi_1) & \cdots & sin(\varphi_2-\varphi_n) \\ \vdots & \vdots & \vdots \\ sin(\varphi_n-\varphi_1) & \cdots & sin(\varphi_n-\varphi_n) \end{bmatrix} = \sqrt{I-\tilde{X}^2}^T \cdot \tilde{X} - \tilde{X}^T \cdot \sqrt{I-\tilde{X}^2} \quad (5)$$

where, I = [1, 1, , 1]. From the above equations, we can see that the matrix elements move from the upper left to the lower right over time, the time dimension is encoded into the geometry of the matrix, and the matrix diagonal elements are single angle values, therefore, the approximate reconstruction of time series can be achieved. However, the expansion of matrix size will increase the computational complexity. To solve this problem, the piecewise aggregation approximation is introduced, which can not only maintain the sequence tendency but also can greatly reduce the sequence size.

Take $y = sin\, t$ as an example, which is a clean signal without noise, wherein $t \in [0, 4\pi]$ and the number of sampling points $n = 400$ (in Figure 1). The time series are converted into 2-dimensional images by using the GAF (in Figure 2).

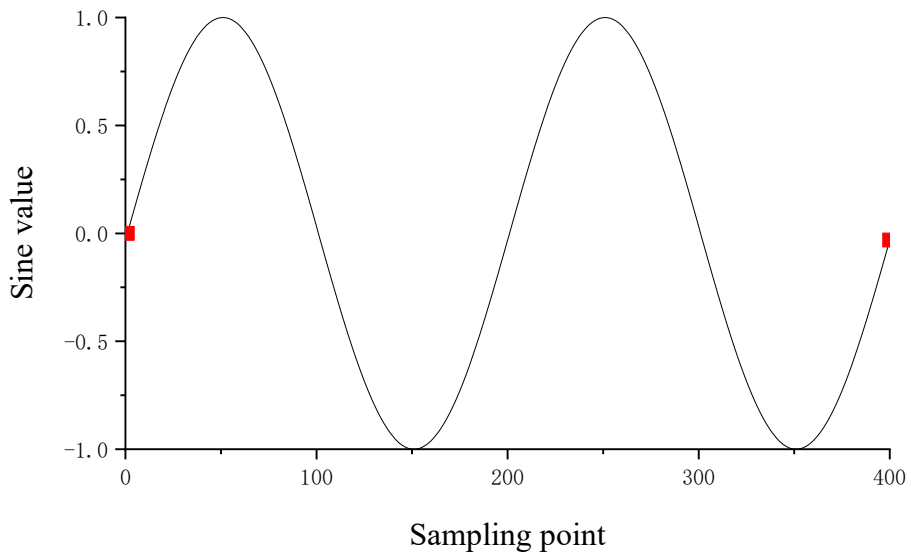


Figure 1. Image of sine sequence signal.

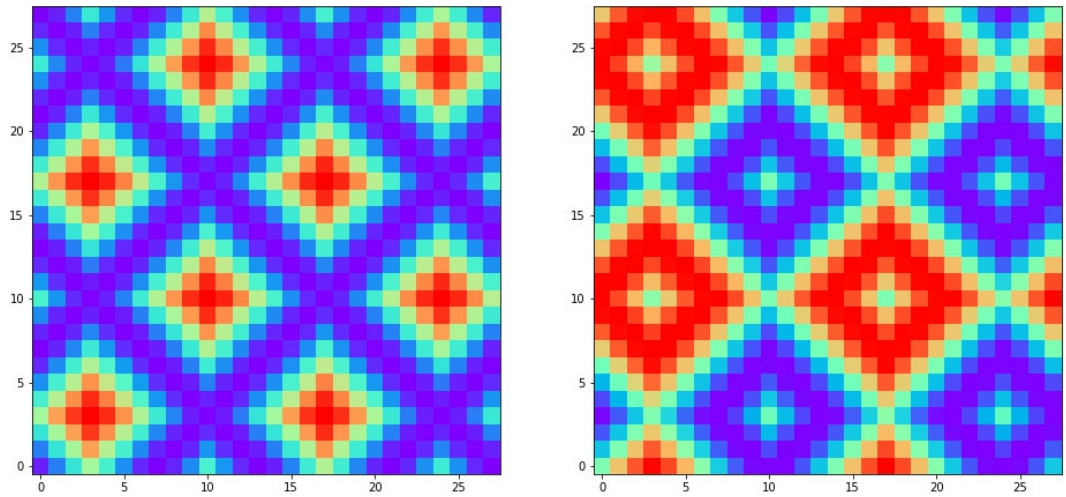


Figure 2. Mapping images of the sine sequence signal by using GAF. (a) GASF, (b) GADF.

In order to realize the reconstruction of the original time series and avoid the loss of characteristic information, the GAF is used to preprocess the original signal. It can clearly show the difference of data features, which is helpful for the subsequent image recognition and classification. In fact, other approaches might also be suitable for the conversion from 1-dimensional signals into a 2-dimensional image, we concentrate on the GAF to perform the mission in this paper.

3. Fault diagnosis based on DenseNet

3.1. DenseNet

DenseNet is a dense CNN, wherein the layers are pairwise connected. The outputs of each layer are the inputs of subsequent layers to maximize the characteristic information between these two layers. In addition, the gradient disappearance in the training process can also be alleviated. The DenseNet can reuse the features and the CNN architecture is easy to train. Each layer is designed to be very narrow when specifying the network structure, and only a few new features are needed to learn at a time. The 5-layer network of Figure 3 is shown as a classical model that has won the world award [27]. According to Figure 3, each layer maps the features as the characteristic inputs and pass them to other layers. For the bearing vibrational data $\boldsymbol{x}_0$, the output characteristic matrix $\boldsymbol{x}_l$ of the l layer module can be expressed as

$$\boldsymbol{x}_l = H_l([\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{l-1}]) \tag{6}$$

where, $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{l-1}$ are the outputs of each dense layer. H_l is a composite function, which includes batch normalization (BN), ReLU and 3×3 convolution function.

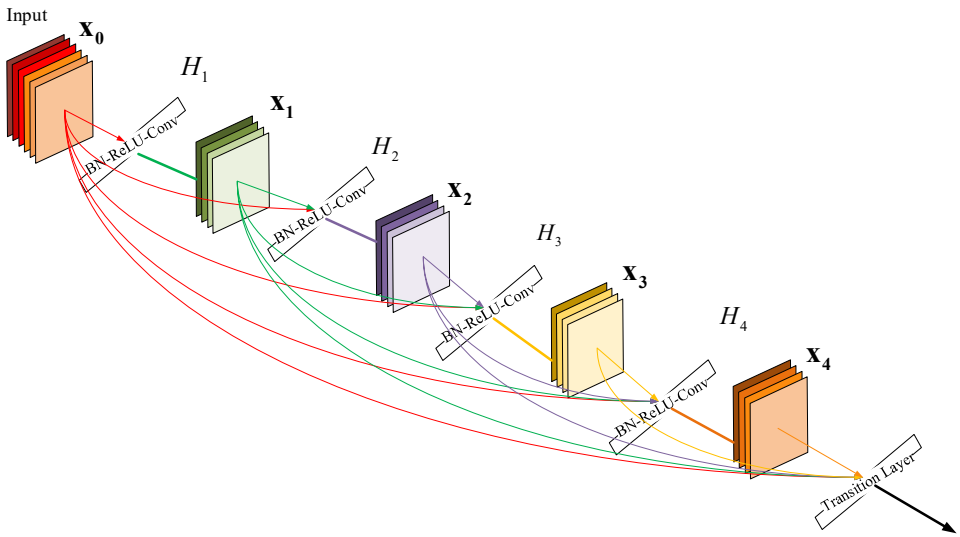


Figure 3. 5-layer dense.

When the feature size is changed, Eq (6) is no longer applicable. The pooling layer, which is included in the transition layer, is needed to reduce the dimensional size of the feature. Assuming that H_l generates additional k features, and k is also the DenseNet growth rate, then the l layer has $k_0 + k(l-1)$ features, where k_0 is the number of input features channels. Therefore, $k = 4$ in Figure 3.

The DenseNet has fewer parameters with small storage requirement and easier network training. Therefore, it is more suitable for feature extraction and fault diagnosis. However, CNN is a data-driven model, which requires a large number of labeled data to obtain satisfactory results, which leads to limitations in some applications.

3.2. Transfer learning

In this section, the TL is introduced to address the sample dependency of the DenseNet in training models. In combination of CNN and TL, the network framework trained on specific data sets can be applied to new problems and new fields that need to be solved. The TL can generalize the source field knowledge to solve the target field task [28–30]. The definition is as follows.

Definition 1. Given the source definition domain D_s, the learning task T_s, the object definition domain D_t and the learning task T_t. The knowledge of D_s and T_s are used to improve the results of object prediction function $f_t(\cdot)$, which is on D_t, wherein $D_s \neq D_t$ and $T_s \neq T_t$.

Remark 1. The domain $D = \{X, P(X)\}$, the task $T = \{Y, P(Y|X)\}$, $f(\cdot) = P(Y|X)$, therefore $D_s \neq D_t$ represents $X_s \neq X_t$ or $P_s(X) \neq P_t(X)$, the same as T.

In this paper, the source dataset is the CIFAR (Canadian Institute for Advanced Research), and the target dataset is the CWRU (Case Western Reserve University). The combination of TL and CNN implementation process is as follows:

a) the CIFAR dataset is used to train the random initialization parameters;

b) the trained network framework is applied to the specific problem to be solved, and the data features are extracted automatically;

c) the extracted features are input into the CNN to diagnose and classify bearing faults.

The combination of TL and CNN can greatly improve the accuracy of network fault diagnosis, reduce the complexity of network training, avoid the dependence of artificial parameters on expert experience. It can also handle the insufficient samples problem, reducing training time and memory consumption. At the same time, the inherent pattern of the network model is changed, which enhances the network model applicability. Therefore, we can train a general model with insufficient samples in the absence of supervision.

4. Data preprocessing

4.1. Data sets selection

The method is verified on the CWRU bearing data center, which has become the evaluation criterion for rolling bearing fault diagnosis [31,32].

The CWRU bearing experimental platform, which consists of a 2 horsepower (1.5 kw) motor, a torque sensor/decoder and a power tester, is shown in Figure 4 [31]. One motor end is connected to the fan end bearing, and the other end is connected to the drive end bearing. The acceleration sensor, which receives the bearing vibrational signal, is installed on each bearing. An electrical discharge machining (EDM) is used to cause inner/outer ring and rolling element faults and the faults diameters are 0.007, 0.014 and 0.021 inches, respectively. The single point fault of rolling bearing is reloaded to the test motor. The vibrational acceleration signal data, which have 0, 1, 2 and 3 horsepower loads, respectively, are collected by using a 16-channel data recorder. In addition, the sampling frequency is 12 kHz.

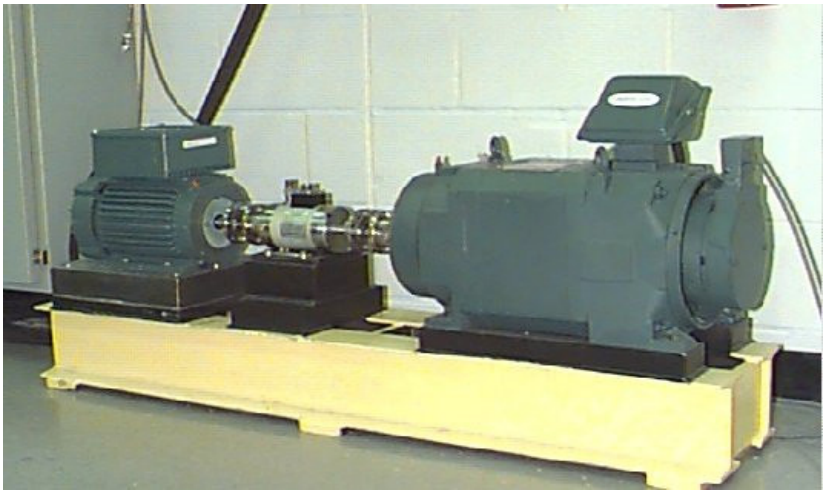


Figure 4. Bearing experimental platform of CWRU.

The samples include 8 normal samples, 23 inner ring fault samples, 53 outer ring fault samples and 11 rolling element fault samples. In this paper, we use 9 samples, which include three different fault diameters and locations, to train the CNN.

4.2. Data processing

In this paper, the samples of fan end without load, which composes 240,000 normal data and 120,000 fault data of each type, are used. The non-overlapping sliding window is selected, and the image size is 300×300. Figure 5 shows the GASF and GADF images by using the same samples.

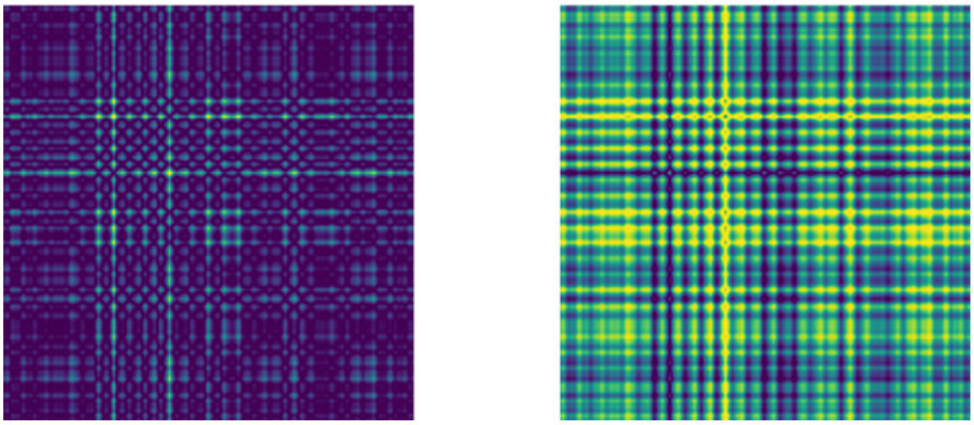


Figure 5. GAF image comparison.

According to Figure 5, the GADF image is easy to distinguish. Therefore, GADF image can be selected for subsequent fault diagnosis. The fault-free GADF image is numbered 0 (in Figure 6), the GADF images corresponding to the different faults are numbered from 1 to 9 (in Figure 7), respectively.

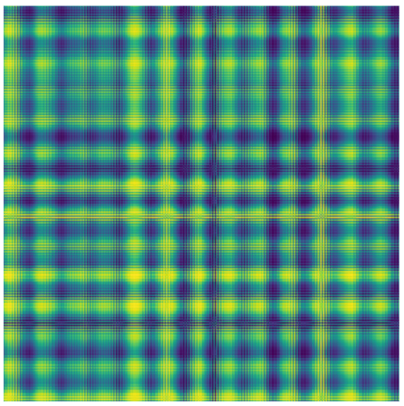


Figure 6. GADF image of normal bearing.

(a) RE, diameter is 0.007 (b) RE, diameter is 0.014 (c) RE, diameter is 0.021

(d) IR, diameter is 0.007 (e) IR, diameter is 0.014 (f) IR, diameter is 0.021

(h) OR, diameter is 0.007 (i) OR, diameter is 0.014 (g) OR, diameter is 0.021

Figure 7. 9 fault types expressed by using GADF.

5. Simulation results

The fault detection of the bearing is carried out by using the Tensorflow. The simulation platform is configured as follows: 64-bit Windows10 operating system, i7-10700k (CPU), NVIDIA Tesla P100 (GPU), the code is written in python3.7. The compilation software is Spyder and Jupyter Notebook. The Pytorch and Scikit-learn are used to train the CNN.

The analysis steps are as follows:

a) the 1-dimensional time series samples can be transformed into 2-dimensional images by using GAF;

b) the samples are divided, and then the CNN is trained. By changing different batch sizes, the results

of different optimization methods can be obtained;
c) the sample datasets are extended by different coincidence rate, and CNN is trained with different batch sizes of samples;
d) the 2/3 coincidence rate of the samples are selected to train the DenseNet121 and ResNet18 networks in combination with the TL.

5.1. Fault diagnosis results

The GADF images of bearings (Figures 6 and 7), which are divided into 70% training set and 30% test set, are put into the CNN with the structure parameters as shown in Table 1. The initial learning rate is 10^{-3}, the epochs is 5, and the batch sizes are 20, 10 and 2, respectively. Two optimizers, the stochastic gradient descent (SGD) and the Adams, are selected. Table 2 shows the training accuracy results, and the training loss values are shown in Table 3.

Table 1. The structure parameters of CNN.

Convolution layer	Input (N, C, W, H)	Convolution kernel (size, input, output)	Activation function	Output (N, C, W, H)
1	[128, 4, 300, 300]	[5 × 5, 4, 16]	ReLU	[128, 16, 150, 150]
2	[128, 16, 150, 150]	[5 × 5, 16, 32]	ReLU	[128, 32, 75, 75]
3	[128, 32, 75, 75]	[5 × 5, 32, 64]	ReLU	[128, 64, 36, 36]
4	[128, 64, 36, 36]	[5 × 5, 64, 64]	ReLU	[128, 64, 4, 4]

Table 2. Accuracy of the training model by using SGD and Adams.

Batch size	SGD	Adam
20	83.03%	73.26%
10	83.33%	72.58%
2	86.44%	71.46%

Table 3. Loss value of the training model by using SGD and Adams.

Batch size	SGD	Adam
20	0.5040	1.0498
10	0.5522	1.6858
2	0.5817	1.8479

According to Tables 2 and 3, the accuracy is not high enough and the training loss value cannot be reduced to a reasonable range. Although all the samples have been used for model training, the results are still not satisfactory. Therefore, the samples need to be expanded to obtain a useful training model.

5.2. Expanded sample validation

The repetition rate of the sliding window is set to 2/3, then the normal samples (Figure 8) and different fault samples (Figure 9) can come to 2398 and 9 × 1198, respectively.

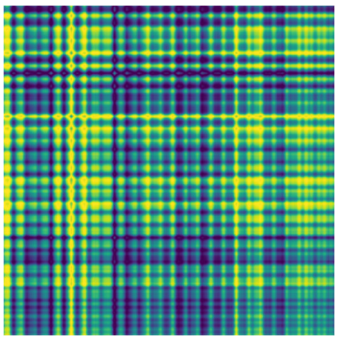
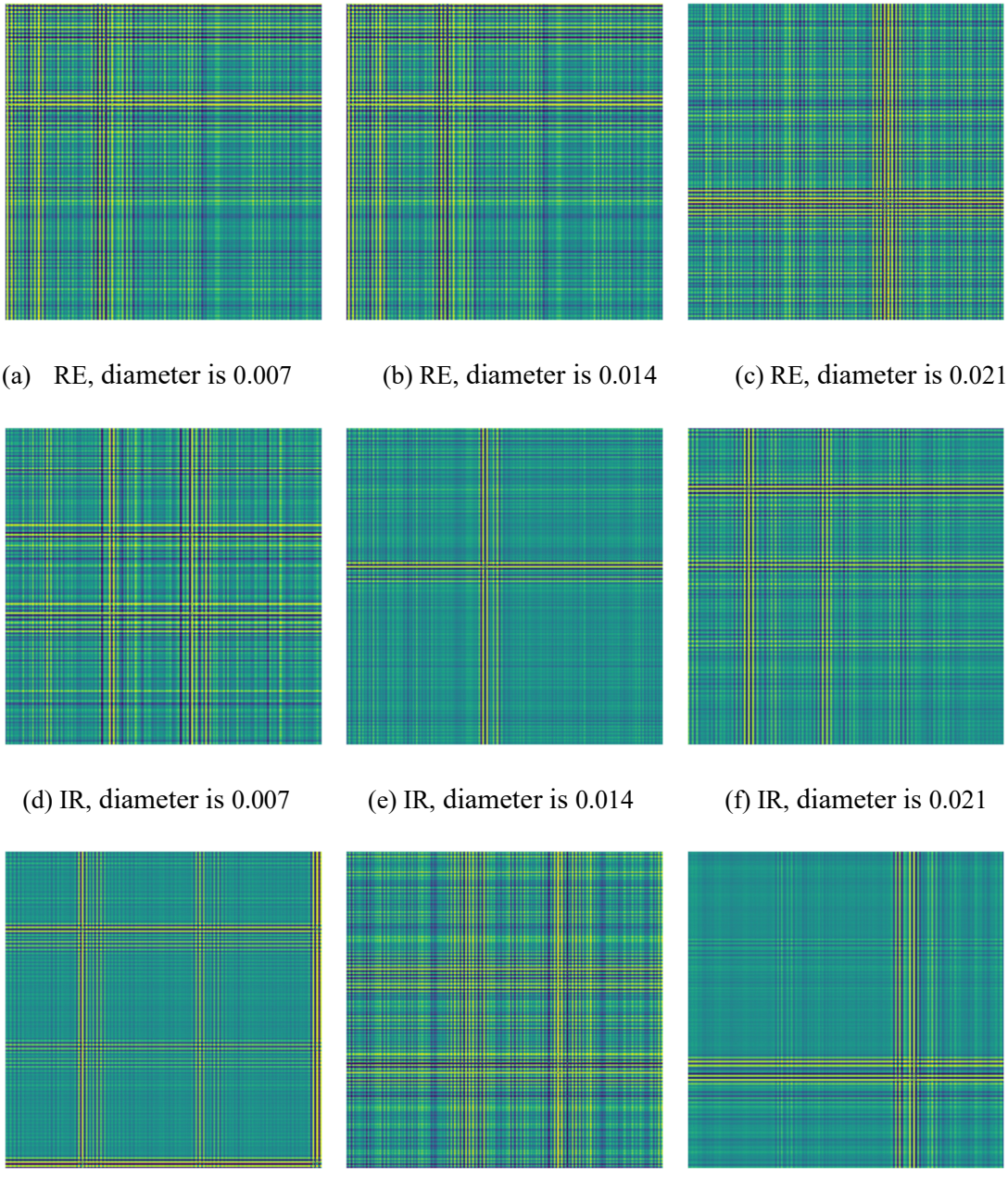


Figure 8. Extended GADF image of normal bearing.



(a) RE, diameter is 0.007 (b) RE, diameter is 0.014 (c) RE, diameter is 0.021

(d) IR, diameter is 0.007 (e) IR, diameter is 0.014 (f) IR, diameter is 0.021

(h) OR, diameter is 0.007 (i) OR, diameter is 0.014 (g) OR, diameter is 0.021

Figure 9. Extended GADF images of 9 fault types.

In addition, 4795 normal samples and 9 × 2395 different fault samples are obtained by setting 5/6 repetition rate. The CNN with the same structure is used for model training. The training accuracy and loss values for different batch sizes and repetition rates are shown in Tables 4 and 5, respectively.

Table 4. Accuracy of training model by using extended samples.

Batch size	Repetition rate: 2/3	Repetition rate: 5/6
20	92.10%	96.28%
10	91.69%	94.70%
2	92.65%	90.34%

Table 5. Loss function of training model by using extended samples.

Batch size	Repetition rate: 2/3	Repetition rate: 5/6
20	0.3478	0.1201
10	0.4082	0.1886
2	0.3655	0.3635

According to Tables 4 and 5, the model accuracy can be greatly enhanced and the loss function can also reach within an acceptable range when the number of samples increases. The best accuracy can reach 96.28% through the basic CNN. Also each index has reached at a reasonable range. However, further improvement is needed.

To achieve better results, the ResNet and the DenseNet are trained by using the extended GADF images, wherein the TL is employed. The structure parameters of the two networks are shown in Tables 6 and 7, respectively.

Table 6. Structure parameters of ResNet18.

Convolution layer	Parameters	Output
Conv 1	7 × 7, 64, step length = 2	112 × 112
Conv 2	$\begin{bmatrix} 3\times3, & 64 \\ 3\times3, & 64 \end{bmatrix}\times2$	56 × 56
Conv 3	$\begin{bmatrix} 3\times3, & 128 \\ 3\times3, & 128 \end{bmatrix}\times2$	28 × 28
Conv 4	$\begin{bmatrix} 3\times3, & 256 \\ 3\times3, & 256 \end{bmatrix}\times2$	14 × 14
Conv 5	$\begin{bmatrix} 3\times3, & 512 \\ 3\times3, & 512 \end{bmatrix}\times2$	7 × 7
Pooling layer	Average pooling	1 × 1

Table 7. Structure parameters of DenseNet121.

Layer	Parameter	Output Size
Convolution layer	7 × 7, conv, stride = 2	112 × 112

Pooling	3 × 3 max pooling, stride = 2	56 × 56
Block1	$\begin{bmatrix}1\times1\\3\times3\end{bmatrix}\times6$	56 × 56
Transition layer 1	1 × 1 conv	56 × 56
	2 × 2 average pooling, stride = 2	28 × 28
Block 2	$\begin{bmatrix}1\times1\\3\times3\end{bmatrix}\times12$	28 × 28
Transition layer 2	1 × 1 conv	28 × 28
	2 × 2 average pooling, stride = 2	14 × 14
Block 3	$\begin{bmatrix}1\times1\\3\times3\end{bmatrix}\times24$	14 × 14
Transition layer 3	1 × 1 conv	14 × 14
	2 × 2 average pooling, stride = 2	7 × 7
Block 4	$\begin{bmatrix}1\times1\\3\times3\end{bmatrix}\times16$	7 × 7
Taxonomic hierarchies	7 × 7 global average pooling	1 × 1
	A fully connected layer of 1000 dimensions, Softmax	

The results show that the training model accuracy can be as high as 99.69% (ResNet18) and 99.83% (DenseNet121), respectively; and the loss functions are also reduced to 0.013 (ResNet18) and 0.005 (DenseNet121). Compared with the existing basic CNN method, the augmented CNN approach based on expanded samples and the ResNet, the proposed scheme can achieve the best performance with regard to the accuracy and loss values.

6. Conclusions

In this paper, a bearing fault diagnosis method based on the GAF and the CNN, which could realize the classification diagnosis, was proposed. The GAF could convert time series information into a GADF image, which makes the fault diagnosis to be more intuitive. In addition, the GAF could also ensure the integrity of feature information and handle the time-dependent issue, which is suitable for the processing of non-stationary signals such as bearing vibration signals. The DenseNet was trained by using the GADF images, which could enhance the low classification accuracy caused by measurement noise in the original bearing signal. The TL, which could avoid the inaccurate training model caused by insufficient samples, was applied in the DenseNet. The simulation results demonstrate the effectiveness of the proposed method.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number 62073177, 61973175, 52175038).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Q. Ni, J. Ji, K. Feng, B. Halkon, A novel correntropy-based band selection method for the fault diagnosis of bearings under fault-irrelevant impulsive and cyclostationary interferences, *Mech. Syst. Sig. Process.*, **153** (2021), 107498. https://doi.org/10.1016/j.ymssp.2020.107498
2. R. L. Larson, R. C. Searle, M. C. Kleinrock, H. Schouten, R. T. Bird, D. F. Naar, et al., Roller-bearing tectonic evolution of the Juan Fernandez microplate, *Nature*, **356** (1992), 571–576. https://doi.org/10.1038/356571a0
3. T. Wang, Z. Liu, G. Lu, J. Liu, Temporal-spatio graph based spectrum analysis for bearing fault detection and diagnosis, *IEEE Trans. Ind. Electron.*, **68** (2021), 2598–2607. https://doi.org/10.1109/TIE.2020.2975499
4. A. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mech. Syst. Sig. Process.*, **20** (2006), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012
5. Q. Ni, J. Ji, K. Feng, B. Halkon, A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis, *Mech. Syst. Sig. Process.*, **164** (2022), 108216. https://doi.org/10.1016/j.ymssp.2021.108216
6. T. Wang, G. Lu, P. Yan, A novel statistical time-frequency analysis for rotating machine condition monitoring, *IEEE Trans. Ind. Electron.*, **67** (2020), 531–541. https://doi.org/10.1109/TIE.2019.2896109
7. S. Yang, Z. Liu, G. Lu, Early change detection in dynamical bearing degradation process based on hierarchical graph model and adaptive inputs weighting fusion, *IEEE Trans. Ind. Electron.*, **17** (2021), 3186–3196. https://doi.org/10.1109/TII.2020.3007653
8. V. Sugumaran, V. Muralidharan, K. Ramachandran, Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing, *Mech. Syst. Sig. Process.*, **21** (2007), 930–942. https://doi.org/10.1016/j.ymssp.2006.05.004
9. J. Wang, S. Liu, R. X. Gao, R. Yan, Current envelope analysis for defect identification and diagnosis in induction motors, *J. Manuf. Syst.*, **31** (2012), 380–387. https://doi.org/10.1016/j.jmsy.2012.06.005
10. N. K. Verma, V. K. Gupta, M. Sharma, R. K. Sevakula, Intelligent condition based monitoring of rotating machines using sparse auto-encoders, in *2013 IEEE Conference on Prognostics and Health Management (PHM)*, (2013), 1–7. https://doi.org/10.1109/ICPHM.2013.6621447
11. G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science*, **313** (2006), 504–507. https://doi.org/10.1126/science.1127647
12. B. Buxton, D. Goldston, C. Doctorow, M. Waldrop, Big data: science in the petabyte era, *Nature*, **455** (2008). Available from: https://www.mendeley.com/catalogue/3c1a840c-af73-3ada-a109-ac33fdead72e/
13. Y. Lei, H. R. Karimi, X. Chen, A novel self-supervised deep LSTM network for industrial temperature prediction in aluminum processes application, *Neurocomputing*, **502** (2022), 177–185. https://doi.org/10.1016/j.neucom.2022.06.080
14. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: efficient channel attention for deep

convolutional neural networks, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 11531–11539. https://doi.org/10.1109/CVPR42600.2020.01155
15. H. Ahmed, M. Wong, A. Nandi, Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features, *Mech. Syst. Sig. Process.*, **99** (2018), 459–477. https://doi.org/10.1016/j.ymssp.2017.06.027
16. P. Jayaswal, S. Verma, A. Wadhwani, Development of EBP-artificial neural network expert system for rolling element bearing fault diagnosis, *J. Vib. Control*, **17** (2011), 1131–1148. https://doi.org/10.1177/1077546310361858
17. O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, et al., Convolutional neural network based fault detection for rotating machinery, *J. Sound Vib.*, **377** (2016), 331–345. https://doi.org/10.1016/j.jsv.2016.05.027
18. S. G. Kumbhar, E. Sudhagar, An integrated approach of adaptive neuro-fuzzy inference system and dimension theory for diagnosis of rolling element bearing, *Measurement*, **166** (2020), 108266. https://doi.org/10.1016/j.measurement.2020.108266
19. A. Ajagekar, F. You, Quantum computing assisted deep learning for fault detection and diagnosis in industrial process systems, *Comput. Chem. Eng.*, **143** (2020), 107119. https://doi.org/10.1016/j.compchemeng.2020.107119
20. W. Mao, L. Ding, S. Tian, X. Liang, Online detection for bearing incipient fault based on deep transfer learning, *Measurement*, **152** (2020), 107278. https://doi.org/10.1016/j.measurement.2019.107278
21. M. Zhao, M. Kang, B. Tang, M. Pecht, Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes, *IEEE Trans. Ind. Electron.*, **65** (2018), 4290–4300. https://doi.org/10.1109/TIE.2017.2762639
22. D. Wu, S. Zhang, H. Zhao, X. Yang, A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing, *IEEE Access*, **6** (2018), 35042–35056. https://doi.org/10.1109/ACCESS.2018.2834540
23. W. Yu, C. Zhao, Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net, *IEEE Trans. Control Syst. Technol.*, **28** (2020), 1083–1091. https://doi.org/10.1109/TCST.2019.2897946
24. Z. Pu, C. Li, S. Zhang, Y. Bai, Fault diagnosis for wind turbine gearboxes by using deep enhanced fusion network, *IEEE Trans. Instrum. Meas.*, **70** (2021), 2501811. https://doi.org/10.1109/TIM.2020.3024048
25. Y. Yang, H. Zheng, Y. Li, M. Xu, Y. Chen, A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network, *ISA Trans.*, **91** (2019), 235–252. https://doi.org/10.1016/j.isatra.2019.01.018
26. V. Tran, B. Yang, F. Gu, A. Ball, Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis, *Mech. Syst. Sig. Process.*, **38** (2013), 601–614. https://doi.org/10.1016/j.ymssp.2013.02.001
27. G. Huang, Z. Liu, L. Maaten, K. Q. Weinberger, Densely connected convolutional networks, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 4700–4708. https://doi.org/10.1109/CVPR.2017.243

28. S. J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.*, **22** (2010), 1345–1359. https://doi.org/10.1109/TKDE.2009.191
29. K. Weiss, T. M. Khoshgoftaar, D. Wang, A survey of transfer learning, *J. Big Data*, **3** (2016), 9. https://doi.org/10.1186/s40537-016-0043-6
30. G. Wilson, D. J. Cook, A survey of unsupervised deep domain adaptation, *ACM Trans. Intell. Syst. Technol.*, **11** (2020), 1–46. https://doi.org/10.1145/3400066
31. K. A. Loparo, Case western reserve university bearing data center, *Bear. Vib. Data Sets, Case West. Reserve Univ.*, **2012** (2012), 22–28.
32. W. A. Smith, R. B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study, *Mech. Syst. Sig. Process.*, **64–65** (2015), 100–131. https://doi.org/10.1016/j.ymssp.2015.04.021

AIMS Press