Intelligent Health Inspection for Road Multi-part Covers Based on Vibration Feature Encoding and Denoising Diffusion Model

Junping Zhong, *Member IEEE*, Yuk Ming Tang, Ka Chun Ng, Kai Leung Yung

***Abstract*—Road multi-part covers (MPCs) are installed to seal the entrance ports of large drains. Due to the long-term impacts of traffic vehicles, MPC damages may occur. While manual inspection is effective, it can lead to traffic disruptions and is inefficient. In this paper, we present a non-invasive approach that analyzes acoustic emissions generated by vehicle-MPC impacts. Specifically, we propose an effective deep learning-based method named VFEDDM (Vibration Feature Encoding and Denoising Diffusion Model), which includes three successive stages: (1) The process of "peak window truncation-> scale normalization-> direction aligned RGB encoding" is proposed to appropriately form the key characteristics of vibrations in different propagation directions. This process ensures robustness against variations in vehicle running conditions and changes in measurement distances. (2) The recent generative AI, denoising diffusion model, is introduced to synthesize high-quality RGB feature images, achieving data augmentation. This can address the model training issue caused by data imbalances between normal and defective data. (3) A deep CNN is constructed and trained by utilizing the augmented RGB image set to learn MPC status-discriminative patterns, which are used to assess the health status of the test MPCs. The effectiveness of VFEDDM is verified in the dataset collected from real MPC sites in Hong Kong. It achieves an accuracy of 0.93 for the test MPCs, and the diagnostic results are well visualized by t-SNE. This would provide support for MPC maintenance decision-making, and significantly improve the inspection efficiency while reducing traffic interference rendered by road closures.**

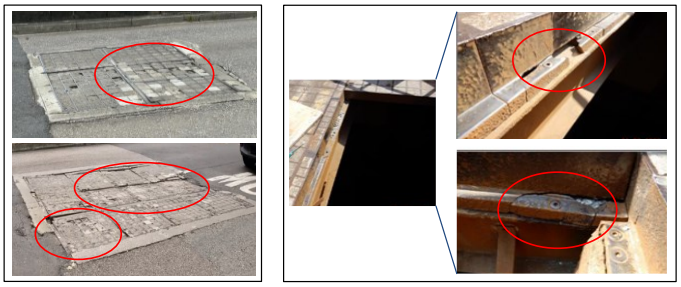
***Index Terms*—Multi-part covers, health inspection, acoustic emission, vibration feature encoding, denoising diffusion model, convolutional neural network**

I. Introduction

THE multi-part cover (MPC) plays a role in covering the entrances to the desilting openings of drainage facilities. In urban areas, there are numerous MPCs installed and some are laid across the main roads with heavy traffic, as shown in Fig. 1(a). Due to the long-term impact or a sudden increase of loading from the vehicles, surface damages and underneath structural metal-frame damages may occur, as shown in Fig. 2(a) and Fig. 2(b) respectively. Currently, workers from the department of drainage services need to open the covers for manually regular inspections and maintenance. However, such opening inspection approach is costly, because it is inefficient and needs to close the traffic road, as shown in Fig. 1(b). Therefore, it is required to develop a non-invasive MPC inspection system to improve efficiency and avoid traffic disruptions.

(a) MPCs on the road (b) Manually inspection with road closure

Fig. 1 MPCs and the costly manually inspection



(a) Surface damages (b) Underneath structure damages

Fig. 2 MPC damages caused by vehicle impacts and environment

For an MPC, heavy impact force would produce obvious acoustic emission. When surface damage or/and internal damage occur, the resulting acoustic emission differs from that of the normal condition. As shown in Fig. 3, the acoustic emission-based approach shows an advantage, it just needs to collect the vibration signals produced by the contraction impacts between the vehicle and MPC from a safe distance, and then diagnose MPC status by signal analyzing algorithm, which will not interfere with the traffic. In real traffic scenarios, different types of vehicles may randomly impact an MPC at different speeds, the sensor distance may change for different MPC sites, these make the collected vibration signals of an MPC various and complex. Therefore, developing an effective signal analyzing algorithm for MPC health inspection is a challenging task.

This research was supported by the Innovation and Technology Fund (ITF) of the Hong Kong Special Administrative Region, China, Project Ref.: ITS/123/21FP. (Corresponding author: Yuk Ming Tang).

Junping Zhong, Yuk Ming Tang, Ka Chun Ng, and Kai Leung Yung are with the Department of Industrial Systems Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China (e-mail: junping.zhong@polyu.edu.hk; yukming.tang@polyu.edu.hk; hadyn1.ng@polyu.edu.hk; kl.yung@polyu.edu.hk).

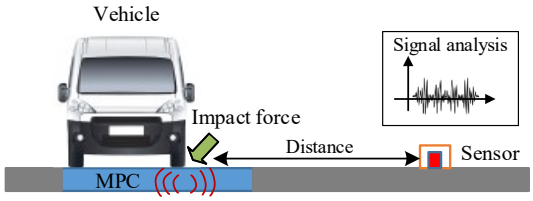


Fig. 3 Acoustic emission-based inspection approach for MPC

In recent years, AI-aid inspection and management technologies have greatly developed and applied in different scenarios [1-3]. In particular, AI techniques integrated with signal and image processing have achieved great success in non-invasive industrial inspection tasks, and they can be classified into two categories. The first category methods [4-7] directly use the one-dimension feature, such as raw time-series, frequency, and spectrum, then send it to the signal processing AI diagnosis models. Commonly, the full length or a segment of the raw time-series is sent to neural networks such as RNN, LSTM, 1D-CNN, and DBN [8] for status label prediction. For multiple sensors, the raw time-series can be combined as a three-dimension point cloud format and processed by a network like PointNet [9]. Several multiple vibration data fusion techniques [10, 11] that combine different data at the feature-level or decision-level are adopted for mechanical fault diagnosis. The second category methods [12-15] transform the raw time-series signal into a visual image, and then send it to the image processing AI diagnosis models. The time-frequency transformation, such as STFT [14] and CWT [15], are commonly applied to extract the time-frequency feature images, which are used for status classification by 2D-CNN [16, 17] network or Vision-transformer [18]. **While the above AI-based diagnostic methods follow the same pipeline of "Vibration Feature Extraction + Status Classification", the feature extraction technique and classification network need to be specially tailored for different scenes and inspection tasks**.

Another problem that may occur in an AI-based diagnostic approach is the imbalance of data for model training [19]. In many practical scenarios, the available defect samples are limited compared to the normal samples. While using imbalanced data for training may be effective by tailored optimization tricks [20, 21], it has an inevitable shortage in generalization ability. At present, the GAN (Generative Adversarial Networks) [22] and its variants [23, 24] are the most widely used techniques for data synthetic, achieving data augmentation. However, the mode collapse could make the training unstable, which leads to unsatisfactory performance. Recently, the diffusion model [25, 26], an advanced generative AI technique, has achieved remarkable success in image synthesis [27-29]. Inspired by the theory of nonequilibrium thermodynamics [30] and theoretically explained by score-based generative models [31, 32], diffusion models view image reconstruction as the process of image degradation and denoising restoration. This process can be repeated multiple times to generate diverse images with realistic textures and details. Several diffusion models [33, 34] based on DDPM (Denoising Diffusion Probabilistic Model), have been developed for intelligent defect detection for industrial equipment, demonstrating exceptional performance.

At present, there is scarcely any relevant literature on road MPC inspection based on acoustic emissions. Considering the intricate measurement environment in traffic and the challenges associated with collecting defect data, MPC inspection faces the following hurdles:

1) The variability in vehicle running conditions (such as weight and speed) and the changing sensor-MPC distances result in diverse and intricate collected signals, posing a challenge in their handling.

2) The available defect samples are restricted in comparison to the normal samples, which creates the data imbalance issue for diagnostic model training.

3) Lack of an effective intelligent diagnosis framework.

In this paper, we propose VFEDDM (Vibration Feature Encoding and Denoising Diffusion Model) to tackle the above challenges. **The contributions are summarized as follows.**

1. A processing scheme "Peak window truncation-> Scale normalization-> Direction aligned RGB encoding" is proposed to appropriately form the key characteristics of vibrations in different propagation directions. The formed RGB feature image is robust against variations in vehicle running conditions and changes in measurement distances.

2. The recent generative AI, denoising diffusion probabilistic model, is introduced to synthesize high-quality defect RGB feature images for CNN diagnostic model training, which can address the data imbalance issue and enhance diagnostic performance.

3. Experimental validation on the test MPCs verified the effectiveness of the proposed VFEDDM, showing promise for practical use and helping in the prevention of traffic disruptions caused by manual inspections. To the best of our knowledge, this is the first work presenting an intelligent health diagnosis framework for road MPC inspection.

The rest of this paper is structured as follows. Section II introduces the experimental setups for data collection. Section III presents the details of the proposed method VFEDDM. It gives the processes of the RBG image encoding for vibration features, the diffusion model for data augmentation, and the deep CNN model for health status patterns mining. In Section IV, the details of the experiments and result analysis in a real MPC dataset are given. Finally, Section VI concludes and gives future work.

II. Experimental Setup for Data Collection

The experimental setup for data collection is shown in Fig. 4, which mainly includes the triggering unit, sensor unit, data amplifier & storage unit, and the auxiliary camera. The *d* is the distance between the MPC and the sensor. When a vehicle passes by, the triggering unit triggers the unit of data amplifier & storage to simultaneously record the X, Y, Z direction vibration signals, which are regarded as one test sample. For an MPC, numerous samples could be collected and used for diagnosis algorithm analysis.

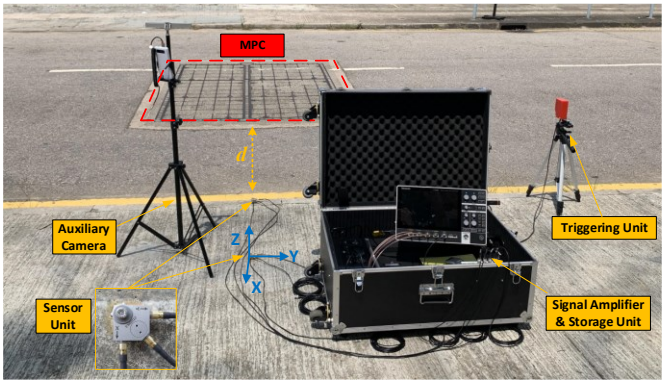


Fig. 4 On-site experimental setups for data collection

- *Sensor Unit*: It is equipped with a CT1020LS 3-axis accelerometer to capture vibration signals in horizontal directions (X and Y) and vertical direction (Z) simultaneously, which can comprehensively capture the acoustic waves produced from the impact between vehicle and MPC.
- *Signal Amplifier & Storage*: A Tektronix MSO 24 Bench Portable Oscilloscope for signal conversion. The sample rate is set to 100kHz, record 2s with 200,000 sample points in total.
- *Distance d*: The sensor is only allowed to be installed along the roadside so as not to interrupt the traffic. Therefore, the *d* usually varies with road scenes.

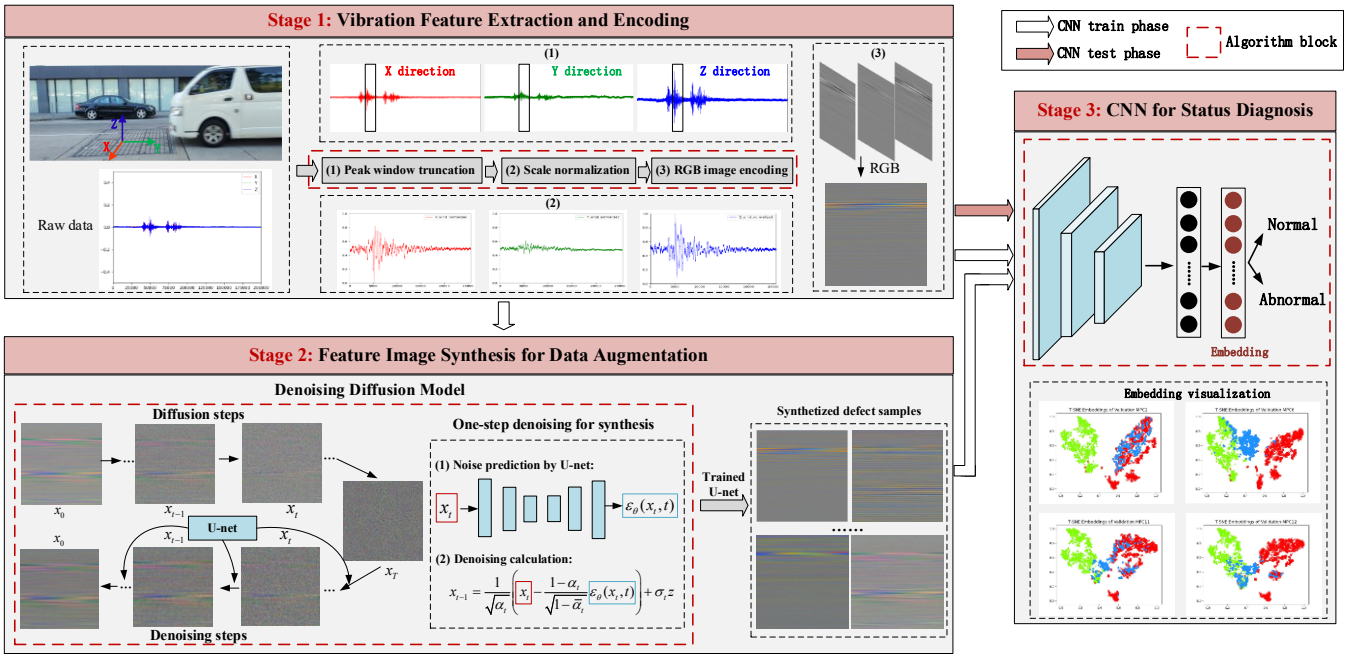


Fig. 5 Overview of the proposed method VFEDDM

III. Methodology

A. Overall Framework of VFEDDM

As shown in Fig. 5, the overall framework of the proposed VFEDDM mainly comprises three successive stages: *Vibration feature extraction and encoding*, *Feature image synthesis for data augmentation*, and *Health status diagnosis*. The overall workflow is described as follows.

Stage 1: Extract the key time-series features of vibration signals in X, Y, and Z directions respectively, and then encode them as a directions-aligned RGB image. Specifically, 1) Segment the strongest impact signal section from the whole complex signal, which concentrates on the main response features of the impact. 2) Normalize the segmented signals to address the scale issue caused by vehicle weight and the measured distance. 3) Convert the three normalized signals into a directions-aligned RGB image. These form a robust input for mining status features in the health diagnosis model.

Stage 2: Based on the encoded RGB feature images, a generative AI called the denoising diffusion model is trained for defective RGB feature image synthetic, and then the synthesized images are used for CNN diagnostic model training in a supervised learning way. The data augmentation makes the CNN training under balanced data, which can address the issue of lacking defective samples and enhance the diagnosis performance.

Stage 3: A deep CNN is constructed to further process the encoded RGB feature image. The CNN is trained by using the real and synthesized RGB feature images, mining the status-discriminative patterns on the training dataset. In the test phase, for an MPC, the trained CNN predicts the status labels for all acoustic emission samples of this MPC, and then assigns the dominant status label as its final diagnostic result.

B. Vibration Feature Extraction and Encoding

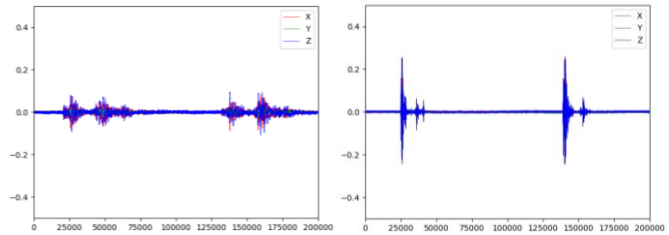
The response of MPC-vehicle collision is greatly influenced by the vehicle running condition and the measurement environment. Several factors, as shown in Table I, make the collected raw data various and complex. They are primarily summarized in the following two aspects.

Table I. Factors influencing the vibration waveform properties

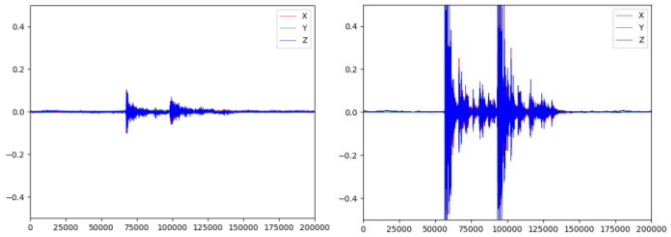
Influence variable	Vehicle weight	Vehicle speed	Sensor-MPC distance *d*
Peak position	--	✓	--
Duration	✓	✓	--
Scale	✓	✓	✓

First, even for the same MPC, different vehicles would pass over it in live traffic, and different vehicle running conditions (vehicle weight and speed) will greatly influence the collision waveform properties (such as *peak position*, *duration*, and *scale*), as shown in Fig. 6(a).

Second, even for the same vehicle running condition, different measurement environments (sensor-MPC distance *d*) will greatly influence the collision waveform *scale*, as shown in Fig. 6(b).



(a) Waveform properties (peak position, duration, and scale) difference caused by vehicle conditions



(b) Waveform scale difference caused by sensor-MPC distance *d*

Fig. 6 Various and complex raw vibration signals collected from the same MPC

In addition, vibrations in different directions (horizontal and vertical) should be appropriately integrated for useful feature mining, which requires an effective means to handle.

To address these issues, we focus on the waveform section around the strongest impact of an X-Y-Z signal, and convert the truncated and normalized signal into a direction-aligned RGB image for CNN feature mining. We propose the processing scheme "**Peak window truncation** -> **Scale normalization** -> **Aligned RGB image encoding**", as described below:

1) **Peak window truncation**. Though the raw vibration data are various and complex. However, the time-series responses primarily concentrated in a limited peak band, **we tend to observe the signal band around the impact moments, because useful vibration information comes from the impacts.** When a vehicle passes the MPC, one or two times impacts may happen, but we tend to focus on the more obvious one. Therefore, for any collected waveform, we find the index t^* with the max amplitude as follows.

$$t^* = \arg\max_t \{X(t), Y(t), Z(t)\} \tag{1}$$

This index moment is considered the strongest impact happens, and the waveform $[t^* - rT, t^* + (1-r)T]$ is truncated as a vibration signal in a duration T that includes the collision characteristics, as shown in Fig. 7. The parameter $r \in (0,1)$ can control the proportion of damping time in T, it is empirically set 0.2.

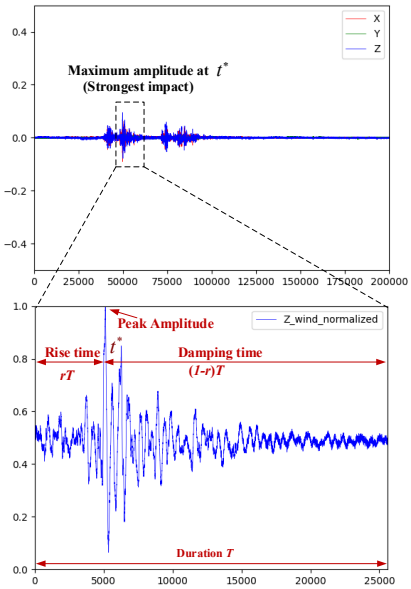


Fig. 7 Peak window truncation operation

2) **Scale normalization**. For each truncated X-Y-Z waveform, the scale normalization is performed to address the scale difference caused by vehicle weight and sensor-MPC distance *d*. Take the X channel as an example, the scale normalization is as follows.

$$X(t_{wind})_{normalized} = \frac{X(t_{wind}) - \min\{X(t_{wind}), Y(t_{wind}), Z(t_{wind})\}}{\max\{X(t^*), Y(t^*), Z(t^*)\} - \min\{X(t_{wind}), Y(t_{wind}), Z(t_{wind})\}} \tag{2}$$

The Fig. 8 shows the scale normalized X-Y-Z waveforms of the Fig.7 case.

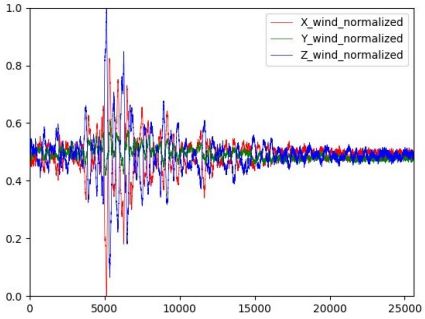
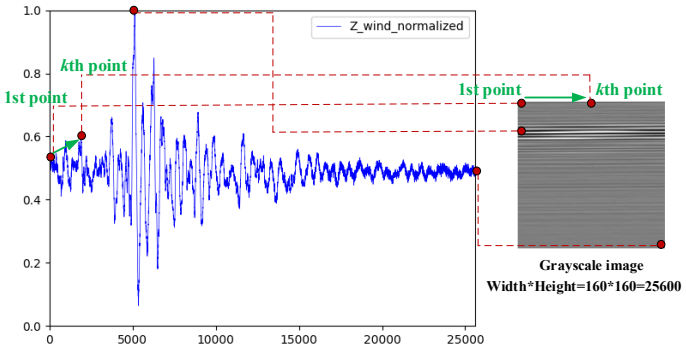
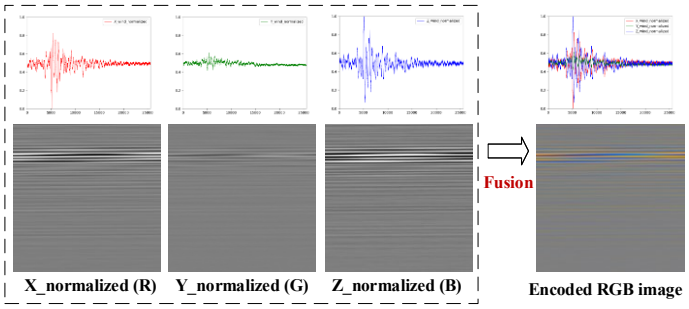


Fig. 8 Scale normalized waveforms of the truncated X-Y-Z signal

3) **Aligned RGB image encoding**. The recent work [35] found visual data processing paradigms like deep neural networks are effective for time series forecasting. In this paper, we integrate different direction signals by combining the truncated and normalized X, Y and Z aligned into an RGB image, as shown in Fig. 9. It encodes X, Y, and Z direction signals into an appropriate sample for a deep neural network to ming the MPC status differences in a supervised data-driven approach.



(a) Encoding the singel channel into a grayscale image



(b) Channels aligned fusion as a RGB image

Fig. 9 Encoding multi-channel vibration signals into RGB image

C. Diffusion Model for Feature Image Augmentation

Due to the lack of defect samples for AI model training, the data synthesis technique is adopted for data augmentation. The recent generative AI, denoising diffusion probabilistic model [25, 26], is introduced for high-quality defect RGB feature image synthetic. The paradigm of the proposed diffusion model is described as follows:

Given an image x_0 that is sampled from a real data distribution $q(x)$, namely $x_0 \sim q(x)$. The forward diffusion process is defined as a Markov chain that gradually adds Gaussian noise to the image x_0 over T timesteps, producing a sequence of noised images x_1, x_2, ..., x_T. The step sizes are controlled by a noise variance schedule $\{\beta_t \in (0,1)\}_{t=1}^{T}$. Mathematically, each diffusion step and the entire diffusion process can be expressed as follows, respectively.

$$q(x_t \mid x_{t-1}) = N(x_t;\ \sqrt{1-\beta_t}x_{t-1},\ \beta_t\mathbf{I}) \tag{3}$$

$$q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t \mid x_{t-1}) \tag{4}$$

Where N represents the Gaussian distribution, and $\mathbf{I}$ is the unit standard deviation of the Gaussian distribution.

The noise variance factors $\{\beta_t \in (0,1)\}_{t=1}^{T}$ are a sequence of positive noise scales, and satisfied $\beta_1 < \beta_2 < ... < \beta_t < ... < \beta_T$, which help to accelerate the diffusion process. On the other hand, the step-by-step operation is inflexible, the diffusion latent result x_t can be directly obtained by merging multiple Gaussian results from x_0 according to the additivity property of Gaussian distribution. Thus, let define $\alpha_t = 1-\beta_t$, then the Formula (4) can be updated as follows

$$q(x_t \mid x_0) = N(x_t;\ \sqrt{\alpha_t}x_0,\ (1-\alpha_t)\mathbf{I}) \tag{5}$$

and $\bar{\alpha}_t = \prod_{i=1}^{t}(1-\beta_i)$. Therefore, x_t can be expressed as a linear combination of x_0 and ϵ.

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \tag{6}$$

Given a RGB image of defect sample x_0, during the diffusion process, it gradually loses its distinguishable texture features. Fig. 10 shows the diffusion process for an encoded RGB image of an abnormal vibration signal.

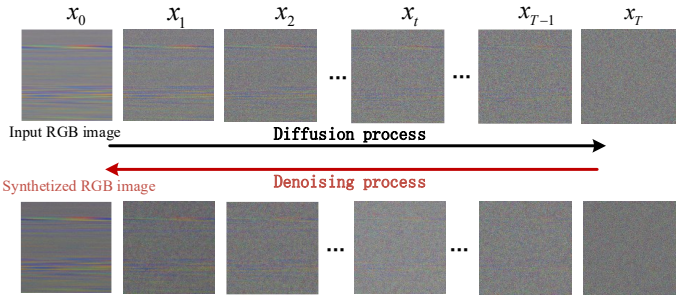


Fig. 10 Diffusion model for RGB feature image synthetic

Though it is easy to corrupt the image, recovering the corrupted image x_T to x_0 is intractable because we cannot easily estimate $q(x_{t-1} \mid x_t)$ that needs to know the distribution of the entire dataset [24]. Fortunately, a neural network p_θ can be used to approximate this conditional probability, then each reverse step can be expressed as follows

$$p_\theta(x_{t-1} \mid x_t) = N(x_{t-1};\ \mu_\theta(x_t,t), \sigma_t^2\mathbf{I}) \tag{7}$$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t)\right) + \sigma_t\mathbf{z} \tag{8}$$

Where $p(x_T) = N(x_T;0,\mathrm{I})$ is a standard Gaussian. Here, a deep neural network U-Net [33, 34] is used to predict the noise $\epsilon_\theta(x_t,t)$, its specific achitecture setting is shown in Fig.11. The U-Net can be trained by optimizing the following objective:

$$\begin{aligned} &\arg\min_\theta D_{\mathrm{KL}}(q(x_{t-1} \mid x_t,\ x_0) \,\|\, p_\theta(x_{t-1} \mid x_t)) \\ &= \arg\min_\theta D_{\mathrm{KL}}(N(x_{t-1};\ \mu_q,\ \Sigma_q(t)) \,\|\, N(x_{t-1};\ \mu_\theta,\ \Sigma_q(t))) \\ &= \arg\min_\theta \frac{1}{2\sigma_q^2(t)}\frac{(1-\alpha_t)^2}{(1-\bar{\alpha}_t)\alpha_t}\left[\left\|\epsilon_0 - \hat{\epsilon}_\theta(x_t,\ t)\right\|_2^2\right] \end{aligned} \tag{9}$$

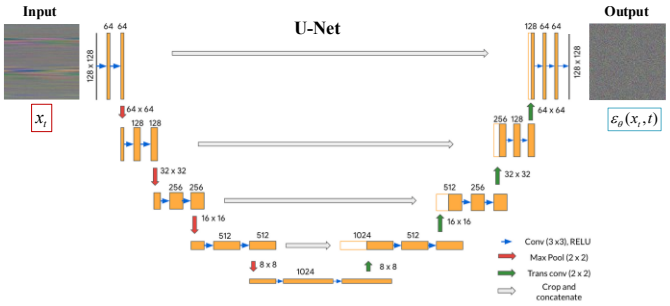


Fig. 11 Diffusion model for RGB feature image synthetic

Optimizing the above formula can force the U-net to predict a standard deviation $\hat{\epsilon}_\theta(x_t,t)$ close to the standard deviation label $\epsilon_0(x_t,t)$, which can be obtained from the pre-defined noises during the diffusion process. Therefore, each denoising step refines the noise-polluted image, allowing the U-net to learn the underlying textures and structures of the training images, which makes the generated images relatively consistent with the patterns of training images. The training process of denoising diffusion model is presented as Algorithm 1.

Algorithm 1 Training process of the denoising diffusion model

1: **Repeat**
2: $x_0 \sim q(x_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
4: $\epsilon \sim N(0, \mathbf{I})$
5: Take gradient descent operation by
6: $\nabla_\theta \left\| \epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t) \right\|^2$
7: **Until** converged

D. Deep CNN for Health Status Diagnosis

To furtherly identify a problematic MPC based on the encoded RGB images of time-series signals, a deep convolutional neural network is constructed and predicts its status label, as shown in Fig. 12. In the training phase, the normal and abnormal images whose health status labels have been manually confirmed are used to train the deep CNN by optimizing the following cross entropy loss function, learning the status patterns (normal and abnormal).

$$H(y,\hat{y}) = -\frac{1}{N}\sum_{1}^{N} y_i \log(\hat{y}_i) \tag{10}$$

Where y is a true label (Normal is 0 and Abnormal is 1), $\hat{y}$ is the predicted label probability, N is the batch size number of model training.

During the test phase, the trained CNN extracts embedding features from the input RGB image and uses it for status label prediction. We utilize the ResNet18 structure [17] as the CNN, as shown in Fig. 11. The embedding is the vector obtained in the last layer, and its similarity with the learned status patterns can be visualized by t-SNE [36].

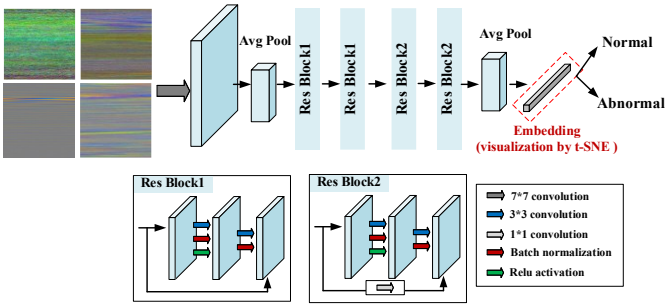


Fig. 12 Deep convolutional neural network for MPC health diagnosis

In diagnosing a single MPC, as vehicles in live traffic pass over it, a series of vibration samples are gathered. Instead of relying on the predicted label of a single sample, the final diagnostic outcome is determined by the predominant label among the predicted status labels from a pool of data samples, This consolidated result can be provided to the DSD department for maintenance decision-making.

IV. Experiment Results and Analysis

A. Dataset and Training Settings

We use the experimental setup described in Section II to collect data on the live traffic in Hong Kong. In our dataset, 29 MPCs located in different road environments are used, and their health status is checked by the DSD (Drainage Services Department). Among them, 14 MPCs including 1,047 normal and 1,113 **(613 real and 500 syntheses)** abnormal samples are used for training, and 14 test MPCs are used for model effectiveness validation, as shown in Table II. These MPCs in the validation can validate the generalization ability of the proposed VFEDDM.

Table II. Train dataset and validation dataset

	MPC number	Normal number	Abnormal number
Train	15	1047	1113
Validation	14	2348	646

For the training of the denoising diffusion model, the hyperparameters are set as follows: learning rate is 0.0001, weight decay is 0, batch size is 4, and iteration number is 60,000. The hyperparameters of deep CNN model training: learning rate is 0.03, weight decay 0.00003, batch size is 512, and iteration epoch is 2,000. The software environment is as follows: Linux Centos 7.9, PyTorch 1.7.1, CUDA 11.2, and NVIDIA GPU RTX 3090.

B. Effectiveness Validation of VFEDDM

(1) Validation results

The widely used evaluation metrics *Accuracy* and *F1-score* are adopted. The *Accuracy* is the proportion of correctly classified items to the total test number, while the *F1-score* is the harmonic average of precision and recall for abnormal samples.

The validation results of VFEDDM for all 14 test MPC cases are shown in Table III. It shows all abnormal MPCs (MPC1,

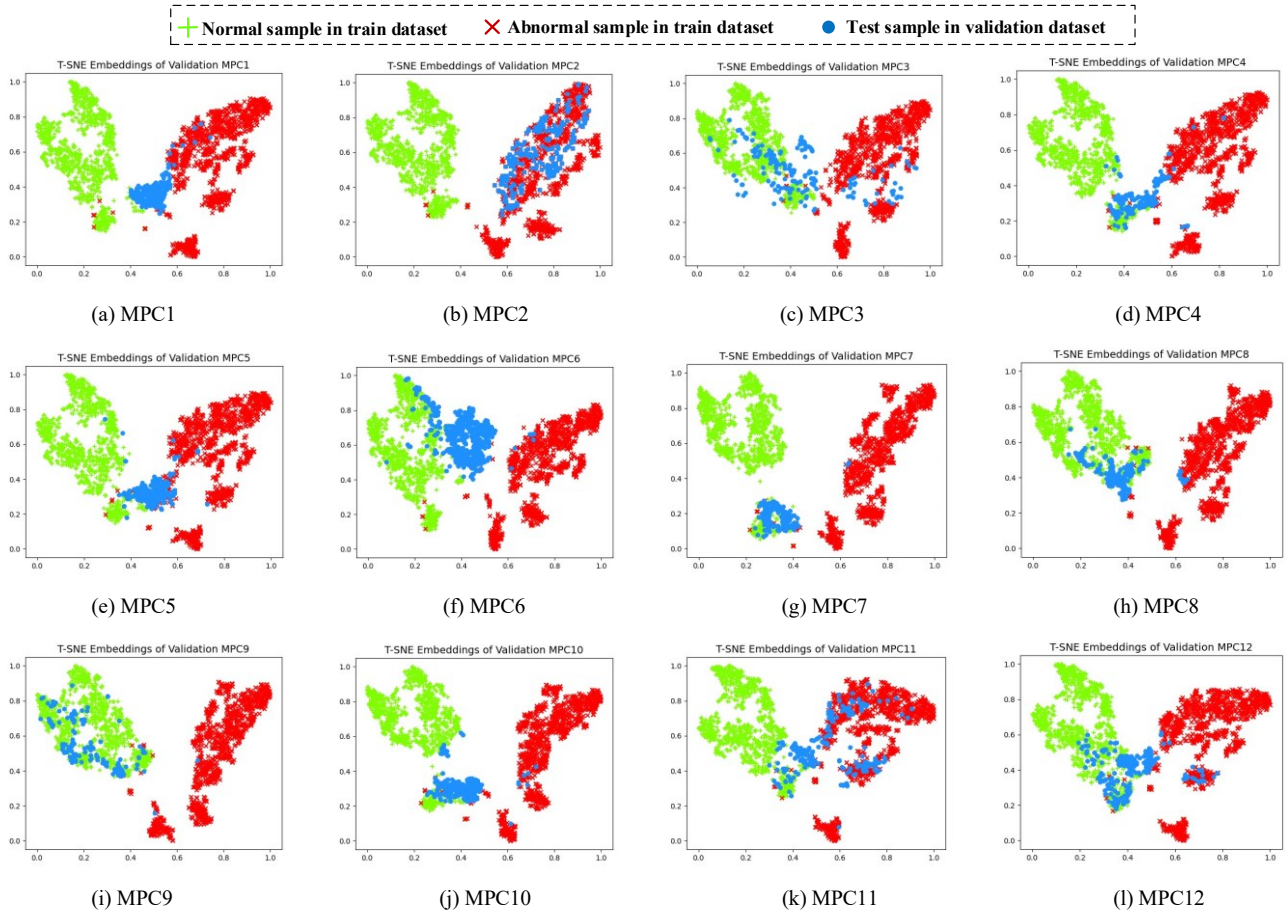
MPC2, MPC5) are correctly diagnosed. For the rest 11 normal MPCs, only the (MPC11) is mistakenly predicted as abnormal. The overall diagnostic *Accuracy* is 0.93 and the *F1-score* is 0.86. Especially, a high *F1-score* shows the proposed model is capable of identifying the abnormal MPCs, meanwhile has good performance for normal MPCS. Finding out the abnormals is prior to misjudging the normals, which is more acceptable for DSD maintenance.

To support the validation results obtained in Table III, the t-SNE [36] is utilized to visualize the distribution of feature embeddings (vectors) extracted by the trained CNN model for each test MPC, as shown in Fig. 13. In these figures, the green and red points represent Normal and Abnormal of training samples respectively, while the blue points are validation samples. It can be seen that from Figs. 13(a)(b)(e) that most of the validated samples are close to the red cluster **(Abnormal)**, thus the MPC1, MPC2, MPC5, and MPC11 are given abnormal labels. Figs. 13(c~d)(f~j)(l~n) show most of the validated samples are close to the green cluster **(Normal)**, thus these MPCs are given normal labels. The t-SNE results show the distributions of feature embeddings are consistent with the obtained results in Table II, which gives interpretability of VFEDDM for the status label prediction.

Table III. Validation results of the proposed VFEDDM in testset

Test MPCs	**MPC1**	**MPC2**	**MPC3**	**MPC4**	**MPC5**	**MPC6**	**MPC7**
True status label	Abnormal	Abnormal	Normal	Normal	Abnormal	Normal	Normal
VFEDDM label	Abnormal	Abnormal	Normal	Normal	Abnormal	Normal	Normal

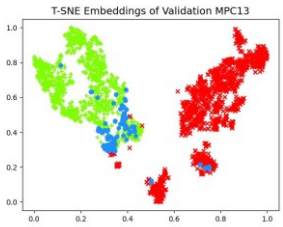
Test MPCs	**MPC8**	**MPC9**	**MPC10**	**MPC11**	**MPC12**	**MPC13**	**MPC14**
True status label	Normal	Normal	Normal	Normal	Normal	Normal	Normal
VFEDDM label	Normal	Normal	Normal	Abnormal	Normal	Normal	Normal



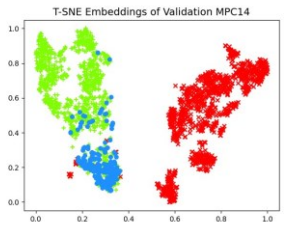
(a) MPC1 (b) MPC2 (c) MPC3 (d) MPC4

(e) MPC5 (f) MPC6 (g) MPC7 (h) MPC8

(i) MPC9 (j) MPC10 (k) MPC11 (l) MPC12



(m) MPC13



(n) MPC14

Fig. 13 t-SNE embedding distributions of samples for each validation MPC

(2) Effect of window truncation size T and ratio r

As the raw vibration signals are irregular, our method only focuses on a segment of the signal that is around the maximum response, which aims to capture the rising and damping features around the maximum response (see Fig. 7). Therefore, the selection of window size T and ratio parameter r will affect the final performance. In the experiment, r is initially fixed at 0.2 empirically (as damping time is commonly considered much longer than the rising time), meanwhile T is set increasingly as 10,000 (image size 100×100), 16,900 (image size 130×130), 25,600 (image size 160×160), 32,400 (image size 180×180), 40,000 (image size 200×200), in both training and validation processes, the results are shown in Table IV. The optimal T=160×160 is selected. Furtherly, T is fixed meanwhile r is set increasingly as 0.1, 0.2, 0.3, and 0.4 in both training and validation processes, the obtained results are shown in Table V. Based on these experiment results, we select the optimal parameter configurations for T=160 and r=0.2.

Table IV. Evaluation results of different r

VFEDDM method	F1-score	Accuracy
RGB_CNN (0.1 & 0.9)	0.67	0.79
RGB_CNN (0.2 & 0.8)	**0.86**	**0.93**
RGB_CNN (0.3 & 0.7)	0.75	0.86
RGB_CNN (0.4 & 0.6)	0.60	0.71

Table V. Evaluation results of different duration T

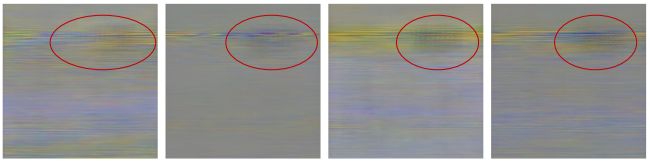
VFEDDM method	F1-score	Accuracy
RGB_CNN (100×100)	0.50	0.86
RGB_CNN (130×130)	0.67	0.79
RGB_CNN (160×160)	**0.86**	**0.93**
RGB_CNN (180×180)	0.80	0.79
RGB_CNN (200×200)	0.50	0.71

(3) Effect of data augmentation

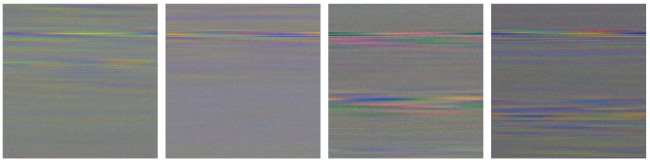
To verify the effect of data augmentation, the VFEDDM without data augmentation and the widely used DCGAN are compared with the diffusion model in validation MPC dataset. The obtained results are presented in Table VI, compared with the model without data augmentation, diffusion model can greatly improve the final diagnostic performance. In addition, diffusion model has higher *accuracy* and *F1-score* than DCGAN. Fig. 14(a) and Fig. 14(b) show some synthesized RGB images of defect samples by DCGAN and diffusion model, respectively. It can be seen there are blur and distortion in the image regions marked in red circles. The diffusion model synthetizes higher-quality RGB images for data augmentation.

Table VI. Evaluation results of different data augmentation methods

VFEDDM	F1-score	Accuracy
Without data augmentation	0.50	0.86
Data augmentation by DCGAN	0.80	0.79
Data augmentation by Diffusion Model	**0.86**	**0.93**



(a) Synthetized images of defect samples by DCGAN



(b) Synthetized images of defect samples by diffusion model

Fig. 14 RGB images synthesis of defect samples by DCGAN and diffusion model

C. Comparison with Other Approaches

The VFEDDM focuses on the specific window-sized time-series characteristic. To further compare it with other potential feature extraction methods that apply on the whole signal, we investigate methods that integrate time-frequency features or original signature features with a deep learning model. Here, four extra diagnosis methods named STFT_CNN [14], CWT_CNN [15], PointNet [9], and DBN [8] are chosen for comparison. The STFT_CNN and CWT_CNN integrate the STFT (Short-Time Fourier Transform) image and CWT (Continuous Wavelet Transform) image with deep CNN, respectively. While the PointNet and DBN (Deep Belief Network) use original signature features. The specific details of the compared approaches and implementations are described below.

- STFT_CNN

Perform the Short-Time Fourier Transform on the X, Y, and Z signals individually, combining the resulting three time-frequency maps into an RGB image. Subsequently, feed the RGB image into a CNN model that is trained with dataset augmented by a diffusion model. Implementation details: Use

the Hamming window, sample rate 100k, nfft 2,560, and overlap 1,250 parameter settings for STFT. The architectures and training parameters of diffusion model and CNN follow those outlined in Section III and Section IV.A.

- CWT_CNN

Perform the Continuous Wavelet Transform on the X, Y, and Z signals individually, combining the resulting three time-frequency maps into an RGB image. Subsequently, feed the RGB image into a CNN model that is trained with dataset augmented by diffusion model. Implementation details: Use the sample rate 100k and wavelet function "sym7" settings for CWT. The architectures and training parameters of diffusion model and CNN follow those outlined in Section III and Section IV.A.

- PointNet

Combine the orignal X, Y and Z vibration signals as into a three-dimensional point cloud data. In the point cloud space, each point (X, Y, Z) represents the normalized amplitude values of X, Y and Z. The combined point cloud (X, Y, Z) keeps the original signature information, and can be directly input to the 3D classification deep learning model PointNet. The training parameter setting of PoineNet follows [9].

- DBN

Apply a DBN that contains 6 Restricted Boltzman Machine layers to process the original X, Y, and Z vibration signals, and directly classify the input vibration signals. DBN training parameters: learning rate 0.01, iteration epoch 200, batch size 64.

Table VII. Comparison results of different methods for the test MPCs

Validation MPC ID	True status label	DBN	PointNet	CWT_CNN	STFT_CNN	VFEDDM
MPC1	Abnormal	Abnormal	Abnormal	Abnormal	Abnormal	Abnormal
MPC2	Abnormal	Abnormal	Abnormal	Abnormal	Abnormal	Abnormal
MPC3	Normal	Abnormal	Normal	Abnormal	Normal	Normal
MPC4	Normal	Abnormal	Normal	Normal	Normal	Normal
MPC5	Abnormal	Abnormal	Normal	Normal	Abnormal	Abnormal
MPC6	Normal	Normal	Normal	Normal	Normal	Normal
MPC7	Normal	Normal	Abnormal	Normal	Abnormal	Normal
MPC8	Normal	Abnormal	Normal	Abnormal	Normal	Normal
MPC9	Normal	Normal	Abnormal	Normal	Normal	Normal
MPC10	Normal	Abnormal	Normal	Normal	Normal	Normal
MPC11	Normal	Abnormal	Normal	Abnormal	Abnormal	Abnormal
MPC12	Normal	Abnormal	Normal	Abnormal	Normal	Normal
MPC13	Normal	Abnormal	Normal	Normal	Normal	Normal
MPC14	Normal	Normal	Abnormal	Normal	Abnormal	Normal
F1-score		0.46	0.50	0.44	0.67	**0.86**
Accuracy		0.43	0.71	0.62	0.79	**0.93**

The evaluation results of all the compared methods are presented in Table VII. The results show the VFEDDM has the highest *Accuracy* 0.93 and *F1-score* 0.86. It seems both the time-frequency feature-based methods STFT_CNN, CWT_CNN and the original signature features-based methods PointNet, DBN are not effective for the unseen MPCs in the validation dataset. We conjecture that these methods are not capable of fitting the intricate collected data that are greatly affected by the variability in vehicle running conditions and sensor-MPC distances. While the proposed scheme "Peak window truncation-> Scale normalization-> Direction aligned RGB encoding" focuses on the key characteristics of vibrations, uniforms the various signal scales, and appropriately forms RGB feature images for status patterns mining, making the VFEDDM more robust against variations in vehicle running conditions and measurement distances. The comparison results show the proposed VFEDDM performs state-of-the-art for urban road MPC health diagnosis.

V. Conclusions

This work trends to diagnose the healthy status of multi-part covers in urban roads by analyzing acoustic emissions of vehicle-MPC impact. An effective method VFEDDM based on vibration feature encoding and generative AI technology is proposed. It can address several existing issues in practice. First, an RGB image encoding method is proposed to extract robust features from the vehicle-MPC impact, it can formulate effective features under complex traffic conditions and measurement distances. Second, a denoising diffusion model is proposed to synthesize high-quality RGB feature images for data augmentation, this can address the data imbalance issue in deep learning model training. Third, a deep CNN is constructed and trained by utilizing the augmented RGB image set to learn MPC status-discriminative patterns, which are used to assess the health status of the test MPCs. The effectiveness of VFEDDM is verified by applying MPC samples in validation

experiments, the obtained results show its promise to improve MPC inspection efficiency while reducing traffic interference rendered by road closures.

For future research, the following limitations warrant further exploration:

1) Given the challenge of acquiring defect samples, the current method is limited to distinguishing between normal and defective items. As the number of negative samples grows, incorporating the data augmentation techniques suggested by VFEDDM will be crucial for enhancing the model to differentiate between defect types and severity.

2) Despite data augmentation, acquiring different defect data remains difficult, requiring manual opening for data labeling, which poses challenges for model updating and upgrading.

3) In places with low traffic flow, due to the limited data acquisition, it is difficult to detect or detect in a timely manner.

ACKNOWLEDGEMENT

This research was supported by the Innovation and Technology Fund (ITF) of the Hong Kong Special Administrative Region, China (Project Ref.: ITS/123/21FP), for the research, authorship and/or publication of this article.

REFERENCES

[1] Ameri, Rasoul, Chung-Chian Hsu, and Shahab S. Band. "A systematic review of deep learning approaches for surface defect detection in industrial applications," *Engineering Applications of Artificial Intelligence*, 130, pp. 107717, 2024.

[2] Geda, M. W., Tang, Y. M., & Lee, C. K. M. "Applications of artificial intelligence in orthopaedic surgery: A systematic review and meta-analysis," *Engineering Applications of Artificial Intelligence*, 133, 108326, 2024.

[3] Lam, H. Y., Tang, V., Wu, C. H., & Cho, V. "A multi-criteria intelligence aid approach to selecting strategic key opinion leaders in digital business management," *Journal of Innovation & Knowledge*, 9(3), 100502, 2024.

[4] Zollanvari, Amin, Kassymzhomart Kunanbayev, Saeid Akhavan Bitaghsir, and Mehdi Bagheri. "Transformer fault prognosis using deep recurrent neural network over vibration signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021.

[5] H. Liu, H. Zhao, J. Wang, S. Yuan and W. Feng, "LSTM-GAN-AE: A promising approach for fault diagnosis in machine health monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-13, 2022.

[6] C. Yang, J. Zhang, Y. Chang, et al, "A novel deep parallel time-series relation network for fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-13, 2023.

[7] H. Cui, J. Li, Q. Hu and Q. Mao, "Real-time inspection system for ballast railway fasteners based on point cloud deep learning," *IEEE Access*, vol. 8, pp. 61604-61614, 2020.

[8] Jiao, Jian, and Xue-jiao Zheng. "Fault diagnosis method for industrial robots based on DBN joint information fusion technology," *Computational intelligence and neuroscience*, no. 1: 4340817, 2022.

[9] Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H., Elhoseiny, M. and Ghanem, B. "Pointnext: Revisiting pointnet++ with improved training and scaling strategies," *Advances in Neural Information Processing Systems*, no. 35, pp. 23192-23204, 2022.

[10] M. S. Kim, J. P. Yun and P. Park, "Deep learning-based explainable fault diagnosis model with an individually grouped 1-D convolution for three-axis vibration signals," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 12, pp. 8807-8817, 2022.

[11] T. Xie, X. Huang and S. -K. Choi, "Intelligent mechanical fault diagnosis using multisensor fusion and convolution neural network," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3213-3223, 2022.

[12] Bai Y, Yang J, Wang J, Zhao Y, Li Q. "Image representation of vibration signals and its application in intelligent compound fault diagnosis in railway vehicle wheelset-axlebox assemblies," *Mechanical Systems and Signal Processing*, 152:107421, 2021.

[13] C. Lian, Y. Zhao, T. Sun, et al, "A new time series data imaging scheme for mechanical fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-11, 2024.

[14] G. Xin, Z. Li, L. Jia, et al. "Fault diagnosis of wheelset bearings in high-speed trains using logarithmic short-time fourier transform and modified self-calibrated residual network," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7285-7295, 2022.

[15] T. Li, Z. Zhao, C. Sun, et al. "WaveletKernelNet: an interpretable deep neural network for industrial intelligent diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 4, pp. 2302-2312, 2022.

[16] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, et al. "Deep residual learning for image recognition," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778. 2016.

[17] N. Jia, Y. Sun and X. Liu. "TFGNet: Traffic salient object detection using a feature deep interaction and guidance fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 3, pp. 3020-3030, 2024.

[18] Dosovitskiy, Alexey. "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint* arXiv:2010.11929, 2020.

[19] Ren, Z., Lin, T., Feng, K., et al. "A systematic review on imbalanced learning methods in intelligent fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp.1-35, 2023.

[20] J. Zhang, K. Zhang, Y. An, et al, "An integrated multitasking intelligent bearing fault diagnosis scheme based on representation learning under imbalanced sample condition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 5, pp. 6231-6242, 2024.

[21] Luo, F., Ma, J., & Ho, G. T. S. "An instance-level data balancing method for object detection via contextual information alignment," *Image and Vision Computing*, 149, 105155, 2024.

[22] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, at al. "Generative adversarial networks," *Communications of the ACM* 63, no. 11, pp. 139-144, 2020.

[23] J. Yang, J. Liu, J. Xie, et al, "Conditional GAN and 2-D CNN for bearing fault diagnosis with small samples," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-12, 2021.

[24] Yang, J., Zhang, G., Chen, B. and Wang, Y. "Vibration signal augmentation method for fault diagnosis of low-voltage circuit breaker based on W-CGAN," *IEEE Transactions on Instrumentation and Measurement*, 72, pp.1-11, 2023.

[25] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840-6851, 2020.

[26] C. Luo. "Understanding diffusion models: An unified perspective," *arXiv preprint* arXiv:2208.11970, 2022.

[27] Dhariwal, P. and Nichol, A. "Diffusion models beat gans on image synthesis," Advances in Neural Information Processing Systems, no. 34, pp.8780-8794, 2021.

[28] Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. "High-resolution image synthesis with latent diffusion models," In Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684-10695. 2022.

[29] Zhu, Ye, Yu Wu, Zhiwei Deng, Olga Russakovsky, and Yan Yan. "Unseen image synthesis with diffusion models," *arXiv preprint* arXiv:2310.09213, 2023.

[30] Sohl-Dickstein, Jascha, et al. "Deep unsupervised learning using nonequilibrium thermodynamics," *International Conference on Machine Learning*, PMLR, 2015.

[31] Song Yang, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint* arXiv:2011.13456, 2020.

[32] Y. Song, L. Shen, X. Lei, and S. Ermon., "Solving inverse problems in medical imaging with score-based generative models," *arXiv preprint* arXiv:2111.08005, 2021

[33] P. Chen, C. Xu, Z. Ma and Y. Jin, "A mixed samples-driven methodology based on denoising diffusion probabilistic model for identifying damage in carbon fiber composite structures," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-11, 2023.

[34] X. Yang, T. Ye, X. Yuan, W. Zhu, X. Mei and F. Zhou, "A novel data augmentation method based on denoising diffusion probabilistic model

for fault diagnosis under imbalanced data," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 5, pp. 7820-7831, 2024.
[35] Yang, Luoxiao, Yun Wang, Xinqi Fan, Israel Cohen, Yue Zhao, and Zijun Zhang. "ViTime: a visual intelligence-based foundation model for time series forecasting," *arXiv preprint* arXiv:2407.07311, 2024.
[36] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE," *Journal of Machine Learning Research* , no. 11, pp. 2579-2605, 2008.

Junping Zhong (Member, IEEE) received his Ph.D. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2022. From Oct 2019 to Oct 2020, he was a Ph.D. visitor in the Department of Structural Engineering at Delft University of Technology, Netherlands. He works as a Postdoctoral Fellow in the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University. His research interests include image processing, signal processing, and their applications in fault detection of industrial infrastructure. He serves as a reviewer for IEEE TII, IEEE TITS, IEEE TIM, and Applied Soft Computing. He was selected as the Outstanding Reviewers of IEEE Transactions on Instrumentation and Measurement in 2021 and 2022, respectively.

Yuk Ming Tang received his B.Sc., M.Phil., and Ph.D. degrees from The Chinese University of Hong Kong. Following his graduation, he worked as a Postdoctoral Fellow at the Faculty of Medicine at CUHK. He is currently a Senior Lecturer in the Department of Industrial and Systems Engineering at the Hong Kong Polytechnic University. Dr. Tang's research interests encompass Artificial Intelligence (AI), Virtual Reality (VR), the Internet of Things, Digital Twin technology, blockchain, and sustainable technology within Industry 4.0, as well as healthcare applications. He serves as the Lab-in-Charge and is a key member of several laboratories and Joint Research Centres focused on the development of projects involving AI and other advanced technologies. Dr. Tang has published over 80 articles in internationally refereed journals, including numerous contributions to top-ranked publications.

Ka Chun Ng holds an MSc in Information Technology and a BEng in Product Engineering with Marketing from the Hong Kong Polytechnic University. After gaining diverse industry experience in research, product engineering, and 3D solutions, he currently serves as a research assistant in the Department of Industrial and Systems Engineering at the Hong Kong Polytechnic University. His research interests span a broad spectrum of cutting-edge topics, including extended reality, the metaverse, artificial intelligence and machine learning, product engineering, and innovative pedagogical strategies. Hadyn's multidisciplinary background enables him to approach research challenges from unique perspectives, focusing on leveraging emerging technologies to enhance educational experiences and industrial processes.

Kai Leung Yung received the B.Sc. degree in electronic engineering, the M.Sc. degree in automatic control systems, and the Ph.D. degree in microprocessor applications in process control, U.K., in 1975, 1976, and 1985, respectively. He became a Chartered Engineer (C.Eng. and M.I.E.E.), in 1981. He was working at the U.K. for companies, such as BOC Advanced Welding Company Ltd., the British Ever Ready Group, and the Cranfield Unit for Precision Engineering. In 1986, he returned to Hong Kong to join the Hong Kong Productivity Council as a Consultant and subsequently switched to academia to join The Hong Kong Polytechnic University, where he is currently with the Department of Industrial and Systems Engineering. He has a wealth of experience in making sophisticated space tools for different depth space exploration missions. These include the Space Holinser Forceps for the MIR Space Station, the "Mars Rock Corer" for the European Space Agency's Mars Express Mission in 2003, the "Soil Preparation System" for the Sino-Russian Phobos-Grunt Mission, in 2011, and advanced precision robotic systems for the China Lunar Exploration Missions such as the Camera Pointing System used on the surface of the moon.