Applied Intelligence
https://doi.org/10.1007/s10489-024-05643-3

Application of a dense fusion attention network in fault diagnosis of centrifugal fan

Ruijun Wang[1] · Yuan Liu[2] · Zhixia Fan[1] · Xiaogang Xu[1] · Huijie Wang[1]

Accepted: 23 June 2024

Abstract
Although the deep learning recognition model has been widely used in the condition monitoring of rotating machinery. However, it is still a challenge to understand the correspondence between the structure and function of the model and the diagnosis process. Therefore, this paper discusses embedding distributed attention modules into dense connections instead of traditional dense cascading operations. It not only decouples the influence of space and channel on fault feature adaptive recalibration feature weights, but also forms a fusion attention function. The proposed dense fusion focuses on the visualization of the network diagnosis process, which increases the interpretability of model diagnosis. How to continuously and effectively integrate different functions to enhance the ability to extract fault features and the ability to resist noise is answered. Centrifugal fan and rotor fault data are used to verify this network. Experimental results show that the network has stronger diagnostic performance than other advanced fault diagnostic models.

Keywords Centrifugal fan vibration · Deep learning · Disperse attention · Dense fusion attention network · Condition monitoring · Fault diagnosis

1 Introduction

Centrifugal fans are fluid-driven rotary machine that finds extensive application in exhaust and filtering systems in factories. Its fault will lead to a reduction in the power plant's energy generation efficiency and potential damage to other equipment [1]. Consequently, the detection and maintenance of the fan is particularly significance. Real-time monitoring of the fan status is required to ensure the safe and efficient operation of the generator set.

It is a popular diagnosis method to extract fault pulse features from the vibration signal of rotating machinery for intelligent recognition of the state, such as artificial neural network (ANN) [2], support vector machine (SVM) [3, 4]. However, the outcome of conventional intelligent diagnostic approaches relies heavily on the quality of feature extraction [5], and it has not yet achieved full automation in the extraction of valuable information. The simple network structure is difficult to illustrate complex nonlinear relationships [6].

With the development and application of artificial intelligence, deep learning has gradually become the mainstream algorithm [7]. The biggest advantage of the deep learning compared to the traditional learning model is that it can learn deeper signal features independently [8, 9], such as convolutional neural network (CNN) and autoencoder (AE). Zhao et al. [10] conducted experimental studies on the performance of deep belief network (DBN) and applied it to machine health monitoring. Li et al. [11] proposed an end-to-end adaptive multiscale fully convolutional network (AMFCN) for fault identification. Jiao et al. [12] developed a two-way complementary data-dense convolutional network. This network was designed to alleviate feature loss and gradient disappearance, facilitating the transmission of fault feature

Ruijun Wang and Yuan Liu are co-first authors of the article.

✉ Zhixia Fan
fzx.ncepu@outlook.com

Ruijun Wang
wrj.ncepu@outlook.com

Yuan Liu
ly.xsyu@outlook.com

Xiaogang Xu
xxg@ncepu.edu.cn

Huijie Wang
whj@ncepu.edu.cn

[1] School of Energy Power and Mechanical Engineering, North China Electric Power University, Baoding 071003, China

[2] School of Electronic Engineering, Xi'an Shiyou University, Xi'an 710065, China

Published online: 10 August 2024

and more comprehensive capture. Existing deep learning methods usually learn all features equally. Therefore, the integration of attentional mechanisms is more focused on effective information expression. The attention mechanism obtains several weighted scores [13, 14] to generate learning differences for fault information. The selective kernel module (SKNet) proposed by Li et al. [15] introduced a global channel attention mechanism [16], which improved the model by dynamically weighted averaging multiple kernels or groups from the same layer. Wang et al. [16] added the channel and spatial attention mechanism to a one-dimensional CNN (1DCNN) to form a multi-attention CNN (MA1DCNN). It can recalibrate the spatial and channel features of each layer to strengthen the expression of fault pulses. Gao et al. [17] proposed the dual attention dense residual network (DADRN). The dual attention module realizes self-adaptive feature refinement and then inputs it to the dense network, making the extracted features more discriminative. Fan et al. [18] constructed a multi-scale and multi attention feature fusion (MAFF) structure with cross layer fusion function. By performing two-step dynamic calibration on feature sets of different paths, different functions were effectively integrated to achieve better diagnostic performance. Shen et al. [19] proposed multiscale attention feature fusion network for rolling bearing fault diagnosis under variable speed conditions. A concatenated multi-scale multi-attention module was designed to adaptively extract fault features for efficient diagnosis. Zhu et al. [20] developed a lightweight multivariate and multi-directional induction network (LM-MDINet) for gearbox fault diagnosis. Its core, multivariate and multidirectional induction layer (M-MDI), guides the model towards expressing interested information in multiple directions to improve performance. In conclusion, adding attention mechanisms [21] can lead the computational resources of deep learning model to the part with the largest amount of fault information in the input signal, enabling efficient representation of useful features.

Therefore, embedding attention mechanisms in diagnostic models has notable advantages in capturing discriminative information. Regarding the pulse characteristics of fan faults are sparsely distributed within the overall signal, representing a minority of the signal. Deep learning models' powerful learning and fusion capabilities often generate an abundance of redundant information, which can significantly overshadow the discrete discriminative information. The lack of an effective way to fuse these two kinds of information can lead to models falling into learned redundancy and bias. The attention mechanism can guide the representation of weights. However, the selection mechanism such as SKNet only emphasis on the soft feature selection of the single layer, and does not solve the cross-layer fusion of different paths and scales. Because fault information learning often produces feature maps under multiple scales [22], the existing fusion methods (such as cascade and addition) directly merge features [23], forming a large amount of redundancy in the feature combination [24]. Therefore, it is still a challenge to selectively integrate effective features of different sizes. In addition, the dynamic weighted average limits the weight distribution of the up and down paths, which has a certain impact on the generation of objective attention matrices. As we all know, the acquired vibration signals include the mechanical vibrations of the fan and the interactions and coupling effects between relevant components. Therefore, it is apparent that the obtained vibration signal incorporates numerous inherent oscillation modes and regularly encounters noise across various frequencies. This implies that a single-scale non-linear dynamic parameter may be insufficient to characterize fan fault signals [16] in deep learning terms. The multi-scale attention mechanism of the above MAFF module may be able to better characterize multimodal information. But the characteristics of concatenating attention are still controversial. For the case of concatenating, multi attention networks (including single scale attention, such as MA1DCNN), although increasing spatial attention can be complementary to channel attention, the linear connection of attention modules makes the weight distribution affect each other. The two attention functions have different characteristics leading to response principles differently for feature weight values. This will lead to ineffective activation of feature points.Although single attention mechanisms (such as M-MDI) do not generate these additional issues, they always feel inadequate when compiling complex data.

We have proposed a dense fusion attention network (DFANet) to tackle the issues of feature extraction and representation. During the extraction process, this method utilizes a compact densely connected backbone and optimizes its fusion mode, thereby allowing the model to effectively emphasize discrete fault information, ultimately enhancing the diagnostic ability of the network. In the model, the characteristic of dense connections is employed to cascade the integrated features with the original features after the comprehensive function extraction. The cascade operation simply stacks the features mechanically, resulting in a significant amount of redundant information within the features. Although bottleneck layers are used to reduce channel numbers, the inherent characteristic of signal redundancy has not been effectively addressed. There is no guarantee that it will focus on learning useful information in the signal when a traditional dense network fuses features. Indeed, the two paths of the dense network show inherent complementarity in learning, thus balancing the stability and exploratory nature of the network. Therefore, effectively combining these two functionalities is crucial for enhancing the performance of dense networks. During the expression process, in order to alter the state that the calibration principle of the current multi-attention mechanism that linearly follow the states of the channel and spatial dimensional distributions, we create dispersed attention module (DAM) with the criterion of reducing the irrelevant response of the space cross-channel function. Attempts to solve the problems of traditional fusion mechanisms with fixed weights, feature redundancy and insensitivity of multi-scale features. Attempts to solve the leading to invalid weight values problem due to significant differences of attentional functions, different feature weight response principles, and linear distribution mechanisms.

Deep learning models are commonly referred to as "black box models". In most fault identification processes, the user needs to explain the monitoring procedure and outcomes. This allows for the adjustment of relevant parameter settings according to actual needs, facilitating accurate decision-making and predictions. Consequently, improving model interpretability is crucial for users' trust and comprehension of deep learning models. MA1DCNN improved network recognition interpretability by illustrating the monitoring pulse extraction process. Typically in engineering applications, noise interferes with signals. Therefore, relying on a single pulse monitoring method is insufficient to interpret the model's operational status. There is a need to propose a model interpretation method that can be used in any diagnostic environment. As is well known, In model training, parameters are changing constantly. We hold the view that comprehending the iterative direction and outcomes of the information in the network is a superior interpretative strategy.

Based on the above-mentioned problems, this paper proposes DFANet for centrifugal fan fault diagnosis. First, the zoom spatial self-attention module (ZSSAM) and the channel attention module (CAM) are constructed. The zoom operation can map at multiple scales and enlarge feature information, which is more conducive to paying attention to key features. At the mean time, the channel attention module tends to focus on the correlation of channel features. Secondly, the DAM is proposed. The channel and the spatial attention are focused on the feature expressions they need. Decoupling the existing multi-attention weights restrict each other to avoid the failure of weight activation. Finally, dense fusion block (DF-Block) is created. It can not only pay attention to the feature information from different angles in a scattered manner, so that the reconstructed feature has a more obvious information tendency, but also can use the dense connection feature to achieve the effect of multiple mixed attention. Furthermore, we seek to find the regularity of weight correction by analyzing the dynamic diagram of weight variation and diagnostic trend. It is used as a reference for exploring the dynamic response intervals of the model and also provides a basis for interpretation of the model. The contributions of this paper can be summarized as follows: 1) Explore the performance of the dispersed attention mechanism initially. It has the multi-attention function and at the same time alleviates the negative influence of the mutual activation of the attention matrix to reinforce the performance of the attention mechanism. 2) A DF-Block is proposed to automatically adapt the weight output while densifying the feature map. It effectively ignores the redundancy and noise. 3) This paper creates a novel method of model interpretation. It can explain the operation mode of the network in different environments, and at the same time be able to research the dynamic response of its parameters and internal structure. 4) DFANet is applied to centrifugal fan fault identification for the first time. And Bayesian algorithm is utilized for hyperparameter optimization to achieve the optimal network configuration.

The remanent part of this paper proceeds as follows: In the Section 2, the design of DFANet is proposed in detail. In the Section 3, under different diagnostic conditions, Uses centrifugal fan and rotor operating condition data for experimental validation and further compares seven representative diagnostic models. Section 4 summarizes the work done throughout the paper.

2 Methodology

With the fast development of computer vision, recognition networks are constantly being proposed or updated. However, vibration signals are used as data input in the field of rotating machinery fault identification, and it is necessary to propose a diagnosis model suitable for signal features. The original network recognition structure is created and changed to focus more on fault pulse feature extraction. Therefore, this paper proposes DFANet. The feature map is densified while the network redundancy is reduced, forming an adaptive recognition model for one-dimensional signals.

2.1 Theoretical basis

Before introducing DFANet and its internal modules, we need to understand some basic theories.

2.1.1 CNN

As a classifier, CNN involves multiple filtering processes. It has advantages such as position invariance, parameter sharing, sparse connections, high-dimensional feature learning, adaptability, and parallelization acceleration. CNN generally consists of convolutional layers, pooling layers, and fully connected layers (FCL).
1) Convolutional layer
The core extraction module in the convolutional layer is called a convolutional kernel (filter), which extracts features by sliding on the data.

The feature map is generated through filter within the convolutional layer and is subsequently activated by the activation function before being output. Let Ψ represent the input to the convolution. The output of the i-th convolutional kernel is denoted as

$$C_i = f(\sum \Psi * k_i + \beta_i) \tag{1}$$

where $*$ is the convolution operation, k_i represents the i-th convolution kernel, β_i represents the bias, i represents the amount of channels, and f represents the activation function.
2)Pooling layer
Pooling layer usually appears along with convolution to reduce feature dimensions and redundancy, and improve computational efficiency.

For the i-th channel of the feature with γ-length in the convolutional layer, the output of the pooling layer is:

$$P_i(n) = \max_{0 \le n \le \frac{\gamma}{S}} \left\{ C_i(n\widehat{W}, (n+1)\widehat{W}) \right\} \tag{2}$$

where C_i is the input, $\widehat{W}$ is a width of the pooling window, and S represents stride size.
3) FCL
The FCL can integrate distributed features. Assume that the input to the FCL l-1 is $K^{l-1} \in R^{1\times d}$, then the output is:

$$K^l = f(K^{l-1}\Omega^l + \beta^l) \tag{3}$$

where $\Omega^l \in R^{d\times e}$ represents the weight. β^l represents the bias. The output set of the l layer is $K^l \in R^{1\times e}$.

2.1.2 Attention mechanism

Attention mechanism is a deep learning technique that mimics human attention mechanisms. In the attention mechanism, the model can learn to give different weights or levels of attention to different parts when processing input data. In this way, the model can focus more on important information related to the task, while ignoring or reducing the processing of irrelevant information, thereby improving the efficiency and performance of the model.

Set the input feature as Z_o, obtain important feature slices through the Squeeze function $F_{sq}(\cdot)$, and then input them into the Excitation function $F_{ex}(\cdot)$ to obtain feature response weights. Obtain the weighted feature set $Z_{o\omega}$ and the overall formula is as follows:

$$Z_{o\omega} = F_{ex}(F_{sq}(Z_o)) \otimes Z_o \tag{4}$$

where, "$\otimes$″ is the calibration weight.

2.2 Zoomed spatial self attention module (ZSSAM)

As the vibration signal is a time series, the spatial dimension contains information about the inherent feature of the signal. ZSSAM uses zoomed mapping to weaken background information and enhance target features. The intrinsic connection is existed between the features of the low-high. It can understand this relationship well to enhance the spatial multiscale attention of fault pulse signals. Figure 1 shows the structure of ZSSAM. The feature vector group Z=[$z_1, z_2, ..., z_c$] is used as the input of the ZSSAM module, $z_i \in R^{W\times 1}$ represents the vector for the i-th channel of the input Z. The average-pooling operations and max-pooling are applied along the channel axis to extract the most sensitive feature Z_M, and the global feature Z_A, from various channels with the same time domain information.

$$z_M^j = max\{z^j(1), z^j(2), ..., z^j(c)\} \tag{5}$$

$$z_A^j = Avgpool(z^j) = \frac{1}{1\times c}\sum\nolimits_{i=1}^{c} z^j(i) \tag{6}$$

$z^j(i)$ stands for the j-th position of the i-th channel. ${z_M}^j$, ${z_A}^j$ represent local and global attention elements at the j-th spatial position of all channels respectively. $Z_M = [{z_M}^1, {z_M}^2, ..., {z_M}^W]^T$, and $Z_A = [{z_A}^1, {z_A}^2, ..., {z_A}^W]^T$ are gotten. They act that output converged feature mapping the spatial information to obtain an valid feature combination in sharing space as $Z_1{}' = Z_M \oplus Z_A$, $Z_1{}' \in R^{W\times 1}$. "$\oplus$″ is the superposition process.

2.2.1 From left to right

In ZSSAM, the level of feature map forming by convolution increases in order. The features with higher levels contain more deep information but lack the expression of shallow information. $Z_1{}'$ serves as the input, and the output of the top-level feature Z'' is

$$Z'' = F_n(F_{n-1}...(F_3(F_2(F_1(Z_1{}'))))) \tag{7}$$

$F_1(\cdot), F_2(\cdot), ..., F_{n-1}(\cdot), F_n(\cdot)$ are convolution layers and the number of channels is 1. The sliding step length is 2.

2.2.2 Vertical transmission

Vibration signals from various health states may exhibit highly similar features. As the number of convolution layers increases, they are projected into a closed region through nonlinear mapping, resulting the phenomenon about high-level features deviate more and more from the expression of original features [25]. Therefore, the features of different levels formed by each convolution layer are used to guide the decoding features of the corresponding layer, so as to

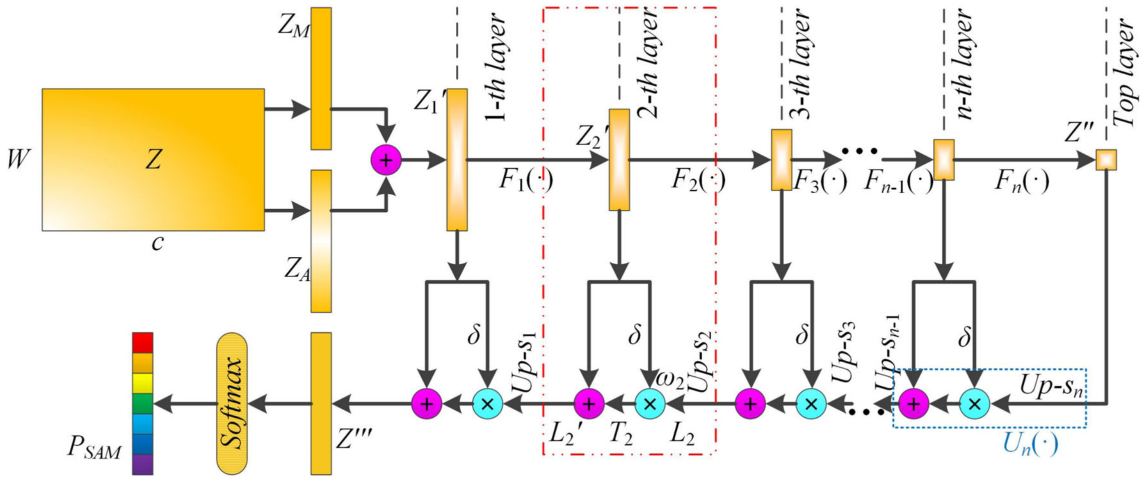


Fig. 1 Structure of ZSSAM

correct the deviation caused when the features are continuously expressed in abstraction. Taking the second layer as an example, the coding feature of the second layer is $Z_2{}'$. The *softmax* function is called to generate tracking weights, that is, $\omega_2 = \delta(Z_2{}')$. The decoding operation reconstructs the features to obtain L_2 in the second layer, and re-weights the weight feature map $T_2 = \omega_2 \otimes L_2$ with the vertical weight ω_2 of the same layer. "$\otimes$" is the calibration weight. The dilution of irrelevant information and the thickening of sensitive features are achieved. At the same time, the introduction of residual connections promotes the mixing of the advantages of features of the low-high, so that the output features contain global information without losing the reflection of detailed information. The vertical mixed output of the same layer features is $L_2{}' = Z_2{}' + T_2$.

2.2.3 From right to left

Up-sampling is based on interpolation (original value filling) algorithm [26], as the decoding method of ZSSAM. The output element is expanded through the up-sampling kernel to increase the information density. Suppose the number of channels is 1, and for the i-th layer, the up-sampled output set based on jv position is

$$\{V_i^j\} = \left\{ \begin{array}{l} w_i^t, v \neq jv \\ w_i^t, v = jv \end{array} \right\} j \in [\sigma t - \sigma, \sigma t] \qquad (8)$$

where $t \in [1, w]$ represents the original position. jv represents the original positions and their corresponding locations after up-sampling. j represents the filling position after up-sampling. The sum output of up-sampling is $V_i = [\{V_i{}^1\}, \{V_i{}^2\}, ..., \{V_i{}^W\}]^T$, $V_i \in R^{\sigma W \times 1}$. σ represents the size of the up-sampling kernel. The up-sampling operation is used to decode the top-layer feature Z'' to attain non-uniform expansion and enhance sensitive features. In ZSSAM, the constructed "up-sampling-weighting-sum" the reconstruction function $U(\cdot)$ and the output of the reconstructed information Z'' is

$$Z''' = U_1(U_2(U_3(...(U_{n-1}(U_n(Z''))))))\qquad (9)$$

$U_1(\cdot)$, $U_2(\cdot)$, ..., $U_{n-1}(\cdot)$ and $U_n(\cdot)$ are econstruction functions of different layers. In the end, the *softmax* function is introduced to output the spatial weight ratio of ZSSAM, that is, $P_{SAM} = \delta(Z''')$.

2.3 Channel attention module (CAM)

The channel features generated by different convolution kernels has different emphases. The quantity of target information contained in the channel affects the effect of feature learning. Consequently, the goal of CAM is to enhance the network's response to different features by explicitly capturing the interdependencies between channels so that it can automatically recognize the importance of the features [11]. Figure 2 is a structural diagram.

Assume that the CAM's input feature map combination is $Z = [z_1, z_2, ..., z_c]$, $z_i \in R^{W \times 1}$.

$$z_{Gi} = Avgpool(z_i) = \frac{1}{(1 \times W)} \sum_{j=1}^{W} z_i(j) \qquad (10)$$

With global average pooling, global time information can be condensed into the channel and thus acquire global channel information $Z_G = [z_{G1}, z_{G2}, ..., z_{Gc}] \in R^{1 \times c}$. At the same time, the global maximum pooling is employed in parallel to reduce the redundancy of input information and maintain the essential characteristics. For the feature Z, which consists of c-length, we can extract the information of the i-th channel,

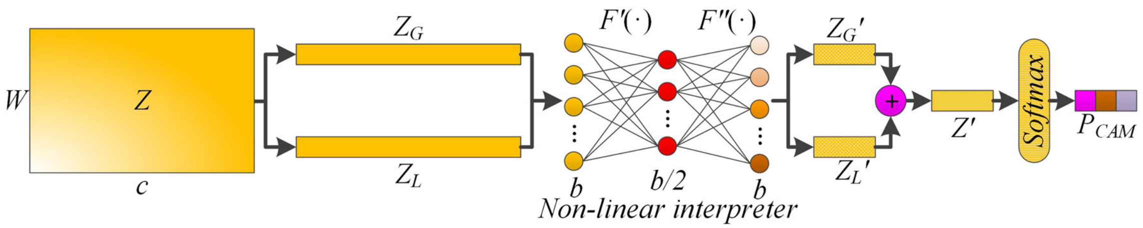


Fig. 2 Structure of CAM

the output is:

$$z_{Li} = max\{z_i^1, z_i^2, ..., z_i^W\} \tag{11}$$

The local information from different channels $Z_L = [z_{L1}, z_{L2}, ..., z_{Lc}] \in R^{1\times c}$ is obtained. W represents the width of the feature map. z_{Li} represents the maximum element of the i-th channel vector of Z. To smooth out the merged channel features, we re integrate local and global information and alter the number of channels.

In CAM, the nonlinear interpreter comprises two convolution layers. It has a powerful nonlinear integration function and fitting ability. Therefore, it is used to map channel position information and alter the number of channels. The features of the global channel information Z_G and the local channel information Z_L respectively outputted by the nonlinear interpreter are $Z_G', Z_L' \in R^{1\times b}$. After they are fused, the output feature Z' is:

$$Z' = Z_G' \oplus Z_L' = (F''(F'(Z_G))) \oplus (F''(F'(Z_L)) \tag{12}$$

among them, where $F'(\cdot)$ is the full connection layer (first layer) operation and $F''(\cdot)$ is the second layer operation. Fusion of global and local features to express the previous coding information is more comprehensive and effective. *Softmax* can dynamically compress the input data to [0,1] and the total of the probabilities is 1 [12]. This points to the fact that adding differences in the weights of the information generated on every channel is more advantageous to strengthening the information with the discriminative function. Taking $Z' = [z_1', z_2', ..., z_b'] \in R^{1\times b}$ as input, the weight output of CAM is obtained, that is, $P_{CAM} = \delta(Z')$.

2.4 Dispersed attention module (DAM)

The calibration principle of the current multi-attention mechanism adheres linearly to the distribution across spatial and channel dimensions. However, due to the large difference between the two attention functions and the different response principles to the weight value of the feature, an invalid weight set will be generated. Therefore, we create a dam to reduce the irrelevant response of space cross channel function and enhance the efficiency of computing resources. The proposed DAM has the function of synchronizing attention and adaptive cross-path feature fusion for the features of different channel numbers. The structure of DAM is shown in Fig. 3.

The multi-channel feature combination $X = [x_1, x_2, ..., x_a]$ and the few channel feature combination $Y = [y_1, y_2, ..., y_b]$ (a > b, a + b = c) form a mapping cascade feature $Z = Conv([X; Y])$. It is input into ZSSAM and CAM respectively, and output their respective weights, namely $P_{SAM} = M_S(Z)$, $P_{CAM} = M_C(Z)$. The features X and Y input to the module are calibrated and a calibrated weighted feature map is obtained. Nevertheless, too much calibration of feature weights can trigger overrespond to depth features. Therefore, input features and calibration features are superimposed on the output by introducing the residual connection [27]. Retain the previous features of the input module to avoid a decrease in the value of the weight response. The feature combination Z^* output by DAM is

$$Z^* = [X'; Y'] = [(P_{SAM} \otimes X) \oplus X; (P_{CAM} \otimes Y) \oplus Y] \tag{13}$$

X', Y' are the final results of each module after tracking and calibration.

2.5 Dense fusion block (DF-Block)

DAM is embedded in the feature connections of different channel numbers to form a DF-Block. DF-Block makes the network have the function of attention and makes the information flow directional. The DAM block focuses on the inherent features of the channel and the spatial dimension. The existing linear calibration principle of following the spatial and channel dimensional distribution is changed. The mutual influence of the attention module on feature activation is eliminated to improve the efficiency of learning. From the analysis of the overall structure of the DF-Block, the dense connection mode provides a special function that the front layer features affect the expression of the back layer. The relevant features noted by DAM will be noted again in the next operation. This way can achieve the effect of multiple mixed concerns. The features of the

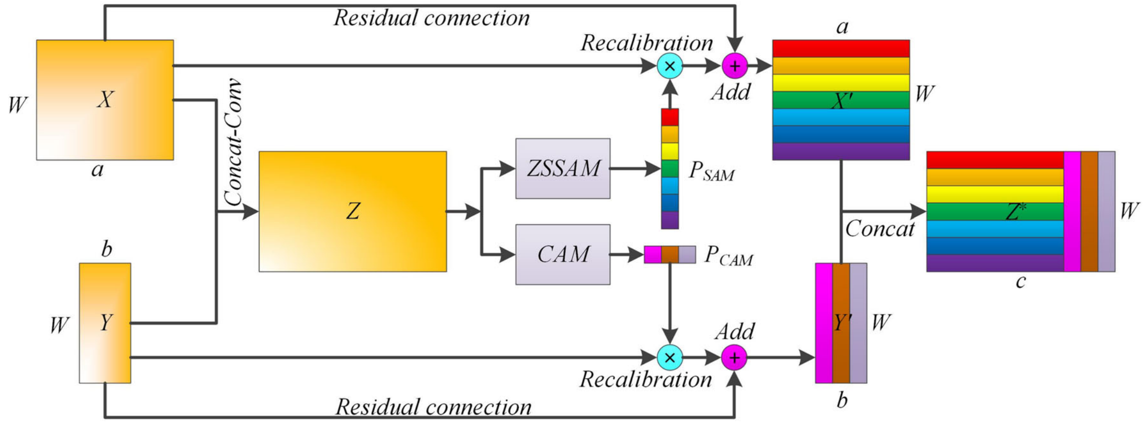


Fig. 3 Structure of DAM

network are activated multiple times and multi-angle coding is performed. The overall diagnosis model has better recognition performance. Figure 4 shows the schematic diagram of the structure. Suppose the aggregation function is Φ, $\Phi^0(\theta_0) = \theta_0$, $\Phi^1(\theta_0, \theta_1) = DAM(\theta_0, \theta_1)$, $\Phi^2(\theta_0, \theta_1, \theta_2) = DAM(DAM(\theta_0, \theta_1), \theta_2)$ then $\Phi^m = DAM(\ldots(DAM(DAM(\theta_0, \theta_1), \theta_2), \ldots), \theta_m)$. As the N layer receives all feature maps from the preceding layers., the output after encoding is:

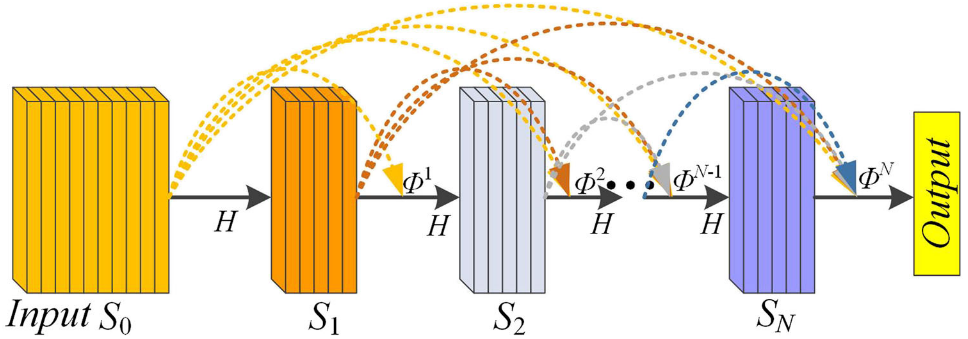
$$S_N = H[\Phi^{N-1}(S_0, S_1, ..., S_{N-1})] \tag{14}$$

where $\Phi^{N-1}(S_0, S_1, ..., S_{N-1})$ means that the aggregation function Φ encodes the feature layers $S_0, S_1, ..., S_{N-1}$. H is the feature extraction comprehensive function (*BN-f-BF-BN-f-EF*) (*BN*: Batch Normalization, *BF*: Bottleneck filter, *EF*: Effective filter). In the extraction of one-dimensional signals, when the extraction performance of several small convolution kernels is comparable to that of one large convolution kernel, the amount of network parameters of multi-layer convolution is larger. Therefore, we prefer to set only one layer of convolution in H.

2.6 Model establishment

The zoomed operation in ZSSAM can combine the fault information of different receptive fields. Up-sampling is used to extend the features after the densification mapping to the original range. At the same time, the vertical connection performs weighted summation to standardize the weight distribution and merge the features of different levels. In addition, the operation of self-attention makes it more self-regulating when strengthening or weakening the feature response value. The spatial module of zoomed self-attention improves the extraction of information in the time domain, and forms a dimensional complementarity with the distributed channel attention module. The jointly constructed DAM structure achieves cross-path feature fusion, resulting in more aggregated intra-class distribution features and more apparent inter-class discrimination basis in the dimensionality-reduced data. The original fault signal obtains a set of feature maps through convolution and maximum pooling operations to input into the DF-Block. DF-Blocks are connected in series by connecting layers. The connection layer includes the squeeze layer and the maxpooling layer. The squeeze layer is made up of BN, f, and 1×1

Fig. 4 Structure of DF-Block



convolution. It compresses the number of channels to half of the original. Work with the maxpooling layer to reduce feature redundancy. In the end, a compact DFANet structure can not only pay attention to information of different dimensions, but also use dense connections to achieve mixed attention. In the final stage, the network utilizes the *Softmax* classifier to generate outputs corresponding to different fault types, effectively executing the fault identification task. The structure of DFANet is shown in Fig. 5.

Figure 6 shows the diagnosis process diagram. The centrifugal fan fault diagnosis process based on DFANet is introduced. First, the fan's vibration signal is obtained. Then the signal is introduced into the model. Offline training is performed in the model. Test the performance of the model online. Finally, the working states are identified through the classifier.

3 Experimental verification and results

3.1 Experiment description

Training and testing conditions for diagnostic models are as follows: Software: Pytorch 1.8.0, Hardware: Ryzen R5 3600 CPU, GTX 1080Ti GPU, and 48 GB of RAM. The batch size is set to 512. White Gaussian noise is added to the signal to quantify the noise intensity.

The training data, validation data, and test data are divided in the ratio of 6:2:2. After, perform overlapping sampling on each data [28]. The length of each sample is set to 1024. Meanwhile, the Xavier normal distribution initializer and the regular normal distribution initializer are introduced in this network, and they are used to initialize the weight parameters of the convolutional layer and the full connection layer, respectively. In addition, we have constructed an early stopping strategy, retaining 20 epochs as thresholds, which means that when the network converges to a temporary maximum, there are still 20 opportunities to converge to a larger size. Therefore, the epoch value we specify will far exceed the training convergence requirements, allowing the network to converge completely. Each run will train the model 10 times in a loop, select the optimal set of network parameters, and save them. The test set is used to evaluate the actual diagnostic performance of the model accuracy.

White Gaussian noise is introduced into the original signals, simulating the interference. The signal to noise ratio (SNR) is defined as

$$SNR_{db} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{15}$$

where P_{signal} represents the average power of the signal. P_{noise} represents the average power of the noise.

3.1.1 Description of the fan data

The experiments on different mechanical vibration measurements were carried out on a centrifugal fan test rig (Model: 4-73 No. 8D), as shown in Fig. 7 for the fan experimental device. The main equipment of the centrifugal fan experimental device is shown in Table 1. Figure 8 shows the fan structure and sensor installation position.

The fan speed is 1200RPM. The sampling frequency is 1600Hz. The simulation experiments cover thirteen fault states (including one no-fault state) of misalignment, unbalance, and different parts and bearing looseness at different severity levels, details of which are shown in Table 2.

3.1.2 Description of rotor data

Spectra Quest's Mechanical Fault Simulator (MFS) simulates six different states. Details of which are shown in Table 3. For more information, please visit https://www02.smt.ufrj.br/~offshore/mfs/page_01.html.

3.2 Parameter selection

This section discusses the hyperparameter selection process using centrifugal fan data to obtain the best combination of parameters to meet the model requirements. To have a

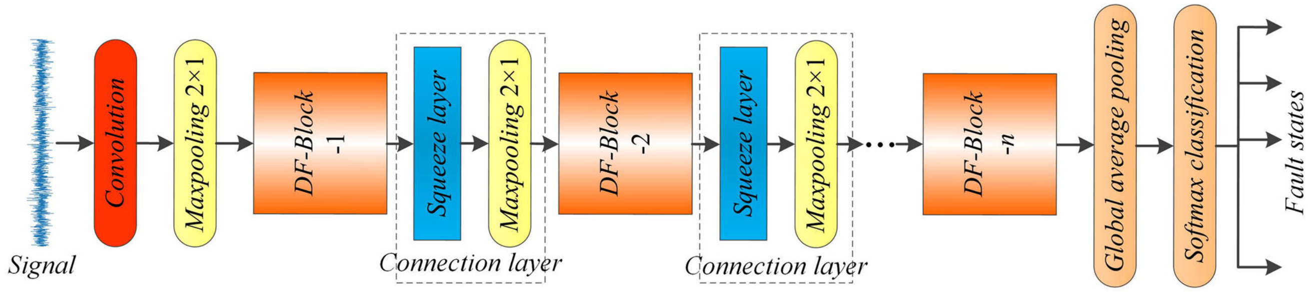


Fig. 5 Structure of DFANet

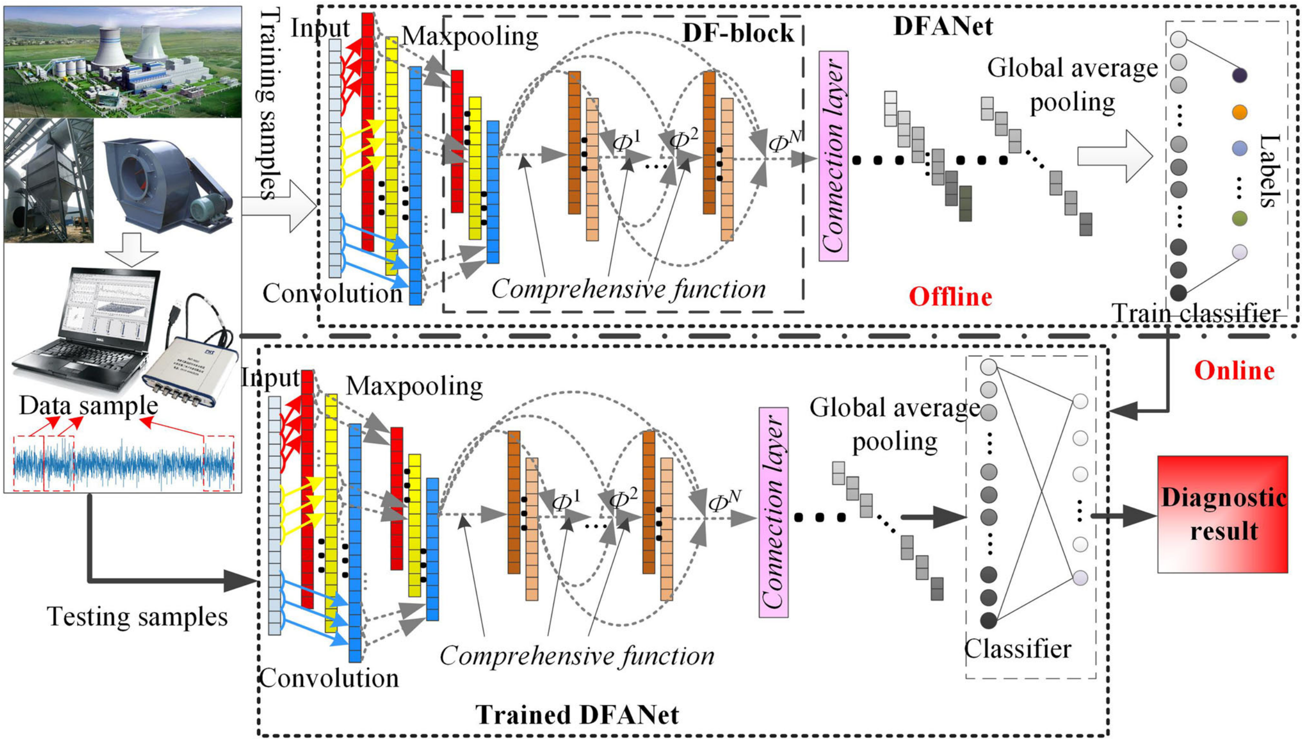


Fig. 6 DFANet's logic diagram for diagnosing fan faults

discriminative criterion and to exclude other influencing factors, first, we assume k=8 (k=b), the size of the *EL* within the DF-Block is 5 × 1, in the ZSSAM structure, the convolution layer is named *ZC* and set to be 2, and all of size is 3 × 1, the number of DF-Block is 3, the number of DAM is 2 in the DF-Block, and the number of initial filter (IF) output channels is 32.

3.2.1 Manual Selection of Model Hyperparameter

A larger convolution kernel [16] should be selected to extract the initial signal features. Thus, the filter can not only cover more complete information of the fault-excited signal segments and achieve the noise reduction effect, but also consider the balance between calculation complexity and accuracy.

Fig. 7 G4-73No8D fan experimental setup

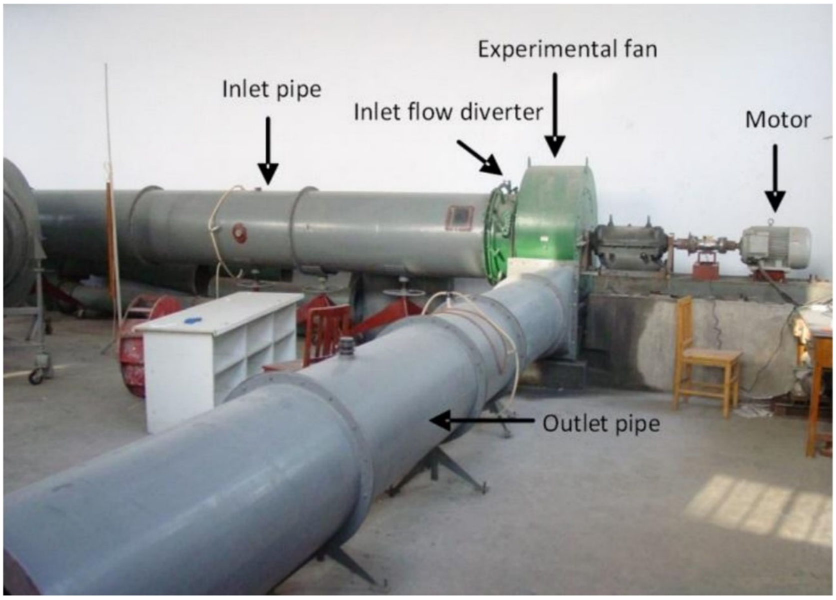


Table 1 Experimental device description of fan

Fan experimental device	Specifications	Installation location
Axial guide vane (Adjust flow)	Adjustment range: 0-90.	The axial guide vane is installed in front of the air inlet.
Motor (Drive the axial blades to rotate)	The motor model is Y180L-4. The rated power is 22 kW. The frequency converter is used to adjust the rotational speed and the precision is 0.3 rotations.	Connected to axial guide vane via coupling.
Eddy current displacement sensor (Measure vibration signals)	(IN-81, Schenck, Germany); The sensitivity is 8mV/mm; Measurement range is 0-1.5 mm and operating frequency range is 0-10 kHz.	Five eddy current sensors were mounted on both sides of the fan bearing, which non-contact measured the horizontal direction, vertical direction and axial vibration displacement signal of the bearing.
Piezoresistive pressure sensors (Measure pressure changes)	Model: SMI5552; Measurement range: 1-20 kPa; Accuracy: 10 Pa; Time response: minimum 2 ms.	Five piezoresistive pressure sensors were arranged on the inner surface of the fan casing with an angular spacing of 60.

At 0dB and -4dB data, Table 4 experiments show that the overall recognition ability of the model is improving with increasing convolutional kernel width. It shows that the large filter positively affects the model's fitting ability. In terms of combined computational complexity and accuracy, the model reaches the balanced feature learning ability when the kernel width is 29. Therefore, the size of the IF is set to 29 × 1 tentatively.

Next, select the hyperparameter combination named Combination Parameter: [k, kernel width of EF, kernel width of ZC_1 and kernel width of ZC_2].

Based on the initial assumption of k value, Table 5 shows that the model performance still needs to be improved. We know the extraction ability of the model increasing as the value of k increasing. Therefore, we discuss the situation of nearby values after doubling the value of k, and set a convolutional kernel width with not significantly different to verify the overall impact on the model when there is a small difference between the two. In addition, ZC_1 and ZC_2 are mainly used to compile the weights, so only need to have some learning effect and not be too large.

Firstly, the experimental results in the Table 5 confirm that the kernel width of ZC_1 and ZC_2 has little effect on the total network. Consequently, the convolutional kernel width taking an intermediate value is 5. Secondly, these combinations have almost no regular impact on the performance of the model. This means that smaller changes in the filter kernel width and the number of channels have little effect on the model. Under the premise of considering the GFOLPs, the number of parameters and the accuracy, if try to understand which factor of the filter kernel width and the number of channels has a greater affect when the model is extracting the features, we analyze two different combinations of the form in the following experiments: 1. A large filter paired with fewer number of channels; 2. A regular sized filter paired with a regular number of channels. The cases of large

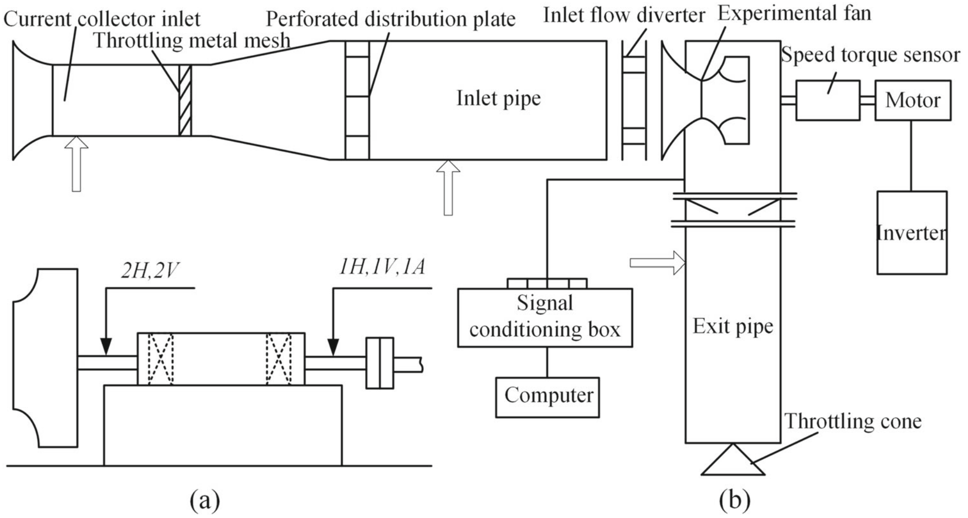


Fig. 8 (a) is a schematic diagram of the sensor installation position. (b) is a schematic diagram of the fan structure

Table 2 Fault status description of fan

Condition	Label Position		Commentary
Normal	0	–	–
Rotor unbalanced (1)	1	Outer edge of the front disc	Installed masses at edge of the front disc to simulate the unbalanced faults of the rotor with different severity and positions.
Rotor unbalanced (2)	2		
Rotor unbalanced (3)	3		
Rotor unbalanced (4)	4		
Angular misalignment	5	Rigid coupling	The experiment simulated faults of the angular misalignment, parallel misalignment (light) and parallel misalignment (serious) for the coupling.
Parallel misalignment (light)	6		
Parallel misalignment (serious)	7		
Bearing loose (all loose slight)	8	Different positions of the bearing	For 10 bolts in different parts of the bearing, changed their tightness to simulate the looseness of the bearing at different degrees and positions.
Bearing loose (all loose serious)	9		
Bearing loose (the left loose)	10		
Static and kinetic friction (1)	11	Between collector and front disc	Different degrees of static and kinetic friction were tested for single-point, multi-point and local face.
Static and kinetic friction (2)	12		

Table 3 Rotor fault data

Condition	Label	Commentary
Horizontal Misalignment	0	This type of fault was induced into the MFS by shifting the motor shaft horizontally of 0.5 mm. Used the same range for the rotation frequency as in the normal operation for each horizontal shift.
Imbalance	1	Simulated imbalanced fault with 6g load.
Normal	2	There were 49 sequences without any fault, each with a fixed rotation speed within the range from 737 to 3686 r/min with steps of approximately 60 r/min.
Overhang Bearing	3	It was placed in locations on the MFS bench: in the external position?having the rotor between the bearing and the motor (overhang position).
Underhang Bearing	4	It was placed in locations on the MFS bench: between the rotor and the motor (underhang position).
Vertical Misalignment	5	This type of fault was induced into the MFS by shifting the motor shaft horizontally of 0.51 mm. Used the same range for the rotation frequency as in the normal operation for each vertical shift.

Table 4 The impact of if size on classification accuracy

Kernel width	0dB		-4dB		Computational complexity	
	Vali	test	vali	test	GFOLPs	Params(M)
25	88.43	90.8	72.11	71.91	1.55	0.07
27	90.1	90.5	73.43	73.36	1.56	0.07
29	**92.31**	**91.29**	**76.80**	**76.14**	**1.58**	**0.07**
31	92.7	90.06	75.98	76.10	1.60	0.07
33	92.55	92.12	76.73	76.57	1.61	0.07
35	93.62	91.23	73.75	74.88	1.63	0.07

Table 5 The impact of combination parameter on classification accuracy

Combination Parameter	0dB		-4dB		Computational complexity	
	Vali	test	Vali	test	GFOLPs	Params(M)
16,17,5,3	92.61	90.39	77.82	76.68	4.57	0.24
18,21,5,3	92.25	91.23	76.45	77.93	6.05	0.32
16,21,5,3	92.71	92.12	77.34	76.8	5.04	0.27
16,17,5,5	92.19	91.83	77.46	76.5	4.58	0.24
16,21,7,5	**93.06**	**92.06**	**78.29**	**77.51**	**5.05**	**0.27**
14,19,7,5	92.67	91.85	77.28	75.59	3.96	0.21
14,25,7,5	92.58	92.32	77.28	75.84	4.50	0.23

Table 6 Combination parameter selection

Parameter combination	0dB		-4dB		Computational complexity	
	Vali	test	Vali	test	GFOLPs	Params(M)
29-32-21-16	92.95	92.77	78.11	77.89	5.04	0.27
99-16-95-8	**93.68**	**93.97**	**80.38**	**80.52**	**3.83**	**0.18**

Table 7 DF-Block quantity selection

Number of blocks	0dB		-4dB		Computational complexity	
	Vali	test	vali	test	GFOLPs	Params(M)
1	92.44	92.49	77.94	78.89	2.31	0.06
2	92.67	93.73	79.67	80.50	3.32	0.12
3	**93.68**	**93.97**	**80.38**	**80.52**	**3.83**	**0.18**
4	92.92	93.14	79.61	80.79	4.09	0.24
5	92.17	92.54	78.89	79.27	4.21	0.30

Table 8 DAM quantity selection

Number of blocks	0dB		-4dB		Computational complexity	
	Vali	test	Vali	test	GFOLPs	Params(M)
2	**93.68**	**93.97**	**80.38**	**80.52**	**3.83**	**0.18**
3	93.16	93.28	79.97	80.67	5.88	0.29
4	92.89	93.29	79.58	79.88	8.25	0.43
5	91.84	92.54	78.73	79.22	11.01	0.58

Table 9 Bayesian optimization ideas

Algorithm: Optimization process

Require: AC ← acquisition function; f ← corresponding function relationship; RH ← hyper-parameter search space; GP ← Gaussian Process; ε_t ← a zero-mean Gaussian distribution.

Procedure:

```
for $t$ in 1,2, ...,$T$ do
    Optimizing $AC$ on GP: $x_t \leftarrow \arg\max_{X \in RH} AC(x \mid D_{1:t-1})$;
    Function value: $y_t \leftarrow f(x_t) + \varepsilon_t$;
    Update the data $D_{1:t} \leftarrow D_{1:t-1} \cup (x_t, y_t)$;
end for
```

filters with a regular number of channels and regular filters with a small number of channels are not considered, because they either have excessive computational complexity or substandard accuracy. Consequently, based on the above experimental results, we set two sets of parameter combinations [kernel width of IF - number of output channels of IF - kernel width of EF - k] for comparison.

Table 6 comparison results reveal that, on the one hand, the combination of large filter with fewer number of channels has a better recognition ability than regular combinations under both 0dB and -4dB data conditions. In addition, with increasing noise, from 0dB to -4dB, the test accuracy of the regular combination decreases by about 15%, while the accuracy of the combination of a large filter with fewer number of channels decreases by about 13%. This is because large filters are more conducive to capturing feature variations over long time scales contained in fault signal data, and the ability to detect and identify global patterns and trends in fault signals helps to improve model identification in noisy environments. On the other hand, the combination of large filters with fewer number of channels is significantly better than the regular combination on GFOLPs and Params. This further indicates that the framework and parameters of the combination are appropriate for the model and can effectively balance the complexity and generalization ability. And more shows reducing the channels is a good choice if we want to not increase the model volume and computational complexity.

After determining the number of filters and channels in the model structure, we proceed to determine the number of DF-Block structures.Table 7 experimental shows that the overall diagnostic ability of the model is best when the number of DF-Block is 3. And the computational complexity is at least 0.26 GFOLPs and 0.06M parameter quantity lower compared to other blocks with excellent performance.

Based on 3 DF-Blocks, we determine the number of DAM. In DF-Block, in order to reflect the role of dense connection, the number of DAM and H needs to be set $\geq$ 2. Table 8 shows that as the number of DAM increases, the recognition ability of the model decreases, which is the same reason why the accuracy of the model decreases when continuing to increase the number of DF-Block. Too large of non-linear mappings lead to overfitting phenomenon, so choose 2 DAMs for best. Compared to 3 DAMs, there is a slight disadvantage in accuracy, but the difference in GFOLPs is about 2, and the difference in Params is 0.11M, which is a clear advantage in computational complexity. In summary, the model performance is more prominent when 2 DAMs are chosen.

3.2.2 Bayesian optimization models automatically select model hyperparameter

Due to the fact that it is unrealistic to expect any learning model to achieve the best generalization on all datasets, it is necessary to automatically optimize hyperparameters. In this paper, hyperparameter optimization is achieved by employing a Bayesian network with Gaussian regression processes as a proxy models. Table 9 shows the pseudo-code. Table 10 shows some selected results. Table 11 shows the structure and parameters of the diagnostic network. Experimental comparisons present the process and results of parameter selection very clearly, however, the process is complicated and time consuming. The neural network optimization approach rapidly attains the best optimization values but lacks an understanding of the internal selection rules. Combining these two methods is more effective in providing a better understanding of network configuration.

3.3 Ablation experiment

We use centrifugal fan data with SNR = 0dB and SNR = -4dB to verify the distributed attention hybrid attention function in the designed DFANet. So as to avoid the influence of other factors on the experiment, the DenseNet with the same hyperparameters are selected for comparison.Meanwhile, set an early stop strategy during model training, and set the threshold to 20, which means that when the maximum value is reached and the last 20 values are not higher than this maximum value, the model stops. Figure 9(a) shows that the proposed network reaches its maximum value of 80.38% at 80 Epochs, while DenseNet reaches its maximum value of 77.43% at 89 Epochs. This explains that the proposed model is more likely to converge to the optimal solution in fewer epochs. The accuracy of the proposed model is about 3% higher than the DenseNet. This is because the concatenation fusion mechanism in DenseNet does not have the ability to selectively extract features, and too many repeated extractions and non-intervention mergers lead to a large amount of redundancy, where the interfering information is dense and the percentage of useful information is reduced. In addition, it can also be concluded from Fig. 9(b) that the accuracy of the proposed model decreases slower than that of DenseNet from 0dB to -4dB, reflecting the fact that the proposed model performs well in capturing generalized patterns in the data in a strongly noisy environment, rather than overly relying on specific samples in the training set. Therefore, it is concluded that the strategy of adding distributed attention hybrid

Table 10 Results of bayesian network optimization

The selected results	Commentary
Activation function	*SeLU*
Learning rate	0.00011
Optimizer	Adam
Loss function	Classification cross entropy

Table 11 The layers of the network and the main attributes

Layer	Output Shape	Connected to
Input Layer	(None, 1, 1024)	–
Conv1D	(None, 16, 512)	Input Layer
MaxPooling1D	(None, 16, 256)	Conv1D
DF-Block_1	(None, 24, 256)	MaxPooling1D
	(None, 32, 256)	
Connection layer_1	(None, 16, 128)	DF-Block_1
DF-Block_2	(None, 24, 128)	Connection layer
	(None, 32, 128)	
Connection layer_2	(None, 16, 64)	DF-Block_2
DF-Block_3	(None, 24, 64)	Connection layer
	(None, 32, 64)	
G_AvePooling1D	(None, 32)	DF-Block_3
Softmax	(None,n-classes)	G_AvePooling1D

attention function in dense can improve the generalization ability of the model in different environments.

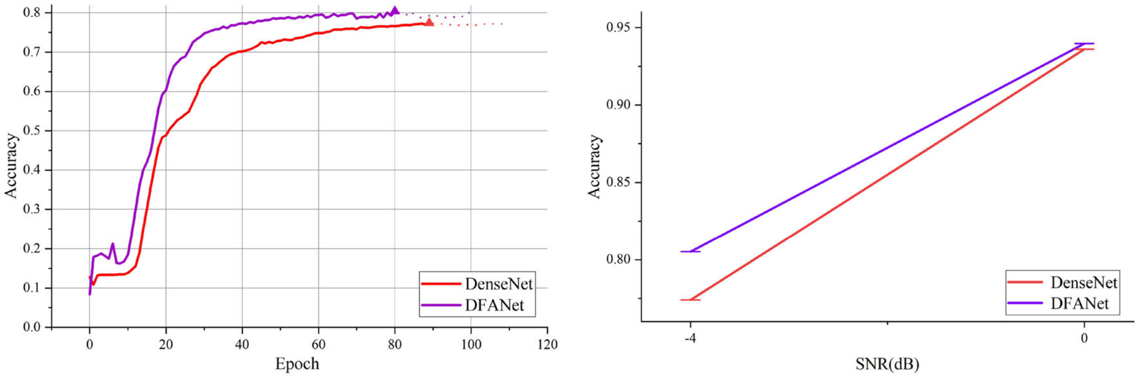
Continue to discuss the influence of the structure in DFANet on the quality of fault identification, including the spatial attention and channel attention of different connection methods, the ZSSAM structure and the CAM structure. This experiment compares 4 network structures: (1) This article proposes DFANet; (2) D-CAM is defined as a dense network with channel attention; (3) D-ZSSAM is defined as a dense network with zoomed space self-attention; (4) D-JAM [16] is a single-scale linearly connected multi-attention dense network. Conduct ten experiments and select the best trained set of models for comparison.

Figure 10(a) results show that DFANet has higher recognition accuracy than D-ZSSAM, D-CAM and D-JAM by 1.12%, 2.02% and 1.46% respectively. It is important to mention that the accuracy of D-ZSSAM is 0.35% higher than that of D-JAM. The feature expression of the ZSSAM module is stronger than that of the linear multi-attention module. As the results show, the linear multi-attention module has the problem of feature activation failure, which results in a decrease in diagnostic efficiency. Moreover, the DAM composed of ZSSAM and CAM performs distributed attention, while using a dense structure to achieve mixed attention. This method can greatly enhance the diagnostic quality of the network. It is worth mentioning that D-JAM converges to the optimum at 37 iterations, which is related to the structure of JAM and the location of the embedding into DenseNet. After applying JAM to densely concatenation, directly compiling weights for the complete feature set can enhance the non-linear mapping capability of JAM. Therefore, it achieves optimal solutions with fewer epochs, but at the same time a large amount of computational complexity is incurred. In addition, Fig. 10(b) shows that the accuracy of the proposed model decreases slower than the other three ablation models from 0 dB to -4 dB, reflecting that the proposed structure is more robust in a noisy environment.

3.4 Comparative study

This section uses -4dB centrifugal fan data to compare a total of 8 typical networks, including the method proposed in the paper.

CDCN [12] is a densely connected convolutional neural network with multiple inputs used for analyzing multi-sensor data. The input of the original paper is the vibration signal collected by the accelerometer and the speed signal collected by the rotary encoder. The network convolution kernel has a width of 9 and k=16. The two path structures are the same, each consisting of three dense blocks, each containing two abstract mappings. Because our fan data do not collect speed signals, we use vibration data from sensors 1 and 3 as input for the networks multipath. Secondly, we conduct a two-stage experiment on this network. In the first stage, the original article version is not modified, and in the second stage, the convolution kernel is increased from 9 to 91,



(a) (b)

Fig. 9 (a) Convergence curves of the validation set for DFANet and DenseNet. (b) Anti noise performance of DFANet and DenseNet

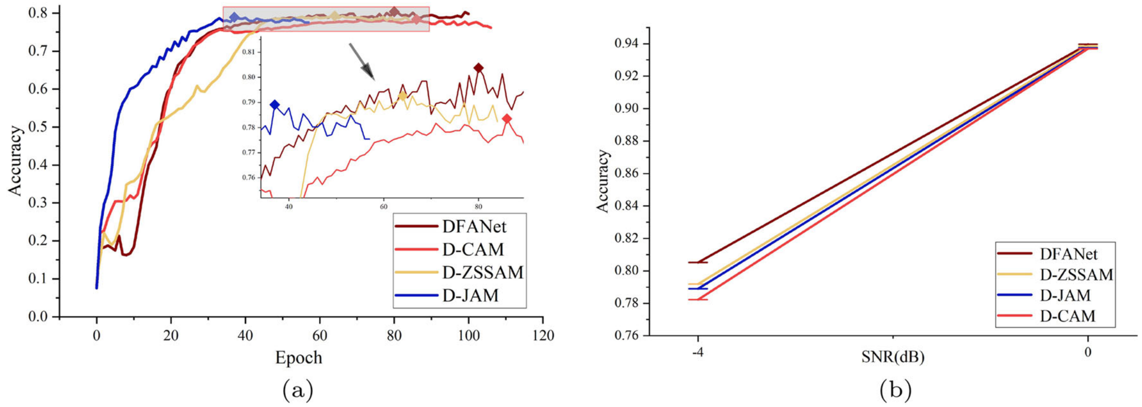


Fig. 10 (a) Comparison of ablation experiments. (b) Anti-noise experiment of ablation network

while the rest remained unchanged. CNN-CBAM is a combined network of traditional CNN and CBAM. CBAM is a classical concatenated multi-attention module, which is quite different from the attention mechanism of the DAM of the proposed network, allowing for cross-analysis. The network is initially set up with four convolutional layers and no pooling layer. Reducing the size of the convolutional layers mainly sets the convolutional kernel sliding step to 2 to achieve the effect of reducing the risk of overfitting. The number of output channels of the convolutional layers is set to 32-64-64-128. The CBAM [29] module is installed after the last convolutional layer of the CNN to calibrate the attentional weights on the feature set before decision making. We perform three stages of validation for this network, the first, second and third stages, with the convolutional layer filter sizes set to 17 × 1,51 × 1,91 × 1, and the number of nonlinear compressions in the CBAM set to 16,4,16, respectively. The network training is preset for 500 epochs. MSCNN [30] uses a traditional multi-scale mapping model, but its convolutional kernel width uses large size 100, which is like the large convolutional kernel of the proposed method for extracting features, and thus can be compared together.IMS-FACNN [31] is an improved network to MSCNN by adding the effect of an attention mechanism in the scale fusion layer, which reduces network redundancy. LM-MDINet [20] is a lightweight densely connected convolutional neural network. It introduces a hybrid strategy of deep separable convolution and regular convolution. It can effectively balance the accuracy and volume. At the same time, it adds a M-MDI attention layer to try to find discriminative information from the interference signal. The JAM module in MA1DCNN is also a kind of tandem multi-attention, which can be regarded as an upgraded version of the CBAM module. The compilation function within the attention mechanism is lightweight optimized and residual connections are added to improve the attention effect of the whole module. Wen-CNN [32] is a traditional approach to fault diagnosis using image data, and has a very typical reference to all of the above networks with different data patterns. None of the above five models construct a stage analysis strategy, and the model structure is the same as the original.

The Table 12 shows the best recognition results of 8 networks in ten experiments. Under -4dB fan data, the accuracy of IMS-FACNN, MSCNN, and the proposed model DFANet is about 80%, demonstrating outstanding noise resistance performance. However, the GFLOPs of IMS-FACNN and MSCNN are 2.6 times and 9.0 times higher than DFANet, respectively. The parameter count of IMS-FACNN and MSCNN is 2.2 times and 329.4 times that of DFANet, respectively. So DFANet could better balance computational

Table 12 Experimental results

Network name	SNR=-4dB Accuracy %	GFLOPs	Params(M)
DFANet	**80.52**	**3.83**	**0.18**
CDCN	65.83-68.44	13.7-120.32	0.19-1.77
CNN-CBAM	68.04-76.03-81.10	13.86-41.53-74.08	0.94-2.8-5
IMS-FACNN	79.73	10.00	0.40
MSCNN	80.86	34.29	59.29
LM-MDINet	74.06	2.75	0.15
MA1DCNN	77.10	36.94	1.19
Wen-CNN	66.92	10.05	19.03

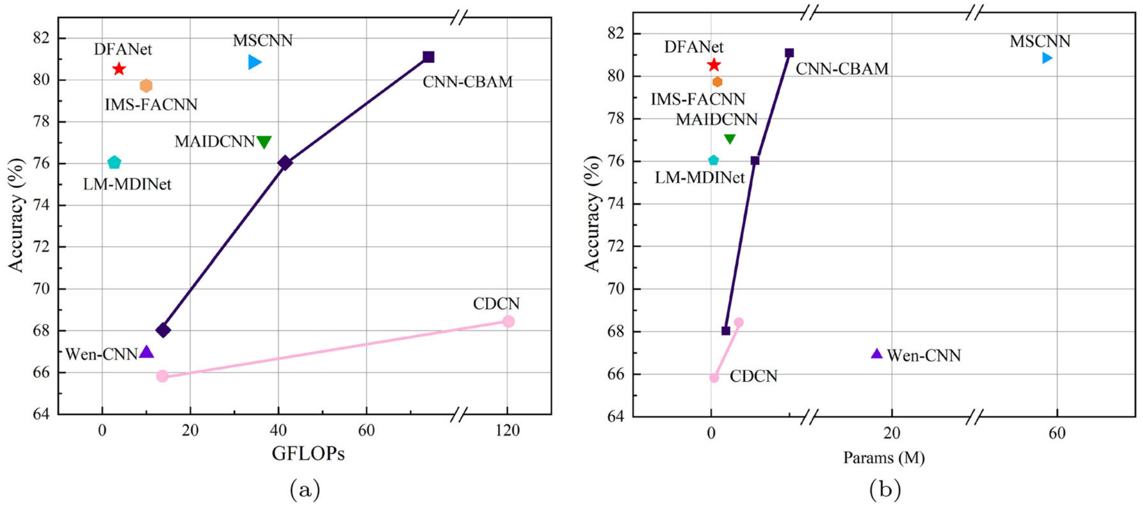


Fig. 11 Comparison results. (a) X-axis is the computational complexity, which is measured using GFLOPs, and the Y-axis is the accuracy. (b) X-axis is the memory occupation, which is measured using Params, and the Y-axis is the accuracy

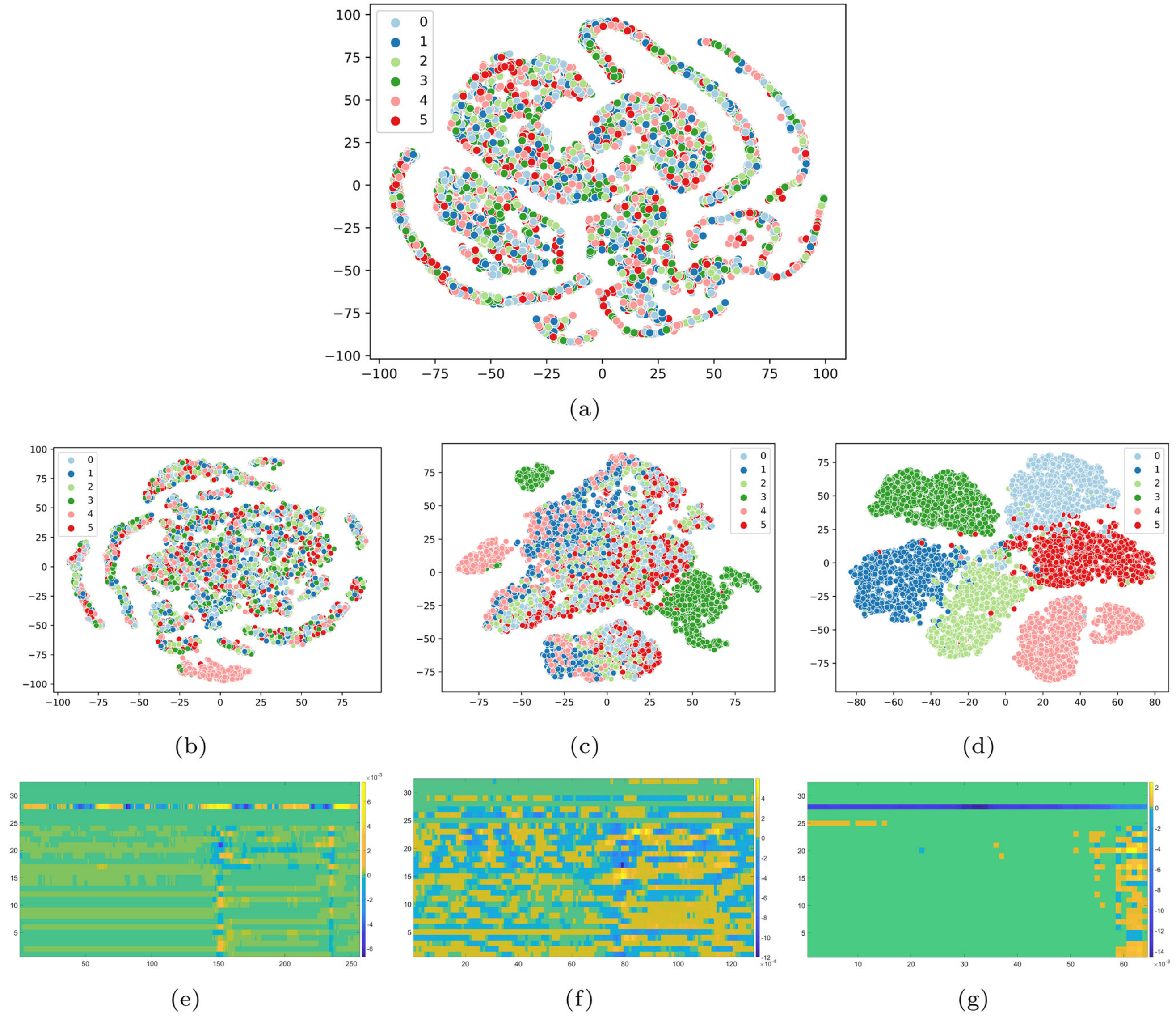
complexity and accuracy. The results of LM-MDINet and MA1DCNN are also acceptable, as the attention mechanism added to the network can guide the model towards capturing the characteristics of various inherent oscillation modes, but its ability is slightly insufficient. The size and quantity of convolutional and corresponding learned feature maps in MA1DCNN are negatively correlated, resulting in a low number of parameters. Meanwhile, non-linear weight calibration using multiple attention modules will increase model computation. The remaining CDCN, CNN-CBAM, and Wen CNN have the lowest overall performance. The reason for the poor performance of Wen CNN is that the incoming data transforms the fault signal into an image, resulting in information loss during conversion.

We increased the convolution kernels of CDCN and CNN-CBAM to verify that large filters help improve the generalization ability of the model, while keeping the other variables unchanged, resulting in results at different stages. As shown in the Fig. 11, increasing the convolutional kernel can enhance the recognition ability of CDCN and CNN-CBAM. However, the improvement level of CDCN is not as good as that of CNN-CBAM, because CDCN is a multi-input network, and the features of multiple sensors have their own trends. This trend will not decrease with changes in the convolution kernel, and may cause confusion during fusion. Perhaps standardizing the data and bringing the clustering centers of the two sensor features closer together can solve this problem. In addition, the computational complexity is shown to explode as the convolutional kernel grows larger. This is because a large convolution extracting a feature once will contain a very large number of multiplication operations and the number of channels in the setup is relatively large and shows a doubling with the increase in abstraction ability. Although the method proposed in this article also uses a large filter, the network has a very small k and the compression rate of the transfer layer is set to 0.5. This keeps the number of input channels per dense block constant and 16. In summary, a small number of channels with a large filter improves the accuracy of the model without creating too much computational burden.

3.5 Experimental verification

Next, we will discuss the interpretability in the diagnostic process of the model. The researchers want to employ human knowledge structures to interpret the portions of machine decisions that diverge from human intentions, thereby enhancing the confidence of the model. Previously, Ref. [16] displayed that the model captures the impulse excitation and improves the interpretability of the model. Similarly, Yang et al. [33] used the method of labeling the signal weight output values. Nevertheless, these methods are only effective when the pulse segment is clear. Due to the strong noise that often overwhelms the collected fault signals, it's challenging to highlight the pulse segment. As a result, these methods are less reliable. As we all know, the rationality of the model determines the quality of feature extraction. The expression of features is a continuous iterative process of weights. Consequently, when researching the interpretability of the model, we attempt to comprehend how attention modules guide feature weights during the model's operation. On this basis, we calculate the changes in calibration error and, in conjunction with diagnostic trends, illustrates the process of weight calibration for the multi-attention mechanism. Referring to the weight change process chart with the best diagnostic performance, identify unusual points or areas in the space and channel weight responses. Subsequently, factors influencing model identification are identified and used to enhance the interpretability of the learning model.

The rotor data with SNR=50dB is used for experimental verification, and the process of diagnosing rotor faults using DFANet model is displayed. As illustrated in Fig. 12, the ordinate is the spatial dimension, and the abscissa is the channel dimension. We jointly reflect the correction of feature weights from the spatial and channel dimension. Figure 12 (e), (f) and (g) show the visualization results, corresponding to the second, fourth and sixth DAM dynamic calibration before and after respectively. The weight information of every DAM module is Cp before calibration and Ca after calibration, and the calibration value Ce=Ca-Cp. Ce>0 means that the network will strengthen the display of the feature, and Ce<0 means that the feature will be weakened. The process is one of continually revising the expression of information. The result of t-SNE shown in Fig.13 further prove that this process is conducive to achieving excellent diagnostic performance. Clusters of different colors represent various types of rotor fault classification results. Figure 12 (a), (b), (c) and (d) are the original data, and the second, fourth, and sixth DAM blocks output feature visualization results. Apparently, the same type of fault keep clustering together as the DAM encodes the information multiple times. It shows that in DFANet's DAM, the space and channel disperse attention combining dense connections to form mixed attention can better selectively express fault information.



(a)

(b)

(c)

(d)

(e)

(f)

(g)

Fig. 12 Visualization display

4 Conclusion

The DFANet proposed in this article is used for operating status diagnosis of centrifugal fan. Experiments verified the superiority of the proposed method. The conclusions are summarized as follows: (1) In the DAM module, the fault features are dispersedly noticed the channel dimension and the space dimension to eliminate the mutual influence between the linear multi-attention modules. The DAM module replaces the traditional dense connection method. The formed DFANet has the effect of multiple mixed attentions, which can more effectively highlight device fault features and suppress irrelevant features. (2) Through experimental verification, the constructed network solves the problem of the current diagnosis model proposed in the first part. (3) The constructed interpretability method focuses on dynamic explanation in the process and attempts to establish a connection between macro diagnostic effects and micro theory. In future research, we will study the transfer learning of this network in the source and target domains of different data distributions. And continue to explore other methods, focusing on the feature extraction of time-domain signals.

Author Contributions Ruijun Wang: Conceptualization, Methodology, Writing-Original Draft. Yuan Liu: Software, Validation, Visualization, Formal analysis. Zhixia Fan: Methodology, Software, Validation, Visualization. Xiaogang Xu: Supervision, Funding acquisition. Huijie Wang: Project administration, Investigation.

Data Availability The datasets generated or analyzed in the process of this paper can be obtained from the corresponding author upon reasonable request for a report of the current study.

Declarations

Declaration of interest statement We declare that there is no actual or potential conflict of interest between us and any other person or organization, including financial, personal or other relationships with others or organizations, which may improperly affect or believe to affect their work.

References

1. Xu X, Qi M, Liu H (2019) Real-time stall detection of centrifugal fan based on symmetrized dot pattern analysis and image matching. Measurement 146:437–446
2. Peng B, Xia H, Lv X, Annor-Nyarko M, Zhu S, Liu Y, Zhang J (2022) An intelligent fault diagnosis method for rotating machinery based on data fusion and deep residual neural network. Appl Intell 52:3051–3065
3. Xue Y, Dou D, Yang J (2020) Multi-fault diagnosis of rotating machinery based on deep convolution neural network and support vector machine. Measurement 156:107571
4. Zhang D, Xiang W, Cao Q, Shiyi C (2021) Application of incremental support vector regression based on optimal training subset and improved particle swarm optimization algorithm in real-time sensor fault diagnosis. Appl Intell 51:3323–3338
5. Lei Y, Jia F, Lin J, Xing S, Ding SX (2016) An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Trans Industr Electron 63:3137–3147
6. Niu G, Liu E, Wang X, Ziehl P, Zhang B (2023) Enhanced discriminate feature learning deep residual cnn for multitask bearing fault diagnosis with information fusion. IEEE Trans Industr Inf 19:762–770
7. Hinton EGR, Salakhutdinov R (2006) Reducing the dimensionality of data with neural networks. Science 313:504–507
8. Khorram A, Khalooei M, Rezghi M (2021) End-to-end cnn + lstm deep learning approach for bearing fault diagnosis. Appl Intell 51:736–751
9. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
10. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX (2019) Deep learning and its applications to machine health monitoring. Mech Syst Signal Process 115:213–237
11. Li F, Wang L, Wang D, Wu J, Zhao H (2023) An adaptive multiscale fully convolutional network for bearing fault diagnosis under noisy environments. Measurement 216:337–346
12. Jiao J, Zhao M, Lin J, Ding C (2019) Deep coupled dense convolutional network with complementary data for intelligent fault diagnosis. IEEE Trans Industr Electron 66:9858–9867
13. Wang P, Sertel E (2021) Channel–spatial attention-based pan-sharpening of very high-resolution satellite images. Knowl-Based Syst 229:107324
14. Wang X, Tang M, Yang T, Wang Z (2021) A novel network with multiple attention mechanisms for aspect-level sentiment analysis. Knowl-Based Syst 227:107196
15. Li X, Wang W, Hu X, Yang J (2019) Selective kernel networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognitin (CVPR), 510–519
16. Wang H, Liu Z, Peng D, Qin Y (2020) Understanding and learning discriminant features based on multiattention 1dcnn for wheelset bearing fault diagnosis. IEEE Trans Industr Inf 16: 5735–5745
17. Hongmin Gao YYXC, Mingxia Wang, Li C (2021) Hyperspectral image classification with dual attention dense residual network. Int J Remote Sens 42:5604–5625
18. Fan Z, Xu X, Wang R, Wang H (2022) Fan fault diagnosis based on lightweight multiscale multiattention feature fusion network. IEEE Trans Industr Inf, 4542–4554
19. Shen D, Jiayi Zhao, Liu S, Cui Z (2024) Multiscale attention feature fusion network for rolling bearing fault diagnosis under variable speed conditions. Signal, Image and Video P, 8–21
20. Zhu X, Wang R, Fan Z, Xia D, Liu Z, Li Z (2022) Gearbox fault identification based on lightweight multivariate multidirectional induction network. Measurement, 110977
21. Mathew MPSFJAFME J, Bouchard PJ (2017) Through-thickness residual stress profiles in austenitic stainless steel welds: A combined experimental and prediction study. Metall Mater Trans A 48(12):6178–6191
22. Block SB, Silva RD, Dorini LB, Minetto R (2021) Inspection of imprint defects in stamped metal surfaces using deep learning and tracking. IEEE Trans Industr Electron 68(5):4498–4507
23. Shang R, He J, Wang J, Xu K, Jiao L, Stolkin R (2020) Dense connection and depthwise separable convolution based cnn for polarimetric sar image classification. Knowl-Based Syst 194:105542
24. Cai Y, Zhang Z, Yan Q, Zhang D, Banu MJ (2021) Densely connected convolutional extreme learning machine for hyperspectral image classification. Neurocomputing 434(86):21–32
25. Zhao M, Zhong S, Fu X, Tang B, Dong S, Pecht M (2021) Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis. IEEE Trans Industr Electron 68(3):2587–2597
26. Yu J, Zhou X (2020) One-dimensional residual convolutional autoencoder based feature learning for gearbox fault diagnosis. IEEE Trans Industr Inf 16(10):6347–6358

27. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. 2016 IEEE Comput Soc Conf Comput Vis Pattern Recognit (CVPR), 770–778
28. Xu X, Wang R, Fan Z, Ma X, Zhao Z, Wang H (2024) Ms-drt: A multilevel and multiscale branch learning scheme for fault diagnosis of rotating machinery. IEEE Trans Industr Inf, 2799–2811
29. Woo S, Park J, Lee J-Y, Kweon IS (2018) Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV), 3–19
30. Jiang G, He H, Yan J, Xie P (2019) Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. IEEE Trans Industr Elec, 3196–3207
31. Xu Z, Li C, Yang Y (2021) Fault diagnosis of rolling bearings using an improved multi-scale convolutional neural network with feature attention mechanism. ISA Trans 110:379–393
32. Wen L, Li X, Gao L, Zhang Y (2018) A new convolutional neural network-based data-driven fault diagnosis method. IEEE Trans Industr Elec, 5990–5998
33. Yang Z-b, Zhang J-p, Zhao Z-b, Zhai Z, Chen X-f (2020) Interpreting network knowledge with attention mechanism for bearing fault diagnosis. Appl Soft Comp 97:106829

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ruijun Wang was born in Inner Mongolia Autonomous Region, China. He received the B.S. degree in building environment and energy application engineering from the Inner Mongolia University of Technology, Hohhot, China, in 2019, and the master's degree in energy and power from North China Electric Power University, Baoding, China, in 2023. His current research interests include computer vision and pattern recognition, machine learning, artificial intelligence, industrial image defect detection, and fault diagnosis of industrial equipment.

Yuan Liu was born in Inner Mongolia Autonomous Region, China. He obtained a bachelor's degree from Xi'an Shiyou University in 2021 and a master's degree in Instrument Science and Technology from Xi'an Shiyou University in 2024. His current research interests include power electronic condition monitoring, machine learning, artificial intelligence, and industrial equipment fault diagnosis.

Zhixia Fan was born in Inner Mongolia Autonomous Region, China. She received the B.S. degree in transportation specialty from the Inner Mongolia University of Technology, Hohhot, China, in 2017, and the master's degree in power engineering from North China Electric Power University, Baoding, China, in 2022. Her current research interests include computer vision and pattern recognition, machine learning, artificial intelligence, industrial image defect detection, and fault diagnosis of industrial equipment.

Xiaogang Xu received the B.S. and M.S. degrees in automation and pattern recognition and intelligent system major, and the Ph.D. degree in thermal energy engineering from North China Electric Power University, Baoding, China, in 2002, 2005, and 2014, respectively. He was a Post-doctor with the School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing, China. He is currently a Senior Engineer and Master Tutor. His main research interests include power plant thermal system performance calculation, system optimization, and intelligent fault diagnosis algorithm research.

Huijie Wang received the M.S. and Ph.D. degrees in thermal energy engineering from North China Electric Power University, Baoding, China, in 1999 and 2009, respectively. His main scientific research interests include thermal economic performance analysis of power plant thermal system, calculation software development, thermal equipment characteristic modeling method, operation optimization, and key technologies of comprehensive utilization of solar photovoltaic; photothermal system economic performance analysis.