Predicting the acoustic influence map of the ultrasonic total focusing method using a depth-conditional diffusion model

Dong Ming [a,b]*, Wang Ze [a], Chen Yuan [a,b], Zhang Guang-ming [c], Cao Xian-gang [a,b], Wan Xiang [a,b]

[a]School of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China;
[b]Shaanxi Key Laboratory of Mine Electromechanical Equipment Intelligent Detection and Control, Xi'an 710054, China;
[c] School of Engineering, Liverpool John Moores University, Byrom Street, L3 3AF, United Kingdom

*Corresponding author: Dong Ming，E-mail: dongm@xust.edu.cn

Abstract

The Acoustic Influence Map (AIM) shows the acoustic energy distribution of total focusing images at different spatial positions. This assists users to select the correct propagation mode, or wave set, for their inspection. To address the computational complexity and low efficiency of existing AIMs, a deep-conditional diffusion model is proposed. The full matrix data (FMC) of side-drilled holes (SDHs) at various positions are simulated and acquired by K-space pseudo-spectral method, and a dataset of total focusing images are created. An improved U-Net is adopted as the backbone network, and a four-level encoder-decoder structure is constructed. The input channels are expanded to 34, and discrete depths are converted into 32-channel feature maps to match the image dimensions by a learnable embedding layer. The mask-depth embedding dual-condition input mechanism is realised by a single-channel mask image and a noise image. The generated images are evaluated using the Structural Similarity Index (SSIM) and Pearson Correlation Coefficient (CC). The range of SSIM value is from 0.86 to 0.93, indicating the model can effectively preserve the overall structure and details of the images. The range of CC value is from 0.68 to 0.86, demonstrating that the model is able to reconstruct the geometric structure of total focusing images. Using simulated and generated images as pre-data, depth gates are set to extract the maximum amplitude along the Z-axis direction, the composite amplitude profile at each depth are obtained. The AIM is then derived using key point-based spline interpolation and Hermite interpolation methods. The normalized amplitudes predicted by the AIM are compared with the total focusing images of holes. The results show a correlation coefficient of 0.96 between the

two datasets, indicating a strong positive correlation between the predicted values of the acoustic influence map and the actual imaging results.

Key words：Total Focusing Method, Acoustic Influence Map, K-space Pseudospectral Method, Conditional Diffusion Model, Ultrasonic Testing

1 Introduction

Ultrasonic arrays (UA) are used for a wide range of inspections in nondestructive testing. Conventional UA is based on beamforming technology and uses electronic time delays to create steerable and focused beams **Error! Reference source not found.**. The FMC acquisition technique allows the capture of all the information possibly measurable by the probe. The Total Focusing Method (TFM) is an advanced ultrasonic imaging technique [2]. TFM creates high-resolution images of flaws in a variety of materials by post-processing the FMC data. This technique synthesizes focus at every point in the image by coherently summing the time-shifted signals from all transmit-receive pairs, thereby achieving superior resolution and signal-to-noise ratio compared to conventional phased array methods [3]. Consequently, TFM is widely employed in demanding inspection scenarios such as aerospace components, nuclear power plants, and railway tracks, where high defect detectability and characterization accuracy are critical [4].

The acoustic field distribution can be used to provide guidance on adjusting scanning strategies, thus optimizing the signal-to-noise ratio and the probability of detection [5-6]. The distribution of acoustic field of conventional UA can be determined using acoustic modelling and simulation [7-9]. TFM is a post-processing imaging technique by focusing at every pixel, superseding conventional phased array techniques. Acoustic Influence Map (AIM) is introduced to predict the acoustic field energy distribution of TFM. The AIM is influenced by multiple factors, such as material properties, probe parameters, imaging algorithms, defect types, orientations and so on [10]. Chi-Hang Kwan et al. [11] proposed a semi-analytical acoustic model for predicting the AIM under different

detection modes, and the modeling tool was integrated into OmniScan X3 series of flaw detectors produced by Olympus. Guan et al. [12] acquired TFM images of SDHs in various positions to construct an amplitude map covering the entire imaging region, with the aim of achieving a quantitative evaluation of defects. This mapping, which was created using extensive TFM image datasets, provides a reliable quantitative standard for the components with complex geometries. Budyn et al. [13] developed a semi-analytical ultrasonic model to generate sensitivity images for predicting the TFM amplitudes of small defects. These images have been thoroughly validated through experimentation and accurately depict the variation in defect response in relation to position and inspection viewpoint. While both experimental and simulation images can be utilized to develop AIM, significant challenges remain in obtaining large volumes of TFM images.

The use of generative image technology is becoming increasingly important in research aimed at reducing reliance on physical experiments and traditional numerical simulations. [14]. To address the issue of low diagnostic accuracy caused by limited fault samples, Xin Yang et al. [15] proposed a damage imaging framework based on a fine-tuned conditional diffusion model. By integrating digital acoustic simulation images from ultrasonic arrays with X-ray images of fatigued composites, the framework enables the generation of high-resolution images. To address the challenge of data scarcity in surface defect detection under small-sample conditions, Xiaoqiao Wang et al. [16] proposed an uncertainty-guided dual-stage conditional diffusion model (UGDS-CDM). The model employs a pre-trained U-Net classifier and expert knowledge to generate high-fidelity industrial defect images for data augmentation. To overcome the limitation of scarce defect data in training generative models, Yichun Tai et al. [17] proposed a method for generating layouts of defects. This method uses pre-trained, layout-conditioned diffusion models to synthesise defect images in specified categories and regions. This expands the training dataset and enhances detection performance. Yu Liu et al. [18] introduced a 3D medical image segmentation method based on Mamba-enhanced diffusion model. Semantic-level embedding machines were employed to capture the relationship between

noisy labels and image data. A global, slice-aware Mamba layer was also designed to improve the accuracy of medical image segmentation. Dongwon Lee et al. [19] proposed DiffectNet, a diffusion-enabled conditional target generation network designed to produce high-fidelity synthetic defect images for phased array ultrasonic testing. While the condition mechanisms in the models mentioned above primarily focus on precisely controlling the attributes such as the category, morphology, or saliency of the generated defects. However, the impact of defect position variations on image quality remains unexplored, and it is still challenging to generate predicted TFM images for defects in numerous new positions.

To addressing the challenge of inefficient AIM generation caused by insufficient TFM images, a depth-conditioned diffusion model (DCDM) was proposed to generate ultrasonic TFM images. The proposed methodology utilises a refined U-Net as its fundamental architecture, this features a four-level encoder-decoder structure and an input expanded to 34 channels. The incorporation of depth conditions is facilitated by a learnable embedding layer, which converts discrete depth values into 32-channel feature maps aligned with the image dimensions. This enables the conditional input of depth information on a channel-by-channel basis. These feature maps are then concatenated channel-wise with a single-channel mask image and a noise images to form a mask-depth dual-condition input mechanism. This enables the model to identify differences in images of SDH at different positions, and to generate images for new positions. By combining the TFM images of SDH at new positions generated by the model, an AIM for the specific inspection configuration can be constructed efficiently.

2 Method

2.1 K-space pseudospectral method

The K-space pseudospectral method is employed to acquire FMC data of SDHs at difference positions. The Fourier transform is used to calculate the global spatial

derivatives in the wavenumber domain. The computation is efficient if the number of spatial sampling points satisfies the Nyquist sampling theorem, which states that there should be at least two grid points per wavelength. In contrast. While traditional finite difference methods [20] and finite element methods [21] generally require at least 10 grid points per wavelength. The K-space pseudospectral method offers computational efficiency and reduced memory requirements. In the context of a homogeneous and lossless medium, the second-order wave equation is:

$$\nabla^2 p(x,t) - \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2} p(x,t) = 0 \tag{1}$$

where p is the sound pressure, and c_0 is the sound speed in the medium. The K-space pseudospectral method is a numerical method for solving partial differential equations. It employs the Fourier transform to map data from the spatial domain to the wavenumber domain.

$$F_I\left\{\frac{\partial f(I)}{\partial I}\right\} = -\frac{1}{2\pi}\int f(I)(-ik_I)e^{-ik_I I} dI = ik_I F_I\{f(I)\} \tag{2}$$

where F_I is the spatial Fourier transform, i is a spatial direction in the Cartesian coordinate system. In three-dimensional cases, I=x,y,z, and $\partial/\partial i$ denotes the spatial partial derivative with respect to the I direction. The symbol $F_I\{\}$ is the one-dimensional Fourier transform operator in the I direction.

In the time domain, a finite difference scheme is employed to calculate the time derivative. For a homogeneous medium, Eq.(1) is transformed into the wavenumber domain via spatial Fourier transformation, the second-order difference of the time derivative is calculated. A relatively small time step can effectively control the calculation error. To achieve higher computational efficiency, a larger time step is adopted, based on the homogeneous solution of the ideal acoustic wave equation:

$$p(\mathbf{k},t) = \cos(c_0 kt)\, p(\mathbf{k},0) \tag{3}$$

An exact pseudospectral scheme for Eq.(1) can be derived by substituting Eq.(3) into the leapfrog finite difference $\mathrm{p}(\mathbf{k},t+\Delta t)$-$2p(\mathbf{k},t)$+$p(\mathbf{k},t-\Delta t)$. After some rearrangement,

this yields the relationship

$$\frac{p(\mathbf{k},t+\Delta t)-2p(\mathbf{k},t)+p(\mathbf{k},t-\Delta t)}{\Delta t^2 sinc^2\left(c_0 k\Delta t/2\right)}=-\left(c_0 k\right)^2 p(\mathbf{k},t) \tag{4}$$

In the given equation $k^2=\mathbf{k}\cdot\mathbf{k}=k_x^2+k_y^2+k_z^2$, where $\mathbf{k}$ is the wavevector, Δt is the time step.

For relatively large time steps, the *sinc* term provides an accurate solution while avoiding numerical dispersion. When deriving the fundamental system of first-order acoustic wave propagation equations, a *sinc* correction factor can be introduced to mitigate errors arising from time-domain finite differences. Notably, for the standard linear wave equation in a homogeneous lossless medium, an exact finite difference scheme for the time derivative exists. The aforementioned process enables the approximate discretisation of the second-order wave equation, the concomitant first-order conservation equations, and the pressure-density relationship. This discretisation is achieved through the implementation of a spatial operator, the *sinc* correction factor [22]:

$$\kappa = sinc^2\left(c_{ref} k\Delta t/2\right) \tag{5}$$

where c_{ref} is the reference sound speed in the medium.

2.2 Total Focusing Method

The TFM is an post-processing imaging technique based on FMC data. The method sequentially excites each array element while all elements receive echo signals to obtain the FMC data. For a phased array probe with N elements, $N\times N$ A-scan signals are acquired. The TFM imaging algorithm focuses the ultrasonic signals at a pixel point within the inspected region and synthesises the amplitude [23]. The schematic diagram of the TFM imaging is shown in Fig. 1.

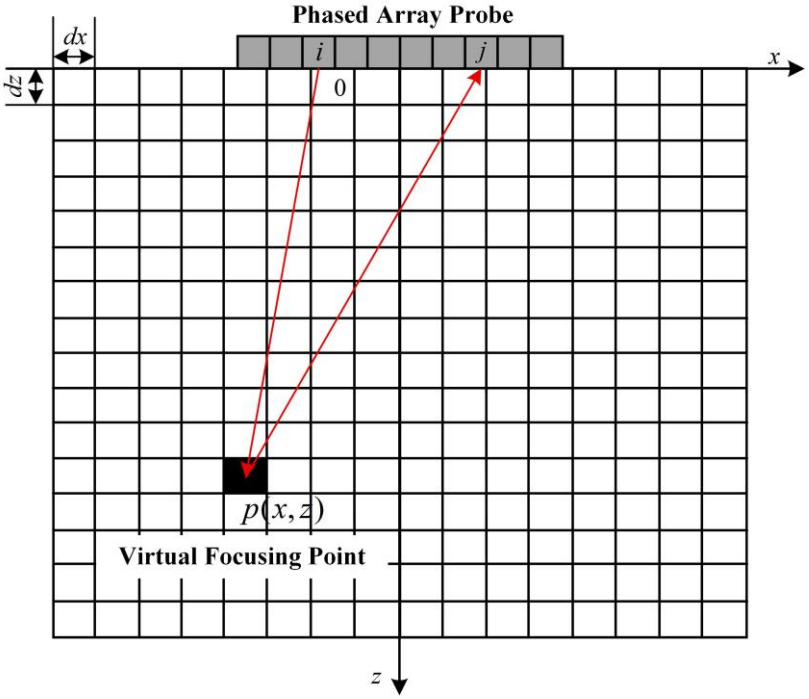


Fig.1. Schematic diagram of the TFM imaging

For any point p within the imaging region, defined by the coordinates (x, z), the amplitude $I(x, z)$ can be expressed as:

$$I(x,z)=\left|\sum_{i=1}^{N}\sum_{j=1}^{N}S_{ij}\left(\frac{\sqrt{(x_i-x)^2+z^2}+\sqrt{(x_j-x)^2+z^2}}{c_0}\right)\right| \tag{6}$$

where x_i and x_j is the transverse coordinates of the transmitting and receiving elements, respectively. S_{ij} is the A-scan signal transmitted by element i and received by element j.

2.3 Deep-Conditional Diffusion Model

This paper proposes a DCDM to overcome the limitation of traditional conditional diffusion models in generating TFM images of SDH at various positions. The model consists of a conditional input processing module and an enhanced U-Net backbone network.

The conditional input processing module employs a dual-condition input fusion mechanism that simultaneously receives two types of information, a position mask and a

32-dimensional feature map derived from depth embedding. The dual-condition input model provides joint constraints on both spatial structure and depth semantics for the subsequent generation model. The position mask generates a binary mask image according to the diameter and location of the SDH, this provides precise constraints for the spatial distribution of the defect. The schematic diagram of the simulated SDH position is shown in Fig. 2, the SDH is depicted within the cross-section by coordinates (a, b), where a is the horizontal direction (x-axis) and b is the depth direction (z-axis). The circular ring area corresponds to the actual size of the SDH. Concurrently, the discrete depth label b is fed into a learnable embedding layer and mapped to a 32-dimensional feature map. Through parameter learning, this embedding layer captures the semantic relationships between different depths, enabling the model to understand the impact of depth variations on image features. The depth feature has dimensions of 32×H×W, where H and W are height and width, respectively.

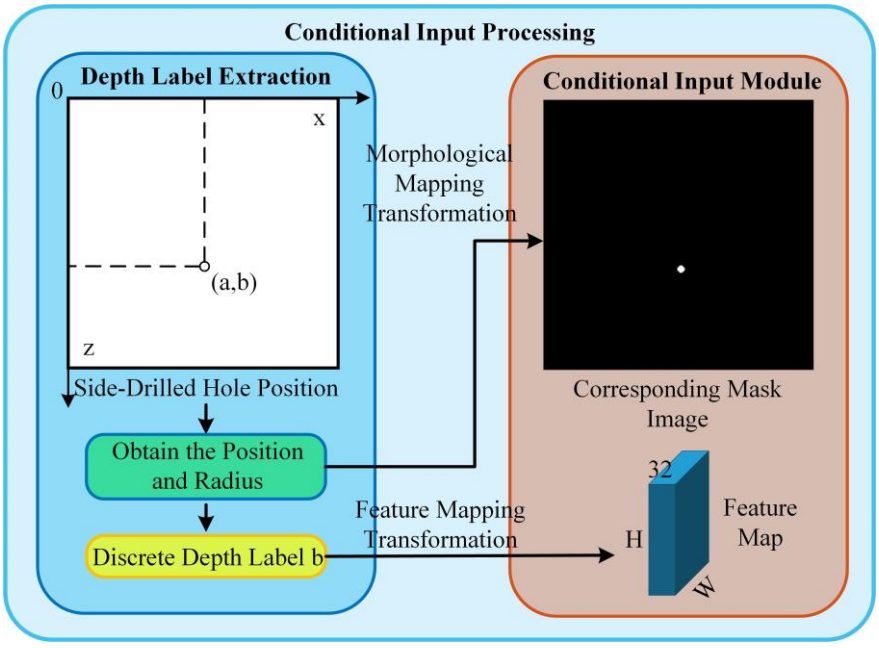


Fig.2. The conditional input processing module

The masks corresponding to the TFM images of SDH at different positions and the depth condition feature maps are concatenated with the noisy TFM images to form $x_{in}^{(t)}$:

$$x_{in}^{(t)} = Concat\left[M, x^{(t)}, \varepsilon(d)\right] \in \mathrm{R}^{(1+1+32)\times \mathrm{H}\times W} \tag{7}$$

where M is the position mask, $x(t)$ is the noisy image at time step t, and $\varepsilon(d)$ is the embedded feature map at depth d.

Within the framework of conditional diffusion models, the improved U-Net is commonly adopted as the core architecture. However, the input channels and conditional inputs of these models are typically designed for simple condition types in general vision tasks (e.g., image inpainting, text-to-image generation). In the specific context of generating TFM images for SDH at new positions, the conventional model is unable to process the noisy image, mask, and depth information concurrently, and the characteristics of defective images cannot be fully learned. Furthermore, the core convolution operations are limited by a local receptive field, which restricts the ability to model long-range dependencies between depth conditions and TFM image features. This hinders the model's ability to discern and replicate the variation patterns inherent in SDH image features across different depths. Additionally, the standard model has a high number of parameters and is computationally complex, resulting in poor efficiency in practical applications. To address these issues, this paper proposes an enhanced U-Net architecture based on the DDPM. As showed in Fig. 3, the enhanced U-Net backbone employs a four-level encoder-decoder configuration. The number of input channels for the single image is set to 34 (1 channel for the noisy image, 1 channel for the mask, and 32 channels for depth features). By incorporating depth embedding vectors and position mask images as conditional parameters at each network layer, the model ensures comprehensive processing of multi-modal conditional information, enabling precise condition-controlled generation. Self-attention modules are integrated into the second downsampling and upsampling blocks to enhance feature representation and generation quality. The base channel count is set to 32, compared to 64 in the original design, reducing computational complexity while maintaining representational capacity and improving training efficiency.

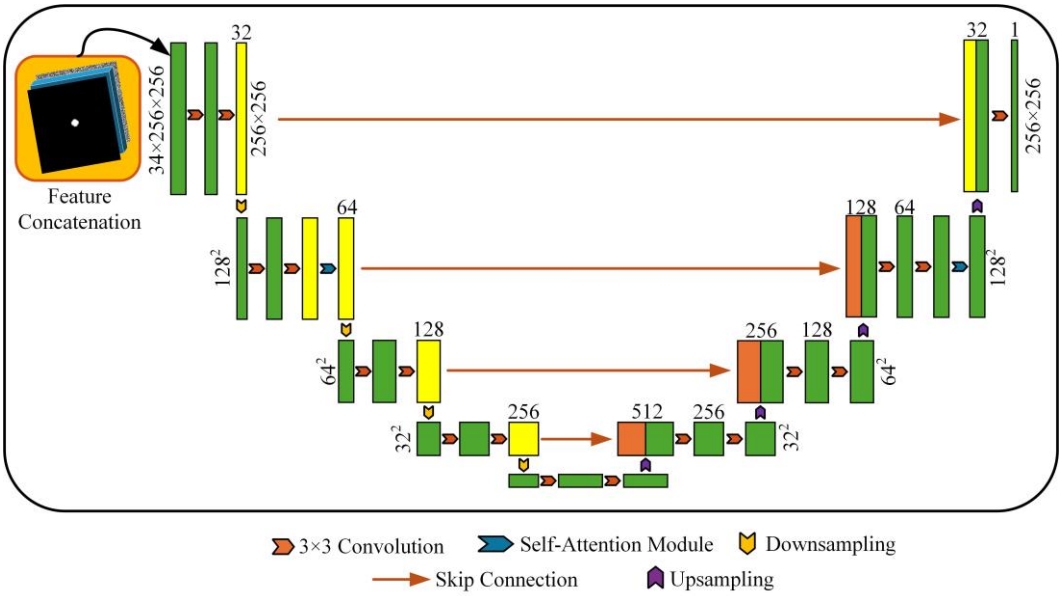


Fig.3. Schematic diagram of the improved U-Net network architecture

The proposed model follows the training paradigm of the DDPM. The objective is to train a denoising network ε_θ to predict the noise added during the forward diffusion process. The loss function is based on mean-squared error (MSE) for noise prediction:

$$J(\theta) = \mathrm{E}_{t,x_0,\varepsilon}\left[\left\|\varepsilon - \varepsilon_\theta(x^{(t)}, M, t, d)\right\|^2\right] \tag{8}$$

where θ is the learnable model parameters, and $\mathbb{E}$ is the expectation averaged over all time steps, training samples, and random noise. The time step t is sampled from a uniform distribution $t \sim U(1,T)$. The noise term ε is sampled from a Gaussian distribution, and d is the input depth condition. The loss function minimizes the MSE between the predicted noise ε_θ and the actual noise ε added to the image.

Fig. 4 shows the network architecture of the DCDM and the AIM generation process. Based on the simulated and generated TFM images, a gate is applied along the z-axis at the depth of the SDH to obtain a TFM strip image. The maximum value of each column along the x-axis direction is then extracted to generate a wave amplitude line. Repeating these steps for TFM images at subsequent horizontal positions yields a set of wave amplitude lines, which are superimposed to form a composite wave amplitude line. Finally, the composite wave amplitude lines at all depths are interpolated along the depth dimension to generate the final AIM.

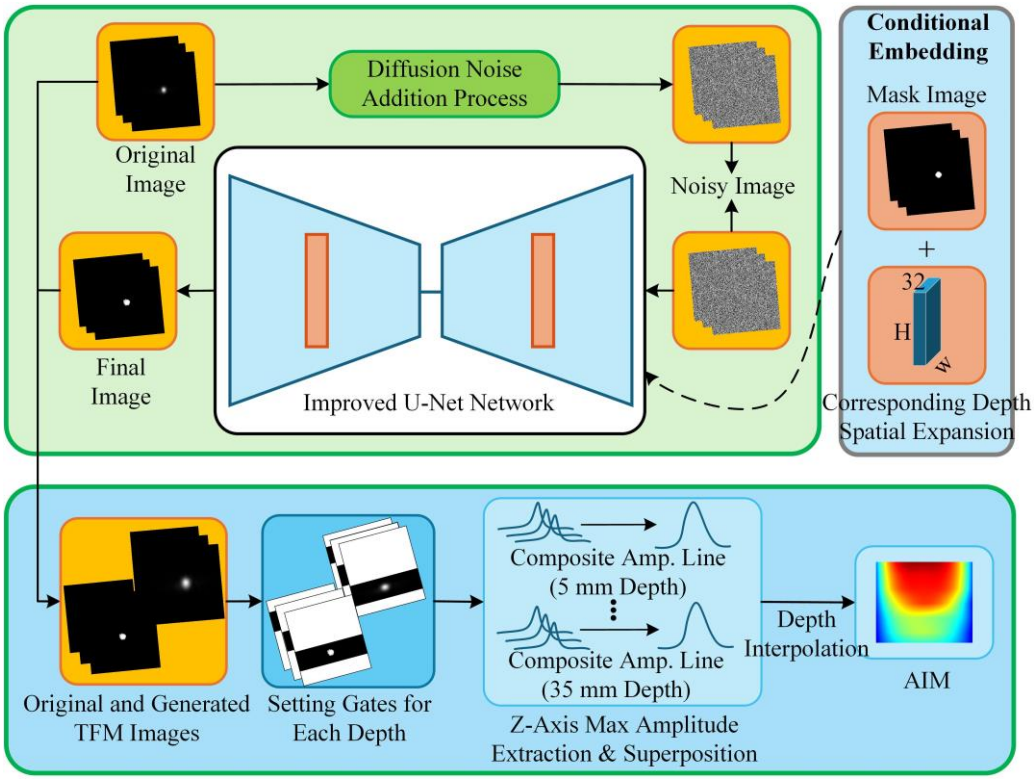


Fig.4. Flowchart for drawing the AIM based on the DCDM

3. Dataset Construction and Model Evaluation

3.1 Constructing the Total Focusing Method Image Dataset

The FMC data acquisition model for a single SDH is shown in Fig. 5. The ultrasonic phased array probe used in this study has a center frequency of 5 MHz, a pitch of 0.5 mm, and consists of 64 elements. The probe is positioned at the center of the top surface of a 40 mm × 40 mm × 5 mm workpiece. The FMC acquisition model is established using k-Wave [24], with a grid spacing of 0.022 mm in both horizontal and vertical directions and a time step of 10.6 ns. The medium is defined as stainless steel, the longitudinal wave velocity is 5790 m/s and the density is 7930 kg/m³. These settings meet the requirement of having more than five grid points per wavelength and a Courant-Friedrichs-Lewy (CFL) number greater than 0.3 [25]. To effectively suppress boundary reflections, an absorbing boundary condition in the form of a Perfectly Matched Layer (PML) with a thickness of 10 grid points is applied in the x, y, and z directions. All computations are performed on

a workstation equipped with an Intel ® Core™ i7-8700K CPU @ 4.00 GHz.

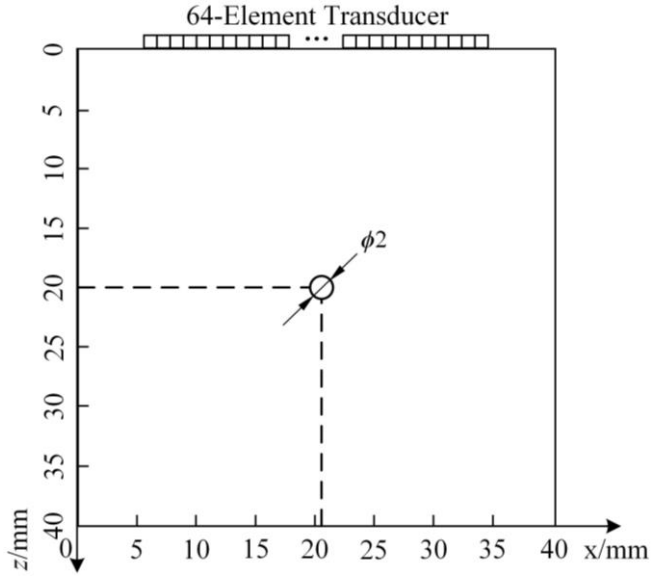


Fig.5. Simulation model for FMC of a single SDH

The excitation signal of the ultrasonic phased array probe can be expressed as a multi-cycle sine wave processed with a Hanning window,

$$F(t) = A\left[1-\cos\left(\frac{2\pi ft}{n}\right)\right]\cdot\sin\left(2\pi ft\right) \tag{9}$$

where A is the amplitude, f is the center frequency of the probe, and n is the number of cycles. When the frequency is set to 5 MHz and the number of cycles is 5, the corresponding excitation signal is shown in Fig. 6.

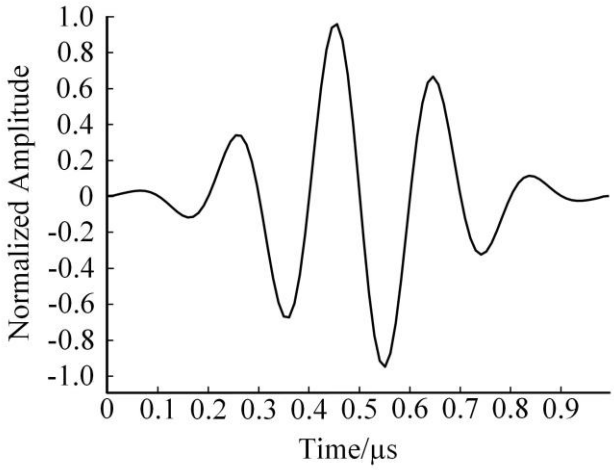


Fig.6. Waveform of the probe excitation signal

The training set of TFM images for SDH was constructed by seven distinct depths in the depth direction (from 5 mm to 35 mm, in steps of 5 mm). At each depth, 50 positions along the horizontal axis, ranging from 2 mm to 38 mm with a 0.72 mm step size, were selected. A total of 350 FMC datasets were acquired sequentially, and 350 TFM images with resolution of 400 × 400 were obtained. The dataset of TFM images for SDHs at different positions is shown in Fig. 7.

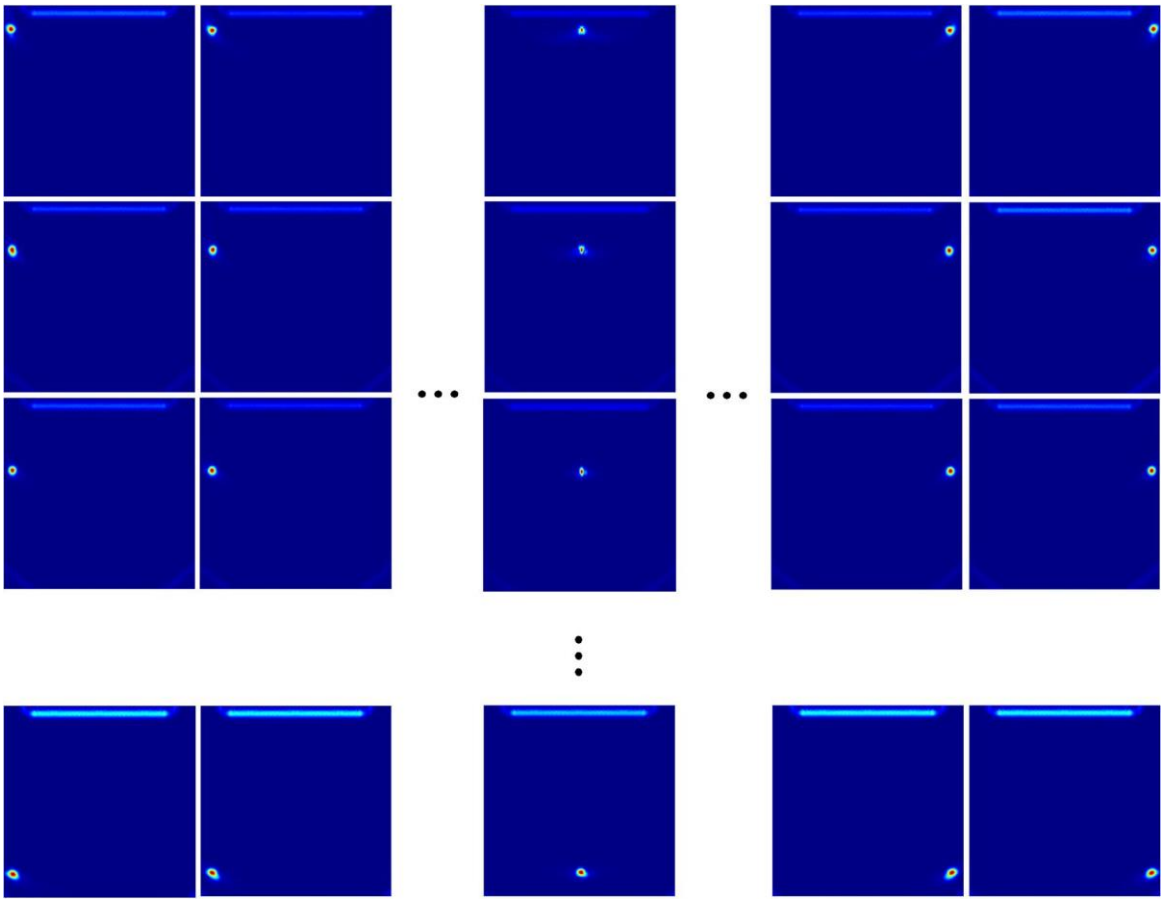


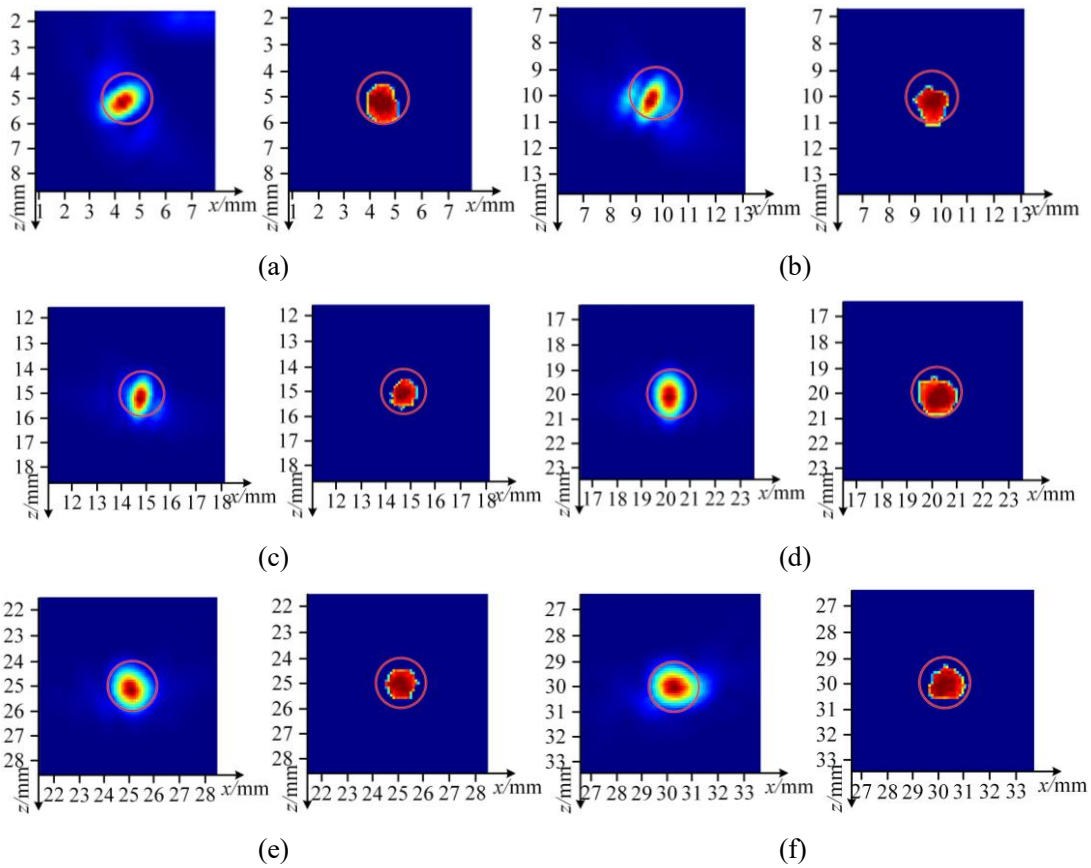
Fig.7. Dataset of TFM images for SDH at various positions

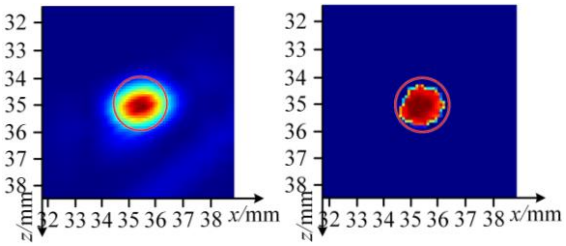
3.2 Evaluation of the Depth-Conditioned Diffusion Model

The model was trained with PyTorch on an NVIDIA GeForce RTX 4060 Ti GPU. Before the training, all TFM images were resized to 256 × 256. For the generative network, the initial learning rate was set to 1×10^{-4}, the batch size to 4, and the total number of training epochs to 50. The linear variance scheduling strategy proposed by DDPM was adopted, with a total of 500 diffusion steps being set to strike a balance between

generation quality and sampling efficiency. The AdamW optimizer was employed, which incorporates decoupled weight decay to improve generalisation and uses the OneCycleLR learning rate scheduler to accelerate convergence. New mask images were inserted at positions intermediate between adjacent training mask positions along the horizontal direction at each depth. The trained network model was used to generate corresponding TFM images of SDH at these novel mask positions, a total of 343 new images across the seven depths were obtained.

To verify the image generation performance of the proposed model, the Comparison of simulated and generated TFM images at different positions were showned in Fig. 8. Both the generated and simulated TFM images have a resolution of 256 × 256. To display the features of the images more clearly, a 50 × 50 region of interest centered on the SDH was cropped and magnified. The simulated TFM image is displayed on the left, whereas the generated TFM image is shown on the right. The red circle in each figure represents the actual SDH.





(g)

Fig.8. Comparison of simulated and generated TFM images at different positions. The coordinates for each subfigure are as follows: (a) (4.57, 5.00), (b) (9.71, 10.00), (c) (14.86, 15.00), (d) (20.00, 20.00), (e) (25.14, 25.00), (f) (30.29, 30.00), (g) (35.43, 35.00).

Structural Similarity Index Measure [26] (SSIM) and Pearson Correlation Coefficient [27] (CC) were adopted to evaluate the quality of the generated images. SSIM quantifies the similarity between two images in terms of luminance, contrast, and structure. CC ensures that the amplitude values extracted from the generated images are strongly correlated with the ground truth. This is essential for achieving quantitative accuracy in amplitude profiles. Meanwhile, SSIM guarantees the structural fidelity of the SDH indications, ensuring the correct localization of peak positions that are essential for the amplitude line analysis. The SSIM is defined as:

$$SSIM\left(x,y\right)=\frac{\left(2\mu_x\mu_y+c_1\right)\left(2\sigma_{xy}+c_2\right)}{\left(\mu_x^2+\mu_y^2+c_1\right)\left(\sigma_x^2+\sigma_y^2+c_2\right)} \tag{10}$$

where μ_x and μ_y are the mean pixel values of the simulated image x and the generated image y, respectively, and these values characterize the luminance component; σ_x and σ_y are the standard deviations of the pixel values of the two images, which characterize the contrast component; c_1 and c_2 are constants introduced to ensure numerical stability and avoid division by zero. The c_1 and c_2 is:

$$c_1=\left(k_1L\right)^2,c_2=\left(k_2L\right)^2 \tag{11}$$

where L is the dynamic range of pixel values, which is typically 255, while k_1 and k_2 are constants far smaller than 1. In this particular instance, k_1 is set to 0.01 and k_2 to 0.03. The SSIM value ranges from 0 to 1, with 1 indicating identical images.

The Pearson Correlation Coefficient (CC) quantifies the linear correlation between

the pixel values of two images, the CC is defined as:

$$CC(x,y)=\frac{\sum_{p=1}^{N}(x(p)-\mu_x)(y(p)-\mu_y)}{\sqrt{\sum_{p=1}^{N}(x(p)-\mu_x)^2}\cdot\sqrt{\sum_{p=1}^{N}(y(p)-\mu_y)^2}} \tag{12}$$

where p is the pixel index, and N is the total number of pixels of the image.

The evaluation results of the generated image are showned in Table 1. The SSIM values ranged from 0.86 to 0.93, indicating high structural similarity between the generated and simulated images. This shows that the proposed DCDM model successfully preserves the global structure and local defect details of TFM images. The SSIM value initially increased and then decreased, reaching its maximum at a depth of 20 mm. This trend is consistent with the spatial sound field distribution characteristics of ultrasonic phased arrays, as the acoustic energy degrades due to wave interference in the near-field zone and acoustic attenuation in the far-field zone. The CC values range from 0.68 to 0.86, with values reaching 0.85 at depths of 15 mm, 25 mm, 30 mm, and 35 mm, indicating a strong linear correlation between the pixel amplitudes of the generated and simulated images. These results confirm that the proposed DCDM model can accurately reconstruct the geometric structure of SDH defects in TFM images, and exhibits excellent and reliable generalization performance.

Tab.1 Quality analysis results of the generated images

Position(x,z)	SSIM	CC
(4.57,5.00)	0.87	0.70
(9.71,10.00)	0.89	0.68
(14.86,15.00)	0.92	0.82
(20.00,20.00)	0.93	0.72
(25.14,25.00)	0.91	0.86
(30.29,30.00)	0.89	0.85
(35.43,35.00)	0.86	0.86

4. Acoustic Influence Map Generation and Validation

4.1 Acoustic Influence Map Generation

The AIM was constructed from a total of 693 TFM images, including 350 images acquired via numerical simulation and 343 images synthesized by the well-trained DCDM model. The AIM generation process comprised three sequential steps. Firstly, a depth gate was applied along the z-axis to the TFM image of the SDH at a specific depth, the TFM strip image was obtained and was shown in Fig. 9. Secondly, the maximum amplitude value of each column along the x-axis was extracted to a single wave amplitude line was generated, as shown in Fig. 10. This process was repeated for TFM images at every horizontal positions, and all extracted wave amplitude lines were superimposed to form a composite wave amplitude line. Finally, the corresponding composite wave amplitude lines were obtained for the seven pre-defined depths (5 mm, 10 mm, 15 mm, 20 mm, 25 mm, 30 mm, and 35 mm).

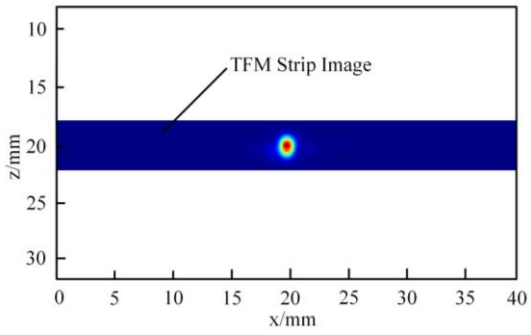


Fig.9. Gating at z= 20 mm

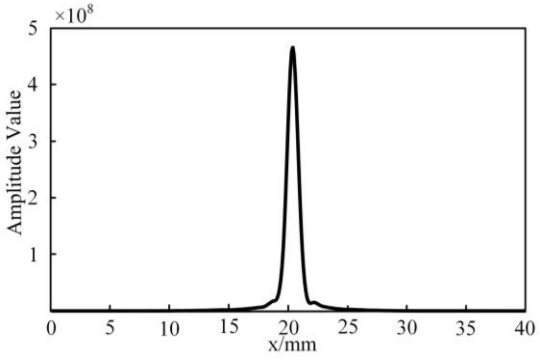


Fig.10. Amplitude line for a single SDH at z=20 mm

As shown in Fig. 11,the raw data for the composite wave amplitude lines at each depth fluctuates greatly, resulting in poor-quality AIM that does not depict the sensitivity distribution accurately. Therefore, a smoothing method based on key point spline interpolation was adopted to process the raw data. This method suppresses any sharp fluctuations while preserving the overall trend and key features of the original curves. Key points were selected on each composite wave amplitude line, and a smooth cubic polynomial curve was then generated between sharp fluctuations points. The piecewise cubic Hermite interpolation method was employed, which guarantees that the resulting smooth curve passes strictly through all predefined key points.

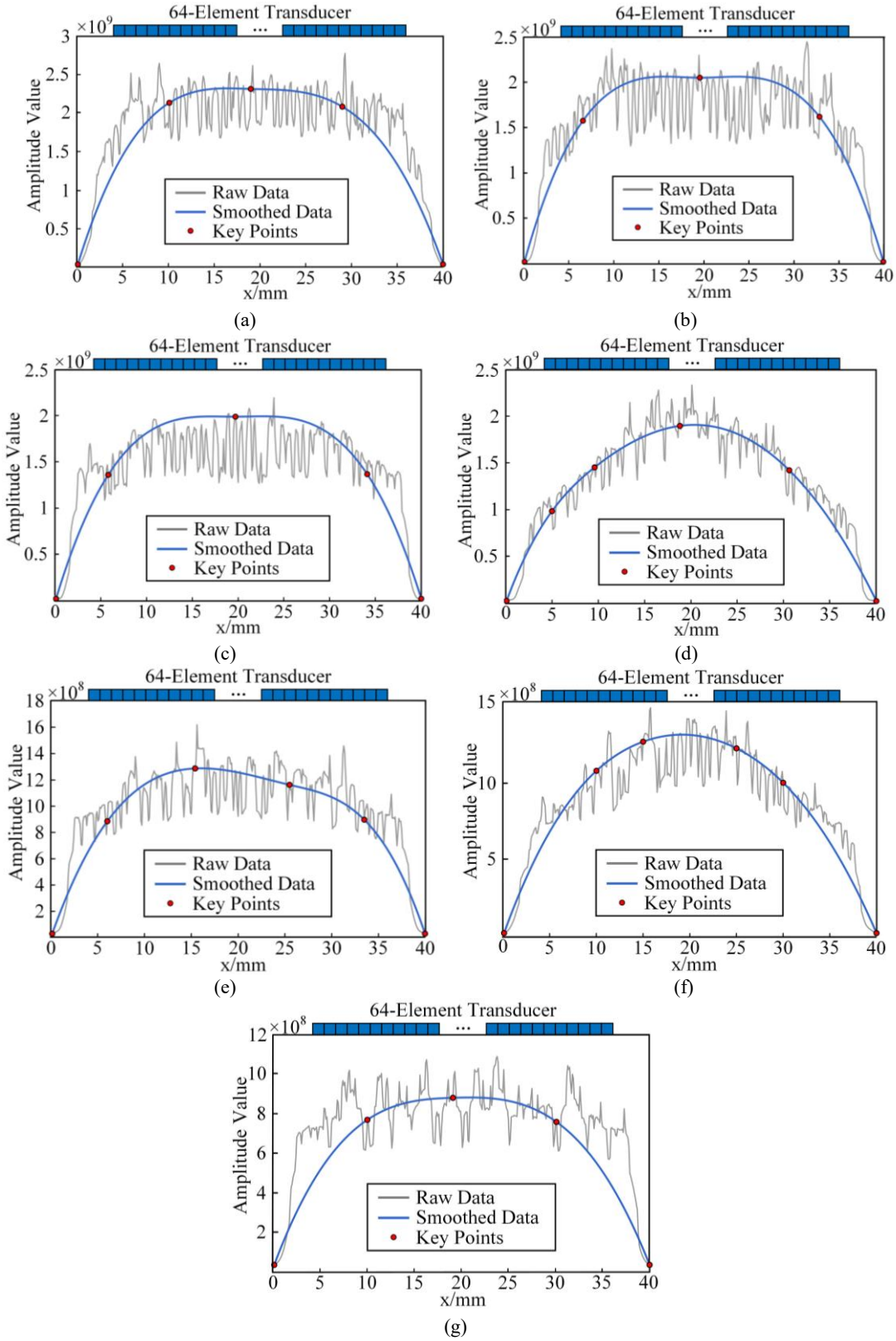


Fig.11. Comparative of key point selection and smoothing at various depths (5 mm to 35 mm) . (a) 5 mm, (b) 10 mm, (c) 15 mm, (d) 20 mm, (e) 25 mm, (f) 30 mm, (g) 35 mm.

Depth-wise interpolation was performed on the seven preprocessed composite wave amplitude lines to generate the AIM. The energy sensitivity of TFM imaging at various positions was clearly shown in Fig. 12. In the near-field region close to the phased array probe was the high-sensitivity region, the delay-and-sum (DAS) beamforming process of TFM enhances the coherent superposition of echoes extracted from FMC data. This is due to the short ultrasonic propagation path and the low acoustic energy attenuation. As the ultrasonic propagation distance increases, beam spreading becomes more severe, leading to increased echo attenuation. Meanwhile, the coherence of signals received by each array element decreases, which weakens the synthetic focusing performance and further reduces the detection sensitivity. This region is characterised by medium-sensitivity. As the ultrasonic wave propagates into the far-field region, the energy sensitivity gradually decreases, this region is characterised by low-sensitivity. Based on the constructed AIM, it is possible to quantitatively evaluate the detection capability of the TFM system.

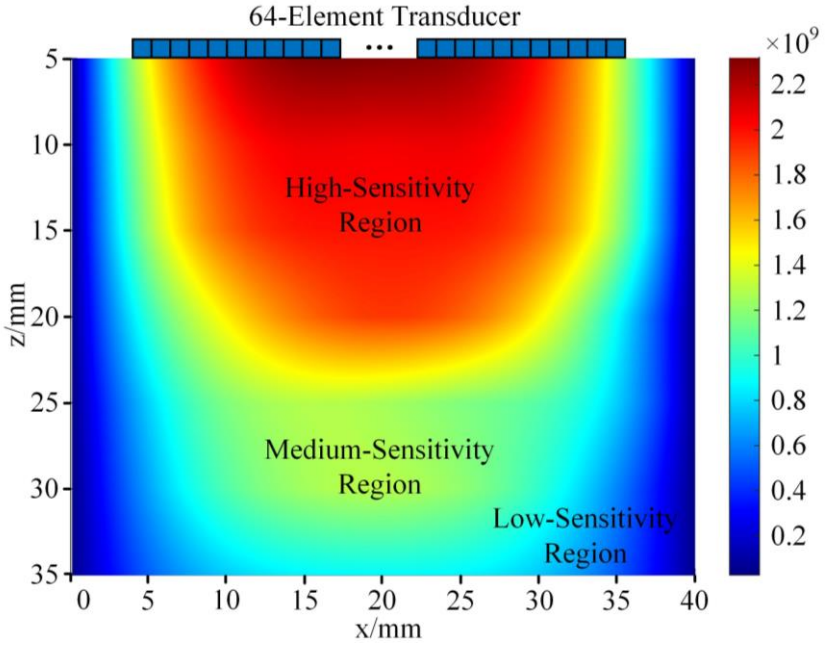
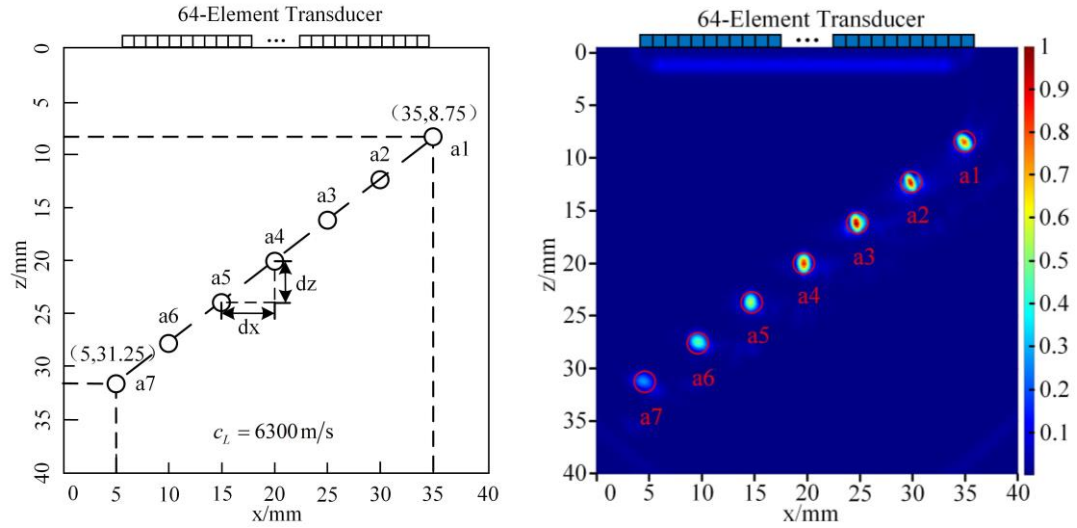


Fig.12. The obtained AIM

4.2 Validation

A series of simulation experiments were conducted to validate the reliability of the

AIM generated by the DCDM. The simulation model was shown in Fig. 13(a). The SDHs were labeled a1, a2, …, each with a diameter of 2 mm. The horizontal position interval was 5 mm, and the depth position interval was 3.75 mm. FMC data were obtained, and the TFM image was shown in Fig. 13(b). A1, A2, A3 and A4 were clearly visible, and the amplitude was relatively high. A5 and A6 were also clear, although their amplitude had decreased. However, the energy effect of A7 was poor, making it almost impossible to distinguish from artefacts.



(a) (b)

Fig.13. Schematic diagram of the simulation (a) and the TFM image (b)

As shown in Fig. 14, the SDHs are annotated on the constructed AIM. the a1, a2, a3, and a4 are located in the high-sensitivity region. a5 and a6 are lie in the medium-sensitivity region, and a7 falls within the low-sensitivity region.

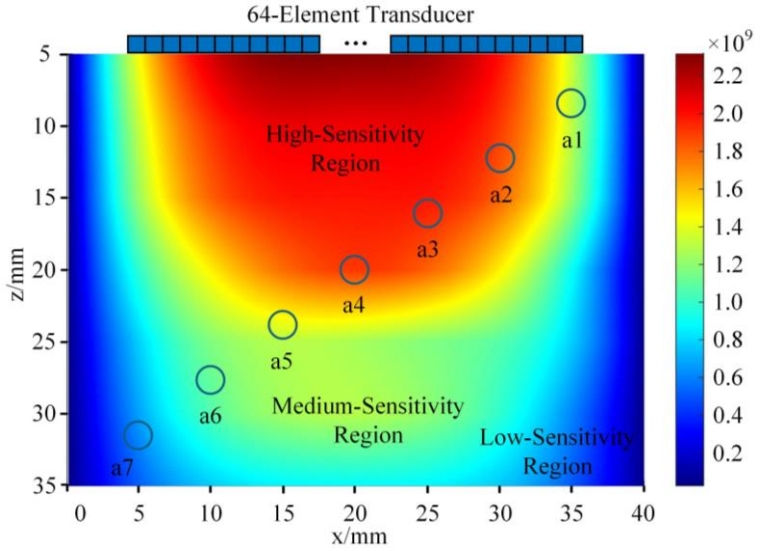


Fig.14. Sensitivity zoning of the AIM with marked obliquely arranged holes

The average amplitude values at the positions of the SDHs were extracted from both the AIM and the TFM image. As shown in Fig. 15, both curves demonstrate a high level of consistency in terms of their trends. The correlation coefficient between the two curves is 0.96, indicating a very strong positive correlation between the AIM results and the TFM imaging results.

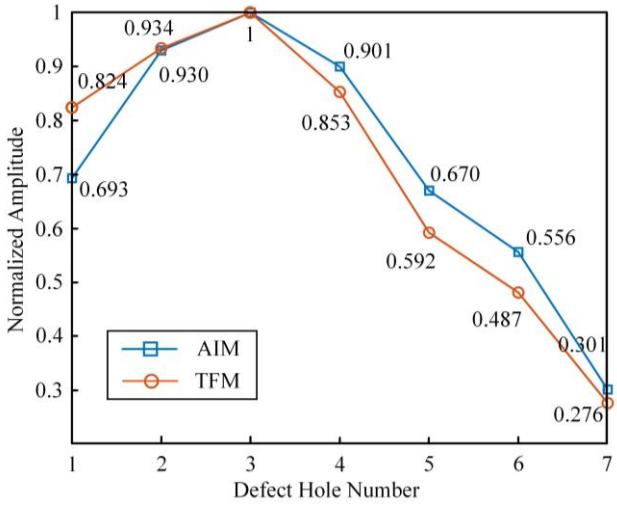


Fig.15. Amplitude comparison of the AIM and TFM images

5. Conclusion

In this paper, a high-efficiency construction method for the acoustic influence map (AIM) of ultrasonic total focusing method (TFM) imaging based on a conditional

diffusion model (DCDM) to solve the problem of low efficiency and high computational cost in traditional AIM generation.

The core of the proposed method lies in the dual-condition input mechanism of the DCDM. The model adopts an enhanced U-Net architecture with a four-level encoder-decoder structure, and converts discrete depth values into depth feature maps matching the image size through a learnable embedding layer. The depth features are then fused with the defect mask and noise image in the channel dimension, enabling the model to accurately capture the spatial variation law of TFM images of side-drilled holes (SDHs) at different positions, and realize high-fidelity TFM image generation for unseen defect positions without repeated full matrix capture (FMC) data acquisition and simulation.

The FMC datasets of SDHs at different positions were obtained via the K-space pseudospectral method, and the TFM image dataset was constructed for model training and validation. The evaluation results show that the SSIM and CC of the TFM images generated by the trained model for new positions reach up to 0.93 and 0.86, respectively, which proves that the DCDM can accurately reconstruct the geometric structure and amplitude characteristics of TFM images. The wave amplitude lines were extracted from both simulated and generated TFM images at each depth to construct composite wave amplitude lines, and the final AIM was obtained through key point spline smoothing and depth-wise interpolation. Quantitative comparison results show that the amplitude values derived from the constructed AIM are highly consistent with the actual amplitude of TFM images at the corresponding SDH positions, with a correlation coefficient as high as 0.96, which fully verifies the reliability and accuracy of the AIM generated by the proposed method.

The proposed method avoids massive repeated FMC simulation and TFM imaging calculations in the traditional AIM generation process, and provides an efficient and reliable tool for the sensitivity evaluation of TFM inspection systems and the optimization of practical inspection schemes. However, the current network has a relatively large parameter scale, resulting in limited training efficiency. Therefore, future work will focus

on the lightweight optimization of the network structure, to accelerate the model convergence and inference speed while maintaining the high quality of image generation.

Competing interests: All authors declare that we have no any competing interests as defined by Springer, or other interests that might be perceived to influence the results and /or discussion reported in this paper.

Disclosure statement： No potential conflict of interest was reported by the author(s)

Funding: This work is supported by the National Natural Science Foundation of China [Grant No.51705418], Scientific Research Program Funded by Education Department of Shaanxi Provincial Government [Grant No. 25JC068].

Reference

[1] Lyu D, Xiao X, Hu H, et al. Orientation characterisation of branched crack-like defects in curved components using ultrasonic array vector total focusing method[J]. Nondestructive Testing and Evaluation, 2026, 41(1): 185-202. https://doi.org/10.1080/10589759.2025.2461728

[2] Wu J ,Hu H ,Song Y , et al. Enhanced defect characterization in unidirectional CFRP composites using multi-modal TFM imaging with self-baseline subtract-and-weight techni que [J]. Measurement, 2026, 260 119916-119916. https://doi.org/10.1016/j.measurement.2025.119916

[3] Zhang H ,Zhang H ,Zhu Q , et al. Ultrasonic total focusing imaging of wrinkles in co mposites with maximum eigenvalue vector coherence factor weighting. [J]. Ultrasonics, 2025, 159 107847. https://doi.org/10.1016/j.ultras.2025.107847

[4] Yang H ,Li S ,Shu J , et al. Data-driven high-resolution total focus imaging from array ultrasonic time-domain signals of reinforced concrete material [J]. Construction and Bui lding Materials, 2025, 492 143048-143048.

[5] Zhang H,Si J,Zhang H, et al. Ultrasonic detection of wrinkles in composites with gradual phase shift migration. [J]. Ultrasonics, 2024, 148 107557. https://doi.org/10.1016/j.ultras.2024.107557

[6] Minari H V ,Anoni G L ,Haach G V , et al. Assessing bonding quality and detecting defects in CLT panels using ultrasonic wave propagation analysis [J]. Construction and Building Materials, 2025, 505 144789-144789. https://doi.org/10.1016/j.conbuildmat.2025.144789

[7] Wang H,Cheng J,Chen W, et al. Defect recognition and classification from ultrasonic p hased array total focusing method imaging based on domain adaption [J]. Journal of Ph ysics: Conference Series, 2024, 2822 (1): 012186-012186. DOI**:** 10.1088/1742-6596/2822 /1/012186

[8] Shi Y H ,Li C ,Zhou P , et al. Sound Field Simulation to Inspect Flat Bottom Defects in Titanium Alloy Small Diameter Bar Based on Ultrasonic Phased Array Technology [J]. Russian Journal of Nondestructive Testing, 2025, 61 (3): 295-308. https://doi.org/1

0.1134/S1061830924603556

[9] Gueugnaut D ,Tessier M ,Bechrouri M , et al. Acceptance Criteria for Defects in Polyethylene Welds, Coupling Phased Array Ultrasonic Testing and Destructive Tests [J]. Journal of Nondestructive Evaluation, 2025, 44 (1): 26-26. https://doi.org/10.1007/s10921-025-01165-1

[10] Wei D . Study on ultrasonic phased array detection of butt angle weld [J]. Journal of Physics: Conference Series, 2025, 2990 (1): 012008-012008. DOI：10.1088/1742-6596/2990/1/012008

[11] CHI-HANG K, GUILLAUME P, BENOIT L. TFM Acoustic Influence Map[C]. 2019 ASNT Research Symposium, USA, 2019.

[12] Guan S ,Hua L ,Wang X , et al. Ultrasonic TFM imaging inspection of large complex rings using defect quantitative evaluation mappings [J]. Nondestructive Testing and Evaluation, 2025, 40 (8): 3377-3398. https://doi.org/10.1080/10589759.2024.2402552

[13] Budyn N S. Imaging and defect characterisation using multi-view ultrasonic data in nondestructive evaluation[D]. University of Bristol, 2020. http://research-information.bristol.ac.uk

[14] Jiang D ,Tao N ,Zhu K , et al. ImbDef-GAN: Defect Image-Generation Method Based on Sample Imbalance [J]. Journal of Imaging, 2025, 11 (10): 367-367. https://doi.org/10.3390/jimaging11100367

[15] Yang X ,Chinchilla C S ,Moradi M , et al. Damage imaging in structural health monitoring with fine-tuned conditional diffusion model [J]. Mechanical Systems and Signal Processing, 2025, 236 112996-112996. https://doi.org/10.1016/j.ymssp.2025.112996

[16] Wang X ,Luo K ,Liu M , et al. UGDS-CDM: A novel uncertainty-guided dual-stage co nditional diffusion model and vision mamba-KAN integrating expert knowledge for surfa ce defect detection under small samples [J]. Advanced Engineering Informatics, 2026, 6 9 (PC): 104011-104011. https://doi.org/10.1016/j.aei.2025.104011

[17] Tai Y ,Zhang Z . Data expansion for industrial defect detection with diffusion prior [J]. Engineering Applications of Artificial Intelligence, 2025, 160 (PA): 111835-111835. https://doi.org/10.1016/j.engappai.2025.111835

[18] Liu Y ,Feng Y ,Cheng J , et al. Mamba Diff: Mamba-Enhanced Diffusion Model for 3D Medical Image Segmentation. [J]. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society, 2025, PP. DOI: 10.1109/TIP.2025.3607615

[19] Lee D ,Lee J H ,Park S C , et al. DiffectNet: diffusion-enabled conditional target generation of internal defects in ultrasonic non-destructive testing [J]. Mechanical Systems and Signal Processing, 2025, 240 113454-113454. https://doi.org/10.1016/j.ymssp.2025.113454

[20] Zhou Z ,Zhou S ,Li S , et al. Tunnel lining quality detection based on the YOLO-LD algorithm [J]. Construction and Building Materials, 2024, 449 138240-138240. https://doi.org/10.1016/j.conbuildmat.2024.138240

[21] Ren Y ,Fu X ,Zhang J , et al. Laser ultrasonic detection of defects in a regular tetrahedral lattice structure based on finite element modeling [J]. Mechanics Research Communications, 2025, 148 104466-104466. https://doi.org/10.1016/j.mechrescom.2025.104466

[22] Chupova D D ,Rosnitskiy B P ,Solontsov V O , et al. Compensation for Aberrations When Focusing Ultrasound Through the Skull Based on CT and MRI Data [J]. Acousti cal Physics, 2024, 70 (2): 288-298. https://doi.org/10.1134/S1063771024601651

[23] Liu Z ,Li Z ,Lyu D , et al. Imaging method of surface defect using leaky Rayleigh wave with

ultrasonic phased array[J]. Nondestructive Testing and Evaluation, 2025, 40 (6): 2311-2328. https://doi.org/10.1080/10589759.2024.2377726.

[24] Kale KD, Naik A. Single-crystal ultrasound transducer for cancellous bone health assessment: A k-wave simulation approach. International Journal of Advanced Technology and Engineering Exploration. 2025;12(128):1162-1174. DOI: 10.19101/IJATEE.2024.111101422

[25] He H,Zheng L,Changshan H, et al. Auto-tuning numerical method for acoustic wave si mulation using analytical solution. [J]. Annual International Conference of the IEEE Eng ineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2023, 2023 1-4. DOI: 10.1109/EMBC40787.20 23.10341031

[26] Illya B ,Marco B ,Raimondo S , et al. Structural similarity index (SSIM) revisited: A d ata-driven approach [J]. Expert Systems With Applications, 2022, 189. https://doi.org/1 0.1016/j.eswa.2021.116087

[27] Deng S ,Zhang J ,Huang Y , et al. A revisit to Pearson correlation coefficient under multiplicative distortions [J]. Communications in Statistics - Simulation and Computation, 2025, 54 (8): 3061-3083. https://doi.org/10.1080/03610918.2024.2333352