ISA Transactions 145 (2024) 387–398

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Practice article

Fault anomaly detection method of aero-engine rolling bearing based on distillation learning

Yuxiang Kang [a], Guo Chen [b,*], Hao Wang [c], Jiajiu Sheng [a], Xunkai Wei [c]

[a] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b] College of General Aviation and Flight, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[c] Beijing Aeronautical Engineering Technical Research Center, Beijing 100076, China

ARTICLE INFO

Keywords:
Distillation learning
Rolling bearings
Anomaly detection
Vision transformer
Aero-engine

ABSTRACT

In this study, we address the issue of limited generalization capabilities in intelligent diagnosis models caused by the lack of high-quality fault data samples for aero-engine rolling bearings. We provide a fault anomaly detection technique based on distillation learning to address this issue. Two Vision Transformer (ViT) models are specifically used in the distillation learning process, one of which serves as the teacher network and the other as the student network. By using a small-scale student network model, the computational efficiency of the model is increased without sacrificing model accuracy. For feature-centered representation, new loss and anomaly score functions are created, and an enhanced Transformer encoder with the residual block is proposed. Then, a rolling bearing dynamics simulation method is used to obtain rich fault sample data, and the pre-training of the teacher network is completed. For anomaly detection, the training of the student network is completed based on the proposed loss function and the pre-trained teacher network, using only the vibration acceleration samples obtained from the normal state. Finally, the trained completed network and the designed anomaly score function are used to achieve the anomaly detection of rolling bearing faults. The experimental validation was carried out on two sets of test data and one set of real vibration data of a whole aero-engine, and the detection accuracy reached 100 %. The results show that the proposed method has a high capability of rolling bearing fault anomaly detection.

1. Introduction

Rolling bearings are a common general-purpose component found in a variety of rotating machinery and equipment, and their operating conditions often affect the accuracy, reliability, and life of the whole equipment. For instance, once a rolling bearing in an aero-engine fails, the cost of repairs and maintenance would rise, as will the likelihood of equipment failure or even casualties. According to statistics, dozens of aircraft air parking or forced landing incidents caused by main bearing failure have resulted in billions of dollars worth of economic losses for a specific type of aero-engine. To ensure flight safety and save repair and maintenance costs, it is crucial to investigate precise, effective, and intelligent monitoring methods to identify rolling bearing problems in aviation engines as soon as possible [1].

The methods for diagnosing faults in rolling bearings have evolved from conventional signal analysis + machine learning-based approaches to end-to-end deep learning-based techniques in recent years, driven by advancements in next-generation artificial intelligence, computer technology, big data, and other related fields. As a result, they have become a current research hotspot [2]. In recent years, Convolutional Neural Networks (CNN) [3], Transfer Learning (TL) [4], Deep Belief Networks (DBN) [5], Autoencoder (AE) [6], and other methods have been widely used in the field of rolling bearing fault diagnosis. For example, Zhao. et al. [7] develop a model-driven deep unrolling method to design the interpretable DL model with ante-hoc interpretability, which is also against noise attacks, and its core is to unroll a corresponding optimization algorithm of a predefined model into a neural network. Wang. et al. [8] propose a method based on a stacked sparse autoencoder (SSAE) combined with a softmax classifier.The aforementioned methods have yielded the most optimal diagnostic outcomes. Nevertheless, these methods solely pertain to diagnostics conducted within a controlled testing environment. Further validation is required in order to apply them effectively for fault diagnosis of actual aero-engine rolling bearings. Additionally, it should be noted that these methods are predicated

* Corresponding author.
E-mail address: cgnuaacca@163.com (G. Chen).

https://doi.org/10.1016/j.isatra.2023.11.034
Received 25 April 2023; Received in revised form 21 November 2023; Accepted 21 November 2023
Available online 25 November 2023

on the assumption that the distribution and quantity of sample data in normal and faulty states are comparable, necessitating prior labeling of all sample data to successfully accomplish the task of fault diagnosis. However, for systems such as aero-engines that are unable to function in an impaired state, the normal operational duration of the system exceeds that of the faulty state. Consequently, acquiring normal class samples is often more feasible while obtaining typical data under various fault conditions becomes more challenging or even unattainable. Moreover, manual data labeling proves to be a burdensome and time-consuming task. Given this reality, it is more practical to accomplish the task of rolling bearing fault detection by employing unsupervised anomaly detection methods that solely rely on training with normal samples.

The current stage witnesses a shift in unsupervised anomaly detection, as it transitions from conventional approaches to deep learning methods [9]. Traditional methods mainly include cluster analysis methods [10], one-class classification methods [11], the hypersphere distance discrimination algorithm [12], support vector description [13], etc. The primary limitation of the aforementioned methods lies in their reliance on manual feature extraction for rolling bearing fault anomaly detection, rendering them ill-equipped to handle diagnosis with a large volume of sample data.

In recent years, deep learning-based anomaly detection has made significant breakthroughs in various fields such as image processing, video analysis, and finance. It has also been applied to early fault detection of rolling bearings. These approaches primarily encompass data reconstruction-based anomaly detection methods, deep one-class classification techniques, Generative Adversarial Networks (GAN), and transfer learning [14]. H XIN [15] et al.proposed a memory residual regression autoencoder model for rolling bearing fault diagnosis and verified the proposed model on IMS rolling bearing life data set and XJTU-SY rolling bearing data set, reaching 97.97 % and 93.51 % diagnostic accuracy, respectively. ZHAO [16] proposed a network model for abnormal fault diagnosis of rolling bearings by combining sparse autoencoder and transfer learning.The method based on data reconstruction establishes the reconstruction error in the model training stage and uses the gradient descent algorithm to train the model. In addition, many anomaly detection models based on data reconstruction, such as Variational Autoencoder (VAE) [17] and Deep Convolution Neural Networks Autoencoder DCNNVE [4], have good performance in early fault detection of rolling bearings. The method based on data reconstruction is usually a symmetric structure of encoding + decoding, which increases the parameters of the model and reduces the calculation speed. Deep One Class Neural Networks (Deep OC-NN) [18] are a class of anomaly detection models that have emerged in recent years. DSVDD (Deep Support Vector Data Description) [19] is representative of this kind of method ，and the application of DSVDD to rolling bearing fault anomaly detection [20] has achieved good results. At the same time, Deep-OCNN [21] and others have been applied in the early fault diagnosis of rolling bearings. The current methods still possess the following limitations when directly applied to anomaly detection: Firstly, there is scope for further enhancement in diagnostic accuracy; secondly, an excessive number of hyperparameters restricts the model's generalization capability.

As a typical representative of transfer learning, distillation learning [22] has also played a certain role in anomaly diagnosis in recent years. To improve the accuracy of anomaly detection, References [23,24] introduced the teacher-student network structure based on distillation learning, only used the image data of normal samples to train the student network(SN), and judged whether it was an anomaly through the difference in output characteristics between the teacher network(TN) and the SN. However, the existing anomaly detection methods based on knowledge distillation learning primarily focus on addressing image data anomaly detection tasks, with limited reports on rolling bearing fault anomaly detection. Meanwhile, in order to further enhance the accuracy, stability, and generalization of this method in detecting rolling bearing faults, the following challenging issues still need to be addressed.

(1) In the MKDAD [23] method, both TN and SN possess identical structures, resulting in a significant increase in the number of network parameters. Consequently, this leads to a reduction in operational efficiency of the model. Henceforth, finding ways to enhance computational efficiency while ensuring detection accuracy has become an arduous problem that needs resolution.
(2) How can we effectively leverage the feature representation acquired through TN learning to enhance the performance of model anomaly detection? Currently, in knowledge distillation learning, anomaly detection is solely based on comparing the output of a specific layer from TN and SN, which fails to fully exploit the comprehensive feature representation across the entire network.
(3) In the anomaly detection of image data, the TN often uses a network pre-trained by large-scale data sets such as Imagenet. However, for rolling bearings under different operating conditions and environments, it is difficult to obtain a standard pre-training data set (including fault samples under various operating conditions). Therefore, how to carry out the pre-training of TN is another problem to be solved in this paper.

In summary, this paper proposes a new knowledge distillation anomaly detection method based on Vit [25] (Vit-KDAD). It is verified on two sets of tester sample data and real aero-engine rolling bearing data sets. The results show that the proposed method has better anomaly detection performance. The innovation of the method mainly includes the following aspects.

(1) The dynamic numerical simulation method is used to obtain sufficient fault sample data sets to solve the problem of TN pre-training.
(2) Two Vit models are used as a TN and an SN. By using a small-scale SN, the computational efficiency of the model is increased without sacrificing model accuracy
(3) A new loss function and anomaly score function of feature center representation is proposed, which increases the difference between the TN and SN.
(4) A new enhanced Transformer encoder that fuses residual blocks is proposed to solve the problem of attention collapse caused by the Vit model with the increase of training.

We have been open source about the program of dynamic modeling and anomaly detection method in the article. https://github.com/xiaokangyu/KDAD.

1.1. Knowledge distillation learning anomaly detection model

The anomaly detection based on knowledge distillation learning primarily involves the examination of trained TN and SN models. Typically, VGG, Resnet, and other models are used for both TN and SN. During training, the SN learns the feature representation of normal samples from the TN through knowledge distillation. During testing, the current input's state is determined by evaluating the discrepancy between output characteristics of TN and SN. The framework for anomaly detection based on knowledge distillation learning is illustrated in Fig. 1. Compared to existing networks like VGG, CNN, Resnet, etc., Vit model demonstrates superior classification performance on large-scale datasets; hence it is employed as the backbone network in this study. Distinct structures are adopted for both TN and SN.

In Fig. 1, $L(T_i,S_i)$ is the characteristic loss value of the i layer. T_i is the feature output of the i-layer TN, and S_i is the feature output of the i-layer SN.

Let $S_i(x_i;\theta)$ be the output of layer i of the SN with weight θ and $T_i(x_i;w)$be the output of layer i of the TN with weight w. The specific loss function is shown in Eq. (1):

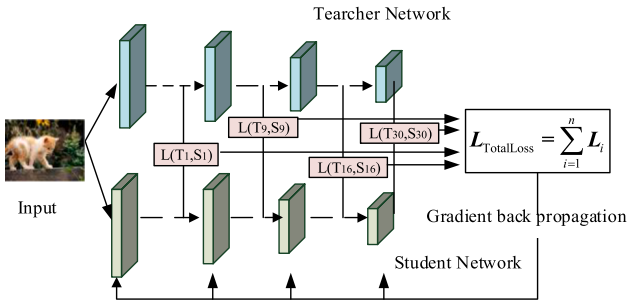

Fig. 1. Distillation learning anomaly detection principle.

$$L = \sum_{i=1}^{n}\left\{\frac{1}{m}\sum_{j=1}^{m}\left\|S_{i,j}(x_i,\theta) - T_{i,j}(x_i,w)\right\|^2 + \lambda\left(1 - \frac{< S_i(x_i,\theta), T_i(x_i,w) >}{\|T_i(x_i,w)\| \times \|S_i(x_i,\theta)\|}\right)\right\} \tag{1}$$

In Eq. (1): j is the j-th feature of the current layer output; m is the total number of features output by the current layer; n is the number of network layers used to calculate the loss function; $< \cdot,\cdot >$ is the function of calculating inner product; $\|\cdot\|$ calculate the function of the module length;$\lambda$is a hyperparameter.

2. The proposed method

2.1. Loss function with embedded central features and Wasserstein distance

The Euclidean distance and cosine similarity currently serve as the two primary components of the loss function Eq.(1) employed in the anomaly detection model based on knowledge distillation learning. However, both methods have inherent limitations. For instance, cosine similarity solely considers the angle between vectors without taking into account their magnitudes. To address these shortcomings in loss value calculation, we propose incorporating the Wasserstein distance [26], which can measure distances between distributions with no intersection whatsoever. The computation of Wasserstein distance is illustrated in Eq.(2).

$$W_p\left(\mu,v\right) = \left(\inf_{\gamma\in\Gamma(\mu,v)}\int_{\chi\times\chi}\|x-y\|^p \mathrm{d}\gamma\left(x,y\right)\right)^{1/p} \tag{2}$$

Where:inf is to take down the exact boundary;$\Gamma(\mu,v)$is the set of joint distributions γ on $\chi\times\chi$; μ,v is the joint distribution parameter; p is the calculated distance dimension, usually $p = 2$. x and y are the input.

To make the SN better learn the feature distribution of the TN, this paper adds the center feature constraint to the original loss function. The process of adding the central feature constraint is as follows: first, calculate the feature center of the output $T_i(x_i;w)$ of the TN in the whole data set, denoted as $o_i = \frac{1}{\mathbf{m}}\sum_{j=1}^{\mathbf{m}} T_{i,j}(x_i;w)$. Then, the Euclidean distance between all SN outputs $S_i(x_i;\theta)$and the feature center $O_i$is calculated as the center feature constraint loss, which is recorded as:$L_O = \sum_{i=1}^{n}\|o_i - S_i(x_i;\theta)\|^2$.

In summary, the final designed loss function of the embedding center feature and Wasserstein distance is shown in Eq.(3):

$$L_{new} = L_o + L + W_p \tag{3}$$

After completing the training of the model, the Anomaly Score of all samples was calculated according S=L_{new}, and the model was evaluated by the AUC index.

2.2. A new enhanced Transformer encoder fusing residual blocks

The specific architecture of the Vit model is illustrated in Fig. 2. Vit employs multiple Transformer encoders to extract features from input

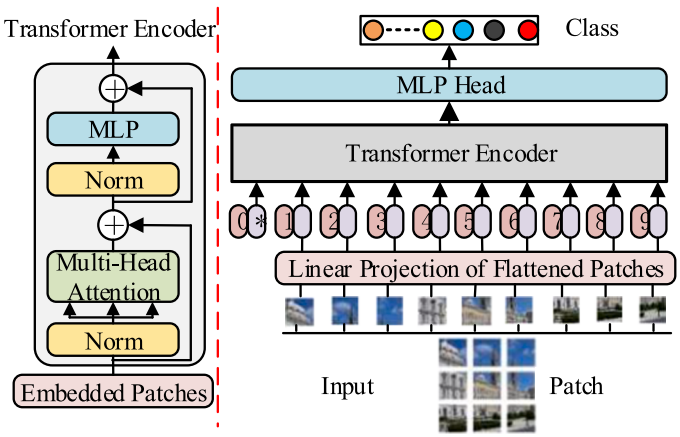

Fig. 2. Vision Transformer principle.

data, with the self-attention module enabling association between pixels that are widely separated across the entire image.

In the computation process, Vit first divides the input $x \in \mathbf{R}^{w\times h\times c}$ into N image blocks of size $p\times p\times c$, denoted as $x_p \in \mathbf{R}^{N\times p\times p}$, where h, w, c, p are the width, height, number of channels of the input image, and the edge length of the divided image block, respectively. In this paper, it is divided into pixel blocks of size 16×16. Then, the resulting N image blocks x_p are mapped into embedding vectors and combined with the classification tokens cls of each image block, denoted as z_0, as the input to the transformer encoder model. The transformer encoder consists of a multi-headed attention module, a feedforward neural network module, and a residual connection.

The multi-headed attention module is extended from the self-attention mechanism. In the self-attention mechanism, three matrices, Key(K), Value (V), and Query (Q) can be calculated from the input z_0. Then, the correlation degree of Q and K is calculated using the dot product and scaled to obtain the weight coefficients, while V is weighted to obtain the self-attention output vector. The calculation process of the self-attentive weights is shown in Eq. (4):

$$\mathrm{Attention}(Q,K,V) = \mathrm{soft\ max}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{4}$$

Where,d is the length of the input sequence z.

The reference[27,28] points out that as the number of layers of the Vit model increases, the performance saturates, i.e., the problem of attention collapse occurs. To solve this problem, an enhanced transformer encoder structure that fuses residual blocks is proposed in this paper, as shown in Fig. 3.

The enhanced transformer encoder structure mainly adds a residual block structure that allows the network to learn deep features without network degradation. The residual block contains a convolutional layer (Cov), a batch normalization layer (BN), an activation layer, and other structures. The inspiration for this structure comes from reference [28]. In reference [28], two augmented shortcuts are added in parallel to the multi-headed attention structure, to avoid the collapse of the model. Based on this, we add a residual block structure in parallel in multi-headed attention. The input of the residual block remains unchanged as the original input. In each layer, the output of the multi-headed attention is combined with that of the residual blocks to enhance the subsequent layers' multi-headed attention. This continuous enhancement process effectively prevents model collapse, as even if a certain layer's multi-headed attention fails, the augmented residual block can still provide corresponding features.

In addition, the multi-headed attention module and the feedforward neural network module incorporate the original inputs to augment the features of each layer in the network.

In summary, the output of the Vit model after the L-th encoder is

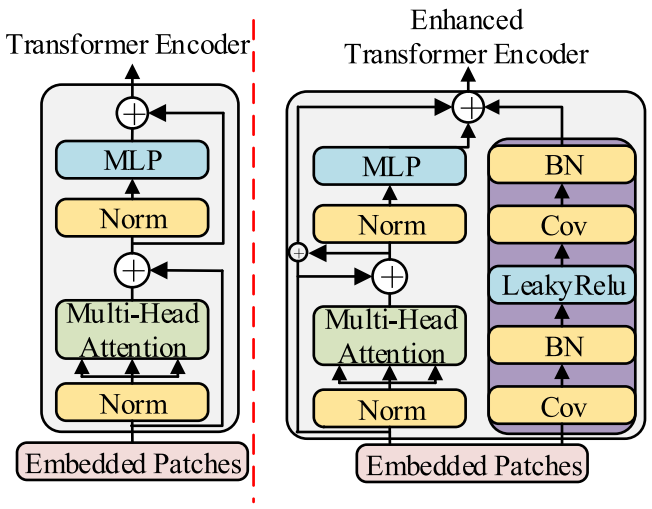

Fig. 3. Enhanced transformer encoder.

calculated as shown in Eq. (5):

$$\begin{aligned} z_l^{mha} &= MHA\left(LN(z_l)\right) + z_l \\ z_{l+1}^{mlp} &= FFN\left(LN\left(z_l^{mha}\right)\right) + z_l^{mha} + z_0 \\ z_{l+1}^{res} &= f(z_l) + z_l \\ y &= FC\left(LN\left(z_L^0\right)\right) \end{aligned} \tag{5}$$

Where $LN(\cdot)$ is the layer normalization and y is the final output of the transformer after multiple encoders. z_l is the input of the l-th layer encoder.*FFN*() is feedforward neural network.

2.3. Model architecture

In Fig. 4, the number of network layers of the TN in the Vit-KDAD method is 12, the size of the input data is 224 × 224 × 3, the size of the patch is 16, and the number of attention heads is 12. The number of network layers of the SN is 4, and the number of attention heads is 2. The Vit-KDAD model used for the process of anomaly detection is:

(1) Firstly, numerical simulation is utilized to acquire the rolling bearing fault data in four distinct states: normal, outer ring fault, inner ring fault, and rolling body fault. Additionally, the TN model undergoes pre-training with initial parameters derived from imagenet model training.

(2) The parameters of the pre-trained completed TN are inputted into the Vit-KDAD model. Only the training data from normal class samples is utilized to complete the training of SN. The threshold is determined by selecting the maximum S value from the output of normal sample.

(3) The trained TN and the SN are used to implement the rolling bearing fault detection.

Algorithm. Vit-KDAD describes the detailed calculation process.

(*continued on next page*)

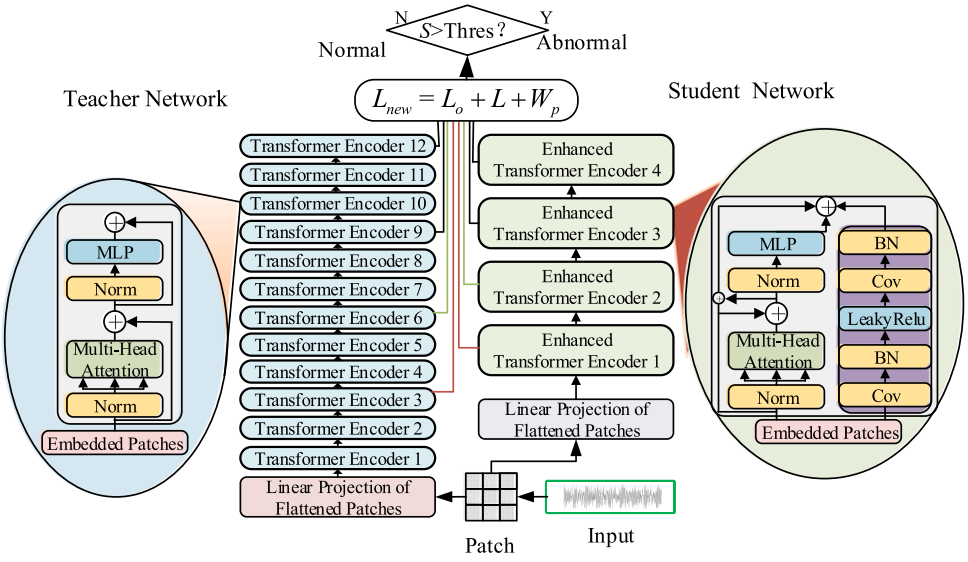

Fig. 4. Vit-KDAD.

(*continued*)

Require: Teacher network(TN) pre-trained by Imagenet dataset.
Require: Rolling bearing simulation dynamics model. Generate simulation data.
Require: Real data sequence,X.
TN retraining:
1. The vibration acceleration data is generated by the simulation.
2. Data preprocessing (one-dimensional data into two-dimensional data)
3. Finish TN training
Vit-KDAD model training:
4. Read the parameters of TN.
5. Randomly initialize a student network SN.
6. Initialize Adam with a learning rate of 10^{-3} and a weight decay of 10^{-5} for the parameters of SN.
7. for iteration = 1, . . . , 70 000 do
8. Choose a random training sequence X_{train} from X, and data preprocessing.
9. $Y' \leftarrow TN\ (X_{train})$
10. $Y \leftarrow SN\ (X_{train})$
11. Compute the loss value between Y ′ and Y according to Eq.3
12. Update the union of the parameters of SN.
13. end for
14. Return SN and TN
Anomaly detection:
15. Read the parameters of TN and SN.
16. Read the data X_{test} and perform data preprocessing.
17 . $Y' \leftarrow TN\ (X_{test})$
18. $Y \leftarrow SN\ (X_{test})$
19. Compute the *S* value.
20. Return anomaly detection results

3. Rolling bearing fault simulation dynamics model

Chen [29] created a dynamic model of a system linking a rolling bearing defect in the rotor with the casing as shown in Fig. 5. The model includes a casing, bearing, bearing seat,and other components. Eq.(6) illustrates the differential equation for system dynamics derived from Newton's second law. Only the right half of the kinetic equation is kept in Eq.(6) because of the presence of a symmetric structure.

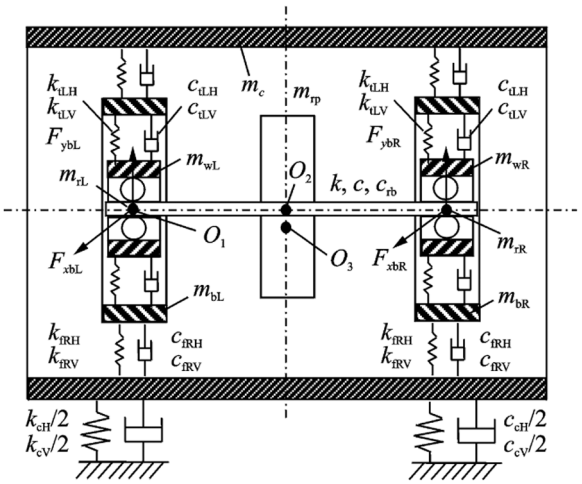

Fig. 5. Rolling bearing coupling system model.

Where,m_{rp}represent the equivalent mass of the rotor at the turntable.m_{bR} represent the right bearing support quality. m_{rR}represent the equivalent mass of the rotor at the high and low-pressure rotor bearings.m_{wR} represent right bearing outer ring mass.m_c, k_{cH}, c_{cH}represent the mass of the casing, the stiffness and damping of the connection between the casing and the foundation. k, c, c_{rb} represent the shaft stiffness, rotor damping at the disc, and rotor damping at the bearing.$k_{tLH}, k_{tRH}, c_{tLH}, c_{tRH}$ are the support stiffness and squeeze film damping between the outer ring of the left and right bearings and the bearing support. $k_{fLH}, k_{fRH}, c_{fLH}, c_{fRH}$ are the support stiffness and damping between the magazine and the left and right end bearing supports, respectively.O_1, O_2, O_3 are the bearing geometry center, rotor geometry center, and rotor center of mass, respectively ; e is the mass eccentricity.F_{xbR}, F_{ybR} are the support reaction forces of the right end bearing.F_{xbL}, F_{ybL} are the support reaction forces of the left end bearing.

$$\begin{cases} m_{rp}\ddot{x}_{rp} + k\left(x_{rp} - x_{rR}\right) + k\left(x_{rp} - x_{rL}\right) + c\dot{x}_{rp} = m_{rp}e\omega^2\cos\omega t \\ m_{rp}\ddot{y}_{rp} + k\left(y_{rp} - y_{rR}\right) + k\left(y_{rp} - y_{rL}\right) + c\dot{y}_{rp} = m_{rp}e\omega^2\sin\omega t - m_{rp}g \\ m_{bR}\ddot{x}_{bR} + k_{fRH}\left(x_{bR} - x_c\right) + c_{fRH}\left(\dot{x}_{bR} - \dot{x}_c\right) + k_{tRH}\left(x_{bR} - x_{wR}\right) \\ +c_{tRH}\left(\dot{x}_{bR} - \dot{x}_{wR}\right) = 0 \\ m_{bR}\ddot{y}_{bR} + k_{fRH}\left(y_{bR} - y_c\right) + c_{fRH}\left(\dot{y}_{bR} - \dot{y}_c\right) + k_{tRH}\left(y_{bR} - y_{wR}\right) \\ +c_{tRH}\left(\dot{y}_{bR} - \dot{y}_{wR}\right) = -m_{bR}g \\ m_{rR}\ddot{x}_{rR} + k\left(x_{rR} - x_{rp}\right) + c_{rb}\dot{x}_{rR} - F_{xbR} = 0 \\ m_{rR}\ddot{y}_{rR} + k\left(y_{rR} - y_{rp}\right) + c_{rb}\dot{y}_{rR} - F_{ybR} = -m_{rR}g \\ m_{wR}\ddot{x}_{wR} + k_{tRH}\left(x_{wR} - x_{bR}\right) + c_{tRH}\left(\dot{x}_{wR} - \dot{x}_{bR}\right) + F_{xbR} = 0 \\ m_{wR}\ddot{y}_{wR} + k_{tRH}\left(y_{wR} - y_{bR}\right) + c_{tRH}\left(\dot{y}_{wR} - \dot{y}_{bR}\right) + F_{ybR} = -m_{wR}g \\ m_c\ddot{x}_c + k_{cH}x_c + c_{cH}\dot{x}_c + k_{fRH}\left(x_c - x_{bR}\right) + k_{fLH}\left(x_c - x_{bL}\right) \\ +c_{fRH}\left(\dot{x}_c - \dot{x}_{bR}\right) + c_{fLH}\left(\dot{x}_c - \dot{x}_{bL}\right) = 0 \\ m_c\ddot{y}_c + k_{cH}y_c + c_{cH}\dot{y}_c + k_{fRH}\left(y_c - y_{bR}\right) + k_{fLH}\left(y_c - y_{bL}\right) \\ +c_{fRH}\left(\dot{y}_c - \dot{y}_{bR}\right) + c_{fLH}\left(\dot{y}_c - \dot{y}_{bL}\right) = -m_c g \end{cases} \tag{6}$$

According to the Hertz contact theory, the bearing force of the rolling bearing can be seen in reference [29].

In this paper,the Runge-Kutta method is used to solve the problem by manual programming in Matlab 2019. The step size (sampling frequency) of the simulation data is the same as the sampling frequency of the corresponding bearing measured data. About the dynamic model of rolling bearing established in this paper, we have open-sourced the relevant code. https://github.com/xiaokangyu/KDAD.

4. Test verification

To verify the effectiveness of the Vit-KDAD method in rolling bearing fault anomaly detection. Rolling bearing fault diagnosis dataset from Case Western Reserve University, USA, rolling bearing fault test dataset from the Intelligent Diagnosis and Expert System (IDES) research laboratory of the Nanjing University of Aeronautics and Astronautics (NUAA) for aero-engine rotor tester with the magazine, and a real aero-engine rolling bearing fault data set were verified.

In this paper, the GPU is NVIDIA GTX1660 6 G; i5–9600 K processor; the operating system is Windows 10; 8 G memory; the programming language is python3.7; the framework of all deep learning models is Pytorch1.11; using an Adam optimization algorithm, the learning rate is 0.001.

4.1. Data set introduction

4.1.1. Case Western Reserve University rolling bearing test data

The driving end data of the rolling bearing fault diagnosis dataset of Case Western Reserve University, USA, whose corresponding bearing model is SKF6205, is selected, and the data sampling frequency is 12 kHz. There are 3 machining defect faults of the bearing inner ring, outer ring, and rolling body, together with the normal state, there are 4 states. The bearing parameters are shown in Table 1.

4.1.2. IDES rolling bearing failure data set

The aero-engine rotor tester, equipped with magazine rolling bearing

Table 1
Rolling bearing parameter information.

Bearing	Internal diameter (mm)	Major diameter (mm)	Ball diameter (mm)	Pitch diameter (mm)	Number of balls
6205	25	52	7.94	39.04	9
6206	30	62	9.5	46	9
Main bearing	133.35	201.73	22.23	167.54	20

test, is conducted on the platform depicted in Fig. 6, which is a 1:3 scale replica of an actual engine. The test platform effectively demonstrates the attenuation characteristics of vibration signals from the aero-engine during transmission. In this test, single-row deep groove ball bearings (bearing model 6206) are employed as per the parameters listed in Table 1. The EDM cutting method is utilized to introduce fault defects including 6 mm wide cracks on both outer and inner rings, as well as depressions with a radius of 0.5 mm and depth of 2 mm on rolling bodies.

The vibration acceleration sensor B& K4805 and NI USB9234 data collected are used in the test. The sampling frequency is 10,240 Hz and the data point of a single sample is 8192. The test speeds are 1500, 1800, 2000, and 2400 (r/min). The sensor installation position is shown in Fig. 6.

4.1.3. A real aero-engine rolling bearing fault data set

The outer ring spalling fault test of the rolling bearing (main bearing) was carried out on a real aero-engine. The size of the outer ring fault is about 15 mm × 7 mm, and the bearing parameters are shown in Table 1, Fig. 7 is a sketch of an aero-engine and rolling bearing outer ring peeling fault. The test was carried out for about 5 h, and the test was finally stopped due to excessive engine vibration. The sampling frequency of this test is 200,000 Hz, a single sample has 200,000 data points, the sampling interval is 2 s, and the data of the engine intermediate casing measuring point is analyzed. To reduce the amount of calculation, only part of the sample data with a speed greater than 13,000 r/min is selected for anomaly detection. In the process of data preprocessing, the data is first downsampled, and the downsampling frequency is 128,000 Hz. According to the method of continuous division, each 50,176 (224 × 224) point is a sample.

4.2. Simulation signal verification

The simulation parameters of the three types of bearings in the simulation calculation are shown in Table 2. f is the fault characteristic frequency. To simplify the model, the contact angle is set to 0 degrees in the main bearing simulation, and only the outer and inner ring failures are simulated.

To illustrate the accuracy of the simulation model, the simulation results and the actual test data are compared to the example of the outer ring failure. The simulation and actual comparison results are shown in Fig. 8. Fig. 8(1)(2)(3) shows the vibration acceleration data of three types of bearings obtained from simulation and actual test, Fig. 8 (4) (5) (6) shows the envelope spectrum obtained after discrete wavelet

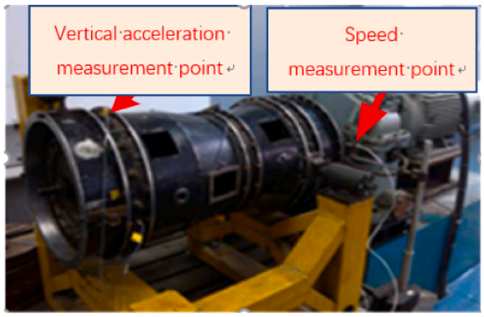

Fig. 6. Acro-engine rotor tester.

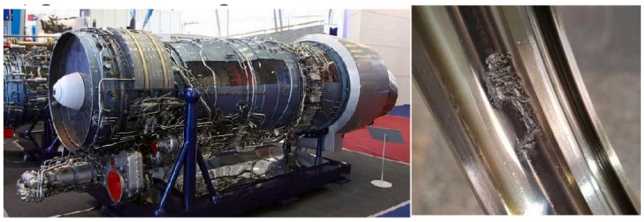

Fig. 7. A certain type of turbofan aero-engine and its rolling bearing outer ring fault.

Table 2
Simulation parameters of rolling bearing.

Bearing	Fault location	L_D (mm)	a (mm)	Rotational speed /rpm	f/Hz
6205	Inner ring	0.5334	2.794	1800	162.5
	Outer ring	0.5334	2.794	1797	107.4
	Rolling element	0.5	0.5	1800	70.6
6206	Inner ring	1	2	2400	217.2
	Outer ring	1	2	2400	142.8
	Rolling element	0.5	0.5	2400	92.7
Main bearing	Inner ring	16	7	14,675	2770.4
	Outer ring	16	7	14,675	2121.7

transform(DWT) and Hibert transform of the vibration acceleration signal, where the wavelet base is db8. For 6205 and 6206 bearing signal decomposition 2 layers, the aero-engine main bearing signal decomposition layer is 5 layers, and the first layer reconstructed signal is chosen uniformly for wavelet envelope spectrum analysis.

The results show that the waveforms of the simulated and measured signals of the three bearing models are in good agreement, and the difference in amplitude is caused by the different structural parameters of the system. From the wavelet envelope spectrum, it can be seen that the simulation model can accurately generate the bearing outer ring fault characteristic frequency and its multiplier frequency, and it can correspond to the measured signal envelope spectrum one by one, which shows the effectiveness of the simulation model.

In particular, it should be noted that the real aero-engine magazine vibration signal contains the weak periodic shock signal caused by the main bearing failure, and is superimposed with the noise components of aero-engine aerodynamics and combustion, high and low-pressure speed and its multiplier frequency, gear meshing frequency, and its multiplier frequency, rotor passing frequency and its multiplier frequency, etc., which makes the collected magazine vibration signal components extremely complex. It is difficult to obtain the aero-engine vibration data that matches the real situation by pure simulation modeling. This is the reason for the difference in the individual frequency components of the wavelet envelope spectra of the two signals in Fig. 8(6).

4.3. Pre-training of TN

(1)For the 6205 bearings, when performing TN training, 12,100 data points were selected as one sample, which was first converted to a grayscale map of 110 × 110 × 1 size [30]. For 6206 bearings, 8100 data

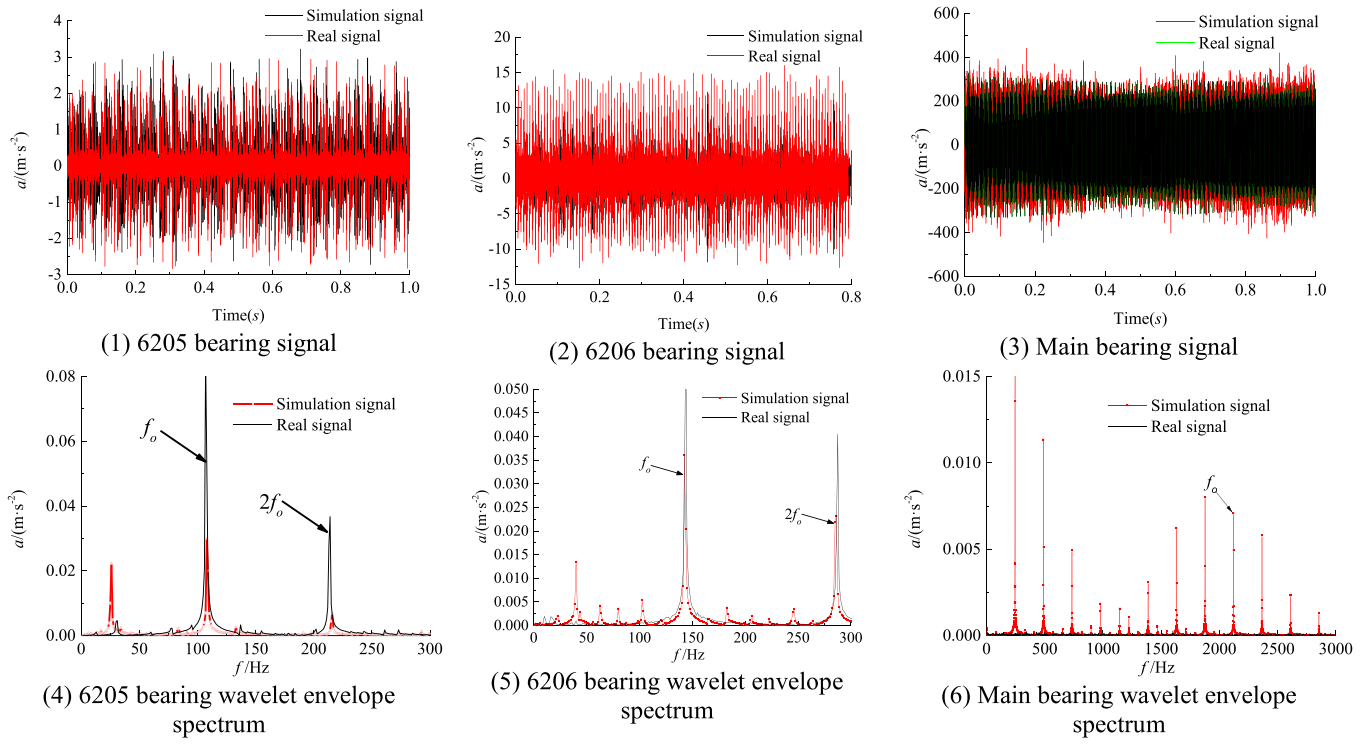

Fig. 8. The results of the simulation and measured signal comparison of three types of bearings.

points are selected as one sample of TN training, which is first converted to a grayscale map of size $90 \times 90 \times 1$. For the aero-engine rolling bearings in the TN training, 50,176 data points were selected as one sample, which was converted to a grayscale map of $224 \times 224 \times 1$ size. All samples ware expanded to a grayscale map of $224 \times 224 \times 3$, with the same data in 3 channels. In this pre-trained model, the length and width of the input image are both 224. To be consistent with the parameters in this model and achieve the purpose of fast training, we also selected samples of the same size.

In the simulation data of three types of bearings, the number of training samples containing each type of fault is 3000, and the number of test samples for each type of fault is 300. The batch size is set to 16, and the number of training times (epochs) is 100.

Fig. 9(1) shows the variation in the test accuracy obtained by the "train-and-test" method. To further demonstrate the classification effect of the TN on the simulated dataset. The T-SNE method was used to visualize the output of the TN of 6206 bearing and main bearing.

The results demonstrate that TN can accurately identify the fault type of simulated signals from three types of bearings. Fig. 9(2)(3) illustrates that TN effectively classifies the simulation data, exhibiting a significant inter-class distribution distance and a minimal intra-class distribution distance. This further confirms the precise classification capability of TN for simulation data.

4.4. Comparative validation of different anomaly detection methods

Only the measured normal samples were used and combined with the trained completed TN to distill learning for the SN in Vit-KDAD. The information on the measured samples of the three types of bearings is shown in Table 3, where both the 6206 bearings and the main bearing use vibration acceleration data based on the magazine measurement points for fault detection.

Table 3
Three types of rolling bearing sample information.

Bearing	Normal	Anomaly		
		Inner ring	Outer ring	Rolling element
6205	140	159	120	157
6206	517	441	461	488
Main bearing	423	*	423	*

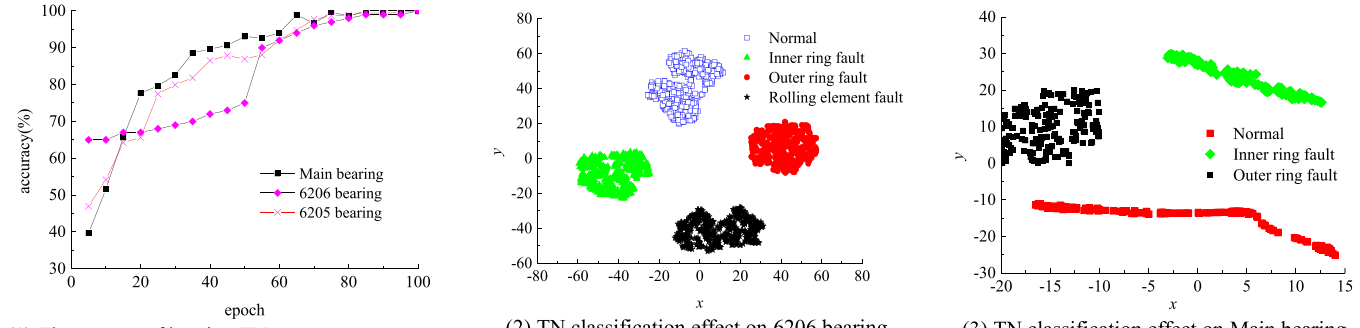

Fig. 9. The training result of TN.

1) Model parameters

The parameters of the TN in the training process are derived from the pre-training parameters obtained through Imagenet training, ensuring consistency with the original network in terms of Patch and Heads. As for the SN parameters in this paper, they need to have consistent patch sizes with TN for loss calculation purposes. Considering both training speed and accuracy, we set 4 layers and 2 heads for the SN. During testing, normal data is divided into test sets and training sets at a ratio of 7:3. As shown in Table 4, the computational complexity and AUC index value of the model are used to determine the number of network layers and Heads of the SN. It can be seen from the table that for the 6205 bearings, when the selection parameters are the number of layers of the SN 3,4,5,6 and heads of 2, the highest detection accuracy is achieved. At this time, from the perspective of accuracy and computational complexity(FLOPs), the two parameters selected.3 and 2 are the best combinations. For 6206 bearings, this optimal combination is 4 and 2. For the main bearing, the best combination is also selected as 4 and 2. In summary, this paper selects the number of layers and heads of the student network as 4 and 2 respectively.

2) Ten-Fold Cross-Validation

To obtain the reliable unbiased test results of this method, the ten-fold cross-validation method was used to verify the model on three types of bearings. In the ten-fold cross-validation process, the normal samples are randomly divided into 10 parts. Then, 9 of them are selected to participate in the training each time, and the remaining 1 part and the fault sample set are used as the test set. The results are shown in Fig. 10. The results show that after ten-fold cross-validation, high detection accuracy is achieved on all three types of bearings, and the average AUCROC value on the three types of bearings reaches 99.9 %, indicating that the method in this paper has the reliable anomaly detection ability.

3) Verification of different data division ratios

To further illustrate the validity of the proportion of the training set and test set for normal samples in this paper, the normal classes are divided according to the following proportions,and then the divided data are tested and verified.The results are shown in Fig. 11. The results show that the detection results change with the change in the partition ratio. When the proportion of samples in the training set increases, the detection effect is better. When the ratio is 6:4, the 6206 bearing reaches the best test result. When the ratio is further increased to 7:3, it reaches the maximum value on all three bearings. Therefore, we divide the training and test sets in a ratio of 7:3 during the verification process.

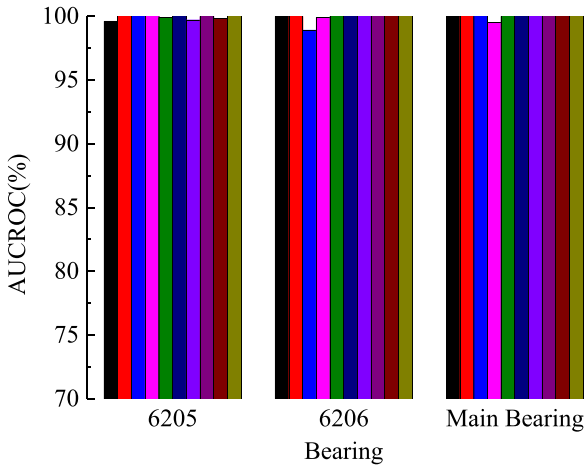

Fig. 10. Ten-fold CV method results.

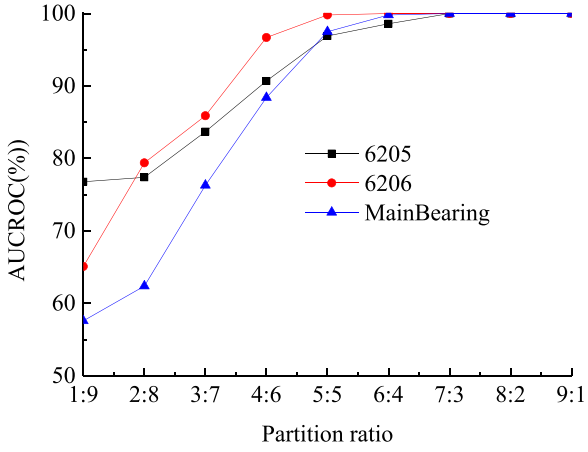

Fig. 11. The verification results of different data division ratios.

4) Comparison Validation

To illustrate the advantages of the Vit-KDAD model in rolling bearing fault anomaly detection, several typical anomaly detection models such

Table 4
The test results of different parameters of the SN.

Bearing	Layers	Number of attention heads											
		1		2		3		4		5		6	
		AUC (%)	FLOPs (G)	AUC (%)	FLOPs (G)	AUC (%)	FLOPs (G)	AUC (%)	FLOPs (G)	AUC (%)	FLOPs (G)	AUC (%)	FLOPs (G)
6205	2	96.4	1.4	99.7	2.5	98.4	3.7	92.3	4.9	89.6	6.1	94.5	7.2
	3	98.6	2.1	**100.0**	3.7	98.6	5.6	97.6	7.2	95.7	8.8	99.8	10.5
	4	95.4	2.8	**100.0**	4.9	**100.0**	7.2	94.2	9.7	97.8	12.2	**100.0**	14.8
	5	96.2	3.4	**100.0**	6.3	**100.0**	9.3	98.6	12.2	98.9	15.3	**100.0**	18.1
	6	93.8	4.1	**100.0**	7.6	**100.0**	11.2	99.4	14.7	96.6	18.3	**100.0**	21.7
6206	2	89.3	1.4	96.7	2.5	90.2	3.7	90.3	4.9	91.1	6.1	90.6	7.2
	3	92.9	2.1	99.6	3.7	95.7	5.6	89.6	7.2	97.8	8.8	98.2	10.5
	4	90.8	2.8	**100.0**	4.9	**100.0**	7.2	95.1	9.7	**100.0**	12.2	97.3	14.8
	5	94.3	3.4	**100.0**	6.3	**100.0**	9.3	96.3	12.2	98.7	15.3	97.1	18.1
	6	92.6	4.1	**100.0**	7.6	**100.0**	11.2	94.8	14.7	**100.0**	18.3	**100.0**	21.7
Main bearing	2	93.2	1.4	94.3	2.5	92.7	3.7	91.8	4.9	92.4	6.1	91.9	7.2
	3	96.6	2.1	96.7	3.7	98.8	5.6	98.5	7.2	**100.0**	8.8	94.7	10.5
	4	97.4	2.8	**100.0**	4.9	99.6	7.2	96.7	9.7	**100.0**	12.2	95.8	14.8
	5	99.2	3.4	**100.0**	6.3	**100.0**	9.3	99.4	12.2	98.9	15.3	96.2	18.1
	6	97.6	4.1	99.1	7.6	**100.0**	11.2	100.0	14.7	**100.0**	18.3	**100.0**	21.7

as MKDAD, DSVDD, OC-NN, and SVDD were compared and validated under the same data division. Among them, the backbone network of MKDAD uses the VGG16 network, and the backbone networks of DSVDD and OC-NN models use the CNN network. To illustrate that the algorithm in this paper does not lose the advantage of computational speed while improving the detection accuracy, the computational time was also compared with the current MKDAD algorithm, which has the best detection effect, on the test set. In the verification process, the normal samples are divided into a training set and a test set according to the ratio of 7:3, and all the fault samples are tested at the same time. The comparison results are shown in Table 5.

For the 6205 bearings: Vit-KDAD was able to achieve a 100 % AUC index, indicating that Vit-KDAD was able to fully differentiate normal sample data from faulty sample data. However, among the remaining four algorithms compared, the best performance was achieved by MKDAD, with a detection accuracy of 98.73 %, which was 2.27 % lower than that of Vit-KDAD. 92.79 % was achieved by DSVDD, which was 7.21 % lower than that of SVDD. 92.32 % was achieved by OC-NN, which was 7.68 % lower than that of SVDD. The worst performance of SVDD was 87.98 %. The comparison results show that Vit-KDAD can accurately identify abnormal rolling bearing failures.

The calculation time of Vit-KDAD was 25.37 s, while that of MKDAD was 21.32 s, which was 4.05 s higher than the calculation time of MKDAD. This is mainly because the number of parameters in the Vit model of the TN in Vit-KDAD is much larger than the number of parameters in the VGG16 model.

For the 6206 bearings: Vit-KDAD was able to achieve 100 % AUC metrics, indicating that Vit-KDAD was able to fully differentiate normal sample data from faulty sample data. However, among the remaining four algorithms compared, the best performing algorithm was MKDAD, with a detection accuracy of 96.38 %, a decrease of 3.62 % compared to Vit-KDAD, DSVDD, with a detection accuracy of 83.24 %, a decrease of 16.76 %, and OC-NN, with a detection accuracy of 89.57 %, a decrease of 10.43 %. The worst performance of SVDD was 80.65 %. The comparison results show that Vit-KDAD can accurately identify the abnormalities of rolling bearing faults. It is worth pointing out that the training of the TN was done by using the bearing seat signal obtained from the simulation, while the test was done by using the magazine measurement point signal during the validation process, but Vit-KDAD was still able to accurately achieve the abnormality detection, which also shows that the SN learned the knowledge of the TN.

The calculation time of Vit-KDAD was 68.16 s, while the calculation time of MKDAD was 52.37 s, which was 15.79 s slower than the calculation time of MKDAD. This is mainly because the number of parameters of the Vit model of the TN in Vit-KDAD is much larger than the number of parameters of the VGG16 model.

For the main bearing: the AUC metric of Vit-KDAD was able to reach 100 %, indicating that Vit-KDAD was able to fully differentiate normal sample data from faulty sample data. However, among the remaining four algorithms compared, the best performer is MKDAD, with a detection accuracy of 87.26 %, which is 12.74 % lower than that of Vit-KDAD. 85.47 % is the detection accuracy of DSVDD, which is 14.53 % lower than that of SVDD. 79.69 % is the detection accuracy of OC-NN, which is 20.31 % lower than that of SVDD. The worst performance of SVDD was 62.38 %. The comparison results show that Vit-KDAD can accurately identify the abnormalities of rolling bearing faults. Compared with the remaining algorithms, the algorithm proposed in this paper has the highest diagnostic accuracy, indicating that Vit-KDAD has certain advantages in the detection of aero-engines rolling bearing faults.

Table 5
Anomaly detection comparison results AUROC (%).

Bearing	Index	Vit-KDAD	MKDAD	DSVDD	OC-NN	SVDD
6205	AUC	100.0	97.73	92.79	92.32	87.98
	Time	25.37 s	21.32 s	*	*	*
6206	AUC	100.0	96.38	83.24	89.57	80.65
	Time	68.16 s	52.37 s	*	*	*
Main bearing	AUC	100.0	87.26	85.47	79.69	62.38
	Time	153.28 s	129.54 s	*	*	*

In terms of the comparison of computation time, the computation time of Vit-KDAD is 153.28 s, while the computation time of MKDAD is 129.54 s, which is 23.74 s slower compared to the computation time. As with the previous findings, this is mainly because the number of parameters of the Vit model in the TN in Vit-KDAD is much larger than the number of parameters of the VGG16 model.

To further illustrate the advantages of Vit-KDAD in terms of detection accuracy, Figs. 12 and 13 visualize the abnormal scores(S) of both Vit-KDAD and MKDAD, as well as the changes in detection accuracy during the test. Additionally, Table 6 presents the corresponding confusion matrix. From Fig. 12, it is evident that Vit-KDAD exhibits a more distinct differentiation between normal and abnormal S-value results compared to MKDAD. Conversely, the S-value curve of MKDAD demonstrates an unstable and fluctuating performance with no clear intuitive distinction between normal and abnormal cases. This disparity elucidates why Vit-KDAD excels at accurately identifying abnormalities.

From Fig. 13, it can be seen that for the 6205 bearing training started Vit-KDAD can achieve 100 % diagnostic accuracy and has maintained that diagnostic accuracy during the period. However, MKDAD increased from the initial 92.38–97.73 % diagnostic accuracy, with large fluctuations during the period. In comparison, Vit-KDAD was more stable and performed better.

For the 6206 bearings, the starting recognition result of Vit-KDAD was 88% at the beginning of the training, and the diagnostic accuracy steadily increased as the training progressed, eventually reaching 100 %. However, the initial value of MKDAD was only about 85 %, and stabilized at 96 % as the training proceeded, with large fluctuations in diagnostic accuracy during the period, indicating that the generalization performance of the MKDAD network was not strong. In comparison, Vit-KDAD has a stronger anomaly detection capability and the model has higher stability and better performance.

For the main bearing, the starting recognition accuracy of Vit-KDAD was about 78% at the beginning of training, and the diagnostic accuracy steadily improved as the training proceeded, eventually reaching 100 %. However, the initial diagnostic accuracy of MKDAD was only about 65 % and stabilized at 87 % as the training proceeded. In comparison, Vit-KDAD has a stronger abnormality detection ability and the model has higher stability and better performance.

It should be noted that during the graphing process, the S values of the main bearing test results of MKDAD were all shifted downward by 30 for clarity and intuition.

4.5. Ablation experiment

Two sets of ablation experiments were designed to further verify the role of each component in Vit-KDAD.

(1) The remaining parameters of Vit-KDAD are kept unchanged, and only the conventional Transformer encoder (denoted as Vit_01) and the enhanced Transformer encoder proposed in this paper are used for the SN, respectively. The detection results of both are analyzed and used to verify the effectiveness of the enhanced Transformer encoder.
(2) Under the premise of constant network structure parameters, the loss function in which only L loss, only $L+W_p$ loss, and only L_o+L loss are considered, and the loss function proposed in this paper is compared and verified. It is used to illustrate the effectiveness of the loss function proposed in this paper.

Two sets of ablation experiments with five different test cases were performed, and the results are shown in Table 7. The results show that the proposed Vit-KDAD can achieve optimal detection accuracy

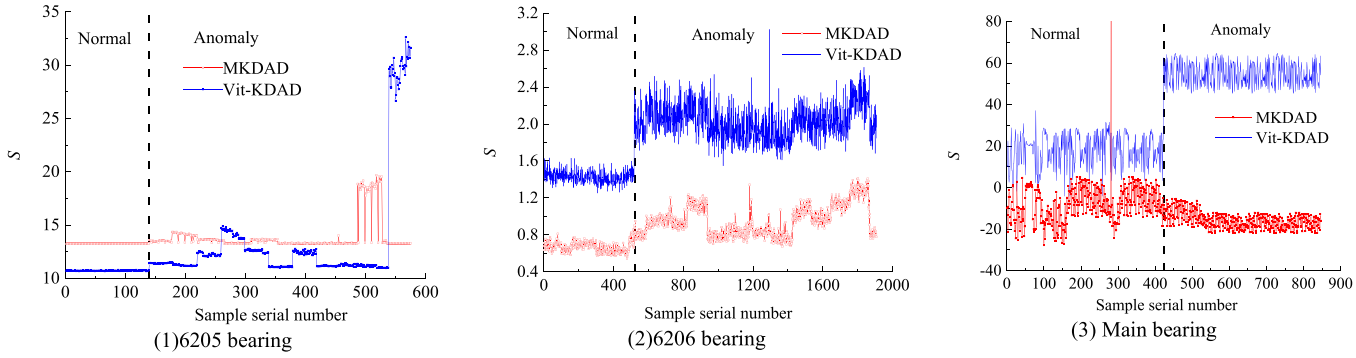

Fig. 12. The change in the anomaly score *S*.

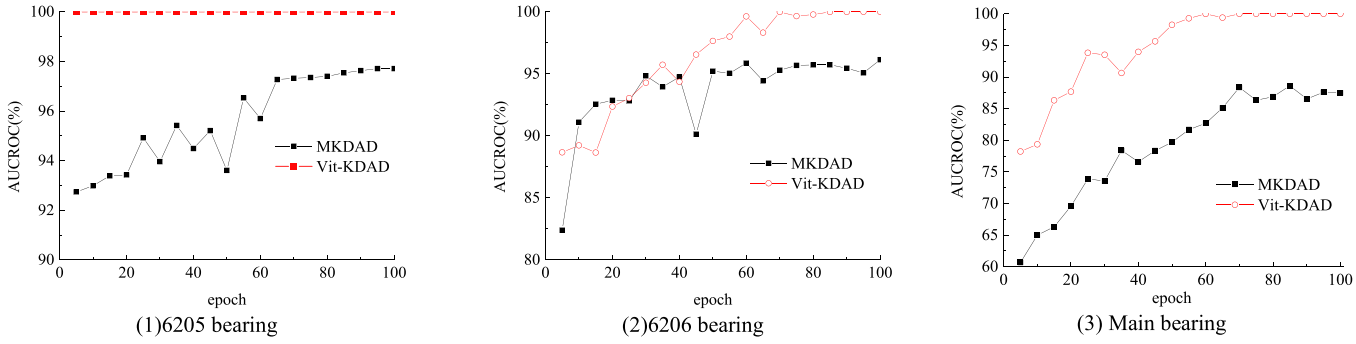

Fig. 13. The change in detection result.

Table 6
Confusion Matrix of Vit-KDAD and MKDAD Methods on Three Types of Rolling Bearing Data.

6205 Bearing and Vit-KDAD			6205 Bearing and MKDAD		
Actual class	Predicted class		Actual class	Predicted class	
	Healthy	Abnormal		Healthy	Abnormal
Healthy	140	0	Healthy	138	2
Inner ring fault	0	159	Inner ring fault	11	147
Outer ring fault	0	120	Outer ring fault	2	112
Ball fault	0	157	Ball fault	8	148
6206 Bearing and Vit-KDAD			6206 Bearing and MKDAD		
Actual class	Predicted class		Actual class	Predicted class	
	Healthy	Abnormal		Healthy	Abnormal
Healthy	517	0	Healthy	481	36
Inner ring fault	0	441	Inner ring fault	16	225
Outer ring fault	0	461	Outer ring fault	31	430
Ball fault	0	488	Ball fault	6	482
Main Bearing and Vit-KDAD			Main Bearing and MKDAD		
Actual class	Predicted class		Actual class	Predicted class	
	Healthy	Abnormal		Healthy	Abnormal
Healthy	423	0	Healthy	357	66
Outer ring fault	0	423	Outer ring fault	61	362

Table 7
Results of ablation experiment AUROC (%).

Bearing	Vit_01	L	$L+W_p$	$L+L_o$	Vit-KDAD
6205	97.32	98.23	99.45	99.02	100.00
6206	98.47	97.64	98.92	98.35	100.00

compared to the other four cases, regardless of the bearing type. Specifically, the enhanced Transformer encoder proposed in this paper has stronger feature extraction advantages than the traditional Transformer encoder, e.g., the detection accuracy obtained by using the traditional transformer encoder on the 6205 bearings is only 97.32 %, while the accuracy can reach 100 % accuracy. On 6206 bearings, the detection accuracy is only 98.47 % with the traditional transformer encoder, but 100 % with the enhanced transformer encoder. In the cases of considering only L loss, considering only$L+W_p$ loss, and considering onlyL_o+L loss, the detection results on both bearings are reduced. The comparison results show that each component of the model proposed in this paper plays an important role in the detection results.

4.6. Generalization Performance Verification

To further illustrate the practicability and generalization performance of the proposed method, the proposed method is verified on the rolling bearing fault data sets of three other turbofan aero-engines of the same type. Among them, the rolling bearing state of engine No.01 is normal, the rolling bearing state of engine No.02 is the outer ring fault, and the composite fault of the outer ring and inner ring spalling exists on the rolling bearing of engine No.03. The sample data of the three aero-engines are shown in Table 8. The data speed range used in the experiment is 14,200–14,675 r/min. The outer ring fault is shown in Fig. 14.

The test uses the model trained in Section 4.4, and the *S* value of the model output is shown in Fig. 15. Compared with the results in Fig. 12 (3), the *S* value of Vit-KDAD on these three engines is more stable. When the threshold is set to 40, the normal and anomaly states can be completely distinguished. The comparison results further prove the

Table 8
The number of samples of three real aero-engines.

Engine number	01	02	03
Rolling bearing states	Normal	Anomaly	
Sample amount	12,635	300	869

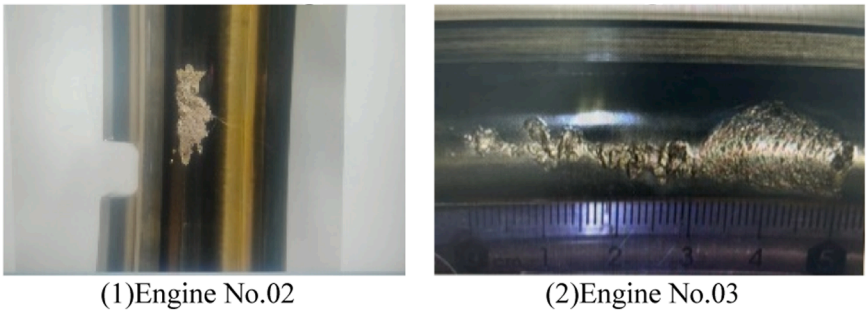

Fig. 14. The outer ring spalling fault of aero-engine No.02 and No.03.

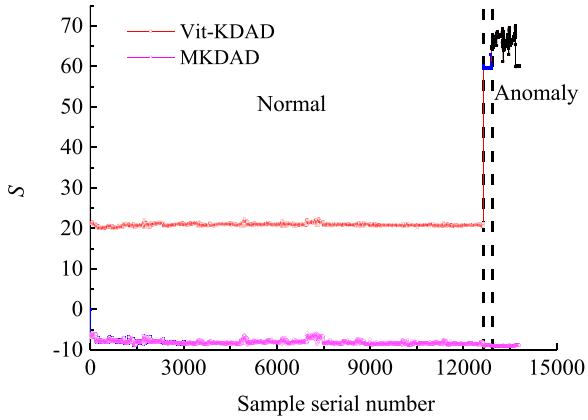

Fig. 15. Fault anomaly detection scores *S* of different engines.

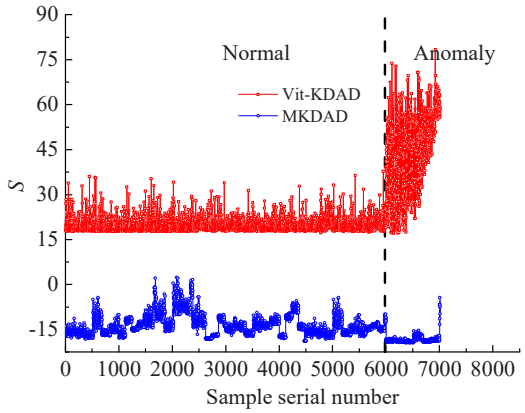

Fig. 16. Outputs of both models on real aero-engine data.

effectiveness and practicability of the proposed method. For the MKDAD method, although the *S* value is more stable than that in Fig. 12(3). However, it does not distinguish between normal and anomaly. Through comparison, it is found that the *S* value fluctuates greatly in Fig. 12(3), while the results on the three engines are more stable. The main reasons are as follows:1) Different vibration acceleration sensors; 2) The data acquisition equipment is different. At the same time, the sampling frequency of the vibration data of the three engines is 256,000 Hz, 204,800 Hz, and 200,000 Hz. To have the same sampling rate as the data in Section 4.4, the data of No.01 engine and No.02 engine are down-sampled, and the down-sampling frequency is 200,000 Hz.

4.7. Rolling bearing fault evolution monitoring

To further verify the Vit-KDAD method. Experiments were carried out on a full-life fault data set of an aero-engine rolling bearing of the same type. The aero-engine test was completed in Beijing from September to November 2021. The total duration of the test was about 150 h, the sampling frequency was 200,000 Hz, the data length of a single sample was 1 s, and the storage interval of the sample was 3 s. During the test, the Endevco vibration acceleration sensor was installed at the position of the intermediate casing of the aero-engine, and the final test was stopped due to the excessive vibration parameters of the whole machine. After disassembly, it was found that the outer ring of the three-point rolling bearing had a serious spalling fault. After that, after expert analysis, it was found that about 120 h or so, the outer ring peeling fault of the three-point rolling bearing began to appear. To reduce the amount of calculation, about 50 sets of samples with a speed range of 14,000–14,675 r/min were randomly selected per hour during the verification process, and a total of 7010 sets of data samples were obtained. The model trained in Section 4.4 is used for verification. The final comparison test results with various methods are shown in Fig. 16.

The comparative results demonstrate that the fault evolution detection test, utilizing the trained model, continues to yield favorable detection outcomes. It can be seen from the results in Fig. 15 that the *S* value also increases with the progress of the test, indicating that the bearing spalling is expanding steadily at this time. In this data set, the value of the normal state is still less than 40, and after the spalling fault occurs, the *S* value is greater than 40, indicating that when the threshold is 40, Vit-KDAD can distinguish between normal and abnormal states. The AUC value on this dataset is 96.63 %. The value of the abnormal score *S* of MKDAD varies greatly in both normal and abnormal stages, and it fluctuates greatly. It is believed that it is mainly caused by the change in speed during the engine test. It also reflects from the side that the features extracted by TransFormer are significantly better than those extracted by traditional CNN structures.

5. Conclusion

In this paper, a knowledge distillation anomaly detection method based on the Vit model is proposed and applied to fault anomaly detection in rolling bearings, yielding promising results.

1) Various comparative analysis results demonstrate that the proposed anomaly detection method achieves superior performance in both tester data and real aero-engine rolling bearing fault data, with a detection accuracy of up to 100 %.
2) Ablation implementation results indicate that each component of the model presented in this paper has a certain impact on the detection accuracy, with the enhanced Transformer encoder having the greatest influence.
3) The test results confirm that by utilizing distillation learning theory, simulation acceleration data for rolling bearings can be obtained through simulation methods and used for training TN models. This approach effectively addresses the lack of fault samples in rolling bearing datasets while improving fault anomaly detection accuracy.

4) The Vit-KDAD model demonstrates practical value based on its successful application to aero-engine rolling bearing fault anomaly detection.

From a comprehensive perspective, this method offers several advantages: 1) The combination of numerical simulation and rolling bearing fault anomaly detection enhances model accuracy. Numerical simulation eliminates the need for additional tests to acquire data, reducing reliance on test benches and saving costs.2) This study focuses specifically on aero-engine rolling bearings. From practical applications, it is evident that this method achieves higher detection accuracy.

Declaration of Competing Interest

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

Acknowledgements

This work has been supported by the National Science and Technology Major Project (J2019-IV-0004-0071); National Natural Science Foundation of China (52272436).

References

[1] Xunkai Wei, Li Yang, LiGuang Zhan, et al. Aeroengine Prediction And Health Management. Beijing: National Defense Industry Press; 2014.
[2] Lei Yaguo, Yang Bin, Jiang Xinwei, Jia Feng, Li Naipeng, Nandi Asoke K. Applications of machine learning to machine fault diagnosis: a review and roadmap (Invited Review Paper). Mech Syst Signal Process 2020;138:106587.
[3] Xu G, Liu M, Jiang Z, et al. Online fault diagnosis method based on transfer convolutional neural networks. IEEE Trans Instrum Meas 2019;(99):1–12 (PP).
[4] CAO Hongru, SHAO Haidong, ZHONG Xiang, et al. Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds. J Manuf Syst 2022;62:186–98.
[5] SHAO Haidong, JIANG Hongkai, ZHANG Haizhou, et al. Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network. IEEE Trans Ind Electron 2018;65(3):2727–36.
[6] Zeng Mengjie, Li Shunming, Li Ranran, Lu Jiantao, Xu Kun, Li Xianglian, et al. A hierarchical sparse discriminant autoencoder for bearing fault diagnosis. Appl Sci 2022;12(818):818.
[7] Zhao Zhibin, Li Tianfu, An Botao, Wang Shibin, Ding Baoqing, Yan Ruqiang, et al. Model-driven deep unrolling: towards interpretable deep learning against noise attacks for intelligent fault diagnosis. ISA Trans 2022;129(Part B):644–62.
[8] Wang1 Junxiang, Zhan1 Changshu, Yu2 Di, Zhao2 Qiancheng, Xie32 Zhijie. Rolling bearing fault diagnosis method based on SSAE and softmax classifier with improved K-fold cross-validation. Meas Sci Technol 2022;33(No.10):105110.
[9] Chalapathy R, Menon AK, Chawla S. Robust, Deep and Inductive Anomaly Detection. Cham: Springer; 2017.
[10] Zeng M, Yang Y, Luo S, et al. One-class classification based on the convex hull for bearing fault detection. Mech Syst Signal Process 2016;81:274–93.
[11] Yong Liu, Chao Wang, Ping Zhou. An Early Warning Method for Rolling Bearing Fault of Civil Aero-Engine. J Propuls Technol 2022;43(02):295–304. https://doi.org/10.13675/j.cnki.tjjs.200284.
[12] Lin T, Chen G, Ouyang W, et al. Hyper-spherical distance discrimination: a novel data description method for aero-engine rolling bearing fault detection. Mech Syst Signal Process 2018;109:330–51.
[13] Pan Y, Chen J, Guo L. Robust bearing performance degradation assessment method based on improved wavelet packet–support vector data description. Mech Syst Signal Process 2009;23(3):669–81.
[14] Chalapathy Raghavendra, Chawla Sanjay. Deep learning for anomaly detection: a survey. CoRR 2019:03407.
[15] Huang X., Wen G., Dong S., et al. memory residual regression autoencoder for bearing fault detection. IEEE Trans Instrum Meas; 2021, PP(99): 1–1.
[16] Zhao X, Jia M, Liu Z. Fault diagnosis framework of rolling bearing using adaptive sparse contrastive auto-encoder with optimized unsupervised extreme learning machine. IEEE Access 2020;8:99154–70.
[17] Shen Zhang;Fei Ye;Bingnan Wang;Thomas G. Habetler. Semi-supervised learning of bearing anomaly detection via deep variational autoencoders; 2019.
[18] Ruff L., Vandermeulen R., A., N. Görnitz, et al. Deep one-class classification. In: Proceedings of the international conference on machine learning; 2018.
[19] Ruff L., Vandermeulen R., A., Grnitz N., et al. Deep semi-supervised anomaly detection. In : Proceedings of the international conference on learning representations; 2019.
[20] Mao W, Chen J, Liang X, et al. A new online detection approach for rolling bearing incipient fault via self-adaptive deep feature matching. IEEE Trans Instrum Meas 2020;69(2):443–56.
[21] Chalapathy R., Menon A.K., Chawla S. Anomaly detection using one-class neural networks. 2018.
[22] Bergmann P, Fauser M, Sattlegger D, Steger C. Uninformed students: student-teacher anomaly detection with discriminative latent embeddings. In: Proceedings of the 2020 IEEE/CVF conf. on computer vision and pattern recognition. Seattle: IEEE; 2020. p. 4182–91 [Doi: 10.1109/CVPR42600.2020.00424.
[23] Salehi M, Sadjadi N, Baselizadeh S, Rohban MH, Rabiee HR. Multiresolution knowledge distillation for anomaly detection. In: Proceedings of the 2021 IEEE/CVF conf. on computer vision and pattern recognition. Nashville: IEEE; 2021. p. 14897–907 [Doi: 10.1109/CVPR46437.2021.01466].
[24] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv 2015;1503:02531.
[25] Vaswani A, Shazier N, Parmar N, Uszkoreit J, Jones L, Gomez NA, et al. Attention is all you need. In: Proceedings of the thirty first int'l conf. on neural information processing systems. Long Beach: Curran Associates Inc.; 2017. p. 6000–10 [Doi: 10.5555/3295222.3295349].
[26] Martin Arjovsky;Soumith Chintala;L.éon Bottou. Wasserstein GAN. Statistics; 2017.
[27] Zhou D., Kang B., Jin X., et al. DeepViT: towards deeper vision transformer; 2021.
[28] Yehui Tang, Kai Han, Chang Xu, An Xiao, Yiping Deng, Chao Xu, et al. . Augmented shortcuts for vision transformers.In: Proceedings of the thirty fifth conference on neural information processing systems, NeurIPS; 2021.
[29] Guo Chen. .Vibration modeling and verification for whole aero-engine. J Sound Vib 2015;349(4):163–76.
[30] Wu Jingyao, Zhao Zhibin, Sun Chuang. Fault-attention generative probabilistic adversarial autoencoder for machine anomaly detection. IEEE Trans Ind Inform 2020;16(12):7479–88.