

Semi-supervised consistency models for automated defect detection in carbon fiber composite structures with limited data

Peng Chen^{1,2,*} , Junxiao Ma¹, Changbo He³ , Yaqiang Jin^{4,5} and Shuai Fan⁶

¹ College of Engineering, Shantou University, Shantou 515063, Guangdong, People's Republic of China

² Key Laboratory of Intelligent Manufacturing Technology, Shantou University, Shantou 515063, Guangdong, People's Republic of China

³ College of Electrical Engineering and Automation, Anhui University, Hefei 230601, Anhui, People's Republic of China

⁴ School of Qilu Transportation, Shandong University, Jinan 250061, Shandong, People's Republic of China

⁵ Qingdao Mingserve Tech, Qingdao 266041, Shandong, People's Republic of China

⁶ School of Mechanical and Electrical Engineering, Chengdu University of Technology, Chengdu 610059, Sichuan, People's Republic of China

E-mail: pengchen@alu.uestc.edu.cn and dr.pengchen@foxmail.com

Received 18 December 2024, revised 9 March 2025

Accepted for publication 13 March 2025

Published 25 March 2025



Abstract

Deep learning has increasingly been adopted in the non-destructive testing industry for detecting defects in carbon fiber composite structures (CFCS), particularly in CFCS-cored aluminum conductor composite core (ACCC) wires. However, the effectiveness of these approaches is often limited by the availability of sizable annotated failure datasets for training purposes. Addressing this challenge, this paper presents a semi-supervised model employing consistency strategies to automate defect detection in CFCS, compensating for the real-world scarcity of samples. It proposes a multi-faceted approach combining synthetic sample generation, transformer-based feature fusion, and a DenseNet architecture-based detection module. Initially, the model generates a large set of synthetic data to mitigate the issue of limited real-world sample availability. These synthetic samples, produced through consistency strategies, are then engineered to complement actual test data. The following step involves a transformer architecture that blends features from synthetic and real samples, refining the dataset for improved damage identification. The final stage features a detection module based on DenseNet architecture, particularly designed for assessing the integrity of CFCS within ACCC wires. Experiments conducted in real-world field scenarios have shown the model's effectiveness, demonstrating that the hybrid use of synthetic and real samples substantially enhances the training process and damage detection capabilities.

Keywords: fault diagnosis, health condition monitoring, semi-supervised learning, damage diagnosis, carbon fiber composite structures

* Author to whom any correspondence should be addressed.

1. Introduction

Carbon fiber composite structures (CFCS) are widely employed across various industries such as power grid infrastructure, wind energy, aerospace, and telecommunications [1–3]. Significantly, CFCS-cored aluminum conductor composite core (ACCC) wires have become prominent in electrical grid applications, offering superior conductivity, lightness, high tensile strength, and resistance to corrosion. Nevertheless, these wires are prone to damage due to adverse environmental stresses, which can lead to costly and unplanned maintenance. Consequently, to prevent such disruptions and improve the efficiency and reliability of the power transmission network, fault diagnostics and health condition monitoring of power lines is imperative.

In the past, inspections to determine the integrity of ACCC wires have traditionally been conducted by manual methods, which present challenges including high costs, potential hazards, and significant labor time. To improve on these manual techniques, a range of Non-destructive evaluation (NDE) technologies have been developed for the inspection of overhead ACCC wires. These technologies encompass a variety of sensors, some attached directly to the wires, such as ultrasonic gilded wave systems [4], and others that form part of power line carrier systems. Additionally, robotic platforms have been equipped with an array of sensors including infrared [5, 6], electromagnetic induction [7], eddy current [8, 9], stress [10], and acoustic wave detectors [11–13]. Power line companies favor unmanned aerial vehicles (UAVs) over helicopters for inspections due to their compact size [14]. Equipped with sensors and cameras, UAVs capture detailed images and navigate close to power lines without disturbances. Nevertheless, the effectiveness of these NDE technologies can sometimes be compromised by environmental conditions like lighting or by the material properties inherent to the ACCC wires, which can be especially challenging for methods like those based on eddy currents. To mitigate these compatibility issues, sometimes a magnetic sleeve must be applied to the wire's composite core, ensuring that the non-destructive detection methods remain functional in grid applications [15].

While many non-destructive testing (NDT) methods are available, their reliability in detecting damage has been inconsistent. Recent studies have shown that x-ray imaging holds promise in identifying internal damages in ACCC wires by allowing inspection of thicker materials and spotting both external and internal flaws. One of its advantages is that it can transmit x-rays through thick composite components, making it possible to inspect thicker materials. Moreover, x-ray imaging can detect both surface and subsurface defects and provide a record of the inspection. The vision-based methodologies such as camera-sensor or infrared-sensor are limited with difficult to accommodate with early failure with different scales and tiny sizes, Song *et al* [16] introduced Deformable you only look once X, improving detection accuracy and speed through strategies like decoupling head and anchor-free design. Adaptively spatial feature fusion resolves feature inconsistency in different scales. Chen *et al* [17] improved YOLOv8 by using content-aware up-sampling for

small objects, a dedicated detection head for small insulator defects, and the SIOU loss function to enhance both mAP and inference speed. Wang *et al* [2] presented an innovative automated detection system that addresses the challenge of identifying internal defects in aluminum conductor composite core (ACC) wires. This system incorporates an x-ray inspection robot and employs an anchor-free object detection model. Different backgrounds (water, ground, trees, buildings, towers, etc) have diverse impacts on the appearance of the power line components. Moreover, variations in illumination conditions also occur because of the presence of sunlight in the outdoor situation, which depends on the daytime and the weather conditions on the day of data collection. Wang *et al* [18] developed a network that initially focuses on insulator segmentation before conducting defect detection on the segmented insulator image. This segmentation step eliminates the background, thereby mitigating interference. In a similar vein, Abdelfattah *et al* [19] utilized generative adversarial networks (GANs) for extracting thin power lines from intricate backgrounds. Their proposed approach involved joint training of a generator and a semantic decoder.

Deep learning [20–24], especially algorithms for fault diagnosis and object detection, has been increasingly applied to enhance NDT, with robots now being developed to carry portable x-ray systems and automated diagnostic tools for identifying wire defects. [25], x-ray imaging has demonstrated potential in identifying internal ACCC wire damage using non-destructive detection techniques. with researchers developing robots equipped with portable x-ray generators and deep learning-based defect diagnostic systems to detect ACCC wire defects. Gao *et al* [26] developed a robot that includes a portable x-ray generator, a DR detecting pane for gathering x-ray images, and a deep learning-based defect diagnostic system combined with manual evaluation to discover defects to ACCC wires. Wei *et al* [27] provide an automatic defect detection approach based on deep convolution network (DCN) for defect detection of the aluminum conductor carbon fiber composite core through x-ray images. However, these approaches, particularly the supervised deep learning algorithms mentioned above, require a large number of labeled failure data for training. In practice, obtaining sufficient amounts of labeled failure samples can be difficult, making these approaches somewhat challenging to implement in real-world industrial applications.

In real-world testing, the quality of captured x-ray images is often compromised by noise, spurious edges, and other environmental interference. This can be especially problematic when examining CFCS, as the low contrast of the carbon core and the high energy absorption of the outer steel core can make it difficult to identify cracks. As known that the effectiveness of deep learning methods is contingent on the scale of the training dataset. A larger annotated dataset corresponds to improved accuracy in classification or detection for deep learning models. The issue of data imbalance and data insufficiency in real-world measurements, where the number of samples for a specific type of defect may be limited, leading to potentially misleading deep learning performance estimates. Wang *et al* [28] modified the contribution of gradients from negative and positive samples in the loss function. The subsequent section

dives into various methodologies employed by the power line inspection community to address issues of data scarcity and class imbalance within datasets. In a study by Zhai *et al* [29], a mirror-flipping augmentation strategy was adopted to expand the dataset, coupled with the utilization of domain knowledge to tackle class imbalance. The incorporation of domain knowledge with visual features proved effective in enhancing the detection of small objects. Researchers [30] have explored ways to design deep learning models to avoid such inaccurate estimates, particularly in cases of small sample sizes. GANs [31] have frequently been regarded as state-of-the-art generative models in terms of sample quality, but they are known for being difficult to train and prone to mode collapse [32]. On the other hand, diffusion models (DM) [33], which are likelihood-based, provide more diversity and a steady training process, may be a more practical solution for dealing with these issues.

The restrictions of the aforementioned can be summarized as follows:

1. Manual inspection drawbacks: Engaging in conventional manual inspection practices presents significant challenges due to their considerable expense, inherent safety risks, and the extensive time commitments they impose.
2. NDE limitations: The introduction of innovative NDE techniques, backed by sensors and advanced robotic systems like UAVs, aims to augment the efficiency of ACCC wires inspection. However, these methods are not immune to difficulties; they may encounter performance issues in certain environmental settings or may not fully adapt to the unique material traits of the wires.
3. Obstacles in model training: The process of procuring a substantial volume of accurately labeled training data for the construction of supervised deep learning algorithms presents a formidable challenge. This scarcity of training data serves as a barrier to the broader application and effectiveness of these computational models in real-life scenarios.
4. Issues with data availability: The imbalance and lack of sufficient data, especially regarding the range and occurrence of material defects, can lead to inaccuracies in the machine learning algorithms designed to detect and classify such defects. This imbalance has a direct impact on the reliability of deep learning outcomes.

As discussed above, there have been few investigations addressing the challenge of small sample sizes in the NDT of CFCS-cored ACCC wires. This study explores novel methodologies for overcoming the limitations of small sample sizes in the NDE of CFCS-cored ACCC wires. The research proposes a semi-supervised model based on consistency strategies to detect damage within CFCS. This model is composed of a module for generating samples using consistency models, a module for feature fusion using a transformer architecture, and a damage detection module. The model is particularly adept at utilizing a limited number of actual test samples to produce an extended set of synthetic counterparts, effectively circumventing the constraints associated with scarce sample data. These synthetic samples are then amalgamated with genuine samples within the feature fusion transformer module to refine their

attributes. Subsequent to this fusion, the samples are processed through a module built on DenseNet architecture, culminating in a specialized module targeting the detection of damage within ACCC wires in CFCSs.

The key contributions of this research are articulated as below:

1. Synthetic data generation and training facilitation: This study proposes a model that utilizes consistency models to generate a large quantity of training data. This addresses the long-standing issues of data insufficiency and imbalance in damage detection. The synthetic samples not only solve the problem of scarce data but also overcome the obstacle of limited training data, which restricts the wide-scale application and effectiveness of computational models in practical scenarios. By incorporating these synthetic samples, the model can be trained more comprehensively, thus improving its performance.
2. Emphasized feature fusion for enhanced detection: Our model incorporates a transformer-based framework, with a primary focus on the feature fusion process. This architecture serves as a crucial medium for integrating the features of real and synthetic data. Through this fusion, the model can extract more comprehensive and discriminative features. This integrated feature set significantly enriches the information available for the model during training.
3. Viability and adaptability of proposed methods: The proposed approach is robust, not only in its immunity to difficulties often presented by certain environmental factors or in adapting to unique material characteristics of structures but also in its straightforward applicability to real-world situations. The methods incur minimal costs, impose low safety risks, and are time-efficient, making them particularly suitable for field applications.

The organization of this paper is systematically presented as follows: Initially, section 2 revisits the fundamental principles underlying the denoising diffusion probabilistic model (DDPM) and consistency models (CM), which play a crucial role in shaping the methodology proposed in this paper. Following that, section 3 is dedicated to present the proposed Semi-supervised CM (SSCM) providing a detailed exposition of their theory and implementation. Subsequently, a comparative analysis, along with validation utilizing a case study from actual field measurements, is detailed in section 5. The paper concludes with section 6, which provides a comprehensive overview of the research's foundational approach and encapsulates the principal findings.

2. Related theory

This section will begin by revisiting CM [34], influenced significantly by the theory of continuous-time DM. Rooted in the extensive exploration of continuous-time DM in previous studies [35, 36], CM present a complex yet intriguing topic. The section will provide concise introductions to both DM and CM to lay the necessary groundwork for later discussions. This

foundational understanding will pave the way for addressing the proposed SSCM in subsequent sections.

DMs adopt a progressive strategy by systematically introducing perturbations to data through Gaussian perturbation. These models then generate samples from the perturbed data via sequential denoising steps, starting with the original data distribution denoted as $p_{\text{data}}(x)$. Utilizing a forward Stochastic Differential equation (SDE) [35, 36], DMs gradually disperse the probability distribution $p_{\text{data}}(x)$.

$$dx_t = \mu(x_t, t) dt + \sigma(t) d\omega_t, \quad (1)$$

where $t \in [0, T]$ and $T > 0$ is a predetermined constant, the functions $\mu(\cdot, \cdot)$ and $\sigma(\cdot)$ represent the drift and diffusion coefficient, respectively. The term $\{\omega_t\}_{t \in [0, T]}$ presents the standard Brownian motion, emphasizing the stochastic nature inherent in the diffusion process.

The SDE can be reversed at each time step. Consequently, the *Reversed SDE* provides a means to systematically transform noise into data. Notably, an ordinary differential equation (ODE) known as the *Probabilistic flow (PF) ODE* is equivalent to the Reversed SDE [35].

$$dx_t = \left[\mu(x_t, t) - \frac{1}{2} \sigma(t)^2 \nabla \log p_t(x_t) \right] dt, \quad (2)$$

The distribution of x_t , denoted as $p_t(x)$, is represented by $p_0(x)$ at time $t=0$, which is equivalent to the data distribution $p_{\text{data}}(x)$. As the system evolves over time, the distribution at time T , $p_T(x)$, approaches a tractable Gaussian distribution $\pi(x)$. The score function of $p_t(x)$ is given by $\nabla \log p_t(x)$.

To facilitate the sampling process, it begins by training a *score model* denoted as $s_\phi(x, t) \approx \nabla \log p_t(x)$ using the *score matching* technique. The resulting $s_\phi(x, t)$ is then integrated into equation (2), enabling the derivation of an empirical estimate of the Partial Differential equation (PF ODE) in the following form.

$$\frac{dx_t}{dt} = -ts_\phi(x_t, t). \quad (3)$$

To begin the empirical PF ODE, start by sampling \hat{x}_T from $\pi = \mathcal{N}(0, T^2 I)$. Utilize any numerical ODE solver to retrospectively solve it, yielding the solution trajectory $\{\hat{x}_t\}_{t \in [0, T]}$. The initial value \hat{x}_0 serves as an approximate sample from the data distribution $p_{\text{data}}(x)$. To address numerical instability, it is common to terminate the ODE solver at $t = \epsilon$, where ϵ is a predefined small positive value.

Consistency models, a novel type of generative model, support both one-step and multi-step iterative sample generation. Training can occur in two modes: distillation and isolation. In the distillation mode, pre-trained DMs are crucial as consistency models distill knowledge for a single-step sampler. Conversely, in the isolation mode, consistency models can be trained independently, distinguishing them as an autonomous class within generative models.

The core of the consistency strategy lies in the self-consistency property of the consistency function $f: (x_t, t) \mapsto x_\epsilon$, building upon the trajectory solution $\{\hat{x}_t\}_{t \in [\epsilon, T]}$ provided

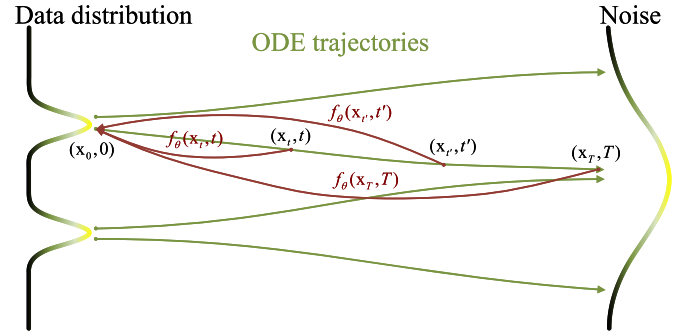


Figure 1. Consistency models correlate points on any trajectory of PF ODE with the origin of that trajectory.

by equation (2). This property ensures $f(x_t, t) = f(x_{t'}, t')$ for all $t, t' \in [\epsilon, T]$, maintaining uniformity within the same PF ODE trajectory. In practical terms, this means that for a given set of data points that lie on the same trajectory of the PF ODE, the consistency function will map them to the same reference point x_ϵ . For example, consider two data points x_{t1} and x_{t2} at different time steps $t1$ and $t2$ within the range $[\epsilon, T]$ but on the same PF ODE trajectory. The consistency function f will output the same result for both points, regardless of their different time—step positions. This property is fundamental for the consistent handling of data during sample generation and feature fusion processes in the proposed semi-supervised model. It helps to standardize the data representation and makes the model more robust to noise and variations in the input data. The goal of the consistency model, denoted as f_θ , is to estimate this consistency function through data-driven learning and reinforce its self-consistency property, as illustrated in figure 1.

For any consistency function $f(\cdot, \cdot)$, it must satisfy the boundary condition $f(x_\epsilon, \epsilon) = x_\epsilon$, indicating that $f(\cdot, \epsilon)$ acts as an identity operation. This constraint, known as the ‘boundary condition,’ is essential for validating a consistency model. In the context of consistency models based on deep neural networks, let us consider a flexible network $F_\theta(x, t)$ with output matching the dimensionality of x . A practical approach to parameterize the consistency model involves using skip connections.

$$f_\theta(x, t) = c_{\text{skip}}(t)x + c_{\text{out}}(t)F_\theta(x, t). \quad (4)$$

The differentiability of functions $c_{\text{skip}}(t)$ and $c_{\text{out}}(t)$ is crucial, ensuring $c_{\text{skip}}(\epsilon) = 1$ and $c_{\text{out}}(\epsilon) = 0$. This configuration maintains consistency model differentiability at $t = \epsilon$, contingent upon the differentiability of $F_\theta(x, t)$ and scaling coefficients, pivotal for effective training of continuous-time consistency models.

A well-trained consistency model $f_\theta(x, t)$ streamlines sample generation by drawing from $\hat{x}_T \sim \mathcal{N}(0, T^2 I)$. The model evaluates $\hat{x}_\epsilon = f_\theta(\hat{x}_T, T)$ for one-step noise-to-sample generation. Iteratively introducing noise further improves the model’s evaluation and enhances the overall sample quality.

3. SSCM

The section proposes SSCMs for detecting damage in CFCSs, as illustrated in the referenced figure 2. The model comprises a sample generation module utilizing consistency models, a feature fusion transformer module, and a damage identification module. It leverages a small set of real-world test samples to generate numerous synthetic samples, thereby addressing the issue of limited sample availability. Subsequently, these synthetic samples are integrated with real ones in the feature fusion transformer module to enhance their characteristics. The enhanced samples undergo processing in a DenseNet-based module, and finally, a module specifically designed for diagnosing damage in the CFCSs of ACCC wires.

The process begins with the original datasets being subjected to a diffusion process as described by the SDE in equation (1). This involves augmenting the data with noise, which is subsequently used for training consistency models. It is important to note that the absence of an existing pre-trained DM precludes the strategy of leveraging a pre-trained scored model, represented by $s_\phi(x, t)$, for training consistency models through distillation. Instead, the training of the consistency models is conducted independently. This independent training relies on the employment of an unbiased estimator for the gradient of the log probability $\nabla \log p_t(x_t)$ as follows.

$$\nabla \log p_t(x_t) = -\mathbb{E} \left[\frac{x_t - x}{t^2} \middle| x_t \right] \quad (5)$$

where $x \sim p_{\text{data}}$ and $x_t \sim \mathcal{N}(x; t^2 \mathbf{I})$, a Monte Carlo estimate of $\nabla \log p_t(x_t)$ can be formed with $-(x_t - x)/t^2$. This estimate effectively replaces the pre-trained DM in consistency distillation, particularly when utilizing the Euler method, like the Heun Solver [36], as the ODE solver in the limit of $N \rightarrow \infty$ [34].

After undergoing training with a neural network to acquire a consistency models characterized by effective self-consistency, a novel sample \tilde{x}_0 can be generated by introducing \tilde{x}_T , drawn from a normal distribution $\mathcal{N}(0, \mathbf{I})$, into the established consistency model, as shown in figure 2.

The dataset synthesized from the consistency model is combined with a limited real dataset using a feature fusion transformer module, producing fused data with enhanced information. Both synthesized and real data undergo processing through four encoder blocks, each consisting of a 3×3 kernel convolution layer followed by ReLU activation and max-pooling. The Spatio-Transformer fusion network merges features from both data at multiple scales. The spatial branch captures local features with convolution layers and a bottleneck layer, while the transformer branch utilizes an axial attention-based transformer block for global context through self-attention. Finally, a nested decoder network, based on RFN-Nest architecture [37], is trained to generate fused

images using the combined features. In figure 1, DenseNet is employed to improve damage identification accuracy. This improvement is achieved through the optimization of information flow, the enhancement of reusable features, and the overall improvement of features themselves. The calculation of the output of the r th layer, denoted as z_r , and it can be calculated as below.

$$z_r = H_r([z_0, z_1, \dots, z_{r-1}]) \quad (6)$$

where the function H_r (with r taking values of 1, 2, 3, ...) is primarily constituted by batch normalization, the Rectified Linear Unit (ReLU), and convolution. The features generated in layers 0, 1, ..., $r-1$ undergo concatenation and are denoted as $[z_0, z_1, \dots, z_{r-1}]$.

Subsequently, the \tilde{z}_r for a small batch of data, denoted as B , can be computed using the following method.

$$z_{BN} = \frac{\hat{z}_r - \mu_B}{\sqrt{\sigma_B^2 + \alpha}} \quad (7)$$

$$\tilde{z}_r = \gamma \cdot z_{BN} + \beta \quad (8)$$

where the input data undergoes normalization, denoted by z_{BN} , the r -th layer input of Batch Normalization, $\hat{z}_r = [z_0, z_1, \dots, z_{r-1}]$, is considered, with mean (μ_B) and variance (σ_B^2) being crucial for calculation. The parameter α is a small constant used to prevent division by zero. Learnable parameters, β and γ , scale and shift, enabling the network to learn and restore the original feature distribution.

The dataset denoted as x_D , utilized for damage identification, amalgamates a small portion of authentic data with a substantial volume of fused data processed through feature fusion transformer module. Within the damage diagnosis module, a soft-max classifier is employed to undergo training on feature vectors derived from a DenseNet-based model mentioned earlier. These feature vectors are extracted from a mixed sample, denoted as $\{(x_D^{(i)}, l^{(i)}); i \in 1, \dots, N, l^{(i)} \in 0, \dots, k-1\}$, where $x_D^{(i)}$ represents the i -th feature vector of the input sample, and $l^{(i)}$ corresponds to the label. The probability of a sample belonging to a specific category is determined for each sample in this set using the following equation.

$$p(l^{(i)} = j | x_D^{(i)}; \theta) = \frac{e^{\theta_j^T x_D^{(i)}}}{\sum_{j=1}^k e^{\theta_j^T x_D^{(i)}}} \quad (9)$$

where $p(l^{(i)} = j | x_D^{(i)}; \theta)$ denotes the probability that $x_D^{(i)}$ belongs to the j th category, signifying the segmentation of the sample into different damage types. This probability is equivalent to the likelihood of the sample being categorized under the j th classification.

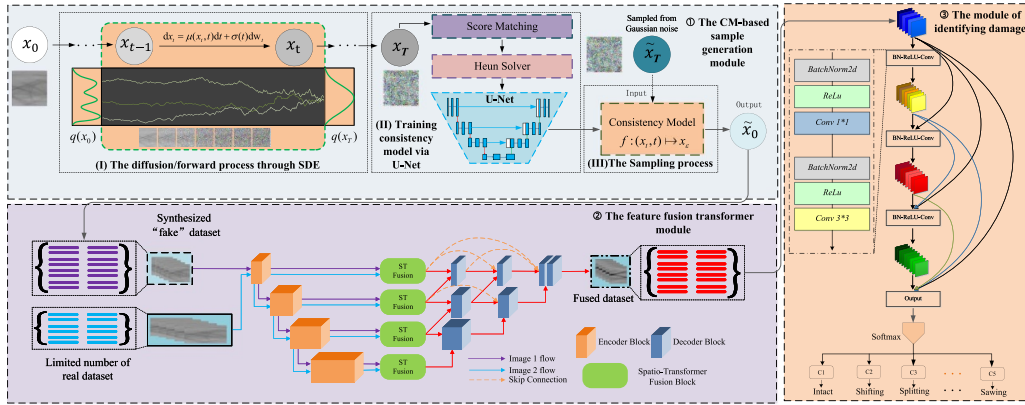


Figure 2. The proposed semi-supervised consistency model.

The method for assessing damage through the use of soft-max is articulated as follows.

$$\begin{aligned}
 h_{\theta} \left(x_D^{(i)} \right) &= \begin{bmatrix} p \left(l^{(i)} = 1 \mid x_D^{(i)} ; \theta \right) \\ p \left(l^{(i)} = 2 \mid x_D^{(i)} ; \theta \right) \\ \vdots \\ p \left(l^{(i)} = k \mid x_D^{(i)} ; \theta \right) \end{bmatrix} \\
 &= \frac{1}{\sum_{j=1}^k e^{\theta_j^T x_D^{(i)}}} \begin{bmatrix} e^{\theta_1^T x_D^{(i)}} \\ e^{\theta_2^T x_D^{(i)}} \\ \vdots \\ e^{\theta_k^T x_D^{(i)}} \end{bmatrix}
 \end{aligned} \tag{10}$$

where θ represents the model parameters and $\sum_{j=1}^k e^{\theta_j^T x_D^{(i)}}$ is the operation of the probability normalization.

The algorithm’s comprehensive workflow is delineated in algorithm 1, providing a high-level overview of its operation. Additionally, a more detailed description of the implemented procedure is available for reference.

4. Testing setup and data measurement

A data acquisition system for quantifying x-ray images of ACCC wires with carbon fiber composite core structures, comprehensively illustrated in figure 3, has been developed. The acquisition robot (figures 3(a) and (b)), composed of data collection, transmission, and analysis modules (figure 3(c)), is equipped with a portable cold cathode x-ray source (TXR-C1R150P - 08) and designed to climb ACCC wires. It integrates an x-ray generator and a digital radiography dashboard at its base for obstacle-free navigation, and enables wireless transmission of x-ray image data to cloud storage via WIFI for subsequent fault detection and analysis on a workstation; in-depth details are available in our previous work [3].

5. Validation case study: real-world field measurements

In this section, the effectiveness of damage assessment with the use of CFCS-cored ACCC wires is validated through real-world field measurements. Following the demonstration in section 5.1, an initial comparison is drawn between actual measurements and synthesized data. The subsequent preparation and examination of the data are comprehensively outlined in section 5.2. Finally, section 5.3 provides a detailed comparative analysis focused on defect-induced damage.

5.1. Comparing data between actual measurements and synthesized data

The examination of synthetic samples generated by the proposed SSCM involves a comparative analysis with real-world data, as illustrated in figure 4. This analysis aims to objectively assess any disparities or deviations between the synthetically produced samples and actual data, aiding in the identification of potential inconsistencies. Specifically, figure 4 contrasts real and synthesized representations of four failure types. Despite minor discrepancies, the majority of instances exhibit similar failure characteristics, which are pivotal for comprehending failure mechanisms and enhancing the output of the SSCM. Quantitatively, the comparison between real and synthetic data employs cosine similarity as a metric, with the results presented in table 1 indicating strong correlations exceeding 0.996. These results affirm the synthetic samples’ close approximation to real-world data, thereby bolstering confidence in the effectiveness of the SSCM model. Furthermore, to provide a comprehensive evaluation of the model’s performance across varying degrees of similarity, we conducted an in-depth analysis of cases where synthetic samples exhibited notable divergence from real measurements. As illustrated in figure A1, we present a detailed visualization of these divergent cases, complemented by corresponding quantitative metrics in table A1 for systematic evaluation. Our analysis revealed that synthetic data encompassing the four failure types (fracture, sawing, shifted, and splitting) demonstrated relatively lower similarity with an average cosine similarity value of 0.95. Specifically, the selected synthetic-real

Algorithm 1. The algorithm for implementing the proposed SSCM method.

```

1: procedure SampleGeneration( $f_{\theta}(\cdot, \cdot), \{\tau_1, \tau_2, \dots, \tau_{N-1}\}, \hat{\mathbf{x}}_T$ )
2:    $\mathbf{x} \leftarrow f_{\theta}(\hat{\mathbf{x}}_T, T)$ 
3:   for  $n = 1$  to  $N - 1$  do
4:      $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
5:      $\hat{\mathbf{x}}_{\tau_n} \leftarrow \mathbf{x} + \sqrt{\tau_n^2 - \epsilon^2} \mathbf{z}$ 
6:      $\mathbf{x} \leftarrow f_{\theta}(\hat{\mathbf{x}}_{\tau_n}, \tau_n)$ 
7:   end for
8:   return  $\mathbf{x}$ 
9: end procedure
10: procedure FeatureFusion( $\mathbf{x}_R, ST, E_J, D$ )
11:    $\mathbf{F}_J \leftarrow E_J(\mathbf{x}, \mathbf{x}_R)$ 
12:    $\mathbf{F}_{J_{fused}} \leftarrow ST(\mathbf{F}_J)$ 
13:    $\mathbf{x}_{fused} \leftarrow D(\mathbf{F}_{J_{fused}})$ 
14:   return  $\mathbf{x}_{fused}$ 
15: end procedure
16: procedure Classification( $\{\mathbf{x}_{fused_1}, \mathbf{x}_{fused_2}, \dots, \mathbf{x}_{fused_M}\}, L, k, C$ )
17:    $\mathbf{x}_{in} \leftarrow \{\mathbf{x}_{fused_1}, \mathbf{x}_{fused_2}, \dots, \mathbf{x}_{fused_M}\}$ 
18:   Initialize Conv1 with ReLU activation
19:   Initialize Dense Blocks and Transition Layers
20:   for  $l = 1$  to  $L$  do
21:      $\mathbf{x}_{in} \leftarrow \mathbf{x}_{in}$ 
22:      $\mathbf{x}_{out} \leftarrow \text{DenseBlock}_l(\mathbf{x}_{in}, n_l)$ 
23:      $\mathbf{x}_{in} \leftarrow [\mathbf{x}_{in}, \mathbf{x}_{out}]$ 
24:      $\mathbf{x}_{in} \leftarrow \text{TransitionLayer}_l(\mathbf{x}_{in})$ 
25:   end for
26:   Global Average Pooling
27:   Flatten the output
28:   Fully Connected Layer with Softmax Activation for  $C$  classes
29:   return Predicted Class Probabilities  $\{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_N\}$ 
30:   for  $i = 1$  to  $N$  do
31:     Predicted Label for Image  $i$ :  $\arg \max(\mathbf{P}_i)$ 
32:     if Predicted Label = 5 then
33:       Output for Image  $i$ : ‘Image  $i$  belongs to class 5’
34:     else
35:       Output for Image  $i$ : ‘Image  $i$  does not belong to class 5’
36:     end if
37:   end for
38: end procedure

```

sample pairs exhibited varying degrees of dissimilarity across different defect types, with cosine similarity values ranging from 0.931 to 0.986: fracture (0.943), sawing (0.986), shifting (0.931), and splitting (0.976).

5.2. Preparation and examination of data

To achieve consistency in training data, it is essential to employ methods such as segmentation, straightening, and gray-level normalization (as illustrated in figure 5(b)). Deviations in ACCC wires resulting from gravitational effects are mitigated by employing straightening techniques, which horizontally align lines using centered pixel coordinates. To account for variations in x-rays and detectors, corrections are made to the brightness and contrast. Discontinuities are

reduced by substituting vertical integration with convolution processes. Additionally, histogram specification is utilized to standardize contrast across various images, ensuring a uniform color distribution.

In analyzing figure 5, defect positions are determined from the central part of raw x-ray images. Additionally, a cross-sectional diagram illustrating the ACCC wire can be observed in figure 5(a). The images are divided into patches using a sliding window for analysis. Figure 5(b) shows overlapping patches, preventing misinterpretations caused by interference from exterior aluminum wire gaps. Figure 6 depicts ACCC wires with CFCS cores in both undamaged condition alongside four different types of damage states. Four defect types: shifting, splitting, fracture, and sawed are identified based on morphological properties (figure 4). Real-world testing introduces noise and interference, reducing image quality. Limited x-ray images lead to the use of SSCM to generate substantial varieties of samples, resulting in 6000 raw samples from 400 images. Table 2 details their distribution for training and testing, each sample sized at $128 \times 128 \times 1$.

5.3. Evaluation and comparative analysis of defect-induced damage

The optimization of hyperparameters constitutes a pivotal stage in ensuring the effective training of the proposed architecture, a notion supported by insights gleaned from the literature [20, 38]. Following each pooling and fully-connected layer, the ReLU is employed as the activation function, alongside batch normalization applied with a momentum of 0.9 across both input and activation layers. The convolution layers are outfitted with 12 input filters. The architecture, depicted in figure 2, encompasses 16 Basic Blocks and 16 Transition Blocks, all interlinked with each Dense Block. During training, the Adam optimizer is utilized, with a learning rate set at 0.0001, spanning 50 epochs and a growth rate of 12. An examination of the relationship between the number of epochs and loss reveals a notable decline initially, which stabilizes around the 40th epoch, suggesting a plateauing of model performance beyond a certain epoch count, a behavior akin to that observed in conventional models like ShuffleNet [39]. As depicted in figure 7(b), other models such as ShuffleNet [39], ResNet-based [40], and DenseNet [41] exhibit superior performance.

For the implementation of these algorithms, PyTorch is utilized on a computing platform equipped with an AMD Ryzen 9 5950X 16-core CPU, 64 GB DDR4 memory, and an Nvidia RTX3090Ti GPU, facilitating the training of the proposed architecture.

This section conducts a comprehensive comparative analysis of several state-of-the-art deep learning architectures to evaluate their efficacy in detecting structural damage within CFCS-cored ACCC wires, which are crucial components in modern power transmission systems. The evaluation encompasses well-established architectures such as ShuffleNet [39], which is known for its computational efficiency; RepVGG [42], which features a straightforward yet effective architectural design; ResNet50 and ResNet101

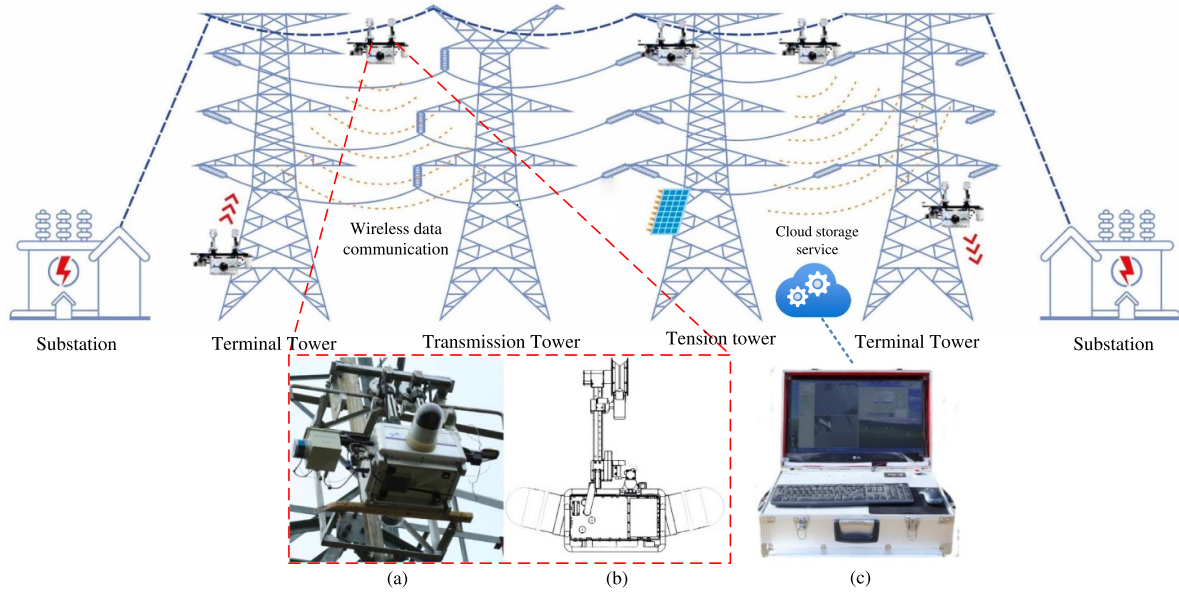


Figure 3. Schematic of damage detection during field deployment measurements: (a) Inspection robot, (b) Inspection robot sketch, (c) Data analysis system.

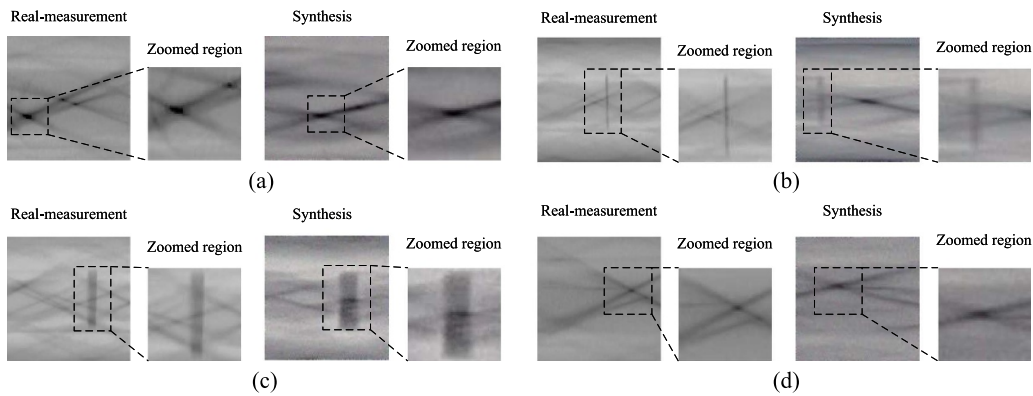


Figure 4. Comparison of Samples: Synthesized vs. Real Measurements. (a) X-ray Image of Fractured Sample with Zoomed Region. (b) X-ray Image of Sawed Sample with Zoomed Region. (c) X-ray Image of Shifted Sample with Zoomed Region. (d) X-ray Image of Splitting Sample with Zoomed Region.

Table 1. Quantitative comparison of real measurement and synthesis.

Type of defect	Cosine similarity
Fracture	0.985
Sawing	0.996
Shifting	0.986
Splitting	0.996

[40, 43], which implement residual learning frameworks; DenseNet [41], which utilizes dense connectivity patterns; and a DDPM-based method [3], which leverages DMs for enhanced feature extraction. Furthermore, to rigorously evaluate the effectiveness of the feature fusion module proposed in this research, the study extends its comparative analysis to include state-of-the-art image feature fusion models, specifically CoCoNet [44] and CDDFuse [45], which have demonstrated notable performance in similar computer vision tasks. The primary metric for evaluation is test accuracy. ShuffleNet

achieves a peak accuracy of approximately 90.2%, but exhibits its instability, resulting in fluctuations and an average accuracy of 90.0%, as shown in figure 8(a). In contrast, ResNet-based algorithms (ResNet50 and ResNet101) demonstrate greater stability, consistently maintaining an accuracy rate near 91.2%. DenseNet, with its direct layer connections, achieves a slightly higher maximum accuracy of 92.0%. The proposed SSCM surpasses the benchmark set by DenseNet, achieving a test accuracy of around 96.2% and an average accuracy of 94.0%, as depicted in figures 7(a) and 8(a). Table 3

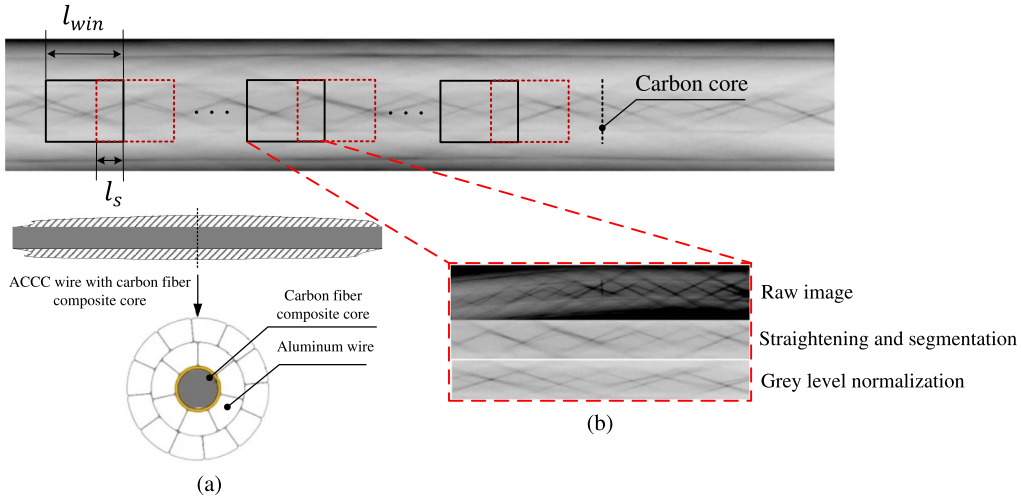


Figure 5. (a) Cross-section diagram of ACCC wires, (b) Schematic illustration of data pre-processing.

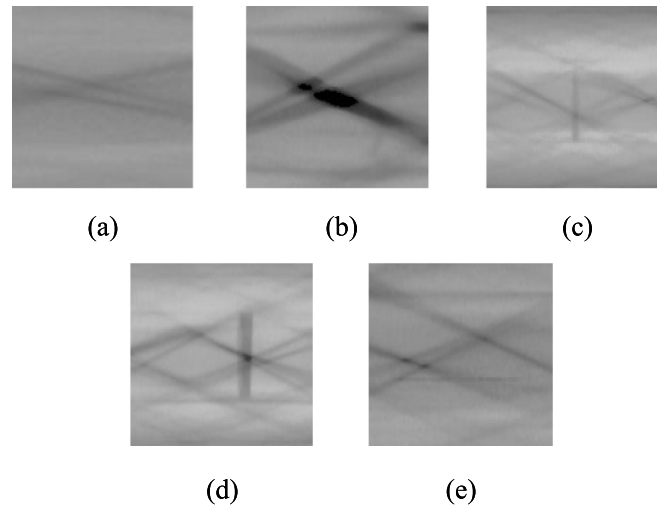


Figure 6. Typical damaged ACCC wires with CFCS cores. (a) X-ray image of an intact sample. (b) X-ray image of a fracture sample. (c) X-ray image of a sawed sample. (d) X-ray image of a shifting sample. (e) X-ray image of a splitting sample.

Table 2. Organization of training and test data sets.

Models	Categories	Training datasets	Test datasets
Existing models	Fracture	400 (real)	100 (real)
	Intact	1200 (real)	100 (real)
	Sawing	400 (real)	100 (real)
	Shifting	400 (real)	100 (real)
	Splitting	400 (real)	100 (real)
Proposed model	Fracture	400 (real) + 800 (synthetics)	100 (real)
	Intact	1200 (real)	100 (real)
	Sawing	400 (real) + 800 (synthetics)	100 (real)
	Shifting	400 (real) + 800 (synthetics)	100 (real)
	Splitting	400 (real) + 800 (synthetics)	100 (real)

Note: Existing models are denoted as ShuffleNet, RepVGG, ResNet50, ResNet101, DenseNet, DDPM-based method, CoCoNet, and CDDFuse.

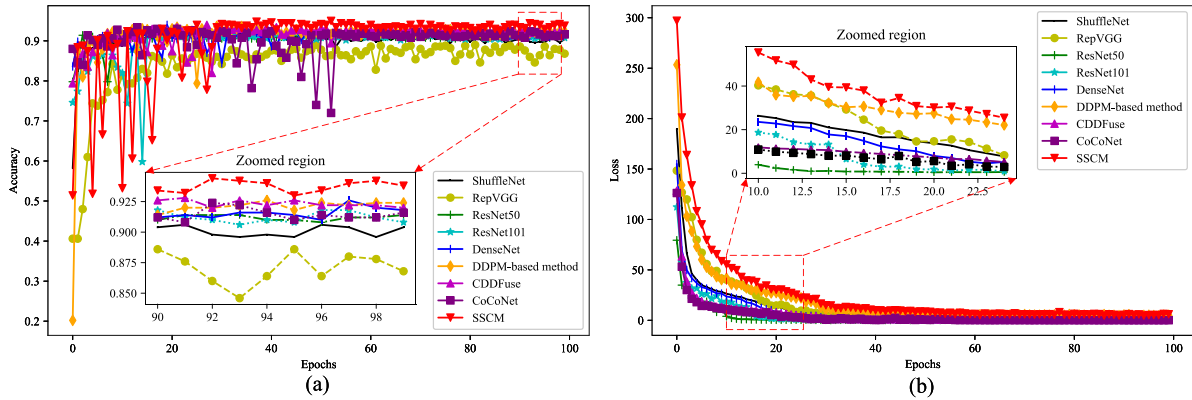


Figure 7. (a) Comparative analysis of model accuracy among various models, (b) Training loss.

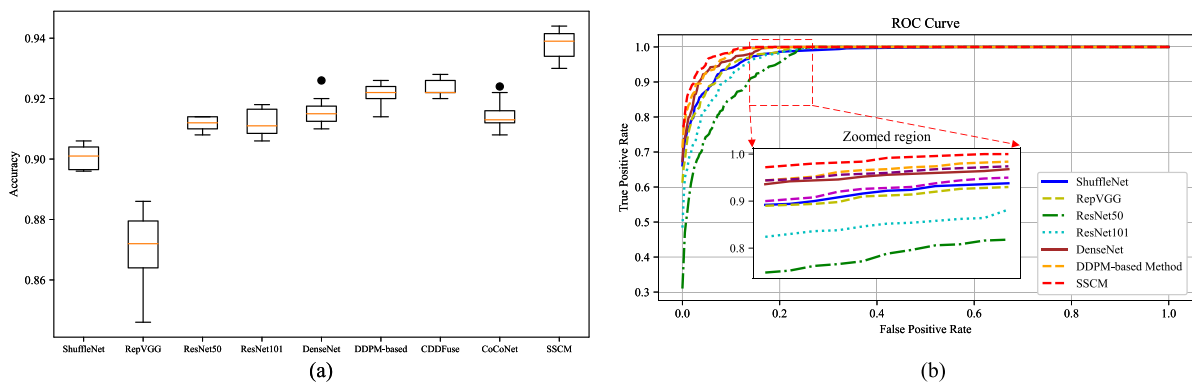


Figure 8. (a) Accuracy comparison through box plots, (b) Evaluating various models with ROC curves.

Table 3. Accuracy comparison of various models after 100 epochs of training.

Models	ShuffleNet	RepVGG	ResNet50	ResNet101	DenseNet	DDPM	CDDFuse	CoCoNet	SSCM
Accuracy	0.904	0.867	0.914	0.908	0.918	0.924	0.920	0.916	0.937

provides a detailed display of the accuracies of different methods in figure 8(a) when trained for 100 Epochs. This enhanced performance is attributed to the incorporation of feature manifold learning, augmenting diagnostic capabilities by integrating synthetic and real field-testing samples. Examining the detailed view in figure 7(a), it is observed that ShuffleNet and RepVGG exhibit a broader range of test accuracies, contrasting with ResNet50, DenseNet, and the proposed SSCM, which demonstrate minimal variation in their results, as indicated by the box plot in figure 8(a). When evaluating various feature fusion methodologies, the proposed SSCM demonstrates significantly enhanced performance in integrating features from both real and synthetically generated samples. Specifically, while CDDFuse and CoCoNet achieve average accuracy rates of 92.3% and 91.4% respectively, our SSCM approach exhibits superior fusion capabilities, as evidenced by the comparative analyses presented in figure 7(a) and further substantiated by the ROC curves and accuracy distributions illustrated in figure 8(a). Furthermore, these empirical results underscore the robust nature of SSCM’s feature integration mechanisms, particularly in maintaining semantic

consistency while preserving spatial relationships between the input samples. The receiver operating characteristic (ROC) [46] curve serves as a tool to assess the effectiveness of classification models by analyzing their shape and the area under the curve (AUC). An examination of the ROC curve, as presented in figure 8(b), reveals that the proposed model attains a higher AUC. This observation suggests an enhanced capability of the proposed model to distinguish accurately between the true and false categories of features derived from CFCS.

The t-SNE [22, 24] dimensionality reduction technique serves as a valuable tool for the visual analysis of latent features encapsulated by various models, as exhibited in figures 9 and 10. This approach significantly improves the comprehension of data patterns and streamlines the identification of clusters, consequently supporting the decision-making process. Meanwhile, these observations are quantitatively validated using confusion matrices in figures 11 and 12. In the case of models like ShuffleNet and RepVGG, as depicted in figures 9(a) and (b), clear separation among scatter points representing intact conditions signals precise

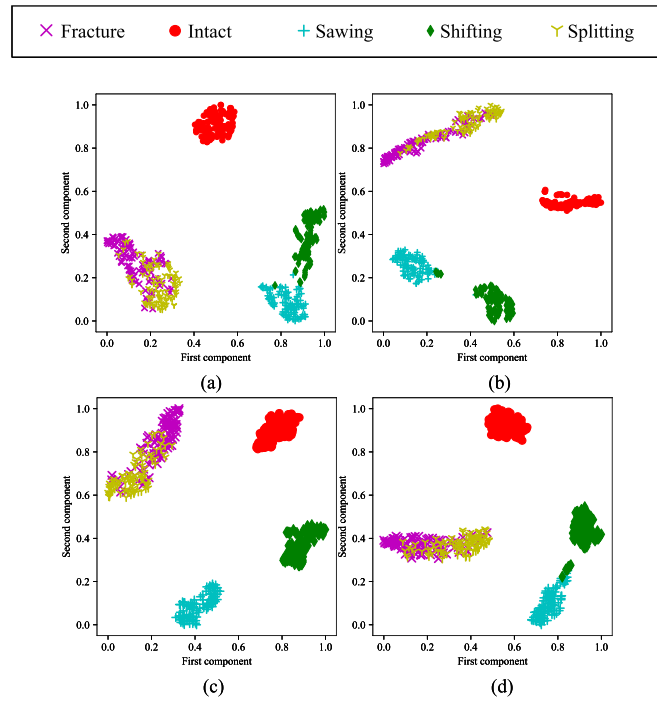


Figure 9. Visualizing features with t-SNE: (a) ShuffleNet, (b) RepVGG, (c) ResNet50, (d) ResNet101.

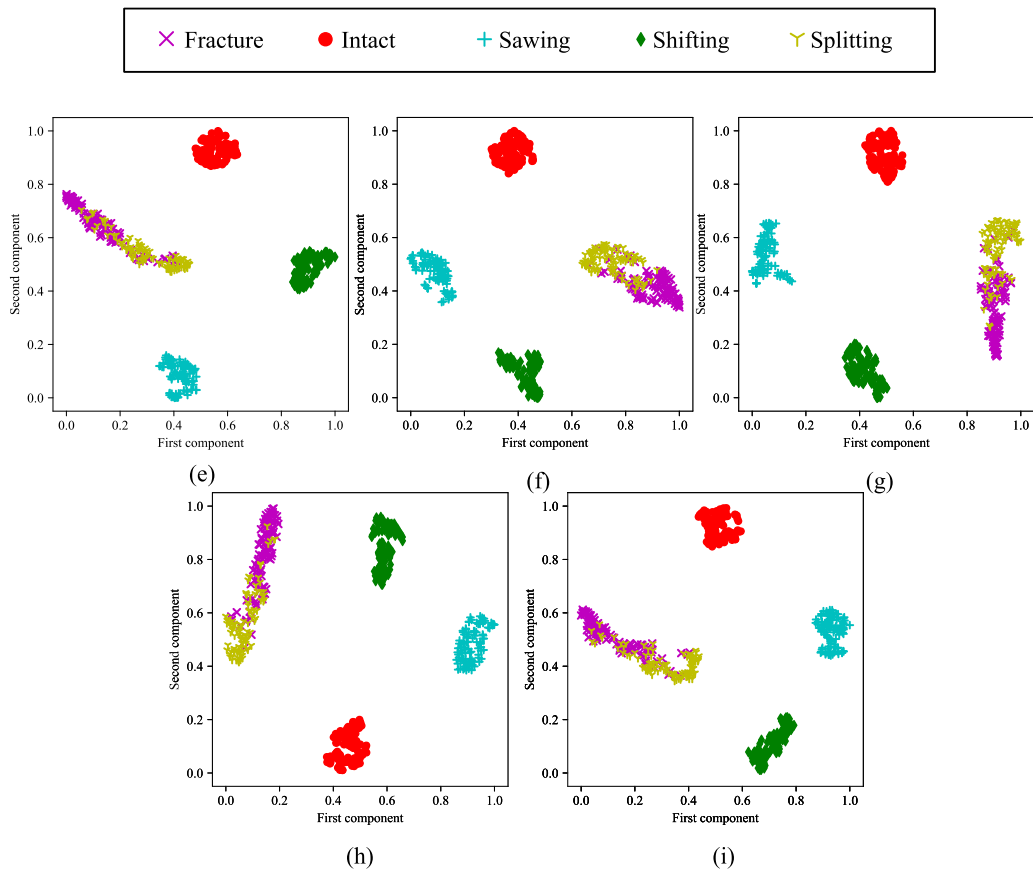


Figure 10. Visualizing features with t-SNE: (e) DenseNet, (f) DDPM-based method, (g) SSCM, (h) CDDFuse, (i) CoCoNet.

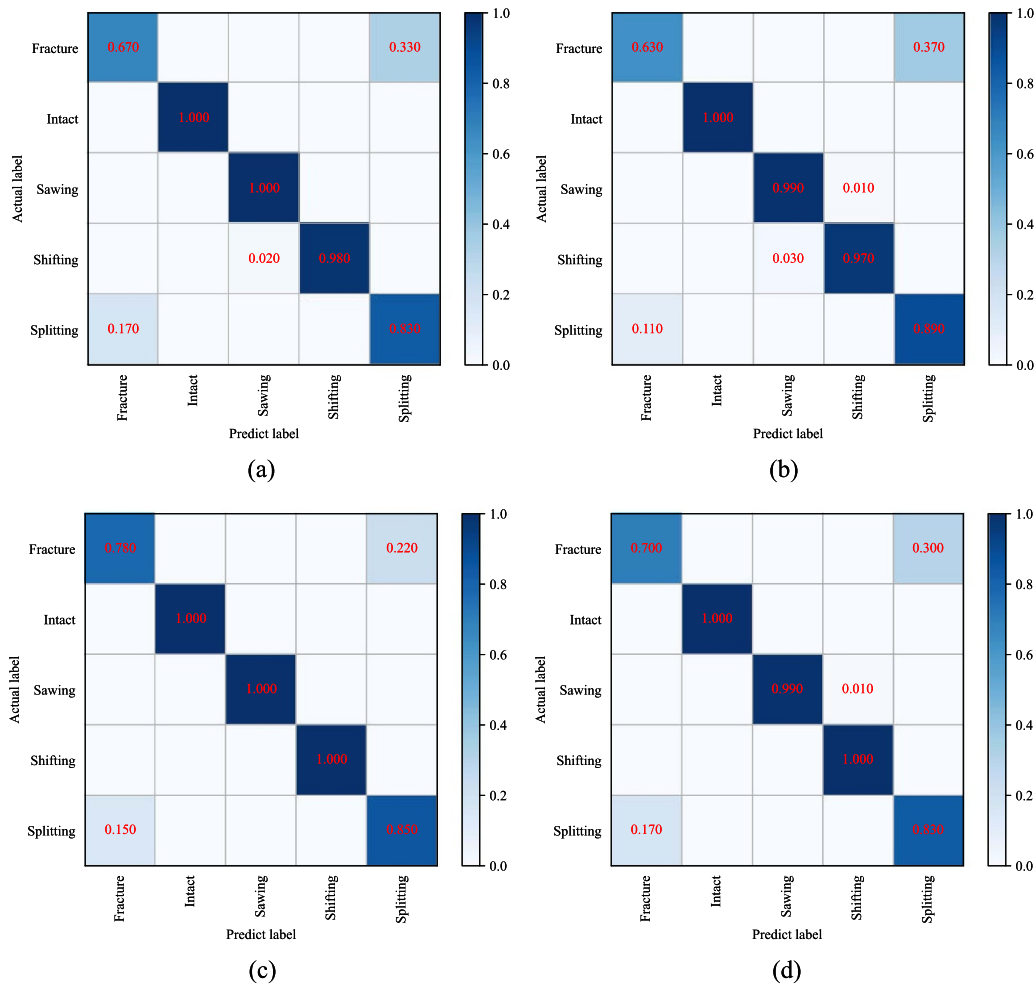


Figure 11. Evaluating classification performance with a Confusion Matrix: (a) ShuffleNet, (b) RepVGG, (c) ResNet50, (d) ResNet101.

recognition capabilities. Conversely, the overlapping scatter points for fracture and splitting modes reveal the models’ limitations in differentiating between these two states. Similarly, the ResNet-based models, illustrated in figures 9(c) and 10(d), demonstrate pronounced overlaps among failure manifestations, highlighting their struggle to discern critical patterns effectively. The feature visualization results obtained through t-SNE dimensionality reduction are presented in figures 10(h) and (i) for the CDDFuse and CoCoNet-based models, respectively. Upon analysis, it becomes evident that a substantial proportion of data points belonging to distinct categories exhibits significant overlap in the feature space. Consequently, this pronounced intersection of feature distributions poses a considerable challenge in establishing clear decision boundaries between fracture and splitting fault classifications, thereby potentially compromising the models’ discriminative capabilities in fault type identification. The proposed SSCM, distinctively outlined in figures 10(g) and 12(g), shows a proficient separation of manifold distributions for nearly all failure types, thus evidencing its superior ability in accurately differentiating between the various conditions. Nonetheless, recognizing splitting and fracture failures remains a pervasive challenge across all examined models, possibly attributable

to the inherent complexity within the data set. In summary, the application of feature manifold visualization offers valuable perspectives on the comparative performance of the models, with particular emphasis on the proposed model’s adeptness at capturing and interpreting latent features with higher precision.

During the ablation study, we systematically adjusted the proportion of synthetic samples in the training set. Specifically, we began with a base case in which synthetic samples made up 20% of the total training samples. Subsequently, we incrementally raised this proportion in 20% increments, ultimately reaching 100%. Throughout these experiments, we kept the model architectures consistent, including the number of layers and neurons in each layer. The training parameters, such as a learning rate of 0.0001, the number of epochs set at 50, and a batch size of 20, remained unchanged. Learning rate schedulers were used to adjust the learning rate adaptively during training to prevent overfitting and accelerate convergence. Additionally, data pre-processing steps, like segmentation, straightening, and gray-level normalization, were uniformly applied to all datasets. This was done to isolate the impact of the synthetic sample proportion on the model’s classification accuracy.

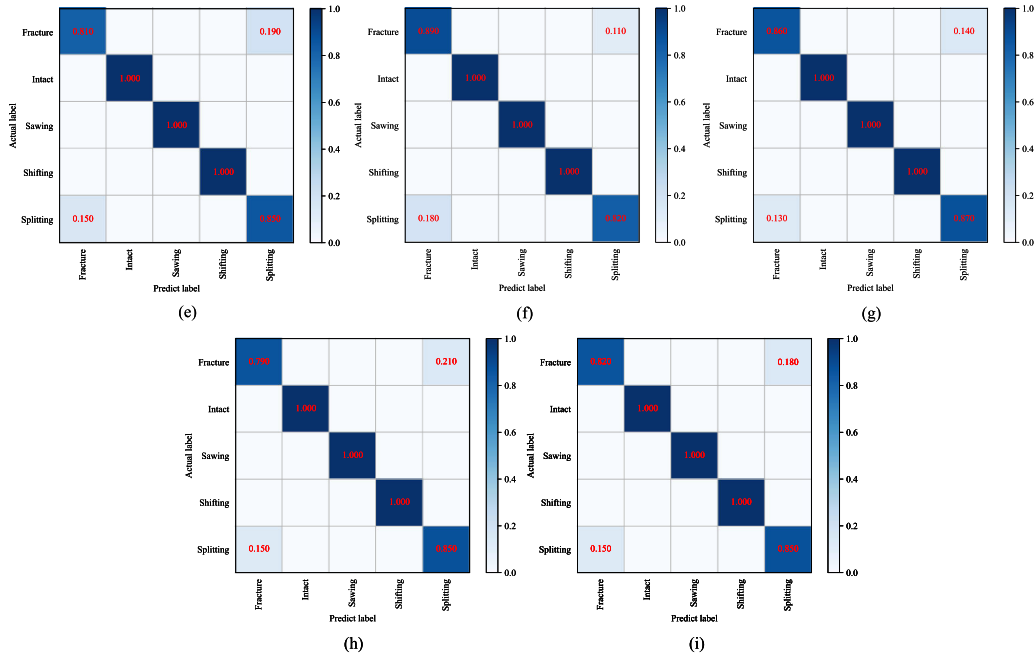


Figure 12. Evaluating classification performance with a Confusion Matrix: (e) DenseNet, (f) DDPM-based method, (g) SSCM, (h) CDDFuse, (i) CoCoNet.

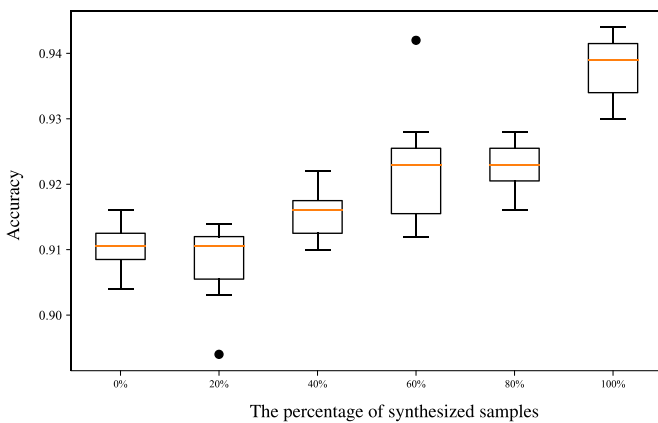


Figure 13. A box plot illustrating results from the ablation experiment.

As depicted in figure 13, when the proportion of synthesized samples increased from 20% to 100%, the model’s classification accuracy for detecting different types of defects, such as fractures and sawing, also increased steadily. Specifically, figure 13 demonstrates that the model achieves its optimal average accuracy of 94.0%, with the maximum accuracy reaching 96.0% when the sample is composed entirely of synthetic data. In contrast, when synthesized samples account for only 20% of the total sample, the accuracy notably declines, reaching a minimum of 89.6%. This indicates that the increase in synthetic samples enhanced the model’s overall performance in defect detection.

Other metrics like recall and F1-score were also examined for the overall training, as shown in table 4. The overall recall improved from 0.919 when there were 0% synthetic samples

Table 4. Table of F1-Score and Recall by Synthetic Sample Proportion.

Proportion	0%	20%	40%	60%	80%	100%
F1-Score	0.919	0.923	0.927	0.943	0.945	0.952
Recall	0.919	0.924	0.928	0.944	0.946	0.953

in the training data to 0.953 when the training data consisted of 100% synthetic samples. The F1 - score also increased during this process. These improvements provide more evidence of the positive impact of synthetic samples on the model’s performance in the overall training. Compared to other studies, our model benefits significantly from synthetic samples. It outperforms many models that face challenges caused by limited data, demonstrating its superiority in handling the scarcity of data during the training process.

To comprehensively evaluate the impact of lower-similarity synthesized data (average cosine similarity of 0.95) on model performance across four distinct defect types, we conducted a systematic analysis using Confusion Matrix methodology. This quantitative assessment examined classification effectiveness across a spectrum of synthetic sample proportions ranging from 0% to 100%, with detailed results presented in figure A2. When utilizing exclusively authentic measurement data (0% synthetic samples), the model established baseline classification accuracies of 85% and 82% for fracture and splitting defects, respectively. However, upon introducing synthetic samples, we observed notable fluctuations in classification performance. Specifically, at a 20% synthetic sample proportion, the model demonstrated diminished performance with classification accuracies of 72%, 76%, and 92% for fracture, splitting, and sawing defects,

respectively. Furthermore, our longitudinal analysis revealed that increasing the proportion of synthetic samples did not yield the anticipated progressive enhancement in model performance. Even when employing a completely synthetic dataset (100% synthetic samples), the classification accuracy showed only marginal improvements, reaching 87% for fracture detection and 86% for splitting defect identification. These empirical findings demonstrate that the incorporation of lower-similarity synthesized data failed to substantially enhance the model's discriminative capabilities. Moreover, the observed performance patterns across various synthetic data ratios indicate that synthetic data augmentation did not contribute meaningfully to the overall model effectiveness.

6. Conclusion

This paper represents an advancement in the detection of defects within CFCS-cored ACCC wires, which are crucial components within NDT industries. The introduced SSCM employs innovative consistency strategies to address the limitations imposed by the scarcity of annotated failure datasets, a prevalent issue in this field. By synthesizing data and integrating transformer-based feature fusion with a DenseNet architecture-based detection module, the model offers a robust solution that effectively compensates for the shortage of real-world samples. The promising results of experiments conducted in authentic field settings validate the efficacy of combining synthetic with real samples. This integration not only enriches the training dataset but also enhances the model's precision in identifying damage.

Key strengths of this study include its use of synthetic data generation to overcome data insufficiency, the incorporation of a transformer-based feature fusion process to improve detection accuracy, and its practical applicability demonstrated in real-world conditions. However, the study also has limitations, including its reliance on the quality of synthetic data, a narrow focus on CFCS-cored ACCC wires, and potential challenges in scaling the approach to broader applications. Future research could explore broader applicability by

extending the SSCM to various types of composite materials and defects beyond CFCS-cored ACCC wires, thereby increasing its relevance and utility within the NDT industry.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

Partial funding for this research has been provided by several sources, including the National Natural Science Foundation of China under Grant 52105111, and 52305085, the Guangdong Basic and Applied Basic Research Foundation under Grant 2025A1515012256, the Shantou University (STU) Scientific Research Initiation Grant under Grant NTF21029, the China Postdoctoral Science Foundation under Grant 2023M740021, and the Natural Science Foundation of Anhui Province under Grant 2108085QE229.

Appendix. Supplementary information on experimental data and analysis

This appendix offers supplementary details regarding the experimental data and analyses conducted in this study.

Table A1. Cosine Similarities of Synthetic-Real Sample Pairs with the Lowest Similarity for Different Types of Defects.

Type of defect	Cosine similarity
Fracture	0.9434 792
Sawing	0.9859 342
Shifting	0.9308 711
Splitting	0.9760 003

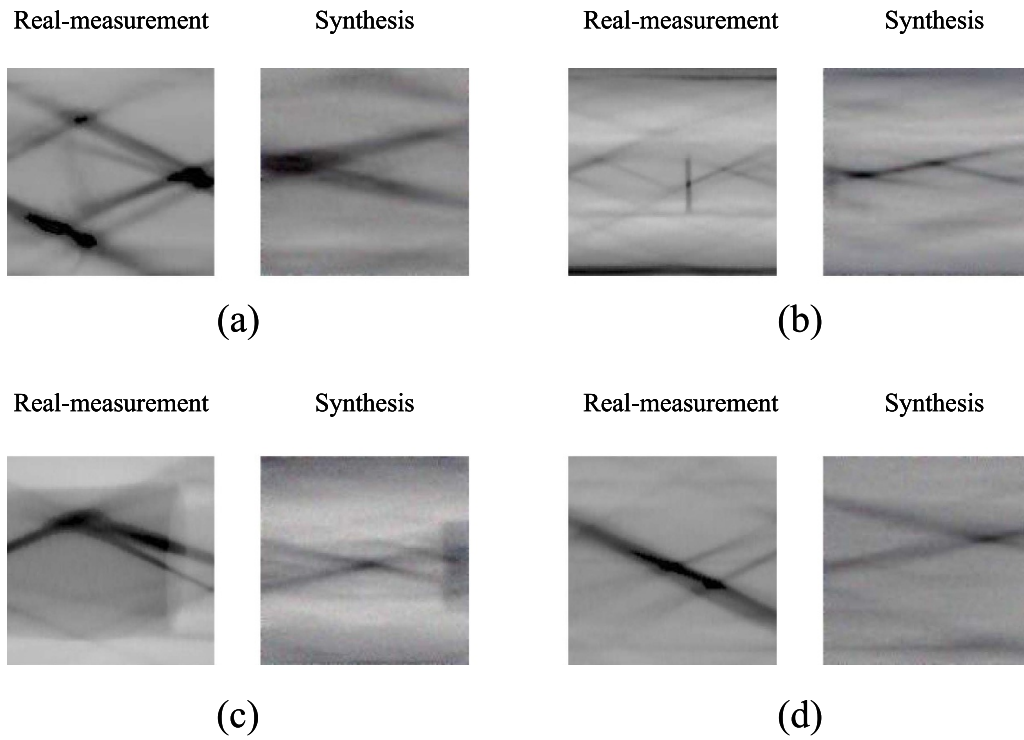


Figure A1. Images of synthetic—real sample pairs with the lowest cosine similarities for different types of defects: (a) Fractured, (b) Sawed, (c) Shifted, (d) Splitting.

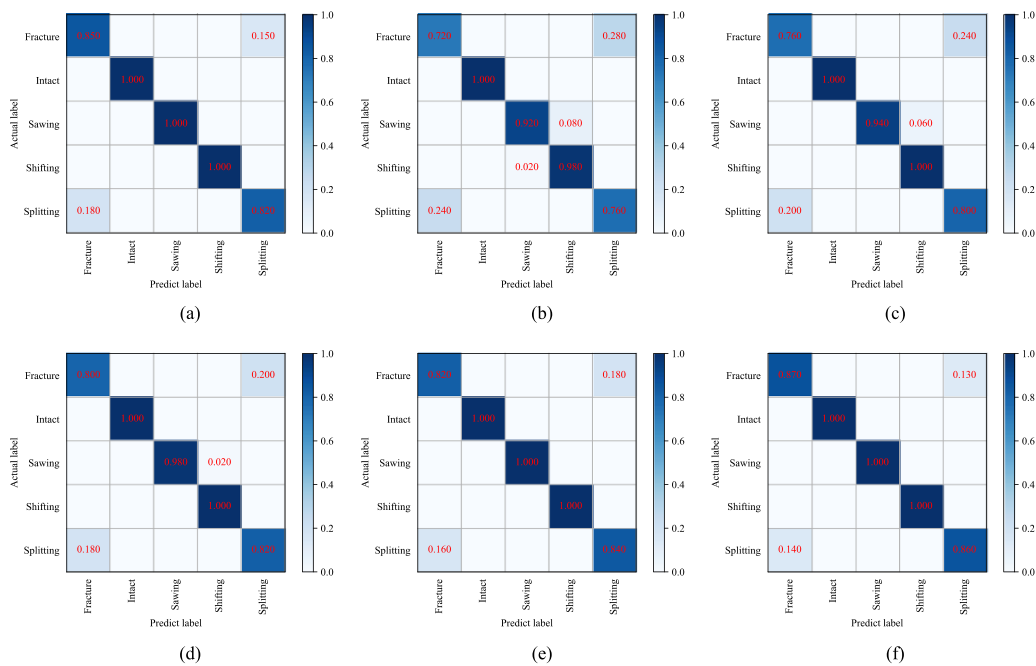


Figure A2. Confusion matrices for synthetic samples with low cosine similarity to real data under different percentages : (a) 0%, (b) 20%, (c) 40%, (d) 60%, (e) 80%, (f) 100%.

ORCID iDs

Peng Chen  <https://orcid.org/0000-0002-3265-3079>
 Changbo He  <https://orcid.org/0000-0003-4180-5334>

References

- [1] Sharma P, Saurav S and Singh S 2024 Object detection in power line infrastructure: a review of the challenges and solutions *Eng. Appl. Artif. Intell.* **130** 107781
- [2] Wang F et al 2023 Internal defect detection of overhead aluminum conductor composite core transmission lines with an inspection robot and computer vision *IEEE Trans. Instrum. Meas.* **72** 3512516
- [3] Chen P, Xu C, Ma Z and Jin Y 2023 A mixed samples-driven methodology based on denoising diffusion probabilistic model for identifying damage in carbon fiber composite structures *IEEE Trans. Instrum. Meas.* **72** 3513411
- [4] Yücel M K, Legg M, Kappatos V and Gan T-H 2017 An ultrasonic guided wave approach for the inspection of overhead transmission line cables *Appl. Acoust.* **122** 23–34
- [5] Chen P et al 2025 Scale-aware domain adaptation for surface defects detection on machine tool components in contaminant measurements *IEEE Trans. Instrum. Meas.* **1–9**
- [6] Zhang H, Sfarra S, Sarasini F, Ibarra-Castanedo C, Perilli S, Fernandes H, Duan Y, Peeters J, Avdelidis N P and Maldague X 2017 Optical and mechanical excitation thermography for impact response in basalt-carbon hybrid fiber-reinforced composite laminates *IEEE Trans. Ind. Inform.* **14** 514–22
- [7] Wang L 2017 Electromagnetic induction holography imaging for stroke detection *J. Opt. Soc. Am. A* **34** 294–8
- [8] Cheng L and Tian G Y 2011 Surface crack detection for carbon fiber reinforced plastic (CFRP) materials using pulsed eddy current thermography *IEEE Sens. J.* **11** 3261–8
- [9] Chen P, Wu Y, Xu C, Jin Y and Zhou C 2024 Markov modeling of signal condition transitions for bearing diagnostics under external interference conditions *IEEE Trans. Instrum. Meas.* **73** 1–8
- [10] de Santos Sierra A, Sanchez Avila C S, Casanova J G and Del Pozo G B 2011 A stress-detection system based on physiological signals and fuzzy logic *IEEE Trans. Ind. Electron.* **58** 4857–65
- [11] Chen P, Wu Y, Fan S, He C, Jin Y, Qi J and Zhou C 2025 Adaptive signal regime for identifying transient shifts: a novel approach toward fault diagnosis in wind turbine systems *Ocean Eng.* **325** 120798
- [12] Fan Y, Kong P, Qi H, Liu H and Ji X 2018 A surface acoustic wave response detection method for passive wireless torque sensor *AIP Adv.* **8** 015321
- [13] Chen P et al 2025 Interference suppression of nonstationary signals for bearing diagnosis under transient noise measurements *IEEE Trans. Reliab.* **1–15**
- [14] Gudmundsson I and Falco G 2022 Porting computer vision models to the edge for smart city applications: enabling autonomous vision-based power line inspection at the smart grid edge for unmanned aerial vehicles (UAVS) 1–10
- [15] Lv F, Zhang P, Tang Z, Yue Y and Yang K 2019 A guided wave transducer with sprayed magnetostriuctive powder coating for monitoring of aluminum conductor steel-reinforced cables *Sensors* **19** 1550
- [16] Song Z, Huang X, Ji C and Zhang Y 2023 Deformable YOLOX: detection and rust warning method of transmission line connection fittings based on image processing technology *IEEE Trans. Instrum. Meas.* **72** 1–21
- [17] Chen Y, Liu H, Chen J, Hu J and Zheng E 2023 INSU-YOLO: an insulator defect detection algorithm based on multiscale feature fusion *Electronics* **12** 3210
- [18] Wang Q, Guo C, Zhao Z, Wang Y, Hu L and Zhai Y 2023 Sirm: an iterative reasoning network for transmission lines based on scene prior knowledge *Eng. Appl. Artif. Intell.* **125** 106656
- [19] Abdelfattah R, Wang X and Wang S P 2023 Generative adversarial networks for power-line segmentation in aerial images *IEEE Trans. Image Process.* **32** 6248–59
- [20] Chen P, Li Y, Wang K, Zuo M J, Heyns P S and Baggeröhr S 2021 A threshold self-setting condition monitoring scheme for wind turbine generator bearings based on deep convolutional generative adversarial networks *Measurement* **167** 108234
- [21] Chen P, Zhang R, Fan S, Guo J and Yang X 2024 Step-wise contrastive representation learning for diagnosing unknown defective categories in planetary gearboxes *Knowl.-Based Syst.* **309** 112863
- [22] Chen P, Li Y, Wang K and Zuo M J 2021 An automatic speed adaption neural network model for planetary gearbox fault diagnosis *Measurement* **171** 108784
- [23] Chen P, Ma Z, Xu C, Jin Y and Zhou C 2024 Self-supervised transfer learning for remote wear evaluation in machine tool elements with imaging transmission attenuation *IEEE Internet Things J.* **11** 23045–54
- [24] Chen P, Li Y, Wang K and Zuo M J 2020 A novel knowledge transfer network with fluctuating operational condition adaptation for bearing fault pattern recognition *Measurement* **158** 107739
- [25] Gong Y, Luo J, Shao H and Li Z 2022 A transfer learning object detection model for defects detection in x-ray images of spacecraft composite structures *Compos. Struct.* **284** 115136
- [26] Gao Y, Song G, Li S, Zhen F, Chen D and Song A 2020 Linespyx: a power line inspection robot based on digital radiography *IEEE Robot. Autom. Lett.* **5** 4759–65
- [27] Wei R et al 2020 Defect detection for aluminium conductor composite core x-ray image with deep convolution network *J. Phys.: Conf. Ser.* **1633** 012166
- [28] Wang J et al 2021 Seesaw loss for long-tailed instance segmentation pp 9695–704
- [29] Zhai Y, Yang X, Wang Q, Zhao Z and Zhao W 2021 Hybrid knowledge R-CNN for transmission line multifitting detection *IEEE Trans. Instrum. Meas.* **70** 1–12
- [30] Sohl-Dickstein J, Weiss E, Maheswaranathan N and Ganguli S 2015 *Deep unsupervised learning using nonequilibrium thermodynamics* (PMLR) pp 2256–65
- [31] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2020 Generative adversarial networks *Commun. ACM* **63** 139–44
- [32] Bond-Taylor S, Hessey P, Sasaki H, Breckon T P and Willcocks C G 2022 *Unleashing Transformers: Parallel Token Prediction With Discrete Absorbing Diffusion for Fast High-Resolution Image Generation From Vector-Quantized Codes* (Springer) pp 170–188
- [33] Ho J, Jain A and Abbeel P 2020 Denoising diffusion probabilistic models *Advances in Neural Information Processing Systems* vol 33 pp 6840–51
- [34] Song Y, Dhariwal P, Chen M and Sutskever I 2023 Consistency models (arXiv:2303.01469)
- [35] Song Y et al 2020 Score-based generative modeling through stochastic differential equations (arXiv:2011.13456)
- [36] Karras T, Aittala M, Aila T and Laine S 2022 Elucidating the design space of diffusion-based generative models *Advances in Neural Information Processing Systems* vol 35 pp 26565–77

- [37] Li H, Wu X-J and Kittler J 2021 RFN-Nest: an end-to-end residual fusion network for infrared and visible images *Inf. Fusion* **73** 72–86
- [38] Cheng X, Ma G, Wu Z, Zu H and Hu X 2023 Automatic defect depth estimation for ultrasonic testing in carbon fiber reinforced composites using deep learning *NDT&E Int.* **135** 102804
- [39] Zhang X, Zhou X, Lin M and Sun J 2018 Shufflenet: an extremely efficient convolutional neural network for mobile devices pp 6848–56
- [40] Xu J et al 2022 Regnet: self-regulated network for image classification *IEEE Trans. Neural Networks Learn. Syst.*
- [41] O’Neill D, Xue B and Zhang M 2021 Evolutionary neural architecture search for high-dimensional skip-connection structures on densenet style networks *IEEE Trans. Evol. Comput.* **25** 1118–32
- [42] Ding X et al 2021 Repvgg: making VGG-style convnets great again pp 13733–42
- [43] Yu J, Shen Z and Zheng X 2020 Joint feature and label adversarial network for wafer map defect recognition *IEEE Trans. Autom. Sci. Eng.* **18** 1341–53
- [44] Liu J, Lin R, Wu G, Liu R, Luo Z and Fan X 2024 Coconet: Coupled contrastive learning network with multi-level feature ensemble for multi-modality image fusion *Int. J. Comput. Vis.* **132** 1748–75
- [45] Zhao Z et al 2023 CDDFUSE: correlation-driven dual-branch feature decomposition for multi-modality image fusion pp 5906–16
- [46] Chang C-I 2020 An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis *IEEE Trans. Geosci. Remote Sens.* **59** 5131–53