Engineering Applications of Artificial Intelligence 162 (2025) 112500



Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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





Image-enhanced generative model for industrial scanning acoustic super-resolution in non-destructive testing

Trong Nhan Nguyen [a,1], Vu Hoang Minh Doan [b,1], Tan Hung Vo [b], Dinh Dat Vu [a], Le Hai Tran [a], Thi Thu Ha Vu [a], Truong Tien Vo [c], Quoc Dung Nguyen [a], Sudip Mondal [d], Jaeyeop Choi [b], Byeongil Lee [a,d,*], Junghwan Oh [a,b,d,e,**]

[a] Industry 4.0 Convergence Bionics Engineering, Department of Biomedical Engineering, Pukyong National University, Busan, 48513, Republic of Korea
[b] Smart Gym-Based Translational Research Center for Active Senior's Healthcare, Pukyong National University, Busan, 48513, Republic of Korea
[c] Department of Engineering and Technology, Pham Van Dong University, Quang Ngai, Vietnam
[d] Digital Healthcare Research Center, Pukyong National University, Busan, 48513, Republic of Korea
[e] Ohlabs Corp., Busan, 48513, Republic of Korea

ARTICLE INFO

Keywords:
Scanning acoustic microscopy
Super resolution
Generative adversarial network
Synthetic dataset
Model deployment

ABSTRACT

Despite significant progress in super-resolution (SR) studies using generative adversarial networks (GANs), the application of these technologies in industrial non-destructive testing (NDT) remains limited owing to the strict requirements. The scarcity of training datasets in this field, primarily due to the confidentiality of scanning images and their inherently low visual diversity, often results in poor SR performance. Meanwhile, the demand for high-resolution (HR) imaging within short acquisition times remains critical in industrial contexts. To address this, we introduce an approach to generate synthetic low-resolution (LR) datasets from HR images using a high-order degradation model. In addition, we propose the Scanning Acoustic Microscopy Generative Adversarial Network (SAMGAN) framework, a model designed to improve the resolution of industrial ultrasound images. The SAMGAN model incorporates transfer learning and a dynamic loss function that combines pixel-wise, perceptual, and adversarial losses. While the quantitative evaluation of the synthetic training dataset remains a challenge, our method achieves an average Peak Signal-to-Noise Ratio (PSNR) of 33.978 dB and Structural Similarity Index Measure (SSIM) of 0.906 on industrial testing data, representing 10.93 % and 5.1 % improvement, respectively, over the most competitive baseline models. Integrated into scalable web-based and software deployment frameworks, SAMGAN enables up to **a** 47.2 % reduction in scanning time without sacrificing image fidelity. These results highlight SAMGAN's practical feasibility as a deployable tool for accelerating HR inspection workflows across industrial environments. To further improve the model's generalizability, future work will focus on addressing the limitations of synthetic data fidelity through enhanced degradation modelling and broader validation.

1. Introduction

Ensuring the structural integrity and reliability of products is a fundamental objective of industrial non-destructive testing (NDT) (Abd Al Rahman and Mousavi, 2020; Nsengiyumva et al., 2021; Wang et al., 2020). The revolution in semiconductor technology and the advent of powerful graphics processing units (GPUs) have significantly enhanced the demands and applications of NDT methods. As the semiconductor industry evolves, the need for precise, high-resolution (HR) inspection techniques has become more critical (Fang et al., 2023). The miniaturization of components and increasing complexity of modern microelectronics necessitate imaging technologies that can detect the smallest

* Corresponding author. Industry 4.0 Convergence Bionics Engineering, Department of Biomedical Engineering, Pukyong National University, Busan, 48513, Republic of Korea.
** Corresponding author. Industry 4.0 Convergence Bionics Engineering, Department of Biomedical Engineering, Pukyong National University, Busan, 48513, Republic of Korea.
E-mail addresses: bilee@pknu.ac.kr (B. Lee), jungoh@pknu.ac.kr (J. Oh).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.engappai.2025.112500
Received 24 September 2024; Received in revised form 17 September 2025; Accepted 23 September 2025
Available online 15 October 2025
0952-1976/© 2025 Published by Elsevier Ltd.

defects and irregularities (Kumar and Arnold, 2022; Silva et al., 2023).

Among the various imaging techniques used in NDT, X-ray microscopy is well-known for its high spatial resolution and ability to reveal fine internal features in dense materials such as metals and semiconductors (Fevola et al., 2024). However, it comes with substantial practical and safety-related constraints. X-ray systems require heavy radiation shielding, specialized infrastructure, and adherence to safety protocols due to the use of ionizing radiation, which poses health risks and limits portability. Besides that, X-ray techniques tend to underperform when applied to low-density materials or multilayered polymeric structures, where contrast is poor due to minimal absorption differences among layers. Terahertz (THz) imaging offers an alternative that is non-ionizing, contactless, and safe for biological and composite materials (Chopard et al., 2021). Despite these advantages, its practical utility in industrial NDT is restricted by two key limitations: (1) its relatively low spatial resolution, as THz waves have longer wavelengths that inherently limit the level of detail resolvable in images; and (2) its poor penetration in conductive or metallic materials, which are common in NDT industries. These constraints reduce the versatility of THz imaging in HR internal inspections of many industrial components.

In contrast, Scanning Acoustic Microscopy (SAM) presents a compelling balance of safety, resolution, and material compatibility (Yu, 2020). By leveraging high-frequency acoustic waves, SAM can achieve high resolution, particularly well-suited for detecting delamination, voids, and cracks in layered or composite structures, such as semiconductor packages or bonded joints. Importantly, unlike X-rays, SAM does not rely on ionizing radiation, and unlike THz imaging, it performs well in both soft and hard materials, including metals, polymers, and ceramics. This makes SAM a preferred modality in industries requiring internal inspection of complex assemblies without destructive sectioning.

Along with the above advantages, SAM also comes with several drawbacks when applied to HR industrial inspections. First, its reliance on a mechanical raster scanning strategy inherently leads to prolonged scanning durations. In this method, the transducer physically traverses the sample in discrete steps to acquire spatial data. When the step size is small (for HR imaging), the number of required scan points increases dramatically, which directly translates to longer acquisition times (Somani et al., 2023). For example, halving the step size results in doubling the scanning time. Second, SAM's performance is affected by depth-dependent acoustic signal attenuation. As ultrasound waves penetrate deeper into a material, particularly those with high acoustic impedance or internal layering (e.g., semiconductor packages or composite structures), signal strength diminishes significantly, reducing contrast and clarity in deeper regions (Martin et al., 2015). Third, SAM is highly sensitive to sample geometry and coupling conditions. Variations in surface curvature, tilt, or coupling medium uniformity can introduce signal artifacts or degradation, affecting the fidelity and repeatability of scans (Yu and Boseck, 1995). These factors present a bottleneck in achieving fast, high-resolution, and reliable imaging in industrial settings. This trade-off between resolution and efficiency poses a challenge in industrial applications where both HR imaging and rapid inspection are crucial (Liu et al., 2020; Xiao et al., 2020).

Traditionally, several approaches have been employed to address the need for HR imaging while managing scanning time. One common method involves the use of high-frequency transducers, which provide a higher resolution for the same scanning parameters (Moore et al., 2020; Morokov et al., 2021). However, this approach is limited by the high cost of high-frequency transducers and the reduced penetration of acoustic waves, because higher frequencies suffer from increased absorption (Kim, 2017). This limitation is particularly problematic when inspecting thick, multi-layered samples, which are essential in many industrial applications. Another advanced technique is Ultrasound Localization Microscopy (ULM), which achieves super-resolution (SR) by enhancing the resolution of traditional ultrasound imaging techniques (Chen et al., 2022; van Sloun et al., 2020). However, ULM is unsuitable for imaging non-biological materials, such as industrial samples, because it relies on microbubble contrast agents within the vascular system that are specific to biological tissues (Christensen-Jeffries et al., 2014; Hingot et al., 2021; Viessmann et al., 2013). Thus, alternative methods are required to increase image resolution without compromising scanning time and penetration depth in industrial environments. The comparison of these existing imaging and SR methods in biomedical and industrial NDT applications is shown in Table A1.

In recent years, generative models - particularly generative adversarial networks (GANs) - have shown remarkable effectiveness in various imaging applications (Deng et al., 2021; Lan et al., 2020; Liu et al., 2023; Ren et al., 2024; Zhou et al., 2023), including SR tasks (Demiray et al., 2021; Gu et al., 2020; Jiang et al., 2021; Lyu et al., 2024; Tian et al., 2022; Zhao et al., 2022). GANs are uniquely suited for SR problems because they combine pixel-wise accuracy with perceptual realism, allowing them to recover high-frequency textures and structural details that are often lost in low-resolution (LR) scanning. This capability is especially critical in SAM-based NDT, where subtle acoustic signal variations correspond to small defects such as delamination or microcracks. Unlike traditional interpolation or CNN-based SR methods that tend to smooth or blur high-frequency content, GANs can synthesize realistic textures in degraded regions and enhance defect visibility in noisy or imbalanced SAM data. This results in significantly improved image interpretability and diagnostic reliability for industrial applications.

Over the past decades, the ultimate objective in SAM has been to image every layer of an industrial sample at micron-level resolution while keeping inspection times to a minimum. However, SAM inherently suffers from a resolution–efficiency trade-off: reducing scan step size yields finer details but multiplies acquisition time, making high-throughput inspections impractical. Traditional approaches - such as switching to higher-frequency transducers or applying standard interpolation and filtering techniques - bring their drawbacks: elevated hardware costs, shallower penetration depths, and the blurring or loss of critical defect features. These limitations highlight the need for a fundamentally different approach that can deliver both speed and detail. To address these limitations, we propose a GAN-based SR framework specifically designed for industrial SAM imaging. Our method utilized a hybrid dataset that includes both real SAM image pairs and synthetically degraded LR data, created through a high-order degradation process that mimics real-world scanning artifacts. The GAN learns to reconstruct structurally and perceptually faithful SR images from LR inputs, enabling a substantial reduction in scanning time without compromising feature visibility or diagnostic reliability. This approach not only overcomes the resolution-efficiency bottleneck but also supports scalable deployment in practical industrial environments.

The main contributions of this research are mentioned below in detail.

(1) GANs are data-hungry models requiring extensive image datasets for effective training. However, in an industrial context, data are often limited and occasionally inaccessible. To address this issue, we proposed a framework that generates synthetic LR datasets from HR images using a high-order degradation model. This synthetic LR dataset was then combined with real LR data to enhance data diversity, leading to improved model performance.
(2) We developed the scanning acoustic microscopy generative adversarial network (SAMGAN), which includes both generator and discriminator models. The generator was built on the ESRGAN backbone, with modifications such as adding batch normalization and changing the activation function. The generator model was then subjected to transfer learning and fine-tuned to better suit the industrial image generation tasks. The discriminator model implemented using the UNet architecture

was designed to determine whether the output image from the generator was real or fake.

(3) We proposed a dynamic loss function to enhance the performance of the proposed model. This loss function combines pixel-wise, perceptual, and adversarial losses in optimized proportions to achieve improved results in the SR task.

(4) We deployed the proposed SAMGAN model and integrated its functionality into our web-based or software platforms to form a comprehensive SR framework. The model functions as an Application Programming Interface (API) within our industrial SAM system.

(5) Our results show that SAMGAN not only improves visual and structural quality metrics (PSNR and SSIM) but also reduces scanning time by up to 47.2 %, enabling faster inspections without compromising image quality or diagnostic reliability.

The remainder of this paper is organized as follows: Section 2 discusses related works. Section 3 presents the methodology used in our study, including data collection, synthetic LR image generation, network architecture, training method, model deployment, and evaluation metrics. Section 4 presents and discusses the results. Section 5 provides the conclusion, and Section 6 outlines the limitations and directions for future work.

2. Related works

Research on SR for industrial ultrasound images remains limited, with most studies being conducted in simulated environments or using simple datasets. Recently, Gao et al. proposed the LESS-CycleGAN, a label-enhanced semisupervised CycleGAN network model designed to improve defect characterization in ultrasonic testing images by eliminating diffraction artifacts (Gao et al., 2024). This model achieved high-precision defect reconstruction using paired cross-domain image samples and a novel authenticity loss function. However, their model was trained using simulated 2-dimensional (2D) images and tested on artificial defects in steel samples, with the final evaluations performed only on cross-sectional images. Similarly, Banerjee et al. introduced a deep learning approach to enhance the resolution of acoustic microscopy images (Banerjee et al., 2024). Their study demonstrated the advantages of transformer-based models and transfer learning in enhancing ultrasound image resolution. They used high-frequency transducer images as the ground truth (GT) and low-frequency transducer images as inputs. Despite these innovations, their study was limited to the top surface of coin samples because of shallow penetration issues. Both studies highlight considerable advancements but require validation with real industrial samples to assess their reliability in practical applications.

In the domain of SR techniques, GANs have shown remarkable success with several notable findings. The Super-Resolution Generative Adversarial Network (SRGAN), introduced by Ledig et al., pioneered the application of GANs for SR using a perceptual loss function to generate HR images with sharp details from LR inputs (Ledig et al., 2017). Building on this, the Enhanced SRGAN (ESRGAN), developed by Wang et al., improved upon SRGAN by incorporating Residual-in-Residual Dense Blocks (RRDB) and a Relativistic Discriminator, which together enhanced the network's ability to learn finer image details and produce more realistic results (Wang et al., 2018b). Further refining these concepts, the Real-ESRGAN extended ESRGAN's capabilities to real-world applications by incorporating more realistic degradation models and diverse data augmentation techniques during training, thereby achieving robust performance on a wide variety of real-world images (Wang et al., 2021). While SRGAN was pioneering in applying perceptual loss to produce sharp textures, it tends to generate artifacts and lose structural accuracy, especially in edge regions. ESRGAN improves detail recovery but requires extensive datasets and is sensitive to training variance, limiting its robustness in small-sample industrial domains. Real-ESRGAN tends to oversmooth outputs, potentially sacrificing fine defect features critical in NDT. The backbone architectures, key strengths, and limitations of three representative GAN-based SR models - SRGAN, ESRGAN, and Real-ESRGAN are summarized in Table A2. In this study, we developed a model built upon the ESRGAN backbone and reimplemented several related SR architectures, including SRGAN, ESRGAN, and Real-ESRGAN, to enable a comprehensive comparison with our framework. Additionally, other conventional SR models, such as Nina super-resolution (Nina-SR) (Gouvine, 2021), Cascading Residual Network (CARN) (Ahn et al., 2018), Residual Channel Attention Networks (RCAN) (Zhang et al., 2018b), and enhanced deep super-resolution network (EDSR) (Lim et al., 2017), have been utilized to offer a more objective evaluation and demonstrate the efficacy of our model in SR imaging tasks for industrial ultrasound applications.

3. Methodology

3.1. Data collection

All images were acquired using a SAM-modeled TSAM-400 (Ohlabs Corp, Busan, Korea) (Pham et al., 2022). Fig. 1 shows the scanning images process of a Power MOSFET. First, the ultrasound transducer was moved along the X-, Y-, and Z-axis to focus the acoustic beam and obtain an ultrasound signal at the desired position (A-scan signal). Subsequently, we fixed the Z-axis position of the transducer and scanned the sample along the Y-axis to acquire a B-scan, which was considered the

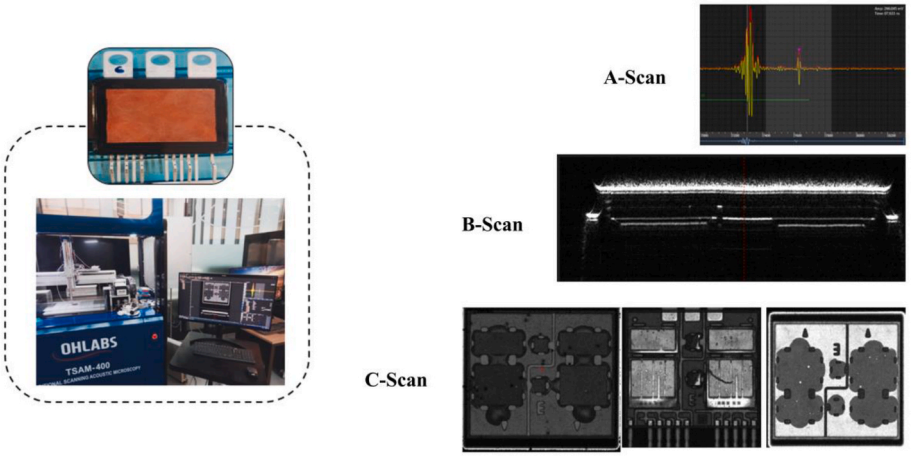


Fig. 1. Scanning data collection using TSAM-400 system.

Table 1
Sample name, size, and corresponding scanning times.

Samples no.	Name	Sample Size (Width x Height) (mm x mm)	Step Size (μm)	Scanning Time (Minutes)
1	Power MOSFET	42 x 70	50	4.84
			100	2.42
2	FPGA Chip	22 x 22	50	1.84
			100	0.92
3	Transistor	15 x 40	50	2.00
			100	1.00
4	Power Semiconductor	80 x 68	50	10.00
			100	5.00

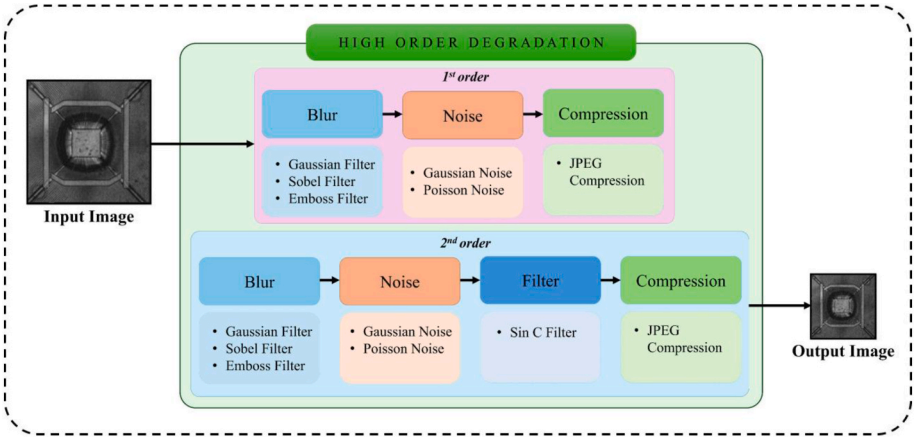


Fig. 2. Synthetic dataset generator process.

cross-section of our sample. The transducer was then adjusted to a certain step size along the X-axis and scanned again along the Y-axis. This process was repeated until the C-scan or the entire ultrasound image of the sample was obtained. Four different types of industrial samples, including Power MOSFETs, Field-Programmable Gate Array (FPGA) Chips, Transistors, and Power Semiconductors, were inspected using the echo mode and the aforementioned raster scan strategy above.

To ensure transparency and avoid data leakage, distinct physical samples were reserved for the training, validating, and testing subsets, meaning that all images derived from a given sample exclusively belonged to one subset. This strategy guarantees that the evaluation was performed on entirely unseen samples, thereby providing a more realistic assessment of the model's generalization ability. Details of the sample distribution across training, validating, and testing, as well as the number of corresponding scanned images, are summarized in Table A3 in the Appendix section. The defects present in these scanned images include micro-cracks, delamination, voids, and solder joint failures (Fig. A1). These defects vary in size, ranging from 20 μm to 500 μm, and their depths span from 10 % to 80 % of the total sample thickness. This diversity in defect characteristics ensures that the dataset captures a wide range of real-world failure scenarios commonly encountered in industrial applications. These scanning images were divided into two subsets based on their resolutions: HR images dataset - 50 μm step size and LR images dataset - 100 μm step size. The names, sizes, and corresponding scanning times of the industrial samples are listed in Table 1.

3.2. Synthetic LR image generation

Synthetic LR images generated from HR images via a degradation process were utilized as part of the training dataset for the SAMGAN model (Gong et al., 2023). The degradation process involves two stages, as illustrated in Fig. 2. In the first stage, the GT image undergoes random blurring (Gaussian filter, Sobel filter, or Emboss filter) using a predefined kernel ($kernel_1$). Gaussian or Poisson noise was then added based on the specified probabilities and noise ranges, and JPEG compression was applied using the DiffJPEG module with a randomly chosen quality factor. The second stage optionally applies another blur kernel ($kernel_2$) based on the given probability. Additional Gaussian or Poisson noise was introduced according to the new probabilities and noise ranges. Finally, a sinc filter was applied, and another JPEG compression was used to finalize this degradation step. This two-stage degradation pipeline was designed to more accurately mimic the complex, compound degradation effects observed in real industrial scanning environments, which often include layered distortions from multiple sources such as transducer interference, electronic noise, and lossy image compression. This approach extends the high-order degradation strategy introduced in Real-ESRGAN, emphasizing not just random ordering but intentional repetition of degradations to simulate realistic noise patterns. The increased diversity and realism in degradation patterns allow the model to generalize more effectively across various real-world conditions, improving restoration quality and robustness for industrial SAM data. The calculation for this process is provided in Equation (1) (Zhang et al., 2018a), as follows:

$$I_{LR} = D_{k,n,j}(I_{HR}) = [(k \otimes I_{HR})\downarrow s + n]_{jpeg} \quad (1)$$

where $D_{k,n,j}$ denotes the degradation method with a blur kernel, noise injection, and JPEG compression. k is the blur kernel and n is the noise injection method. I_{HR} is the HR image. I_{LR} is the LR image. $\downarrow s$ is the downsampling after calculating the convolution between the blur kernel and HR (s = 1, which means we maintain the same size of the images after convolution). *jpeg* denotes the JPEG compression.

3.3. Network architecture

3.3.1. Generator model

We developed the SAMGAN by enhancing the network architecture of the ESRGAN model, as illustrated in Fig. 3a. Our model utilized small kernels and incorporated 23 RRDB blocks (Fig. 3b). Each block within the network—including the convolutional layers, batch normalization, and Parametric ReLU (PreLU) activation—was meticulously designed to

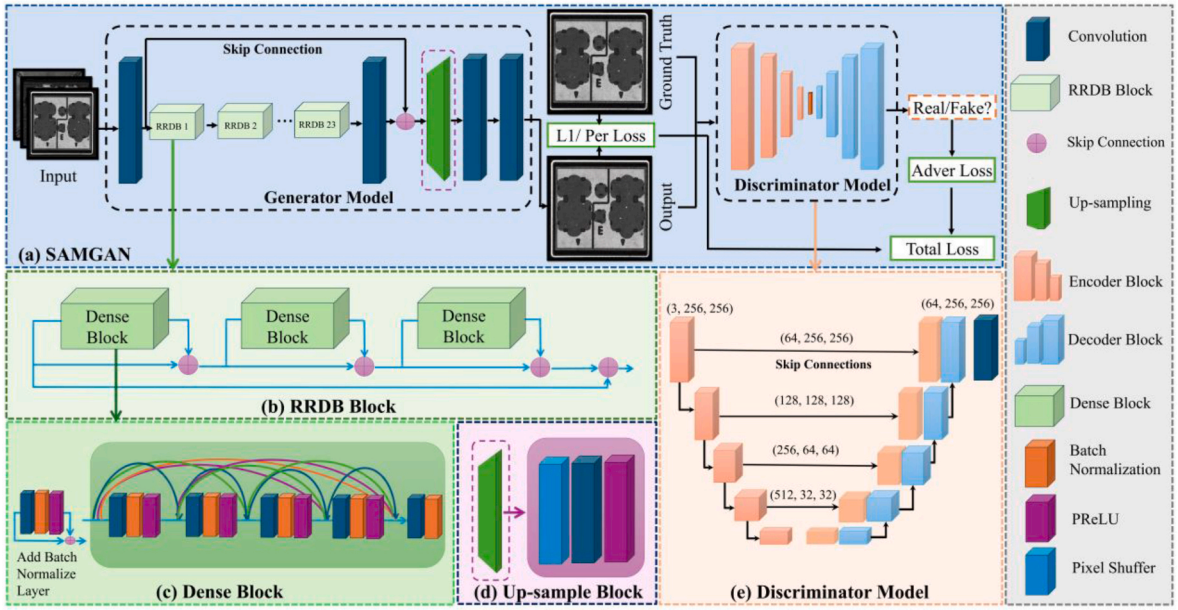


Fig. 3. Architecture of the proposed SAMGAN model. (a) Overall model framework comprises the Generator and Discriminator networks. (b) Residual-in-Residual Dense Block (RRDB) structure. (c) Dense block structure. (d) Up-sampling block structure. (e) Discriminator network structure.

manage the feature extraction (Fig. 3c). The up-sampling block in Fig. 3d utilizes a pixel shuffler block to increase the image size.

3.3.2. Discriminator model

As shown in Fig. 3e, we adopted a UNet-based discriminator architecture (Schonfeld et al., 2020), adapted from the Relativistic GAN used in ESRGAN, to better address the broad degradation characteristics inherent in industrial SAM images. The discriminator model uses a fixed input size of 256 × 256 × 3 and an initial feature depth of 64. Unlike a standard discriminator, which classifies an image as real or fake in a holistic manner, our UNet-based discriminator provides pixel-level discrimination through its encoder–decoder design with skip connections. This structure enhances the model's sensitivity to local spatial inconsistencies, which are critical in identifying subtle defects and material boundaries in SAM images. The rich multi-scale feedback enables the generator to learn finer structural details and preserve edge-level features, thus significantly contributing to the restoration fidelity required in industrial inspection scenarios.

3.3.3. Loss functions

Our SAMGAN model was trained using three primary loss functions: pixel-wise loss L_1 (Equation (2)) (Ledig et al., 2017), perceptual loss L_{Percep} (Equation (3)) (Ledig et al., 2017), and adversarial loss L_{Adver} (Equation (4)) (Wang et al., 2018b). The L_1 loss is effective in reconstructing image details, whereas the L_{Percep} loss focuses on capturing high-level features. The L_{Adver} loss enhances the realism of the generated images.

$$L_1 = \|x_{sr} - x_{hr}\|_1 \tag{2}$$

here, $\|\cdot\|$ denotes the L_1 norm, which computes the sum of absolute differences across all pixels, x_{sr} is the SR image generated by the model and x_{hr} is the GT image.

$$L_{Percep} = \|\phi(x_{sr}) - \phi(x_{hr})\|_2{}^2 \tag{3}$$

here, the $\|\cdot\|_2^2$ term calculates the squared L_2 norm, measuring the Euclidean distance between the feature representations of the generated and reference images. x_{sr} is the SR image generated by the model and x_{hr} is the GT image, $\phi(\cdot)$ is the VGG feature extractor network applied to the images.

$$L_{Adver} = -E_{x_{hr}}[1 - log(D(x_{hr}))] - E_{x_{sr}}[log(D(x_{sr}))] \tag{4}$$

where x_{sr} is the SR image generated by the model and x_{hr} is the GT image, $D(x_{hr})$, $D(x_{sr})$ is the discriminator output and $-E_{x_{sr}}$, $-E_{x_{hr}}$ is the negative expectation.

3.4. Training method

The dataset, comprising 1265 pairs of LR and HR scanning images, was divided into three subsets: training (80 %), validation (10 %), and testing (10 %). Initially, the generator model backbone was pre-trained using the DIV2K dataset (Agustsson and Timofte, 2017; Wang et al., 2018a) for 100,000 iterations. This pre-training enabled the model to learn general image features before fine-tuning on our dataset. Subsequently, the SAMGAN model was trained for 400,000 iterations at a learning rate of 0.0001, initialized with pretrained weights from the DIV2K dataset. The model was trained using our dataset, which included original images, multi-scale images, and sub-images randomly cropped from the dataset. The model was trained on synthetic images generated using a high-order degradation model. A separate set of 127 scanning images was used as the test dataset to evaluate the performance of the proposed model.

As shown in Fig. 4, our SAMGAN model was trained using a combination of L_1, L_{Percep}, and L_{Adver} losses, with a learning rate of 0.0001. The proportion of each loss in the combination was adjusted to 0.1, 1, and 0.005 to obtain the optimal SR images. The hardware platform employed in this study consisted of a high-performance computer equipped with a high-speed NVIDIA GeForce RTX 3090Ti GPU and 12 GB of graphics memory. The networks were configured using Python 3.9 and processed with CUDA 12.5. PyTorch 2.3.1, served as the backend for model development and training. This procedure provides the necessary computational resources and software environments to conduct the experiments effectively. The training procedure for the SAMGAN is described in detail in **Algorithm 1**.

Transfer learning with DIV2K was motivated by its rich feature diversity, enabling faster convergence and improved generalization on our limited industrial dataset compared to training from scratch. Similarly, the combination of L_1, perceptual, and adversarial losses was designed

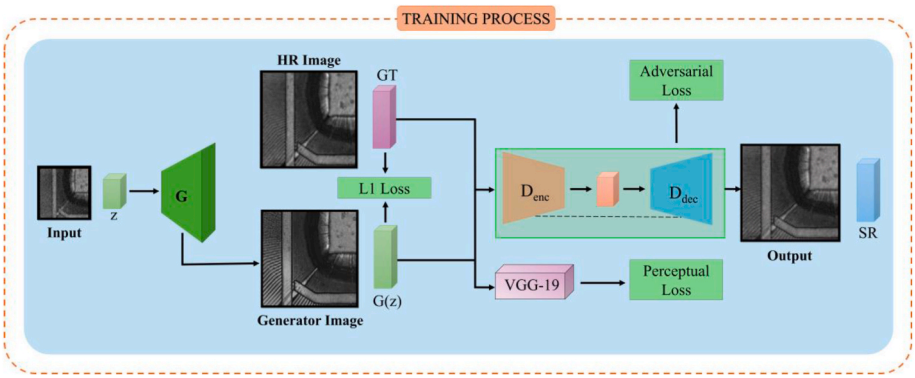


Fig. 4. Training process.

to balance pixel-level accuracy, structural fidelity, and visual realism, addressing the complex degradation patterns in industrial scanning images.

3.5. Model deployment

The deployment framework of the proposed SAMGAN is illustrated in Fig. 5. After training using PyTorch, the SAMGAN model was converted into an Open Neural Network Exchange (ONNX) format. This conversion enhances interoperability and allows the model to be executed across multiple platforms without relying on the PyTorch runtime. In particular, the ONNX format enables more efficient and lightweight packaging for deployment on containerized platforms, contributing to faster inference and smoother scalability. The ONNX model file was then packed into Docker for seamless deployment across different environments without encountering system configuration issues. The containerized model is stored in the Azure Container Registry, facilitating efficient management and distribution. To enable scalable and reliable deployment of the SAMGAN model, we utilize Azure Kubernetes Service (AKS) as the core orchestration layer. AKS automates the deployment and lifecycle management of containerized models by pulling Docker images from Azure Container Registry and distributing them across cluster nodes. It supports horizontal and vertical autoscaling to dynamically adjust compute resources in response to inference workloads, ensuring consistent performance. Integrated with Azure API Management, AKS enables secure and high-throughput API serving, while direct connections to Azure Data Lake provide robust support for large-scale data handling.

The model is accessible via APIs, with the input data and model results stored in the Azure Data Lake, which is a robust system for managing and processing large datasets. The Azure API Management secures, stabilizes, and scales APIs, enabling front-end applications to send requests and receive processed images through an API endpoint. Users can interact with the model through a custom web interface or software. To ensure optimal operation, the Azure Monitor tracks performance metrics, such as API response times, resource usage, and errors. Azure Cost Management provides real-time cost analysis and optimization by identifying unused or overused resources. The Azure Key Vault secures sensitive information, including encryption keys, API tokens, and certificates, which are crucial for protecting industrial scanning data and controlling access to the deployed SAMGAN model. It also enables role-based access control, ensuring that only authorized services can retrieve and use these secrets during inference.

To bridge the AI model with practical industrial usage, the scanning system was designed to support direct interaction with the deployed SAMGAN model. The user interface (UI) of the scanning system is developed in C# using DevExpress. The model integration is facilitated through a RESTful API, where the scanning software sends ultrasound scanning data to the deployed SAMGAN model via HTTP requests. The processed images are then returned and displayed within the existing UI framework. The API backend ensures high-speed communication between the scanning software and the deployed model, enabling seamless

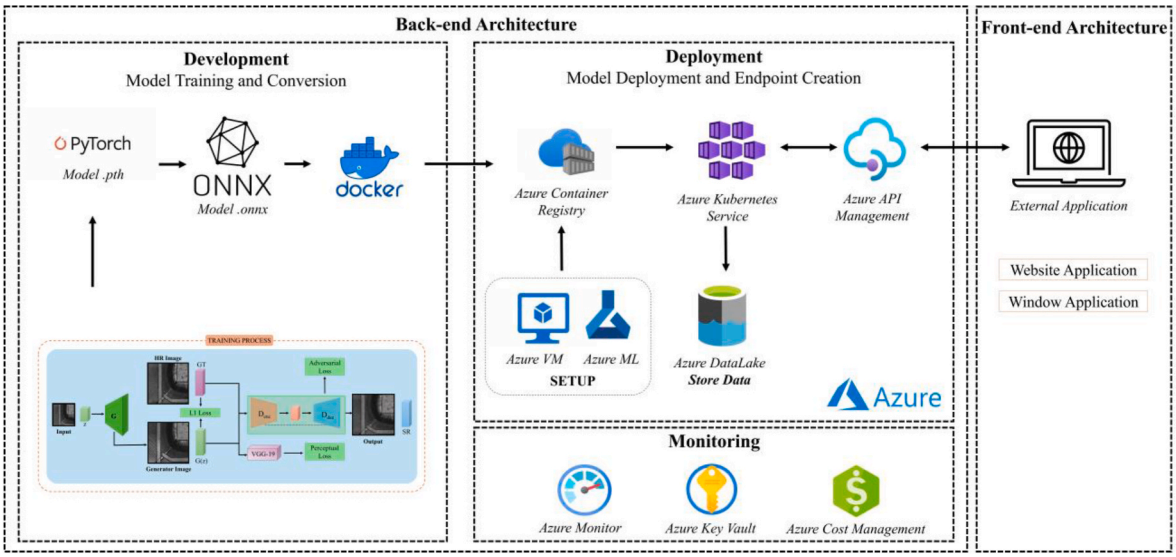


Fig. 5. Model deployment framework.

integration without modifying the core scanning application. In addition to cloud-based deployment, ONNX also enables local deployment within the C++-based environment of the scanning software. By initializing an inference session using ONNX Runtime in C++, the application can run the converted SAMGAN model directly on CPU or GPU hardware. The inference results are returned as OrtValue, the default output type in ONNX Runtime for C++. This setup allows for real-time processing even in offline or resource-constrained environments, providing a practical alternative to cloud deployment.

3.6. Evaluation metrics

We employed two widely used metrics, the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM), to assess the performance of our SAMGAN model in the context of industrial scanning acoustic SR. PSNR quantifies the reconstruction quality by measuring the ratio between the maximum possible pixel value and the mean squared error between the SR and GT images. A higher PSNR indicates a closer approximation to the reference image. In contrast, SSIM evaluates the perceptual quality by comparing the luminance, contrast, and structural information between images. This aligns better with human visual perception by considering the local patterns of pixel intensities. Using both the PSNR and SSIM, we ensured that our evaluation captured both the numerical accuracy and perceptual similarity of the SR images, providing a comprehensive assessment of the effectiveness of the model in enhancing acoustic images. The equations for PSNR and SSIM are shown in Equation (5) (Gonzales and Wintz, 1987) and **6** (Wang et al., 2004), respectively.

$$PSNR(x,y) = 10\,log_{10}\left(\frac{R^2}{MSE(x,y)}\right) \tag{5}$$

where R is the maximum possible pixel value of the image. For an 8-bit image, R is 255. MSE represents the mean squared error between the GT image x and SR image y.

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{6}$$

where μ_x and μ_y are the mean pixel values of images x and y, respectively. σ_x^2 and σ_y^2 are the variances of images x and y, respectively. σ_{xy} is the covariance between images x and y. C_1 and C_2 are constants to stabilize the division with weak denominators.

To provide a more comprehensive evaluation of the SAMGAN model's performance, particularly in industrial applications, we have incorporated computational efficiency metrics, including Floating Point Operations (FLOPs) (to measure computational complexity), Total Parameters (to assess model complexity), and Inference Time (to determine processing time per image). Together with PSNR and SSIM, these metrics offer a balanced perspective on both restoration quality and computational demands, which is essential for assessing the model's practicality in real-world deployment scenarios.

4. Results and discussions

4.1. Synthetic image generation

The main purpose of this degradation process was to mimic the characteristics of LR images. Another objective is to increase the diversity of the limited training dataset. The generative model is data-hungry, and its accuracy improves as more diverse and independent data are fed into the model. As illustrated in Fig. 6, the synthetic LR images were generated by applying the second high-order degradation method. Images at each degradation step were created using the corresponding techniques.

To overcome the challenge of limited HR paired data in industrial SAM imaging, we leveraged a synthetic data generation strategy as a critical augmentation method. By simulating realistic degradations - including blur, noise, and compression artifacts - we enriched the diversity of the training set and introduced controlled variations that mimic actual imaging conditions. This approach proved highly effective, as evidenced by the consistent improvement in both PSNR and SSIM metrics when synthetic data was combined with industrial data during training. According to the results from Table A4 and Fig. A2, the model trained with this hybrid dataset achieved an average PSNR increase of 5.6831 dB and an SSIM improvement of 0.1795 compared to using industrial data alone. These gains demonstrate not only enhanced

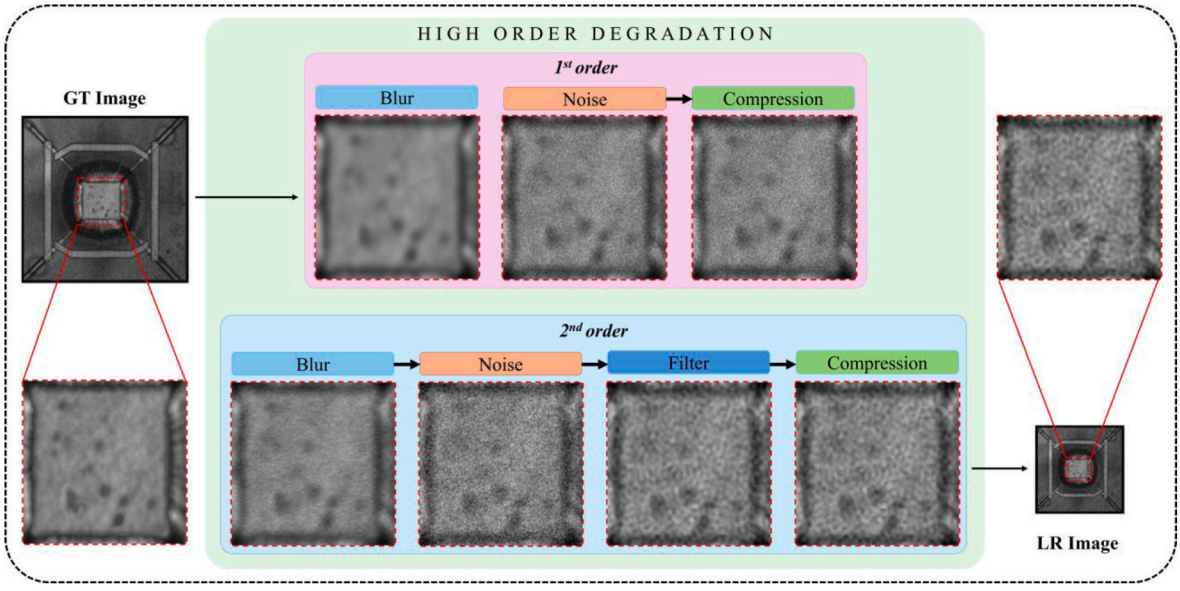


Fig. 6. Result of the synthetic LR image generator.

reconstruction quality but also stronger generalization capabilities. Along with the convergence behavior illustrated in Fig. A3, we also observe that the inclusion of synthetic data enables the model to converge faster and more stably, indicating improved learning efficiency in data-constrained scenarios. Furthermore, loss curve analysis revealed that synthetic data contributes to greater training stability, especially in the later phases, with reduced perceptual loss oscillation and lower final values. Collectively, these results affirm the meaningful role of synthetic data in mitigating the limitations of small industrial datasets and improving GAN-based reconstruction under real-world constraints.

4.2. Training results

The training process for our SAMGAN model took 9 days, completing 400,000 iterations. We used the Adam optimizer (learning rate = 0.0001, $\text{beta}_1 = 0.90$, $\text{beta}_2 = 0.99$, weight decay = 0) with a batch size of 8 to efficiently update the model's parameters. The loss curves, shown in Fig. 7, provide insights into the model's learning progression.

The L_1 loss (Fig. 7a) rapidly decreased during the initial iterations, indicating fast learning in the early stages. It then stabilized within a narrow range of 0.00015–0.0003, suggesting that the model reached an optimal state with no signs of significant overfitting. The final low L_1 loss (~0.0002) demonstrated the effectiveness of the model in performing SR tasks.

As shown in Fig. 7b, the L_{Percep} gradually decreased, indicating the model's ability to reconstruct the HR features in the SR images. Despite fluctuations owing to the complexity of the features extracted from VGG-19 and the diverse training dataset, the amplitude of these fluctuations decreased over time, reflecting the stability of the model in generating complex features. The perceptual loss remained relatively low, ranging from 0.02 to 0.08, with a typical value of approximately 0.05, showing an effective reduction in feature space differences.

Fig. 7c illustrates the fluctuation in the L_{Adver} loss throughout the entire training process. This is typical for GAN models because of the dynamic competition between the generator and discriminator (Saqur and Vivona, 2020). This fluctuation suggests that the model approaches the Nash equilibrium (Zhang, 2022), where the generator produces images that are nearly indistinguishable from the real images. This also indicated that the model did not overfit. Despite the lack of a clear convergence, the fluctuation amplitude was small over the training period, indicating a balanced and stable training process.

In summary, the composite loss function, which combines the L_1, L_{Percep}, and L_{Adver} losses, effectively balances the pixel-level accuracy, high-level feature reproduction, and image realism, resulting in a stable and well-performing model.

The training process for the proposed SAMGAN model is shown in Fig. 8. This figure shows the enhancement of the model's image quality as the number of epochs increases. During the initial training iterations, the model can quickly reconstruct the overall structure of the image, but tends to prioritize main features over pixel-level accuracy. Pixel-wise L_1 loss helps the model converge quickly; however, the overall background is distorted. By incorporating the total loss of the SAMGAN, the model corrects most pixel distortions in the image. However, some pixel artifacts remained in the brightest and darkest areas. In this phase, the model refines these regions to improve the accuracy. Perceptual loss captures perceptual similarities, whereas adversarial loss contributes to the generation of more realistic and sharper textures. During this period, background contrast and sharpness were fine-tuned to achieve the desired image quality.

As summarized in Table A5 and illustrated in Fig. A4, the model fine-

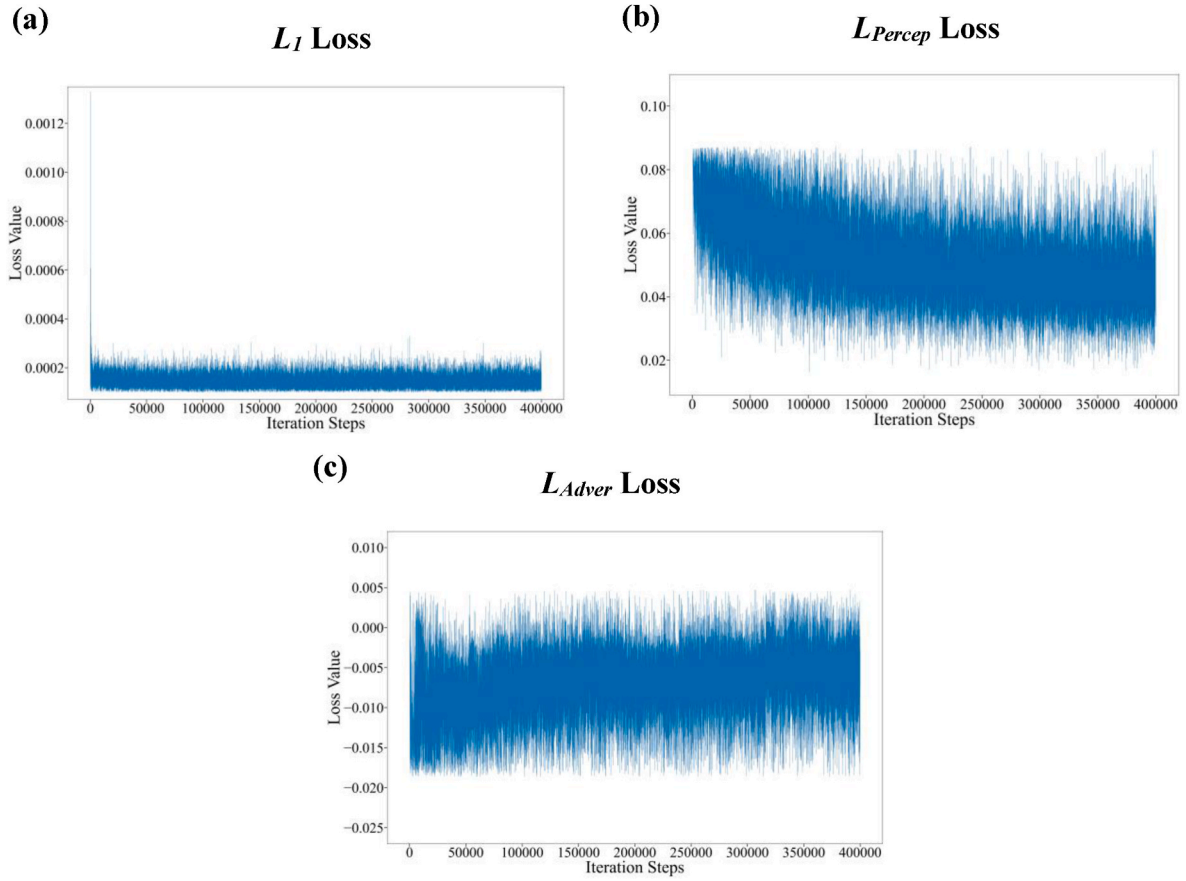


Fig. 7. Loss function curves of SAMGAN. (a) L_1 loss, which preserves fine image details. (b) L_{Percep} loss, which emphasizes high-level perceptual features. (c) L_{Adver} loss, which improves the realism of the generated images.

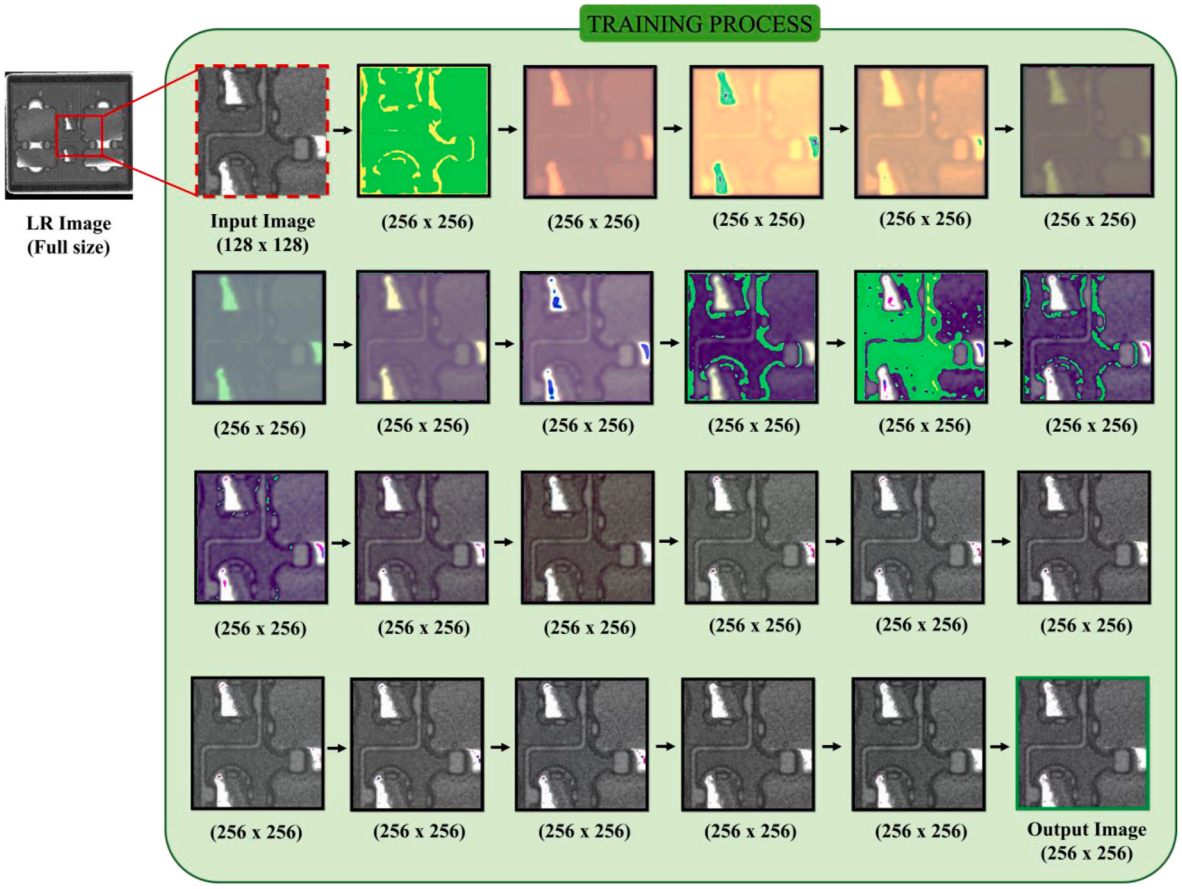


Fig. 8. The training process through iterations helps restore the SR images from the LR images. The results show that the restored image has a higher resolution, with sharper details compared to the input LR image.

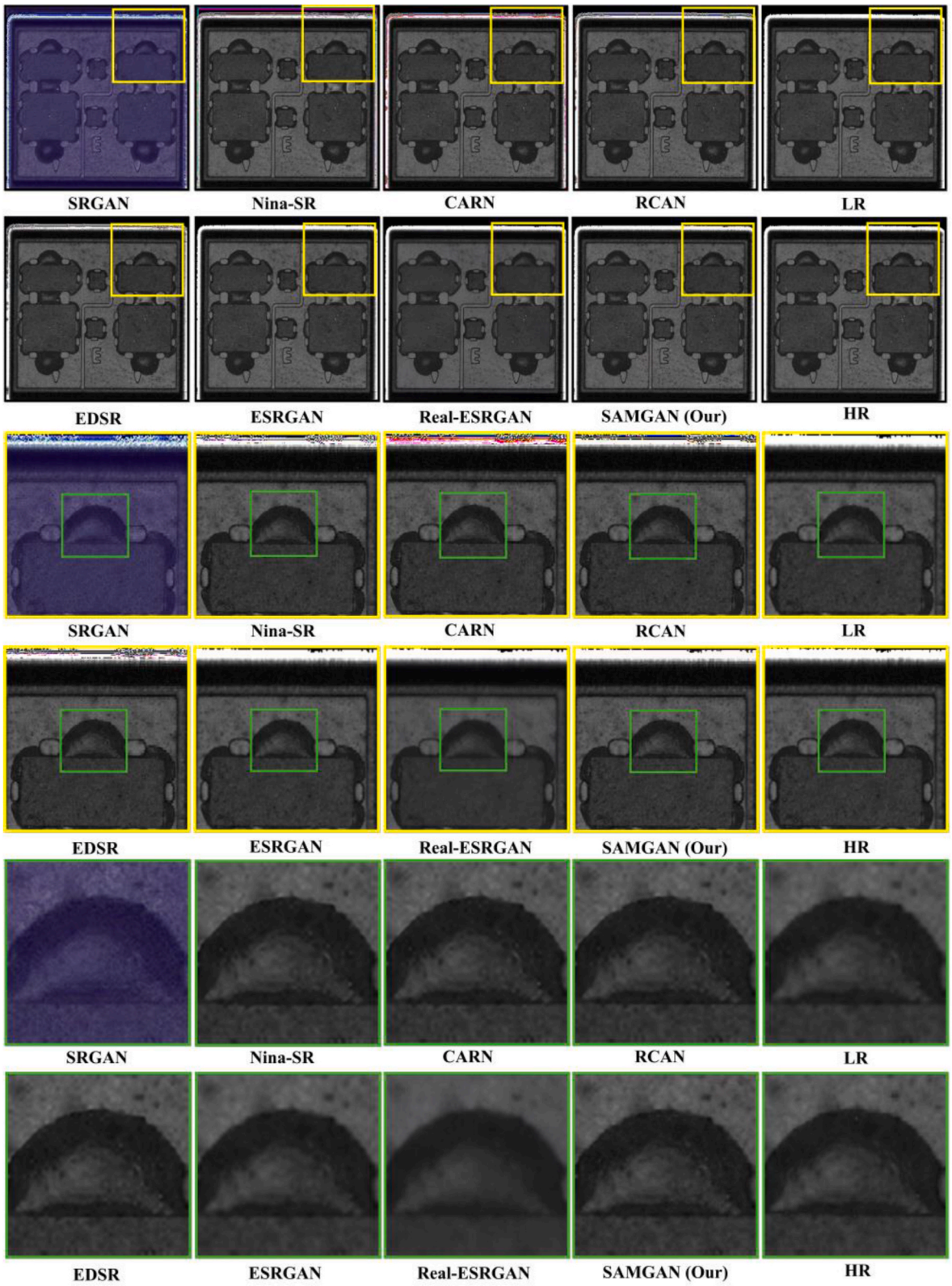


Fig. 9. Qualitative comparisons between different models on the testing dataset. The comparison includes SR images generated by our SAMGAN model and other SOTA models, alongside the corresponding LR and HR reference images. Selected regions are zoomed in to better highlight the visual differences among the models.

tuned with transfer learning consistently achieved the highest PSNR and SSIM scores over 100,000 iterations, outperforming both the DIV2K-only and self-built dataset models. Specifically, transfer learning led to a performance gain of 4.2999 dB in PSNR and 0.1333 in SSIM compared to the DIV2K-only model, and an improvement of 1.2017 dB/0.0363 in PSNR and SSIM, respectively, over training on the custom dataset alone. Additionally, qualitative comparisons provided in Fig. A5 demonstrate that the fine-tuned model more effectively preserves defect boundaries and mitigates over-smoothing—an issue observed in the DIV2K-only and self-trained variants. These visual and quantitative findings validate the

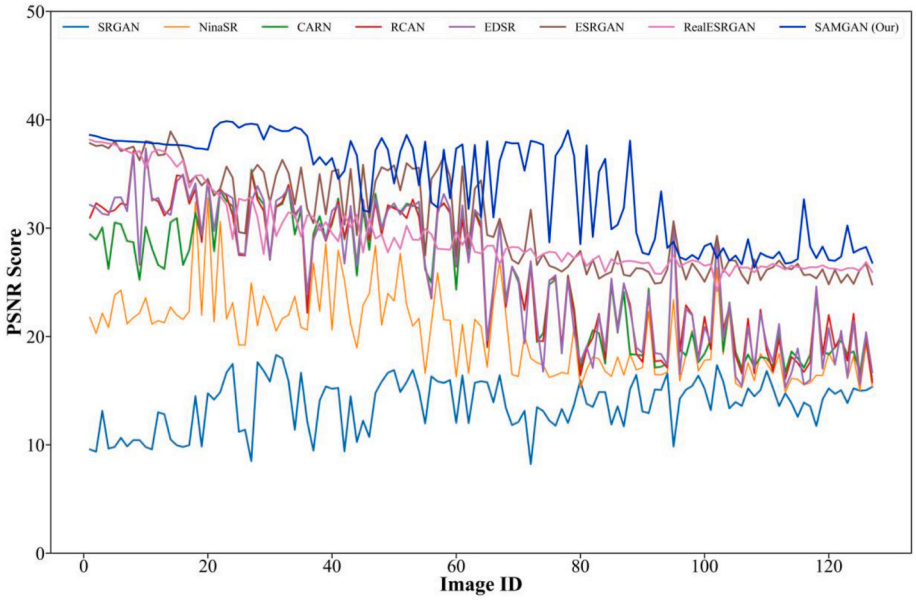


Fig. 10. PSNR scores of the testing dataset across various models.

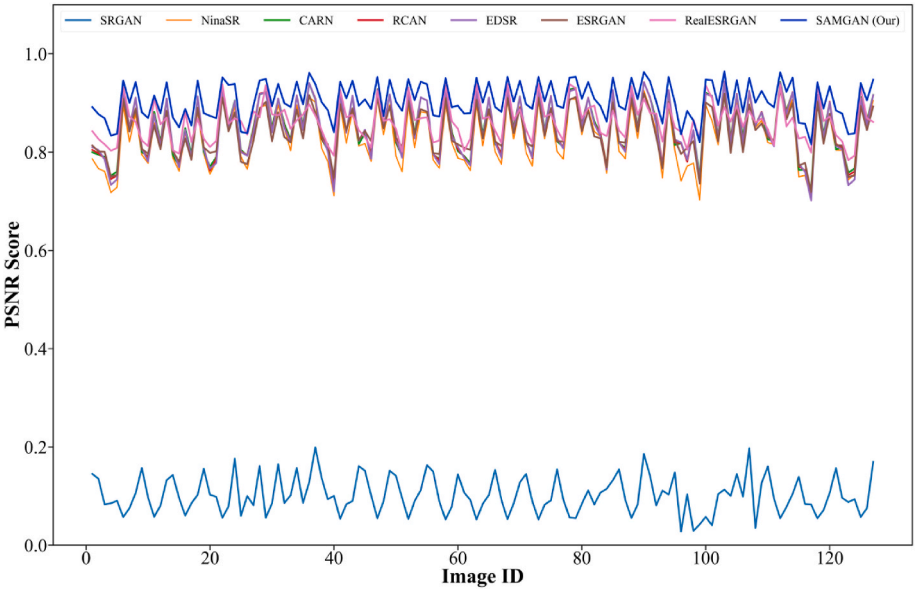


Fig. 11. SSIM scores of the testing dataset across various models.

Table 2
Comparison of PSNR metric between different models sorted from high to low.

Image no.	SRGAN	Nina-SR	CARN	RCAN	EDSR	ESRGAN	Real-ESRGAN	SAMGAN (Ours)
1	16.661	21.609	32.503	32.984	33.131	32.161	35.677	**39.867**
2	17.467	23.114	32.015	31.193	32.757	28.976	34.659	**39.789**
3	14.864	30.623	33.412	33.558	33.356	32.961	33.658	**39.731**
4	8.492	24.927	35.385	35.153	32.378	32.866	35.038	**39.632**
5	11.396	19.201	27.643	27.472	27.558	32.532	29.511	**39.555**
…	…	…	…	…	…	…	…	…
123	13.858	16.123	18.622	18.114	16.65	26.533	26.197	**26.813**
124	15.345	15.386	16.006	15.769	16.693	25.94	24.801	**26.804**
125	14.76	14.857	16.663	16.068	15.272	26.165	26.325	**26.725**
126	13.6	15.258	17.073	16.578	15.418	26.346	25.72	**26.645**
127	14.478	15.925	17.195	16.602	17.115	25.911	27.107	**26.349**
Avg	13.717	20.284	25.23	25.976	25.934	29.414	30.63	**33.978**

Table 3
Comparison of SSIM metric between different models sorted from high to low.

Image no.	SRGAN	Nina-SR	CARN	RCAN	EDSR	ESRGAN	Real-ESRGAN	SAMGAN (Ours)
1	0.055	0.926	0.943	0.941	0.941	0.936	0.929	**0.962**
2	0.052	0.899	0.926	0.924	0.923	0.939	0.910	**0.950**
3	0.053	0.906	0.929	0.927	0.926	0.938	0.910	**0.953**
4	0.056	0.916	0.927	0.924	0.924	0.938	0.911	**0.952**
5	0.055	0.903	0.927	0.925	0.924	0.934	0.910	**0.951**
…	…	…	…	…	…	…	…	**…**
123	0.085	0.718	0.752	0.745	0.733	0.803	0.750	**0.834**
124	0.094	0.756	0.767	0.760	0.744	0.793	0.753	**0.838**
125	0.083	0.704	0.723	0.713	0.701	0.799	0.719	**0.816**
126	0.100	0.711	0.738	0.732	0.720	0.792	0.748	**0.841**
127	0.088	0.743	0.758	0.753	0.732	0.784	0.749	**0.836**
Avg	0.102	0.832	0.852	0.851	0.849	0.862	0.842	**0.906**

Table 4
Quantitative comparison of SAMGAN and state-of-the-art super-resolution models in terms of total parameters, computational complexity (FLOPs), inference time, and average image quality metrics (PSNR and SSIM, reported as mean ± standard deviation). All differences from SAMGAN are statistically significant ($p < 0.001$, paired *t*-test).

Model	Total Parameter ($M = 10^6$)	FLOPs[a] ($G = 10^9$)	Inference Time[b] (Mili-Seconds)	Model Performance (Average)	
				PSNR (dB)	SSIM
SRGAN	1.40M	95.14G	31.70	13.72 ± 2.37	0.10 ± 0.04
NinaSR	0.10M	6.33G	6.50	20.28 ± 3.75	0.83 ± 0.06
CARN	1.59M	63.78G	4.80	25.23 ± 5.71	0.85 ± 0.06
RCAN	15.44M	1008.27G	93.80	25.98 ± 6.05	0.85 ± 0.06
EDSR	40.73M	2688.87G	9.40	25.93 ± 6.27	0.85 ± 0.06
ESRGAN	16.00M	93.67G	10.78	29.41 ± 3.62	0.86 ± 0.04
Real-ESRGAN	40.26M	294.26G	74.50	30.63 ± 4.59	0.84 ± 0.05
SAMGAN (Our model)	**24.39M**	**69.03G**	**7.03**	**33.98 ± 4.70**	**0.91 ± 0.04**

[a] **FLOPs:** Calculated on an NVIDIA GeForce RTX 3090 Ti with an input size of (256 × 256 × 3).
[b] **Inference Time:** Measured as the average time to process 127 test images using the models on an RTX 3090 Ti GPU (excluding model loading time and cloud service latency; only image generation time is considered).

effectiveness of our transfer learning strategy in addressing domain-induced performance variability and confirm its crucial role in adapting to the complex degradation characteristics of industrial SAM images.

4.3. Comparison with state-of-the-art (SOTA) methods

To demonstrate the effectiveness of our proposed model, we compared the generated super-resolution images with those produced by several state-of-the-art (SOTA) methods, including Nina-SR, CARN, RCAN, EDSR, SRGAN, ESRGAN, and Real-ESRGAN. For an objective and fair comparison, all models were retrained or fine-tuned on our dataset under the same conditions in terms of dataset, number of epochs, and training strategy. Hyperparameters were set according to their original publications to ensure the use of the best-known configurations. The configuration details are shown in Table A6 in the Appendix section.

Our SAMGAN model inherits the strengths of ESRGAN in preserving the main features from the GT while leveraging real-ESRGAN's ability to enhance image sharpness. The primary contribution of the proposed SAMGAN is to increase the sharpness of the generated images while retaining the main features, including minimal defects, which are crucial for industrial ultrasonic inspection. As shown in Fig. 9, the SR image generated by the SRGAN model was heavily distorted, indicating poor model performance. The Nina-SR, CARN, RCAN, and EDSR models struggle to process the white-colored region at the top of the sample image, which is likely owing to insufficient feature information in this area. This occurred because of the ineffectiveness of the skip connection mechanism within the residual blocks of these models. In contrast, the ESRGAN, Real-ESRGAN, and our SAMGAN models outperform these approaches in retaining image features. This advancement is attributed to the combined use of L_1, L_{Percep}, and L_{Adver} losses, as opposed to relying solely on L_1 loss, similar to as in conventional techniques. However, Real-ESRGAN tends to smooth the images excessively, causing the loss of fine details, which can be critical. Although ESRGAN preserves most image features, it does not significantly improve the image resolution compared to the original LR and HR images. Our SAMGAN model was fine-tuned to enhance contrast and sharpness while maintaining all key image features, as demonstrated in the zoomed-in section of Fig. 9. Additional comparisons shown in Figs. A6, A7, A8, A9, and A10 further support our findings.

Figs. 10 and 11 present the evaluation of the studied models on the test dataset in terms of the PSNR and SSIM metrics. Tables 2 and 3 present the quantified PSNR and SSIM scores, showing the top five highest, lowest, and average scores across all 127 images for the different models. Our SAMGAN model achieved the highest average PSNR score of 33.978, followed by ESRGAN with a score of 30.630. In terms of SSIM, our model also obtained the best score of 0.906, compared to 0.862 for the second-highest, Real-ESRGAN. As shown in Table A7, SAMGAN achieves approximately 10.93 % higher PSNR and 5.1 % higher SSIM compared to the most competitive baseline models - ESRGAN and Real-ESRGAN, respectively. This quantification aligns with the strengths of ESRGAN in preserving key image features and Real-ESRGAN in enhancing visual quality, as previously discussed. Our SAMGAN model inherits the advantages of both and demonstrates superior performance in generating industrial scanning acoustic SR images compared with existing GAN models based on the PSNR and SSIM metrics.

In industrial settings, computational indicators such as parameter count, FLOPs, and inference time are crucial for evaluating the deploy ability and efficiency of AI models. As shown in Table 4, the SAMGAN model, with 24.39 million parameters and 69.03 GFLOPs, demonstrates a lightweight computational footprint, significantly reducing hardware demands and operational costs when implemented on inspection systems. Despite not being the absolute fastest, SAMGAN achieves an impressive inference time of 7.03 ms, outperforming heavy models like

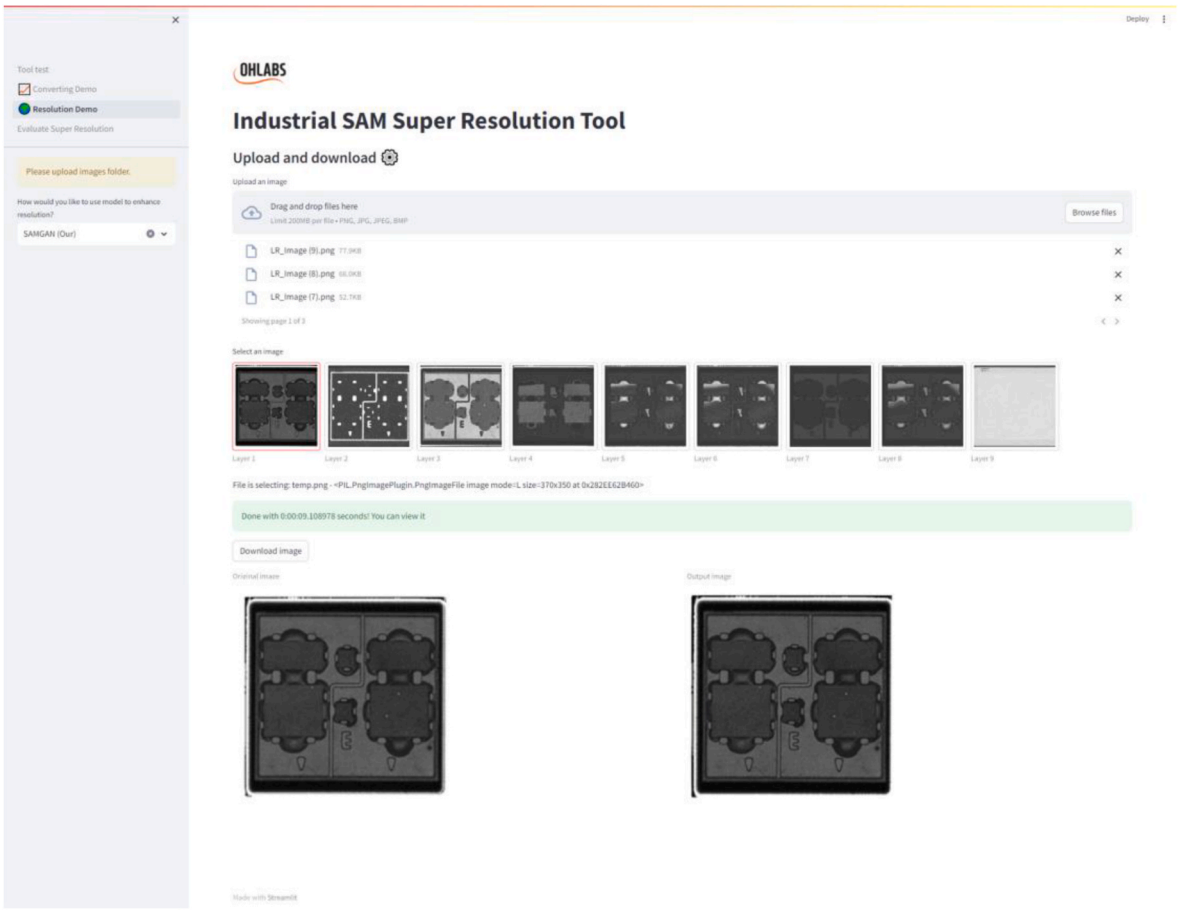


Fig. 12. SAMGAN deployment on the web application. Left: LR image obtained from scanning with a step size of 100 μm. Right: SR image generated by our model.

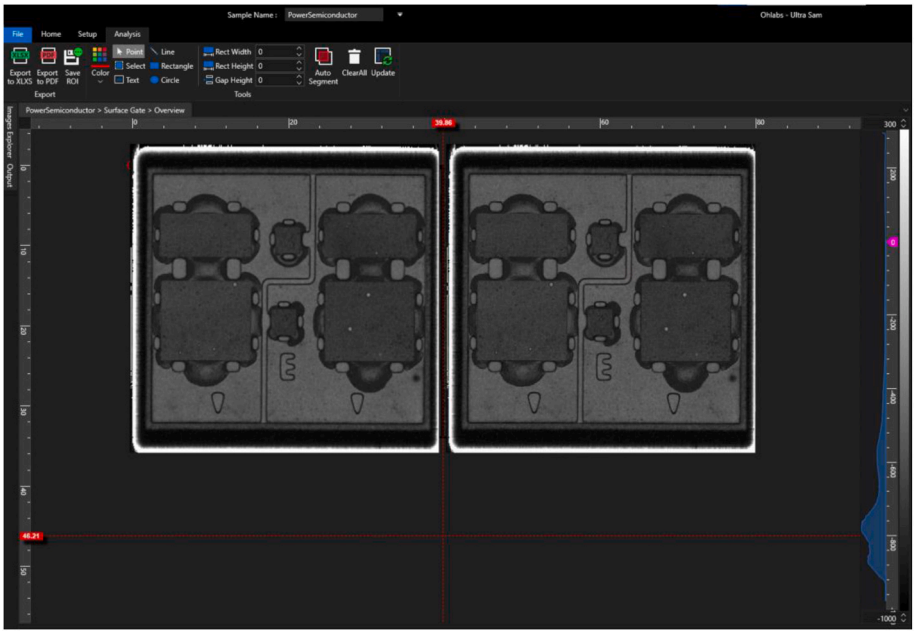


Fig. 13. Deployment of SAMGAN on the TSAM-400 software. Left: LR image obtained from scanning with a step size of 100 μm. Right: SR image generated by our model.

RCAN (93.80 ms), EDSR (9.40 ms), ESRGAN (10.78 ms), and Real-ESRGAN (74.50 ms) in both speed and accuracy, making it well-suited for real-time processing in industrial NDT pipelines.

Importantly, this computational efficiency does not compromise output quality. SAMGAN achieves the highest PSNR (33.98 $\pm$ 4.70 dB) and SSIM (0.91 $\pm$ 0.04) across all models, ensuring high-fidelity image

Table 5
Scanning time with and without using our SAMGAN model.

Samples no.	Name	Scanning Time (seconds)			
		LR	SR	HR	LR + SR
1	Power MOSFET	145	10	290	**155**
2	FPGA Chip	55	6	110	**61**
3	Transistor	60	7	120	**67**
4	Power Semiconductor	300	17	600	**317**

reconstruction essential for quality assessment. Statistical analysis over 127 test images confirmed that these improvements are highly significant ($p < 0.001$, paired *t*-test), demonstrating the consistency and reliability of SAMGAN's performance. By balancing speed, efficiency, and image reconstruction quality, SAMGAN provides a practical and scalable solution for smart manufacturing systems, enabling high-throughput, low-latency inspection and supporting real-time quality control in industrial NDT applications.

4.4. Model deployment results

To validate the feasibility and reliability of our study, our SAMGAN model was deployed in two environments: web-based and software applications. Users can use this model as a standalone web-based tool that specializes in image SR processing (Fig. 12). Additionally, the model can be integrated into user-scanning software as a compact SR framework (Fig. 13). As listed in Table 5, the scanning time is reduced by approximately 44.2 %–44.5 % for small samples, such as FPGA Chips or Transistors, when using our SR technique. The reduction is even greater for larger samples, such as Power MOSFETs and Power Semiconductors, with scanning times being approximately 46.6 %–47.2 % faster than those of conventional scanning strategies.

The deployment of our SAMGAN model on a scanning image of a MOSFET is illustrated in Fig. 14. We highlight three regions of interest marked in red, green, and yellow boxes. Each region contains different types of defects: the red box contains voids; the green box includes both a crack and voids; and the yellow box targets a solder-related defect.

5. Conclusions

In this study, we proposed the SAMGAN, a generative adversarial network framework tailored for enhancing Scanning Acoustic Microscopy (SAM) images in industrial NDT. To address the scarcity of paired training data, we introduced a high-order degradation model that generates synthetic LR images from HR data. The SAMGAN model integrates transfer learning, a UNet-based discriminator, and a dynamic loss function combining pixel-wise, perceptual, and adversarial terms to effectively learn from limited and synthetic data. Quantitative results demonstrate the effectiveness of SAMGAN, achieving an average PSNR of 33.978 dB and SSIM of 0.906. On average, SAMGAN outperforms the most competitive baseline models (ESRGAN or Real-ESRGAN) by approximately 10.93 % in PSNR and 5.1 % in SSIM. These improvements translate into visually sharper images that retain critical structural details, enhancing clarity and user interpretability.

Beyond image quality, SAMGAN supports up to nearly 50 % reduction in scanning time without compromising diagnostic fidelity, addressing a key industrial requirement. The deployment results demonstrate that SAMGAN significantly reduces scanning time, with a 44.2 %–44.5 % reduction for smaller samples (FPGA Chips, Transistors) and an even greater 46.6 %–47.2 % reduction for larger samples (Power MOSFETs, Power Semiconductors). These reductions contribute directly to improved operational efficiency and cost savings in industrial inspection processes. In summary, our work provides value not only in academic research but also in practical industrial SAM applications. Beyond that, the proposed framework is adaptable to other imaging modalities that face similar resolution–efficiency trade-offs, including biomedical ultrasound, thermal imaging, and industrial X-ray or terahertz systems.

6. Limitations and future works

While the results of this study are promising, there are several limitations that need to be discussed. First, a primary limitation of this study is the absence of a quantitative benchmark for assessing how closely the synthetic LR images replicate real-world degradations. Since the model's effectiveness in resolution enhancement and defect detection depends heavily on the similarity between synthetic and real SAM

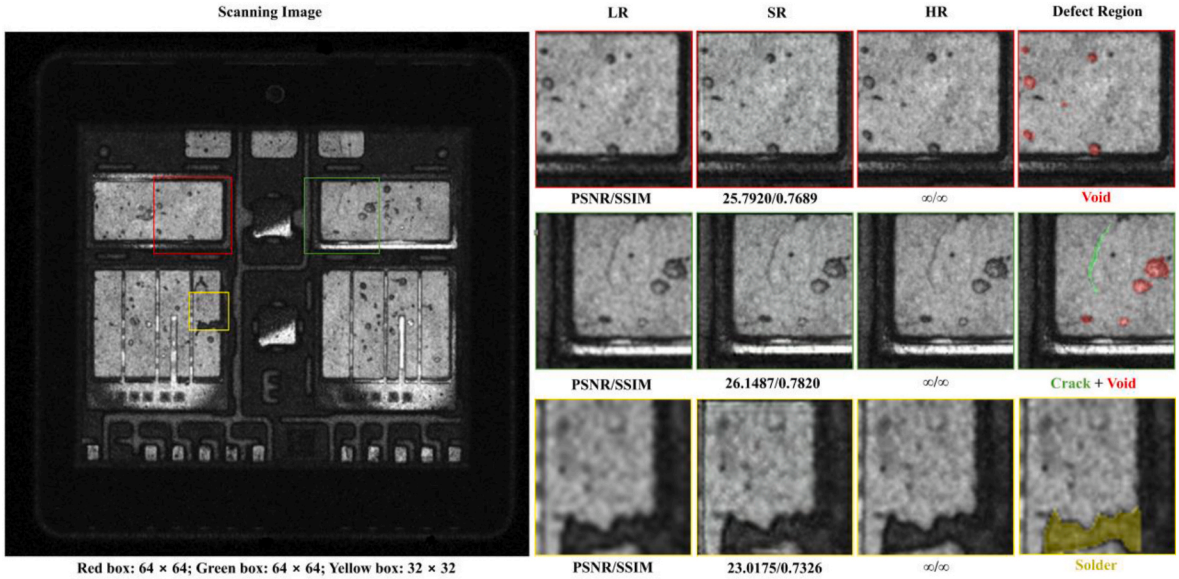


Fig. 14. The SR results on an industrial SAM image, including defect regions. Right-hand panels show zoomed-in patches of these regions. The zoomed-in regions consist of voids, cracks, and solder defects in the sample.

images, this remains a critical gap. Currently, the degradation model is evaluated only indirectly through the performance of the final SR outputs. A more direct assessment of synthetic data realism would provide stronger validation of the proposed approach. Second, although our dataset encompasses a diverse range of industrial samples and common defect types (voids, cracks, solder, delamination), the variability of real-world defects extends beyond the current scope of our data. Further works focus on expanding the dataset to include a wider range of materials and defect scenarios to improve the robustness and applicability of SAMGAN across diverse inspection environments. Lastly, the model is currently evaluated using standard image quality metrics such as PSNR and SSIM. While these metrics are informative, they do not directly capture task-specific performance in NDT.

In addition, although the current model architecture has demonstrated the ability to effectively restore and enhance resolution for various samples such as Power MOSFETs or ICs, we acknowledge that for samples with highly intricate line structures or complex textures, artifacts may still be introduced during the SR process. This is a recognized limitation of GAN-based models, which may hallucinate non-existent features under challenging conditions. Moreover, due to the inherent complexity of SAM images, a domain gap may still exist between synthetic and real degraded data. Therefore, future work will aim to expand the diversity of training samples and apply style transfer techniques to reduce the disparity between the synthetic and real datasets. Quantitative benchmarks will be used to evaluate the similarity between synthetic and real datasets. Adaptive degradation models or semi-supervised learning could be explored to better handle diverse industrial ultrasound data. Investigating SAMGAN's robustness to noisy or corrupted LR inputs is also essential, given the variability in real-world imaging conditions. Future studies should incorporate domain-specific evaluation criteria, such as defect localization accuracy, to better assess the model's impact on inspection reliability and diagnostic accuracy. Next, an ablation study on the dynamic loss components - specifically the roles of L_1, perceptual, and adversarial loss will be conducted to provide valuable insight into their individual and combined contributions to SR performance. Additionally, exploring k-fold cross-validation could be considered to further enhance model stability and ensure more reliable generalization across diverse sample types. Finally, we will focus on optimizing the SAMGAN architecture for real-time deployment, making it more suitable for high-throughput industrial environments. These directions collectively aim to improve the model's utility, robustness, and scalability across a wide range of practical NDT applications.

CRediT authorship contribution statement

Trong Nhan Nguyen: Software, Methodology. **Vu Hoang Minh Doan:** Writing – original draft, Methodology, Conceptualization. **Tan Hung Vo:** Visualization, Conceptualization. **Dinh Dat Vu:** Methodology. **Le Hai Tran:** Methodology. **Thi Thu Ha Vu:** Methodology. **Truong Tien Vo:** Writing – review & editing. **Quoc Dung Nguyen:** Methodology. **Sudip Mondal:** Writing – review & editing. **Jaeyeop Choi:** Resources, Project administration. **Byeongil Lee:** Resources, Project administration. **Junghwan Oh:** Resources, Project administration, Funding acquisition.

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.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A5A8023404).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.engappai.2025.112500.

Appendix

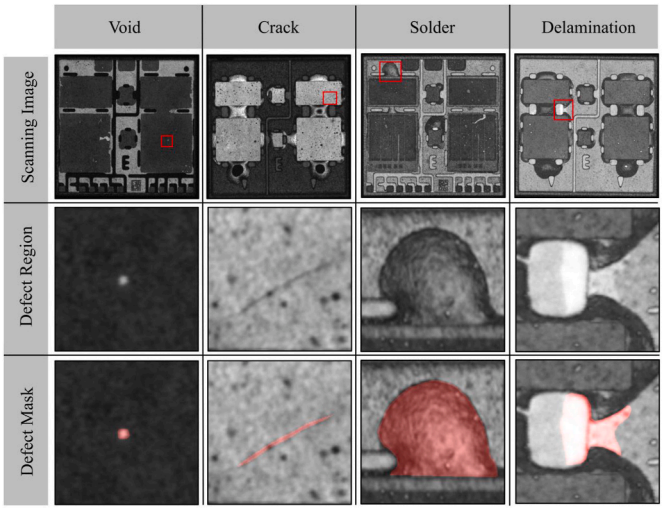


Fig. A1. Scanning images with 4 types of defects: void, crack, solder, and delamination.

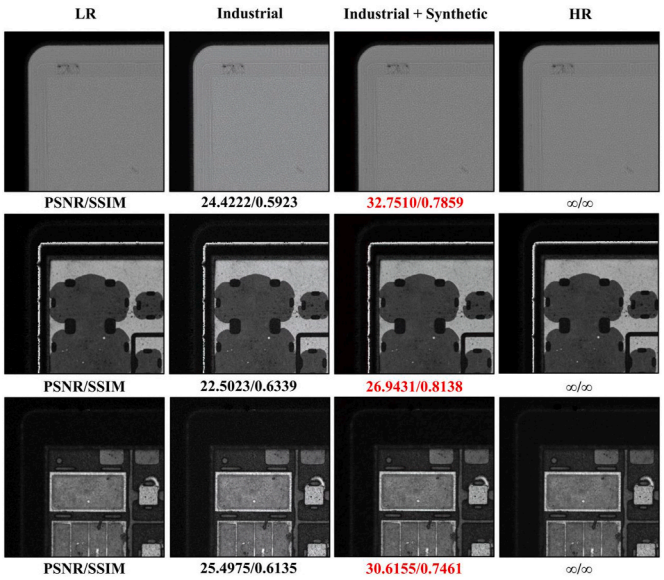


Fig. A2. Visual comparison of reconstruction results of different training datasets: Industrial only and Industrial + Synthetic with PSNR/SSIM metrics.

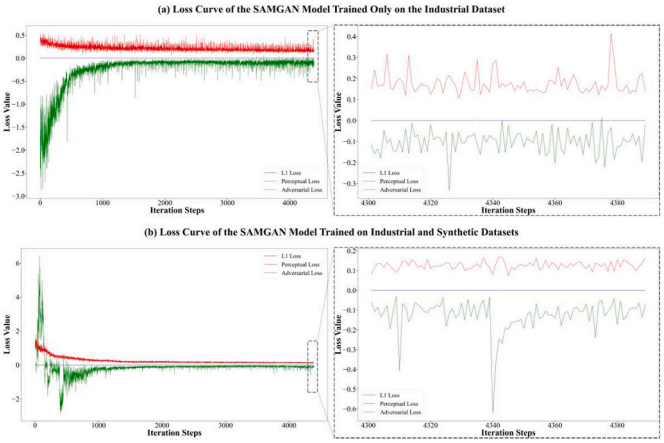


Fig. A3. Loss curve comparison between training on industrial-only and industrial + synthetic datasets.

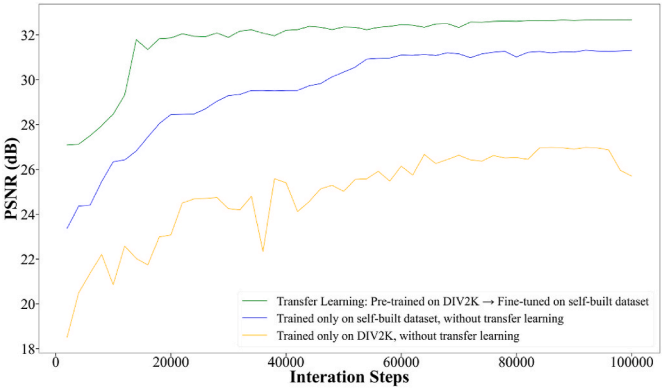


Fig. A4. Comparison of PSNR performance on the validation dataset among three training strategies: training only on DIV2K without transfer learning (orange), direct training on the self-built dataset without transfer learning (blue), and transfer learning by pre-training on DIV2K followed by fine-tuning on the self-built dataset (green).

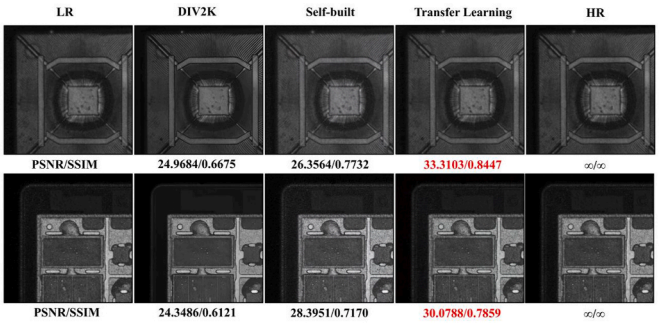


Fig. A5. Qualitative and quantitative comparison of SAMGAN SR performance using different training strategies. Visual and PSNR/SSIM comparison on two representative industrial samples (from the Validation dataset) restored using models trained on DIV2K, the self-built dataset, and fine-tuned with transfer learning.

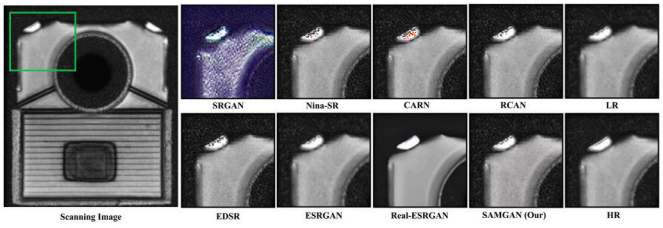


Fig. A6. Qualitative comparison of corresponding SR images generated by SAMGAN and other models on a randomly selected scanning image from the test dataset.

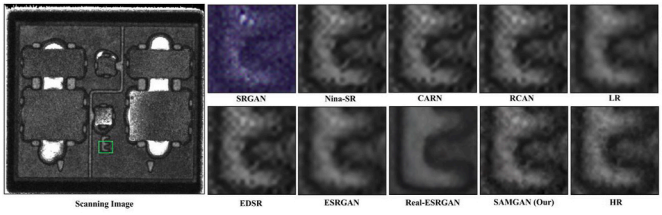


Fig. A7. Qualitative comparison of corresponding SR images generated by SAMGAN and other models on a randomly selected scanning image from the test dataset.

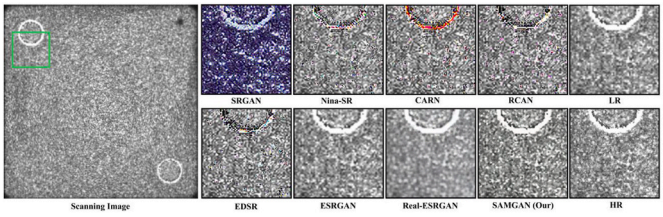


Fig. A8. Qualitative comparison of corresponding SR images generated by SAMGAN and other models on a randomly selected scanning image from the test dataset.

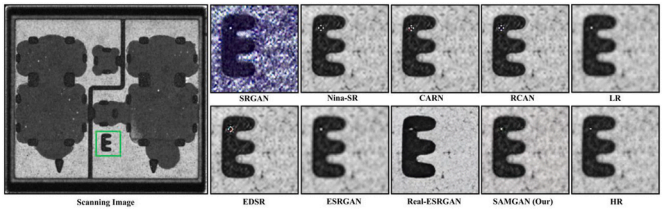


Fig. A9. Qualitative comparison of corresponding SR images generated by SAMGAN and other models on a randomly selected scanning image from the test dataset.

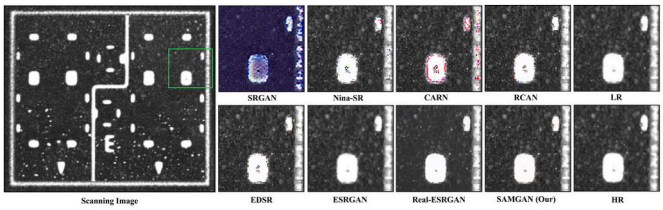


Fig. A10. Qualitative comparison of corresponding SR images generated by SAMGAN and other models on a randomly selected scanning image from the test dataset.

Table A1
Comparison of existing imaging and SR methods in biomedical and industrial NDT applications (The limitation of each approach is highlighted in bold).

Methods	Raster-scan	High-frequency Transducer-based	ULM	SR
Dataset	Real	**Simulated**	Real	Real
Applications	Biomedical and industrial	Biomedical and industrial	**Biomedical**	Biomedical and industrial
Speed	**Slow**	Fast	Fast	Fast
Penetration Depth	Deep	**Shallow**	Deep	Deep
Deployment Feasibility	**Not reported**	**Not reported**	**Not reported**	Yes
Reference	You et al. (2023)	(Moore et al., 2020; Morokov et al., 2021)	(Chen et al., 2022; van Sloun et al., 2020)	Our work

Table A2
SRGAN, ESRGAN, and Real-ESRGAN's backbone, key strengths, and limitations comparison

Model	Backbone Architecture	Key Strengths	Limitations
SRGAN	ResNet-style generator + Relativistic discriminator	First GAN-based SR; produces sharp textures	Prone to artifacts; structural accuracy suffers at edges
ESRGAN	RRDB (23 blocks) generator + Relativistic discriminator	Improved detail via residual-in-residual dense blocks	Sensitive to training variance; requires large datasets
Real-ESRGAN	ESRGAN backbone + realistic degradation	Robust to real-world image degradations	Tends to oversmooth fine defects; aliasing causes twisted lines in restorations; may lose small structural features

Table A3
Overview of the industrial component dataset used in this study

Category	Power Semiconductor	Power MOSFET	Transistor	FPGA Chip
Number of Samples	49	28	57	9
Number of Scanned Layers	9	5	3	4
Number of Samples for Training and Validating	39	23	54	7
Number of Samples for Testing	10	5	3	2
Number of Scanned Images	882	140	171	72

Table A4
Quantitative evaluation of model performance using industrial-only and industrial + synthetic training data.

Image no.	Only Industrial dataset		Industrial dataset + Synthetic dataset	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM
1	26.3564	0.7732	33.3103	0.8447
2	24.5387	0.5946	33.1519	0.7962
3	24.3556	0.6099	33.0625	0.7956
4	21.6501	0.7322	31.9569	0.8147
5	23.7925	0.6185	31.8861	0.7579
...	...	...	...	...
72	18.5339	0.1768	21.0375	0.2289
73	16.5182	0.1261	17.6771	0.1738
74	12.9103	0.0892	13.4308	0.1335
75	12.2529	0.2593	12.7443	0.3452
76	12.0036	0.2139	12.0583	0.2941
Avg	**22.7114**	**0.4946**	**28.3945**	**0.6741**

Table A5
Performance enhancement of SAMGAN over SOTA models across PSNR and SSIM metrics (%).

Performance Enhancement	SRGAN	Nina-SR	CARN	RCAN	EDSR	ESRGAN	Real-ESRGAN
PSNR	147.707	67.511	34.673	30.805	31.017	15.516	10.930
SSIM	788.235	8.894	6.338	6.463	6.714	5.104	7.601

Table A6
Training configurations and implementation details of the compared SOTA models.

Model	SRGAN	Nina-SR	CARN	RCAN	EDSR	Real-ESRGAN	ESRGAN
Iterations (K = 1000)	400K	400K	400K	400K	400K	400K	400K
Batch size	16	16	16	16	16	12	8
Learning rate	1e-4	1e-4	1e-4	1e-4	1e-4	2e-4	2e-4
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam	Adam
Pretrained-weight?	No	Yes	Yes	Yes	Yes	Yes	No
Compute Platform	NVIDIA GeForce RTX 3090Ti GPU						Tesla T4

Table A7
Performance Evaluation of the SAMGAN Model Using Three Different Training Strategies Over 100,000 Iterations on the Test Dataset (50 Images) across PSNR and SSIM metrics (%).

Performance Metrics	DIV2K	Self-built Dataset	Transfer Learning
PSNR	22.4220	25.5202	**26.7219**
SSIM	0.4889	0.5859	**0.6222**

Data availability

The Python code implementation of SAMGAN and the model checkpoints used in this study are available on GitHub at https://github.com/nhannt69/SAMGAN. A portion of the scanning dataset will be made available upon request.

References

Abd Al Rahman, M., Mousavi, A., 2020. A review and analysis of automatic optical inspection and quality monitoring methods in electronics industry. IEEE Access 8, 183192–183271. https://doi.org/10.1109/ACCESS.2020.3029127.

Agustsson, E., Timofte, R., 2017. Ntire 2017 challenge on single image super-resolution: dataset and study. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 126–135. https://doi.org/10.1109/CVPRW.2017.150.

Ahn, N., Kang, B., Sohn, K.-A., 2018. Fast, accurate, and lightweight super-resolution with cascading residual network. Proceedings of the European Conference on Computer Vision (ECCV), pp. 252–268. https://doi.org/10.48550/arXiv.1803.08664.

Banerjee, P., Akarte, S.M., Kumar, P., Shamsuzzaman, M., Butola, A., Agarwal, K., Prasad, D.K., Melandsø, F., Habib, A., 2024. High-resolution imaging in acoustic microscopy using deep learning. Mach. Learn.: Sci. Technol. 5, 015007. https://doi.org/10.1088/2632-2153/ad1c30.

Chen, X., Lowerison, M.R., Dong, Z., Han, A., Song, P., 2022. Deep learning-based microbubble localization for ultrasound localization microscopy. IEEE Trans. Ultrason. Ferroelectrics Freq. Control 69, 1312–1325. https://doi.org/10.1109/TUFFC.2022.3152225.

Chopard, A., Cassar, Q., Bou-Sleiman, J., Guillet, J.-P., Pan, M., Perraud, J., Susset, A., Mounaix, P.J.N., International, E., 2021. Terahertz Waves for Contactless Control and Imaging in Aeronautics Industry, 122, 102473.

Christensen-Jeffries, K., Browning, R.J., Tang, M.-X., Dunsby, C., Eckersley, R.J., 2014. In Vivo Acoustic Super-resolution and Super-resolved Velocity Mapping Using Microbubbles, vol. 34, pp. 433–440.

Demiray, B.Z., Sit, M., Demir, I., 2021. D-SRGAN: DEM super-resolution with generative adversarial networks. SN Computer Sci. 2, 1–11. https://doi.org/10.48550/arXiv.2004.04788.

Deng, Y., Huang, D., Du, S., Li, G., Zhao, C., Lv, J., 2021. A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis. Comput. Ind. 127, 103399. https://doi.org/10.1016/j.compind.2021.103399.

Fang, T., An, J., Chen, Q., He, Y., Wang, H., Zhang, X., Evaluation, 2023. Progress and comparison in nondestructive detection, imaging and recognition technology for defects of wafers, chips and solder joints. Nondestr. Test. Eval. 1–56. https://doi.org/10.1080/10589759.2023.2274007.

Fevola, G., Ossig, C., Verezhak, M., Garrevoet, J., Guthrey, H.L., Seyrich, M., Brückner, D., Hagemann, J., Seiboth, F., Schropp, A.J.A.S., 2024. 3D and Multimodal X-Ray Microscopy Reveals the Impact of Voids in CIGS Solar Cells, vol. 11, 2301873.

Gao, F., Li, B., Chen, L., Wei, X., Shang, Z., Liu, C., 2024. Ultrasound image super-resolution reconstruction based on semi-supervised CycleGAN. Ultrasonics 137, 107177. https://doi.org/10.1016/j.ultras.2023.107177.

Gong, Z., Xu, Y., Liu, H., 2023. Research on image super-resolution restoration algorithm based on high order degradation model. 2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE). IEEE, pp. 586–590. https://doi.org/10.1109/NNICE58320.2023.10105790.

Gonzales, R.C., Wintz, P., 1987. Digital Image Processing. Addison-Wesley Longman Publishing Co., Inc.

Gouvine, G., 2021. NinaSR: efficient small and large ConvNets for super-resolution. https://doi.org/10.5281/zenodo.4868308.

Gu, Y., Zeng, Z., Chen, H., Wei, J., Zhang, Y., Chen, B., Li, Y., Qin, Y., Xie, Q., Jiang, Z., Applications, 2020. MedSRGAN: medical images super-resolution using generative adversarial networks. Multimed. Tool. Appl. 79, 21815–21840. https://doi.org/10.1007/s11042-020-08980-w.

Hingot, V., Chavignon, A., Heiles, B., Couture, O., 2021. Measuring image resolution in ultrasound localization microscopy. IEEE Trans. Med. Imag. (40), 3812–3819. https://doi.org/10.1109/TMI.2021.3097150.

Jiang, M., Zhi, M., Wei, L., Yang, X., Zhang, J., Li, Y., Wang, P., Huang, J., Yang, G., 2021. FA-GAN: fused attentive generative adversarial networks for MRI image super-resolution. Comput. Med. Imag. Graph. 92, 101969. https://doi.org/10.1016/j.compmedimag.2021.101969.

Kim, L.T., 2017. Principles of ultrasound and applied ultrasound physics relevant for advanced sonographers. In: Milas, M., Mandel, S.J., Langer, J.E. (Eds.), Advanced Thyroid and Parathyroid Ultrasound. Springer International Publishing, Cham, pp. 37–47. https://doi.org/10.1007/978-3-319-44100-9_5.

Kumar, A., Arnold, W., 2022. High resolution in non-destructive testing: a review. J. Appl. Phys. 132. https://doi.org/10.1063/5.0095328.

Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., Chen, Y., Zhou, X., 2020. Generative adversarial networks and its applications in biomedical informatics. Front. Public Health 8, 164. https://doi.org/10.3389/fpubh.2020.00164.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., 2017. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690. https://doi.org/10.48550/arXiv.1609.04802.

Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K., 2017. Enhanced deep residual networks for single image super-resolution. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 136–144. https://doi.org/10.48550/arXiv.1707.02921.

Liu, S., Liu, F., Yang, Y., Li, L., Li, Z., 2020. Nondestructive evaluation 4.0: ultrasonic intelligent nondestructive testing and evaluation for composites. Res. Nondestr. Eval. 31, 370–388. https://doi.org/10.1080/09349847.2020.1826613.

Liu, Y., Chen, J., Hou, J.-a., 2023. Learning position information from attention: end-to-end weakly supervised crack segmentation with GANs. Comput. Ind. 149, 103921. https://doi.org/10.1016/j.compind.2023.103921.

Lyu, Y., Xu, Y., Yang, C., Jiang, X., Zhao, X., Zhu, X., 2024. WSRGAN: a wavelet-based GAN for super-resolution of plane-wave ultrasound images without sampling loss. Eng. Appl. Artif. Intell. 127, 107384. https://doi.org/10.1016/j.engappai.2023.107384.

Martin, D.J., Wells, I.T., Goodwin, C.R.J.A., Medicine, I.C., 2015. Phys. Ultrasound 16, 132–135.

Moore, M.J., Bodera, F., Hernandez, C., Shirazi, N., Abenojar, E., Exner, A.A., Kolios, M. C., 2020. The dance of the nanobubbles: detecting acoustic backscatter from sub-micron bubbles using ultra-high frequency acoustic microscopy. Nanoscale 12, 21420–21428. https://doi.org/10.1039/D0NR05390B.

Morokov, E., Levin, V., Chernov, A., Shanygin, A., 2021. High resolution ply-by-ply ultrasound imaging of impact damage in thick CFRP laminates by high-frequency acoustic microscopy. Compos. Struct. 256, 113102. https://doi.org/10.1016/j.compstruct.2020.113102.

Nsengiyumva, W., Zhong, S., Lin, J., Zhang, Q., Zhong, J., Huang, Y., 2021. Advances, limitations and prospects of nondestructive testing and evaluation of thick composites and sandwich structures: a state-of-the-art review. Compos. Struct. 256, 112951. https://doi.org/10.1016/j.compstruct.2020.112951.

Pham, V.H., Vo, T.H., Vu, D.D., Choi, J., Park, S., Nguyen, D.T., Lee, B.-I., Oh, J., 2022. Development of scanning acoustic microscopy system for evaluating the resistance spot welding quality. Res. Nondestr. Eval. 33, 123–137. https://doi.org/10.1080/09349847.2022.2073415.

Ren, Z., Huang, K., Zhu, Y., Feng, K., Liu, Z., Fu, H., Hong, J., Glowacz, A., 2024. Progressive generative adversarial network for generating high-dimensional and wide-frequency signals in intelligent fault diagnosis. Eng. Appl. Artif. Intell. 133, 108332. https://doi.org/10.1016/j.engappai.2024.108332.

Saqur, R., Vivona, S., 2020. Capsgan: using dynamic routing for generative adversarial networks. In: Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC). Springer, pp. 511–525. https://doi.org/10.1007/978-3-030-17798-0_41, 21.

Silva, M.I., Malitckii, E., Santos, T.G., Vilaça, P., 2023. Review of conventional and advanced non-destructive testing techniques for detection and characterization of small-scale defects. Prog. Mater. Sci. 138, 101155. https://doi.org/10.1016/j.pmatsci.2023.101155.

Somani, A., Banerjee, P., Rastogi, M., Habib, A., Agarwal, K., Prasad, D.K., 2023. Image inpainting with hypergraphs for resolution improvement in scanning acoustic microscopy. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 3113–3122. https://doi.org/10.1109/CVPRW59228.2023.00313.

Tian, C., Zhang, X., Lin, J.C.-W., Zuo, W., Zhang, Y., Lin, C.-W., 2022. Generative adversarial networks for image super-resolution: a survey. arXiv preprint arXiv. https://doi.org/10.48550/arXiv.2204.13620.

van Sloun, R.J., Solomon, O., Bruce, M., Khaing, Z.Z., Wijkstra, H., Eldar, Y.C., Mischi, M., 2020. Super-resolution ultrasound localization microscopy through deep learning. IEEE Trans. Med. Imag. 40, 829–839. https://doi.org/10.1109/TMI.2020.3037790.

Viessmann, O., Eckersley, R., Christensen-Jeffries, K., Tang, M.-X., Dunsby, C.J.P.i.M., Biology, 2013. Acoustic Super-resolution with Ultrasound and Microbubbles, vol. 58, p. 6447.

Wang, B., Zhong, S., Lee, T.-L., Fancey, K.S., Mi, J., 2020. Non-destructive testing and evaluation of composite materials/structures: a state-of-the-art review. Adv. Mech. Eng. 12, 1687814020913761. https://doi.org/10.1177/1687814020913761.

Wang, X., Xie, L., Dong, C., Shan, Y., 2021. Real-esrgan: training real-world blind super-resolution with pure synthetic data. Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1905–1914. https://doi.org/10.1109/ICCVW54120.2021.00217.

Wang, X., Yu, K., Chan, K., Dong, C., Loy, C., 2018a. BasicSR: open source image and video restoration. https://github.com/XPixelGroup/BasicSR.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C., 2018b. Esrgan: enhanced super-resolution generative adversarial networks. Proceedings of the European Conference on Computer Vision (ECCV) Workshops. https://doi.org/10.1007/978-3-030-11021-5_5, 0-0.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image Quality Assessment: from Error Visibility to Structural Similarity, vol. 13, pp. 600–612.

Xiao, X., Gao, B., Tian, G.Y., Cai, Z.G., qing Wang, K., 2020. Physical perspective forward-inverse learning for ultrasonic sensing diagnosis in small diameter and thin-wall tube. Ultrasonics 105, 106115. https://doi.org/10.1016/j.ultras.2020.106115.

You, X., Liu, J., Li, Y., Jiang, Y., Liu, J.J.J.o.M., 2023. 3D Microscopy in Industrial Measurements, vol. 289, pp. 137–156.

Yu, H., 2020. Scanning acoustic microscopy for material evaluation. Appl. Microsc 50, 25. https://doi.org/10.1186/s42649-020-00045-4.

Yu, Z., Boseck, S.J., 1995. Scanning Acoustic Microscopy and its Applications to Material Characterization, vol. 67, p. 863.

Zhang, K., Zuo, W., Zhang, L., 2018a. Learning a single convolutional super-resolution network for multiple degradations. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3262–3271.

Zhang, S., 2022. On the Nash equilibrium of moment-matching GANs for stationary Gaussian processes. Mathematical and Scientific Machine Learning. PMLR, pp. 113–128. https://doi.org/10.48550/arXiv.2203.07136.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y., 2018b. Image super-resolution using very deep residual channel attention networks. Proceedings of the European Conference on Computer Vision (ECCV), pp. 286–301. https://doi.org/10.48550/arXiv.1807.02758.

Zhao, M., Wei, Y., Wong, K.K., 2022. A generative adversarial network technique for high-quality super-resolution reconstruction of cardiac magnetic resonance images. Magn. Reson. Imaging 85, 153–160. https://doi.org/10.1016/j.mri.2021.10.033.

Zhou, T., Li, Q., Lu, H., Cheng, Q., Zhang, X., 2023. GAN review: models and medical image fusion applications. Inf. Fusion 91, 134–148. https://doi.org/10.1016/j.inffus.2022.10.017.