Identification of weld defects from ultrasonic signals using GASF and an improved DCGAN-ResNet network

Yuan Chen[1,3,*], Hao Shen[2], Guangming Zhang[4], Ming Dong[2,3], Xiang Wan[2,3]

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

2College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi,710054, China

3Shaanxi Key Laboratory of Mine Electromechanical Equipment Intelligent Detection and Control, Xi'an University of Science and Technology, Xi'an, Shaanxi, 710054, China

4 General Engineering Research Institute, School of Engineering, Faculty of Engineering and Technology, Liverpool John Moores University, Byrom Street, L3 3AF, United Kingdom

**Email: chenyuan1030@126.com*

Abstract: In this paper, a method for weld defect identification from ultrasonic signals using the Gramian Angular Summation Field (GASF) and an improved deep convolutional generative adversarial network and residual network (DCGAN-ResNet) is proposed to overcome the problems of small-sample imbalance of echo signals as well as the low identification accuracy and poor efficiency of traditional convolutional neural networks (CNN). Firstly, the DCGAN model is improved based on Wasserstein distance and spectral normalization, and the augmented dataset is used to validate its effectiveness. Then, the residual block for the ResNet model is improved using group convolution to enhance the nonlinear representation of the network and reduce the number of parameters and computations. Finally, the squeeze-and-excitation (SE) attention mechanism is introduced for feature recalibration to enhance attention to important features and recognition efficiency. Experimental results show that the improved DCGAN-ResNet method outperforms other commonly used methods in terms of feature extraction, recognition accuracy and efficiency for weld defects, and its test accuracy reaches 91.99%, which is 14.36% higher than that before dataset augmentation. Thus, the proposed method is effective and feasible for weld defect recognition from ultrasonic signals under small-sample imbalance conditions, and can also be applied to other pattern recognition fields.

Keywords: ultrasonic signal; weld defect recognition; small-sample imbalance; GASF; DCGAN; ResNet

1 Introduction

Welding is one of the most important manufacturing technologies, which has been widely applied in various industrial fields such as aerospace, machinery manufacturing, bridge building, petrochemical, and so on. However, weld defects, including porosity, cracks, slag inclusion and lack of penetration, are inevitable in welded structures due to some factors such as welding materials, process parameters, operating techniques and working conditions. These defects not only affect the mechanical properties of welded joints, but can also lead to structural failure and, in serious cases, even to safety accidents. Therefore, weld inspection is particularly important to ensure the quality of welded structures. As an important non-destructive testing method, ultrasonic testing has some advantages of easy operation, high detection efficiency and precise positioning, making it become a one of the main methods for the quality assessment of welded structures. Thus, the defect identification techniques in ultrasonic testing of welded seams are of great significance for enhancing the safety and reliability of welded structures, thereby ensuring the safe operation of equipment and facilities [1] [2].

Ultrasonic A-scan has been widely used in the field of weld defect detection due to its advantages of real-time, portability and high sensitivity [3]. At present, weld defect identification using ultrasonic signals is mainly based on signal processing and pattern recognition techniques. Feature extraction from ultrasonic defect echo signals is the key to defect recognition. Various signal processing techniques, including short-time Fourier transform, empirical mode decomposition, local mean, variational mode decomposition, wavelet and wavelet packet transform, et al., are usually used to extract the time-frequency features from ultrasonic signals, which are then fed into the classifiers such as K-nearest neighbor discrimination, BP neural network, support vector machine, et al., to achieve automatic defect classification [4]. Zhou [5] et al. investigates the damage evolution of Q245R steel with different welding defects using acoustic emission and infrared thermography technology. The results indicated that the non-destructive testing technology can effectively reveal the effects of different welding defects on the damage evolution of Q245R steel. Xu et al. [6] used wavelet packet to extract the features of the defects from ultrasonic echo signals, and then evaluated and recognized the extracted features using separable criteria and BP neural network with an accuracy of 85.71%. HosseinAbadi et al. [7] applied wavelet packet transforms to extract the statistical features of the denoised ultrasonic signal by discrete wavelet transform, and used a multiclass support vector machine classifier to achieve the structural damage and estimate its severity. Yi et al. [8] proposed a feature extraction method based on wavelet packet algorithm, and applied support vector machine for quantitative classification on the ultrasonic echo signals generated by cracks. Shi et al. [9] used wavelet packet decomposition combined with deep neural network to identify via holes, inclined holes and flat-bottom holes. Experimental results showed that the defect classification accuracy was 21.66% higher than that of BP network. The above methods mainly use signal processing techniques to extract the features from ultrasonic echo signals, which are largely influenced by subjective factors. Moreover, different methods are adopted for different objects, resulting in the lack of a reliable theoretical basis for feature extraction. In addition, the traditional signal processing techniques are often difficult to accurately analyze and process the complex echo signals, leading to poor accuracy and generalization ability.

With the rapid application of artificial intelligence in the field of computer vision, deep learning has been widely used in object detection and image classification, which provides a new idea for achieving end-to-end defect recognition in ultrasonic testing, and can effectively improve the recognition accuracy and robustness [10]. Shi [11] et al. presented a welding defect detection model based on multiscale feature enhancement and aggregation. Experiments showed that the proposed model can effectively identify welding defects, with an improvement in mAP of about 1.2%, an increase in average recall of about 5.3%, a slight increase in parameter volume, and a significant decrease in Flops of 62.6% compared to the original model. Pan et al. [12] proposed a new transfer learning model based on MobileNet to achieve welding defect classification, and the experimental results indicated that the model achieved a prediction accuracy of 97.69%. Zhi et al. [13] established a deep learning fusion network model and pattern classification method based on waveform sequence features and image features, and the experiments showed that this method has a higher defect recognition rate than other methods. Hou et al. [14] proposed a deep convolutional neural network (DCNN) model to classify the weld defects from X-ray images, achieving an accuracy of 97.2%. Sonwane et al. [15] presented a multi-scale dense crossed block network (MDCBNet) to improve the detection and classification of weld defects, and the experimental results demonstrated its superiority in weld defect classification compared to existing methods.

The above methods achieved better effects on weld defect identification, however, most deep learning-based pattern recognition methods use images as input. Therefore, defect identification using A-scan ultrasonic signals should transform them into images. An image can be represented as matrices in a computer, and transforming a time series into an image is equivalent to transforming a time series into a matrix. There are mainly two methods for converting time series data into two-dimensional images. One involves time-frequency analysis techniques such as the short-time Fourier transform (STFT) [16] and wavelet transform [17], and so on. Another focuses on image coding, using techniques such as Markov transition field (MTF) [18], Recurrence Plots (RP) [19], Gramian angular summation field (GASF), Gramian angular difference field (GADF) [20], and so on. In this paper, GASF is used to merge the angular representation of the time series with the Gram matrix to create an image that captures the global features of the time series, taking advantage of image recognition in computer vision. Chen [21] et al. utilized recurrence and gradient maps to transform one-dimensional ultrasound signals into two-dimensional images, and then used an enhanced VGG-16 network for weld defect classification. The results showed that the VGG-16 network outperformed VGG-19, ResNet34, and ResNet50, achieving an impressive recognition rate of 97.5%. Xu et al. [22] converted the ultrasonic signals into the wavelet time-frequency images using continuous wavelet transform, which were used as input to train a CNN classification model to achieve automatic surface defect depth classification, with an average accuracy rate of 97.3% in the test. Ding et al. [23] proposed a ResNet model with wavelet transform to identify the depth of surface defects using surface acoustic wave signals obtained by the transient thermal grating (TTG) method, and achieved a classification accuracy of 96.21%. Miao et al. [24] proposed an integrated weld defect identification approach combining eddy current detection with 3D laser scanning based on CNN, and continuous wavelet transform was used to draw 2D time-frequency maps for eddy current signals as input to CNNs, which significantly improved the recognition efficiency while maintaining the recognition accuracy.

However, deep learning-based defect recognition methods require a large number of training samples. For small-sample imbalance data in practical applications, generative adversarial network (GAN) can effectively balance and augment the dataset [25]. Yang et al. [26] proposed an attention self-supervised learning auxiliary classifier generative adversarial network (ASSL-ACGAN) algorithm to expand the samples based on X-ray images of weld defects, and the experimental results showed that ASSL-ACGAN based data augmentation is superior to Wasserstein GAN (WGAN), WGAN gradient penalty (WGAN-GP) and ACGAN. Liu et al. [27] proposed a deep convolutional generative adversarial network (DCGAN) method to expand the small-sample fault dataset of rolling bearing. The experimental results showed that the test accuracy of the diagnosis model for the test set before and after the expansion of the small-sample dataset was increased from 91.67% to 98.96%. Deng et al. [28] adopted the improved DCGAN to expand the data samples, and an improved ResNet is constructed to train and test the balanced dataset after data augmentation. The experimental results of two different gearing fault datasets demonstrated that this method can generate high-quality samples and achieve a better diagnostic performance than other methods, showing great potential in imbalanced fault diagnosis.

The most common weld defect recognition methods are based on one-dimensional ultrasonic signals or X-ray weld defect images. Nevertheless, there are relatively few studies on deep learning-based weld defect recognition using ultrasonic A-scan signals, especially fewer studies on weld defect recognition under small-sample imbalance conditions. Thus, a weld defect recognition method combining GASF and improved DCGAN-ResNet is proposed in this paper. Firstly, the GASF is used to convert the acquired ultrasonic echo signals of weld defects into the images to form a dataset; then the dataset is augmented using the improved DCGAN; finally, the improved ResNet model is trained and tested using the augmented image dataset of weld defects to verify the effectiveness and feasibility of the proposed method.

2 Theoretical backgrounds

2.1 Convolutional neural network

CNN [29] is used to extract the local features from input data using convolutional layers and reduce spatial dimensionality using a pooling layer to learn complex patterns and hierarchical features. It typically consists of the following:

(1) The convolutional layer is usually used to extract the local features from the input data, which is expressed by:

$$C(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I\,(i+m, j+n) \cdot K(m,n) \tag{1}$$

where $C(\mathrm{I}, \mathrm{j})$ is the element in row i and column j of the output of the convolution operation, $I(i+m, j+n)$ is the element in row $i+m$ and column $j+n$ of the input matrix, $K(m,n)$ is the element in row m and column n of the convolution kernel, and M and N are the number of rows and columns of the input matrix and the convolution kernel, respectively.

(2) The activation layer is used to activate the features extracted by the convolutional layer. The commonly used activation function is rectified linear unit (ReLU), which maps the negative values to zero and retains only the positive values, allowing the network to better learn the features of the input data.

(3) Pooling layer is mainly used to reduce the spatial dimensionality (e.g. height and width) of the input data while preserving its important features. It allows the neural network to keep the transformations such as translation, rotation and scaling somewhat invariant and reduces the sensitivity of the model to small changes in position.

(4) The fully connected layer allows global synthesis of the features extracted from the previous layer in the neural network and maps these features to the final output space for data classification.

2.2 Deep convolutional generative adversarial network

DCGAN is an advanced variant of GAN [30], specifically designed to improve the quality and stability of the generated images by combining the advantages of CNN and GAN.

GAN is similar to the network model of the two-player game, which consists of a generative network G and a discriminative network D. The two networks reach the Nash equilibrium through continuous adversarial training, and the network D is generated to learn the distribution of real data. The schematic diagram of the GAN model is shown in Fig. 1.

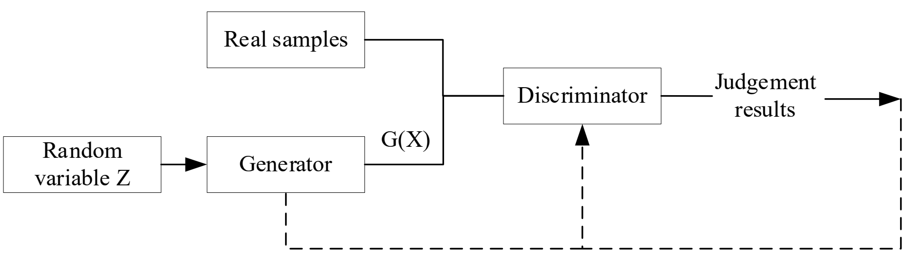


Fig. 1 Schematic diagram of the GAN model.

The GAN model is that the generator G generates data samples by inputting random noise, and then determines whether the samples are from real samples or not through the discriminator D until the discriminator is fooled, whose expression is:

$$min_G max_D V(D,G) = E_{X\sim P_{real}}[logD(x)] + E_{z\sim P_z(z)}\left[log\left(1 - D\big(G(Z)\big)\right)\right] \quad (2)$$

where $V(D,G)$ is the objective function, x is the real sample, whose sample distribution is $P_{real}(x)$, $G(Z)$ is the sample generated by the generator G from the input noise, and $D(x)$ is the output decision probability of the discriminator D, i.e. the probability of judging whether the generated sample is false and whether the real sample is true. Eq. (2) is essentially two optimization problems of the discriminator and the generator, which can be divided into optimization formulas for D and G:

$$max_D V(D,G) = E_{x\sim P_{real}}[logD(x)] + E_{z\sim P_z(z)}\left[log\left(1 - D\big(G(Z)\big)\right)\right] \quad (3)$$

$$min_G V(D,G) = E_{z\sim P_z(z)}\left[log\left(1 - D\big(G(Z)\big)\right)\right] \quad (4)$$

By using convolution and transposed convolution operations, DCGAN effectively preserves spatial structural information in images. The incorporation of batch normalization in the generator and discriminator helps to stabilize the training process and accelerate model convergence, which enables DCGAN to generate high-resolution and high-quality images with a simple architecture that is easily scalable.

DCGAN can learn the distribution of real samples to generate virtual samples to solve the problem of small-sample imbalance for the dataset.

2.3 Residual network model

ResNet [31] is used to solve the problems of gradient vanishing and gradient exploding in deep network training by introducing skip connections, thus improving the training speed and accuracy of the network. The core idea of residual structure is to take the difference between the input and output of the network (i.e. the residuals) as the learning objective, and to better learn the features of the data by learning the residuals, which is expressed as:

$$H(x) = F(x) + x \quad (5)$$

where $H(x)$ is the output of the residual block, and $F(x)$ is the residual function and x is the input.

2.4 Channel attention mechanism

The attention mechanism described in Ref [32], known as SE, first performs dimensionality reduction for the input features through global average pooling (GAP) operation, and compresses and activates the features using two fully connected layers to obtain the interactions between the channels. Finally, the obtained attentional weights are rescaled between 0 and 1 using a sigmoid activation function, which is combined with the original input features to generate a weighted feature representation, as shown in Fig. 2. The inclusion of the SE module in the network improves channel characteristics and model performance.

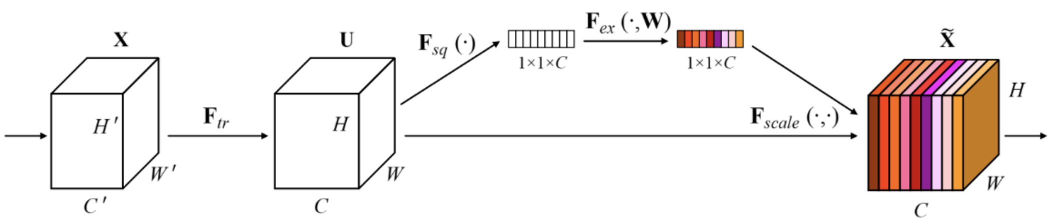


Fig. 2 Schematic diagram of the SE attention mechanism.

3 Dataset preparation

3.1 Weld specimens and signal acquisition

5 weld specimens of 45# carbon steel were designed and fabricated, as shown in Fig. 3, including 2 plate-plate butt, 1 tube-plate butt and 2 tube-tube butt specimens. Among 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; and the tube-plate butt specimen contains two cracks and one each of the other three types of defects, and both the tube-tube butt specimens I and II contain one each of the four types of defects. Fig. 4 gives the dimensions of the weld specimens and the location of the defects.

(a) Plate-plate butt specimen I

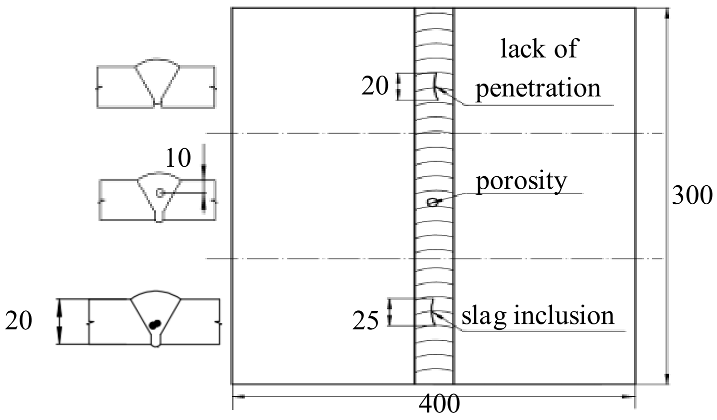
(b) Plate-plate butt specimen II

(c) Tube-plate butt specimen

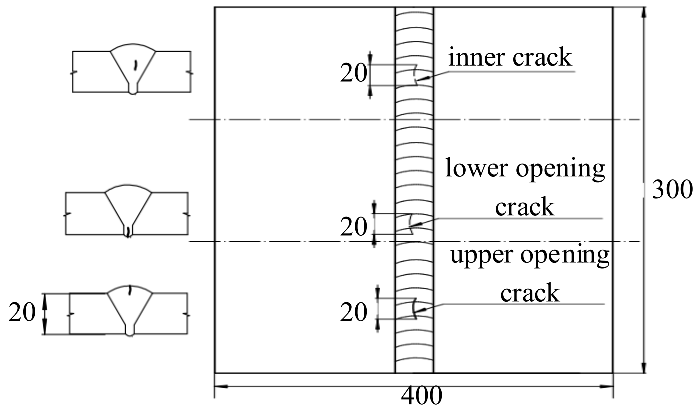
(d) Tube-tube butt specimen I

(e) Tube-tube butt specimen II

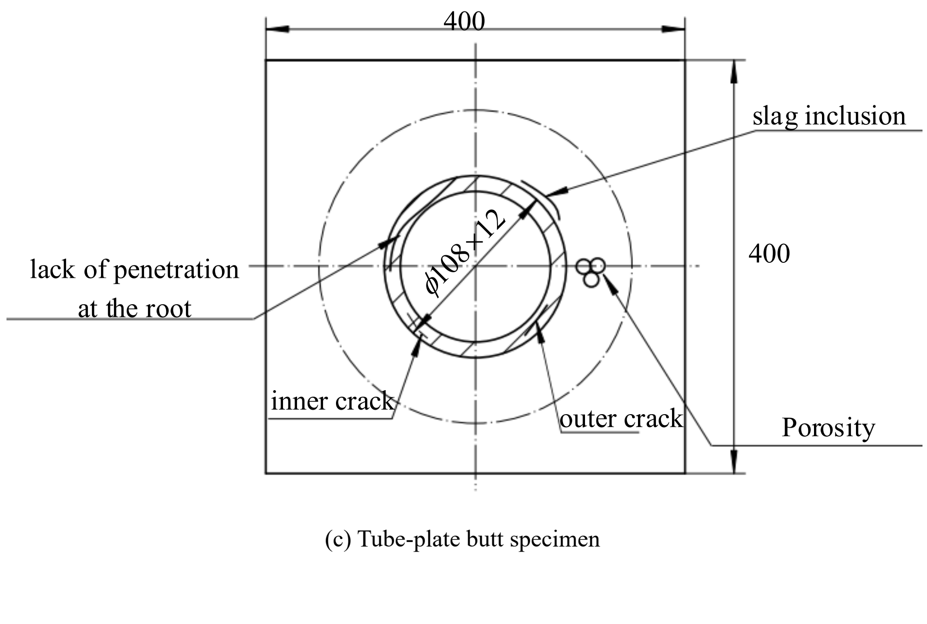
Fig. 3 Weld specimens.



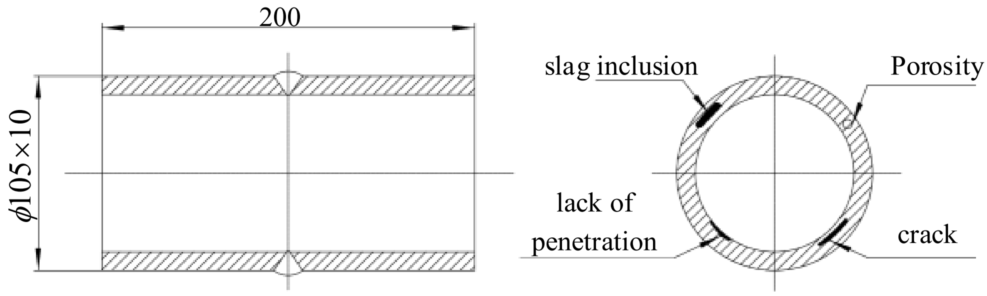
(a) Plate-plate butt specimen I



(b) Plate-plate butt specimen II



(c) Tube-plate butt specimen



(d) Tube-tube butt specimens I and II

Fig. 4 Weld specimen dimensions and defect location.

In the experiments, the A-scan ultrasonic echo signals of defects in the weld specimens were collected using a CTS-4020 ultrasonic flaw detector and five types of angle probes as shown in Fig. 5. The angle probes are 2.5Z10×10A45, 5Z8×13A45, 2.5P13×13A60, 2.5Z10×10A70 and 5Z10×10A70, respectively, as shown in Fig. 6. By changing the detection position and direction, and adjusting the gain of the digital ultrasonic flaw detector, the above weld specimens were detected. A total of 2100 ultrasonic echo signals were collected, as shown in Table 1. For the specimens shown in Fig. 3(a), (b), (d) and (e), the signal acquisition was performed on both sides of the weld, i.e. 2 positions, whereas for the specimen shown in Fig. 3(c), the probes were solely placed on the steel plate for acquisition, i.e. 1 position. The acquisition was performed on 3 directions at each position, i.e. the probe was oriented in the positive direction and offset to the left and right at a certain angle. The acquisition distance was 2, i.e. each before and after. Two different gains were used for each acquisition. In addition, 30 ultrasonic signals were randomly acquired for each type of defect. The total number of ultrasonic echo signals collected is as follows: the number of cracks is $(5\times2\times3\times2\times2+2\times1\times3\times2\times2)\times5+30=750$, and the number of slag inclusion, porosity and lack of penetration are $(3\times2\times3\times2\times2+1\times1\times3\times2\times2)+30=450$, respectively.

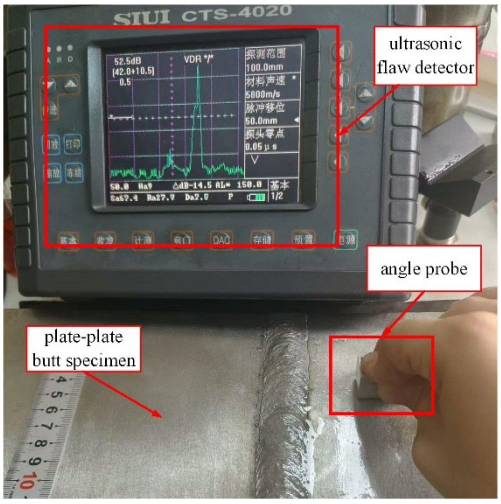


Fig. 5 Schematic diagram of ultrasonic signal acquisition for weld defects.

Fig. 6 Ultrasonic angle probes.

Table 1 Small-sample imbalance dataset.

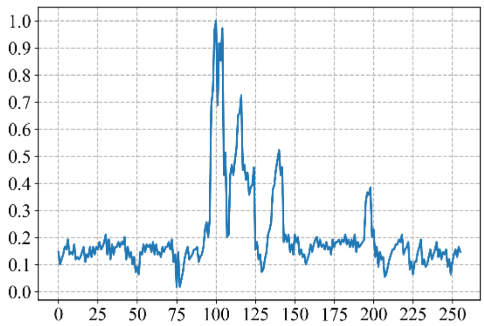
Type of defect	Slag inclusion	crack	porosity	lack of penetration
Number of signals	450	750	450	450

3.2 Weld defect dataset preparation

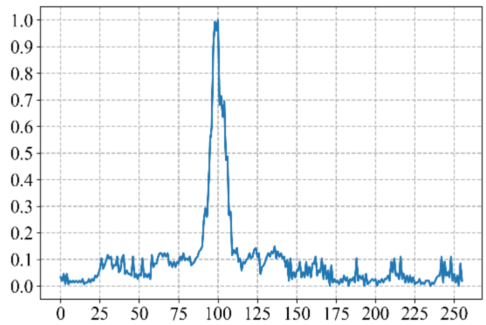
3.2.1 Ultrasonic signal pre-processing

The acquired A-scan ultrasonic echo signals are intercepted and only the echo signals of defects with a length of 256 sampling points are retained. In addition, to eliminate the influence of the signal amplitude on the accuracy of defect recognition, the intercepted ultrasonic echo signals are normalized by Eq. (6). Fig. 7 displays the representative ultrasonic signals for the four types of weld defect after the above pre-processing.

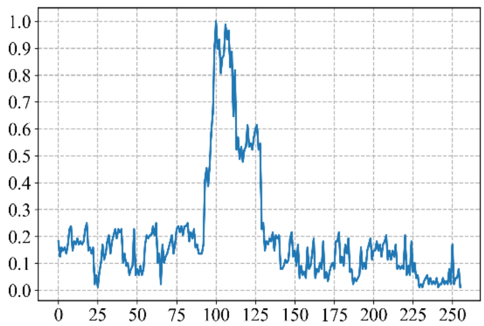
$$X_i = \frac{x_i - min(x)}{max(x) - min(x)} \tag{6}$$

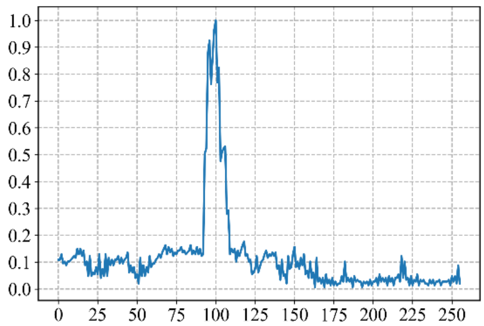


(a) slag inclusion



(b) crack





(c) porosity (d) lack of penetration

Fig. 7 Representative ultrasonic signals for four types of weld defects.

3.2.2 Preparation of weld defect image dataset using GASF

Gramian angular field (GAF) mainly includes Gramian angular summation field (GASF) and Gramian angular difference field (GADF), which maps one-dimensional time series data into a two-dimensional polar coordinate system and generates a matrix that can reflect the time dependence and periodicity of the time series data through the trigonometric functions. The specific steps are as follows:

(1) The sequence is aggregated by segmented aggregation approximation to reduce the size and normalize the one-dimensional data using Eq. (7).

$$\widetilde{X_i} = \frac{X_i - min(X)}{max(X) - min(X)} \times 2 - 1 \quad (7)$$

(2) The normalized data is then encoded as an angular cosine, with the time stamp serving as the radius of the scaled data $\tilde{X}$, which can be represented in polar coordinates as:

$$\begin{cases} \phi_i = arccos(\widetilde{X_i}), 0 \le \widetilde{X_i} \le 1, \widetilde{X_i} \in \tilde{X} \\ r_i = \frac{t_i}{N}, t_i \in N \end{cases} \quad (8)$$

where t_i is the time stamp, N is a constant factor to regularize the span of the polar coordinate system, and ϕ_i is the radians of the scaled data X_i in polar coordinates, i.e. the angular cosine.

(3) In this paper, we use the GASF based on the cosine function of the sum of two angles, whose expression is:

$$G_s = \begin{pmatrix} \cos\phi_1 \\ \vdots \\ \cos\phi_i \\ \vdots \\ \cos\phi_{256} \end{pmatrix} (\cos\phi_1 \cdots \cos\phi_i \cdots \cos\phi_{256}) - \begin{pmatrix} \sin\phi_1 \\ \vdots \\ \sin\phi_i \\ \vdots \\ \sin\phi_{256} \end{pmatrix} (\sin\phi_1 \cdots \sin\phi_i \cdots \sin\phi_{256})$$

$$= \begin{pmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_{256}) \\ \vdots & \ddots & \vdots \\ \cos(\phi_{256} + \phi_1) & \cdots & \cos(\phi_{256} + \phi_{256}) \end{pmatrix} \quad (9)$$

The process of converting the A-scan ultrasonic signal into a two-dimensional time map using GASF is shown in Fig. 8. The 2100 A-scan ultrasonic echo signals of weld defects after the pre-processing above are processed using GASF to generate 2100 images of 64×64 pixels, forming the weld defect dataset.

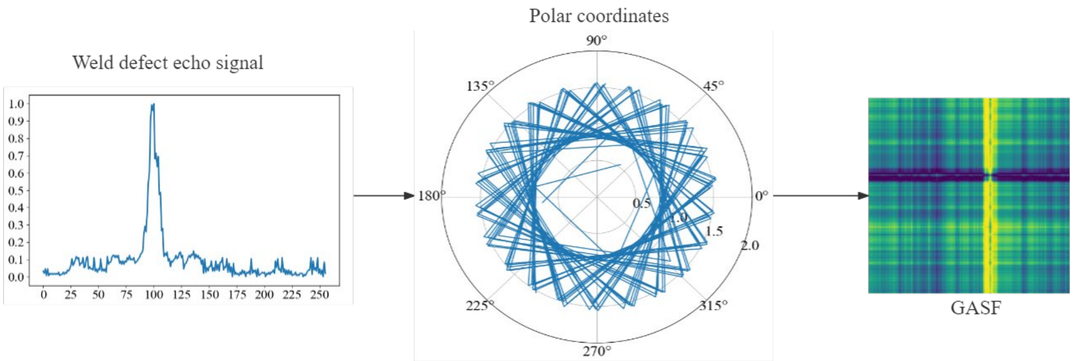


Fig. 8 Flowchart of conversion of the ultrasonic signal into an image using GASF.

4 Improved DCGAN-ResNet network model for weld defect recognition

4.1 Improved DCGAN based on Wasserstein distance and spectral normalization

Compared to other GAN algorithms, DCGAN has significant advantages in terms of training stability and the quality of generated samples, but it still suffers from some problems such as model collapse and gradient vanishing. Therefore, in this paper, DCGAN is improved as follows:

(1) Conventional GANs typically use the Jensen-Shannon (JS) divergence to quantify the difference between the generated distribution and the real data distribution. However, the JS divergence proves to be ineffective in cases where there is no overlap between the two distributions, which can lead to the gradient vanishing problem in the early stages of training, making it difficult for the generator to learn. In contrast, Wasserstein distance [33] measures the distance between two distributions by evaluating the minimal cost required to transform one distribution into the other. It provides a smoother alternative to JS divergence and does not suffer from the gradient vanishing problem when dealing with non-overlapping distributions. Thus, the Wasserstein distance is introduced as the training loss function to measure the difference in distribution between the generated samples and the real samples, which is expressed by:

$$W\left(P_r, P_g\right) = inf_{\gamma \in \prod(P_r, P_g)} E_{(x,y)\sim\gamma}[\| x - y \|] \tag{10}$$

where P_g is the distribution of the generated samples, $\prod\left(P_r, P_g\right)$ is the set of all possible joint distribution probabilities of P_r and P_g, y is the generated sample, and $\|x - y\|$ is the distance between them.

(2) Spectral Normalization [34] is a regularization technique used to constrain the spectral norm (i.e., the maximum singular value of a matrix) of weight matrices in neural network layers. By constraining the spectral norm of each layer's weight matrix, Spectral Normalization indirectly ensures that the entire network has a Lipschitz constant. This constraint is particularly beneficial for computing Wasserstein distance, as WGAN requires the discriminator to be a 1-Lipschitz function. Spectral Normalization achieves this by normalizing the weight matrix of each layer:

$$W_{SN} = \frac{W}{\sigma(W)} \tag{11}$$

where σ(W) is the maximum singular value of the weight matrix.

The Wasserstein distance makes the training of generators and discriminators more stable by providing smoother and more stable gradient signals. It better guides the generator to adapt the generated distribution to the real distribution during training. And spectral normalization can effectively avoid gradient explosion and gradient vanishing by limiting the number of spectral paradigms in the network layer, ensuring stability of training and the improvement of model performance. The combination of the two makes the improved GAN better able to generate high-quality images and show better stability and reliability during training.

In this paper, the above improved DCGAN model is used to perform sample balance and augmentation on the weld defect dataset to eliminate the influence of the small-sample imbalance problem on the defect recognition. The model architecture is illustrated in Fig. 9. The generator in the model consists of four transposed convolutional layers that generate a 64×64×3 color GASF image. The real and generated GASF images are both input into the discriminator, which includes two convolutional layers and a fully connected layer to evaluate the authenticity of the image. When the generator and discriminator are in equilibrium, a GASF image is generated that is very similar to the real image. And the generated GASF images for the weld defects and the real GASF images are shown in Fig. 10.

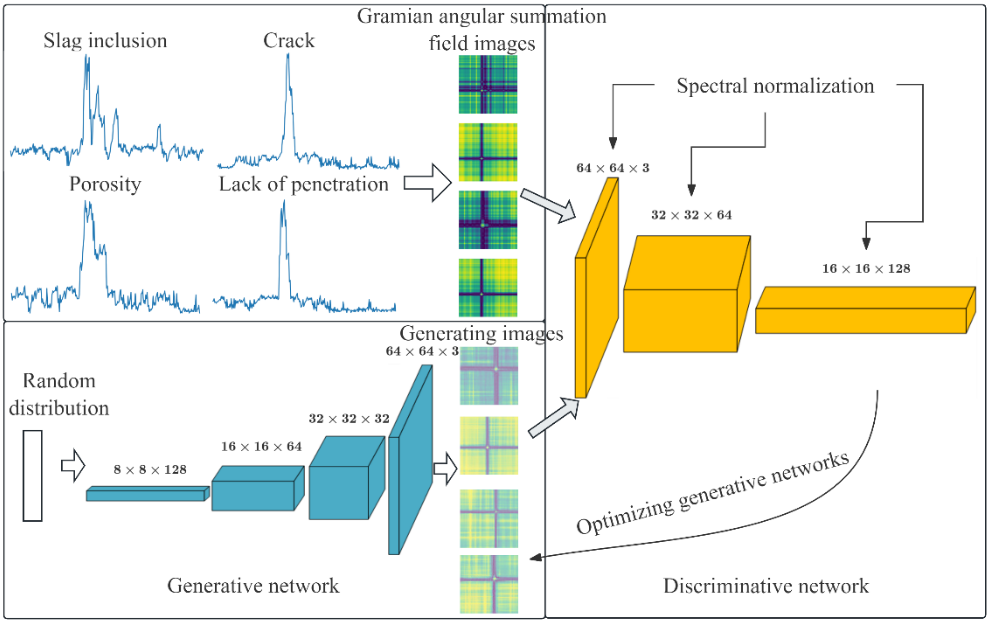
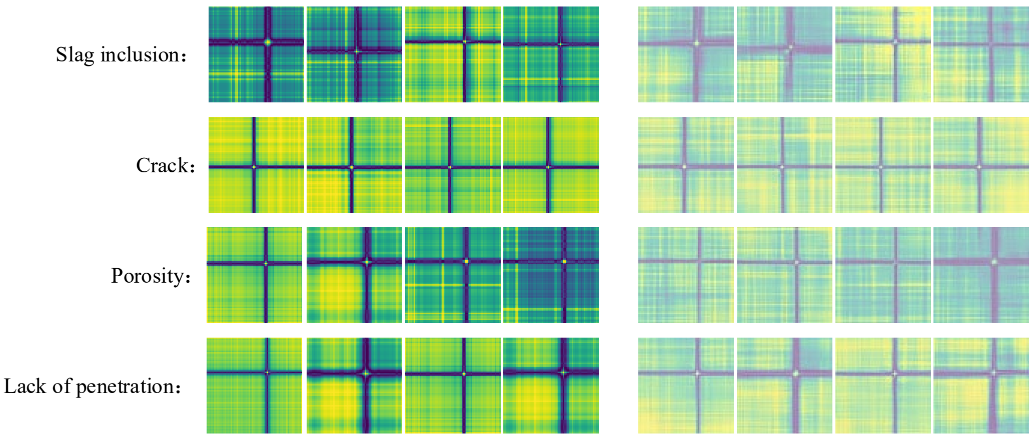


Fig. 9 Schematic diagram of the improved DCGAN model structure.



(a) Real image (b) Generate image

Fig. 10 The real GASF images for weld defects and the images generated by the improved DCGAN.

4.2 Improved ResNet based on group convolution and SE block

The ResNet network model presented in this paper consists of a 7×7 standard convolution layer, a Max pooling layer, a global average pooling layer, a fully connected layer, a Softmax classifier, four standard residual blocks, and four improved residual blocks, whose architecture is shown in Fig. 11.

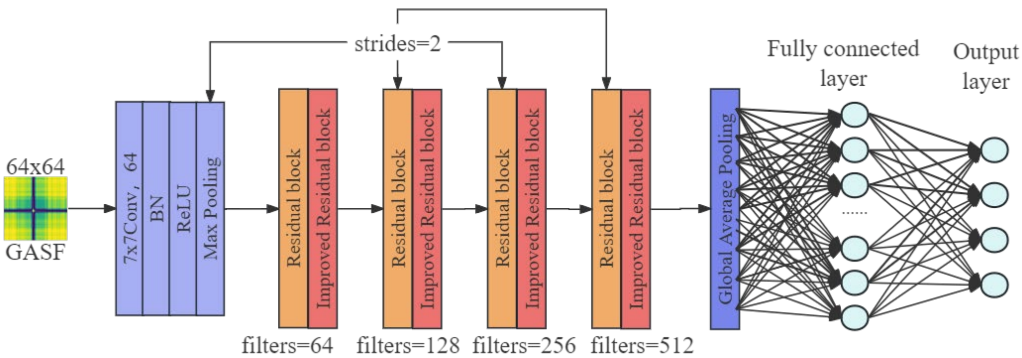


Fig. 11 Schematic diagram of the improved ResNet model structure.

(1) A 7x7 convolutional layer is used to extract the initial features, and the nonlinear transformation is performed using batch normalization (BN) and the ReLU activation functions. The BN layer can speed up the convergence process of the neural network and enhance its stability, which helps to reduce the problems of vanishing or exploding gradients. Meanwhile, the ReLU activation function can help introduce nonlinearity and sparsity into the initial features, mitigating problems such as overfitting and gradient vanishing. These results in more meaningful and diverse feature representations for subsequent layers.

(2) Down-sampling via the Max Pooling layer allows for feature dimensionality reduction, translation invariance and feature selection, thereby enhancing the computational efficiency and generalization capability of the model.

(3) The standard residual block contains two 3×3 convolutional layers with the same number of output channels, each followed by a BN layer and a ReLU activation function. A skip connection is directly added before the final ReLU activation function, which introduces a 1×1 convolutional layer to transform the shape and change the number of channels, as shown in Fig. 12.

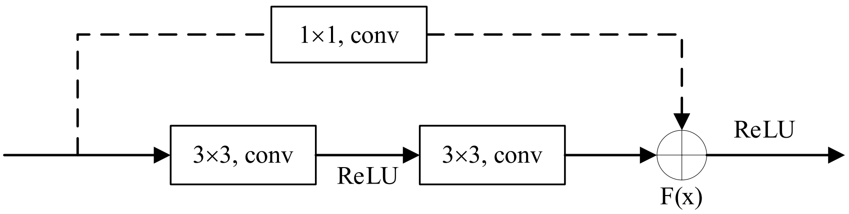


Fig. 12 Standard residual block.

(4) The group convolution and the SE attention mechanism are used to improve the standard residual block. Firstly, in the residual block, the input features are separately convolved into feature maps with different numbers of channels. Group convolution is then utilized for feature extraction to reduce the number of parameters and computational complexity. Finally, the SE attention mechanism is introduced to selectively adjust the importance of each channel to improve the feature representation. The structure of the improved residual block is shown in Fig. 13.

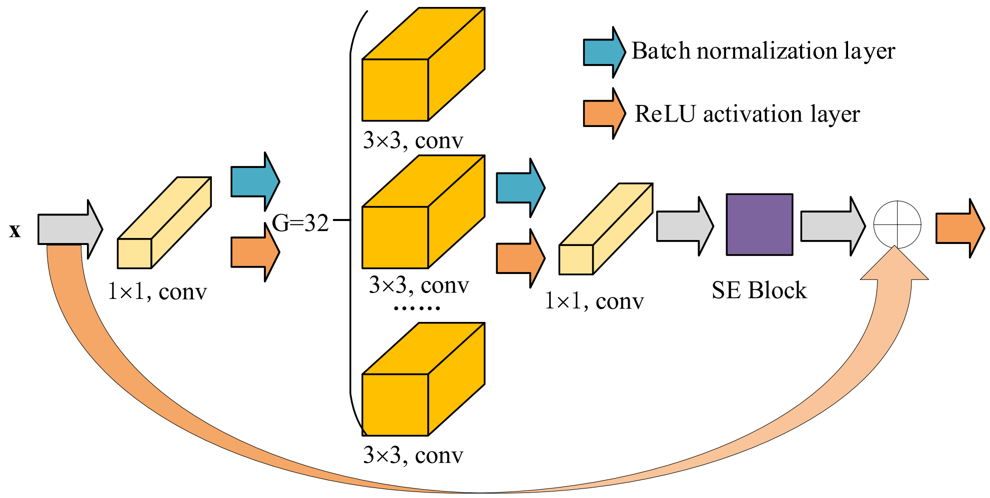


Fig. 13 Schematic diagram of the improved residual block structure.

(5) The global average pooling layer performs a pooling operation on the feature maps by calculating the average value of each feature map, which converts the entire feature map to an average value and can reduce the number of parameters in the model. In addition, the fully connected layer performs the feature combination and transformation by learning the weights and bias

parameters to generate the final feature representation. The Softmax classifier converts the output of the fully connected layer into a probability distribution, making the sum of the probabilities for all categories 1. Classification prediction can be made by determining the probability that the input belongs to each category based on the predicted probabilities.

5 Experimental results and discussion

All experiments in this paper are completed using the TensorFlow 2.6 framework on a Windows 10 operating system with a Python programming environment. Network training and testing are performed on a computer with an Intel Xeon E5-2660 CPU, 16GB RAM memory and an NVIDIA GeForce RTX 2060 GPU.

5.1 Sample balance and augmentation

DCGAN and the improved DCGAN are employed to augment the above dataset of weld defect images, and the quality of the generated images is assessed and analyzed. The network structure parameters of the generators and discriminators of DCGAN and improved DCGAN are identical, as shown in Table 2.

Table 2 Structure parameters of the DCGAN and improved DCGAN model.

	name	Input	Output	Activation	kernel	BN	Dropout
G	Deconv1	100	[8,8,128]	LeakyReLU	5×5	√	—
	Deconv2	[8,8,128]	[16,16,64]	LeakyReLU	5×5	√	—
	Deconv 3	[16,16,64]	[32,32,32]	LeakyReLU	5×5	√	—
	Deconv 4	[32,32,32]	[64,64,3]	Tanh	5×5	√	—
D	Conv1	[64,64,3]	[32,32,64]	LeakyReLU	5×5	—	0.3
	Conv 2	[32,32,64]	[16,16,128]	LeakyReLU	5×5	—	0,3
	FC 3	[16,16,128]	1	—	5×5	—	—

The above two GANs are trained with 2000 cycles, a learning rate of 0.0005 and the RMSprop optimizer. 100 GASF images are generated for each type of defects. The quality of the generated images is quantitatively assessed by two metrics such as Fréchet inception distance (FID) and structural similarity index (SSIM). FID is a metric for quantifying the realism and diversity of images generated by GANs, and the smaller the FID value, the better the diversity and quality of the image, which is calculated as follows:

$$d^2\big((m, C), (m_\omega, C_\omega)\big) = \|m - m_\omega\|^2 + Tr\big(C + C_\omega - 2\sqrt{C \times C_\omega}\big) \tag{12}$$

where m and m_ω represent the feature means of the real and generated images respectively, C and C_ω denote the covariance matrices of the real and generated feature vectors respectively, $\|m - m_\omega\|^2$ refers to the sum of the squared differences between the two mean vectors, and Tr represents the sum of the elements on the main diagonal of the square matrix. Table 3 gives the FID calculation results of the generated samples.

Table 3 FID calculation results of the generated samples.

GANs	Slag inclusion	Crack	Porosity	Lack of penetration
DCGAN	417.05	389.98	484.25	464.59
Improved DCGAN	305.77	225.81	268.15	303.61

SSIM is used as a metric to measure the similarity between two given images, which is more in line with the human visual perception of image quality. The higher the SSIM value, indicating the more the detailed information in the image, and the better the image quality. The SSIM is calculated by:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{13}$$

where x and y denote 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, and σ_{xy} is the covariance of x and y. In addition, $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ are the constants used to maintain image stability, and L is the dynamic range of the pixel values, where $k_1 = 0.01$, and $k_2 = 0.03$.

The SSIM is calculated for each generated image with its corresponding original image of each class, and the real image that is most similar to each generated image is used to calculate the SSIM (i.e., the one with the largest SSIM value), and then divided by the number of generated images to obtain the average value, which can be expressed by:

$$SSIM_{Average} = \frac{1}{N}\sum_{i=1}^{N}[\max_j(SSIM(G_i, R_j)) \tag{14}$$

where G_i is the ith generated image, R_j is the jth real image, N is the total number of generated images, and $\max_j(SSIM(G_i, R_i))$ denotes the largest of all SSIMs computed for each generated image G_i with all real images R_j. Table 4 shows the SSIM calculation results of the generated samples.

Table 4 SSIM calculation results of the generated samples.

GAN	Slag inclusion	Crack	Porosity	Lack of penetration
DCGAN	0.3872	0.4338	0.3659	0.4330
Improved DCGAN	0.4975	0.5508	0.5043	0.5419

As can be seen from Tables 3 and 4, the quality of the images generated by the improved DCGAN is better than that of DCGAN. However, both FID and SSIM metrics have their own limitations. FID utilizes the inception network model to extract the sample features, which can be influenced by the model performance and the number of samples, resulting in image distortion. In addition, SSIM only considers image structure, cannot capture the advanced features of images, and is sensitive to brightness and contrast. Therefore, it is more objective to use them as relative metrics to evaluate the performance of GANs.

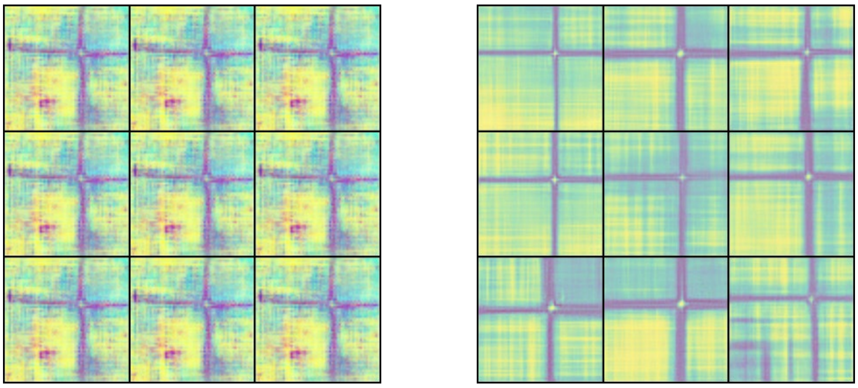
To better evaluate the effect of different image transformations on image similarity and to compare them with the generated samples, we selected one image from each class of real GASF images and applied the following process: (1) addition of Gaussian noise with a variance of 0.01; (2) application of Gaussian blur with a radius of 5; (3) reduction of image contrast by 50%. Then, we compared the above processed images with the corresponding original images by calculating the SSIM, and the results are presented in Table 5.

Table 5 Effects of image transformations on GASF image similarity

	Slag inclusion	Crack	Porosity	Lack of penetration
Gaussian noise	0.5471	0.4937	0.5286	0.4855
Gaussian blurring	0.4058	0.4928	0.4318	0.4996
Reduced contrast	0.8403	0.8351	0.8417	0.8362

As can be seen from Table 5, the SSIM values for the Gaussian noise transformation range from 0.49 to 0.55, for the Gaussian blur transformation from 0.4 to 0.5, and for the contrast reduction transformation from 0.8 to 0.85. As shown in Table 4, the SSIM values for the improved DCGAN-generated samples range from 0.5 to 0.6. A comparison with the SSIM values of the generated samples shows that the SSIM values of the generated samples, ranging from 0.5 to 0.6, are slightly higher than those of the images with Gaussian noise transformation but lower than those of the images with contrast reduction transformation. This indicates that the generated samples can maintain the structural similarity of the images to a certain extent, but still do not reach the similarity level of the images with contrast reduction transformation.

Human visual perception serves as an intuitive assessment of the quality of the generated samples. The images generated by the two models above are shown in Fig. 14 for an example of lack of penetration. It can be evident from Fig. 14 that DCGAN produces the images of poor quality and model collapse, while the improved DCGAN produces the images of significantly better quality.



(a) DCGAN (b) Improved DCGAN

Fig. 14 The generated GASF images of lack of penetration.

The weld defect dataset obtained above contains 2100 GASF images, i.e. the images for crack are 750 and the images for each of slag inclusion, porosity and lack of penetration are 450. The improved DCGAN is employed to perform sample balance and augmentation on the dataset with different ratios. The generated samples together with the real ones form a new dataset, which is then split into training, validation and test sets in the ratio of 8:1:1. The improved ResNet model proposed in this paper is utilized for defect recognition to assess the performance of the improved DCGAN, and the experimental results are presented in Table 6. It can be observed that the defect recognition accuracy is significantly enhanced after sample balance and augmentation. However, it is noted that when the number of samples in the dataset reaches a certain threshold, the recognition accuracy decreases instead.

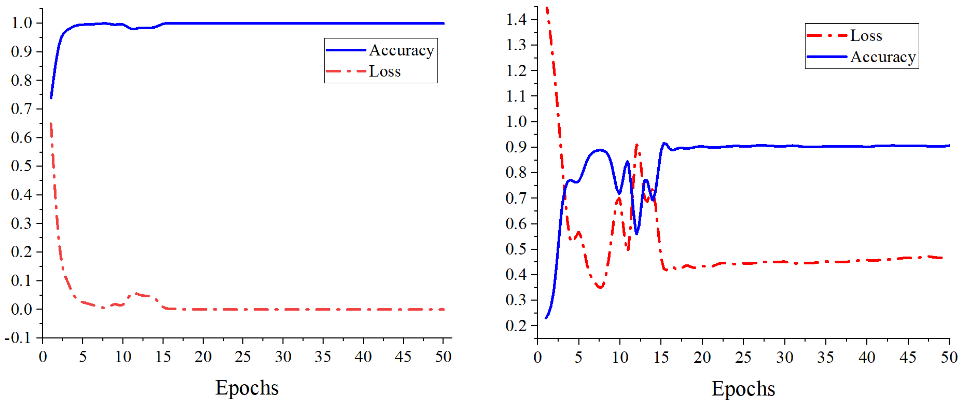
Table 6 Comparison of experimental results of defect recognition for sample augmentation.

Dataset No.	Slag inclusion	Crack	Porosity	Lack of penetration	Total	Accuracy	Increase
1	450	750	450	450	2100	77.63%	—
2	800	800	800	800	3200	86.87%	9.24%
3	1024	1024	1024	1024	4096	88.97%	2.1%
4	1280	1280	1280	1280	5120	91.99%	3.02%
5	1536	1536	1536	1536	6144	91.33%	-0.66%

5.2 Weld defect recognition using improved ResNet

The above dataset 4 is divided into a training set, a validation set and a test set in a ratio of 8:1:1 to train and test the improved ResNet network. The network training parameters are set as follows: the number of input samples per batch is 64, the learning rate is 0.001, and the number of iterations is 50. The Adam optimizer and cross-entropy loss function are used during training.

After the model training is completed, the accuracy of the network for the training and validation sets, as well as the change curve of the cross-entropy loss, are obtained, as shown in Fig. 15. It can be seen that the improved ResNet model converges very quickly and tends to stabilize after 15 epochs of training. Furthermore, the loss value remains consistently low, effectively preventing overfitting.



(a) Training set (b) Validation set

Fig. 15 Accuracy and cross-entropy loss curves for the training and validation sets.

To assess the generalization performance of the trained network, it is tested on the test set, and the test results are visually evaluated using a confusion matrix, as shown in Fig. 16.

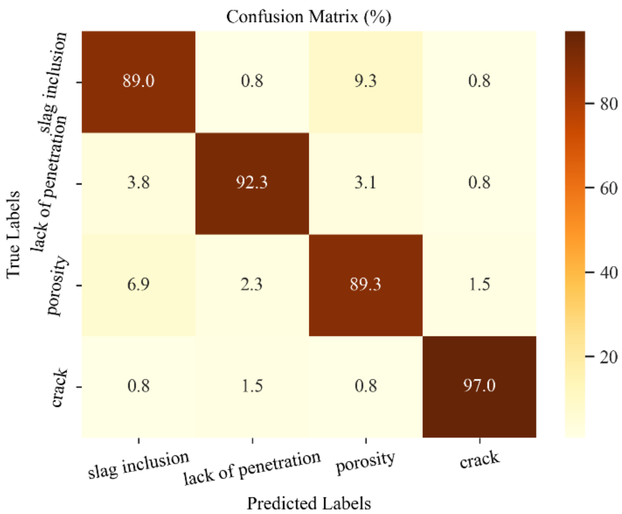
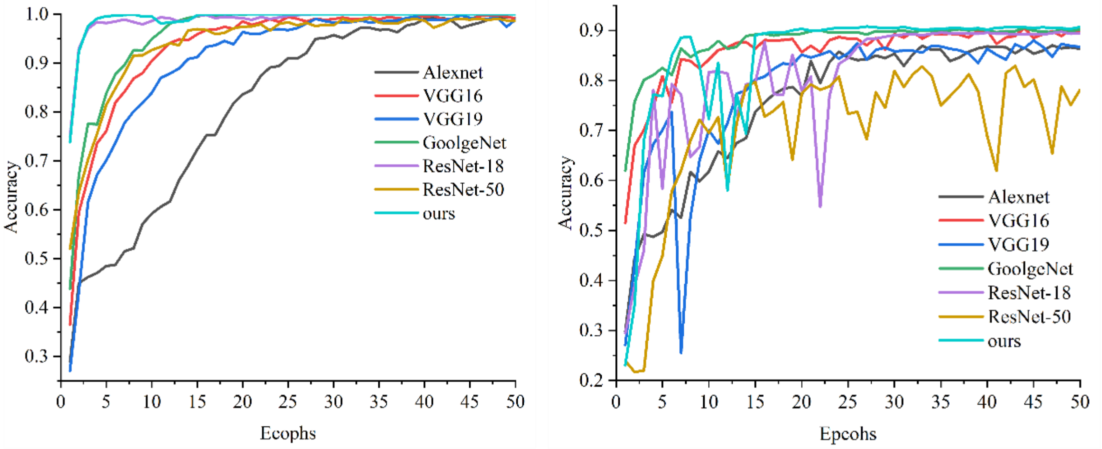


Fig. 16 Visualization of the test results.

It can be observed from Fig. 15 that the recognition accuracies for lack of penetration, slag inclusion, porosity and crack are 92.3%, 89.0%, 89.3% and 97.0% respectively. The average recognition accuracy for four types of defects is 91.9%. Therefore, the improved ResNet network achieves a good recognition effect of weld defect.

To further validate the proposed method, the above dataset 4 is used to train and test the network models such as the improved ResNet, ResNet-18, AlexNet, VGG-16, VGG-19, GoogleNet and ResNet-50. ResNet-18 and ResNet-50 use the standard residual blocks without the SE attention mechanism and group convolution, while AlexNet consists of 5 convolutional layers and 3 fully connected layers. VGG-19 consists of 16 convolutional layers and a final 3 fully connected layers, while VGG-16 consists of 13 convolutional layers and 3 fully connected layers. GoogleNet improves computational efficiency and model performance by introducing the Inception module. The training and validation results of the seven network models are shown in Fig. 17, and their test results are summarized in Table 7.



(a) Training set (b) Validation set

Fig. 17 Experimental results of the training and validation sets using different network models.

Table 7 Experimental results of the test set using different network models.

Network models	Accuracy (%)	Training time (s)	number of parameters (10^6)
AlexNet	86.32	271.98	23.17
VGG-16	90.04	307.93	39.90
VGG-19	87.50	320.06	45.21
ResNet-18	90.04	182.52	11.20
ResNet-50	78.52	298.85	22.47
GoolgeNet	90.14	162.29	5.87
Improved ResNet	91.99%	182.00	5.43

It can be observed from Fig. 17 that the improved ResNet and ResNet-18 have faster convergence rates than AlexNet, VGG-16, VGG-19, GoolgeNet and ResNet-50, indicating better fitting and feature extraction capabilities. In addition, the improved ResNet has faster convergence and greater stability than ResNet-18. We found that VGG-19 and ResNet-50 have large differences in the training and validation sets due to the deeper and more complex network structure, indicating severe overfitting. The proposed method has the least overfitting due to the simplified model complexity. As shown in Table 7, the improved ResNet achieves higher recognition accuracy and the least number of model parameters compared to the other six network models. Therefore, the weld defect recognition method based on the improved ResNet network proposed in this paper is effective and feasible.

5.3 Hyperparameter optimization and cross-validation

Hyperparameter optimization refers to the systematic adjustment of the hyperparameters of a machine learning model, such as learning rate (LR), batch size (BS), optimizer, etc., during the training process to find the best combination of hyperparameters and thereby improve the performance of the model. In this paper, we systematically optimize the hyperparameters of the model to improve its performance and generalization ability. Specifically, the hyperparameter search space is defined as follows: batch size [32, 64], learning rate [0.001, 0.0001], and optimizer [Adam, SGD]. We trained the model with each combination of hyperparameters and determined the optimal settings by comparing the model performance under different combinations. Dataset 4 was split into the training and test sets at a ratio of 4:1, and the results of the training are shown in Table 8.

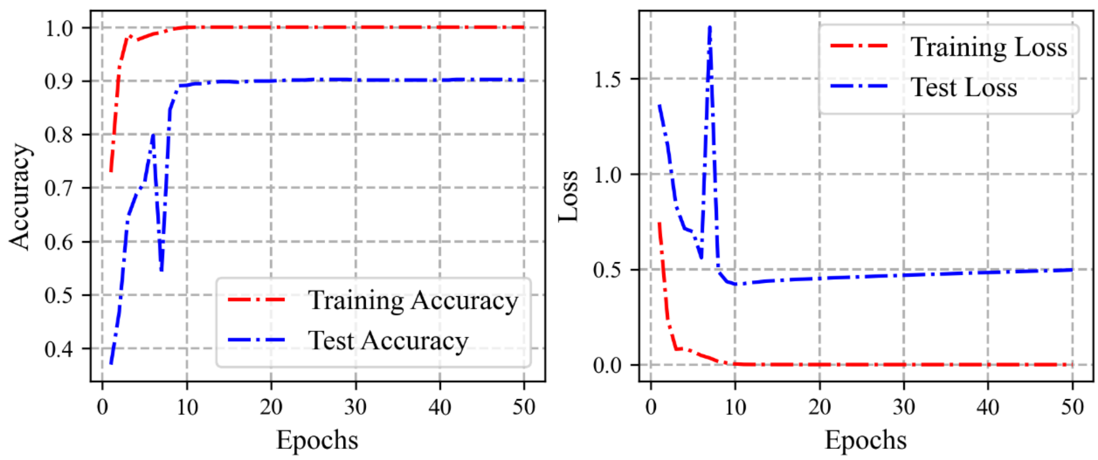
Table 8 Accuracy and training time for different hyperparameter combinations.

BS	32	32	32	32	64	64	64	64
LR	0.001	0.001	0.0001	0.0001	0.001	0.0001	0.001	0.0001
optimizer	Adam	SGD	Adam	SGD	Adam	SGD	Adam	SGD
Accuracy (%)	86.91	88.77	73.54	83.40	90.92	86.43	88.96	80.86
Train time(s)	250.97	241.64	252.99	243.86	171.05	162.43	166.61	164.37

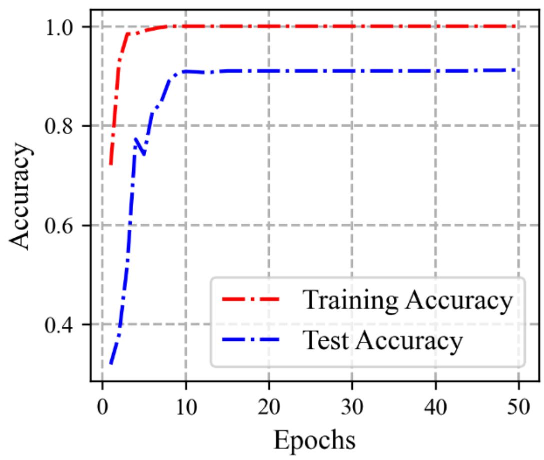
It can be seen from Table 8 that the total training time for batch size 64 is consistently less than batch size 32. In addition, the Adam optimizer outperforms the SGD optimizer for the dataset 4. Of the eight different hyperparameter combinations, the combination of a batch size of 64, a learning rate of 0.001 and the Adam optimizer achieves the best performance and the highest accuracy on the validation set, indicating that this hyperparameter combination is optimal in this study.

Cross-validation is a technique for evaluating the performance of machine learning models by repeatedly splitting the training dataset into training and the validation subsets to obtain a more stable estimate of model performance. Its main purpose is to reduce the risk of overfitting and underfitting of the model, and to provide a more reliable estimate of how the model will perform on new data. In this paper, 5-fold cross-validation is used, and the dataset 4 is divided into 5 parts, namely Fold1, Fold2, Fold3, Fold4, and Fold5. Each fold is used as the test set, while the remaining folds are used in turn as the training set.

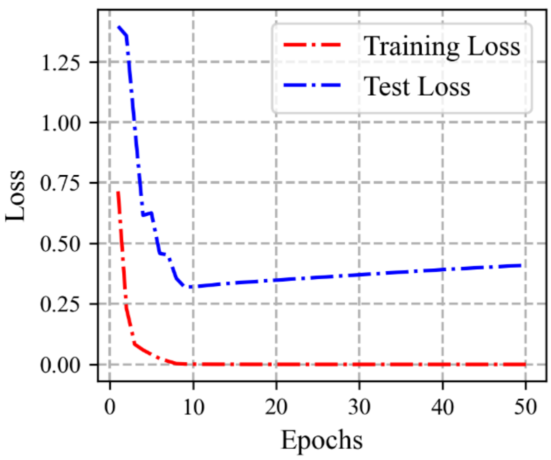
The parameters of network training were set as follows: batch size of 64, learning rate of 0.001 and 50 iterations. The results of the training are shown in Fig. 18 and Table 9.



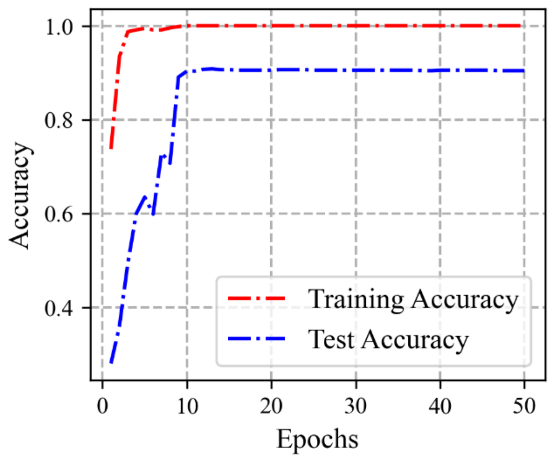
(a)Fold1: Training and test accuracy (b) Fold1: Training and test accuracy



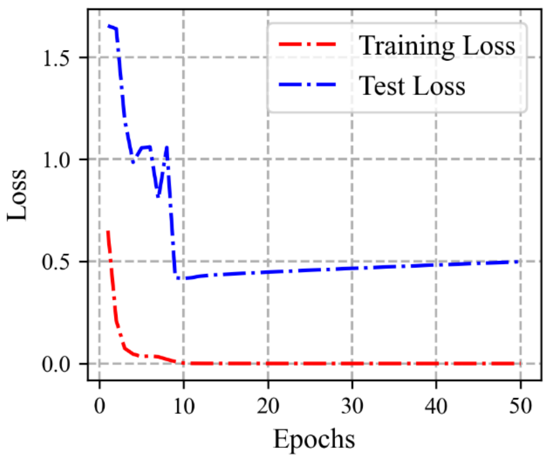
(c) Fold2: Training and test accuracy



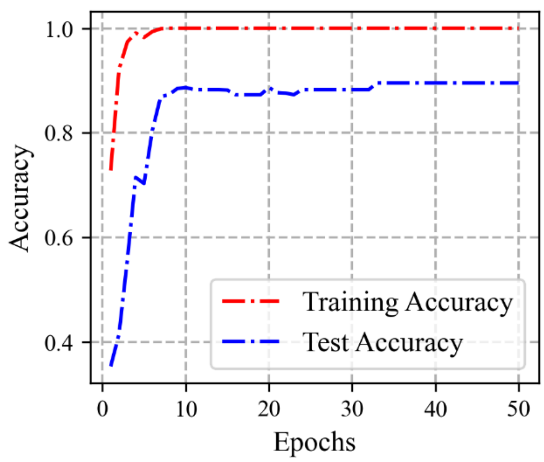
(d) Fold2: Training and test accuracy



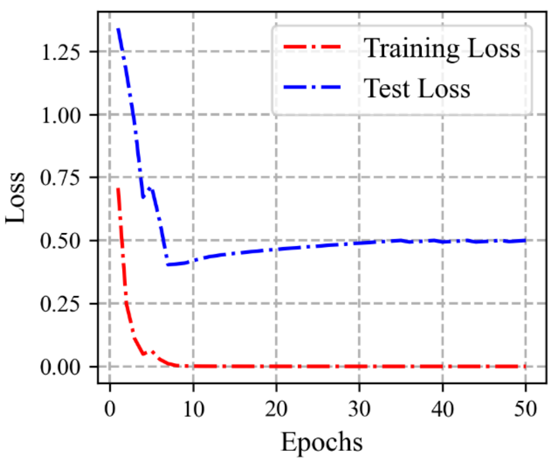
(e) Fold3: Training and test accuracy



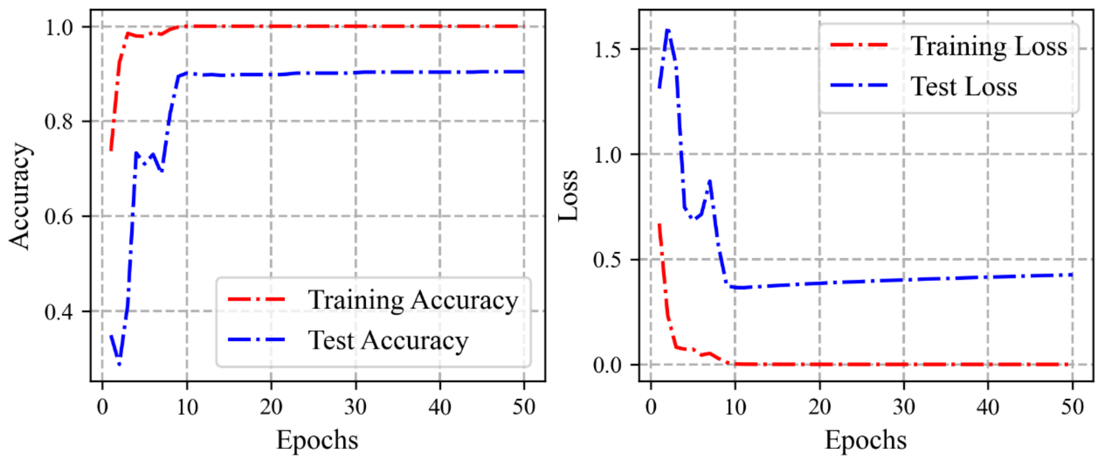
(f) Fold3: Training and test accuracy



(g) Fold4: Training and test accuracy



(h) Fold4:Training and test accuracy



(i) Fold4: Training and test accuracy (j) Fold4: Training and test accuracy

Fig. 18 Training and test results for 5-fold-cross-validation

Table 9 Cross-validation accuracy and training time.

	Fold1	Fold2	Fold3	Fold4	Fold5	**Average**
Accuracy (%)	90.14	91.21	90.43	89.55	90.43	**90.35**
Train time(s)	147.10	144.99	144.96	141.64	142.56	**144.66**

From Fig. 18 and Table 9, it can be observed that the entire dataset is effectively utilized to increase the reliability of the model evaluation through cross-validation, further demonstrating that it has a good stability and generalization ability for weld defect classification.

6 Conclusions

A weld defect recognition method from ultrasonic signals based on GASF and an improved DCGAN-ResNet is proposed in this paper, and the experiments are performed to verify its effectiveness and feasibility. The key findings are as follows:

(1) Ultrasonic A-scan signal of weld defect is a type of time series signal, which is converted into an image using GASF for feature extraction and recognition by a convolutional neural network.

(2) The DCGAN model is improved by introducing Wasserstein distance as the training loss function and spectral normalization to achieve sample balance and augmentation, and the high-quality data samples are generated to improve the accuracy of the classification model.

(3) Group convolution is utilized to improve the residual block by dividing the input features into multiple groups for feature extraction within each group separately, and the SE attention mechanism is then used for feature-weighted fusion, which reduces the number of network parameters and thus enhances the representational capability of the network.

(4) The improved Resnet model is trained and tested on the balanced and augmented dataset, achieving a recognition accuracy of 91.99%. Compared to other methods, the proposed approach has a higher recognition accuracy, efficiency, faster training speed, and has a potential applicability in other image recognition fields.

(5) The proposed method has significant potential for practical applications that can potentially be integrated into intelligent measurement systems or operate on a stand-alone computer by modularizing the steps such as data preprocessing, feature extraction, model training and prediction. Although it may be necessary to retrain the model when using a different material on the welded plates or changing the measurement system parameters, this challenge can be addressed by some techniques such as transfer learning, etc. Thus, future research could further explore solutions based

on the above considerations.

Acknowledgements

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

References

[1] Chi D Z, Qi C C. Research on NDT based on acoustic in welding defects testing in China[J]. Journal of Netshape Forming Engineering, 2018, 10(1):74-81.

[2] Dwivedi S K, Vishwakarma M, Soni A. Advances and researches on nondestructive testing: A review[J]. Materials Today: Proceedings, 2018, 5(2): 3690-3698.

[3] Fan W, Wan D Y, Xu Z Y, et al. Feature extraction of echo signal of weld defect guided waves based on sparse representation[J]. IEEE Sensors Journal, 2019, 20(5): 2692-2700.

[4] Han X D, Li G Y, Hu Y N, et al. Research on time-domain ultrasonic signal recognition technology based on one-dimensional CNN [J]. Electronic Measurement Technology, 2022, 45(12): 20-25.

[5] Zhou W, Du Y G, Wang S D, et al. Effects of welding defects on the damage evolution of Q245R steel using acoustic emission and infrared thermography[J]. Nondestructive Testing and Evaluation, 2023, 38(2): 189-210.

[6] Xu J M, Ke L M. Characteristic feature based on wavelet packet pattern recognition for ultrasonic inspection signals from defects in FSW Joints Using artificial neural network[J]. Nuclear Power Engineering, 2020, 41(1):163-166.

[7] HosseinAbadi H Z, Amirfattahi R, Nazari B, et al. GUW-based structural damage detection using WPT statistical features and multiclass SVM[J]. Applied Acoustics, 2014, 86(12): 59-70.

[8] Yi Q J, Wang H T, Guo R P, et al. Laser ultrasonic quantitative recognition based on wavelet packet fusion algorithm and SVM[J]. Optik, 2017, 149: 206-219.

[9] Shi C L, Shi F F, Zhang B X. Defect type analysis based on deep neural network and wavelet packet transform [J]. Acta Acoustica, 2016, 41(04): 499-506.

[10] Yang D M, Cui Y R, Yu Z Y, et al. Deep learning based steel pipe weld defect detection[J]. Applied Artificial Intelligence, 2021, 35(15): 1237-1249.

[11] Shi L, Zhao S, Niu W. A welding defect detection method based on multiscale feature enhancement and aggregation[J]. Nondestructive Testing and Evaluation, 2023: 1-20.

[12] 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.

[13] Zhi Z L, Jiang H Q, Yang D Y, et al. A deep learning fusion model of wave and image data for weld defect recognition [J]. Journal of Xi'an Jiaotong University, 2021, 55(5): 73-82.

[14] Hou W H, Wei Y, Jin Y, et al. Deep features based on a DCNN model for classifying imbalanced weld flaw types[J]. Measurement, 2019, 131: 482-489.

[15] Sonwane S, Chiddarwar S. Enhancing weld defect detection and classification with MDCBNet: a multi-scale dense cross block network for improved explainability[J]. NDT & E International, 2024, 142: 103029.

[16] Ahmadi H R, Mahdavi N, Bayat M. A novel damage identification method based on short time Fourier transform and a new efficient index[J]. Structures, 2021, 33: 3605-3614.

[17] Tian C, Zheng M, Zuo W, et al. Multi-stage image denoising with the wavelet transform[J]. Pattern Recognition, 2023, 134: 109050.

[18] Yan J, Kan J, Luo H. Rolling bearing fault diagnosis based on Markov transition field and residual network[J]. Sensors, 2022, 22(10): 3936.

[19] Afonso L C S, Rosa G H, Pereira C R, et al. A recurrence plot-based approach for Parkinson’s disease identification[J]. Future Generation Computer Systems, 2019, 94: 282-292.

[20] Wang, Z G and Oates T. Imaging time-series to improve classification and imputation[C]. The 24th International Joint Conference on Artificial Intelligence, 2015, 3939-3945.

[21] 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, 2024: 1-20.

[22] Xu Z X, Guan S Y, Yang F, et al. Laser ultrasonic surface defects detection method based on 2D-CNN [J]. Journal of Applied Optics, 2021, 42(1): 149-156.

[23] Ding L, Wan H P, Lu Q B, et al. Using deep learning to identify the depth of metal surface defects with narrowband SAW signals[J]. Optics & Laser Technology, 2023, 157: 108758.

[24] Miao R, Shan Z T, Zhou Q Y, et al. Real-time defect identification of narrow overlap welds and application based on convolutional neural networks[J]. Journal of Manufacturing Systems, 2022, 62: 800-810.

[25] Luo J, Huang J Y. Generative adversarial network: an overview[J]. Chinese Journal of Scientific Instrument, 2019, 40(3): 74-84.

[26] 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.

[27] Liu Y Q, Cai H Y, Li W K et al. Research on deep convolutional generative adversarial networks diagnosis method for bearing fault under small sample condition[J]. Journal of Vibration, Measurement & Diagnosis, 2023, 43(4):817-823+836.

[28] Deng C Y, Deng Z H, Lu S, et al. Fault diagnosis method for imbalanced data based on multi-signal fusion and improved deep convolution generative adversarial network[J]. Sensors, 2023, 23(5): 2542.

[29] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[C]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[30] Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: An overview[J]. IEEE Signal Processing Magazine, 2018, 35(1): 53-65.

[31] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 770-778.

[32] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 7132-7141.

[33] Adler J, Lunz S. Banach Wasserstein Gan[J]. Advances in neural information processing systems, 2018, 31.

[34] Miyato T, Kataoka T, Koyama M et al. Spectral Normalization for Generative Adversarial Networks[C]. ICLR, Vancouver Convention Center, Vancouver, BC, Canada, 2018, April 30-May 3.