Damage imaging in structural health monitoring with fine-tuned conditional diffusion model

Yang, Xin; Cantero-Chinchilla, Sergio; Moradi, Morteza; Komninos, Panagiotis; Fang, Chen; Liao, Yunlai; Kundu, Pradeep; Zarouchas, Dimitrios; Chronopoulos, Dimitrios

DOI
10.1016/j.ymssp.2025.112996

Publication date
2025

Document Version
Final published version

Published in
Mechanical Systems and Signal Processing

Citation (APA)
Yang, X., Cantero-Chinchilla, S., Moradi, M., Komninos, P., Fang, C., Liao, Y., Kundu, P., Zarouchas, D., & Chronopoulos, D. (2025). Damage imaging in structural health monitoring with fine-tuned conditional diffusion model. *Mechanical Systems and Signal Processing*, *236*, Article 112996. https://doi.org/10.1016/j.ymssp.2025.112996

Mechanical Systems and Signal Processing 236 (2025) 112996

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Full length article

Damage imaging in structural health monitoring with fine-tuned conditional diffusion model

Xin Yang [a,*], Sergio Cantero-Chinchilla [b], Morteza Moradi [c], Panagiotis Komninos [c], Chen Fang [a,d], Yunlai Liao [a,e], Pradeep Kundu [a], Dimitrios Zarouchas [c], Dimitrios Chronopoulos [a]

[a] Department of Mechanical Engineering & Division of Mechatronic System Dynamics (LMSD), KU Leuven, Belgium
[b] School of Electrical, Electronic and Mechanical Engineering, University of Bristol, Bristol, BS8 1TR, United Kingdom
[c] Center of Excellence in Artificial Intelligence for Structures, Prognostics & Health Management, Aerospace Engineering Faculty, Delft University of Technology, Kluyverweg 1, Delft, 2629 HS, The Netherlands
[d] Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia
[e] School of Aerospace Engineering, Xiamen University, Xiamen, Fujian 361005, PR China

ARTICLE INFO

Communicated by S.D. Rosa

Dataset link: https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/, https://huggingface.co/stabilityai/stable-diffusion-v2, https://huggingface.co/laion/CLIP-ViT-H-14-laion2B-s32B-b79K

Keywords:
Structural health monitoring
Damage imaging
Conditional diffusion models
Fine-tuning technique
Delay-and-sum beamforming
Ultrasonic sparse array imaging

ABSTRACT

Damage imaging plays a crucial role in structural health monitoring (SHM) systems for fast and efficient damage assessment. Delay-and-sum (DAS) beamforming is a widely used algorithm in non-destructive testing for damage imaging, but its effectiveness is often compromised by the use of sparse ultrasonic transducer arrays and the difficulty in detecting progressive delamination larger than the wavelength using guided wave-based methods under fatigue loading. Although X-ray imaging offers detailed assessments of progressive delamination, its application is still limited due to the need to interrupt fatigue loading cycles and its high operational cost. To this end, we propose a novel **D**amage **I**maging framework that uses the fine-tuned **C**onditional **D**iffusion **M**odel for SHM systems (DI-CDM). Leveraging the powerful image generation capabilities of diffusion models, the framework was fine-tuned by combining DAS beamforming images derived from ultrasonic sparse array data with X-ray images captured during fatigue loading cycles of the composite structures. The proposed approach can generate damage images that reveal the progression of delamination size in the fatigue loading process. The framework was validated through numerical simulations and experimental data from NASA datasets for composite structures, demonstrating its potential and effectiveness by applying diffusion models in SHM applications to enable fast, high-resolution damage imaging.

1. Introduction

Safety and reliability are key considerations for the operation of industrial systems and equipment, including mechanical components, aircraft, and other engineering structures [1,2]. During service life, damage can occur due to long-term and cyclic loads, which make a significant contribution to the failure of critical structures. Structural health monitoring (SHM) is an effective tool that combines non-destructive testing (NDT) technologies (such as vision inspection [3], infrared thermography or thermal-based methods [4], and ultrasonic testing [5]) to identify, locate and assess the severity of damage with permanently attached or embedded sensing systems to monitor the health status of a structure in service [6,7]. These techniques can generate direct damage imaging,

* Corresponding author.
E-mail address: xin.yang@kuleuven.be (X. Yang).

https://doi.org/10.1016/j.ymssp.2025.112996
Received 14 March 2025; Received in revised form 2 June 2025; Accepted 13 June 2025
Available online 26 June 2025

Nomenclature

α_t	Variance level in diffusion process
η	Trainable parameter in denoising process
$\hat{x}_t$	Latent feature in denoising process
$\mathbb{E}$	Expectation
$\mathbb{R}$	The set of all real numbers
$\mathbf{I}$	Identity matrix
$\mathbf{Q}, \mathbf{K}, \mathbf{V}$	Query, Key, Value matrices
$\mathbf{W}$	Weight matrices
$\mathcal{D}$	Kullback–Leibler divergence
$\mathcal{L}$	Loss function
$\mathcal{N}$	Gaussian distribution
μ_θ	Noise parameter
ω_j	Weighting factor
ϕ	Imaging position
σ_t	Noise level
τ_i	Time delay
θ	Neural network parameters
ε	Gaussian noise
ε_θ	Learnable network
φ	Direction of arrival
A, B	Low rank matrices in LoRA
c_t	Context input
d_k	Dimension of the key matrix
$g_{i,j}$	Time-domain signal
N_p	Number of elements
p	Backward transition function
q	Forward transition function
r	LoRA rank
s	Image intensity
x_t	Latent feature in diffusion process
y	Conditioning inputs

allowing for quick and intuitive assessment of damage severity in the inspected structure [8]. Specifically during service, fast and reliable damage imaging plays an important role in robust and real-time damage assessment without interrupting the operating profile. Thus, advanced algorithms that use NDT techniques in SHM need to be developed to facilitate damage imaging.

Among the existing SHM approaches for damage imaging, guided wave techniques based on ultrasonic testing are widely adopted due to their high sensitivity to possible damage [9]. Piezoelectric transducers (PZTs) commonly used in these ultrasonic tests offer advantages such as high sensitivity, low cost, and compact size. Data from a single transducer are often insufficient for comprehensive damage characterization, hence they are typically arranged in arrays on the surface of the structure. The configuration of the PZT array is usually divided into two types: phased arrays and sparse arrays [10]. The former comprise a linear or planar arrangement of PZT elements, allowing ultrasonic beams to be electronically steered, scanned, swept and focused in multiple directions [11]. The computational load can be relatively high when large amounts of data are captured from phased array PZT elements. The latter, sparse arrays utilize fewer transducer elements while achieving comparable resolution with reduced complexity, making them a practical choice for SHM. Specifically, high-resolution damage imaging from sparse array signals relies on advanced postprocessing algorithms such as delay-and-sum (DAS) beamforming [12], which sums delayed signals to obtain focused receive beams in the preferred direction; total focusing method (TFM) [13], offering high resolution via full matrix capture (FMC) and synthetic focusing at all pixels; and plane wave imaging (PWI) [14], which prioritizes imaging speed through unfocused waves transmitted simultaneously by a subset of transducer elements. These imaging algorithms are still regarded as representative post-processing damage imaging algorithms in NDT. Nevertheless, the application of these algorithms is still limited to *off-line* monitoring, rather than SHM scenarios, which need to provide a fast and efficient damage assessment.

Unlike the aforementioned phased-array imaging approaches, which are often referred to as physics-based imaging approaches, data-driven imaging approaches have gained significant attention in recent years. Deep learning (DL) techniques, in particular, are capable of learning features or latent representations directly from data. However, these approaches typically require large amounts of labeled data, which can be impractical for realistic engineering applications. To address this challenge, researchers

proposed hybrid methods that combine physics-based and data-driven approaches for damage assessment. For instance, in the work of Miorelli et al. [6], a convolutional neural network (CNN) was trained using DAS images with varying damage sizes as input. While this study demonstrated the potential of CNNs for damage localization, it focused solely on static conditions and did not explore dynamic damage imaging. Alternatively, as one of the most popular models in the field of generative artificial intelligence (GenAI), generative adversarial networks (GANs) are introduced due to their strong capabilities for new data generation, which serves as an expansion of the limited database. The network is designed to generate new data with the same statistical characteristics as the training samples. As an example, Luleci et al. [15] used a GAN to generate damage-related vibration data, addressing the challenge of insufficient training datasets. In addition, Cheng et al. [16] trained a GAN on numerical simulation datasets containing various defects and validated it using experimental thermal images recorded during infrared thermography, with the aim of defect segmentation. Although GAN-based generative models can be explored for damage imaging, the adversarial loss in GANs can be slow to converge, often leading to training instability and mode collapse [17,18]. Moreover, most of the present works still focus on using GANs for data augmentation to address data imbalance problems, and then the augmented data can be used further for damage detection or classification.

Recently, the denoising diffusion probabilistic model (DDPM [19], also known as the diffusion model) has stood out as a new paradigm of GenAI, surpassing previous GAN-based models in image generation tasks [17]. Diffusion models can generate two types of images: unconditional image generation and conditional image generation. Unconditional images are usually generated by transforming random noise into representative data without any guidance, while conditional image generation refers to generating images based on specific conditions (images, audio, text, etc.). The conditional diffusion model (CDM) is widely adopted because the image generation in the diffusion model denoising process can produce high-quality and domain-specific images based on given reference information. This capability is particularly valuable in SHM systems, where CDMs can integrate diverse formats of sensor data, including vibration acceleration [20] or images derived from Fourier transformations. However, to date in the field of condition monitoring for composite structures or rotational machines, most studies [21–24] utilize diffusion models for data augmentation to address data imbalance issues, which indeed share some similarities with GAN-based approaches. For damage imaging tasks, it is still challenging to control the diffusion models to generate damage images in the SHM field and achieve the desired imaging results.

In this work, we propose a conditional diffusion model for SHM to achieve fast damage imaging (DI-CDM). First, the X-ray images providing progressive delamination information are utilized as input to the DI-CDM framework. The DAS beamforming, as a foundational time-domain ultrasonic imaging technique, is then applied to ultrasonic sparse array data to generate the initial damage image, which serves as a conditioning input to the DI-CDM. The multimodal fusion of X-ray imaging and ultrasonic DAS beamforming enables the DI-CDM to refine binary X-ray images into high-quality damage images. After fine-tuning, the proposed framework operates without the need for X-ray imaging, producing enhanced damage images directly from pure noise during the inference stage. The main contributions of this work are as follows:

- To the best of our knowledge, this work is the first to apply diffusion models for image generation in the realm of SHM, in particular LoRA fine-tuned conditional diffusion models for damage imaging.
- Multimodal fusion of sparse array ultrasonic guided wave data and X-ray images to enhance damage imaging.
- The fine-tuned DI-CDM model enables fast damage imaging without the need for costly X-ray imaging.

The remainder of this work is organized as follows. Section 2 reviews the theoretical background of DAS beamforming for damage imaging and the mathematical foundations of diffusion models. Section 3 details the architecture of the proposed DI-CDM framework, including its conditional diffusion pipeline and multimodal fusion strategy. Section 4 describes the NASA experiment setup and two evaluation metrics for imaging results comparison. Section 5 presents numerical simulations based on the NASA dataset and compares the imaging results of DAS and DI-CDM imaging using simulation data. Section 6 validates the effectiveness of the DI-CDM framework through experimental datasets. Section 7 provides concluding remarks.

2. Preliminaries

Prior to introducing the DI-CDM framework, it is crucial to summarize the traditional DAS beamforming for damage imaging based on sparse array ultrasonic testing, which lies the foundation of damage imaging techniques. The basic diffusion model is formulated in the following subsection.

2.1. DAS beamforming for damage imaging

DAS beamforming is a widely used technique in damage imaging because of its implicit simplicity. Specifically, the basis of DAS beamforming is to directly reconstruct damage images from recorded ultrasonic guided wave signals received across a sparse array of ultrasonic PZT transducer elements. For each pixel position, DAS beamforming synthesizes the received data at the corresponding time-of-flight (ToF) [25]:

$$s(\phi) = \sum_{j=1}^{N_p} \omega_j(\phi) g_{i,j}(t - \tau_i(\phi)) \tag{1}$$

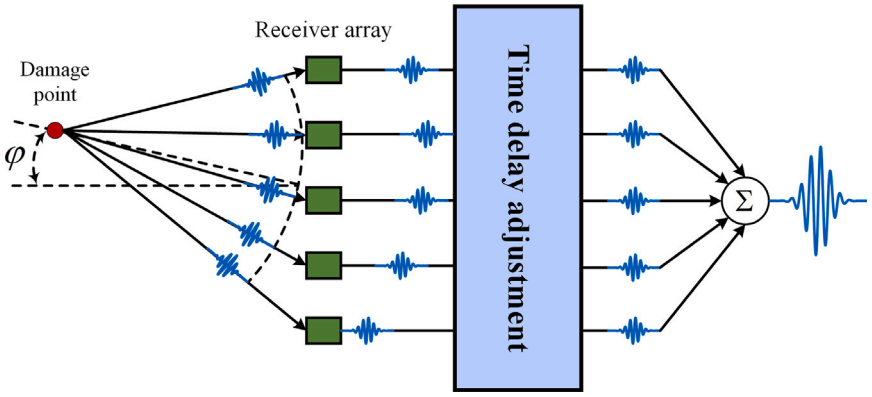

Fig. 1. The generic illustration for DAS beamforming algorithm. The reflected wave scattered by the damage point arrives at different ToF for the linear receiver array. DAS beamforming compensates for these ToF variations by applying a reverse delay to each channel, allowing the delayed signals to be summed effectively.

where $s(\phi)$ denotes the image intensity at the imaging position ϕ, and N_p refers to the number of PZT receiver elements. $\omega_j(\phi)$, $g_{i,j}(t)$, and $\tau_i(\phi)$ are the weighting factors assigned to jth beam shape, the time-domain signal received by the jth receiver (emitted by the ith transmitter), and the time delay at ϕ for the ith transmitter, respectively.

As shown in Fig. 1, DAS beamforming operates by compensating for ToF differences among the waves reflected from the damage point to the receiver array. These ToF differences follow a linear relationship based on the distance of each receiver along the array. The time delay adjustment module in the DAS beamforming compensates for these ToF variations by applying a reverse delay to each receiver signal, such that the highest output signal amplitude can be achieved in the direction φ.

2.2. The fundamentals for diffusion models

Diffusion models implement the forward and backward processes through two Markov chains [19]: The forward diffusion process produces a Markov chain by progressively adding Gaussian noise to the original data, transforming the original clean data into pure noise data. Reversely, the backward denoising process generates a separate Markov chain that reconstructs the desired clean data from these noisy data. The original observed data are denoted as x_0, where the subscript indicates that the data are free from the noise introduced in the diffusion process. The forward Markov chain transforms clean data x_0 to a sequence of noisy data $\{x_1, x_2, \ldots, x_T\}$ over T timesteps [26]:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\alpha_t}x_{t-1}, \alpha_t\mathbf{I}) \tag{2}$$

where q denotes the forward transition probability function. $\{\alpha_t \in (0,1)\}_{t=1}^{T}$ are hyperparameters that indicate the changing variance of the noise level at each timestep, $\mathbf{I}$ refers to the identity matrix, and $\mathcal{N}$ is the general notation for the Gaussian distribution. A notable property of the Gaussian noise added at each timestep is that the x_t can be directly computed from initial x_0 in closed form $q(x_t|x_0)$ [27]:

$$\begin{aligned} q(x_t|x_0) &= \mathcal{N}\left(x_t; \sqrt{\tilde{\alpha}_t}x_0, (1-\tilde{\alpha}_t)\mathbf{I}\right) \\ x_t &= \sqrt{\tilde{\alpha}_t}x_0 + \sqrt{1-\tilde{\alpha}_t}\varepsilon \\ \varepsilon &\sim \mathcal{N}(0, \mathbf{I}) \end{aligned} \tag{3}$$

where $\tilde{\alpha}_t = \prod_{i=1}^{t}(1-\alpha_i)$. Specifically, $\{\tilde{\alpha}_t\}_{t\to T}$ is converged to be approximately 0 such that $q(x_T) \approx \mathcal{N}(x_T; 0, \mathbf{I})$. This result implies that the endpoint of the forward process $q(x_T)$, which also serves as the starting point of the backward process, is approximately a standard normal distribution [28].

The backward denoising process takes the form $p_\theta(\hat{x}_{t-1}|\hat{x}_t)$ as a transition probability function modeled by a neural network parameterized with θ. This process gradually reduces the noise in a sequence of corrupted data $\{x_T, x_{T-1}, \ldots, x_1\}$ to regenerate the clean data $\hat{x}_0$, as described by the Eq. (4) [29,30]:

$$\begin{aligned} p_\theta(\hat{x}_{t-1}|\hat{x}_t) &= \mathcal{N}\left(\hat{x}_{t-1}; \mu_\theta(\hat{x}_t), \Sigma_\theta(\hat{x}_t)\right) \\ p_\theta(\hat{x}_0) &= \int p(\hat{x}_T)\prod_{t=1}^{T} p_\theta(\hat{x}_{t-1}|\hat{x}_t)\mathrm{d}\hat{x}_{1:T} \end{aligned} \tag{4}$$

where μ_θ and Σ_θ represent the mean and variance of the Gaussian distribution, respectively. $\hat{x}_{1:T}$ denotes the sequence $\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_T\}$. The training process of the diffusion model aims to optimize the likelihood objective function, maximizing the probability that $p_\theta(\hat{x}_0)$ can regenerate the observed training data x_0. This training task is accomplished by minimizing the evidence

lower bound (ELBO) of the negative log-likelihood, with the loss function $\mathcal{L}_\theta(\hat{x}_0)$ written as follows [31]:

$$\mathcal{L}_\theta(\hat{x}_0) = \mathbb{E}_q\Big[\underbrace{D_{KL}\big(q(x_T|x_0) \parallel p(\hat{x}_T)\big)}_{\mathcal{L}_T} + \sum_{t>1}\underbrace{D_{KL}\big(q(x_{t-1}|x_t,x_0) \parallel p_\theta(\hat{x}_{t-1}|\hat{x}_t)\big)}_{\mathcal{L}_{t-1}} - \underbrace{\log p_\theta(\hat{x}_0|\hat{x}_1)}_{\mathcal{L}_0}\Big] \tag{5}$$

where $\mathbb{E}$ denotes the expectation and D_{KL} represents the Kullback–Leibler divergence. $\mathcal{L}_T$ and $\mathcal{L}_0$ are the prior loss and reconstruction loss, which are fixed during the training process. Thus, the whole loss function that needs to be optimized is $\mathcal{L}_{t-1}$. Using Bayes theorem, the $q(x_{t-1}|x_t,x_0)$ term in the $\mathcal{L}_{t-1}$ can be rewritten as [32]:

$$\begin{aligned} q(x_{t-1}|x_t,x_0) &= \mathcal{N}\big(x_{t-1};\mu(x_t,x_0),\tilde{\beta}_t\mathbf{I}\big)\\ \mu(x_t,x_0) &= \frac{1}{\sqrt{\alpha_t}}\Big(x_t - \frac{\alpha_t}{\sqrt{1-\tilde{\alpha}_t}}\varepsilon\Big)\\ \tilde{\beta}_t &= \frac{\alpha_t(1-\tilde{\alpha}_{t-1})}{1-\tilde{\alpha}_t} \end{aligned} \tag{6}$$

Given Eqs. (4) and (6), $\mathcal{L}_{t-1}$ can be further simplified to:

$$\mathcal{L}_{t-1} = \mathbb{E}_q\Big[\frac{1}{2\sigma_t^2} \parallel \mu(x_t,x_0) - \mu_\theta(\hat{x}_t) \parallel^2\Big] + \mathrm{C} \tag{7}$$

where C is a constant and σ_t controls the noise level during the denoising process. Previous works [19,30] suggest that the noise parameter μ_θ can be parameterized by:

$$\begin{aligned} \mu_\theta(\hat{x}_t) &= \frac{1}{\sqrt{1-\alpha_t}}\Big(\hat{x}_t - \eta(t)\varepsilon_\theta(\hat{x}_t)\Big)\\ \eta(t) &= \frac{\alpha_t}{\sqrt{1-\tilde{\alpha}_t}} \end{aligned} \tag{8}$$

where ε_θ is a learnable network that predicts the noise ε corresponding to input $\hat{x}_t$ at tth time step in backward process. Then, with Eqs. (6) and (8), the loss function Eq. (7) can be written as:

$$\begin{aligned} \mathcal{L}_{t-1} &= \mathbb{E}_{t,\hat{x}_t,\varepsilon}\Big[\frac{\alpha_t^2}{2\sigma_t^2(1-\alpha_t)(1-\tilde{\alpha}_t)} \parallel \varepsilon - \varepsilon_\theta(\hat{x}_t) \parallel^2\Big]\\ &\propto \mathbb{E}_{t,\hat{x}_t,\varepsilon}\Big[\parallel \varepsilon - \varepsilon_\theta(\hat{x}_t) \parallel^2\Big] \end{aligned} \tag{9}$$

Thus, once training is finished, a chain of samples are removed with noise beginning from the corrupted data $\hat{x}_T \sim \mathcal{N}(0,\mathbf{I})$ to clean data $\hat{x}_0$ by the following equation [28,33]:

$$\begin{aligned} \hat{x}_t &\leftarrow \frac{\hat{x}_{t+1} - \eta(t+1)\varepsilon_\theta(\hat{x}_{t+1})}{\sqrt{1-\alpha_{t+1}}} + \sigma_t\nu\\ t &= T-1, T-2, \ldots, 0 \end{aligned} \tag{10}$$

where $\nu \sim \mathcal{N}(0,\mathbf{I})$ and σ_t controls the noise level during the denoising process.

3. Fine-tuned conditional diffusion model for damage imaging

This section provides a detailed explanation of damage imaging within the DI-CDM framework, which is designed for efficient and high-quality damage imaging. To adapt diffusion models for damage imaging, additional conditional guidance is incorporated into the proposed framework. These conditions constrain the model to generate physically consistent damage images. In this work, DAS imaging and the number of cycles over the lifetime of the composite structures serve as conditioning inputs, providing initial damage location and cyclic loading information to guide the diffusion model for damage imaging.

In Fig. 2, the proposed DI-CDM damage imaging framework is illustrated, which utilizes DAS images derived from ultrasonic sparse array signals captured during the fatigue loading experiment as conditioning inputs to the diffusion model. The inputs to the DI-CDM framework are the X-ray images, which are then fed into a pretrained encoder network for feature extraction. Gaussian noise ε is added to these encoded features to produce the final pure noise. During the backward denoising process, a trainable UNet ε_θ learns the noise distribution and progressively denoises the features, transforming them from pure Gaussian noise to the desired latent features, conditioned on the DAS image. Finally, a pretrained decoder network generates the damage image from these denoised latent features. It is important to note that the fatigue loading cycle, denoted as c_t, also serves as a conditioning context input to the DI-CDM framework. This cycling context provides the model with information about the specific loading stage at which the X-ray images were captured. For simplicity, the symbol c_t is omitted from subsequent mathematical formulations, as it can be regarded as supplementary combined with DAS conditioning input. That is, both DAS image and cycling information collectively constitute the conditioning inputs y.

In standard diffusion models, the core objective is to learn the parameterized approximation of the reverse transition probability $p_\theta(\hat{x}_{t-1}|\hat{x}_t)$ presented in Eq. (4). Conditional diffusion models extend this framework by conditioning the transition probability on

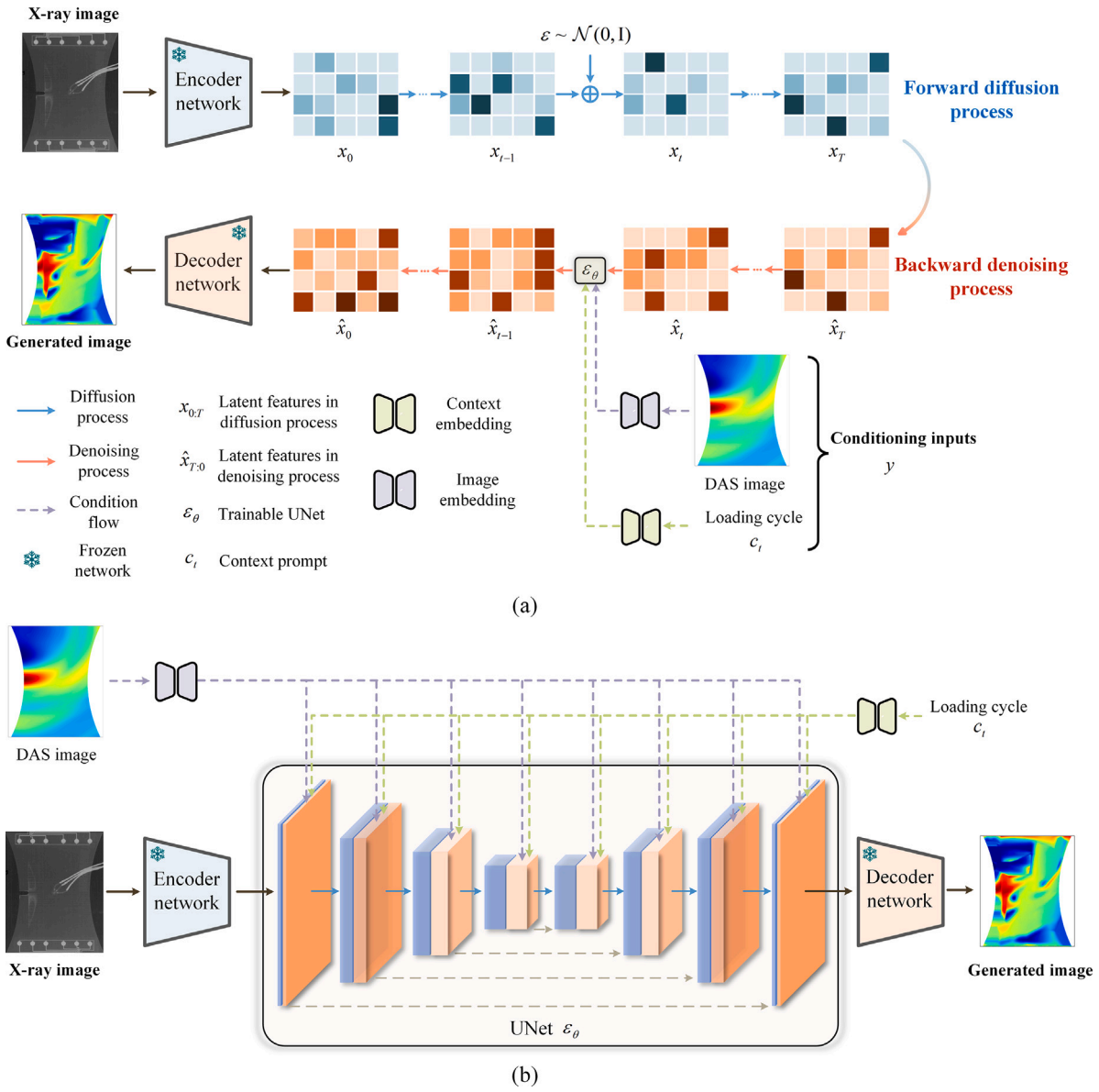

Fig. 2. The proposed DI-CDM framework for damage imaging. (a) The DI-CDM framework leverages DAS images and cycling text prompt as conditioning inputs to guide the denoising process. (b) The multimodal fusion mechanism integrates ultrasonic with X-ray modalities within the UNet ε_θ.

an additional input y, resulting in $p_\theta(\hat{x}_{t-1}|\hat{x}_t, y)$. Then, the loss function and denoising sample will be adapted from Eqs. (9) and (10) as follow [34]:

$$\begin{aligned} \mathcal{L}_{t-1} &= \mathbb{E}_{t,\hat{x}_t,\varepsilon}\left[\| \varepsilon - \varepsilon_\theta(\hat{x}_t, y) \|^2\right] \\ \hat{x}_t &\leftarrow \frac{\hat{x}_{t+1} - \eta(t+1)\varepsilon_\theta(\hat{x}_{t+1}, y)}{\sqrt{1-\alpha_{t+1}}} + \sigma_t \nu \end{aligned} \tag{11}$$

Pretrained Encoder and Decoder network: To harness the representation capabilities of the pretrained model for feature extraction, the encoder and decoder networks remained frozen throughout the fine-tuning process. For the conditioning inputs, we employed a pretrained image encoder (*CLIP-ViT-H-14-laion2B-s32B-b79K* model [35]) to generate DAS image embeddings and a pretrained text encoder (CLIP text encoder [36]) to process fatigue cycling text inputs. The pretrained model for the X-ray modality is based on the *sd-vae-ft-mse-original* variational autoencoder (VAE) [37], which integrates both pretrained encoder and decoder networks. More details for the pretrained models are provided in Data availability section. During the fine-tuning training of the

Table 1
Material properties and layup information in NASA dataset.

E_{11} [Gpa]	E_{22} [Gpa]	E_{33} [Gpa]	G_{12} [Gpa]	G_{13} [Gpa]	G_{23} [Gpa]	ν_{12}	ν_{13}	ν_{23}	ρ [kg/m^3]
137.5	8.4	8.4	6.2	3.092	3.092	0.309	0.309	0.5	1800

Specimens	Layup	Imaging region	PZT transducers	X-ray images
L1	$[0_2/90_4]_s$		✓	✓
L2	$[0/90_2/45-45/90]_s$	$[152 \times 178]$	✓	✓
L3	$[90_2/45/-45]_{2s}$		✓	✓

DI-CDM pipeline, only the UNet model parameters are updated to enhance training efficiency and data efficiency while maintaining the feature representation ability of the pretrained networks.

Cross attention mechanism in UNet: The cross attention layer in UNet facilitates the fusion of ultrasonic and X-ray modalities by dynamically calibrating their attention scores. As illustrated in Fig. 2(b), X-ray images are first processed through a pretrained VAE to extract feature representations. These image features, denoted as z_t, are integrated into the UNet model via cross-attention layers. During this process, conditioning ultrasonic modality input features y_t are fused with z_t to yield the output of the cross attention layer, which is defined as follows [38]:

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{Softmax}(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d_k}})\mathbf{V} \tag{12}$$

where $\mathbf{Q} = z_t\mathbf{W}_q, \mathbf{K} = y_t\mathbf{W}_k, \mathbf{V} = y_t\mathbf{W}_v$ are the query, key, and value matrices of the attention operation respectively, and $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v$ are the weight matrices of the trainable linear projection layers. d_k refers to the dimension of the key matrix $\mathbf{K}$. A detailed visualization of the multimodal fusion mechanism in the UNet is illustrated in Fig. 3. Within the UNet architecture, $\hat{x}_t$ and $\hat{x}_{t-1}$ (the symbols remain consistent with the illustration provided in Fig. 2) denote the latent features at time steps t and $t-1$ during the denoising process, respectively. The variable y represents the conditioning inputs, which in this study corresponds to the ultrasonic modality and fatigue cycling information.

As illustrated in Fig. 3(a), the Transformer block serves as a key component to combine ultrasonic and X-ray modalities. This block enables dynamic fusion between the two modalities by computing attention scores across their feature representations. The mathematical formulation of this fusion process is explicitly defined in Eq. (12), which describes how the cross attention mechanism computes weighted combinations of features through query-key–value operations. To further clarify the implementation details, Fig. 3(b) provides a step-by-step visualization of the tensor operations within the Transformer block. This schematic illustrates the transformation process of multidimensional feature maps from both modalities through three key components: the primary Transformer block, its basic Transformer subblock, and the inner cross attention layer. Within the cross attention layer, latent features are represented as z_t (X-ray modality features) and y_t (ultrasonic modality and loading cycle contextual features), respectively.

LoRA fine-tuning technique: Fine-tuning is the common practice of taking a model that has been trained on a wide and diverse dataset and then retraining this model on smaller and task-specific datasets. This approach is widely used in deep learning and has proven to be highly effective for all manner of models. In this manuscript, we applied the Low-Rank Adaptation (LoRA) [39] technique to fine-tune the DI-CDM framework, leveraging the pretrained diffusion model specifically for damage imaging. The LoRA fine-tuning technique compared to regular fine-tuning is illustrated in Fig. 4. The core architecture of LoRA involves keeping the main pretrained model weights frozen, ensuring that the fundamental knowledge the model has learned remains intact. Rather than retraining all model parameters, LoRA introduces low-rank matrices into the training of each layer of the Transformer architecture introduced in the UNet. These matrices are significantly smaller, reducing the computational load and simplifying the adaptation process [40].

4. Data description and evaluation metrics

4.1. Data description

The fine-tuning training and testing datasets reported in this study are sourced from the fatigue aging test on CFRP composites, conducted by Stanford Structures and Composites Laboratory in collaboration with the Prognostic Center of Excellence (PCoE) at NASA Ames Research Center [41,42]. In this experiment, a set of coupons was subjected to a tension–tension fatigue test conducted under a load-controlled loop of sinusoidal loadings with a frequency of 5 Hz and a load ratio of 0.14. The coupons were laminated with three symmetric layup configurations to account for the effect of ply orientation with each ply measuring 0.132 mm in thickness, as shown in Table 1. For damage monitoring, two sets of six-PZT-sensor SMART layers were attached to the surface of each sample. This configuration provided six actuators and receivers to monitor ultrasonic guided wave propagation through the samples. Each of the 36 actuator–receiver trajectories was actuated at seven evenly spaced frequencies in the range of 150 to 450 KHz.

Fig. 5 illustrates the experimental setup and imaging region. To enhance X-ray absorption, dye penetrant was applied to the samples before capturing X-ray images. Fatigue cycling tests were periodically paused to collect both PZT data and X-ray images. Based on the plate dimensions and the actuator–receiver array spacing, the imaging area was defined as a region measuring 152 pixels in width and 178 pixels in height. DAS beamforming results were obtained from three specimens (L1, L2, and L3). The DAS beamforming images and X-ray images obtained from specimens L2 and L3 were used to fine-tune the proposed DI-CDM pipeline, while the images from specimen L1 were used for testing to evaluate imaging performance.

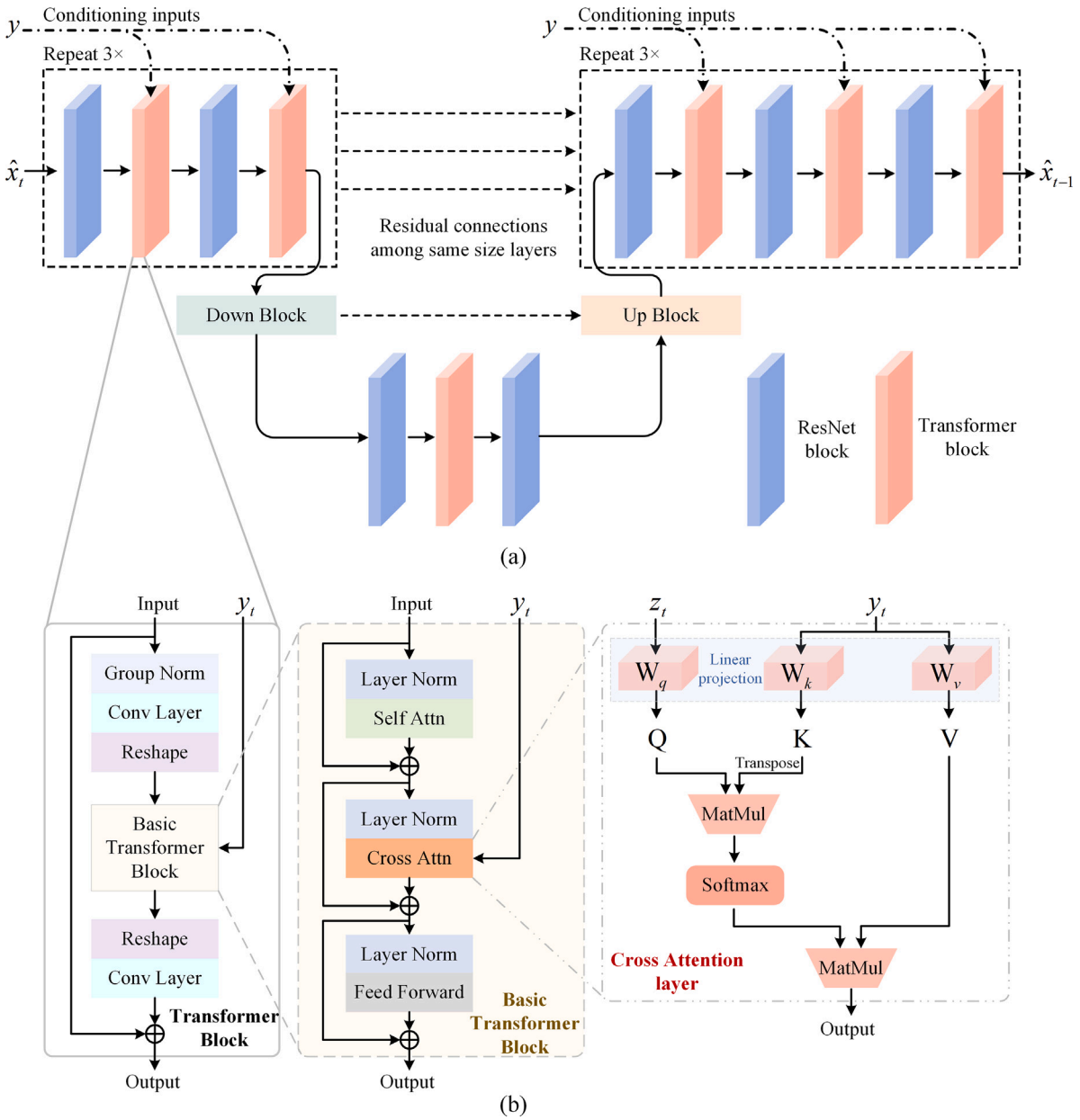

Fig. 3. The multimodal fusion mechanism in UNet ϵ_θ. (a) Ultrasonic and X-ray modality fusion within the Transformer block; (b) A step-in visualization of the Transformer block for cross attention mechanism.

4.2. Evaluation metrics

4.2.1. Intersection over union

Damage imaging results were evaluated based on image assessment metrics. The intersection over union (IoU) is the ratio of the intersection of two delamination regions to their union area. The definition of IoU is shown below:

$$\text{IoU} = \frac{|\mathcal{P} \cap \mathcal{P}_{ref}|}{|\mathcal{P} \cup \mathcal{P}_{ref}|} \tag{13}$$

where $\mathcal{P}$ denotes the delamination region of the normalized damage image generated by the model with pixel intensity equal to or greater than 0.5. $\mathcal{P}_{ref}$ is the ground truth delamination region, which is represented as the delamination region in the X-ray image in this work. The IoU score will be high if there is a significant overlap between the model output delamination region and the ground-truth delamination region. In contrast, a low degree of overlap will result in a low IoU score.

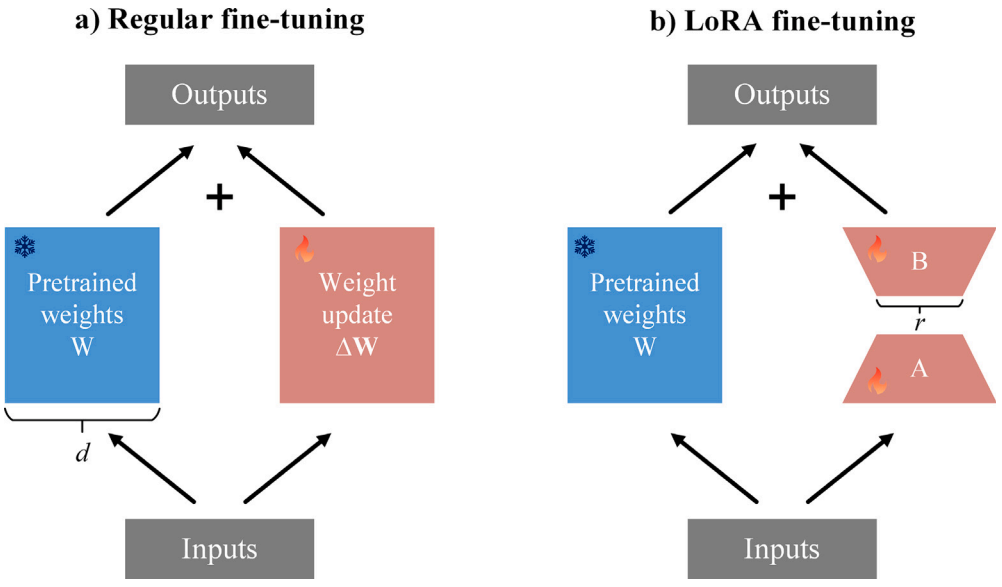

Fig. 4. LoRA fine-tuning technique. Comparing with regular fine-tuning updating pretrained weights $W \in \mathbb{R}^{d\times d}$, LoRA is performed by replacing the weight update matrix ΔW with two smaller matrices ($A \in \mathbb{R}^{d\times r}$ and $B \in \mathbb{R}^{r\times d}$, where $r \ll d$) through low-rank decomposition, which greatly reduces the number of trainable parameters.

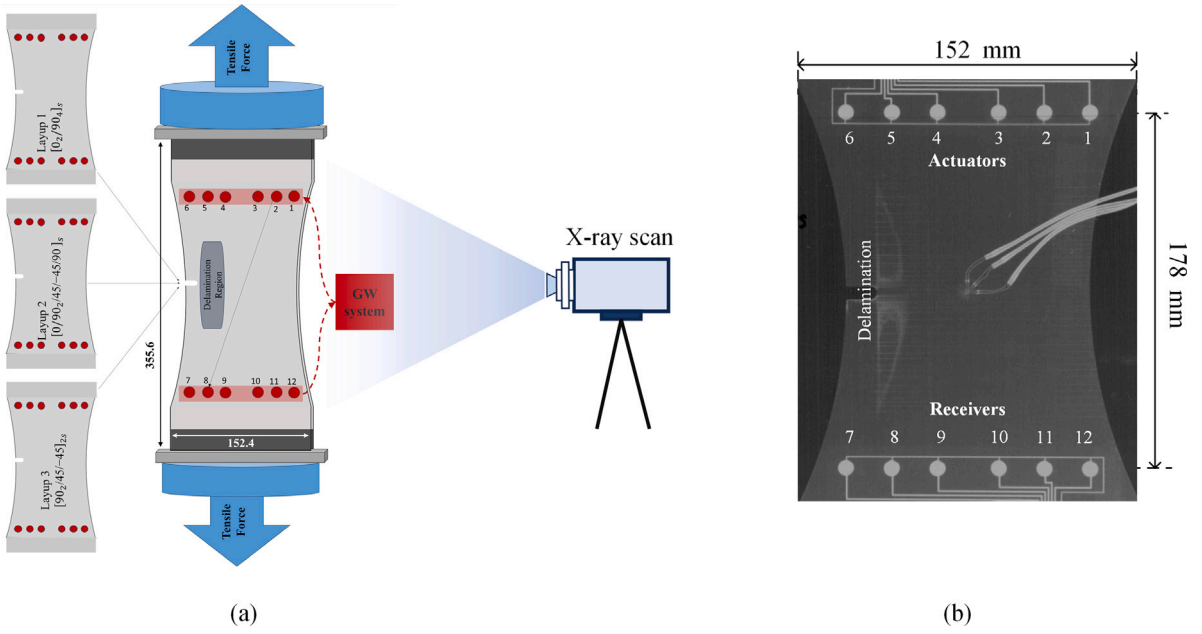

Fig. 5. The experimental setup and imaging region. (a) NASA datasets experiment setup [43]; (b) The imaging region of the composite specimens.

4.2.2. Signal-to-noise ratio

To efficiently evaluate the imaging results from DAS beamforming and the proposed DI-CDM framework, signal-to-noise ratio (SNR) is employed as a dimensionless metric for quantitative assessment. The SNR for damage imaging is calculated by subtracting the average pixel intensity of the background from the average pixel intensity of the damage region and then dividing by the standard deviation of the background pixel intensities. Mathematically, this can be expressed as [44,45]:

$$\mathrm{SNR} = 20\log_{10}\frac{|\mathrm{mean}(\mathcal{A}_D) - \mathrm{mean}(\mathcal{A}_{\mathcal{U}D})|}{\mathrm{std}(\mathcal{A}_{\mathcal{U}D})} \tag{14}$$

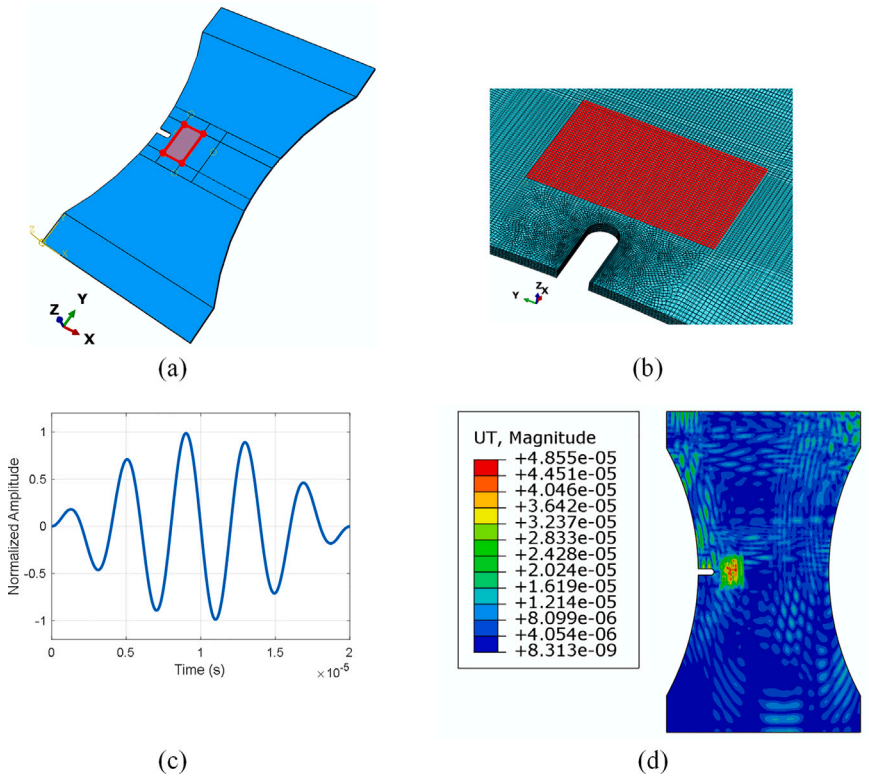

Fig. 6. Numerical simulation setup of composite plates in NASA dataset. (a) The simulated composite plate in Abaqus/Explicit; (b) An enlarged view of delamination area in composite plate; (c) A 250 kHz Hanning-modulated actuation signal as an input; (d) Wave propagation across the delamination region.

where $\mathcal{A}_D$ and $\mathcal{A}_{\mathcal{U}\mathcal{D}}$ denote the pixel intensities of the delamination region and the background (undelaminated region) in the imaging results, respectively. The terms mean and std separately refer to the mean value and standard deviation. The target region of $\mathcal{A}_D$ corresponds to the delamination area generated by the DAS imaging and the proposed DI-CDM approaches, while all pixels outside the target region constitute the background region, which represents the noise in the definition of SNR.

5. Numerical validation

To validate our framework, numerical simulations were first conducted to generate sparse array ultrasonic guided wave data. Delamination growth in composite plates was modeled in the Abaqus/Explicit environment (see Fig. 6(a)), with an enlarged view of the delamination within the composite ply provided for clarity. The C3D8R element was used with a 1 mm mesh size, refined to 0.5 mm near the notch to create a smooth meshing transition (Fig. 6(b)). A 250 kHz Hanning-modulated signal was applied to the actuation transducer, and the interaction of the guided waves with the delamination is visualized in Fig. 6(d). Notably, all transducers in the simulation served as both actuator and receiver elements to obtain better DAS imaging results. Although this approach is often classified as TFM imaging in most literature, TFM itself is a subset of the DAS imaging method. For consistency, we retained the term DAS imaging throughout this study.

5.1. DAS imaging

5.1.1. Group velocity estimation

For DAS imaging, once the distance and wave velocity are known, imaging results can be obtained using the approach detailed in Section 2.1. A critical step in generating the DAS image involves estimating the group velocity from the PZT receiver signals. Given the relatively small size of the coupons in the NASA experiment (15.24 cm × 25.4 cm), the low velocity of the fundamental anti-symmetric A_0 mode limits its utility for distinguishing reflected A_0 mode from the delamination. Therefore, the symmetric S_0 mode was utilized in this study as the primary mode for DAS imaging. Fig. 7 presents the detailed framework for estimating the group velocity of the symmetric S_0 wave mode from PZT receiver signals. A 250 kHz excitation frequency was used as the actuation signal, and actuator–receiver signals were captured around the actuator at angular orientations ranging from 0° to 360° in 30° increments, as shown in Fig. 7(b). The corresponding output signals were processed by a band-pass filter with a passband from 230 to 270 kHz to effectively reduce noise in the received signals. Then, the Hilbert transform was applied to isolate the first arrival of

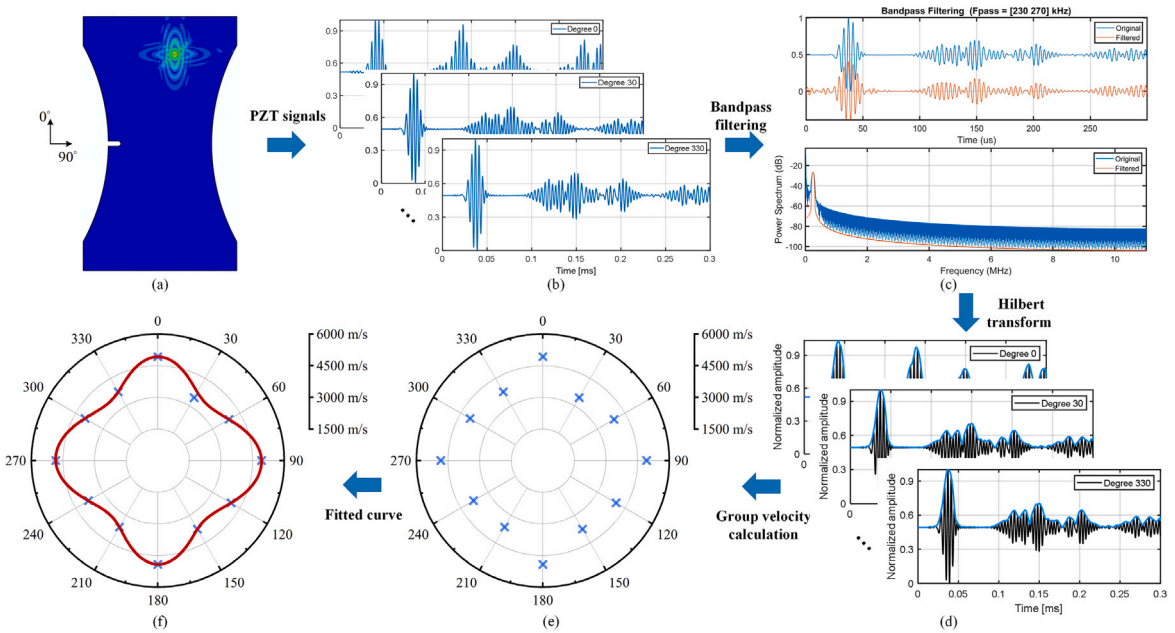

Fig. 7. The symmetric S_0 wave mode group velocity estimation from PZT receiver signals (the burst excitation centered at 250 kHz serves as an illustration for this schematic). (a) Composite plates in NASA experiments; (b) PZT signals received from various actuator–receiver pairs; (c) A band-pass filter applied to the PZT data to reduce noise disturbance; (d) Hilbert transform for filtered signals to extract the arrival time of the S_0 wave; (e) The estimated S_0 wave group velocity visualized in a radar plot; (f) Fitted result for S_0 wave group velocity.

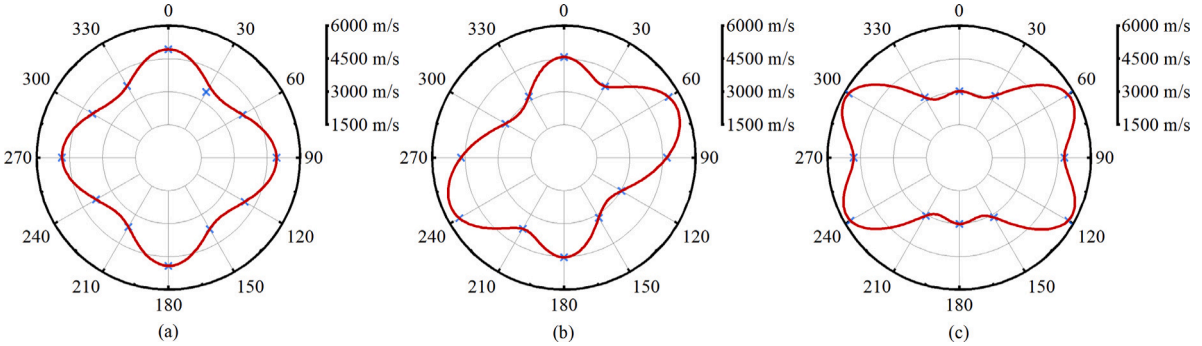

Fig. 8. The S_0 wave mode group velocity estimation results in numerical simulation. (a) L1 specimen; (b) L2 specimen; (c) L3 specimen.

the S_0 wave mode, allowing accurate estimation of arrival time. The group velocity of the S_0 wave mode at the center frequency of 250 kHz was calculated from 0° to 360° with a gap of 30° by the predefined node distance in the simulation and the estimated arrival time from Fig. 7(e). Thus, the estimated velocity of the S_0 wave mode can be fitted in all directions, as illustrated in Fig. 7(f).

As outlined in the flowchart, the estimated S_0 wave mode velocities from numerical simulations of three composite specimens are shown separately in Fig. 8(a–c). By fitting the S_0 wave velocity data across angular orientations from 0° to 360° in 30° increments, the wave velocity can be determined in all in-plate directions. In particular, the velocity exhibits an angular dependence, reflecting the directionally varying propagation characteristics of the composite laminate. This angular-dependent behavior displays nearly symmetrical characteristics, as illustrated in the subfigures, which are consistent with the results reported in Wang's work [46].

5.1.2. Imaging results

The DAS imaging algorithm can be applied in all directions using the estimated angular-dependent velocity of the S_0 wave mode and the known distance between the actuator–receiver pair. Note that the weight factors assigned to each beam shape in the DAS beamforming algorithm are the same, as there is no prior information regarding the desired beam shape. To simulate delamination progression, an increasing size of the delamination was simulated in Abaqus to represent the progression of the delamination region. Ultrasonic guided wave signals were then recorded and subtracted in the time–frequency domain from a baseline (undamaged)

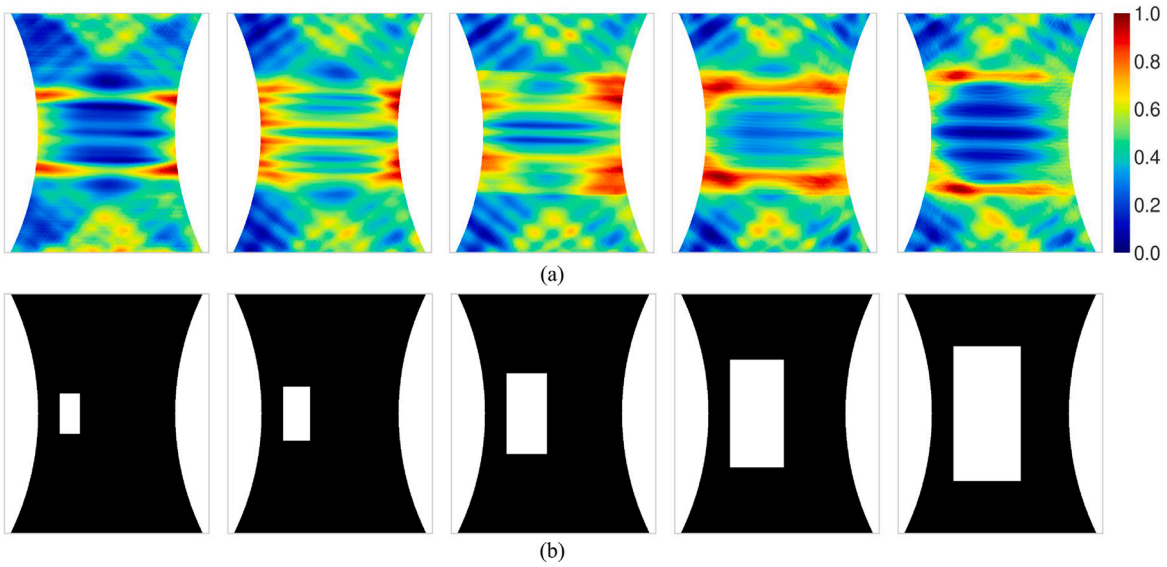

Fig. 9. Damage imaging results based on numerical simulation data. (a) DAS imaging results; (b) Binary images (denoted as Sim#1-5) representing ground truth, derived from the simulated delamination area.

Table 2
LoRA rank and corresponding trainable parameters in UNet.

LoRA ranks	UNet model parameters	Trainable parameters	Percentage (%)
16	868,443,332	2,521,088	0.2903
32		5,042,176	0.5789
64		10,084,352	1.1512
128		20,168,704	2.2761

signal before applying the DAS imaging algorithm, with the results for the simulated delamination shown in Fig. 9(a). As shown in the figure, the DAS image intensities expand radially from the center of the composite plate toward its edges as the simulated delamination grows. This occurs because the expanding delamination causes the actuated guided waves to reflect directly back to the upper transducer array. A similar reflection mechanism occurs at the bottom transducer array, creating symmetric high-intensity regions around the center of the plate. The horizontal symmetry of the high intensity area aligns with the experimental sensor configuration, which included only upper and lower horizontal sparse transducer arrays. This result also suggests that incorporating a vertical sparse transducer array could further enhance the DAS imaging results. The quantitative comparison between DAS and DI-CDM imaging based on the IoU and SNR metrics is given in Table 3.

5.2. DI-CDM imaging

5.2.1. LoRA fine-tuning

Following the DI-CDM imaging pipeline proposed in Section 3, we employed LoRA fine-tuning techniques to fine-tune simulated binary images, with DAS images and cycling information as conditioning inputs. The LoRA fine-tuning carried out in this study was implemented via the PEFT package provided by Hugging Face [47]. The final damage image is generated as a numerical matrix within the range of 0 to 1. Before fine-tuning training, the hyperparameters need to be configured, including the noise schedule, timesteps, gradient accumulation step, LoRA rank, learning rate, and warm-up rate. In this work, we applied the DDPM noise scheduler [19] with 1000 diffusion steps in the fine-tuning training. At inference, the DDIM scheduler [48] was adopted to denoise the noisy latents in 50 steps. The gradient accumulation step was set to 16 to avoid GPU out-of-memory errors for multi-GPU distributed and parallel training based on the Accelerate library [49]. As for LoRA rank r, we selected a value of 32 and compared the training efficiency for different LoRA rank parameters as illustrated in Table 2. The AdamW optimizer was used with the learning rate set as a smaller value of 1e−6 to ensure the model can refine fine-grained details on a customized dataset. A warm-up strategy with 300 steps was implemented to enhance learning stability by gradually increasing the learning rate. The training details were logged and visualized on the Weights&Biases platform, with the loss curve and learning rate variation shown in Fig. 10. All of the fine-tuning training was performed based on KU Leuven High Performance Computing (HPC) clusters with 2×NVIDIA A100 80 GB GPUs.

Fig. 10(a) shows the convergence of the loss curve after approximately 700 training steps. Notably, curves corresponding to larger LoRA ranks exhibit a steeper loss decline than those with smaller ranks. Lower LoRA ranks tend to regulate the model's output toward structured or formatted results, while higher ranks allow the model to capture finer-grained details, thus improving

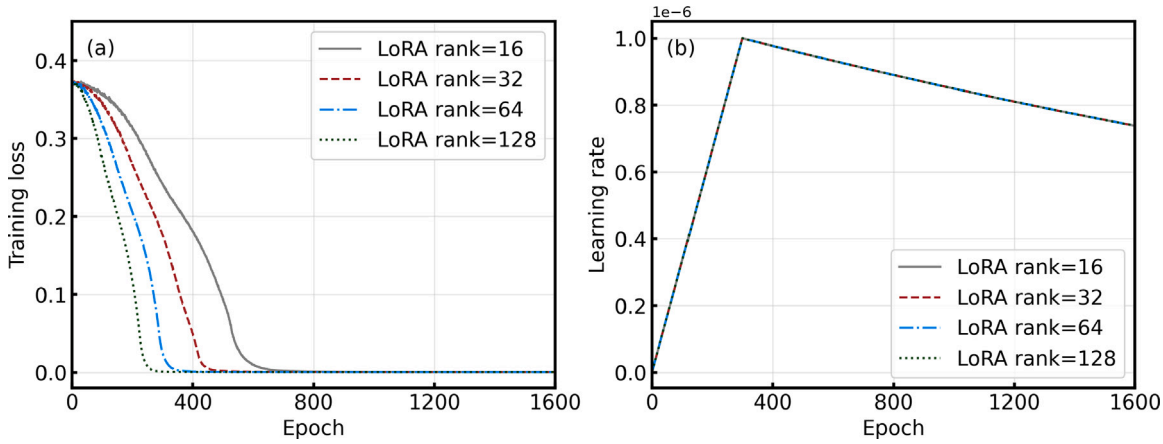

Fig. 10. LoRA fine-tuning training. (a) Loss curve; (b) Learning rate variation with warmup strategy.

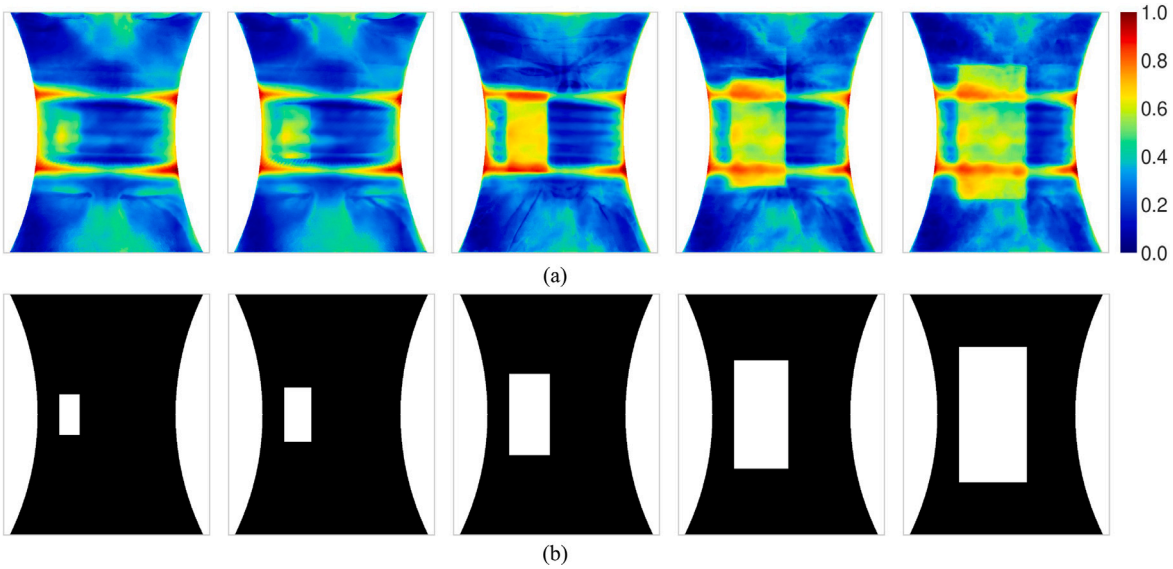

Fig. 11. DI-CDM imaging results on simulation data. (a) DI-CDM imaging results; (b) Binary images (denoted as Sim#1-5) representing ground truth, derived from the simulated delamination area.

performance. However, increasing the LoRA rank introduces more trainable parameters, leading to higher computational costs. The trade-off between model performance and parameter efficiency through optimal LoRA rank selection remains an active research area in efficient diffusion model-based and large language model-based fine-tuning (e.g., dynamic adaptive rank optimization and geometry-aware rank adaptation). For further exploration, readers may refer to recent studies in Refs. [50,51].

5.2.2. *Imaging results*

The DI-CDM imaging results are shown in Fig. 11(a). These results show its ability to capture delamination progression by fusing binary imaging inputs with conditioning DAS images derived from numerical simulation data. Across the subfigures (left to right), increasing the weighting of the conditioning inputs correlates with reduced image intensity outside the delamination regions. However, high-intensity values persist at the boundary edges due to contributions from the conditioning DAS imaging.

Table 3 provides a quantitative comparison based on IoU and SNR metrics. The proposed DI-CDM pipeline outperforms DAS imaging in all IoU metrics for Sim#1 to #5, as it effectively recovers the delamination area (Fig. 11(b)). For SNR metrics, DI-CDM achieves superior results in the first three simulation cases. In Sim#4 and #5, the SNR values (0.1106 and 0.1190) of the DI-CDM are comparable to the DAS imaging results, attributable to the standard deviation in the undamaged area. Overall, the DI-CDM iteratively generates damage images and reveals delamination progression, yielding higher IoU and SNR values than DAS imaging in numerical simulation.

Table 3
Quantitative comparison of simulation results based on proposed evaluation metrics.

Methods	No.	$\mathcal{P}_{ref}/_{(\text{pixels})}$	$\mathcal{P}/_{(\text{pixels})}$	IoU↑	Mean($\mathcal{A}_D$)	Mean($\mathcal{A}_{U\text{-}D}$)	Std($\mathcal{A}_{U\text{-}D}$)	SNR/$_{(\text{dB})}$↑
DAS	Sim#1	450	1653	0.0050	0.5988	0.2810	0.1022	9.8493
	Sim#2	800	4387	0.0556	0.5825	0.3498	0.0985	7.4622
	Sim#3	1800	8229	0.0731	0.6199	0.3684	0.0925	8.6897
	Sim#4	3200	7479	0.1239	0.6485	0.3694	0.0913	9.7011
	Sim#5	5000	6508	0.1173	0.6089	0.3332	0.1110	7.8988
DI-CDM	Sim#1	450	1270	**0.0325**	0.5932	0.2658	0.0965	**10.6096**
	Sim#2	800	1243	**0.0608**	0.5855	0.2647	0.0974	**10.3501**
	Sim#3	1800	2606	**0.6211**	0.5791	0.2356	0.0970	**10.9863**
	Sim#4	3200	2289	**0.3878**	0.5708	0.2539	0.1106	**9.1429**
	Sim#5	5000	2133	**0.2697**	0.5631	0.2689	0.1190	**7.8605**

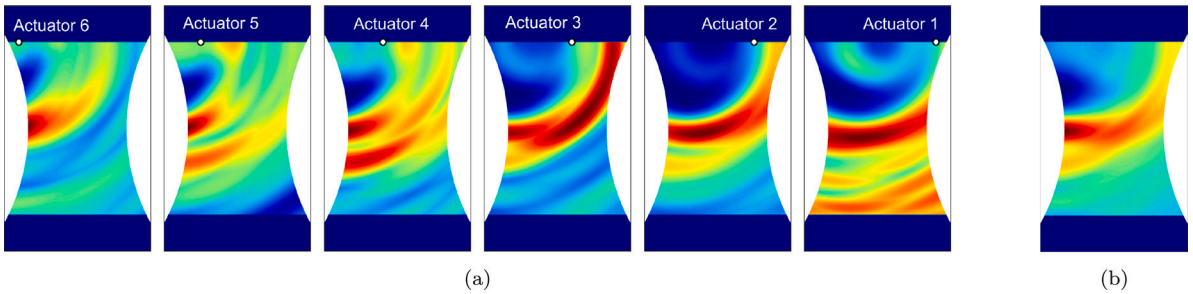

Fig. 12. DAS imaging results at a center frequency of 250 kHz for the L2 composite specimens at the first fatigue cycle. (a) Imaging results generated individually by actuators 1 to 6. (b) Combined imaging results obtained by averaging the DAS imaging results from all six actuators.

6. Experimental validation

To further validate the effectiveness of the proposed DI-CDM framework, the sparse array ultrasonic data and X-ray images representing the real delamination region in the NASA experiment were used to generate DAS imaging and the proposed DI-CDM imaging results.

6.1. DAS imaging

As described in Section 3, in experimental validation, the DAS imaging results acquired from the L2 and L3 samples were utilized to fine-tune the DI-CDM framework. Fig. 12 presents the DAS imaging results for the L2 composite plate at a center actuation frequency of 250 kHz at the first fatigue cycle. For each actuator, these images were generated using the predefined actuator–receiver distance and the estimated velocity of the S_0 wave mode, following the same procedure outlined in the simulation section. This process produced six distinct damage images (one for each of the actuators 1–6), as shown in Fig. 12(a). (In the NASA experiments, six actuators were positioned on the upper transducer array and six receivers on the lower array.) Notably, the DAS image from actuator 1 shows an intensified damage pattern near the lower edge, caused by wave reflections from the plate boundaries. To improve imaging quality, individual images from the six actuators were averaged to create a combined damage image, shown in Fig. 12(b), which served as input for fine-tuning the DI-CDM framework.

6.2. DI-CDM imaging

Following the same fine-tuning training procedures described in the simulation section, the DAS imaging results, the proposed DI-CDM imaging results, and the ground truth X-ray images for the L1 composite specimen from fatigue loading cycles 1 to 40k are shown in Fig. 13(a–c), respectively.

Fig. 13(a) presents the DAS imaging results for fatigue loading cycles ranging from 1 to 40k. At the initial cycle, the DAS imaging result indicates the damage location near the center of the tested plate, which aligns with the initial notch observed in the X-ray image at cycle 1, as shown in the first subfigure of Fig. 13(c). However, even at the initial cycle, the DAS imaging result fails to accurately locate the notch's position. This limitation arises because the ultrasonic transducer array used in the NASA experiment consists of a single linear array. Although incorporating an additional linear ultrasonic transducer array in the vertical direction and making a more accurate estimation of the S_0 wave speed (as illustrated in Fig. 7) could enhance DAS imaging and mitigate imaging errors, this study focuses on damage imaging based on existing DAS imaging results. From loading cycles 10k to 40k, the DAS imaging results show that it becomes difficult to locate the damage, and also that it cannot obtain the damage delamination from the ultrasonic signals. When the delamination region progressively expands, as shown in the subfigures in Fig. 13(c), the actuation guided waves were reflected back rather than transmitting through the delamination region and being received by the receiver

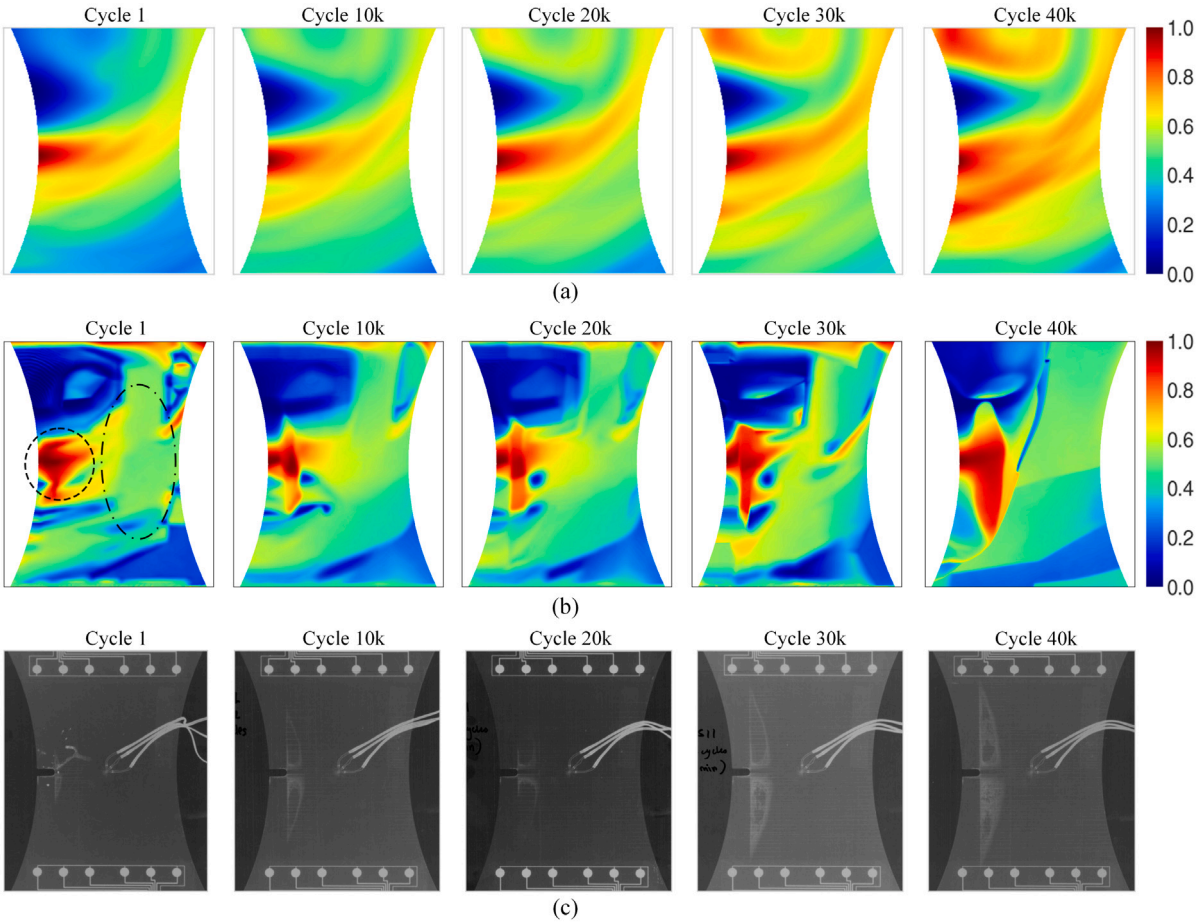

Fig. 13. Damage imaging results for the L1 composite specimen of fatigue loading cycles from 1 to 40,000. (a) DAS imaging results. (b) DI-CDM imaging results. (c) Ground truth X-ray images.

array. This is also the explanation for why the DAS imaging results typically exhibit higher pixel intensities in the damage image to the top left of the subfigures in Fig. 13(a).

In contrast, the damage imaging results generated by the proposed DI-CDM approach demonstrate progressive changes (highlighted by the circle with black dashed line) as the loading cycles increase from 1 to 40,000, as illustrated in Fig. 13(b). Since the DAS imaging result at cycle 0 was used as the input for testing the DI-CDM, the imaging results produced by the fine-tuned DI-CDM exhibit significantly enhanced performance compared to the DAS imaging results at cycle 1. Specifically, the middle left region of the DI-CDM imaging result at cycle 1 (black dashed line) exhibits a higher concentration of intensity values than the corresponding DAS results. This improvement can be attributed to the fact that the proposed DI-CDM was fine-tuned using X-ray image data collected from the L2 and L3 composite plates, enabling its outputs to align with the delamination progression observed in those plates. From cycle 10,000 onward, the damage images generated by the DI-CDM gradually begin to reveal the delamination region, reflecting the well-learned capability of the proposed DI-CDM to illustrate and generate such damage. In the right region (marked by the black dashed–dotted line), pixel intensity values remain stable at approximately 0.4. However, compared to DAS imaging results, the undelaminated area is more distinctly represented. To quantitatively demonstrate the improved performance of the proposed DI-CDM, Table 4 provides a quantitative comparison of the DAS and DI-CDM approaches based on IoU and SNR evaluation metrics.

As shown in Table 4, the reference $\mathcal{P}_{ref}$ value derived from X-ray ground truth images in Fig. 13(c), shows a significant increase from 156 to 1943. This trend aligns with the gradual expansion of the delamination region under fatigue loading cycles. Regarding the IoU values, while DAS imaging produces values consistently below 0.1, the DI-CDM achieves values nearly three times higher than those of the baseline DAS results. Furthermore, starting from cycle 10,000, the IoU values from the DI-CDM indicate that the proposed approach can partially capture the upward trend of the delamination region observed in X-ray images, with values increasing from 0.1322 to 0.3340. Although the IoU values are not exceptionally high, the improved results are achieved without relying on costly X-ray imaging during the testing phase. Additionally, these results suggest that multimodal fusion can indeed enhance the performance of DAS imaging based on ultrasonic sparse array data.

Table 4
Quantitative comparison of experiment results based on proposed evaluation metrics.

Methods	Cycles	$\mathcal{P}_{ref}/_{(\text{pixels})}$	$\mathcal{P}/_{(\text{pixels})}$	IoU↑	Mean($\mathcal{A}_D$)	Mean($\mathcal{A}_{U'D}$)	Std($\mathcal{A}_{U'D}$)	$\text{SNR}/_{(\text{dB})}$↑
DAS	1	156	4090	0.0164	0.6117	0.3390	0.0993	8.7732
	10k	985	4902	0.0623	0.5974	0.4023	0.0995	5.8496
	20k	1614	6676	0.0888	0.5769	0.3946	0.1093	4.4458
	30k	1677	7948	0.0761	0.6065	0.4057	0.1116	5.1008
	40k	1943	8957	0.0883	0.6232	0.3757	0.1137	6.7579
DI-CDM	1	156	5441	**0.0346**	0.6140	0.3010	0.1511	**6.3264**
	10k	985	6364	**0.1322**	0.6054	0.2926	0.1440	**6.7394**
	20k	1614	4846	**0.1872**	0.6245	0.3203	0.1388	**6.8135**
	30k	1677	4710	**0.2234**	0.6258	0.3271	0.1260	**7.4967**
	40k	1943	3993	**0.3340**	0.6698	0.3448	0.1322	**7.8148**

As for the SNR metric, the values for DAS beamforming consistently remain below 7 dB, except for the value at cycle 1. This exception occurs because the undelaminated region in the DAS result has a lower standard deviation, indicating a smoother pixel value distribution. In contrast, the proposed DI-CDM framework achieves better SNR values under higher standard deviation values for the undelaminated region, with all measurements exceeding 6.3264 dB. This demonstrates that DI-CDM more effectively separates the delamination region (the SNR signal component) from background interference (the noise component) as fatigue cycles progress. These results highlight the effectiveness of the proposed approach in maintaining signal clarity during progressive degradation.

7. Conclusion

This study introduces a novel damage imaging framework designed for SHM systems based on a fine-tuned conditional diffusion model (DI-CDM). Given the need for efficient damage imaging to enable fast damage assessment, damage imaging using sparse array ultrasonic transducers during fatigue loading cycles presents a significant challenge. The proposed framework begins with generating initial damage image through DAS beamforming using sparse array ultrasonic data. X-ray imaging, which provides detailed insight into composite delamination, is then used as input to the DI-CDM framework. This approach enables the continuous enhancement of the damage image by integrating the delamination information extracted from X-ray scans. Through rigorous fine-tuning training, the DI-CDM framework achieves fast, high-resolution damage imaging. The main conclusions of this study are summarized as follows.

(i) The DI-CDM framework was validated using NASA datasets in both simulation and experiments. Compared to DAS beamforming, the DI-CDM framework achieves significant improvements in damage imaging accuracy, demonstrating higher IoU and SNR values.
(ii) By integrating X-ray with ultrasonic modality, the DI-CDM framework generated fused damage images that more accurately reflected delamination progression during fatigue loading processes.
(iii) This work validated the effectiveness of applying LoRA fine-tuning in SHM systems, enabling diffusion models pretrained on large-scale datasets to effectively adapt to X-ray images fused with sparse array ultrasonic data for damage imaging.

Future work could be explored in these directions. **Multimodal sensor fusion for enhanced SHM systems:** integrating data from diverse sensors (e.g., acoustic emissions, strain gauges, and digital image correlation) could improve more accurate damage assessment by leveraging complementary data sources across modalities. **Controllable diffusion model inputs:** Conditional guidance in diffusion models requires a deeper exploration to optimize their adaptability for specific SHM systems, such as refining damage imaging under varying operating conditions and sensitivity analysis of conditioning input, i.e., DAS imaging in this work. **Domain-specific diffusion model fine-tuning:** while this study validated the framework using NASA datasets in SHM, the fine-tuning technique is broadly applicable to other research areas.

CRediT authorship contribution statement

Xin Yang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Sergio Cantero-Chinchilla:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Morteza Moradi:** Writing – review & editing, Writing – original draft, Methodology. **Panagiotis Komninos:** Writing – review & editing, Writing – original draft, Methodology. **Chen Fang:** Writing – review & editing, Writing – original draft. **Yunlai Liao:** Writing – review & editing, Writing – original draft. **Pradeep Kundu:** Writing – review & editing, Writing – original draft, Supervision. **Dimitrios Zarouchas:** Writing – review & editing, Writing – original draft. **Dimitrios Chronopoulos:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The NASA composite structure dataset used in this work is publicly available and can be found in the repository: https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/.

For DI-CDM fine-tuning, the code could be shared via the GitHub repository after the potential publication of this manuscript, and the pretrained models used in this work can be found in the Hugging Face repository:

- https://huggingface.co/stabilityai/stable-diffusion-v2.
- https://huggingface.co/laion/CLIP-ViT-H-14-laion2B-s32B-b79K.

References

[1] W. Wu, S. Cantero-Chinchilla, D. Prescott, R. Remenyte-Prescott, M. Chiachío, A general approach to assessing SHM reliability considering sensor failures based on information theory, Reliab. Eng. Syst. Saf. 250 (2024) 110267, http://dx.doi.org/10.1016/j.ress.2024.110267.

[2] A. Farrokhabadi, H. Lu, X. Yang, A. Rauf, R. Talemi, A. Hossein Behravesh, S. Kaveh Hedayati, D. Chronopoulos, Energy absorption assessment of recovered shapes in 3D-printed star hourglass honeycombs: Experimental and numerical approaches, Compos. Struct. 347 (2024) 118444, http://dx.doi.org/10.1016/j.compstruct.2024.118444.

[3] X. Yang, A. Farrokhabadi, A. Rauf, Y. Liu, R. Talemi, P. Kundu, D. Chronopoulos, Transfer learning-based Gaussian process classification for lattice structure damage detection, Meas. 238 (2024) 115387, http://dx.doi.org/10.1016/j.measurement.2024.115387.

[4] L. Radicioni, V. Giorgi, L. Benedetti, F.M. Bono, S. Pagani, S. Cinquemani, M. Belloli, On the performance of data-driven dynamic models for temperature compensation on bridge monitoring data, J. Civ. Struct. Heal. Monit. (2025) 1–16, http://dx.doi.org/10.1007/s13349-025-00920-2.

[5] X. Yang, C. Fang, P. Kundu, J. Yang, D. Chronopoulos, A decision-level sensor fusion scheme integrating ultrasonic guided wave and vibration measurements for damage identification, Mech. Syst. Signal Process. 219 (2024) 111597, http://dx.doi.org/10.1016/j.ymssp.2024.111597.

[6] R. Miorelli, C. Fisher, A. Kulakovskyi, B. Chapuis, O. Mesnil, O. D'Almeida, Defect sizing in guided wave imaging structural health monitoring using convolutional neural networks, NDT E Int. 122 (2021) 102480, http://dx.doi.org/10.1016/j.ndteint.2021.102480.

[7] A. Mardanshahi, A. Sreekumar, X. Yang, S.K. Barman, D. Chronopoulos, Sensing techniques for structural health monitoring: A state-of-the-art review on performance criteria and new-generation technologies, Sensors 25 (5) (2025) http://dx.doi.org/10.3390/s25051424.

[8] M. Barzegar, D.J. Pasadas, A.L. Ribeiro, H.G. Ramos, Baseline-free damage imaging for structural health monitoring of composite lap joint using ultrasonic-guided waves, IEEE Open J. Instrum. Meas. 3 (2024) 1–8, http://dx.doi.org/10.1109/OJIM.2024.3487239.

[9] C. Fang, X. Yang, K. Gryllias, D. Vandepitte, X. Liu, L. Zhang, D. Chronopoulos, Using removable sensors in structural health monitoring: A Bayesian methodology for attachment-to-attachment uncertainty quantification, Mech. Syst. Signal Process. 224 (2025) 111973, http://dx.doi.org/10.1016/j.ymssp.2024.111973.

[10] Q. Huan, M. Chen, Z. Su, F. Li, A high-resolution structural health monitoring system based on SH wave piezoelectric transducers phased array, Ultrason. 97 (2019) 29–37, http://dx.doi.org/10.1016/j.ultras.2019.04.005.

[11] J. Latif, M.Z. Shakir, N. Edwards, M. Jaszczykowski, N. Ramzan, V. Edwards, Review on condition monitoring techniques for water pipelines, Meas. 193 (2022) 110895, http://dx.doi.org/10.1016/j.measurement.2022.110895.

[12] Y. Ma, S. Hedayatrasa, A. Han Orta, K. Van Den Abeele, M. Kersemans, Broadband nonlinear delay and sum (NL-DAS) for baseline-free damage imaging using a sparse sensor network, Mech. Syst. Signal Process. 224 (2025) 111989, http://dx.doi.org/10.1016/j.ymssp.2024.111989.

[13] D. Teng, L. Liu, Y. Xiang, F.-Z. Xuan, An optimized total focusing method based on delay-multiply-and-sum for nondestructive testing, Ultrason. 128 (2023) 106881, http://dx.doi.org/10.1016/j.ultras.2022.106881.

[14] L. Le Jeune, S. Robert, E. Lopez Villaverde, C. Prada, Plane wave imaging for ultrasonic non-destructive testing: Generalization to multimodal imaging, Ultrason. 64 (2016) 128–138, http://dx.doi.org/10.1016/j.ultras.2015.08.008.

[15] F. Luleci, F.N. Catbas, O. Avci, Generative adversarial networks for data generation in structural health monitoring, Front. Built Environ. 8 (2022) http://dx.doi.org/10.3389/fbuil.2022.816644.

[16] L. Cheng, Z. Tong, S. Xie, M. Kersemans, IRT-GAN: A generative adversarial network with a multi-headed fusion strategy for automated defect detection in composites using infrared thermography, Compos. Struct. 290 (2022) 115543, http://dx.doi.org/10.1016/j.compstruct.2022.115543.

[17] P. Dhariwal, A. Nichol, Diffusion models beat GANs on image synthesis, 2021, arXiv:2105.05233 URL https://arxiv.org/abs/2105.05233.

[18] H. Thanh-Tung, T. Tran, On catastrophic forgetting and mode collapse in generative adversarial networks, 2020, arXiv:1807.04015 URL https://arxiv.org/abs/1807.04015.

[19] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, 2020, arXiv:2006.11239 URL https://arxiv.org/abs/2006.11239.

[20] F.M. Bono, A. Argentino, L. Bernardini, L. Benedetti, G. Cazzulani, C. Somaschini, M. Belloli, Automated operational modal analysis of a steel truss railway bridge employing free decay response, J. Infrastruct. Intell. Resil. 4 (1) (2025) 100145, http://dx.doi.org/10.1016/j.iintel.2025.100145.

[21] X. Yang, T. Ye, X. Yuan, W. Zhu, X. Mei, F. Zhou, A novel data augmentation method based on denoising diffusion probabilistic model for fault diagnosis under imbalanced data, IEEE Trans. Ind. Informatics 20 (5) (2024) 7820–7831, http://dx.doi.org/10.1109/TII.2024.3366991.

[22] T. Zhang, J. Lin, J. Jiao, H. Zhang, H. Li, An interpretable latent denoising diffusion probabilistic model for fault diagnosis under limited data, IEEE Trans. Ind. Informatics 20 (8) (2024) 10354–10365, http://dx.doi.org/10.1109/TII.2024.3393002.

[23] C. Guo, Y. Ding, H. Cui, J. Xu, L. Bai, Ultrasonic characterization of defects in polycrystalline materials based on TFM image reconstruction using denoising diffusion probabilistic models, Mech. Syst. Signal Process. 224 (2025) 111997, http://dx.doi.org/10.1016/j.ymssp.2024.111997.

[24] W. Fu, R. Zhou, Y. Gao, Z. Guo, Q. Yu, A diffusion model-based deep learning approach for denoising acoustic emission signals in concrete, Measurement 251 (2025) 117143, http://dx.doi.org/10.1016/j.measurement.2025.117143.

[25] Z. Lan, C. Rong, C. Han, X. Qu, J. Li, H. Lin, Coherence-based phase aberration correction and beamforming for ring-array ultrasound imaging, Mech. Syst. Signal Process. 220 (2024) 111651, http://dx.doi.org/10.1016/j.ymssp.2024.111651.

[26] X. Li, M. Sakevych, G. Atkinson, V. Metsis, BioDiffusion: A versatile diffusion model for biomedical signal synthesis, 2024, arXiv:2401.10282 URL https://arxiv.org/abs/2401.10282.

[27] G. Yariv, I. Gat, L. Wolf, Y. Adi, I. Schwartz, AudioToken: Adaptation of text-conditioned diffusion models for audio-to-image generation, 2023, arXiv:2305.13050 URL https://arxiv.org/abs/2305.13050.

[28] L. Lin, Z. Li, R. Li, X. Li, J. Gao, Diffusion models for time series applications: A survey, 2023, arXiv:2305.00624 URL https://arxiv.org/abs/2305.00624.

[29] J. Ho, C. Saharia, W. Chan, D.J. Fleet, M. Norouzi, T. Salimans, Cascaded diffusion models for high fidelity image generation, 2021, arXiv:2106.15282 URL https://arxiv.org/abs/2106.15282.

[30] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, M. Chen, GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models, 2022, arXiv:2112.10741 URL https://arxiv.org/abs/2112.10741.

[31] J. Wu, R. Fu, H. Fang, Y. Zhang, Y. Yang, H. Xiong, H. Liu, Y. Xu, MedSegDiff: Medical image segmentation with diffusion probabilistic model, 2023, arXiv:2211.00611 URL https://arxiv.org/abs/2211.00611.
[32] H. Zhang, Y. Li, J. Huang, Conditional denoising diffusion probabilistic model for seismic diffraction separation and imaging, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–13, http://dx.doi.org/10.1109/TGRS.2024.3381193.
[33] A. Bansal, H.-M. Chu, A. Schwarzschild, S. Sengupta, M. Goldblum, J. Geiping, T. Goldstein, Universal guidance for diffusion models, 2023, arXiv:2302.07121 URL https://arxiv.org/abs/2302.07121.
[34] G.M. Garcia, K.A. Zeid, C. Schmidt, D. de Geus, A. Hermans, B. Leibe, Fine-tuning image-conditional diffusion models is easier than you think, 2025, arXiv:2409.11355 URL https://arxiv.org/abs/2409.11355.
[35] C. Schuhmann, R. Beaumont, R. Vencu, C.W. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis, M. Wortsman, P. Schramowski, S.R. Kundurthy, K. Crowson, L. Schmidt, R. Kaczmarczyk, J. Jitsev, LAION-5B: An open large-scale dataset for training next generation image-text models, in: Thirty-Sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2022, URL https://openreview.net/forum?id=M3Y74vmsMcY.
[36] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever, Learning transferable visual models from natural language supervision, 2021, arXiv:2103.00020URL https://arxiv.org/abs/2103.00020.
[37] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, 2022, arXiv:2112.10752 URL https://arxiv.org/abs/2112.10752.
[38] O. Petit, N. Thome, C. Rambour, L. Soler, U-net transformer: Self and cross attention for medical image segmentation, 2021, arXiv:2103.06104 URL https://arxiv.org/abs/2103.06104.
[39] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, Lora: Low-rank adaptation of large language models, 2021, arXiv:2106.09685 URL https://arxiv.org/abs/2106.09685.
[40] K.P.V. Srinivasan, P. Gumpena, M. Yattapu, V.H. Brahmbhatt, Comparative analysis of different efficient fine tuning methods of large language models (LLMs) in low-resource setting, 2024, arXiv:2405.13181 URL https://arxiv.org/abs/2405.13181.
[41] M. Chiachio, J. Chiachio, A. Saxena, G. Rus, K. Goebel, Fatigue damage prognosis in FRP composites by combining multi-scale degradation fault modes in an uncertainty Bayesian framework, Proc. Struct. Heal. Monit. 1 (2013).
[42] C. Larrosa, K. Lonkar, F.-K. Chang, In situ damage classification for composite laminates using Gaussian discriminant analysis, Struct. Heal. Monit. 13 (2) (2014) 190–204, http://dx.doi.org/10.1177/1475921713517288.
[43] M. Moradi, F.C. Gul, D. Zarouchas, A novel machine learning model to design historical-independent health indicators for composite structures, Compos. Part B: Eng. 275 (2024) 111328, http://dx.doi.org/10.1016/j.compositesb.2024.111328.
[44] K. Ji, P. Zhao, C. Zhuo, H. Jin, L. Zhang, J. Fu, A high-SNR ultrasonic imaging method for weakly heterogeneous medium, Mech. Syst. Signal Process. 211 (2024) 111249, http://dx.doi.org/10.1016/j.ymssp.2024.111249.
[45] F.J. Madruga, C. Ibarra-Castanedo, O.M. Conde, J.M. López-Higuera, X. Maldague, Infrared thermography processing based on higher-order statistics, NDT & E Int. 43 (8) (2010) 661–666, http://dx.doi.org/10.1016/j.ndteint.2010.07.002.
[46] L. Wang, F. Yuan, Group velocity and characteristic wave curves of lamb waves in composites: Modeling and experiments, Compos. Sci. Technol. 67 (7) (2007) 1370–1384, http://dx.doi.org/10.1016/j.compscitech.2006.09.023.
[47] S. Mangrulkar, S. Gugger, L. Debut, Y. Belkada, S. Paul, B. Bossan, PEFT: State-of-the-art parameter-efficient fine-tuning methods, 2022, https://github.com/huggingface/peft.
[48] J. Song, C. Meng, S. Ermon, Denoising diffusion implicit models, 2022, arXiv:2010.02502 URL https://arxiv.org/abs/2010.02502.
[49] S. Gugger, L. Debut, T. Wolf, P. Schmid, Z. Mueller, S. Mangrulkar, M. Sun, B. Bossan, Accelerate: Training and inference at scale made simple, efficient and adaptable., 2022, https://github.com/huggingface/accelerate.
[50] M. Valipour, M. Rezagholizadeh, I. Kobyzev, A. Ghodsi, Dylora: Parameter efficient tuning of pre-trained models using dynamic search-free low-rank adaptation, 2023, arXiv:2210.07558 URL https://arxiv.org/abs/2210.07558.
[51] A. Ed-dib, Z. Datbayev, A.M. Aboussalah, GeLoRA: Geometric adaptive ranks for efficient LoRA fine-tuning, 2024, arXiv:2412.09250 URL https://arxiv.org/abs/2412.09250.