An intelligent detection approach for multi-part cover based on deep learning under unbalanced and small size samples

Lerui Chen[12], YukMing Tang[1*], Yidan Ma[3], Kai Leung Yung[1]

1 Department of industrial and systems engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China.

2 Zhongyuan-Petersburg aviation college, Zhongyuan University of Technology, Zhongyuan Road NO.41 of Zhengzhou, Zhengzhou,450007, China.

3 State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, 710049, China.

* Corresponding author: Yuk Ming Tang (yukming.tang@polyu.edu.hk)

Abstract: The problem of unbalanced and small samples is main challenge to the application of deep learning in fault detection of complex systems. To address this issue, this paper introduces an intelligent detection approach for multi-part cover (MPC) based on auto-encoder Wasserstein generative adversarial networks (AEWGAN) and structure adaptive adjustment convolution neural network (SAACNN). The proposed approach incorporates data augmentation techniques and a detection algorithm to enhance the accuracy of MPC detection. For the data enhancement, a novel AEGWAN model is proposed to enhance the correlation and reduce the difference between the generated samples and real samples, achieved by replacing the random noise vector in the traditional generative adversarial network (GAN) with hidden variables auto-encoded by real samples. In addition, the Wasserstein distance is utilized to substitute for the Kullback Leibler (KL) divergence or Euclidean Distance (ED) in traditional GAN as the objective function. This substitution helps eases the gradient disappearance and training instability in the training process. For the detection algorithm, although AEWGAN can expand the samples, there are still differences between the generated and real samples due to the limitations of the model. To further ease the effect of the difference for detection accuracy, a novel comprehensive loss function is designed for CNN. On the basis of the new loss function, a novel SAACNN is created to adaptively select the optimal network structure, which speeds up network training progress and improves the detection accuracy. The effectiveness of the proposed approach is verified by experiments with other models, showcasing its superior capabilities in terms of data enhancement, denoising and generalization.

Keywords: Multi-part cover; Fault detection; Unbalanced samples; Small size samples; Deep learning

1 Introduction

Urban road covers play a vital role in providing convenient access for maintenance and repairs, while also ensuring cleanliness and safety, as the demand for underground infrastructure, including power cables and gas pipes, continues to grow alongside expanding urban areas. With the constant growth and complexity of urban infrastructure, manhole covers have gained more widespread application[1]-[2]. The multi-part cover (MPC), a special form of manhole cover, offers the advantages of cost-efficiency and large coverage area, making it the optimal choice on the road in many countries and regions such as United States, Canada, United Kingdom, Hong Kong, Japan. The structure of MPC is shown in Fig.1.

Fig.1 The structure of MPC

Unlike the ordinary manhole cover, MPCs are constructed using concrete and steel strips, forming a component that resembles in a grid pattern. These covers are not a single, solid unit but rather a composite structure, making them both heavy and more intricate in design. Nevertheless, due to the heavy weight and rolling by vehicles for a long time, the cover and the supporting beam under the cover are prone serious damage. In severe cases, the MPC may collapse, posing a hidden risk to vehicle safety. In Hong Kong alone, there are already thousands of MPCs installed on the main roads. Hence, the timely detection of faults in MPCs is both urgent and significant importance.

Currently, manual inspection is the primary method used to assess the health of MPCs. However, this approach suffers from low efficiency. Moreover, manually detecting internal damage in MPCs, such as the supporting beam damage, is challenging, which increases the risk of mis-detection. It is urgent to develop an intelligent approach to detect the MPC health. However, the researches of MPC intelligent detection are rare. Current manhole cover health detection researches could provide references for MPC. Yu *et al* [3] and Yadav *et al* [4] proposed a detection approach for manhole covers suing mobile laser scanning (MLS). They designed a corresponding model to extract the state features from the scanning images. Pasquet *et al* [5] and Liu *et al* [6] put forward a detection approach for manhole using satellite images. By extracting the features from the satellite images, fault detection could be accomplished. These approaches above have some shortcomings, such as, images are greatly impacted by the external environment, which is not fit in cloudy or rainy days. In addition, unlike the manhole cover, MPCs are bigger and have the same material as the road, making it difficult to obtain clear and complete images for detection. What's more, this approach is only suitable for the detection of MPCs with severe surface damage and not internal damage. Some researches adopted radio frequency identification (RFID)[7]-[8] or ground penetrating radar (GPR)[9]-[10]. Although the technologies of RFID and GPR can detect the manhole cover health states, deployment of a large volume of devices is needed, which gives rise to high maintenance costs. With the expansion of urban areas, abundant MPCs are adopted in urban roads. Hence, fast, convenient, efficient and low-cost detection is vital. Vibration signals, which contain many internal health information of the system, are extensively being applied in fault detection and identification[11]-[13] The advantage of vibration lies not only in its convenient acquisition, but also in its low cost. Besides, the entire detection process will create little impact on pedestrians and vehicles. Therefore, this paper adopts the vibrational signals to realize the fault detection of MPC.

Machine learning, especially deep learning, has been widely utilized in fault detection. Compared with traditional machine learning methods that rely on manual extraction of fault features, deep learning implements the fault feature extraction and classification in an end-to-end manner, avoiding the cumbersome operations of manual feature

extraction and enhance feature characterization[13]-[16]. However, the quality and quantity of training samples largely determine the performance of fault detection based on deep learning. At present, most research on deep learning is conducted with a balanced and large number of training samples. However, practical engineering applications often yield skewed dataset that contains a greater proportion of normal data and fewer fault data. This results in data imbalance between different states and subsequently increases the risk of over-fitting in the deep learning network. In addition, this scenario causes the network model to focus on the training of abundant samples, with decision-making ability towards the minority samples. As a result, the imbalance hinders the network's ability to make correct judgments. In response, Goodfellow *et al* [17] proposed the GAN, through the confrontation between the generator and the discriminator to achieve the 'Nash equilibrium'. Ultimately, the generator produces synthetic data that closely adheres to the real data distribution. In recent years, the GAN network has proven to be an emerging and effective data enhancement network, applied in many fields, such as visual enhancement, natural language processing and fault detection. Its advantages include sample generation, data enhancement, feature completion, etc. Liu *et al* [18] proposed a cycle-generative adversarial network (Cycle-GAN) to realize the image enhancement under unsupervised training. Natarajan *et al* [19] introduced the dynamic GAN model for the generation of high-quality videos. Gao *et al* [20] developed a GAN-based approach for fault diagnosis of bearings for data augmentation under small samples. Luo *et al* [21] proposed a two-stage GAN method for fault diagnosis in mechanical systems. Guo *et al* [22] introduced the GAN based on local weights-shared multi-generator for data enhancement and construct a fault diagnosis model. Liu *et al* [23] developed the latent optimized stable GAN for small samples enhancement without prior knowledge. Although the aforementioned GAN-based data augmentation studies have yielded positive results certain limitations exist. For example, GAN training is difficult, and the generator and concurrent convergence of generators and discriminators often proves difficult, which causes issues such as training instability, gradient disappearance, and model collapse. To solve such problems, some researches proposed improvements based on the traditional GAN. Fan *et al* [24] introduced the Wasserstein distance into GAN to create WGAN to perform the fault diagnosis of rolling bearing under imbalanced data. Meng *et al* [25]adopted the auxiliary classification generative adversarial network (ACGAN) to add the label information to input and introduced the Wasserstein distance into the loss function of ACGAN to mitigate gradient disappearance. To solve the mode collapse and gradient vanishing, Li *et al* [26] introduced the Wasserstein distance into GAN to design a new framework called modified auxiliary classifier GAN (MACGAN), successfully applied to the fault diagnosis of rotating machinery. Chen *et al* [27] introduced the pre-training mechanism and Wasserstein into GAN with gradient penalty to proposed the PT-WGAN-GP to achieve the fault diagnosis of rolling bearing. Although these enhanced GAN approaches have greatly improved the resolution of model gradient disappearance and model collapse by integrating Wasserstein distance, the input of generators comprises random noise signals. The correlation between random noise vectors and real samples is relatively weak. This weak correlation hinders the generation of accurate samples. Therefore, there is a pressing need for further improvement in this regard.

In terms of classification, the current research based on GAN model focuses on how to design generators with label information to achieve data enhancement and classification. Examples include the auxiliary classification generative confrontation network (AC-GAN)[28]-[29],conditional adversarial network (C-GAN)[30]-[32], and

CWGAN-GP[33]-[34]. Although these improved GANs can learn image information based on labels, they struggle to capture the distribution characteristics of images when sample features are subtle, and they do not account for the interpretability of the generator's input to the generated samples. Therefore, despite the early-stage expansion of samples through GANs (or improved GANs) at the data level, it only belongs to the data level processing, and differences still exist between the generated samples and the real samples due to the limitations of the generative model. To further mitigate the impact of these differences on the classification results, it is necessary to design a specialized network for mixed data diagnosis. CNN serves as an advanced neural network with characteristics like local link and weight sharing, rendering it excellent for feature extraction and classification. However, CNN still faces some challenges in fault detection. Firstly, most researches remain entrenched in the mode of "data-driven and result-oriented", which simply pursues accuracy enhancement and ignores the research on the network structure. Secondly, certain parameters in the network, such as layer count and node quantity, are selected blindly, with network structure determined through trial and error. At present, studies on the adaptive adjustment of the network structure are scant. Therefore, designing a CNN that can adaptively adjust the network structure (layer and node count) according to input characteristics is of great significance in improving the speed and accuracy of network.

For the problems of data expansion and fault classification, this paper proposes an intelligent detection approach based on AEWGAN and SAACNN for MPC from the aspects of data enhancement and detection algorithms. The AEGAN model is proposed to perform data enhancement on small size samples to obtain the sufficient samples to be detected. The SAACNN model is proposed to accurately classify the mixed samples. The detail contributions are as follows.

(1) The AEWGAN model improves GANs by using hidden variables encoded from real samples to learn the distribution of real data faster and reduce the differences between the generated and the real samples.

(2) A new comprehensive loss function and CNN with structure adaptive adjustment (SAACNN) are developed to minimize the differences between generated and real samples, improving training speed and accuracy.

(3) The proposed approach outperforms other methods when testing on the MPCs.

The rest of this paper is divided into the following sections. Section 2 describes the history and relevant work of CWT, AEGAN, and SAACNN. Section 3 explains the proposed method for MPC detection. Section 4 presents the implementation of experiment verifications and discusses the findings. Section 5 concludes with final thoughts and proposed directions for further research.

2 Methodology

2.1 Continuous wavelet transform (CWT)

Frequency information is important for signal analysis. Time-frequency domain information contains both the time-domain information and frequency-domain information, which yields a more comprehensive understanding of the internal characteristics of the signal. Among many time-frequency analysis methods, CWT can extract the fault features more through the use of suitable wavelet functions, which avoids complex window method selection. Therefore, CWT is proved more suitable for feature extraction of non-stationary signals, such as the vibrational signal[35]-[36].The CWT can be described by Eqs.(1)-(2).

$$CWT(a,b) =< x, \psi_{a,b} >= \frac{1}{\sqrt{a}} \int x(t)\psi^*(\frac{t-b}{a})\, dt \tag{1}$$

Where,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \tag{2}$$

is the wavelet basis's shifted and dilated forms, which resemble a group of variable window functions. The signal to be converted is denote by $x(t)$, the scale factor and the translation factors are denoted by a(a>0)and b respectively. $\psi()$ is the mother wavelet function and $\psi^*(\frac{t-b}{a})$ is the complex conjugate of $\psi(\frac{t-b}{a})$. The operation of the CWT can concentrate on any part of the signal by altering parameters a and b.

2.2 Autoencoder Wasserstein Generative adversarial network (AEWGAN)

2.2.1 Wasserstein Generative adversarial network (WGAN)

GAN, an unsupervised generation network, consists of two key components: Generator and Discriminator. The function of the generator is to learn the distribution characteristics of real samples to generate new samples that closely match the distribution of real samples, typically taking a random noise vector as input. The discriminator is used to compare the difference between the real samples and the generated samples. In the process of network training, the generator and the discriminator undergo alternating training phases and engage in competitive and adversarial interaction until they attain the Nash equilibrium. At this point, the samples generated by the generator are close to the real samples.

GAN can be abstracted as a maximum and minimum problem which is shown as Eq.(3).

$$\min_G \max_D V(D,G) = E_{x-P_{data}(x)}[log\, D\,(x)] + E_{z-P(z)}[log(\,1 - D(G(z)))] \tag{3}$$

Where, E represents the expectation. $D(x)$ represents the probability of determining the real sample to the real. $D(G(z))$ represents the probability of the generated sample to the real. $P_{data}(x)$ represents the distribution of the real sample. P_z represents the distribution of the noise, which is a Gaussian distribution. $V(D,G)$ represents the difference between the real sample and the generated sample.

The objective function of GAN can be decomposed into two optimization functions, which are shown as Eqs.(4)-(5).

$$L_G = \min_G V(D,G) = E_{z-P(z)}[log(\,1 - D(G(z)))] \tag{4}$$

$$L_D = \max_D V(D,G) = E_{x-P_{data}(x)}[log\ D\ (x)] + E_{z-P(z)}[log(\ 1 - D(G(z)))] \tag{5}$$

Where, L_G is the loss function of generator. L_D is the loss function of discriminator.

During network training, the objective is to minimize the loss function of the generator L_G and maximize the loss function of the discriminator L_D. The distribution of real samples will be obtained by the generator causing the discriminator to struggle in differentiating between real and generated samples. The optimal solution of the discriminator can be described as Eq.(6).

$$D^*(x) = \frac{P_{data}(x)}{P_{data}(x) + P_g(x)} \tag{6}$$

Where, $P_g(x)$ is the distribution of samples generated by the generator. When $P_g(x) = P_{data}(x)$, $D^*(x)$ =0.5, it indicates that the state reaches to Nash balance and the global optimal solution is obtained.

However, the adoption of KL or JS distance to optimize Eq.(3) will inevitably lead to gradient disappearance and training instability[37]. Based on GAN, the WGAN model is proposed by many researchers, in which Wasserstein distance is adopted to quantify the difference between the generated and real samples. Its structure is shown in Fig.1.

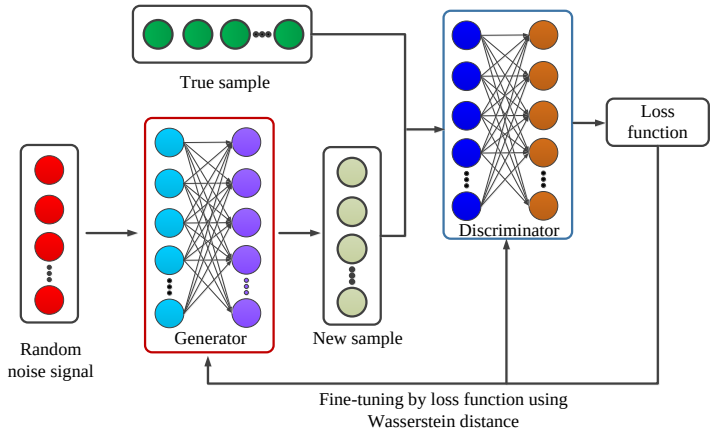


Fig.1 WGAN structure

Adopting the model of Wasserstein distance function can alleviate the problems of gradient disappearance and training instability in the training process. The function of Wasserstein distance can be expressed as Eq.(7).

$$W(P_{data}, P_g) = \inf_{\gamma \in \Pi(P_{data}, P_g)} E_{(x,\tilde{x}) \sim \gamma}[\|x - \tilde{x}\|] \tag{7}$$

Where, $\Pi(P_{data}, P_g)$ represents the joint probability distribution γ between P_{data} and P_g. $\|x - \tilde{x}\|$ represents the distance between the real sample and the generated sample. The smaller the value of $W(P_{data}, P_g)$, the higher the similarity between the real sample and the generated sample.

2.2.2 Auto-encoder (AE)

AE is a single-layer neural network composed of an encoder and a decoder, and the entire calculation includes two processes: encoding and decoding. The specific structure is shown in Fig.2.

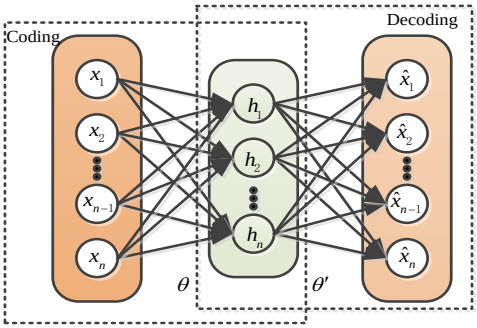


Fig.2 .AE network

The process of encoding can be expressed as Eq.(8).

$$\begin{cases} h_n = f_\theta(x_n) = S_f(W \cdot x_n + b) \\ \theta = \{W,\ b\} \end{cases} \tag{8}$$

Where, x_n is input. θ is the parameter sets. W is the weight matrix; b is the offset.S_f is the activation function.

The process of encoding can be expressed as Eq.(9).

$$\begin{cases} \hat{x}_n = g_{\theta'}(h_n) = S_{\mathrm{g}}(W' \cdot h_n + d\ ) \\ \theta' = \{W',\ d\} \end{cases} \tag{9}$$

Where, $\hat{x}_n$ is input. θ' is the parameter sets. W' is the weight matrix; d is the offset.S_{g} is the activation function.

By continuously training the parameters set $\{\theta, \theta'\}$, making the loss function $L(x, \hat{x})$ minimization, which is shown in Eq.(10).

$$L(x, \hat{x}) = \underset{\theta,\theta'}{argmin} \frac{1}{m} \sum_{i=1}^{m} \|x - \hat{x}\|^2 \tag{10}$$

2.2.3 Design of AEWGAN

This paper utilizes the encoding capacity of AE and the adversarial learning mechanism of WGAN to propose the AEWGAN for sample enhancement. In the model of AEWGAN, the random noise vector is replaced with the encoded hidden variable. This replacement enhances the correlation between the generated and real samples and improves the accuracy of generated samples. AEWGAN includes an auto-encoder network, a generation network, and a discrimination network, as shown in Fig.3.

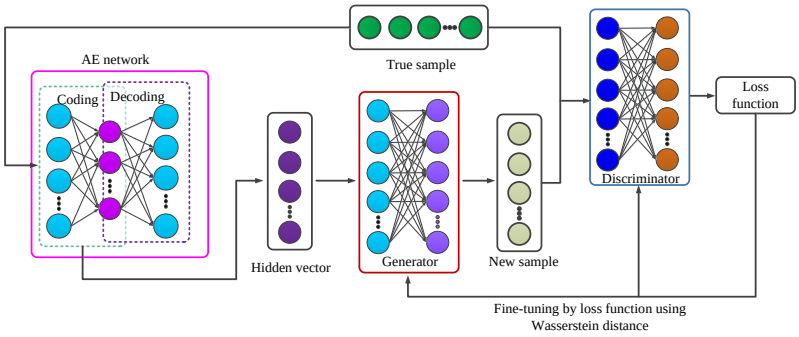


Fig.3 AEWGAN structure

(1) ***AE***: Its function is to obtain the hidden variables associated with real samples through encoding operations, and adopt the hidden variables as input for the generator.

(2) ***Generation network***: This network is composed of several layers of deconvolution transformed convolutions, which increases the dimensionality of low dimensional hidden variables layer by layer, and then generates the new samples that are consistent with the real samples. Its task is to adopt Eq.(4) to generate fake samples to deceive the discriminator.

(3) ***Discrimination network***: This network consists of several convolutional modules and a full connection layer. It performs simultaneous feature extraction and dimensionality reduction for both generated and real samples, and maps these samples to true or false probability values using a nonlinear activation function.

The AEWGAN model designed in this paper combines the encoder from AE with the generator of WGAN, which not only improves the quality of generated samples, but also addresses training instability and gradient disappearance issues.

2.3 Convolution neural network with structure adaptive adjustment (SAACNN)

CNN, an advanced neural network, is composed of convolutional layers, pooling layers, and full connection layers. It is equipped with strong feature extraction capabilities and finds widespread application in pattern recognition. CNN operations mainly include two phases: forward propagation and reverse fine-tuning. Forward propagation is adopted to pre-calculate the sample categories, while reverse fine-tuning continuously corrects and adjusts the network parameters by comparing the error between the pre-calculation results and the real results, which makes the calculation in the direction of error reduction. The loss function in traditional CNN is shown in Eq.(11).

$$J(\omega, b) = \frac{1}{2m}\sum_{i=1}^{m}\left\| g_{\omega,b}(x^{(i)}) - y^{(i)} \right\|^2 \tag{11}$$

Where, $m$is the samples size.$g_{\omega,b}(x^{(i)})$ and $y^{(i)}$are the predicted value and real value of the sample $x^{(i)}$, respectively.

Although the loss function shown in Eq.(11) can guide network training, it overlooks the differences between samples of different types, which leads to defects in convergence speed and accuracy. Therefore, the novel loss function is proposed to optimize network parameters to aggregate samples of the same type and diverge samples of varying types, which enhances the accurate identification of the classifier.

The error function between same type samples is shown in Eq.(12).

$$J_1(\omega, b) = \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{m_j}\left\|g_{\omega,b}(x^{(i,j)}) - M^{(j)}\right\|^2 \tag{12}$$

Where, n is the size of the sample type. m_j and $g_{\omega,b}(x^{(i,j)})$ are the sample size and the sample predicted value of j-th type respectively. The samples mean value of j-th type is denoted by $M^{(j)}$, and calculated using Eq.(13).

$$M^{(j)} = \frac{\sum_{i=1}^{m_j} g_{\omega,b}(x^{(i,j)})}{m_j} \tag{13}$$

The error function between vary type samples is shown in Eq.(14).

$$J_2(\omega, b) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\|M^{(i)} - M^{(j)}\right\|^2 \tag{14}$$

Where, $M^{(i)}$ is the samples mean value of i-th type and $M^{(j)}$ is the samples mean value of j-th type.

The novel comprehensive loss function is shown in Eq.(15).

$$J_{new}(\omega, b) = J(\omega, b) + \gamma J_1(\omega, b) - \beta J_2(\omega, b) \tag{15}$$

Where,$J(\omega, b)$ is the traditional error function. $J_1(\omega, b)$and $J_2(\omega, b)$ are the errors between same types and different types of samples respectively. γ and β are weight coefficients obtained by prior knowledge.

Iterative optimization of the gradient descent method is used to determine the minimum value of $J_{new}(\omega, b)$, which is shown in Eq.(16).

$$\omega_{ij}^{l} = \omega_{ij}^{l} - \eta\frac{\partial}{\partial\omega_{ij}^{l}}J_{new}(\omega, b)$$
$$b_j^l = b_j^l - \eta\frac{\partial}{\partial b_j^l}J_{new}(\omega, b) \tag{16}$$

Where,η is the learning rate.

The selection of network layers and nodes in traditional CNN relies on empirical experience, which can be a blind process. To overcome the blindness of network structure selection, this paper proposes a network structure adaptive adjustment algorithm based on Eq.(15). The procedure is outlined as follows.

(1) Establish a CNN model with a single convolution layer, one pooling layer and one full connection layer and initialize the network parameters, with a node increasing stride set to 2 and the layer increasing stride set to 1.

(2) Feed training samples into the CNN model and calculate the loss function $J_{new}(\omega, b)$. Analyze the change trend of $J_{new}(\omega, b)$. If the loss function decreases smoothly and quickly to a predefined threshold as iterations increase, retain the network structure. If the loss function exhibits instability with increasing iterations, augment the nodes in each layer and continue training. In cases where the loss function decreases rapidly and smoothly to the threshold with the iterations increasing, but until the training task quit, the loss function still surpasses the threshold,

increasing the network layers.

(3) Repeat step (2) several times until the termination condition is reached, then adopt the optimal structure obtained from step (2) for testing.

The network structure adaptive adjustment algorithm can be shown in Fig.4.

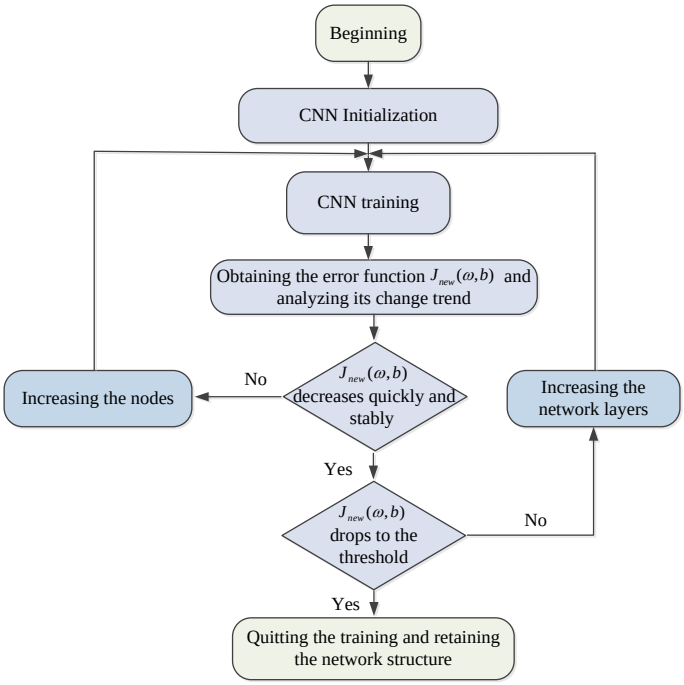


Fig. 4. The structure adaptive adjustment algorithm

2.4 AEWGAN-SAACNN proposed in this paper

The AEWGAN-SAACNN designed in this paper includes two parts: sample enhancement network and classification network, illustrated in Fig.5.The AEWGAN model serves to implement the sample enhancement, while the SAACNN model serves to classify the mixed samples. The two models are connected in series, where the output of the previous network serves as ~~is~~ the input for the next network.

Unlike the traditional GAN, the input of AEWGAN proposed in this paper is the hidden variables generated through encoding real samples rather than the random noise, which enhances the correlation and similarity between the real and generated samples. The generator and discriminator within AEWGAN are composed of a sequence of transposed convolution layers and convolutional layers, and the two modules interact to generate the optimal samples by loss function gradient updates. SAACNN, on the other hand, is constructed with convolution layers, pooling layers, and one full connection layer. Through a series of convolution and pooling operations, fault features hidden in the mixed samples can be deeply extracted. Meanwhile, the SAACNN employs a structure adaptive adjustment algorithm to accelerate the network training process and enhance feature extraction accuracy by adaptively optimizing the network structure. The parameters of two models are shown in Section 4.3.

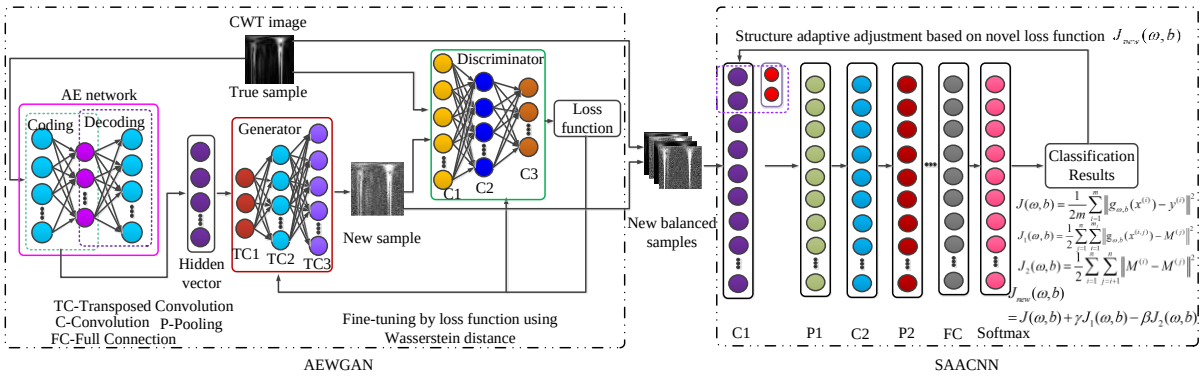


Fig.5 The structure of AEWGAN-SAACNN designed in this paper

3 The overall process of fault detection for MPC

For the MPC fault detection under imbalanced and small size samples, in this paper, AEWGAN is designed for sample enhancement, and SAACNN is proposed for state classification. The overall process of fault detection can be shown in Fig.6.

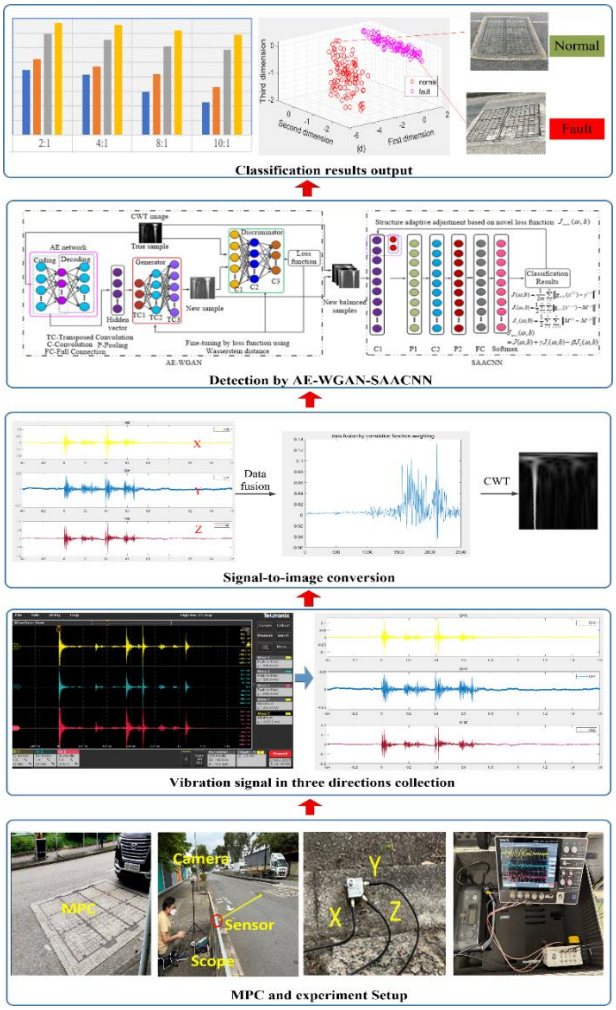


Fig.6 Overall process of MPC health detection

The details are as follows.

（1）***Data collection and processing***. Build an experimental platform to collect the vibrations in the X, Y, and Z directions of MPC under normal and fault conditions (500 normal samples and 500 fault samples), then the correlation function weight method is adopted to fuse the vibrations in three directions. Finally, CWT is adopted to process the fused data to obtain 2D gray images. 80% of images are randomly selected as the training set, and the remaining are taken as the testing set. Remove the fault samples by a random proportion to construct the multiple unbalanced training sets.

（2）***Samples enhancement.*** Put the unbalanced samples with vary degrees into AEWGAN model for training. When it reaches to the maximum times of training, reserved the optimal model. The generated samples could be obtained by the optimal model. Finally, mix the generated samples into the real samples to achieve the small size samples enhancement.

(3) ***Healthy state classification***. Put the mixed training set into the SAACNN model, and adopted the novel comprehensive loss function to guide the network training. The network adaptively adjusts the number of network layers and nodes based on the changes trend of the loss function. After the optimal network is obtained, put the testing set into the network model to achieve the classification of various samples. The training and testing process can be shown in Fig.7.

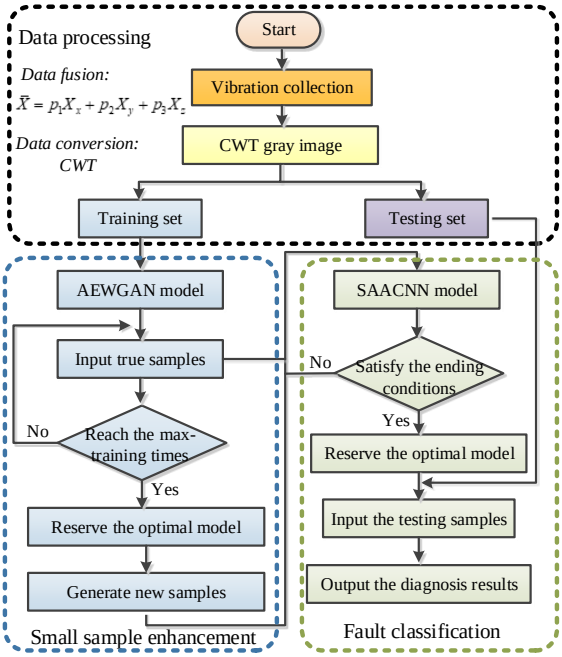


Fig.7 The flow of training and testing

4 Experiment verifications and discussions

To verify the effectiveness of proposed approach, MPCs in HongKong are taken for fault detection, several groups of experiments are implemented and the results are discussed.

4.1 Data acquisition

The data acquisition instrument mainly includes a scope and a vibration sensor, which is shown in Fig.8. The three-directional accelerometer sensor (CT1020LS, CHENGTEC, CHINA) is connected to an oscilloscope (Tektronix MSO24, 4 Analogue Channels, 200 MHz, CHINA) to observe the signal in real-time, where the frequency sampling rate is set as 100 KHz. Besides, the sensor is mounted on the aluminum piece using bolts to fix the sensor. Prior to attaching the sensor, the aluminum piece was set up near the testing site with an adhesive glue (Devcon 22045, USA). When the vehicle passing through the MPC, the vibration data could be recorded by the scope.500 groups of data in each state are collected and the sampling time for each group data is 2s.

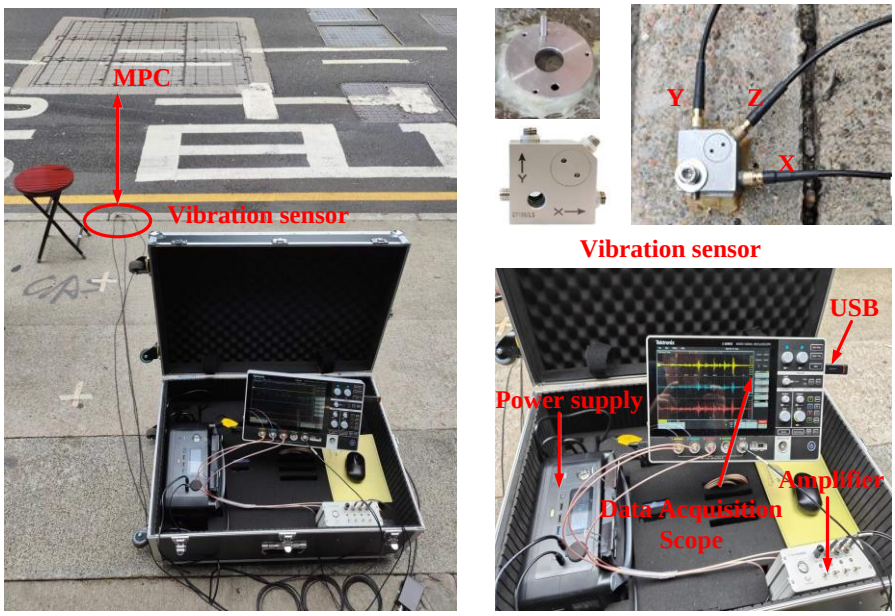


Fig.8 The data acquisition platform

MPCs in vary locations are adopted for research in this paper, normal and fault MPC are shown in Fig.9. For the fault, the fault occurs in the supporting beam under the cover, and the fault type is supporting beam deformation. The vibration of two states is shown in Fig.10.

(a) (b)

Fig.9 MPC: (a) normal; (b) fault

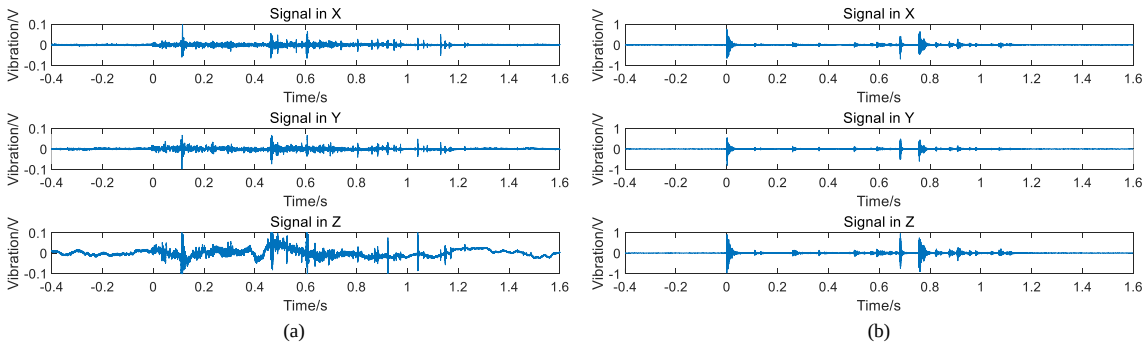


Fig.10 The vibration of MPC: (a)normal;(b) fault

From Fig.10, it can be concluded that the vibration signal in good state is weak and the vibration in bad state is strong. The reason is that when the supporting beam is bent, the MPC is in unbalanced state, when the vehicle rolls the MPC, it will collide the beam.

4.2 Data fusion and conversion

Due to the fact that data in multi-directions contain richer state information than data in single direction[38], The sensor adopted in this paper can simultaneously collect vibrations in the X, Y, and Z, then, the cross-correlation function method in [39] is adopted to fuse the vibration in three directions. Assuming that $x_1(n), x_2(n), x_3(n), \cdots, x_m(n)$ is the measured signal，the cross-correlation function for any two signals is as follows.

$$R_{x_i,x_j}(m) = \frac{1}{N-m}\sum_{i=1}^{N-m} x_i(n)x_j(n+m)\,, m = 0,1,2,\cdots,k \tag{17}$$

The energy of signal can be expressed as follows

$$E_{ij} = \sum_{k=1}^{n}\left[R_{x_i,x_j}(k)\right]^2 \tag{18}$$

Where, E_{ij} is the correlation energy of signal by cross-correlation operation. The energy of signal by the i-th sensor can be expressed as follows

$$E_i = \sum_{j=1,j=i}^{n} E_{ij} \tag{19}$$

Because the weighted p_i is proportional to the energy, therefore

$$p_1 : p_2 : \cdots : p_n = E_1 : E_2 : \cdots : E_n \tag{20}$$

$$p_1 + p_2 + \cdots + p_n = 1$$

Hence, the fused signal can be expressed as follows

$$\bar{X} = p_1 X_1 + p_2 X_2 + \cdots + p_n X_n \tag{21}$$

Adopting the cross-correlation function method to fuse the vibration of MPC as Fig.10, the result can be shown in Fig.11.

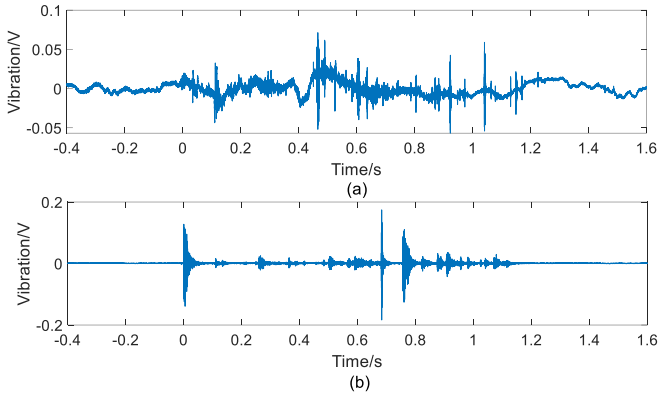


Fig.11 Data fusion by cross correlation function in two states: (a)normal;(b) fault

To meet with the requirement of the model in this paper for input, it is necessary to convert the original 1D vibration into 2D image. In this paper, CWT is adopted to process the vibration, and some of the CWT images are randomly selected, which are shown in Fig.12.

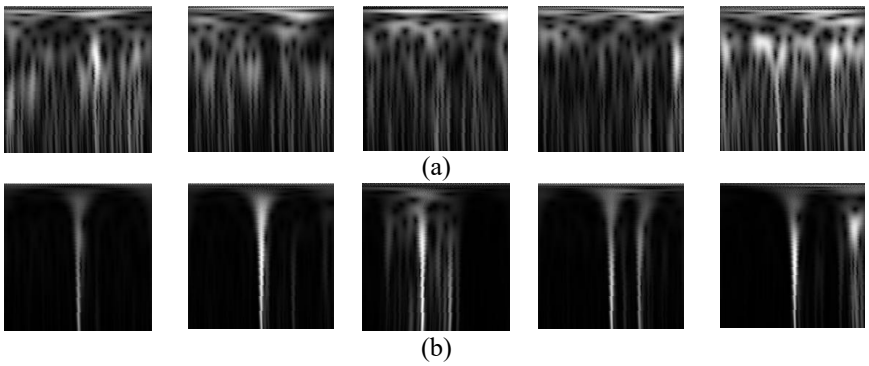


Fig.12 CWT gray images of two states:(a) normal (b) fault

From Fig.12, it can be seen that the CWT gray images under different states have significant differences. Therefore, CWT images can be adopted as the fault information for detection. Select 500 fusion samples under normal and fault states to obtain 500 CWT images for each state. To construct the unbalanced samples that meets the experiment requirements, fault samples are randomly removed by a certain proportion, which is shown in Table 1.

Table 1 Unbalance samples of MPC

Sample sets	Normal	Fault	Unbalance ratio (normal:fault)
Training set A	400 samples	200 samples	2:1
Training set B	400 samples	100 samples	4:1
Training set C	400 samples	50 samples	8:1
Training set D	400 samples	40 samples	10:1
Testing set	100 samples	100 samples	1:1

4.3 Comparative experiments

The software development environment of following experiments is based on Python programming language, and the operation system of computer is Window10. The computer configuration is Intel (R) core (TM) i7-4790 CPU @ 3.6GHz with a memory (RAM) size of 32GB. The configuration of network parameter is shown in Table 2 and Table 3.

Table 2 The structure of AEWGAN

Model	Layer	Variables	Training parameters
AE	Input	200 neurons	The batch size is 32. *ReLu* is adopted as the activation function. The learning rate is set to 0.002. The training times is set to 500
	Hidden	100 neurons	
	Output	200 neurons	
Generator	Deconv1	the kernel size is 3×3, the stride is 2, and the padding is 1	
	Deconv2	the kernel size is 3×3, the stride is 2, and the padding is 1	
	Deconv3	the kernel size is 3×3, the stride is 2, and the padding is 1	
	Deconv4	the kernel size is 3×3, the stride is 1, and the padding is 1	
Discriminator	Conv1	the kernel size is 5×5, the stride is 2, and the padding is 1	
	Conv2	the kernel size is 5×5, the stride is 2, and the padding is 1	
	Conv3	the kernel size is 5×5, the stride is 2, and the padding is 1	
	Conv4	the kernel size is 5×5, the stride is 2, and the padding is 1	

Table 3 The initialization parameters of SAACNN

Input	Layer	Variables					Training parameters
		Activation function	Convolution kernels	**Size** (Stride)	pooling operation	pooling kernel (Stride)	
CWT images with size of 200×200	Conv1+BN+Pooling 1	*ReLu*	8	3×3 (2)	Max	2×2 (2)	The batch size is set to 64, and the initial learning rate is 0.001. The learning rate is reduced to half of its previous value every 20 epochs. The maximum number of epochs is set to
	Conv2+BN+Pooling 2	*ReLu*	16	3×3 (2)	Max	2×2 (2)	
	Conv3+BN+Pooling 3	*ReLu*	16	3×3 (2)	Max	2×2 (2)	

	Conv4 +BN+P ooling 4	*ReLu*	32	3×3 (2)	Max	2×2 (2)	200, and the training iterations are set to 500. In new loss function, $\gamma = 0.6$, $\beta = 0.4$
	Full connect ion	2 nodes with the *ReLu*	/	/	/	/	
	Softma x	/	/	/	/	/	

4.3.1 The quality comparisons of generated samples

The purpose of the sample augmentation network is not only to expand the sample size but more importantly, to generate convincing samples in terms of quality. This emphasis on sample quality is crucial as it directly influences the accuracy of detection network. To verify the effectiveness of generated samples by AEWGAN, the imbalanced ratio 10:1 is considered, as shown in Table 1 for an example (training set D). In this context, this paper utilizes AEWGAN, GAN, WGAN, infoGAN [40], and CWGAN [33] respectively to enhance the fault samples. Four images generated by different approaches are randomly selected, as shown in Fig.13.

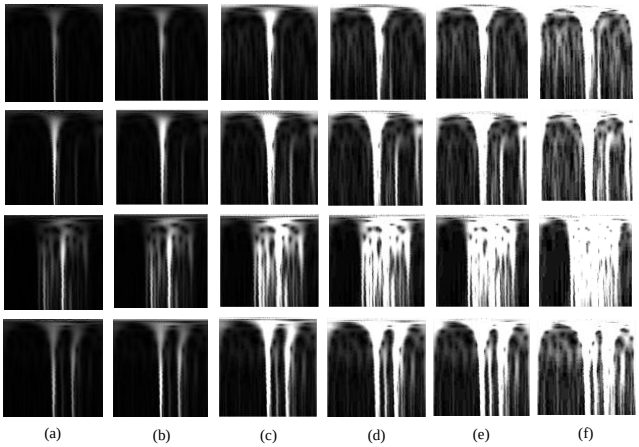


Fig.13 The comparison of real samples and generated samples: (a) real (b) by AEWGAN (c) by infoGAN (d) by CWGAN (e) by WGAN (f) by GAN

From Fig.13, it can be observed that the images generated by AEWGAN bear the closest resemblance to real images. Although not exact replicas, this similarity signifies that the proposed network not only learns feature distributions from the real samples, but also maintains diversity among generated samples, which is beneficial to improve the robustness and generalization capacity of the classification network. In contrast, the samples generated by GAN are different from the real samples both in the key and edge features, indicating GAN's limitations in learning features and generating high-quality samples. To quantify the difference between the generated and the real samples, Kullback Leibler (KL) divergence [41] and Euclidean Distance (ED) [42] are adopted, respectively, to evaluate the samples generated by different sample enhancement networks. The results are detailed in Table 4.

Table 4 The KL and ED values of the generated samples by vary models

Model	KL	ED
GAN	2.4173	10.5847
CWGAN	1.9152	7.2963
WGAN	2.0017	8.6391
infoGAN	1.1532	3.7295
AEWGAN	1.5786	1.7082

From Table 4, it can be seen that among many GAN methods, AEWGAN performs the best in KL and ED indicators, while the traditional GAN performs the worst. The reason may be that the input of traditional GAN is the random noise, which makes it difficult for the generator to learn the feature distribution in real samples. What's more, the traditional GAN adopted KL distance or JS distance to describe the loss function, rather than the Wasserstein distance function, it is easy to encounter the gradients disappearance, resulting in the network parameters stop updating, which makes it difficult to learn useful knowledge. However, AEWGAN avoids the occurrence of the aforementioned problems, resulting in the higher similarity between the generated and the real.

To further demonstrate the effectiveness of AEWGAN, taking the unbalance degree 10:1 as an example, principal component analysis (PCA) is utilized to perform the visualization of dimensional reduction on the unbalanced samples of original and the balanced samples by enhancement, extracting the first three features for visualization. The results are shown in Fig.14.

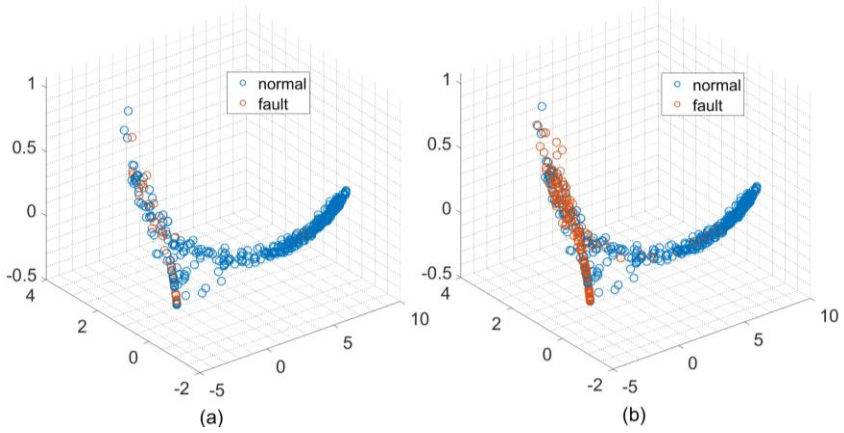


Fig.14 Data feature distribution: (a) unbalanced samples of original (b) balanced samples by enhancement

As illustrated in Fig.14, through the mixture of the generated samples, the number of fault samples is consistent with that of normal samples. In addition, the fault samples generated are clustered around the real fault samples, indicating that the AEWAGN model has learned the feature distribution from the real fault samples. Hence, the distribution of the generated data more closely resembles that of real samples.

4.3.2 Comparisons with single direction vibration

To illustrate the effectiveness of data fusion on MPC fault detection, taking the unbalanced ratio 8:1 as an example, for fairness, this paper adopts the same data enhancement network (AEWGAN) and the same detection network (SAACNN) to test the data in X, Y, Z and the fused data respectively. Each experiment is repeated 10 times, and average accuracies are shown in Fig.15.

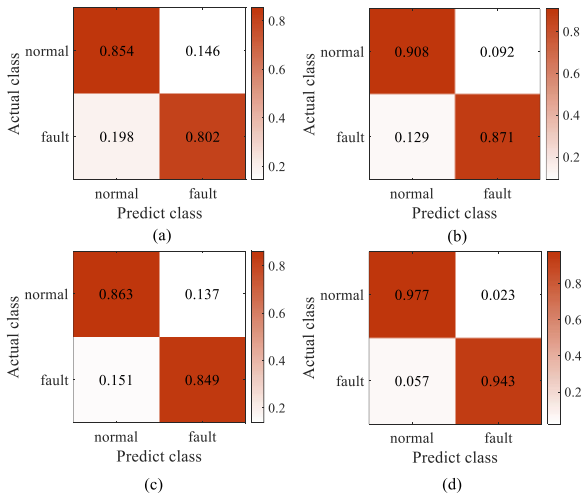


Fig.15 The comparisons with vibration :(a) in X;(b) in Y;(c) in Z;(d) data-fusion

As depicted in Fig.15, compared with the methods using vibration in X, Y, and Z for detection, the average accuracy of approach by data fusion is 96.0%, which is improved by 13.2%, 7.05%, and 10.4%, respectively. Obviously, data fusion method has more advantages. The reason may be that the cross-correlation function fusion can adjust the size of correlation coefficients by the energy of signal, which ensures both the information diversity and the allocation rationality. Compared with the information in single direction, information fusion is advantageous for the classifier to make accurate recognition.

4.3.3 The effectiveness verification of AEWGAN-SAACNN

To verify the performance of AEWGAN-SAACNN with unbalanced samples, CNN, SMOTE-CNN and AEWGAN-CNN are adopted with different unbalanced degrees (2:1,4:1,8:1,10:1) in Table 1 for comparisons. The SMOTE realizes the data expansion by random interpolation in real samples. Each approach is repeat 10 times with different balance degrees, and the results are shown in Fig.16.

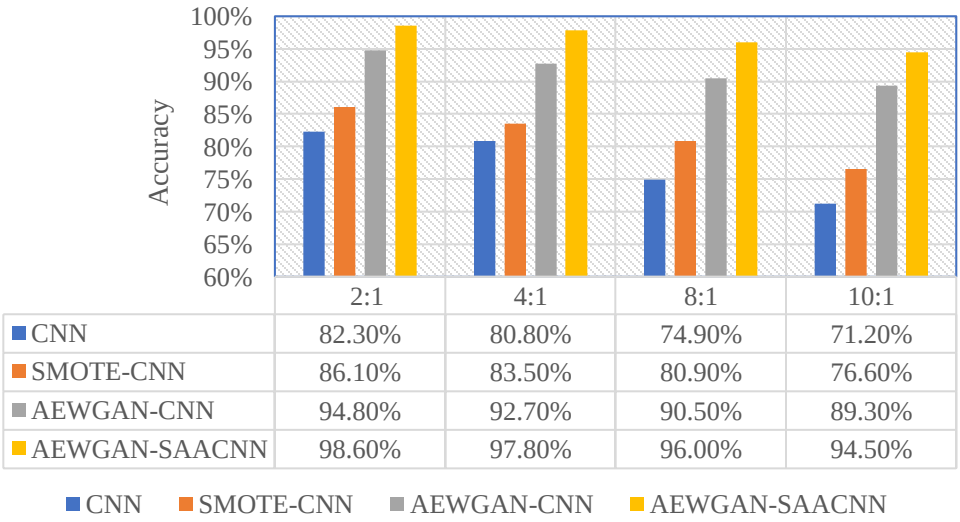


Fig.16 The accuracies of different networks with different unbalance degrees

As can be seen from Fig.16, **firstly**, when unbalanced samples are directly fed into CNN for detection, the accuracy notably lags behind other methods using data enhancement, which verifies the conclusion that unbalanced samples can substantially hamper detection accuracy. **Secondly**, the extent of imbalance has a considerable impact on the classification result. With increasing imbalance, the detection performance of the four approaches decreases, but the AEWGAN proposed in this paper maintains the highest accuracy (above 90%), which further proves its effectiveness in samples enhancement for MPC detection. It is worth noting that SMOTE is a common data enhancement method, but its accuracy in MPC is relatively lower since SMOTE adopts the simple interpolation mechanism to increase samples size, which fails to enhance the diversity of sample features, not conducive to improving the generalization ability of the model.**Thirdly**, compared with CNN, SAACNN can automatically adjust the network structure to optimize performance according to input characteristics, instead of blindly selecting the network structure based on experience. Hence, the highest accuracy can be achieved under varying degrees of sample imbalance. To further illustrate the feature extraction capabilities of the four models, take the unbalanced ratio 8:1 (**Training set C**) in Table 1 as an example for detection. Visualizations of feature extraction results in each network are illustrated in Fig.17.

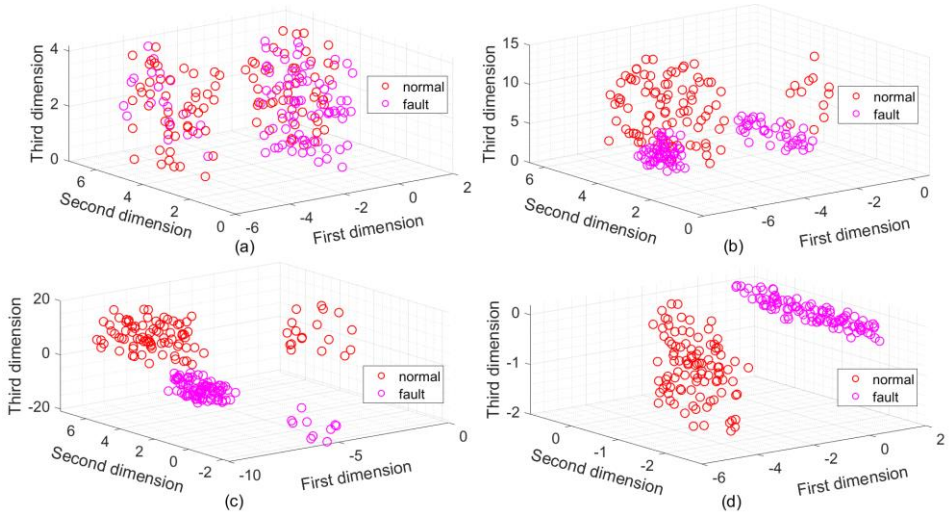


Fig.17 The visualization of various models:(a) CNN;(b) SMOTE-CNN;(c) AEWGAN-CNN;(d) AEWGAN-SAACNN

From Fig.17, it can be seen that the four models have significant differences in feature extraction capabilities. In Fig.17 (a), the fault features are more divergent, resulting in a larger overlap area with the normal features. Compared with Fig.17 (a), the fault features in Fig.17 (b) and Fig.17 (c) are relatively clustered, but overlaps with normal features are still exsited. In Fig17(d), the fault features are the most clustered and have clear boundaries with the normal, which indicates that the proposed approach has strong feature extraction and differentiation capabilities.

4.3.4 Anti-noise performance analysis

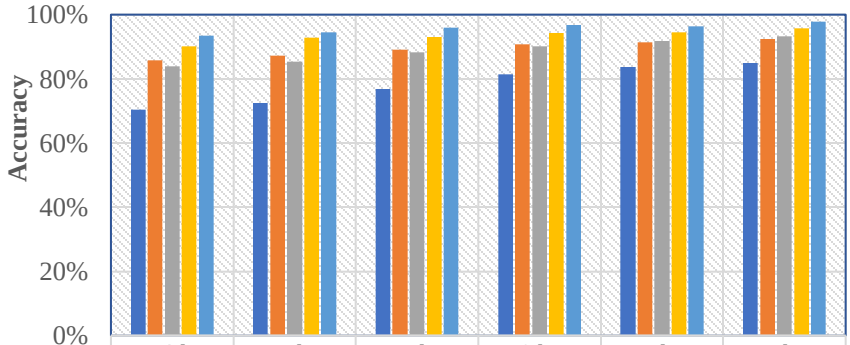
Since MPC located in a noisy environment, the vibration collected is easily polluted by the noise, therefore, the anti-noise ability of the designed network is an important indicator to evaluate its superiority. To verify the anti-noise performance of the network designed in this paper. Gaussian white noise with different signal-to-noise ratio (SNR) is

added to the original vibration to simulate the noise-interference environment, which is defined as Eq.(22).

$$SNR = 10lg^{\frac{P_{signal}}{P_{noise}}} \tag{22}$$

Where, P_{signal} represents the energy of the clean signal. P_{noise} represents the energy of noise.

After the data enhancement by AEGWAN, the noise convolution neural network(NOSCNN)[43], multireceptive field denoising residual convolution neural networks(MFDRCNN) [44], Gramian time frequency enhancement Network (GTFEN)[45] and collaborative fusion convolutional neural network (CFCNN)[46] are adopted as the comparisons. Taking the samples with unbalanced ratio 2:1 in Table 1 as an example for diagnois, 10 experiments are conducted and the average results are shown in Fig.18.



	-6dB	-4dB	-2dB	0dB	1dB	2dB
NOSCNN	70.40%	72.50%	76.80%	81.40%	83.70%	84.90%
MFDRCNN	85.70%	87.30%	89.20%	90.70%	91.30%	92.50%
GTFEN	83.90%	85.40%	88.30%	90.10%	91.80%	93.20%
CFCNN	90.20%	92.90%	93.10%	94.30%	94.50%	95.70%
SAACNN	93.50%	94.60%	95.90%	96.70%	96.30%	97.80%

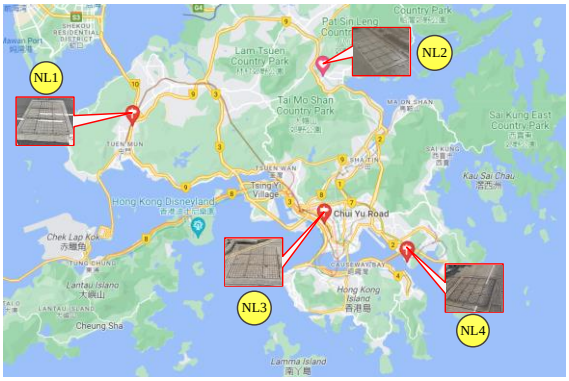
Fig.18 The accuracies of various models with noise

From Fig.18, it can be seen that with the increase in noise intensity, the performance of the five models mentioned above experiences a decline. However, SAACNN consistently maintains the highest average accuracy compared to the other four models. For example, under the strongest noise (SNR=-6dB), SAACNN achieves an average accuracy of 93.5%, surpassing NOSCNN, MFDRCNN, GTFEN, and CFCNN by 23.1%, 7.8%, 9.6%, and 3.3% respectively. This outcome can be attributed to the enhanced loss function model in SAACNN, which optimizes the network parameters in a manner conducive to classification. Meanwhile, SAACNN can adaptively adjust the network structure according to the characteristics of the input to maintain optimal performance even in the noisy environment. Therefore, regardless of fluctuations in noise intensity, SAACNN maintains an accuracy of over 93%. Although the other four models have anti-noise capabilities, their network structures selection has relied on blind and uncertain experience. When the noise intensity changes (the input characteristics are changed), these models lack the adaptability to achieve optimal performance under varying noise conditions. Therefore, the SAACNN proposed in this paper is demonstrated to have strong anti-noise interference ability and robustness.

4.3.5 Comparisons of generalization with other networks

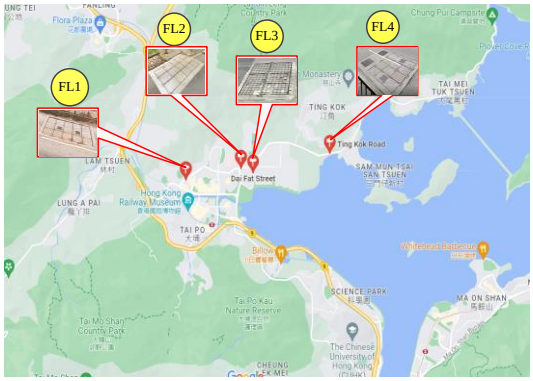
Since there are thousands of MPCs in the city that differ in size and weight.To demonstrate the proposed network

adapts to various types of MPCs, vibrations of MPCs at different positions as shown in Fig.19 are collected.Adopting AEWGAN model for samples enhancement to generate the balanced samples. Then dividing the training and testing sets according to Table 5, Finally, different networks are adopted for detection. Taking the samples with unbalanced ratio 4:1 in Table 1 as an example, each method is tested 10 times and the average accuracy is recorded, which is shown in Fig.20.



NL1:Normal in Tuen Fu Road;**NL2**:Normal in TaiPo Tau Drive;
NL3:Normal in Chui Yu Road;**NL4**:Normal in Shung Shun Street

(a)



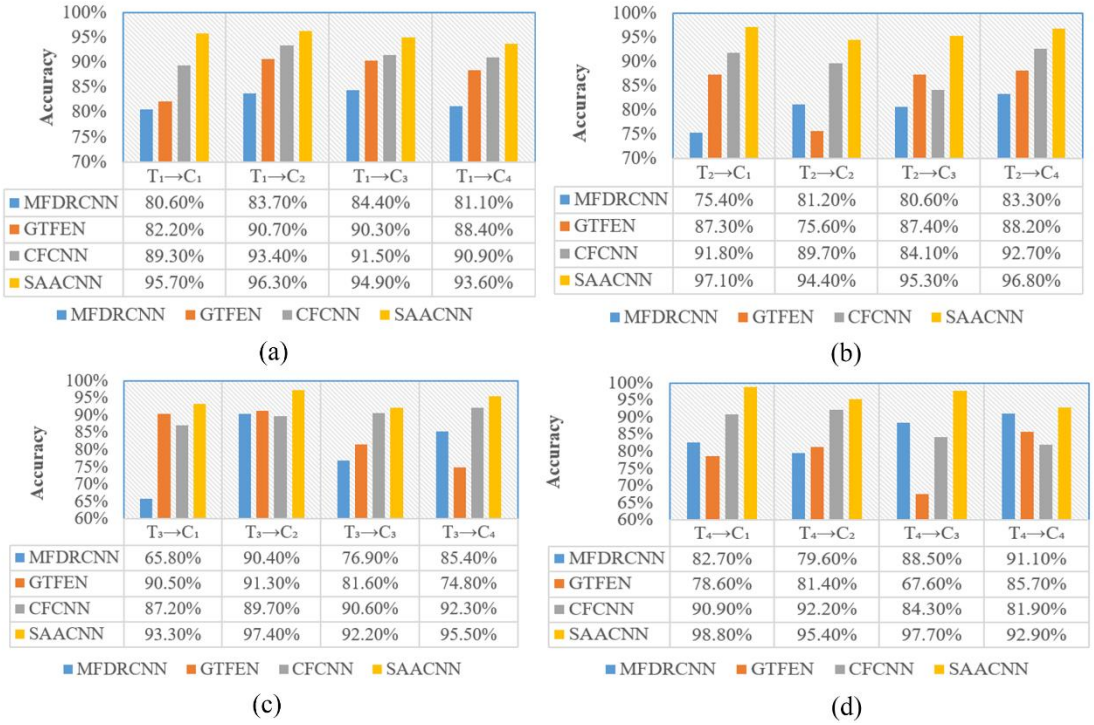
FL1:Fault in Ting Tai Road;**FL2**:Fault in Dai Fuk Street;
FL3:Fault in Dai Fat Road;**FL4**:Fault in Ting Kok Road

(b)

Fig.19 MPCs in different locations: (a) normal MPC;(b) fault MPC

Table 5 The training samples and testing samples setting in different situations

	Training samples					**Testing samples**			
Label	**Normal**	**Number**	**Fault**	**Number**	**Label**	**Normal**	**Number**	**Fault**	**Number**
T_1	NL_1+NL_2+NL_3	400	FL_1+FL_4	100	C_1	NL_4	100	FL_2	100
T_2	NL_1+NL_2+NL_4	400	FL_2+FL_3	100	C_2	NL_3	100	FL_1	100
T_3	NL_1+NL_3+NL_4	400	FL_3+FL_1	100	C_3	NL_2	100	FL_4	100
T_4	NL_2+NL_3+NL_4	400	FL_4+FL_2	100	C_4	NL_1	100	FL_3	100



(a)	$T_1 \rightarrow C_1$	$T_1 \rightarrow C_2$	$T_1 \rightarrow C_3$	$T_1 \rightarrow C_4$
MFDRCNN	80.60%	83.70%	84.40%	81.10%
GTFEN	82.20%	90.70%	90.30%	88.40%
CFCNN	89.30%	93.40%	91.50%	90.90%
SAACNN	95.70%	96.30%	94.90%	93.60%

(b)	$T_2 \rightarrow C_1$	$T_2 \rightarrow C_2$	$T_2 \rightarrow C_3$	$T_2 \rightarrow C_4$
MFDRCNN	75.40%	81.20%	80.60%	83.30%
GTFEN	87.30%	75.60%	87.40%	88.20%
CFCNN	91.80%	89.70%	84.10%	92.70%
SAACNN	97.10%	94.40%	95.30%	96.80%

(c)	$T_3 \rightarrow C_1$	$T_3 \rightarrow C_2$	$T_3 \rightarrow C_3$	$T_3 \rightarrow C_4$
MFDRCNN	65.80%	90.40%	76.90%	85.40%
GTFEN	90.50%	91.30%	81.60%	74.80%
CFCNN	87.20%	89.70%	90.60%	92.30%
SAACNN	93.30%	97.40%	92.20%	95.50%

(d)	$T_4 \rightarrow C_1$	$T_4 \rightarrow C_2$	$T_4 \rightarrow C_3$	$T_4 \rightarrow C_4$
MFDRCNN	82.70%	79.60%	88.50%	91.10%
GTFEN	78.60%	81.40%	67.60%	85.70%
CFCNN	90.90%	92.20%	84.30%	81.90%
SAACNN	98.80%	95.40%	97.70%	92.90%

Fig.19 The generalization comparisons of different networks:(a) Case 1; (b) Case 2; (c) Case 3; (d) Case 4

It can be seen from Fig.20 that when both the training and testing sets change, the accuracy of each network also undergoes alterations. The specific performances are as follows: In Case1, as shown in Fig.20(a), as the testing samples changes, the accuracy of SAACNN remains within the range of 93.60%~96.30%. In Case 2, as shown in Fig.20(b), with changes in testing samples, the accuracy of SAACNN remains stable between 94.40%~97.10%. In Case 3, as shown in Fig.20(c), when the testing samples changes, the accuracy of SAACNN ranges from 92.20%~97.40%. In Case 4, as shown in Fig.20(d), when the testing samples changes, the accuracy of SAACNN maintains at 92.90%~98.80%. Therefore, it can be concluded that there is an inconsistency between the testing set and the training set, the SAACNN model manages to uphold a high accuracy and is less affected by samples variations. The accuracies of the other three networks are lower than that of SAACNN, and the accuracy varies greatly with the change of the samples, possibly because changes in the training set lead to alterations in the optimal structure of each network, resulting in different results. What's more, during the training process , the structure selections of CFCNN, GTFEN and MFDRCNN is manual with an element of subjectivity. Within the limited experience, the network's performance may not reach the optimal level. In contrast, the SAACNN proposed in this paper with the novel loss function model and optimization criteria can adjust the network structure to the optimal level according to the characteristics of the samples set. Therefore, SAACNN exhibits a robust generalization ability.

5 Conclusions

This paper proposes the AEWGAN model for the unbalanced and small size samples enhancement. Compared with the traditional GAN model, it can enhance the correlation with real samples, which improves the quality of the generated samples. For detection, the SAACNN model is developed with a structure adaptive adjustment function, which greatly improves the diagnosis accuracy and generalization ability.The application of this model in MPC fault

detection yields specific conclusions as outlined below.

(1) In terms of sample enhancement capabilities, the AEWGAN model outperforms various other types of GANs, including GAN, WGAN, infoGAN, and CWGAN. The samples generated by AEWGAN proposed are closer to the real samples;

(2) Compared with the vibration in a single direction, the cross-correlation function method is adopted to fuse the data in three directions, which can synthesize the data in different directions to enrich the state information, and has the highest accuracy.

(3) In terms of detection accuracy, compared with CNN, SMOTE-CNN, and AEWGAN-CNN, no matter how the unbalance degree of the dataset changes, the AEWGAN-SAACNN model proposed exhibits the highest accuracy and the excellent stability.

(4) In terms of anti-noise performance, the network model proposed has strong noise resistance, even in the environment with the strongest noise (SNR=-6dB), the average accuracy of proposed approach is 93.5%, which is improved by 23.1 %, 7.8%, 9.6% and 3.3% than that of NOSCNN, MFDRCNN, GTFEN and CFCNN respectively.

(5) In terms of generalization ability, compared with other networks, even the testing set has changed, the network proposed still maintains a high accuracy rate (over 92%), which proves that the network has a strong generalization and feature transfer learning capability.

What's different from the laboratory environment is that the fault type can be set freely in the laboratory, but in the actual environment, the faults of MPC on the road are random, and the health status of most MPCs are unknown (need to be confirmed by the government department). In addition, due to Hong Kong laws, traffic and security restrictions, more data are difficult to be collected, which results in that the current fault types of MPC are not enough. In future, more focus should be put on testing the proposed approaches under the circumstance of more fault types and more MPC, which can further confirm the effectiveness of the proposed approach and guide us to find more reliable conclusions.

Acknowledgments

This research was supported by the Innovation and Technology Fund (ITF) of the HongKong Special Administrative Region, China (Project Ref.: ITS/123/21FP), for the research, authorship and/or publication of this article. It is also supported by the Foreign Expert Project of Henan Province [No. HNGD2023027]. Natural science foundation of Zhongyuan University of Technology [NO. K2023MS020].

References

[1] B.D. Zhou, W.J. Zhao, W.H. Guo, L.C. Li, D.J. Zhang, Q.Z. Mao, Q.Q. Li, Smartphone-based road manhole cover detection and classification, *Autom. Constr*. 140 (2022). https://doi.org/10.1016/j.autcon.2022.104344.

[2] H.S. Zhang, L. Li, X. Liu, Development and test of manhole cover monitoring device using LoRa and accelerometer, *IEEE Trans. Instrum. Meas.* 69 (2020) 2570–2580. https://doi.org/10.1109/TIM.2020.2967854.

[3] Y.T. Yu, H.Y. Guan, Z. Ji, Automated detection of urban road manhole covers using mobile laser scanning data, *IEEE Trans. Intell. Transp. Syst.* 16 (2015). https://doi.org/10.1109/TITS.2015.2413812.

[4] M. Yadav, B. Lohani, S. Goel, Geometric and radiometric constraints-based extraction of urban road manhole covers and their maintenance-related information using mobile laser scanning data, *Geocarto Int*. 37 (2022)

16716–16735. https://doi.org/10.1080/10106049.2022.2115151.

[5] J. Pasquet, T. Desert, O. Bartoli, M. Chaumont, C. Delenne, G. Subsol, M. Derras, N. Chahinian, Detection of manhole covers in high-resolution aerial images of urban areas by combining two methods, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. 9 (2016) 1802–1807. https://doi.org/10.1109/JSTARS.2015.2504401.

[6] G.X. Liu, H.G. Jia, R. Zhang, H.X. Zhang, H.L. Jia, B. Yu, M.Z. Sang, Exploration of subsidence estimation by persistent scatterer InSAR on time series of high resolution terraSAR-X images, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. 4 (2011) 159–170. https://doi.org/10.1109/JSTARS.2010.2067446.

[7] G.Y. Jia, G.J. Han, H.L. Rao, L. Shu, Edge Computing-based intelligent manhole cover management system for smart cities, *IEEE Internet Things J*. 5 (2018) 1648–1656. https://doi.org/10.1109/JIOT.2017.2786349.

[8] S.R. Yang, J.R. Chang, H.M. Lin, K.H. Lin, Identification of manholes beneath pavements using RFID and AR technologies, *J. Test. Eval.* 43 (2015). https://doi.org/10.1520/JTE20140136.

[9] T. Yamaguchi, T. Mizutani, Detection and localization of manhole and joint covers in radar images by support vector machine and hough transform, *Autom. Constr.* 126 (2021) 103651. https://doi.org/10.1016/j.autcon.2021.103651.

[10] N. Kim, K. Kim, Y.K. An, H.J. Lee, J.J. Lee, Deep learning-based underground object detection for urban road pavement, *Int. J. Pavement Eng.* 21 (2020) 1638–1650. https://doi.org/10.1080/10298436.2018.1559317.

[11] L.R.Chen, Y.D. Ma, H.Y. Hu, U.S. Khan. An effective fault diagnosis approach for bearing using stacked de-noising auto-encoder with structure adaptive adjustment[J].*Measurement*.214(2023)112774. https://doi.org/ 10.1016/j.measurement.2023.112774 .

[12] J.G. Jang, C.M. Noh, S.S. Kim, S.C. Shin, S.S. Lee, J.C. Lee, Vibration data feature extraction and deep learning-based preprocessing method for highly accurate motor fault diagnosis, *J. Comput. Des. Eng.* 10 (2023) 204–220. https://doi.org/10.1093/jcde/qwac128.

[13] L.R.Chen,Y.D.Ma,H.Q.Wang,S.J.Wen,L.F.Guo, A novel deep convolutional neural network and its application to fault diagnosis of the squirrel-cage asynchronous motor under noisy environment[J]. *Meas. Sci. Technol.* 34(2023)115113. https://doi.org/10.1088/1361-6501/acea9b.

[14] Y.M. Tang, W.T.Kuo, C.K.M.Lee, Real-time mixed reality (MR) and artificial intelligence (AI) object recognition integration for digital twin in Industry 4.0, *Internet Things,* 23(2023) 100753.

[15] W.Li, Y.M.Tang, Z.Wang,K.M. Yu,S.To, Atrous residual interconnected encoder to attention decoder framework for vertebrae segmentation via 3D volumetric CT images. *Eng. Appl. Artif. Intell.*, 114(2022) 105102. https://doi.org/ 10.1016/j.engappai.2022.105102.

[16] Y.M. Tang, K.Y. Chau, Y.Y. Lau, Z. Zheng, Data-intensive inventory forecasting with artificial intelligence models for cross-border e-commerce service automation. *Appl. Sci.*, 13(2023)3051. https://doi.org/10.3390/app13053051.

[17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, *Commun. ACM*. 63 (2020) 139–144. https://doi.org/10.1145/3422622.

[18] X.C. Liu, T. Zhang, J.W. Zhang, Toward visual quality enhancement of dehazing effect with improved Cycle-GAN, *Neural Comput. Appl*. 35 (2023) 5277–5290. https://doi.org/10.1007/s00521-022-07964-1.

[19] B. Natarajan, R. Elakkiya, Dynamic GAN for high-quality sign language video generation from skeletal poses using generative adversarial networks, *Soft Comput.* 26 (2022) 13153–13175. https://doi.org/10.1007/s00500-022-07014-x.

[20] H.H. Gao, X.R. Zhang, X.J. Gao, F.Y. Li, H.G. Han, ICoT-GAN: Integrated convolutional transformer GAN for rolling bearings fault diagnosis under limited data condition, *IEEE Trans. Instrum. Meas*. 72 (2023). https://doi.org/10.1109/TIM.2023.3271729.

[21] W. Luo, W. Yang, J.L. He, H. Huang, H.D. Chi, J.L. Wu, Y. Shen, Fault diagnosis method based on two-stage GAN for data imbalance, *IEEE Sens. J.* 22 (2022) 21961–21973. https://doi.org/10.1109/JSEN.2022.3211021.

[22] Q.W. Guo, Y.B. Li, Y.J. Liu, S.Y. Gao, Y. Song, Data augmentation for intelligent mechanical fault diagnosis based on local shared multiple-generator GAN, *IEEE Sens. J.* 22 (2022) 9598–9609. https://doi.org/10.1109/JSEN.2022.3163658.

[23] S. Liu, J.L. Chen, C. Qu, R.J. Hou, H.X. Lv, T.Y. Pan, LOSGAN: latent optimized stable GAN for intelligent fault diagnosis with limited data in rotating machinery, *Meas. Sci. Technol.* 32 (2021). https://doi.org/10.1088/1361-6501/abd0c1.

[24] J.G. Fan, X.F. Yuan, Z.M. Miao, Z.H. Sun, X.X. Mei, F.Y. Zhou, Full attention Wasserstein GAN with gradient normalization for fault diagnosis under imbalanced data, *IEEE Trans. Instrum. Meas.* 71 (2022). https://doi.org/10.1109/TIM.2022.3190525.

[25] Z. Meng, Q. Li, D.Y. Sun, W. Cao, F.J. Fan, An intelligent fault diagnosis method of small sample bearing based on improved auxiliary classification generative adversarial network, *IEEE Sens. J.* 22 (2022) 19543–19555. https://doi.org/10.1109/JSEN.2022.3200691.

[26] W. Li, X. Zhong, H.D. Shao, B.P. Cai, X.K. Yang, Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework, *Adv. Eng. Informatics*. 52 (2022). https://doi.org/10.1016/j.aei.2022.101552.

[27] J.Y. Chen, Z.T. Yan, C.Y. Lin, B.Q. Yao, H.J. Ge, Aero-engine high speed bearing fault diagnosis for data imbalance: A sample enhanced diagnostic method based on pre-training WGAN-GP, *Measurement.* 213 (2023). https://doi.org/10.1016/j.measurement.2023.112709.

[28] Q. Sun, F. Peng, X.H. Yu, H.S. Li, Data augmentation strategy for power inverter fault diagnosis based on Wasserstein distance and auxiliary classification generative adversarial network, *Reliab. Eng. Syst. Saf.* 237 (2023). https://doi.org/10.1016/j.ress.2023.109360.

[29] K.C. Shu, S.T. Mao, J.L. Coyle, E. Sejdic, Improving non-invasive aspiration detection with auxiliary classifier Wasserstein generative adversarial networks, *IEEE J. Biomed. Heal. Informatics*. 26 (2022) 1263–1272. https://doi.org/10.1109/JBHI.2021.3106565.

[30] F. Baek, D. Kim, S. Park, H. Kim, S. Lee, Conditional generative adversarial networks with adversarial attack and defense for generative data augmentation, *J. Comput. Civ. Eng.* 36 (2022). https://doi.org/10.1061/(ASCE)CP.1943-5487.0001015.

[31] N.C. Ma, J.K. Wang, J.B. Liu, M.Q.H. Meng, Conditional generative adversarial networks for optimal path planning, *IEEE Trans. Cogn. Dev. Syst.* 14 (2022) 662–671. https://doi.org/10.1109/TCDS.2021.3063273.

[32] Z.C. Wang, H. Xia, J.Y. Zhang, B. Yang, W.Z. Yin, Imbalanced sample fault diagnosis method for rotating machinery in nuclear power plants based on deep convolutional conditional generative adversarial network, *Nucl. Eng. Technol.* 55 (2023) 2096–2106. https://doi.org/10.1016/j.net.2023.02.036.

[33] Y.Z. Kang, L. Chen, N. Jia, W. Wei, J. Deng, H.Z. Qian, A CWGAN-GP-based multi-task learning model for consumer credit scoring, *Expert Syst. Appl.* 206 (2022). https://doi.org/10.1016/j.eswa.2022.117650.

[34] F. Han, S.J. Zhu, Q.H. Ling, H. Han, H.L. Li, X.L. Guo, J.C. Cao, Gene-CWGAN: a data enhancement method for gene expression profile based on improved CWGAN-GP, *Neural Comput. Appl.* 34 (2022) 16325–16339. https://doi.org/10.1007/s00521-022-07417-9.

[35] M.Q. Tran, M.K. Liu, Q. V Tran, T.K. Nguyen, Effective fault diagnosis based on wavelet and convolutional attention neural network for induction motors, *IEEE Trans. Instrum. Meas*. 71 (2022) 3501613. https://doi.org/10.1109/TIM.2021.3139706.

[36] J. Gu, Y.X. Peng, H. Lu, X.D. Chang, G.A. Chen, A novel fault diagnosis method of rotating machinery via VMD, CWT and improved CNN, *Measurement.* 200 (2022) 111635. https://doi.org/10.1016/j.measurement.2022.111635.

[37] H. Liu, Y.L. Tian, C.G. Peng, Z.Q. Wu, Privacy-utility equilibrium data generation based on Wasserstein generative adversarial networks, Inf. Sci. (Ny). 642 (2023). https://doi.org/10.1016/j.ins.2023.119069.

[38] H.D. Shao, J. Lin, L.W. Zhang, D. Galar, U. Kumar, A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance, *Inf. Fusion*. 74 (2021) 65–76. https://doi.org/10.1016/j.inffus.2021.03.008.

[39] G.F. Bin, Z.N. Jiang, X.J. Li, B.S. Dhillon, Weighted multi-sensor data level fusion method of vibration signal based on correlation function, *Chinese J. Mech. Eng.* 24 (2011) 899–904. https://doi.org/10.3901/CJME.2011.05.899.

[40] F.C. Lv, G.L. Liu, Q. Wang, X.Q. Lu, S.F. Lei, S.H. Wang, K. Ma, Pattern recognition of partial discharge in power transformer based on InfoGAN and CNN, *J. Electr. Eng. Technol.* 18 (2023) 829–841. https://doi.org/10.1007/s42835-022-01260-7.

[41] S.Y. Ji, Z.Z. Zhang, S.H. Ying, L.J. Wang, X.B. Zhao, Y. Gao, Kullback-Leibler divergence metric learning, *IEEE Trans. Cybern.* 52 (2022) 2047–2058. https://doi.org/10.1109/TCYB.2020.3008248.

[42] T.A. Terlep, M.R. Bell, T.M. Talavage, D.L. Smith, Euclidean distance approximations from replacement product graphs, *IEEE Trans. Image Process*. 31 (2022) 125–137. https://doi.org/10.1109/TIP.2021.3128319.

[43] P. Peng, J. Wang, NOSCNN: A robust method for fault diagnosis of RV reducer, *Measurement*. 138 (2019)652-658, https://doi.org/10.1016/j.measurement.2019.02.080.

[44] Y.D. Xu, X.A. Yan, B.B. Sun, J.H. Zhai, Z. Liu, Multireceptive field denoising residual convolutional networks for fault diagnosis, *IEEE Trans. Ind. Electron.* 69 (2022) 11686–11696. https://doi.org/10.1109/TIE.2021.3125666.

[45] L.S. Jia, T.W.S. Chow, Y.X. Yuan, GTFE-Net: A gramian time frequency enhancement CNN for bearing fault diagnosis, *Eng. Appl. Artif. Intell.* 119 (2023). https://doi.org/10.1016/j.engappai.2022.105794.

[46] Y.D. Xu, K. Feng, X.A. Yan, R.Q. Yan, Q. Ni, B.B. Sun, Z.H. Lei, Y.C. Zhang, Z. Liu, CFCNN: A novel convolutional fusion framework for collaborative fault identification of rotating machinery, *Inf. Fusion*. 95 (2023)

1–16. https://doi.org/10.1016/j.inffus.2023.02.012.