Identification of welding defects from ultrasonic signals using an improved ACGAN and WOA-ShuffleNet V1

Yuan Chen[1,2,*], Xin Su[3], Hao Shen[3], Guangming Zhang[4], Hongwei Ma[2,3], Ming Dong[2,3], Wenbin Liu[3]

1College of Sciences, Xi'an University of Science and Technology, Xi'an, Shaanxi, 710054, China

2Shaanxi Key Laboratory of Mine Electromechanical Equipment Intelligent Detection and Control, Xi'an University of Science and Technology, Xi'an, Shaanxi, 710054, China

3College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi,710054, China

4School of Engineering, Liverpool John Moores University, Byrom Street, L3 3AF, United Kingdom

**Email: chenyuan1030@126.com*

Abstract: In the ultrasonic detection of weld defects, addressing issues such as small-sample multi-class imbalanced distribution of echo signals and lightweight requirements for classification models, a welding defect ultrasonic signal recognition method based on Gramian angular summation field (GASF), improved auxiliary classifier generative adversarial network (ACGAN), and WOA-ShuffleNet V1 is proposed. First, ACGAN is enhanced by integrating a frequency-aware module, latent space optimization, and optimized loss functions to improve generated sample quality. Then, through comparative analysis of expansion ratios and the impact of generative models on classification results, the effectiveness of the improved ACGAN in augmenting multi-class imbalanced small-sample data is validated. Finally, the whale optimization algorithm (WOA) is employed to optimize hyperparameters of ShuffleNet V1. Experimental results show that on the expanded data, the method maintains 90.75% recognition accuracy while reducing FLOPs, model size and number of parameters, thus achieving a balance between lightweighting and classification performance.

Keywords: Ultrasonic signal; Welding defect identification; ACGAN; WOA; ShuffleNet V1

1 Introduction

Welding is a vital process in modern industrial manufacturing. The quality of the welding plays a pivotal role in determining the mechanical properties, structural integrity and long-term service reliability of welding components. However, due to factors such as material properties, process parameter fluctuations and operator skill levels, typical defects such as porosity, cracks, lack of penetration and slag inclusions are prone to occurring during the welding process. Due to its non-destructive nature, high sensitivity, strong penetration and real-time dynamic testing capability, ultrasonic testing has become a core method for evaluating welding quality. Nevertheless, existing ultrasonic testing methods for defect identification still have shortcomings, including low reliability in signal feature extraction, insufficient numbers of defect samples, imbalanced distribution of defect types, and poor generalization capability in traditional machine learning models. Thus, advancing defect identification techniques in ultrasonic welding testing is essential to ensure the structural integrity and operational reliability of welding components [1-3].

To address the limitations of conventional methods in terms of small-sample learning, feature representation and model generalizability, researchers have leveraged generative adversarial network (GAN). Through data augmentation and feature space optimization, GANs provide an effective solution for defect identification in industrial applications with small-sample imbalance. Posilović et al. [4] developed a novel GAN model that can generate realistic ultrasonic B-scan images. Training this model using a combination of real and synthetic data increased the average accuracy of deep convolutional neural networks (CNNs) to 76%. Fan et al. [5] combined GAN with the least square

inverse time shift algorithm to significantly improve the accuracy and efficiency of defect detection in high-density polyethylene materials. Luo et al. [6] adopted the discriminator of a DCGAN as the backbone of the segmentation network. By leveraging transfer learning and incorporating an adversarial loss to optimize segmentation accuracy, this approach established a novel strategy for pixel-level identification of welding defects, achieving a segmentation intersection-over-union (IoU) of 93.3%. In response to the scarcity of labelled data, Irfan et al. [7] developed a parallel convolutional semi-supervised GAN (PC-SSGAN) that uses time-frequency representations and leverages bottleneck-style blocks to enhance multi-scale feature extraction. This approach achieved an accuracy of over 99.50% in bearing fault classification, while reducing the number of model parameters and the time taken for training by approximately 33%. Yang et al. [8] proposed an attention self-supervised learning auxiliary classifier generative adversarial network (ASSL-ACGAN) algorithm, which significantly improved the accuracy of welding defect recognition in small sample datasets by expanding X-ray film samples and optimizing the conversion of pre-processed data. Yuan et al. [9] presented a data augmented generative adversarial network (DCSGAN) that extracts high-dimensional defect spatial features through bilinear interpolation and a deep network. This effectively solved the problem of scarce ultrasonic welding defect data, and significantly improved the detection rate of defects. Yu et al. [10] developed a zero-sample learning and GAN-based model for detecting welding defects in the pipeline, which achieved efficient and accurate recognition of both visible and invisible defects through cross-modal feature fusion. Peng et al. [11] proposed an ultrasonic Wasserstein generative adversarial network (US-WGAN), which effectively solved the problem of insufficient sample data in pipeline ultrasonic non-destructive testing and achieved a defect classification accuracy of over 97%. The above research indicates that GAN has become a vital technology for addressing the issue of data scarcity and distribution imbalance in welding defect detection.

It should be emphasized that ultrasonic echo signals originating from welding defects represent a form of non-stationary time series data, whose characteristic patterns are intrinsically correlated with specific defect types. To comprehensively extract time-frequency domain characteristics, current research predominantly converts one-dimensional signals into two-dimensional time-frequency representations, which are subsequently processed using CNNs to facilitate efficient feature extraction and defect classification. Miao et al. [12] employed continuous wavelet transform (CWT) to generate two-dimensional time-frequency images, which were subsequently classified using a CNN, achieving a classification accuracy of 96.94% for narrow welding defects. Chen et al. [13] developed a hybrid approach integrating variational mode decomposition (VMD) with recurrence plots and gradient graphs to construct discriminative feature representations, which were subsequently classified using the VGG16 network to achieve a recognition accuracy of 97.5%. Zhang et al. [14] employed a classification framework based on graphical differential field coding of time series data, utilizing a full connected neural network for defect identification. With the rapid advancement of deep learning technology, CNNs and their enhanced variants are finding increasingly widespread application in ultrasonic testing for welding defects. Zhang et al. [15] used a CNN to process ultrasonic guided wave signals, achieving accurate identification of defect characteristics including type, severity and spatial location in complex operational environments. Zhang et al. [16] constructed a dataset combining finite element simulations and experimental measurements. They used wavelet transform images of reflected and transmitted signals as input for a convolutional neural network, achieving precise identification of surface defect penetration depth and subsurface defect width, with error rates of 5.87% and 5.6% respectively. This validated the combined use of time-frequency images and deep learning.

Pawar et al. [17] proposed a classification method utilizing phased array ultrasonic A-scan signals combined with discrete wavelet transform (DWT) features, achieving an average classification accuracy of 80% for weld defect identification. Munir et al. [18] demonstrated enhanced accuracy in welding defect classification through the integration of autoencoder networks with CNNs.

To address the critical need for real-time industrial inspection performance, lightweight neural networks have emerged as a key research focus in ultrasonic welding defect detection. Liu et al. [19] proposed a lightweight network architecture that effectively extracted discriminative features from small samples of welding defects through parameter fine-tuning, hereby significantly enhancing both recognition accuracy and computational efficiency. Meanwhile, Zhang et al. [20] developed a multi-domain and multi-scale deep feature adaptive fusion method for welding defect detection. By constructing a time-domain dataset, designing a cross-domain information fusion model, and optimizing CNN hyperparameters, they achieved a recognition accuracy of 96.54%. Zhang et al. [21] proposed a lightweight, self-optimizing, multi-scale MobileNetV3 (SLM-MobileNetV3) model based on multi-spatial domain datasets, achieving 97.26% recognition accuracy for five welding defect types by constructing a multi-spatial domain dataset and designing with a processing time of merely 2ms per image. Pan et al. [22] developed TL-MobileNet, a transfer learning-based model for welding defect classification, which attained an accuracy of 97.69%. Zhang et al. [23] proposed an improved deep convolutional generative adversarial network (DCGAN) integrated with a lightweight DG-MobileNet architecture for welding defect recognition, achieving 98.78% classification accuracy. Ding et al. [24] developed an improved, lightweight MobileNetV2 model that can recognize in-situ welding surface defects at high speed.

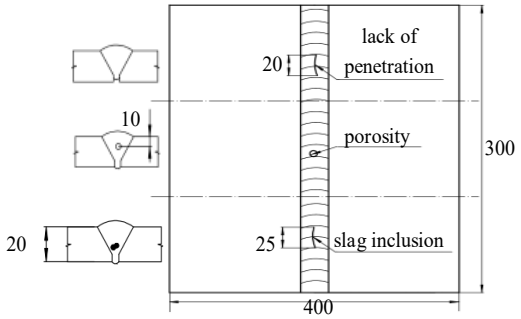
Although traditional GANs can effectively augment samples and balance classes, they suffer from two major limitations: (1) class-specific training requirements lead to computational complexity that scales linearly with the number of classes; (2) dependence on substantial initial sample sizes. Moreover, welding defect recognition remains insufficient under conditions of small-sample and multi-class imbalance. Therefore, this paper proposes a novel welding defect identification method combining GASF, an improved ACGAN and a WOA-ShuffleNet V1. First, the ultrasonic echo signals from welding defects are converted into time-frequency representations using GASF. Then, the dataset is augmented using the improved ACGAN. Finally, WOA is employed to automatically determine the optimal combination of hyperparameters for the ShuffleNet V1 model, achieving an optimal balance between model lightness and classification performance.

2 Dataset preparation

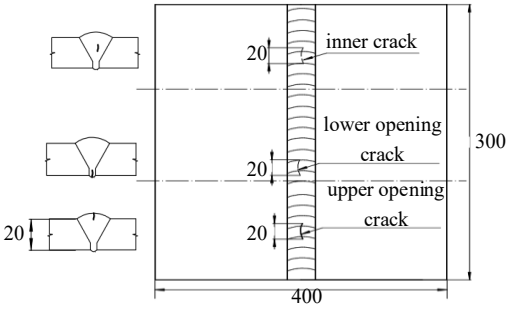
2.1 Welding specimens and signal acquisition

Five welding specimens made of 45# carbon steel are designed and fabricated. Of them, the plate-plate butt specimen I contains one each of three types of defects such as lack of penetration, porosity and slag inclusion. The plate-plate butt specimen II contains three types of crack defects such as upper opening, inner and lower opening cracks. The tube-plate butt specimen contains two cracks and one each of the other three types of defects. Both of the tube-tube butt specimens contain one each of all four types of defects. The detailed dimensions and defect locations are provided in Fig. 1 [25]. The principle of ultrasonic testing is based on wave reflection and scattering at material interfaces or defects. The resulting echoes are received by the probe and converted into electrical signals. Critical features such as the time-of-flight, amplitude, and waveform of these signals are subsequently analyzed to determine the location, size and type of the defects.

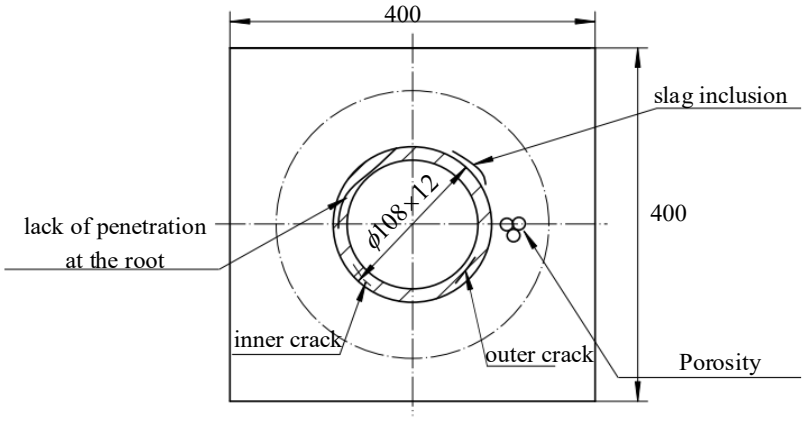
The A-scan ultrasonic echo signals of defects in the welding specimens are collected using a CTS-4020 digital ultrasonic flaw detector with five types of angle probes as shown in Fig. 2. The angle probes are 2.5Z10×10A45, 5Z8×13A45, 2.5P13×13A60, 2.5Z10×10A70 and 5Z10×10A70, respectively. By systematically varying the probe position and direction, and the gain of the digital ultrasonic flaw detector, a total of 2100 ultrasonic A-scan signals are collected during the inspection of the above welding specimens [25]. For the specimens shown in Figs. 1(a, b, d), ultrasonic signals are collected from two locations on both sides of the weld. For the specimen shown in Fig. 1(c), ultrasonic signals are collected from a single location on the steel plate. At each position, the probe is oriented in three distinct directions: aligned directly with the defect, and then angled to the left and right. The collection distance is 2, i.e. each before and after. Two different gains are used for each collection. In addition, five probes are randomly selected to collect 30 ultrasonic signals for each type of defect. The total number of ultrasonic signals collected is as follows: crack: (5×2×3×2×2+2×1×3×2×2)×5+30=750; slag inclusion, porosity and lack of penetration each: (3×2×3×2×2+1×1×3×2×2)×5+30=450.



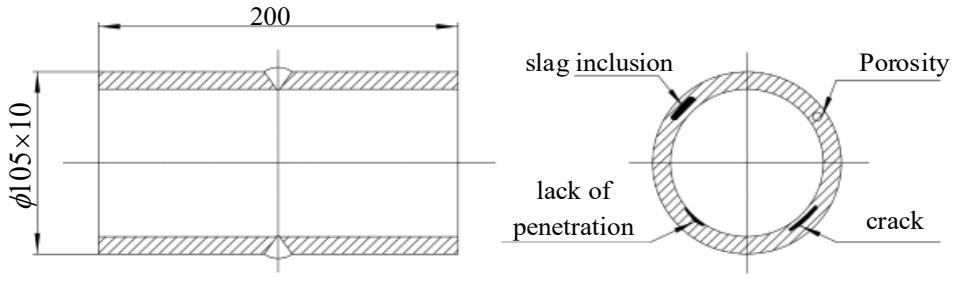
(a) Plate-plate butt specimen I



(b) Plate-plate butt specimen II



(c) Tube-plate butt specimen



(d) Tube-tube butt specimens I and II

Fig. 1 Welding specimen dimensions and defect location.

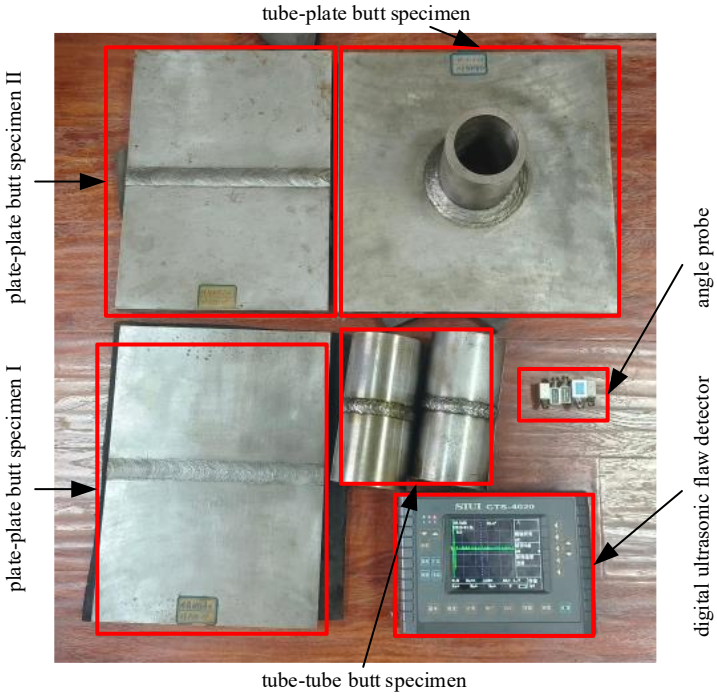


Fig. 2 Welding specimens and ultrasonic signal acquisition.

From the total of 2100 collected signals, a subset of 1800 ultrasonic A-scan signals representing various defects (including slag inclusions, cracks and lack of penetration) is selected for this study. The detailed distribution of these signals across defect types is provided in Table 1.

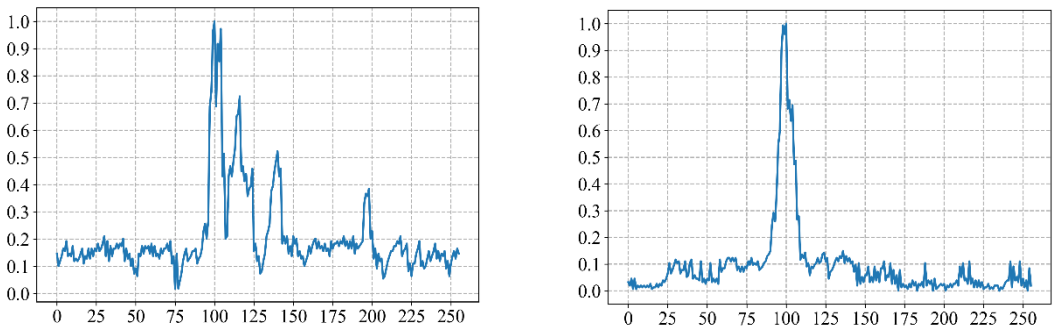
Table 1 Multiclass imbalanced small-sample dataset.

Type of defect	Slag inclusion	Crack	Porosity	Lack of penetration
Number	400	600	450	350

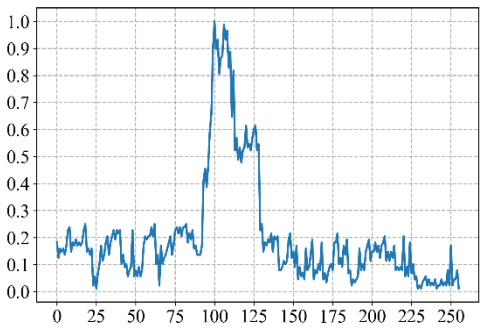
2.2 Preprocessing of ultrasonic echo signals

For the collected 1800 A-scan ultrasonic echo signals, only the segment containing the defect information is retained, with each segment fixed at a length of 256 sample points. This length is determined to be sufficient for capturing the complete echo waveforms of all defect types in the experiments while facilitating subsequent computations. To eliminate the influence of signal amplitude on defect detection, the clipped signals are normalized according to Eq. (1). This normalization yields the A-scan signals for the four welding defect types, as illustrated in Fig. 3 [25].

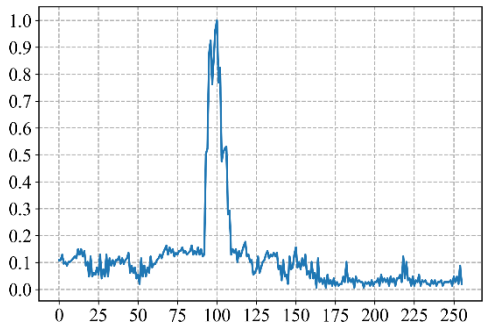
$$X_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{1}$$



(a) slag inclusion (b) crack



(c) porosity



(d) lack of penetration

Fig.3 Representative ultrasonic signals for four types of welding defects.

The ultrasonic signal characteristics of each defect type are determined by their underlying physical mechanisms:

(1) Slag inclusion: The signal consists of clustered, low-amplitude waveforms due to scattering from irregularly shaped slag particles that have an acoustic impedance similar to that of steel.

(2) Crack: The echo is high-amplitude and steep when the planar defect is aligned with the beam, alongside pronounced attenuation of the backwall echo.

(3) Porosity: The response is characterized by symmetric, low-amplitude echoes from spherical pores or bunches of periodic echoes for clustered porosity.

(4) Lack of penetration: The signal is characterized by a stable, singular, high-amplitude echo originating from a continuous linear defect at the center of the weld.

2.3 Preparation of welding defect image dataset using GASF

Gramian angular field [26] (GAF) consists of two variants: GASF and Gramian angular difference field (GADF), which maps one-dimensional time series data into a two-dimensional representation through polar coordinate transformation, employing the trigonometric function to construct a matrix that effectively captures the time dependence and periodicity of the time-series data. The specific steps are as follows:

(a) Reduce the size by approximating the aggregation sequence using segmented aggregation, and normalize the one-dimensional data X using the following formula:

$$\tilde{X}_i = \frac{X_i - \min(X)}{\max(X) - \min(X)} \times 2 - 1 \tag{2}$$

(b) Encode the normalized data as the angle cosine, with the timestamp as the radius, and the scaled data $\tilde{X}$ can be expressed in polar coordinates as:

$$\begin{cases} \phi_i = \arccos(X_i), 0 \le X_i \le 1, X_i \in \tilde{X} \\ r_i = \dfrac{t_i}{N}, t_i \in N \end{cases} \tag{3}$$

where t_i is the timestamp, N is a constant factor to regularize the span of the polar coordinate system, and ϕ_i is the radians of the scaled data $\tilde{X}$ in polar coordinates, i.e. the angular cosine.

(c) Calculate the GASF value using the following formula:

$$G_{GASF} = \begin{pmatrix} \cos(\phi_1 + \phi_2) & \cdots & \cos(\phi_1 + \phi_n) \\ \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{pmatrix} \tag{4}$$

These signals are subsequently converted into images using the GASF algorithm, resulting in a

dataset of 1800 images, each with a resolution of 64×64 pixels.

3 The proposed method

3.1 Improved ACGAN

ACGAN [27] is an advanced generative model that integrates a GAN with an auxiliary classifier. The improved ACGAN learns multi-modal feature representations of welding defects through adversarial training. A key architectural innovation requires the discriminator to perform dual tasks: distinguishing between real and generated samples, and predicting the defect type of each sample. Consequently, the generator learns to map specific defect types onto distinct texture patterns that have a physical meaning. For instance, cracks are associated with linearly continuous textures that reflect their elongated morphology, while porosity is represented by discrete, circular patterns that correspond to its spherical nature. Although ACGAN improves generation diversity and controllability through categorical manipulation, it still suffers from some issues including indistinct image details, insufficient diversity and mode collapse phenomena. Therefore, this paper proposes the following improvements to the ACGAN framework:

(1) Incorporation of frequency-aware module

The structure of this module, illustrated in Fig. 4, employs a dual-path parallel design to optimize feature extraction through frequency transformation, frequency-domain convolution, and channel attention mechanisms. The spatial path preserves local texture and morphological details via standard convolution, while the frequency path maps the input to the frequency domain using a Fast Fourier Transform (FFT) to capture global spectral distributions. Frequency-domain convolution is then applied to isolate key frequency components critical for defect characterization. Meanwhile, an attention mechanism, implemented via global average pooling and fully-connected layers, generates channel-wise weights to emphasize informative features. Finally, the refined frequency information is integrated back with the spatial features to enhance representational power.

The frequency-aware module introduced captures the distinct signatures of different defect types in the frequency domain. For example, the sharp reflective interfaces of cracks result in prominent high-frequency components, while the spherical morphology of porosity produces stronger mid-to-low frequency responses. The ability to analyze these spectral signatures enhances the model's discriminative power by leveraging the fundamental physical properties of defects directly.

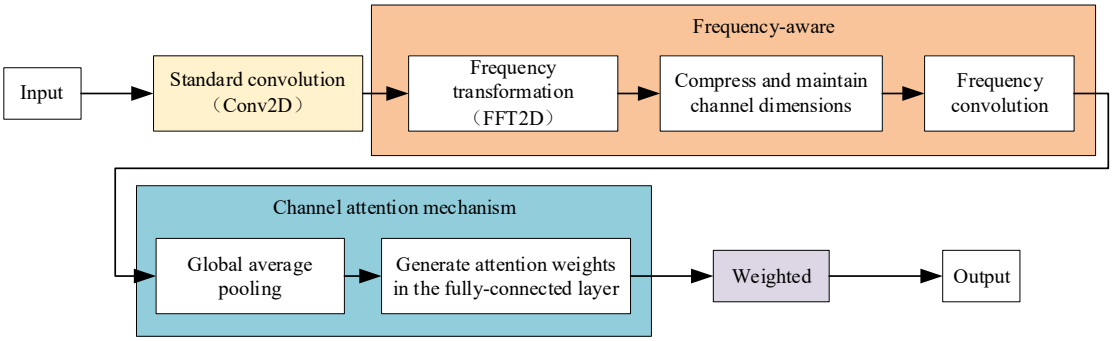


Fig. 4 Schematic of frequency-aware module architecture.

(2) Introduction of latent space optimization [28]

The latent space optimization module is designed to enhance training stability and output quality in the generative adversarial network. It operates by imposing regularization constraints on the latent vectors, which encourages the generation of more realistic images while preventing excessive

deviation during latent space exploration. This mechanism effectively mitigates mode collapse and promotes sample diversity. To this end, the latent space regularization loss constrains the distance between the optimized latent vector z and the initial random sampling vector z_0 using the L2-norm, formulated as:

$$L_{reg} = \beta \cdot E[\| z - z_0 \|^2] \tag{5}$$

where β is the weight coefficient of the regularization term, which controls the strength of the latent space optimization.

(3) Loss function optimization

The discriminator aims to produce the outputs close to 1 for real samples and close to 0 for fake samples, thereby distinguishing between the two. The discriminator loss function comprises the adversarial loss, which determines whether an image is real, and the category loss, which determines the category of an image. Their expressions are as follows:

$$L_{disc} = E_{x\sim p_{data}}[(D(x)-1)^2] + E_{z\sim p_z}[(D(G(z)))^2] \tag{6}$$

$$L_{label} = E_{x\sim p_{data}}[-\sum_i y_i \log(D_{label}(x))] + E_{z\sim p_z}[-\sum_i y_i \log(D_{label}(G(z)))] \tag{7}$$

The total loss of the discriminator can be calculated by the following formula:

$$L_{total\_disc} = L_{disc} + L_{label} \tag{8}$$

where $D(x)$ is the output of the discriminator for the real sample x, p_{data} is the distribution of the real data; $D(G(z))$ is the output of the discriminator for the generated sample $G(z)$, p_z is the distribution of the generated sample, $G(z)$ is the output of the generator, $D_{label}(x)$ and $D_{label}(G(z))$ are the predictions of the discriminator for the real sample and the real sample label, respectively, and y_i is the label.

The task of the generator is to generate fake samples that the discriminator perceives as real. The aim is to produce an output from the discriminator close to 1, i.e., the output of the fake samples should be close to 1. The generator loss is similar to that of the traditional GAN, which mainly comprises the loss from minimizing the discriminator's judgement of the authenticity of the generated images (adversarial loss) and the loss from the generated image category (category loss). They can be expressed as follows:

$$L_{gen} = E_{z\sim p_z}[(D(G(z))-1)^2] \tag{9}$$

$$L_{gen\_label} = -\sum_i y_i \log(G_{label}(G(z))) \tag{10}$$

The total loss of the generator can be calculated by the following formular:

$$L_{total\_gen} = L_{gen} + L_{gen\_label} + L_{reg} \tag{11}$$

As illustrated in Fig. 5(a), the generator uses a transposed convolution structure to gradually convert latent vectors into target images. The discriminator uses a convolution structure to distinguish between the real and the generated images, as illustrated in Fig. 5(b).

The specific parameters of the generator and discriminator are shown in Tables 2 and 3.

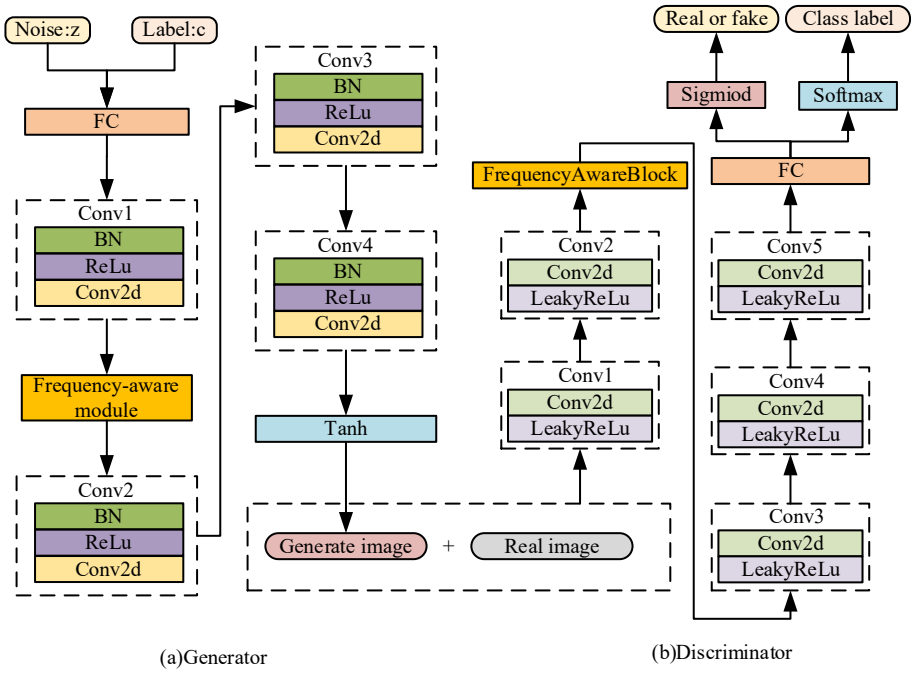


Fig. 5 Schematic diagram of the generator and discriminator structure.

Table 2 Generator parameter settings.

	Layers	Input	Output	Convolution kernel	Stride	Padding	Dropout
Generator	Fully-connected layer	[100,4]	[16,16,128]	—	—	—	—
	Transposed conv1	[16,16,128]	[32,32,128]	5×5	2	1	—
	Frequency-aware module	[32,32,128]	[32,32,128]	3×3	—	—	—
	Transposed conv2	[32,32,128]	[64,64,64]	5×5	2	1	—
	Transposed conv3	[64,64,64]	[64,64,32]	5×5	1	1	—
	Transposed conv4	[64,64,32]	[64,64,3]	5×5	1	1	—

Table 3 Discriminator parameter settings.

	Layers	Input	Output	Convolution kernel	Stride	Padding	Dropout
Discriminator	Conv1	[64,64,3]	[32,32,32]	5×5	2	1	0.2
	Conv2	[32,32,32]	[16,16,64]	5×5	2	1	0.2
	Frequency-aware module	[16,16,64]	[16,16,64]	3×3	1	—	—
	Conv3	[16,16,64]	[8,8,64]	5×5	2	1	0.2
	Conv4	[8,8,64]	[4,4,128]	5×5	2	1	0.2
	Conv5	[4,4,128]	[4,4,256]	5×5	1	1	0.2
	Fully-connected layer	[4,4,256]	[1,4]	—	—	—	—

To mitigate the impact of imbalanced data in small samples on defect recognition results, this paper utilizes an improved ACGAN to augment the dataset. As illustrated in Fig. 6, the model consists of a generator, a discriminator and a frequency-aware module. The process begins by converting ultrasonic echo signals into two-dimensional images using the GASF transform. The generator then

maps a random noise vector, conditioned on defect category labels, to the data space via adversarial training in the latent space. The generated samples are optimized to satisfy two objectives simultaneously: the adversarial loss, which aims to deceive the discriminator; and the category loss, which ensures feature consistency with the target defect type. Finally, the discriminator evaluates a mixture of real and generated samples and is responsible for both classifying the defect type and distinguishing between real and generated samples.

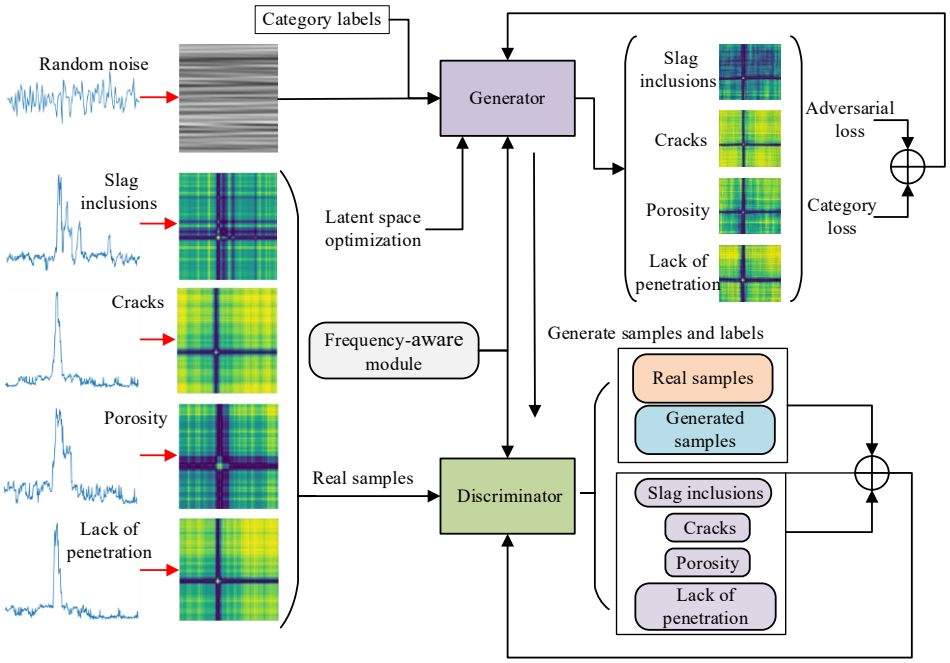


Fig. 6 Schematic diagram of the structure for generating welding defect image samples based on improved ACGAN.

3.2 WOA-optimized ShuffleNet V1 model

ShuffleNet V1 [29] is an efficient lightweight CNN architecture that achieves an optimal balance between computational efficiency and model accuracy through channel shuffle operations and pointwise group convolution. Its structure for welding defect time-frequency images with an input dimensionality of 64×64×3 is shown in Fig. 7. WOA-ShuffleNet V1 identifies defect features in GASF images through a hierarchical feature extraction mechanism. In its initial layers, convolutional kernels learn to detect low-level patterns, including: (1) edge and corner detection for identifying the linear boundaries characteristic of cracks; (2) spot detection for locating the circular contours of porosity; and (3) textural filters for capturing the irregular morphologies of slag inclusions. These basic features are subsequently combined in deeper layers to form more complex and discriminative representations.

Through channel shuffle operations, the model effectively combines these basic features to create more complex, hierarchical patterns. This process enables comprehensive defect morphologies to be learnt by synthesizing diverse feature information. The final layer of the network then distils these integrated patterns into semantically discriminative features, establishing a direct correspondence with the unique characteristics of the four defect types.

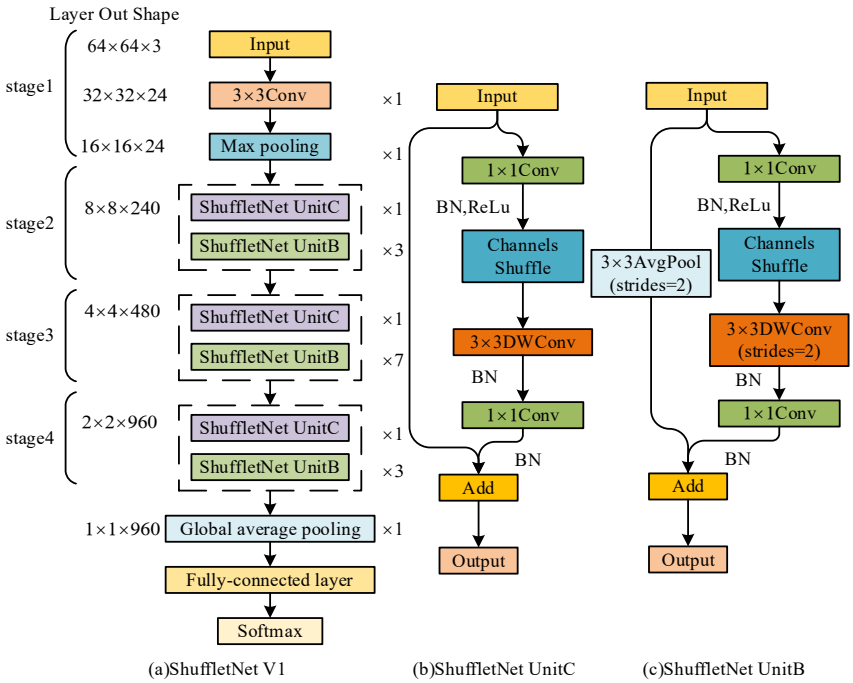


Fig. 7 Schematic diagram of the structure of the ShuffleNet V1 model.

The WOA is an intelligent global optimization algorithm that simulates the hunting behavior of humpback whales [30]. It is straightforward to implement and requires fewer parameters, making it easy to adjust. In this paper, the WOA is used to optimize the key hyperparameters in the training of the ShuffleNet V1 model, including the learning rate, batch size and the number of groups. Its optimization flow is illustrated in Fig. 8.

During the process of hyperparameter optimization process, the fitness function is used to evaluate the performance of the combination of network hyperparameters. In this paper, the fitness function is defined as the negative value of the classification accuracy of the test set, which is expressed as follows:

$$f(X) = -\text{test\_accuracy}(X) \tag{12}$$

where X is a hyperparameter vector containing the learning rate, batch size and number of groups, and $\text{test\_accuracy}(X)$ is the test set classification accuracy.

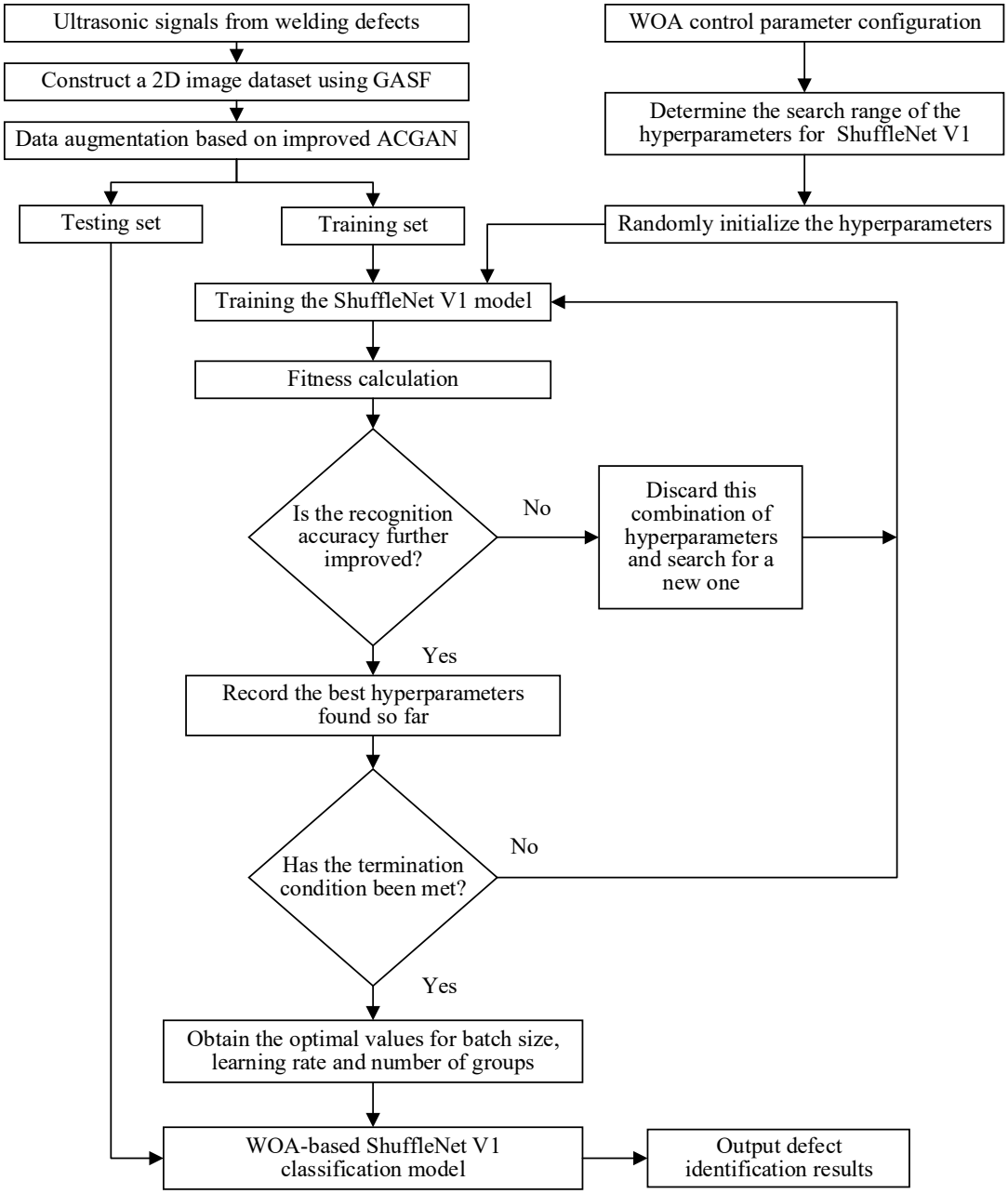


Fig.8 Flowchart of the hyperparameter optimization for the WOA-based ShuffleNet V1 model.

4.Experimental results and analysis

All experiments in this paper are completed using the TensorFlow 2.6 framework on a Windows 11 operating system with a Python programming environment. Network training and testing are performed on a computer primarily configured with an Intel Core i9-13900KF CPU, 32G RAM memory, and an NVIDIA GeForce RTX 4090 GPU.

4.1 Quality assessment of generated samples

For the multi-class imbalanced small-sample welding defect image dataset, sample augmentation and balancing are performed using the improved ACGAN. The experimental results are then compared and analyzed with those of CGAN and ACGAN. The parameter settings are provided in Table 4.

Table 4 Hyperparameter configuration for the three comparative models.

Model	Number of iterations	Batch size	Learning rates	Optimizer
CGAN	4000	64	D: 0.0002 G: 0.0001	RMSprop
ACGAN	4000	64	D: 0.0002 G: 0.0001	RMSprop
Improved ACGAN	4000	64	D: 0.0002 G: 0.0001	RMSprop

Fig. 9 gives real images alongside those generated by CGAN, ACGAN and the improved ACGAN. As can be seen, the generated images from the improved ACGAN are superior to those from the CGAN and ACGAN in terms of visual fidelity.

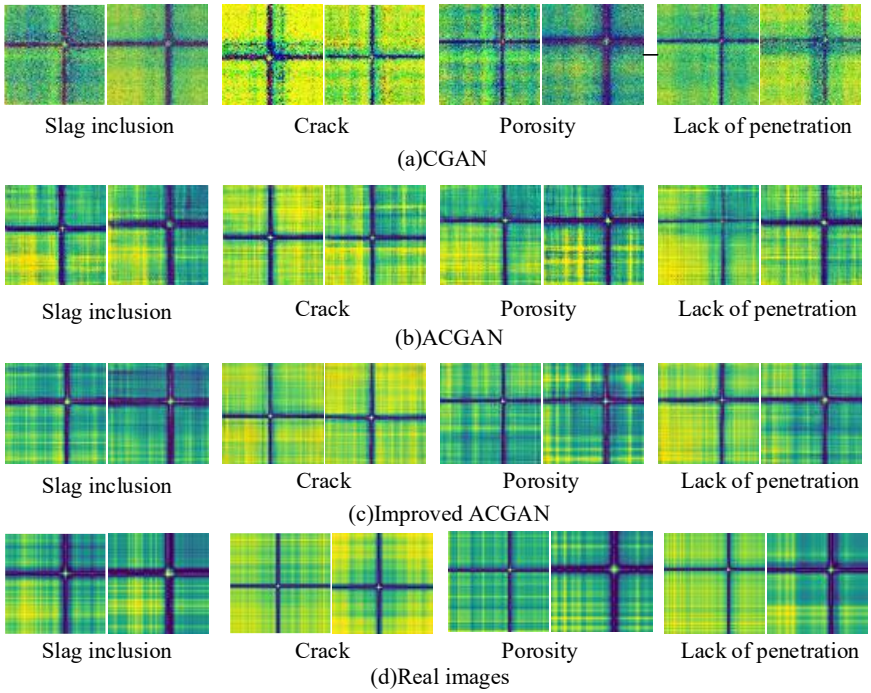


Fig. 9 Real and generated images from three different methods.

To quantitatively evaluate the effectiveness of the improved ACGAN method, in this paper, four metrics such as Fréchet inception distance (FID), structural similarity index (SSIM), peak signal-to-noise ratio (PSNR) and learned perceptual image patch similarity (LPIPS) are adopted to comprehensively analyze the quality of generated images from different perspectives.

(1) FID is an important metric used to quantify the quality of generated images. It evaluates differences in the statistical distribution between real and generated images in the feature space. The smaller the FID value, the better the diversity and quality of image, and the higher its quality, which is calculated as follows:

$$FID = \left\| m - m_{\omega} \right\|^2 + Tr(C + C_{\omega} - 2\sqrt{C \times C_{\omega}}) \tag{13}$$

where m and m_w are the feature means of the real and generated images respectively, C an d C_w are the covariance matrices of the real and generated feature vectors respectively, $\|m - m_w\|^2$ is the sum of the squared differences between the two mean vectors, and T_r is the sum of the elements on the main diagonal of the square matrix.

(2) SSIM is used to evaluate image quality by comparing the luminance, contrast and structure between two given images. The higher the SSIM value, the better the image's structural integrity and quality. It is calculated as follows:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{14}$$

where x and y are the original and the generated images respectively, μ_x and μ_y are the mean values of x and y respectively, σ_x^2 and σ_y^2 are the variances of x and y respectively, σ_{xy} is the covariance of x and y. Additionally, $c_1 = (k_1 L^2)$ and $c_2 = (k_2 L^2)$ are smaller constants that ensure that the denominator is not zero, and L is the dynamic range of the pixel values, where $k1 = 0.01$, and $k2 = 0.03$.

For each type of defect, the SSIM is calculated for each generated image and its corresponding original image. Then, the real image that is most similar to each generated image (i.e., the one with the largest SSIM value) is used to calculate the SSIM, and they are summed up. Finally, the average value is obtained through dividing by the number of generated images, which can be expressed as follows:

$$SSIM_{Average} = \frac{1}{N}\sum_{i=1}^{N}[\max_j(SSIM(G_i, R_j)) \tag{15}$$

where G_i is the ith generated image, R_j is the jth real image, N is the total number of generated images, and $\max_j(SSIM(G_i, R_j))$ is the largest of all SSIMs calculated for each generated image G_i with all real images R_j.

(3) PSNR is an image quality evaluation metric based on mean squared error (MSE) that measures the distortion of the generated image at a pixel level relative to the original image. It can be calculated as follows:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{16}$$

$$PSNR = 10 \cdot \log_{10}(\frac{MAX_I^2}{MSE}) \tag{17}$$

where $I(i,j)$ and $K(i,j)$ are the pixel values of the original and generated images at the coordinate (i,j) respectively, $m \times n$ is the resolution, and MAX_i is the maximum possible value of image pixels. A higher PSNR value indicates smaller distortion between the generated and original images, and the quality is better.

(4) LPIPS is an image quality evaluation metric that uses deep learning to assess the perceived similarity of two images by calculating the distance between them in the deep feature space. It uses a pre-trained CNN to extract image features, and then calculates the weighted average of feature differences. The pre-trained model used in the experiments is VGG. A smaller LPIPS value indicates a greater visual similarity between the generated and generated images in the perceptual feature space. It is calculated as follows:

$$LPIPS(x,y) = \sum_l \frac{1}{H_l W_l}\sum_{h=1}^{H_l}\sum_{w=1}^{W_l} w_l \bullet \| \phi_l(x)_{h,w} - \phi_l(y)_{h,w} \|_2^2 \tag{18}$$

where $H_l \times W_l$ denotes the size of the feature mapping, $\phi_l(x)$ and $\phi_l(y)$ denote the feature mapping of the original image x and the generated image y in the lth layer, respectively, and w_l is the weight of the lth layer.

The number of samples generated for each type of defect is the same as in Table 1. The quality of the generated images is evaluated using the above four metrics, and the results are presented in

Tables 5-8.

Table 5 FID calculation results of samples generated by three models.

Model	Slag inclusion	Crack	Porosity	Lack of penetration
CGAN	476.11	402.23	507.35	345.56
ACGAN	245.32	186.37	300.58	173.23
Improved ACGAN	195.80	143.05	290.71	125.60

Table 6 SSIM calculation results of samples generated by three models.

Model	Slag inclusion	Crack	Porosity	Lack of penetration
CGAN	0.4817	0.5599	0.5801	0.5857
ACGAN	0.5676	0.6547	0.5792	0.6223
Improved ACGAN	0.5839	0.6816	0.6048	0.6494

Table 7 PSNR calculation of samples generated by the three models.

Model	Slag inclusion	Crack	Porosity	Lack of penetration
CGAN	29.22	29.26	29.22	29.26
ACGAN	29.39	29.34	29.33	29.36
Improved ACGAN	29.49	29.39	29.40	29.40

Table 8 LPIPS calculation results of samples generated by the three models.

Model	Slag inclusion	Crack	Porosity	Lack of penetration
CGAN	0.5974	0.5355	0.5289	0.5273
ACGAN	0.4369	0.3686	0.4263	0.3843
Improved ACGAN	0.3922	0.3101	0.3532	0.3340

The number of defects generated for each type is the same as that shown in Table 1. Compared to the ACGAN, the SSIM has improved by 2.79%, 3.95%, 4.23% and 4.17% for slag inclusion, crack, porosity and lack of penetration, respectively.

As detailed in Tables 5-8, which use PSNR to evaluate quality, the improved ACGAN achieves the highest PSNR values for all defect types. Comprehensive evaluation confirms that the images generated by our model have the lowest FID and LPIPS values, as well as the highest SSIM and PSNR values, compared to those generated by all other models.

Notably, for crack defects, the improved ACGAN demonstrates significant improvements in key metrics compared to the ACGAN: a 23.24% reduction in FID, a 3.95% increase in SSIM, a 0.17% improvement in PSNR and a 15.87% decrease in LPIPS. These quantitative advancements confirm that the generated images are clearer, more diverse, and more closely aligned with the original defect signatures.

4.2 Sample augmentation

The improved ACGAN is used to balance the samples and augment the dataset constructed in Section 2.2 at different ratios. The generated and real samples are combined to form a new dataset, which is divided into training, validation and test sets at a ratio of 8:1:1. Subsequently, the ShuffleNet V1 model is used to verify the effectiveness of the improved ACGAN for sample augmentation. The parameter settings are provided in Table 9.

Table 9 Hyperparameter configuration for training the ShuffleNet V1 benchmark model.

Number of iterations	Batch size	Learning rates	Optimizer
50	64	0.001	Adam

The experimental results are shown in Table 10.

Table 10 Experimental results of sample augmentation.

Dataset	Slag inclusion	Crack	Porosity	Lack of penetration	Total	Accuracy	Increase	Training time/s	Model size/MB
1	400	600	450	350	1800	70.56%	—	23.78	35.47
2	450	750	450	450	2100	76.19%	4.68%	24.27	35.47
3	800	800	800	800	3200	85.94%	9.75%	32.92	35.47
4	1000	1000	1000	1000	4000	88.50%	2.56%	38.70	35.47
5	1200	1200	1200	1200	4800	86.87%	-1.63%	45.76	35.47

As can be seen, balancing and augmenting the samples significantly improves the accuracy of defect identification. However, once the dataset contains a sufficient number of samples, the identification accuracy tends to stabilize or even decrease slightly. In addition, as the dataset size increases, the training time increases while the model size remains unchanged.

To further evaluate the performance of the improved ACGAN, the CGAN and ACGAN are also used for sample augmentation, generating a total of 4000 defect samples respectively. The experimental results are shown in Table 11. As can be seen, the improved ACGAN achieves the best augmentation effect of three models.

Table 11 Experimental results of the three models for sample augmentation.

Model	Slag inclusion	Crack	Porosity	Lack of penetration	Accuracy	Recall	AUC	Increase
CGAN	1000	1000	1000	1000	81.25%	80.00%	92.05%	—
ACGAN	1000	1000	1000	1000	84.25%	83.75%	93.36%	3.00%
Improved ACGAN	1000	1000	1000	1000	88.50%	88.50%	95.71%	4.25%

After applying the augmented data generated by the improved ACGAN model, significant improvements are observed across multiple key metrics on the test set. Accuracy increased substantially, and two other core metrics—recall and the area under the ROC curve (AUC)—also showed marked enhancements. Recall, which measures a model's ability to correctly identify all actual defect cases, increased from 83.75% to 88.50%. The AUC metric evaluates the model's overall capacity to distinguish between defect categories across all possible decision thresholds; higher values indicate more stable and reliable performance. In these experiments, the AUC value increased from 93.36% to 95.71%. The substantial improvement in recall signifies a significant reduction in the model's omission rate across various defect detection scenarios. Concurrently, the enhanced AUC value demonstrates the model's improved discriminative capability and better trade-off between detection and false positive rates.

4.3 Welding defect identification

As can be seen in Section 4.2, dataset 4 achieves the best classification results for the ShuffleNet V1 model. Therefore, WOA is used to optimize the hyperparameters of the ShuffleNet V1 model based on dataset 4. The WOA parameters are configured as follows: population size of 10 whales and 10 iterations. The optimal hyperparameters obtained through WOA optimization for the ShuffleNet V1 model are: learning rate of 0.00156, batch size of 110, and group count of 1. Dataset 4 is divided into training, validation and test sets using an 8:1:1 ratio. The training is conducted for 50 epochs, with the remaining parameters consistent with those of the ShuffleNet V1 model in Section 4.2.

Table 12 gives the experimental results for the ShuffleNet V1 model before and after hyperparameter optimization. As can be seen, the accuracy is improved by 2.25% after optimization, demonstrating the effectiveness of optimizing the hyperparameters of the ShuffleNet V1 model using

the WOA. The recall rate also increases to 90.25%, indicating that hyperparameter optimization enhances overall accuracy while reducing the rate at which defects are missed. The AUC rises to 96.88%, reflecting the stable and excellent comprehensive classification performance achieved by the WOA-ShuffleNet V1 model. Furthermore, the increased batch size reduces training time.

Table 12 Experimental results for the ShuffleNet V1 model before and after hyperparameter optimization.

Network model	Learning rate	Batch size	Number of subgroups	Accuracy	Recall	AUC	Training time (s)	Model size (MB)
ShuffleNet	0.001	64	1	88.50%	88.50%	95.71%	38.70	35.47
WOA-ShuffeNet V1	0.00156	110	1	90.75%	90.25%	96.88%	36.04	35.47

To further analyze the generalization ability of the network model, a confusion matrix is used to visualize and evaluate its performance, and the results are shown in Fig. 10. As can be seen, the identification accuracy for slag inclusion, lack of penetration, porosity and crack are 90%, 94%, 83% and 96%, respectively. Of these, crack identification has the highest accuracy rate, while porosity identification has the lowest.

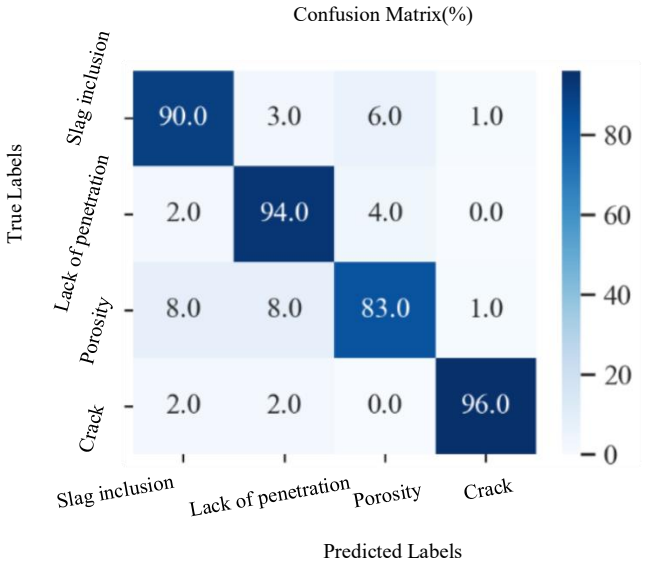


Fig. 10 Visualization of the results.

4.4 Model lightweight

Floating point operations (FLOPs) are an important metric for measuring the computational complexity of deep learning models. Unlike the number of parameters, FLOPs more directly reflect the computational requirements of the model, making them crucial for model optimization, accelerating inference and deploying hardware. Lower FLOPs generally indicate faster computation speeds, particularly in environments with limited resources, such as on mobile devices or in embedded systems. However, the computational process for FLOPs varies across different layers. For instance, the formula for calculating FLOPs in a convolutional layer is as follows:

$$FLOPs = 2 \times H \times W \times C_{in} \times C_{out} \times K^2 \tag{19}$$

where $H \times W$ is feature map size; C_{in} and C_{out} are the number of the input and output channels, respectively, and K is the convolution kernel size.

To further validate the lightweight performance of WOA-ShuffleNet V1, we compared it with

several mainstream convolutional neural network models, including GoogleNet, ResNet-18 and ResNet-50. The experiments use Dataset 4 from Section 4.2 for training, which is divided into training, validation and test sets at a ratio of 8:1:1. The model training parameter configuration is consistent with that provided in Table 9.

The hyperparameters for WOA-ShuffleNet V1 are the optimal parameters determined in Section 4.3, except for the number of groups, which is changed while keeping all other parameters constant. The results are shown in Table 13.

Table 13 Comparison of the model's FLOPs, the number of parameters and the accuracy.

Model	Number of groups	Accuracy	FLOPs (10^6)	Model size (MB)	Number of parameters (10^6)
GoogleNet	—	90.25%	452.20	67.41	5.87
ResNet-18	—	92.25%	299.09	128.47	11.20
ResNet-50	—	90.75%	597.16	257.72	22.47
WOA-ShuffleNet V1	g=1	90.75%	224.25	35.47	2.66
	g=2	82.75%	114.87	20.82	1.37
	g=3	84.50%	78.42	15.90	0.94
	g=4	87.50%	60.19	13.44	0.73

To comprehensively evaluate the trade-off between accuracy and computational complexity in the models, this paper uses the Pareto frontier analysis, a methodology derived from multi-objective optimization theory, to evaluate the trade-off between model accuracy and computational complexity in a comprehensive way. The Pareto frontier identifies the optimal solution set that achieves the best possible balance between competing objectives. Fig. 11 illustrates the distribution of all evaluated models within this two-dimensional design space, clearly showing the optimal region in which our WOA-ShuffleNet V1 architecture outperforms conventional approaches in terms of performance and efficiency.

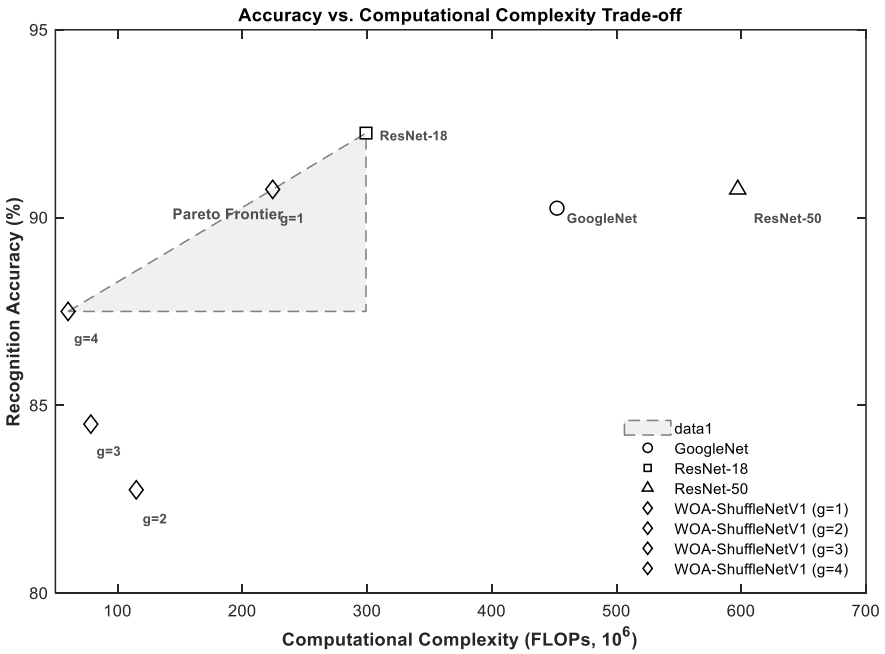


Fig. 11 Performance evaluation based on Pareto frontier analysis comparing model accuracy with computational complexity.

As shown in Fig. 11, both ResNet-18 and WOA-ShuffleNet V1 (g=1) are on the Pareto frontier, which indicates an optimal balance between high accuracy and relatively low computational complexity. However, WOA-ShuffleNet V1 has superior practical advantages due to its substantially

smaller model size and reduced parameter count.

Table 13 shows that WOA-ShuffleNet V1 achieves lower computational costs (FLOPs), a more compact model size and fewer parameters than the other three benchmark models, while maintaining competitive recognition accuracy. Reducing the number of groups in our architecture yields substantial efficiency gains with minimal accuracy loss. Specifically, increasing groups from 1 to 4 results in a 73.16% decrease in FLOPs, a 62.11% decrease in model size, and a 72.56% decrease in parameters, while recognition accuracy decreases by only 3.25%.

5 Conclusions

This paper proposes an intelligent identification method for welding defects from ultrasonic signals that integrates GAS, an improved ACGAN and a WOA-ShuffleNet V1. The effectiveness and feasibility of this method have been verified through the experiments. The key findings are as follows:

(1) The improved ACGAN achieves significant improvements in the FID, SSIM, PSNR and LPIPS metrics by incorporating a frequency-aware module, latent space optimization and an improved loss function. This model effectively addresses the balancing and augmentation issues associated with multi-class, imbalanced and small-samples, thereby enhancing the accuracy of the classification model.

(2) The WOA is used to automatically optimize the hyperparameters of the ShuffleNet V1 model. Experimental results demonstrate that the optimized model achieves a defect identification accuracy of 90.75%. Furthermore, the FLOPs, model size and number of parameters of the WOA-ShuffleNet V1 model are significantly reduced compared to other commonly used classification models.

(3) Adjusting the number of groups effectively reduces the FLOPs, model size and number of parameters of the WOA-ShuffleNet V1 model. When the number of groups is set to 4 rather than 1, the FLOPs, model size and number of parameters are reduced by 73.16%, 62.11% and 72.56% respectively, while still achieving an identification accuracy of 87.5%.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China [Grant No. 52175518].

References

[1] Gupta M, Khan M A, Butola R, et al. Advances in applications of non-destructive testing (NDT): a review[J]. Advances in Materials and Processing Technologies, 2022, 8(2): 2286-2307.

[2] Sun H, Ramuhalli P, Jacob R E. Machine learning for ultrasonic nondestructive examination of welding defects: A systematic review[J]. Ultrasonics, 2023, 127: 106854.

[3] Amarnath M, Sudharshan N, Srinivas P. Automatic detection of defects in welding using deep learning-a systematic review[J]. Materials Today: Proceedings, 2023.

[4] Posilović L, Medak D, Subašić M, et al. Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans[J]. Neurocomputing, 2021, 459: 361-369.

[5] Fan L, Xiao Z, Dong F, et al. Defect detection using integration of ultrasonic least-squares reverse

time migration and generative adversarial network[J]. Nondestructive Testing and Evaluation, 2024: 1-15.

[6] Luo Y, Ling J, Chen F, et al. Weld defect segmentation and detection using semi-supervised transfer learning-based Generative Adversarial Networks[J]. Nondestructive Testing and Evaluation, 2025, 1-23. https://doi.org/10.1080/10589759.2025.2548374.

[7] Irfan M, Khan NA., Mushtaq Z, et al. A computationally efficient method for induction motor bearing fault detection based on parallel convolutions and semi-supervised GAN[J]. Nondestructive Testing and Evaluation, 2024, 40(10): 4717-4743.

[8] Yang W, Xiao Y C, Shen H K, et al. An effective data enhancement method of deep learning for small weld data defect identification[J]. Measurement, 2023, 206: 112245.

[9] Yuan Z, Gao X, Yang K, et al. Performance enhancement of ultrasonic weld defect detection network based on generative data[J]. Journal of Nondestructive Evaluation, 2024, 43(4): 102.

[10] Yu Z, Ma Q, Yuan H, et al. Defect identification method for ultrasonic inspection of pipeline welds based on cross-modal zero-shot learning[J]. Measurement Science and Technology, 2023, 35(2): 025009.

[11] Peng L, Li S, Sun H, et al. A pipe ultrasonic guided wave signal generation network suitable for data enhancement in deep learning: US-WGAN[J]. Energies, 2022, 15(18): 6695.

[12] Miao R, Gao Y, Ge L, et al. Online defect recognition of narrow overlap weld based on two-stage recognition model combining continuous wavelet transform and convolutional neural network[J]. Computers in Industry, 2019, 112: 103115.

[13] Chen R, Hu P, Gui X, et al. An on-line weld inspection method for underwater offshore structure based on an improved deep convolutional network[J]. Nondestructive Testing and Evaluation, 2025, 40(1): 289-308.

[14] Zhang Y, Gan F, Chen X. Motif difference field: an effective image-based time series classification and applications in machine malfunction detection[C]. 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), IEEE, 3079-3083.

[15] Zhang Z, Pan H, Wang X, et al. Deep learning empowered structural health monitoring and damage diagnostics for structures with weldment via decoding ultrasonic guided wave[J]. Sensors, 2022, 22(14): 5390.

[16] Zhang J, Wan H, Sun F, et al. Laser ultrasonic inspection of surface and subsurface defects in materials based on deep learning[J]. Nondestructive Testing and Evaluation, 2024, 40(10): 4744-4763.

[17] Pawar P, Buktar R. Detection and classification of defects in ultrasonic testing using deep learning[C] Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2021. Springer Singapore, 2022: 1-15.

[18] Munir N, Park J, Kim H J, et al. Performance enhancement of convolutional neural network for

ultrasonic flaw classification by adopting autoencoder[J]. NDT & E International, 2020, 111: 102218.

[19] Liu J, Zhao Z, Fu M, et al. Active small sample learning based the pipe weld defect detection method[J]. Chinese Journal of Scientific Instrument, 2022, 43(11): 252-261.

[20] Zhang R, Gao M, Fu L, et al. Weld defect detection based on adaptive fusion of multi-domain and multi-scale deep features[J]. Journal of Vibration and Shock, 2023, 42(17): 294-305+313.

[21] Zhang R, Zhao N, Fu L, et al. Recognizing defects in stainless steel welds based on multi-domain feature expression and self-optimization[J]. Journal of Intelligent Manufacturing, 2023, 34(3): 1293-1309.

[22] Pan H, Pang Z, Wang Y, et al. A new image recognition and classification method combining transfer learning algorithm and MobileNet model for welding defects[J]. IEEE Access, 2020, 8: 119951-119960.

[23] Zhang L, Pan H, Jia B, et al. Lightweight DCGAN and MobileNet based model for detecting X-ray welding defects under unbalanced samples[J]. Scientific Reports, 2025, 15(1): 6221.

[24] Ding K, Niu Z, Hui J, et al. A weld surface defect recognition method based on improved MobileNetV2 algorithm[J]. Mathematics, 2022, 10(19): 3678.

[25] Chen Y, Shen H, Zhang G, et al. Identification of weld defects from ultrasonic signals using GASF and an improved DCGAN-ResNet network[J]. Nondestructive Testing and Evaluation, 2025, 40(7): 2841-2867.

[26] Wang Z, Oates T. Imaging time-series to improve classification and imputation[J]. International Joint Conference on Artificial Intelligence, 2015, 3939-3945.

[27] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs[C]. International Conference on Machine Learning, PMLR, 2017: 2642-2651.

[28] Gu Z, Li W, Huo J, et al. LoFGAN: Fusing local representations for few-shot image generation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 8463-8471.

[29] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848-6856.

[30] Mirjalili S, Lewis A. The whale optimization algorithm[J]. Advances in engineering software, 2016, 95: 51-67.