

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/172139/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Chen, Chong, Wang, Tao, Lu, Kaijie, Liu, Ying and Cheng, Lianglun 2024. Compact convolutional transformers- generative adversarial network for compound fault diagnosis of industrial robot. Engineering Applications of Artificial Intelligence 138 (PartA) , 109315. 10.1016/j.engappai.2024.109315

Publishers page: <http://dx.doi.org/10.1016/j.engappai.2024.109315>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Compact Convolutional Transformers- Generative Adversarial Network for Compound Fault Diagnosis of Industrial Robot

Chong Chen^a, Tao Wang^{a,b*}, Kaijie Lu^a, Ying Liu^c, and Lianglun Cheng^a

^a *Guangdong Provincial Key Laboratory of Cyber-Physical System, Guangdong University of Technology, Guangzhou 510006, China*

^b *School of Automation, Guangdong University of Technology, Guangzhou 510006, China.*

^c *Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK*

**Corresponding author*

E-mail: wangtao_cps@gdut.edu.cn

Abstract

The safe operation of Industrial robots is a major concern in intelligent manufacturing. Accurate compound fault diagnosis is essential to the safe operation of industrial robots, while it is challenging to achieve since the compound fault samples are hard to be collected. Generative adversarial network (GAN) is a useful tool for addressing the data imbalance issue. However, the computation efficiency of GAN in addressing the data imbalance issue has not been investigated. Hence, this study proposes a lightweight GAN named compact convolutional Transformers-GAN (CCT-GAN) to alleviate the data imbalance issue in compound fault diagnosis modelling. Firstly, the feedback current signals collected from the industrial robot are transformed into time-frequency images via continuous wavelet transformation (CWT). Secondly, CCT-GAN is designed to achieve high-quality fake data generation and compound fault diagnosis modelling without large computational costs. Thirdly, the relation between a single fault and the compound fault is considered in the compound fault diagnosis modelling via multi-hot representation to alleviate the data imbalance issue. An experimental study based on the real-world compound fault dataset of industrial robots reveals that the proposed CCT-GAN shows merits in compound fault diagnosis modelling in comparison with the prevailing algorithms. The results indicate that CCT-GAN can performance of compound fault diagnosis when only 100 data samples from each compound fault category are available.

Keywords: Compound fault diagnosis; Generative adversarial network; Industrial robot; Data imbalance.

1. Introduction

The recent advance in automation and intelligence has witnessed the rapid development of industrial robots (Yoon et al., 2023). Industrial robots have been widely deployed in various manufacturing tasks such as assembly (Zhang et al., 2021) and painting (Gleeson et al., 2022). The high precision and reliability of the industrial robot have greatly improved the productivity of production (Chen et al., 2024). For the industrial robot, the fault diagnosis of its transmission system is essential to the operation and maintenance of the industry. The complexity of industrial robots arises from their composition of multiple interconnected subsystems including mechanical and electrical components. This integration results in complex and nonlinear relationships between components. The tight coupling between an industrial robot's transmission system means failures can propagate through multiple components, leading to compound faults. Hence, it is challenging to identify faulty parts since compound faults produce symptoms that can be unclear or misleading. Meanwhile, the complex nature of industrial robots leads to an imbalanced distribution of fault types. The alleviation of the data imbalance issue is helpful to the compound fault diagnosis modelling (Wang et al., 2023a). Therefore, it is necessary to develop an intelligent approach that can address the data imbalance issue in compound fault diagnosis modelling.

Recent research on compound fault diagnosis is mainly based on data-driven approaches (Chen et al., 2023a). With the advance of machine learning and signal processing techniques, researchers have focused on establishing a compound fault diagnosis model based on sensory data (Yan et al., 2024b). Among the data-driven methods, deep learning has gained increasing

attention in fault diagnosis modelling. The existing studies on compound fault diagnosis mainly focus on the gearbox (Yan et al., 2024a) and bearing (Shao et al., 2024), which failure data can be obtained in the run-to-failure test. However, the faulty data for the industrial robot is hard to obtain because the robot is not allowed to work for a long time in faulty conditions. Because of that, the ratio between the normal operational data and the faulty data is extremely dramatic. The data imbalance issue may greatly jeopardize the compound fault diagnosis modelling performance using deep learning approaches (Chen et al., 2023b). In the literature, various methods have been proposed to address the data imbalance issue. The most common way to address this issue is oversampling the minority class. However, it cannot generate samples which are diverse and follow the distribution of the original data. Apart from that, two types of prevailing approaches which are transfer learning and GAN have been widely studied to address the data imbalance issue in fault diagnosis modelling. Transfer learning can transfer the knowledge in a pre-trained model to a new scenario and use limited data to fine-tune the model (Chen et al., 2023c). However, the deployment of transfer learning still requires a large collection of faulty data in the source domain. Another prevailing method towards imbalanced data modelling is GAN, which can generate synthetic samples to expand the minority class (Ren et al., 2023). Compared with the existing data augmentation approaches such as oversampling and mixup, GAN can learn the distribution of the original data and generate fake samples, which is more scaled, adaptive and non-linear (Dixit et al., 2021; Wang et al., 2023b). Among different variants of GAN, Auxiliary classifier GAN (ACGAN) is prevailing in fault diagnosis. With an auxiliary classifier, the loss of fault diagnosis can be used to support the

training of the generator, and the generated samples are further fed into the auxiliary classifier to improve the fault diagnosis performance.

However, there are two challenges for compound fault diagnosis modelling based on imbalanced data: (1) The current GANs used for fault diagnosis primarily rely on autoencoder or convolutional neural network (CNN) with residual connections for their generator and discriminator. However, these networks may not be able to effectively extract temporal-spatial features when they are too shallow. On the other hand, designing a deeper network can lead to larger network sizes and increased complexity in GAN training. Therefore, the challenge lies in balancing the network complexity and the performance of GAN to achieve optimal temporal-spatial feature extraction for sample generation and fault diagnosis; (2) Introducing an auxiliary classifier into GAN can improve the performance of sample generation, while it also increasing the complexity of the training process. In order to achieve better algorithm performance and training stability, a suitable loss function needs to be designed.

Recently, Transformers network has gained increasing attention in fault diagnosis (Jin and Chen, 2021) and remaining useful life prediction (Zhou et al., 2022). Transformers network can effectively capture the faulty patterns in the sensory data. The identification of faulty patterns is helpful to improve the performance of GAN. However, two main drawbacks of the Transformers network are the large computational cost and the requirement for a large number of labelled data. In order to address both issues, a compact convolutional Transformers (CCT) was proposed (Hassani et al., 2021). CCT adopts convolutional blocks for feature reduction and

Transformers blocks for sequential feature learning. Furthermore, a sequence pooling strategy is adopted to improve the sequence learning capability. The advance of CCT can leverage the cost-efficiency of ACGAN for the compound fault diagnosis of industrial robots. Hence, with the advance of CCT and ACGAN, a CCT-GAN is proposed in this study. The proposed CCT-GAN adopts a CCT as the auxiliary classifier to provide the loss information of fault diagnosis to support the training of the generator. The generated samples are further fed into CCT auxiliary classifier to alleviate the data imbalance issue. The main contribution of this study is four-fold: (1) A new CCT-GAN compound fault diagnosis framework is proposed to address the data imbalance issue in compound fault diagnosis modelling; (2) In CCT-GAN, CCTs are designed as the discriminator and classifier, and a ULSAM-CNN is designed as a generator. The lightweight design of CCT and ULSAM-CNN greatly lowers the computational complexity to achieve better performance; (3) An experimental study based on the real-world robotic compound fault dataset is conducted to reveal the effectiveness of CCT-GAN. The rest of this paper is organized as follows. Section 2 reviews the latest research on compound fault diagnosis and the research of GAN in fault diagnosis. Section 3 introduces the problem statement and the proposed method. The experimental study is conducted in Section 4 and the results are demonstrated in Section 5. Finally, Section 6 discusses and Section 7 concludes.

2. Literature Review

2.1. The Research on Compound Fault Diagnosis

Researchers have mainly focused on the data-driven and the signal processing approach for compound fault diagnosis. The recent advance in deep learning has greatly improved the

compound fault diagnosis accuracy for rotating machinery. Rotate vector (RV) reducer is a core part of the mechanical transmission system. In order to achieve compound fault diagnosis for RV reducer, Xu et al. (2022) proposed a convolutional capsule network which composed of one-dimension convolutional neural networks (1D-CNN) and a capsule network. Firstly, 1D-CNN is used to extract the key features from the signal. Secondly, a two-layer stacked capsule network and a dynamic routing algorithm are deployed to identify the compound fault of the RV reducer. Huang et al. (2021) proposed a multi-source fidelity sparse representation method to achieve gearbox compound fault diagnosis. In this approach, the compound fault diagnosis can be enabled using the multi-source fidelity sparse representation method by establishing the generalized sparse representation model. Furthermore, the penalty function and the convexity conditions of the objective function were studied in this study to get better feature representation. Aiming to separate the fault-related features from the vibration data of gearbox, Li et al. (2019) investigated a multiple enhanced sparse decomposition approach to get the harmonic components and transient features from the vibration signals. Besides, a generalized minimax concave penalty was introduced to enhance the decomposition performance. (Liang et al., 2019) designed an integrated approach which adopts wavelet transform for time-frequency feature extraction and CNN for fault diagnosis modelling. Meanwhile, multi-label classification is introduced into CNN to achieve compound fault diagnosis.

Yang et al. (2022) proposed a deep capsule graph convolutional network for the compound fault diagnosis of harmonic drives. Firstly, the frequency spectrum is derived from the multi-sensor data, and then the label relationship map of the adjacency matrix is constructed.

Secondly, the fault-related features are represented using the deep capsule network, and a graph convolutional network is used for learning the relationship between different faulty types. Subsequently, the two networks are combined to get the diagnosis results, which are optimized by a dynamic routing algorithm. Attention mechanism has also been widely studied in compound fault diagnosis. Zhou et al. (2022) proposed an attention-enhanced multi-modal network for the compound fault diagnosis of industrial robots. A multi-head self-attention enhanced convolution neural network module and a long short-term memory network module were adopted to learn the fault-related patterns from different perspectives. Then the extracted features were combined and sent into a fully-connected layer for compound fault diagnosis. Lyu et al. (2022) presented a deep learning algorithm that can address the high-noise issue in bearing compound fault diagnosis modelling. In this algorithm, soft thresholding and global context were combined to extract the faulty features of compound fault in the noisy environment.

The signal process techniques also have been widely studied in compound fault diagnosis. Envelop spectrum analysis is an effective way of compound fault diagnosis. Alavi et al. (2022) proposed a targeted informative frequency band selection method to choose the most informative frequency band for envelope analysis. In this approach, negentropy is derived from lag information based on fractional lower-order autocorrelation and its logarithm-based derivation. This approach can be used to conduct the compound fault diagnosis when the noise is severe. In order to address the faulty feature is too weak in compound fault diagnosis modelling, Zhang et al. (2022) proposed a diagnosis method combining multi-symplectic geometry mode decomposition and multipoint optimal minimum entropy deconvolution

adjusted. Firstly, an improved symplectic geometry mode decomposition is used to separate fault components. Then, the weak fault features are enhanced by the multi-symplectic geometry mode decomposition. VMD is prevailing in compound fault diagnosis. The deployment of variational mode decomposition (VMD) requires the pre-set number of instinct mode functions (IMF), which is challenging to determine. In order to address this issue, Yi et al. (2022) proposed an IMF number estimation approach based on power spectral density so to optimize the performance of VMD.

2.2. The Research of GAN in Fault Diagnosis

In the operation of the machinery, fault happens occasionally, which leads to faulty data is hard to be observed and collected. When the normal operational data is abundant and the faulty data is limited, fault diagnosis modelling is challenging. Researchers have studied various methods to achieve few-shot learning in fault diagnosis. GAN is an effective method in addressing the data imbalance issue. GAN was proposed for the image generation task. The raw sensor signals can be transformed into time-frequency images via wavelet transformed or Fourier transform-based approach (Wang et al., 2018). Yang et al. (2021) investigated an approach that combines a conditional GAN (CGAN) and a 2D-CNN for bearing fault diagnosis when the data samples are limited. CGAN is deployed to generate fake samples to alleviate the insufficient training samples issue. Liu et al. (2022) utilized frequency slicing wavelet transform to generate time-frequency images. A multi-scale residual network structure and a hybrid loss function were designed to improve the performance of GAN. After the synthetic images were generated, the data was then sent into a feature enhancement CapsNet for fault diagnosis modelling. Li et al.

(2022) proposed an improved ACGAN for multi-mode data augmentation of the 2D time-frequency images. An independent classifier was introduced into ACGAN for the label classification of generated images. The Wasserstein distance was used as the loss function of ACGAN to overcome the model collapse issue.

Besides applying GAN for 2D time-frequency image generation, some existing studies also deployed GAN for sensory data generation. Peng et al. (2020) adopted deep autoencoders as the generator and discriminator of ACGAN and realized fault diagnosis based on the imbalanced data. In order to enhance the extraction of critical temporal information, Liu et al. (2021) proposed an LSTM-GAN-AE network, which that combines GAN and autoencoder based on the bidirectional long short-term memory network. Firstly, GAN was used to get the reconstruction residual and learn the discriminative representation. Secondly, an autoencoder was introduced to extract the critical temporal features, which were then fed into the LSTM network to train a fault diagnosis model. Zhou et al. (2020) proposed a global optimization GAN to address the data imbalance issue in the fault diagnosis modelling of bearings. In this network, an autoencoder was adopted as the generator, which can generate fake faulty samples to alleviate the few-shot issue of the faulty category. Two discriminators were designed in this network to identify the fake samples and classify the faulty types. In the training process, the information on fault features and diagnostic errors is introduced into the loss function to enhance the generation performance. Dixit et al. (2021) combined ACGAN with meta learning to achieve fault diagnosis for rotating machines with limited data. With the deployment of model agnostic meta learning, ACGAN can learn the task of data generation using limited

samples. The synthetic data was then used to train a fault diagnosis model via machine learning algorithms such as support vector machine and random forest. In this study, the models used for the generator and discriminator are fully-connected neural networks. Luo et al. (2021) used deep CNN as the generator and discriminator of condition-GAN for faulty data argumentation. After the data samples were generated, a CNN was used to train a fault diagnosis classifier. In the existing studies, most GAN network adopted CNN or autoencoder as the backbone of discriminator and generator due to the good trainability of these network.

2.3. A Brief Summary

Compound fault diagnosis is essential to the operation and maintenance of industrial robots. Most recent studies on compound fault diagnosis mainly focus on the component level such as the bearing and gearbox. Different from bearing and gearbox whose structure is relatively simple, industrial robot is a complex mechatronic system, which has multiple components. The tight coupling of the components in the industrial robot presents a great challenge in compound fault diagnosis. Deep learning is an algorithm that requires a large amount of data to achieve high performance. However, the compound fault data is hard to collect. When the compound data is limited, deep learning may not be able to establish an accurate model based on the imbalanced data. The existing studies have investigated the imbalanced data fault diagnosis modelling using GAN, while most of them deployed simple CNN or autoencoder as generator and discriminator. The trade-off between the model complexity and the algorithm performance of GAN has not been investigated in the existing studies. The simple network may not be able to fully explore the spatial-temporal feature, while the complex network may impede the

training stability of GAN. Hence, in this paper, we propose a CCT-GAN algorithm to leverage the quality of the generated samples, which is beneficial to tackle the data imbalance issue in compound fault diagnosis modelling.

3. Methodology

3.1. Problem Statement

In the actual manufacturing scenario, the uptime of the industrial robot is much longer than the downtime. Meanwhile, the probability of a single fault is higher than that of a composite fault, which leads to the challenge of collecting the compound fault samples. When limited compound fault data is available, the dataset becomes imbalanced and poses a great challenge for compound fault diagnosis modelling.

In order to alleviate the data imbalance issue in compound fault diagnosis modelling, the correlation between single fault and compound fault can be used. It can be seen from Figure 6 that there are multiple components in an industrial robot. Taking the incipient fault of the motor and reducer as examples. It can be seen from the CWT images generated from the feedback current signals that the faulty patterns of reducer 1 and motor 2 are periodic. When a compound fault of reducer 1 and motor 2 occurs, its faulty patterns become complex. The time-frequency image of the compound fault indicates that its faulty patterns are correlated to the single faults of reducer 1 and motor 2. Under this context, it is worthwhile to investigate how to make full use of the faulty correlation in the GAN and compound fault diagnosis modelling.

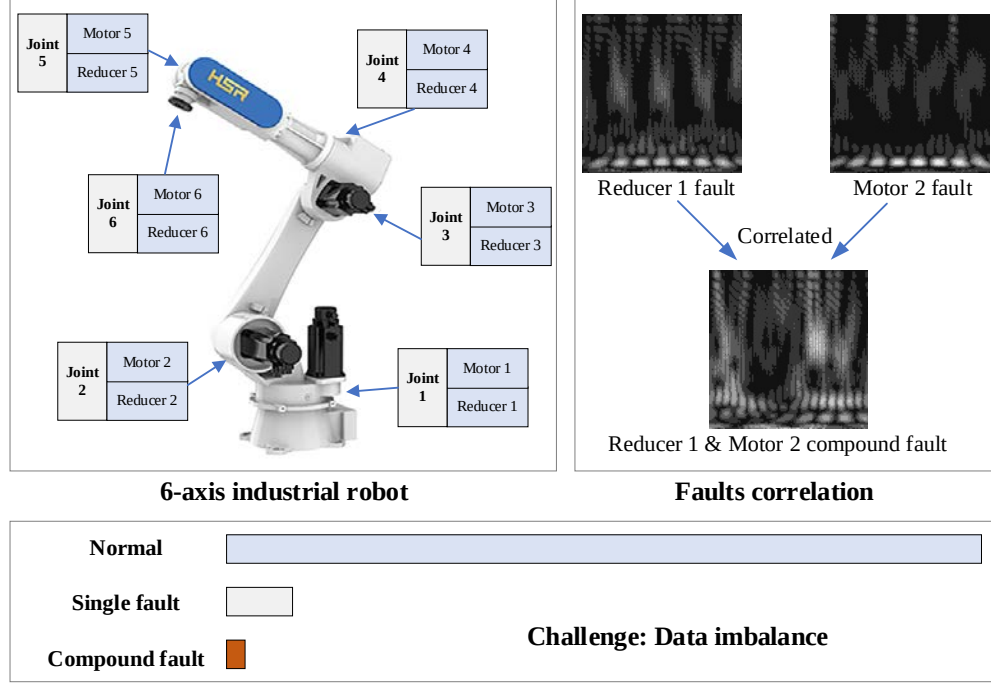


Fig. 1. The demonstration of the fault correlation and the data imbalance issue.

3.2. The Overall Flow of CCT-GAN-Based Compound Fault Diagnosis

The details of the proposed CCT-GAN-based compound fault diagnosis approach are elaborated in this section. As shown in Figure 2, the feedback current data of the industrial robot is collected from the servo driver. In comparison with the vibration signal that requires extra sensors to collect, the feedback current signal can be directly collected from the existing device. The faulty patterns can be reflected in the feedback current signals (Zhou et al., 2022). After the feedback current signals are collected, CWT is adopted to generate the time-frequency images, which can reveal the fault-related patterns. Then the CWT images are split into a training set and a testing set. The imbalanced training data is used to train the discriminator and classifier. The randomly generated noisy data z and the assigned label y' are used to generate the CWT images. In each training epoch, the generated images are different, which provides diversified samples to train the discriminator and classifier. The output of the discriminator is

the probability of being real or fake. The output of the classifier is the multi-hot representation of the compound fault diagnosis results. The loss of discriminator and classier are then sent back to the generator for the guidance of better generation. In the training of CCT-GAN, the gradient penalty and weight clipping are introduced to improve the training stability.

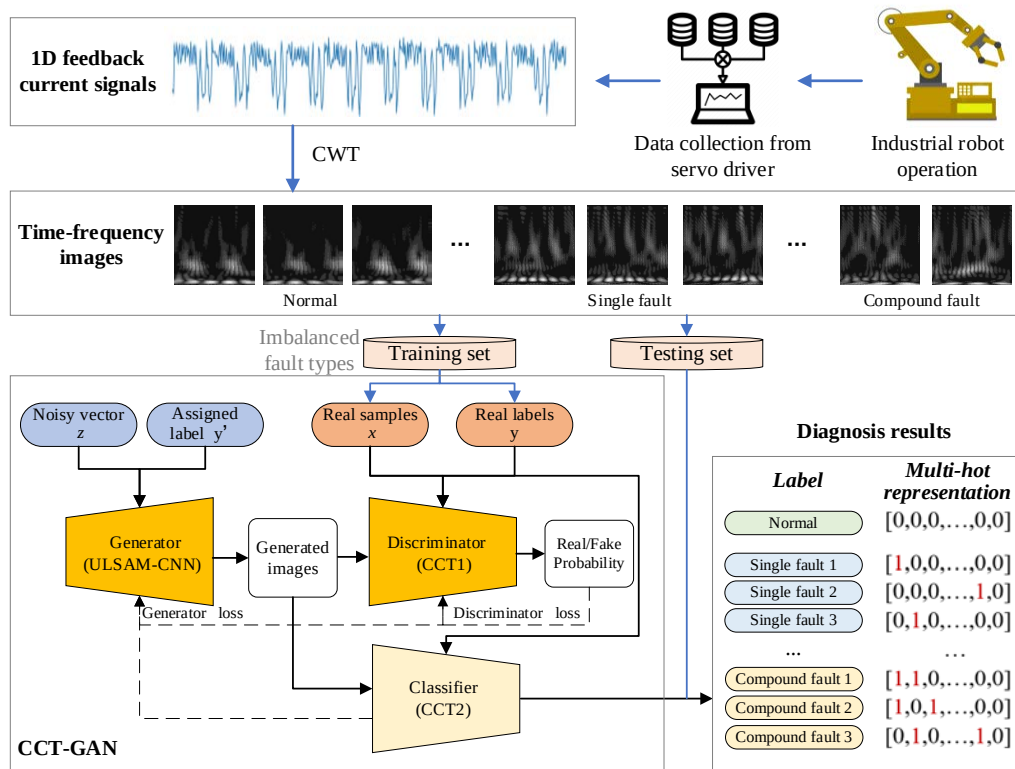


Fig. 2. The flowchart of the proposed compound fault diagnosis approach.

3.3. The Structure of the Proposed CCT-GAN

Computational efficiency is one of the key concerns in the network design of GAN. Traditional ACGAN and WGAN mainly deployed CNN as the backbone of the generator and discriminator. However, the vanilla CNN is short for the capability to learn sophisticated patterns within the data. Although the CNN with residual connection enables the generator and discriminator to learn the higher-level features, its cost-efficiency still needs to be improved. In

the proposed CCT-GAN, in order to achieve better generation performance without significantly increasing the model parameters, a lightweight attention mechanism named ULSAM (Saini et al., 2020) is introduced into the generator. The input of the generator is the concatenation of the randomly generated noise z and the embedding of the assigned label y' . The input is processed by deconvolutional layers for upsampling. The batch normalization and leaky Rectified Linear Unit are adopted in each deconvolutional layer to improve the training stability. The deconvolutional layers can effectively reconstruct the position and shape features of the time-frequency images. However, the local details of generated images can be blurred since the deconvolution operation may not be able to generate distinctly local patterns. In order to address this issue, ULSAM and a convolutional layer are introduced into the high level of the generator.

With the employment of ULSAM, individual attention maps can be learned for each feature subspace together with multiscale and multifrequency feature learning. The feature maps from the deconvolutional layer 4 are denoted as $F \in R^{m \times h \times w}$, where m is the number of input channels, h and w is the spatial dimensions of the feature maps. The focus of ULSAM is to efficiently obtain the cross-channel inter-dependencies in the feature maps. The details of the ULSAM-enhanced CNN are shown in Figure 3. ULSAM divides F into g groups $[F_1, F_2, \dots, F_g]$. Each group has G feature maps. The process of an intermediate feature maps F_i are shown as follows:

$$A_i = \text{softmax}\left(PW^1\left(MP^{3 \times 3, 1}\left(DW^{1 \times 1}(F_i)\right)\right)\right) \quad (1)$$

where $MP^{3 \times 3, 1}$ denotes the max pooling operation with the kernel size is 3×3 and the padding size is 1. $DW^{1 \times 1}$ denotes the depthwise convolution with the kernel size is 1×1 . PW^1 denotes the pointwise convolution with the kernel size is 1×1 . The attention map A_i in each group can learn the dependencies between different feature maps.

Then a gating mechanism with a SoftMax activation is used to refine the feature maps.

$$\hat{F}_i = (A_i \otimes F_i) \oplus F_i \quad (2)$$

where $\otimes$ denotes the element-wise multiplication and $\oplus$ is element-wise addition.

In the next stage, the refined feature maps are concatenated and yield the final output, which can be expressed as follows:

$$\hat{F} = \text{concat}([\hat{F}_1, \hat{F}_2, \dots, \hat{F}_i, \dots, \hat{F}_g]) \quad (3)$$

Finally, the feature maps are sent into a convolutional layer to get the generated image.

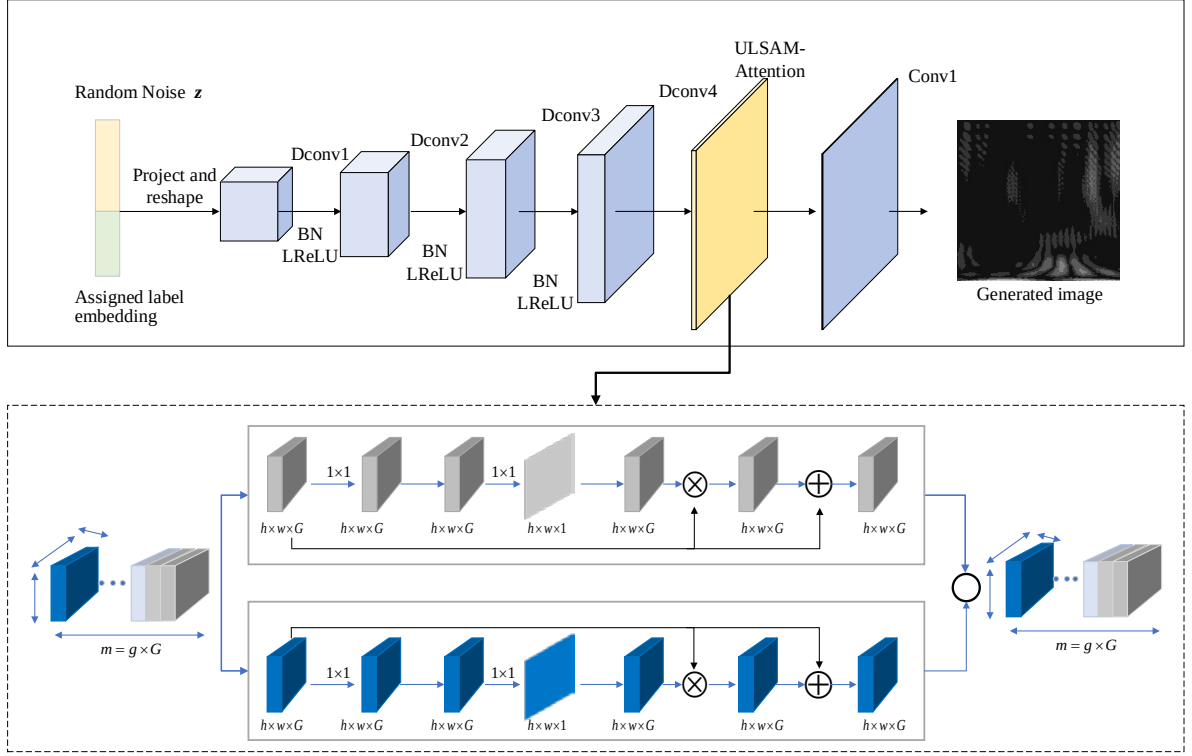


Fig. 3. The details of ULSAM-enhanced CNN.

Different from ACGAN using a network with two outputs to achieve discrimination and classification simultaneously, the discriminator and classifier in CCT-GAN are two independent networks. The networks for discrimination and classification do not focus on the same patterns. Using the same network to achieve both discrimination and classification tasks may lead to the unsatisfactory performance of both tasks since it has critical requirements for feature extraction. The adoption of two independent networks for discrimination and classification leads to better performance, while it increases the computational complexity. Hence, two CCT networks are adopted as the discriminator and classifier in this study. CCT is a type of lightweight vision transformer (ViT) network, which can efficiently achieve image classification with low computational cost.

In the lower level of the CCT, the convolution and max pooling layers are deployed to lower the size of feature maps. Subsequently, the feature maps are fed into the Transformers block for global pattern learning. The output of the Transformers block is then reduced by sequence pooling. Because of the deployment of convolutional blocks in the lower level and the sequence pooling in the higher level, CCT can achieve comparative performance to the large Transformers models with smaller network sizes and lower computational costs. ViT (Dosovitskiy et al., 2020) is well-known for its excellent image classification performance and high computational cost. Different from ViT which gets rid of convolutional operation, CCT adopts convolutional blocks for feature maps reduction. The feature maps with smaller sizes reduce the computational cost of the Transformers module, which leads to higher computational efficiency. The feature extraction by convolutional blocks can be expressed as:

$$F = \text{MaxPooling}(\text{ReLU}(\text{Conv2d}(x))) \quad (4)$$

where x is the input CWT images and F is the output of the convolution block.

In the next stage, the extracted features in conjunction with the positional encoding are sent into a Transformer block for sequential pattern learning. Multi-head self-attention can be deemed as the combination of multiple self-attention modules, which enables the network to focus on the patterns from different perspectives. By determining the relationship between a query and a set of key-value pairs to output, attention weights are obtained. The query vector q , key vector k and value vector v are used to get the attention score of the vanilla self-attention layer. When

an input i is fed into a self-attention attention module, the attention score can be expressed as follows:

$$Score = softmax(q_i * k_i) * v_i \quad (5)$$

In the multi-head self-attention layer, the matrices Q, K and V are used to replace the vectors q, k and v . With the combination of the multi-perspective information, the importance of various fault-related features can be obtained. The Q, K and V can be calculated by a linear transformation of the input data X :

$$Q = XW^q \quad (6)$$

$$K = XW^k \quad (7)$$

$$V = XW^v \quad (8)$$

, where W^q, W^k and W^v are trainable projection matrices.

In addition to the feature maps yielded by convolutional blocks, positional encoding provides sequential information of the feature maps to the Transformer block. The output of a Transformer block is shown as follows.

$$x_T = f(x_0) \in \mathbb{R}^{b \times n \times d} \quad (9)$$

where the dimension of x_T is $b \times n \times d$, b denotes the mini-batch size, n denotes the sequence length, d denotes the embedding dimension of the feature maps, x_0 denotes the output of the max pooling layer, and x_T denotes the output of the Transformers block. Then, x_T is then sent into a fully connected layer. The output is then transposed and activated by the SoftMax function to obtain the output of sequence pooling x_p , which is shown as follows.

$$x_p = SoftMax(Linear(x_T)^T) \times x_T \in \mathbb{R}^{b \times 1 \times d} \quad (10)$$

The 2nd dimension of x_T represents the sequence length. The importance of the sequence can be revealed by pooling the data in this dimension. After the output of sequence pooling is obtained. It is further used to get the compound fault diagnosis result or the discrimination result of the generated image.

As shown in Figure 4, the classifiers and discriminators share the same left and intermediate parts. However, the outputs of both networks are different. The output of the discriminator is the probability of the real image. Hence, a linear layer is adopted as the output layer of the discriminator. The output of the classifier is the compound fault diagnosis results. The compound fault refers to the concurrence of multiple single faults. If a type of compound fault is regarded as a new and independent faulty type, its association with the single faults may be omitted. Hence, we adopted multi-hot encoding to represent the label of the data. The association between compound fault and single faults can be helpful to alleviate the data imbalance issue in compound fault diagnosis modelling. Hence, the multi-hot label is used to represent the fault classes. Each digit in a multi-hot label stands for a component in an industrial robot. The adoption of the multi-hot label also can partially reflect the topology of the components. In order to get the multi-hot output, the fully connected layers with the sigmoid activation are used as the final layer of the classifier. The number of nodes in the last layer represents the number of components considered in the industrial robot.

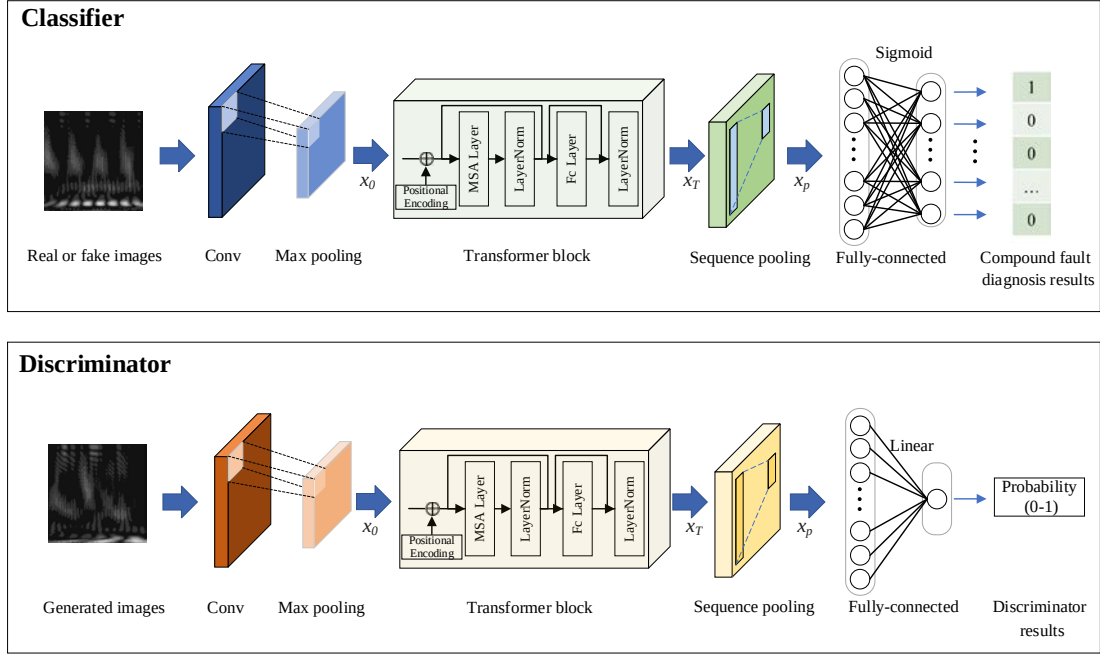


Fig. 4. The details of the classifier and discriminator.

3.4. The design of the loss function

The proposed CCT-GAN takes advantage of ACGAN and WGAN. ACGAN aims to achieve discrimination and classification at that same time. The loss for discrimination L_D and classification L_{Class} can be expressed as:

$$L_D = E_{xP_{r(x)}} [\log D(x)] + E_{P_{z(z)}} \left[\log \left(1 - D \left(G(z, c_g) \right) \right) \right] \quad (11)$$

$$L_{Class} = E_{xP_{r(x)}} [-\log P(c = c_r | x)] + E_{P_{z(z)}} [-\log P(c = c_g | G(z, c_g))] \quad (12)$$

where $P_{r(x)}$ is the distribution of real data, $P_{z(z)}$ is a prior distribution of the noise input, $D(x)$ is the probability that x is selected in the real dataset, $E_{xP_{r(x)}}$ is the expectation of x from $P_{r(x)}$, $E_{P_{z(z)}}$ is the expectation of z from the noise input, $G(z, c_g)$ denotes the generated image from the generator G , c_r and c_g denotes the labels of x and the generated images, $P(c = c_r | x)$ and $P(c = c_g | G(z, c_g))$ denotes the conditional probability distribution of the class labels of the real data and the generated data.

In the adversarial training stage, the generator in ACGAN aims to achieve the minimization of $L_{\text{Class}} - L_{\text{validity}}$, while the discriminator aims to achieve the maximization of $L_{\text{Class}} + L_{\text{validity}}$. ACGAN adopts the JS divergence to measure the distance of the distribution between the synthetic data and real data. The JS divergence can lead to the model collapse in the training process due to its discrete characteristics. In order to address this issue, WGAN (Arjovsky et al., 2017) replaces JS divergence with the Wasserstein distance. It is smooth and continuous, which is helpful for the training of GAN. The Wasserstein distance can be expressed as:

$$W(p_{r(x)}, p_{g(y)}) = \inf_{\gamma \sim \Pi(p_{r(x)}, p_{g(y)})} E_{(x,y) \sim \gamma} [\|x - y\|] \quad (13)$$

where $p_{g(y)}$ is the distribution of generated image y , $\inf_{\gamma \sim \Pi(p_{r(x)}, p_{g(y)})}$ represents the infimum of the joint probability distribution of the real data and generated data, $E_{(x,y) \sim \gamma} [\|x - y\|]$ represents the expected distance between the real data x and the generated data y .

Then the Wasserstein distance is approximated by the Lipschitz continuity condition, which can be expressed by:

$$WD = \min_G \max_{D \in R} V(G, D) = E_{x \sim p_{r(x)}} [D(x)] - E_{y \sim p_{g(y)}} [D(y)] \quad (14)$$

where $D \in R$ represents that the discriminator D should satisfy the Lipschitz continuity condition.

In CCT-GAN, the loss function is redesigned which takes advantage of Wasserstein distance and the loss function in ACGAN. The loss of classifier combines the loss from the real data and the label and the loss from the generated data and the assigned label, which can be expressed as:

$$Loss_C = L_C^R + L_C^G = E_{xP_{r(x)}}[P(c = c_r | x)] + E_{P_{z(z)}}[P(c = c_g | G(z, c_g))] \quad (15)$$

The loss of the discriminator is the same as that in WGAN. The loss of the generator combines the loss of the generator in WGAN and the classification loss of the generated images. The loss of discriminator and generator in CCT can be expressed as:

$$Loss_D = E_{xP_{r(x)}}[D(x)] - E_{yP_{g(y)}}[D(y)] \quad (16)$$

$$Loss_G = L_D^G + L_C^G = E_{P_{z(z)}}(D(G(z, c_g))) + E_{P_{z(z)}}[P(c = c_g | G(z, c_g))] \quad (17)$$

In order to address the weight constraints issue raised in WGAN and further improve the training stability, the gradient penalty (Gulrajani et al., 2017) is introduced into $Loss_D$.

The gradient penalty constrains the gradient norm of the output of the discriminator with respect to its input. The constraint with a penalty is enforced on the gradient norm for random samples $\hat{x} \sim P_{\hat{x}}$. The gradient penalty term can be expressed as:

$$GP = \lambda_{\hat{x} \sim P_{\hat{x}}}^E [\|\nabla \hat{x} D(\hat{x})\|_2 - 1]^2 \quad (18)$$

where λ is the penalty coefficient. The modified $Loss_D$ with gradient penalty can be expressed as:

$$Loss_D = E_{xP_{r(x)}}[D(x)] - E_{yP_{g(y)}}[D(y)] + \lambda_{\hat{x} \sim P_{\hat{x}}}^E [\|\nabla \hat{x} D(\hat{x})\|_2 - 1]^2 \quad (19)$$

4. Experimental Setup

4.1. Data Collection

In order to evaluate the effectiveness of the proposed compound fault diagnosis approach, a compound fault injection experiment was conducted. The feedback current of the third axis was collected from the motor drivers. The faulty reducers and motors collected from the actual

manufacturing scenarios were used to replace the normal components in the experimental industrial robot. The faulty parts suffer from incipient faults such as minor abnormal sounds and oil leakage. After the faulty parts are mounted in the experimental robot and operated, the feedback current signals were collected from the driver. The three-phase current signals were collected by the Hall element in the driver and the signals used for modelling are the currents of the q-axis, which frequency is 1Hz.

The experiments involved three types of single faults and three types of compound fault. The details of the compound fault data are shown in Table 1. The wave from the frequency domain was obtained via fast Fourier transform. In order to obtain the time-frequency images, the data is segmented by the length of 100. Then CWT is adopted to transform the feedback current signals in each segment to a 32*32 time-frequency image. The data volume of the CWT images can be found in Table 1.

Table 1. The details of the compound fault dataset.

Fault type	No.	Faulty parts	Multi-hot label	Data volume (CWT images)
	0	Normal	[0,0,0,0,0,0,0]	3,000
Compound fault	1	Reducer 1 and Motor 2	[1,0,0,0,0,0,1]	1,888
	2	Reducer 1 and Reducer 3	[1,0,1,0,0,0,0]	2,046
	3	Reducer 3 and Reducer 4	[0,0,1,1,0,0,0]	2,131
Single fault	4	Reducer 3	[0,0,1,0,0,0,0]	3,000
	5	Motor 2	[0,0,0,0,0,0,1]	3,000
	6	Reducer 4	[0,0,0,1,0,0,0]	3,000

The time domain and frequency domain of different faulty categories are illustrated in Figure 5. In the time-domain, the raw feedback current signals were inspected for each fault category. The ‘Normal’ condition exhibited a consistent and smooth waveform, with minimal

fluctuations, indicative of the robot's healthy operational state. However, as we delved into the faulty conditions, such as the 'Reducer 3 fault' or 'Motor 2 fault', the signals revealed distinct anomalies. These included oscillations, spikes, and irregularities that were absent in the normal condition. The temporal features, such as the duration and frequency of these anomalies, provided valuable insights into the nature and severity of the faults, highlighting the need for a nuanced interpretation of the data. Transitioning to the frequency domain via Fast Fourier Transform (FFT), the spectral content of the signals painted a clearer picture of the fault characteristics. The 'Normal' condition's spectrum predominantly featured a strong peak, with minimal harmonic content. In contrast, the faulty conditions introduced a spectral complexity that was both revealing and diagnostic. For instance, the 'Reducer 3 fault' condition displayed sidebands around the low frequency, suggesting mechanical disturbances at specific harmonics. The 'Motor 2 fault', on the other hand, showed an increase in high-frequency noise, which could be indicative of electrical issues within the motor. CWT images for the different fault categories revealed the temporal-frequency information that the time and frequency domain analyses alone could not capture. The CWT analysis thus provided a visual and intuitive understanding of the fault dynamics, enhancing the diagnostic capability of the study. The CWT images displayed the signal's energy distribution over time and frequency, with each fault category exhibiting unique patterns. In the 'Normal' condition, the CWT image showed a well-localized and consistent energy signature at the low frequency, indicative of the robot's steady operation. However, in the 'Reducer 3 fault' category, the CWT revealed a series of scattered energy concentrations, suggesting periodic disturbances at specific times that correlated with the fault's mechanical nature. For the 'Motor 2 fault', the image showed a more diffuse spread

of energy across higher frequencies, indicating a broader spectrum of electrical disturbances.

The compound fault categories, such as ‘Reducer 1 and Motor 2 fault’, presented a complex

interplay of these patterns, with multiple energy concentrations at various frequencies and times,

reflecting the compound nature of the faults.

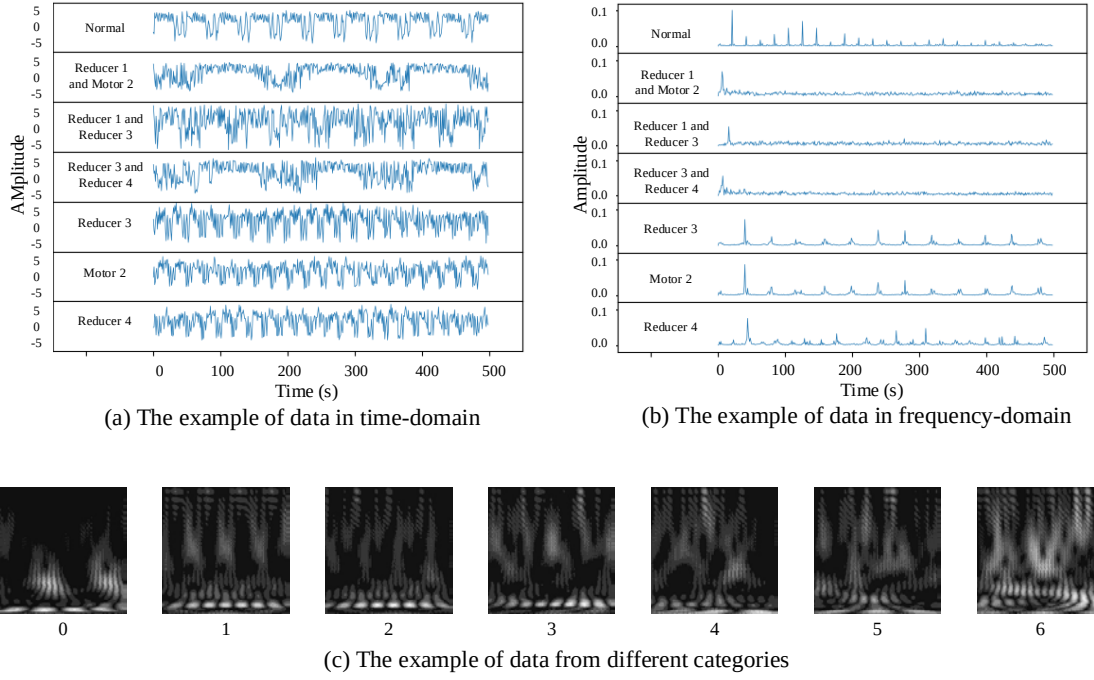


Fig. 5. The illustration of signals in the time domain and frequency domain.

4.2. Benchmarking Experiments

In this study, the performance of CCT-GAN was evaluated in two experiments. The imbalanced dataset adopts all the data from normal and single faults categories. We evaluated the performance of different algorithms using datasets with different ratios of compound fault samples. The CCT developed in this study adopts two convolutional blocks for feature extraction. The convolutional kernel size was set as 3×3 . Two Transformer blocks were adopted for the sequential patterns learning. The number of nodes in the Transformer block was set as 100, and the number of heads in the multi-head self-attention layer was set as six. Adam was adopted as the optimizer and the learning rate was set as 0.001. The Binary Cross Entropy was

used as the loss function. the early stopping strategy was adopted in the training stage. The parameters for the proposed CCT-GAN were determined by conducting a series of trials to find a trade-off between computational efficiency and performance.

In the first experiment, the performance and computational costs of CCT-GAN were evaluated. The prevailing GAN networks applied in fault diagnosis, which are ACGAN and ACWGAN, were used as benchmarking algorithms. Meanwhile, the impact of the prevailing generator and discriminator structure such as vanilla CNN and ResNet on the algorithm performance was also revealed in this experiment. Model size and FLOPs (floating-point operations per second) were adopted to reveal the computational efficiency. Besides, the effect of the gradient penalty was also revealed.

The setup of the second experiment is detailed in Table 2. The setting of CCT adopted in the CCT-GAN is the same as the first experiment. The batch size was set as 128. Adam was selected as the optimizer of the generator, discriminator, and classifier. The learning rate of the optimizer for the generator and discriminator was set as 0.0001, and the learning rate of the optimizer for the classifier was set as 0.001. The early stopping strategy was adopted and the maximum training epoch was set as 1500. In the second experiment, the performance of CCT-GAN in addressing the data imbalanced issue was explored. CCT-GAN was used for compound fault diagnosis modelling based on the different number of instances from the compound fault categories.

Table 2. The details of the setup of the experiment for GAN-based compound fault diagnosis.

Algorithms	Classifier and Discriminator			Generator			Gradient penalty
	CNN	ResNet	CCT	CNN	ResNet	ULSAM-CNN	
CCT-GAN			√			√	√
ACWGAN- ResNet-GP		√			√		√
ACWGAN-GP	√			√			√
ACWGAN	√			√			
ACWGAN-ResNet		√			√		
ACWGAN-ULSAM	√					√	
ACWGAN-ULSAM-CCT			√			√	

In order to fully evaluate the algorithm performance of the proposed CCT-GAN from different aspects, the overall accuracy of the testing set and the accuracy for the three compound fault categories in the testing set were adopted. Precision, recall and F1 values were also used to evaluate the performance of CCT-GAN in imbalance data modelling. Mixup, as a prevailing data augmentation approach, was introduced to evaluate the performance of CCT-GAN. The number of Mixup samples are the same as the data in the original category. Meanwhile, the model parameters were marked and compared to reveal the computational cost of different algorithms. The 5-fold cross-validation was conducted to ensure the results are comprehensive. All the experiments were conducted on an Intel i9-10920X 3.50Ghz CPU with Nvidia GeForce RTX 3090 graphics card.

5. Experimental Results

5.1. The Comparison Between Different GAN-Based Compound Fault Diagnosis

Algorithms

In order to reveal the algorithm performance of CCT-GAN and other prevailing GAN algorithms, the full-size dataset was used for compound fault modelling. The compound fault modelling results and the computational costs are demonstrated in Table 3. CCT-GAN achieves the highest overall accuracy and the highest accuracy for compound fault categories, which are 86.79% and 71.23%. Without the gradient penalty, the algorithm performance ACWGAN-ULSAM-CCT network is slightly lower than that of CCT-GAN. This experiment shows ACWGAN has the lowest overall accuracy and accuracy for compound fault categories, which are 68.49% and 47.83% respectively. In order to reveal the computational efficiency of the CCT-GAN, the model size and FLOPs of different GAN-based algorithms were compared. CCT-GAN and ACWGAN-ULSAM-CCT get the smallest model size, which is 6.8M. Meanwhile, the FLOPs of CCT-GAN are 8144.5M, which is slightly higher than that ACWGAN and ACWGAN-GP.

Table 3. The comparison of the algorithm performance and the computational costs.

	CCT-GAN	ACWGAN- ResNet-GP	ACWGAN	ACWGAN- GP	ACWGAN- ResNet	ACWGAN- ULSAM	ACWGAN- ULSAM- CCT
Overall							
Accuracy (%)	86.79±0.48	83.87±0.61	68.49±0.46	70.94±1.14	83.57±0.33	69.9±1.43	85.4±0.82
Accuracy for compound fault categories (%)	71.23±1.11	67.12±1.52	47.83±1.34	52.69±1.11	64.84±0.82	48.5±0.61	69.73±1.68
Model size (M)	6.8	24.3	17.2	17.2	24.3	17.8	6.8
FLOPs (M)	8144.5	14571.9	8131.7	8131.7	14571.9	8141.7	8144.5

5.2. The Modelling Performance of CCT-GAN Based on Imbalanced Data

To further evaluate the performance of CCT-GAN under the data imbalanced issue, the testing accuracy during the training stage based on the different numbers of compound fault data was marked and compared. Figure 6 illustrates the comparison of the overall accuracy of CCT based on different compound fault data volumes. It is obvious that all the curves tend to converge when the training epoch arrives at 1000. When there are only 20 data samples from each compound fault category are available, CCT-GAN can achieve approximately 76% overall accuracy. The best overall accuracy achieved by CCT-GAN is approximately 87% when full-size data is available.

The accuracy for compound fault categories of CCT-GAN based on different compound fault data volumes is demonstrated in Figure 7. CCT-GAN gets the accuracy for compound fault categories which is approximately 45% when 20 data samples from each compound fault category are available. CCT-GAN's performance in terms of accuracy for compound fault categories increases rapidly as the number of data samples increases. When 300 data samples are available, the accuracy for compound fault categories achieved by CCT-GAN reaches 55%. When the full-size compound fault categories data is adopted, CCT-GAN yields the best performance in this experiment, which is approximately 72%.

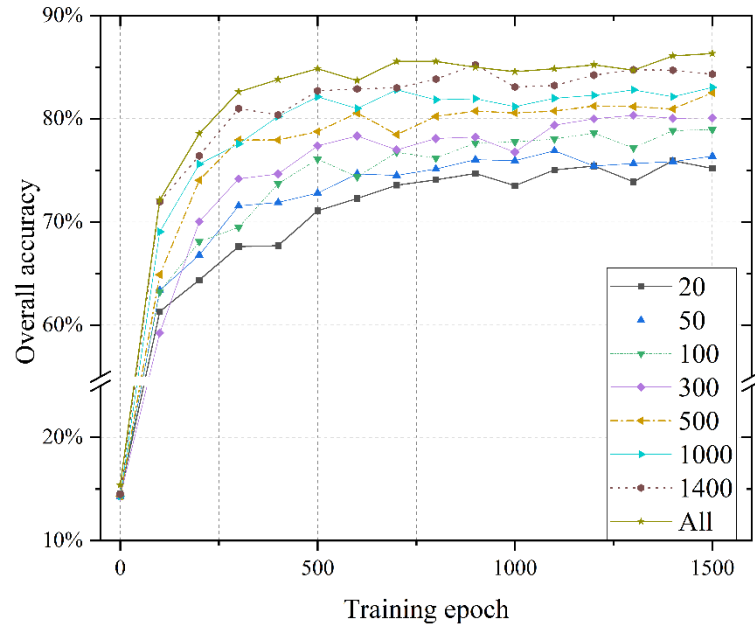


Fig. 6. The overall accuracy of CCT based on different compound fault data volumes.

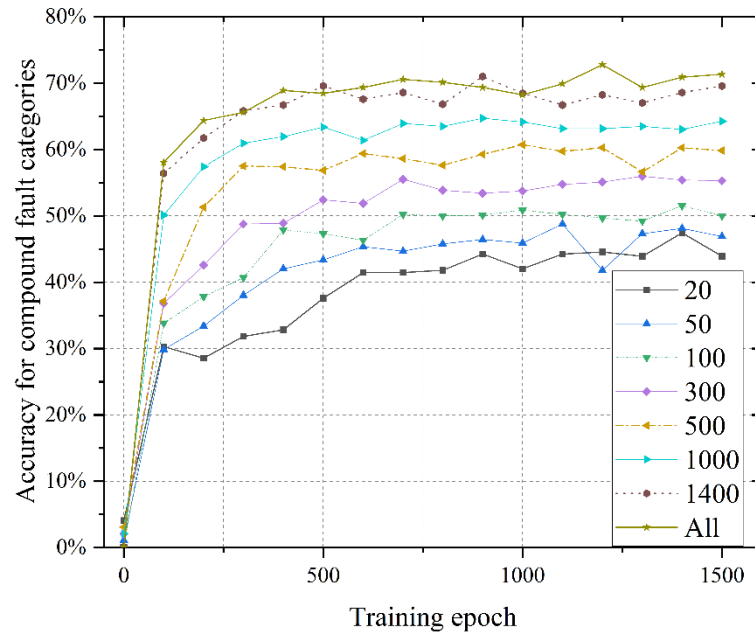


Fig. 7. The accuracy for compound fault categories of CCT-GAN based on different compound fault data volumes.

The algorithm performance of CCT-GAN, CCT, and the CCT enhanced by Mixup based on the imbalanced data are illustrated in Figures 8 and 9. It is obvious that the overall accuracy of CCT-GAN always higher than that of CCT in Figure 8. The margin of overall accuracy between

the two algorithms enlarged with the increase of the instance number from the compound fault categories. When the full-size data is available, the overall accuracy of CCT-GAN is 4.45% higher than that of CCT. The results in Figure 8 indicate that when the instance number from the compound fault categories is lower than 50, CCT-GAN can slightly improve the accuracy for compound fault categories which is less than 4%. After that, the difference between CCT-GAN and CCT in terms of the accuracy for compound fault categories is enlarged. The improvement is up to 7.6% when 500 instances from each compound fault category are adopted. When the full-size data is adopted, the accuracy for compound fault categories obtained from CCT-GAN reaches 71.23%. With the enhancement of Mixup, the performance of CCT in terms of the overall accuracy and the accuracy for compound fault categories 1-3%.

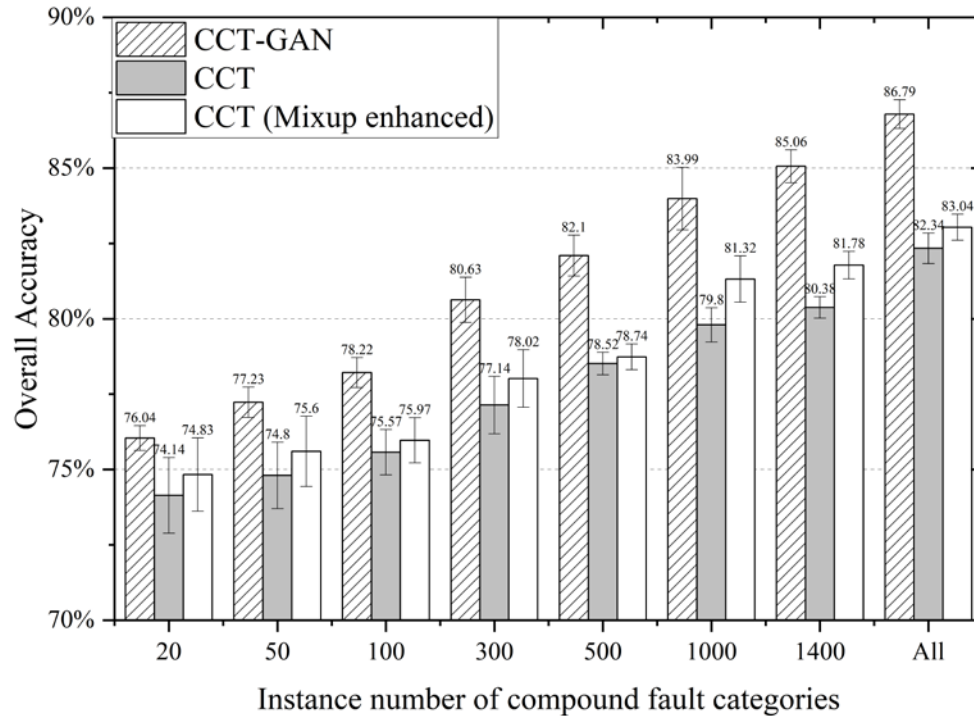


Fig. 8. The overall accuracy of CCT and CCT-GAN based on different compound fault data

volumes.

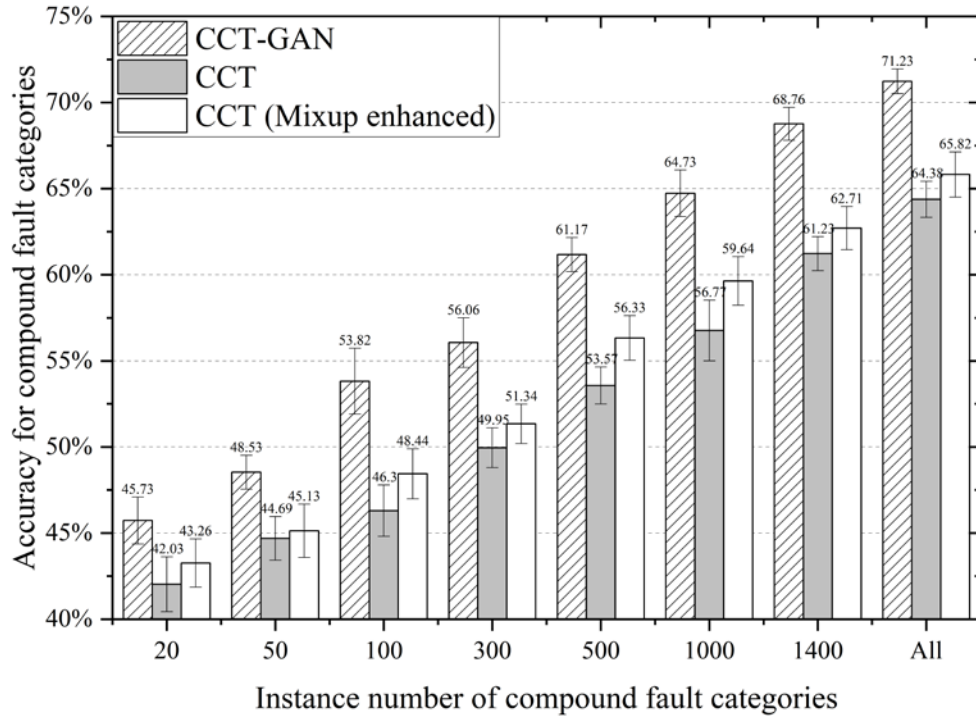


Fig. 9. The accuracy for compound fault categories of CCT and CCT-GAN based on different compound fault data volumes.

The experimental results of the CCT-GAN algorithm across various data volumes in the compound fault category are shown in Table 4. The Precision, Recall and F1 values of CCT-GAN show dramatic advantage in comparison with CCT and CCT (Mixup enhanced). As the data volume increases from 20 to 1400, the precision, recall, F1 value, and overall accuracy of the algorithm show a gradual improvement. Specifically, the precision starts at 0.7498 and rises to 0.8424, indicating a better ability to correctly identify true positives. Similarly, recall, which measures the algorithm's ability to detect all relevant instances, starts at 0.7297 and increases to 0.8334, suggesting an enhanced detection capability with larger datasets. The F1 value also shows a steady increase from 0.7396 to 0.8378, reflecting a balanced performance improvement.

Table 4. The Precision, Recall and F1 values of different algorithms with different data volumes in compound fault category.

	Data Volume in compound fault category	20	30	50	100	300	500	1000	1400	All
CCT-GAN	Precision	0.7498	0.7468	0.7747	0.7738	0.7793	0.7811	0.7986	0.8168	0.8424
	Recall	0.7297	0.7278	0.7533	0.7642	0.7841	0.7845	0.7904	0.8132	0.8334
	F1 value	0.7396	0.7371	0.7638	0.7689	0.7816	0.7827	0.7944	0.8149	0.8378
CCT	Precision	0.6921	0.7048	0.7235	0.7580	0.7525	0.7596	0.7761	0.7961	0.8135
	Recall	0.6888	0.7051	0.7341	0.7529	0.7596	0.7325	0.7693	0.8151	0.8212
	F1 value	0.6904	0.7049	0.7287	0.7554	0.7560	0.7458	0.7726	0.8054	0.8173
CCT (Mixup enhanced)	Precision	0.7171	0.7254	0.7364	0.7324	0.7567	0.7647	0.7803	0.7968	0.8213
	Recall	0.7142	0.7137	0.7221	0.7364	0.7625	0.7541	0.7755	0.7954	0.8246
	F1 value	0.7156	0.7195	0.7291	0.7349	0.7595	0.7593	0.7778	0.7960	0.8229

6. Discussion

An efficient network structure can bring tangible benefits to the performance of GAN. It is difficult for GAN to reach Nash Equilibrium when the model size is too large. However, when the model size is small, the network may not be able to learn the latent patterns within the data. Hence, the computational efficiency of GAN should be investigated. In this study, a cost-efficient network named CCT-GAN is proposed to address the data imbalance issue in compound fault diagnosis modelling.

The results of the first experiment indicate that CCT-GAN can achieve better performance and computational efficiency in comparison with the prevailing GAN-based algorithms. ACWGAN and ACWGAN-GP adopt vanilla CNN as the generator, discriminator, and classifier, which achieves the worse algorithm performance in comparison with other algorithms. Although ACWGAN- ResNet-GP and ACWGAN- ResNet achieve an overall accuracy over 83% and the

accuracy for compound fault categories over 67%, their model size is much larger than that of the other algorithms. In comparison with other algorithms, CCT-GAN achieved the best algorithm performance and the smallest model size. The FLOPs of CCT-GAN are slightly higher than that of the ACWGAN and ACWGAN-GP, which adopted simple CNN as the generator, discriminator and classifier. However, the FLOPs of CCT-GAN are obviously less than that of ACWGAN- ResNet-GP and ACWGAN-ResNet. Both model size and FLOPs reveal that CCT-GAN has satisfactory computational efficiency. With the deployment of ULSAM-CNN, both accuracies of ACWGAN-ULSAM are higher than that of ACWGAN, which are 1.41% and 0.67%, respectively. The lightweight attention does not obviously increase the model complexity while it can promote the algorithm performance of the network. Besides, the gradient penalty strategy deployed in CCT-GAN can promote both accuracies by 1.5%. The vanilla CNN and ResNet are used as the benchmarking algorithms in the GAN-based model because both networks have shown excellent performance in the GAN-based fault diagnosis modelling (Li et al., 2022; Yang et al., 2021). Both vanilla CNN and ResNet are easier to be trained and achieve Nash Equilibrium in the GAN network. Meanwhile, deploying the complex networks into GAN also requires a large computational load, which may not be affordable and not suitable for the actual deployment.

When the imbalanced data was used for modelling, the performance of CCT-GAN is jeopardized when only limited data samples are available. However, in comparison with CCT, the performance of CCT-GAN in terms of both accuracies are better. Especially for the accuracy of compound fault categories, the improvement is obvious when over 100 data

samples from the compound fault categories are available. CCT-GAN can promote both overall accuracy and the accuracy of compound fault categories by 2.65% and 7.52% when only 100 data samples from each compound fault category are available. It also can achieve Precision, Recall and F1 values of over 0.70, which is better than the benchmarking algorithms. The reason is that the generator in CCT-GAN is effectively trained and generates diversified compound fault samples. The generated samples alleviate the data imbalance issue in the classifier. However, when the data samples from the compound fault categories are extremely scarce (only 50 samples or less), the generator may not be able to generate the compound fault samples effectively.

The data imbalance issue impedes algorithms to get an accurate classification. Therefore, our future works will focus on compound fault features extraction and meta learning for compound fault diagnosis. Firstly, the multi-scale feature extraction module that can effectively locate and extract the compound fault-related features will be studied and introduced into CCT to further leverage the algorithm performance. Secondly, meta learning is a type of algorithm that can achieve few-shot learning based on very limited samples. The current version of CCT-GAN is struggled in compound fault diagnosis based on insufficient samples. Hence, the combination of meta learning and CCT-GAN will be further explored to achieve satisfactory compound fault diagnosis accuracy when only limited data is available.

7. Conclusions

The alleviation of the data imbalance issue can be beneficial to the compound fault diagnosis of industrial robots. In order to address the data imbalance issue and the low computational efficiency in the GAN-based compound fault diagnosis modelling, a lightweight GAN called CCT-GAN was proposed in this study. An experimental study based on the real-world industrial robot compound fault dataset was conducted, which indicates that the proposed CCT-GAN can effectively improve the compound fault diagnosis accuracy. The key finding of this research are: (1) CCT-GAN can effectively address the data imbalance issue in compound fault diagnosis modelling; (2) The designed CCT-GAN can achieve better computational efficiency in comparison with the prevailing GAN-based algorithms; (3) When the samples from each compound fault category are over 100, CCT-GAN can effectively improve the compound fault accuracy, which can alleviate the data imbalance issue in compound fault diagnosis modelling. In the actual manufacturing scenarios, the proposed CCT-GAN can be helpful for engineers to lower the maintenance cost of industrial robots.

Acknowledgement

Our work is supported by multiple funds in China, including the Key Program of NSFC-Guangdong Joint Funds (U2001201), the National key R & D project (2021YFB3301802), the Natural Science Foundation of Guangdong Province (2020B1515120010). Our work is also supported by Guangdong Provincial Key Laboratory of Cyber-Physical System (2020B1212060069).

References

- Alavi, H., Ohadi, A., Niaki, S.T., 2022. A novel targeted method of informative frequency band selection based on lagged information for diagnosis of gearbox single and compound faults. *Mechanical Systems and Signal Processing* 170, 108828.
- Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks, *International conference on machine learning*. PMLR, pp. 214-223.
- Chen, C., Liu, C., Wang, T., Zhang, A., Wu, W., Cheng, L., 2023a. Compound fault diagnosis for industrial robots based on dual-transformer networks. *Journal of Manufacturing Systems* 66, 163-178.
- Chen, C., Wang, T., Liu, C., Liu, Y., Cheng, L., 2023b. Lightweight Convolutional Transformers Enhanced Meta Learning for Compound Fault Diagnosis of Industrial Robot. *IEEE Transactions on Instrumentation and Measurement*.
- Chen, D., Lv, P., Xue, L., Xing, H., Lu, L., Kong, D., 2024. Positional error compensation for aviation drilling robot based on Bayesian linear regression. *Engineering Applications of Artificial Intelligence* 127, 107263.
- Chen, X., Yang, R., Xue, Y., Huang, M., Ferrero, R., Wang, Z., 2023c. Deep transfer learning for bearing fault diagnosis: A systematic review since 2016. *IEEE Transactions on Instrumentation and Measurement* 72, 1-21.
- Dixit, S., Verma, N.K., Ghosh, A., 2021. Intelligent Fault Diagnosis of Rotary Machines: Conditional Auxiliary Classifier GAN Coupled With Meta Learning Using Limited Data. *IEEE Transactions on Instrumentation and Measurement* 70, 1-11.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Gleeson, D., Jakobsson, S., Salman, R., Ekstedt, F., Sandgren, N., Edelvik, F., Carlson, J.S., Lennartson, B., 2022. Generating Optimized Trajectories for Robotic Spray Painting. *IEEE Transactions on Automation Science and Engineering* 19, 1380-1391.
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C., 2017. Improved training of wasserstein gans. *Advances in neural information processing systems* 30.
- Hassani, A., Walton, S., Shah, N., Abuduweili, A., Li, J., Shi, H., 2021. Escaping the big data paradigm with compact transformers. *arXiv preprint arXiv:2104.05704*.
- Huang, W., Song, Z., Zhang, C., Wang, J., Shi, J., Jiang, X., Zhu, Z., 2021. Multi-source fidelity sparse representation via convex optimization for gearbox compound fault diagnosis. *Journal of Sound and vibration* 496, 115879.
- Jin, C.-c., Chen, X., 2021. An end-to-end framework combining time–frequency expert knowledge and modified transformer networks for vibration signal classification. *Expert Systems with Applications* 171, 114570.
- Li, N., Huang, W., Guo, W., Gao, G., Zhu, Z., 2019. Multiple enhanced sparse decomposition for gearbox compound fault diagnosis. *IEEE Transactions on Instrumentation and Measurement* 69, 770-781.
- Li, W., Zhong, X., Shao, H., Cai, B., Yang, X., 2022. Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework. *Advanced Engineering Informatics* 52, 1-16.

Liang, P., Deng, C., Wu, J., Yang, Z., Zhu, J., Zhang, Z., 2019. Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform. *Computers in Industry* 113, 103132.

Liu, H., Zhao, H., Wang, J., Yuan, S., Feng, W., 2021. LSTM-GAN-AE: A Promising Approach for Fault Diagnosis in Machine Health Monitoring. *IEEE Transactions on Instrumentation and Measurement* 71, 1-13.

Liu, J., Zhang, C., Jiang, X., 2022. Imbalanced fault diagnosis of rolling bearing using improved MsR-GAN and feature enhancement-driven CapsNet. *Mechanical Systems and Signal Processing* 168, 108664.

Luo, J., Huang, J., Li, H., 2021. A case study of conditional deep convolutional generative adversarial networks in machine fault diagnosis. *Journal of Intelligent Manufacturing* 32, 407-425.

Lyu, P., Zhang, K., Yu, W., Wang, B., Liu, C., 2022. A novel RSG-based intelligent bearing fault diagnosis method for motors in high-noise industrial environment. *Advanced Engineering Informatics* 52, 101564.

Peng, P., Wang, Y., Zhang, W., Zhang, Y., Zhang, H., 2020. Imbalanced process fault diagnosis using enhanced auxiliary classifier gan, 2020 Chinese Automation Congress (CAC). IEEE, pp. 313-316.

Ren, Z., Gao, D., Zhu, Y., Ni, Q., Yan, K., Hong, J., 2023. Generative adversarial networks driven by multi-domain information for improving the quality of generated samples in fault diagnosis. *Engineering Applications of Artificial Intelligence* 124, 106542.

Saini, R., Jha, N.K., Das, B., Mittal, S., Mohan, C.K., 2020. Ulsam: Ultra-lightweight subspace attention module for compact convolutional neural networks, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1627-1636.

Shao, H., Zhou, X., Lin, J., Liu, B., 2024. Few-shot cross-domain fault diagnosis of bearing driven by task-supervised ANIL. *IEEE Internet of Things Journal*.

Wang, C., Gan, M., Zhu, C.a., 2018. Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory. *Journal of Intelligent Manufacturing* 29, 937-951.

Wang, J., Shao, H., Yan, S., Liu, B., 2023a. C-ECAFormer: A new lightweight fault diagnosis framework towards heavy noise and small samples. *Engineering Applications of Artificial Intelligence* 126, 107031.

Wang, X., Jiang, H., Liu, Y., Liu, S., Yang, Q., 2023b. A dynamic spectrum loss generative adversarial network for intelligent fault diagnosis with imbalanced data. *Engineering Applications of Artificial Intelligence* 126, 106872.

Xu, Q., Liu, C., Yang, E., Wang, M., 2022. An Improved Convolutional Capsule Network for Compound Fault Diagnosis of RV Reducers. *Sensors* 22, 6442.

Yan, S., Shao, H., Wang, J., Zheng, X., Liu, B., 2024a. LiConvFormer: A lightweight fault diagnosis framework using separable multiscale convolution and broadcast self-attention. *Expert Systems with Applications* 237, 121338.

Yan, X., Jiang, D., Xiang, L., Xu, Y., Wang, Y., 2024b. CDTFAN: A novel coarse-to-fine dual-scale time-frequency attention fusion network for machinery vibro-acoustic fault diagnosis. *Information Fusion* 112, 102554.

- Yang, G., Tao, H., Du, R., Zhong, Y., 2022. Compound Fault Diagnosis of Harmonic Drives Using Deep Capsule Graph Convolutional Network. *IEEE Transactions on Industrial Electronics*.
- Yang, J., Liu, J., Xie, J., Wang, C., Ding, T., 2021. Conditional GAN and 2-D CNN for Bearing Fault Diagnosis With Small Samples. *IEEE Transactions on Instrumentation and Measurement* 70, 1-12.
- Yi, C., Wang, H., Ran, L., Zhou, L., Lin, J., 2022. Power spectral density-guided variational mode decomposition for the compound fault diagnosis of rolling bearings. *Measurement* 199, 111494.
- Yoon, J., Han, J., Nguyen, T.P., 2023. Logistics box recognition in robotic industrial depalletising procedure with systematic RGB-D image processing supported by multiple deep learning methods. *Engineering Applications of Artificial Intelligence* 123, 106311.
- Zhang, B., Xu, L., Zhang, J., 2021. Balancing and sequencing problem of mixed-model U-shaped robotic assembly line: Mathematical model and dragonfly algorithm based approach. *Applied soft computing* 98, 106739.
- Zhang, J., Zhang, Q., Qin, X., Sun, Y., Zhang, J., 2022. Gearbox compound fault diagnosis based on a combined MSGMD–MOMEDA method. *Measurement Science and Technology* 33, 065102.
- Zhou, F., Yang, S., Fujita, H., Chen, D., Wen, C., 2020. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowledge-Based Systems* 187, 104837.
- Zhou, X., Zeng, H., Chen, C., Xiao, H., Xiang, Z., 2022. An attention-enhanced multi-modal deep learning algorithm for robotic compound fault diagnosis. *Measurement Science and Technology* 34, 014007.