IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 September 2023, accepted 23 October 2023, date of publication 1 November 2023,
date of current version 8 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329130

RESEARCH ARTICLE

Mixed Attention Mechanism Generative Adversarial Network Painting Image Conversion Algorithm Based on Multi-Class Scenes

XINXIN NIE[1,2] AND JING PU[3]
[1]College of Literature and Media, Chengdu Jincheng College, Chengdu 611731, China
[2]College of Creative Arts Studies, University of Technology MARA, Kuala Lumpur 47810, Malaysia
[3]School of Arts and Media, Sichuan Agricultural University, Ya'an 625014, China

Corresponding author: Jing Pu (pujing325@163.com)

ABSTRACT With the rapid development of computer vision and artificial intelligence, image conversion technology has been widely applied in the fields of digital media and art. Based on the Generative adversarial network, this paper proposes a painting image conversion algorithm for multi class scenes. It performs deep feature extraction in the residual module, divides the encoding and decoding parts of the generator into functional parts, and designs them separately. The mixed attention module is inserted into the decoder and encoder to preserve the texture details of the image. The deep network interpolation is incorporated to achieve smooth and continuous conversion of the painting image. The experiment showed that the loss value of the research method decreases to 0.008 after 400 iterations during the loss value testing. Its maximum peak signal-to-noise ratio is 34.9dB when the bit rate increases to 1000kb/s. In the SAR image conversion dataset, the F1 value increases to 97.4 after 200 iterations. The pixel loss when it reaches 100% conversion in outdoor images is 5.38k. The data indicates that the research method has good performance in painting image conversion and can provide effective technical references for image conversion.

INDEX TERMS Image conversion, generative adversarial network, mixed attention, deep network interpolation, global discrimination.

I. INTRODUCTION

Painting image conversion algorithm has a wide range of applications in the field of computer vision and image processing. It can convert images from one style or scene to another, providing powerful tools for image editing and synthesis [1]. With the continuous development of computer graphics and deep learning technology, the painting image conversion algorithm has made remarkable progress [2]. However, existing methods still have some problems, such as insufficient generalization ability and insufficient detail retention in multi-class scenes. This limits the wide use of painting image conversion techniques in practical applications [3]. The acquisition and annotation of training data in multiple scenarios is also a huge project. The lack of data will affect the generalization ability of the model. The traditional image conversion algorithm based on style transfer uses convolutional neural network to input the style of art works. But the conversion result still has style distortion and the calculation cost is high. The image conversion method based on texture synthesis extracts texture features from painting works and applies these textures to input images. However, when extracting complex painting styles, the conversion effect of strokes and lines is poor [4]. The hybrid attention mechanism can help the model focus on different scenes and improve the accuracy and diversity of image transformation. Deep network interpolation can be used to smooth the image conversion process and maintain the consistency and authenticity of the image. By training the antagonistic process of generator network and discriminator network, highly realistic images can be generated, and painting techniques and features in artistic styles can be imitated to produce more realistic artworks [5]. Moreover, the generative adversarial network can not only generate realistic images, but also generate diverse outputs. With random input or noise introduction, the generator can produce multiple images of

The associate editor coordinating the review of this manuscript and approving it for publication was Byung-Gyu Kim.

the same style but slightly different. This adds creativity and variety. Generative adversarial networks allow end-to-end training. This means that the generator and the discriminator can be trained simultaneously. This eliminates the need to manually design the feature extractor and simplifies the entire model building process.

The spatial attention mechanism, multi-head attention mechanism and mixed attention mechanism can be used to optimize the painting image conversion algorithm. Hybrid attention mechanisms combine multiple types of attention mechanisms. By adaptively selecting and tuning the attentional mechanisms according to the nature of the task and the data, they have the flexibility to improve the performance of the model in different situations. In painting image conversion tasks, different styles, elements and features require different types of attention. So mixing attention mechanisms can help models better adapt to these changes and produce more artistic and quality painting effects. In view of this, a painting image conversion algorithm based on hybrid attention mechanism generating adversarial network is proposed to provide a feasible reference scheme for image conversion.

The paper mainly consists of five parts. The first part discusses and summarizes the current research results on image conversion and GAN. The second part mainly designs the PIC algorithm for MAM-GAN facing multi-class scenarios, and elaborates on the technical means and thinking used in the algorithm. The third part analyzes the effectiveness of the research method, and conducts performance testing and application analysis on the research method. The fourth part discusses the research content and experimental results. The final part is a discussion and summary of the entire text.

II. RELATED WORKS

Image is an important carrier of information transmission. With the advancement of information technology, more and more scholars have realized the importance of Image Conversion Technology (ICT). Some scholars have conducted relevant research on this technology. M. Elmezayen et al. proposed a method based on CMOS sensors for analog-to-digital conversion in image capture and conversion. It used a column parallel architecture to predict the conversion time and power in the image, and introduced indicators to evaluate the monotonicity and distortion of the image. The proposed method has good detection sensitivity and operational efficiency, and can generate accurate detection results [6]. Xie and his research team proposed an automatic encoder method for the filling problem in comic style image conversion. It introduced an intermediate domain for screening comic mapping, summarized local texture features, and unified the data attributes of color filling. This method can generate good comic style images [7]. Zhou and other researchers proposed a simplified imaging method for targets in response to the conversion problem of radar images. It introduced image interpolation and coordinate conversion techniques, and used distance algorithms to construct a near field correction model, which has high robustness and can effectively complete image conversion [8]. T Deepa's team proposed an image conversion method based on overlap point recognition for the recognition of normal and abnormal cell images. This method learned the overlapping area of nucleus and cytoplasm, identified the overlapping points when the nucleus overlaps, and then divided the cell image with the learned model. So it has good cell image processing performance [9]. Kosowski et al. and other scholars proposed a conversion method that integrates time and voltage parameters for the conversion of CMOS images. It extracted the image pixels of the light-emitting diode and regulated the reference voltage during the operation of a single slope analog-to-digital converter. The good image conversion speed in low pixel areas has been verified [10].

Some scholars have conducted relevant research on GAN. Scholars such as Li proposed a GAN based method for predicting the temperature of molten steel in electric arc furnaces. It optimized the generator by using the Long short-term memory network, displayd the state of the transformer furnace, and predictd the trend of temperature change. This method has good prediction accuracy and the algorithm occupies less storage space [11]. Niu et al. proposed a classification method that combines GAN to address the classification problem of imbalanced data. It used approximate data distribution to process data and generate new data, and introduced Differential evolution to determine the number of a few data. This method has good classification performance and can perform accurate data classification [12]. Muhammad et al. used classification models to measure model performance by introducing deep convolutional multiple images at different stages for judgment. Based on the results, they optimized and adjusted the model to solve the problem of dataset annotation in deep learning. The optimized model has good peak signal-to-noise ratio performance during runtime and can perform accurate dataset annotation [13]. Scholars like Yu proposed a GAN based pattern generation method to address the bottleneck issue of pattern design in clothing design. It used multi-scale discriminators to process local texture details and utilizes self attention mechanisms to enhance the system's artistic perception ability. This method has good running speed and can effectively generate paper pattern schemes that meet the requirements [14]. Fan et al. proposed a GAN-based perception method for detecting abnormal behavior in videos, which alternates training between generative adversarial and perceptual adversarial algorithms to establish a dual stream structure for policy updates. The proposed algorithm has accurate abnormal behavior detection ability and can still maintain good performance when multiple targets appear [15].

In summary, although GAN has been studied and applied in various fields, research on image conversion is still relatively scarce. In view of this, a GAN-based multi-scene PIC algorithm is raised to provide more feasible technical references for the field of image conversion.

III. PIC ALGORITHM FOR MAM-GAN

ICT can provide more processing space for image processing. This chapter will focus on elaborating the PIC algorithm structure of the designed MAM-GAN.

A. IMAGE CONVERSION ALGORITHM BASED ON GAN

Most computer vision problems can be seen as image conversion problems. The pixel mapping from an image in one region to the corresponding image in another region is called image conversion [16]. Common image conversion operations include mapping colorless images to corresponding colored images for image coloring. The other is to map low resolution images to corresponding high-resolution images for super-resolution conversion [17], [18]. When performing image conversion, the expression of the image will change, and the noise vector is easily ignored during the conversion process, leading to pattern collapse. GAN can automatically conduct confrontation learning on the Loss function after specifying the generated image to enhance the network performance to achieve the task goal. Conditional GAN can solve the problem of uncontrollable image conversion results [19]. The GAN-based image conversion algorithm consists of two parts: a generator and a discriminator. The generator uses a three segment structure, including downsampling, residual module, and upsampling, as shown in Figure 1.

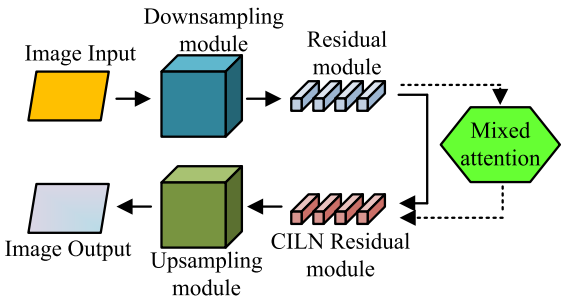


FIGURE 1. Generator network structure.

In Figure 1, the image first enters the downsampling module for preliminary extraction of image features after input, and then extracts depth features in the residual module. The MAM is inserted into the middle section of the residual network, and some residual modules are normalized through compromise instances and layer normalization. Finally, the attributes are input to the upsampling module for feature mapping reconstruction, and then the generated image is output. During the process of converting instances from the source domain space to the target domain space, the generator comes into play. The generator can be divided into two parts based on its functionality: encoder and decoder. The encoder includes downsampling and 9-layer residual module. The decoder includes 9 layers of residual and upsampling modules that have been normalized through compromise instances and layer normalization. The upper sampling module and the lower sampling module each contain three layers of convolution layer and deconvolution layer. Each layer of convolution layer and deconvolution layer is composed of convolution normalization Activation function structure. The last layer of the upsampling module and the first layer of the downsampling module use convolutional kernels of size to capture more details. The structure of the hybrid attention module inserted between the decoder and encoder is Figure 2.

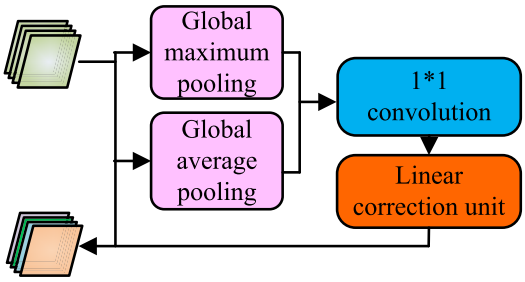


FIGURE 2. Mixed attention module.

The module in Figure 2 includes Global Maximum Pooling (Gmp), Global Average Pooling (Gap), Convolutional and Modified Linear Units. Attention can enhance generator control, enhance attention to objects and edges in the image, retain the texture details of the image, and reduce the loss of details during the conversion process. The original image is input into the Gap layer to calculate the mean of the feature map, reflecting the importance of the feature map in the corresponding channel. The original image is input into Gmp to preserve the maximum value of the image and pay attention to important areas of the image. The new feature maps obtained from the Gmp and Gap layers are cascaded together. The convolution is used to extract potential combination relationships. Then it calculates the weight coefficients of different channel feature maps through a linear correction unit. The coefficient is multiplied by the original feature map to generate a feature map of interest that includes channel and region information. The generation of attention feature maps is Equation (1).

$$\begin{cases} W = F\left(C_E\left(x\right)\right) \\ o\left(x\right) = W * C_E = \left\{w^k * C_E^k\left(x\right) | 1 \le k \le n_c\right\} \end{cases} \tag{1}$$

In equation (1), W represents the weight coefficient obtained through learning. $C_E\left(x\right)$ represents the original feature map. F represents the attention module. $o\left(x\right)$ represents the feature map generated after passing F. $C_E^k\left(x\right)$ represents the encoder output feature map of the k-th channel. w^k represents the weight coefficient obtained from the F-learning of the k-th channel. n_c represents the total number of encoder channels. The normalization calculation of compromise instances and layer normalization is Equation (2).

$$CILN\left(a, \delta, \omega, \varphi\right) = \omega\left(\delta\frac{a-\mu_I}{\sqrt{\sigma_I^2+\varepsilon}} + \left(1-\delta\right)\frac{a-\mu L}{\sqrt{\sigma_L^2+\varepsilon}}\right) + \varphi \tag{2}$$

In equation (2), *CILN* represents compromise instances and layer normalization. a represents input data. μ_I, μ_L and σ_I, σ_L represent the mean and standard deviation of instance and layer normalization. ω represents the scaling parameter obtained from the automatic update of the fully connected layer. φ represents the translation parameter obtained by automatically updating the fully connected layer. δ is a weight parameter dynamically calculated from the fully connected layer. This discriminator adopts a multi-scale global discriminant discriminator. It combines multiple discriminators with the same structure at different image scales into a network and inputs complete images for global discriminant analysis. The study set up a discriminator consisting of three discriminators to obtain Figure 3.

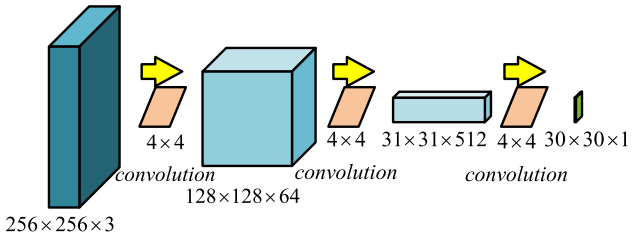


FIGURE 3. Discriminator construct.

In Figure 3, when using three discriminators, the image needs to undergo two image downsampling before making a judgment. The discriminator uses a convolutional kernel with size 4 × 4, and the padding has a size of 1. The image with size 256 × 256 × 3 undergoes two layers of convolution after input to obtain a feature map with size 31 × 31 × 512, and then undergoes a channel reduction convolution to obtain a feature map with size 30 × 30 × 1 for output. The final feature map output pixels are judged by sigmoid for true or false probability, and BCEloss is used to calculate the final loss. The Loss function determines the learning effectiveness of the model. The design of the Loss function consists of L1 regularization loss, generation confrontation loss, perception loss and feature matching loss. The adversarial loss in the optimization Equation obtained using a multi-scale decision maker is Equation (3).

$$L_{GAN}(G, D_k) = E_{(x,y)}\left[\log D_k(y)\right] + E_x\left[\log\left(1 - D_k(G(x))\right)\right] \quad (3)$$

In equation (3), L_{GAN} represents adversarial loss. x represents the source domain image. y represents the target domain image. D_k represents the feature of the k-th discriminator. $G(x)$ represents the image generated by the generator. E represents mathematical expectation. The generated image is constrained to guide the results towards the target domain. The L1 regularization loss is introduced, resulting in equation (4).

$$L_{L1}(G) = E_{(x,y)}\left[\|y - G(x)\|_1\right] \quad (4)$$

In equation (4), L_{L1} represents L1 regularization loss. The feature matching loss is Equation (5).

$$L_{FM}(G, D_k) = E(x, y)\sum_{i=1}^{T}\frac{1}{N_i}\left[\left\|D_k^{(i)}(x, y) - D_k^{(i)}(x, G(x))\right\|_1\right] \quad (5)$$

In equation (5), L_{FM} represents feature matching loss. T represents the total number of layers of the discriminator. N_i represents the number of elements in layer i of the discriminator. $D_k^{(i)}$ represents the i-th layer feature in the k-th discriminator network. The feature matching loss is learned from the multi-layer features of the discriminator, matched with the intermediate features of the synthesized image and the real image for energy efficiency, and the gap is measured. The pre trained perceptual loss network is used to extract image features to enhance the discriminant ability of the discriminator. Since image conversion only considers from source domain to target domain, the setting of perception Loss function is Equation (6).

$$L_{VGG}(G) = \sum_{i=1}^{N}\frac{1}{M_i}\left[\left\|F^{(i)}(y) - F^{(i)}(G(x))\right\|_1\right] \quad (6)$$

In equation (6), L_{VGG} represents perceived loss. $F^{(i)}$ represents the i-th layer of the perceptual loss network in the pre trained model. M_i represents the number of elements in layer i of the perceptual loss network. The final Loss function is the minimum of the sum of L1 regularization loss, generation confrontation loss, perception loss and feature matching loss.

B. CONTINUOUS CONVERSION ALGORITHM FOR DRAWING IMAGES BASED ON DNI FUSION

When performing PIC, users need to obtain varying degrees of preview of conversion results. Using a certain image conversion output cannot meet this requirement [20], [21]. This paper integrates DNI to achieve smooth control of the network and achieve continuous transitions of image conversion effects to varying degrees. Generating a balance value between two mappings can achieve a balanced transition between them. The parameter for DNI parameter fusion is Eq (7).

$$\theta_{fusion} = \alpha\theta_A + (1 - \alpha)\theta_B \quad (7)$$

In equation (7), θ_{fusion} represents the parameters of the fusion model. α represents the fusion coefficient. θ_A and θ_B represent parameter vectors. When the quantity of models increases to N, these model parameters have a certain correlation, and the fusion model parameters can be represented as Equation (8).

$$\theta_{fusion} = \alpha_1\theta_A + \alpha_2\theta_B + \cdots + \alpha_n\theta_N \quad (8)$$

In equation (8), the fusion coefficients of each model are different, and the sum of all fusion coefficients is 1. The fusion model parameters are the linear Convex combination of model parameters, and the continuous conversion effect

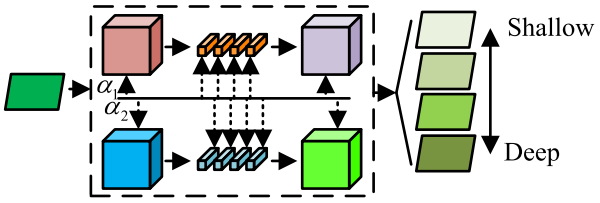


FIGURE 4. Deep network interpolation structure.

can be achieved by adjusting the fusion coefficient. Figure 4 shows the structure of DNI.

In Figure 4, the fusion, convolution, and normalization of the model are performed on a deep network layer with parameter vectors. The convolutional layer contains two operations: bias and convolution, using filter coefficients and bias as calculation parameters. After completing the convolution operation, the bias will be added to the results, and the model will be fused in addition to the coefficients in the convolution kernel. The bias will also be operated on, and the continuous nonlinear activation layer will be affected by the bias. Assuming there is a high-dimensional space containing all natural images, the degradation steps of natural images are uninterrupted in space and can be considered as adjacent in space. Figure 5 shows the data space mapping process.

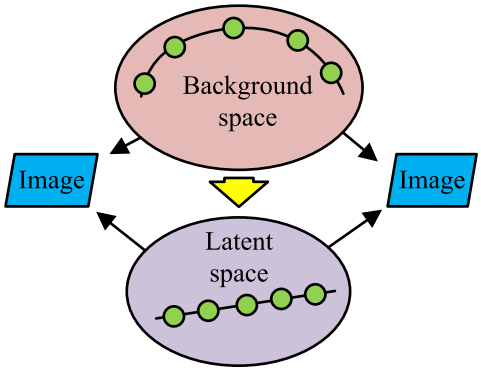


FIGURE 5. Data space mapping.

From Figure 5, in real background space, natural images are located on approximately nonlinear manifolds. If the image difference is too simple, it can lead to unexpected error details such as artifacts appearing in the results. If only analyzing the background space of an image, it is difficult to obtain high-quality continuous image transformation or restoration. Therefore, it is necessary to search for potential high-dimensional spaces and achieve meaningful interpolation. The manifold mapped by the depth neural network is approximated to a linear Euclidean space region, and the two positions in the Potential space are expressed as endpoints. Then the unknown points are expressed as Equation (9).

$$Z_i \approx \alpha_i X_i + (1 - \alpha_i)\, Y_i \tag{9}$$

In equation (9), X_i and Y_i represent two positions in the Potential space. Z_i represents an unknown point. α_i represents the affine coefficient. When performing continuous image conversion, image blurring and low brightness are commonly used functions. The continuous transformation of image blurring helps to provide technical support for operations such as image clarity and sharpening. When performing image blur continuous transformation, the first step is to learn clear images by reviewing the image transformation models of adversarial networks. The mapping relationship of the generated clear image serves as the origin of the continuous transformation. Then, set the foundation for the next training step and learn the convolutional model in advance in the fuzzy transformation model. Afterwards, generate images with 100% concentration blur and learn the mapping relationship between clear images and 100% concentration blur images. Finally, the model obtained from the first two steps of learning is subjected to DNI, and fusion coefficients are introduced to generate a fusion model to complete the continuous transformation of image blur. The continuous conversion of low brightness images can provide technical support for operations such as changing image contrast and saturation. There is a lack of continuous data for comparison when performing image low brightness continuous conversion. Therefore, the first step is to generate a minimum brightness conversion image as a pre trained model, and then retrain the conversion models for normal brightness images and low brightness images. The two models set DNI and introduce fusion coefficients for fusion, resulting in a series of conversion models of varying degrees, achieving continuous conversion from high brightness to low brightness. The brightness fusion model is Equation (10).

$$Net_{fusion} = \alpha Net_{high} + (1 - \alpha)\, Net_{low} \tag{10}$$

In equation (10), Net_{fusion} represents the brightness fusion model. Net_{high} represents the high brightness conversion model. Net_{low} represents the low brightness conversion model. In this study, unsupervised learning is used to train the image transformation model. But when painting transformation is carried out, most images do not have matching objects. At this point, L1 regularization loss, generative adversarial loss, perception loss, and feature matching loss are not applicable. This paper uses a multi-scale global convolutional discriminator for whole feature map discrimination, and introduces identity loss and cyclic consistency loss. The optimized generator structure obtained is Figure 6.

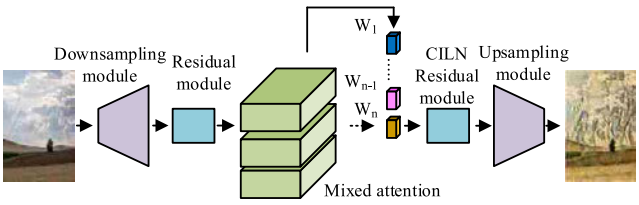


FIGURE 6. Optimize generator structure.

In Figure 6, the mixed attention module is still inserted between the encoder and the encoder. The image from the

source domain is input to the downsampling module on the centerline of the encoder for preliminary feature extraction, which is then extracted by the residual module. The information of the channel and key areas is fused through a mixed attention module. The deep convolution is performed by a balance normalization residual module. Clear images are reconstructed in the upsampling module. The global convolutional identification is completed by an optimized discriminator and belongs to a fully convolutional network. Figure 7 shows its structure.

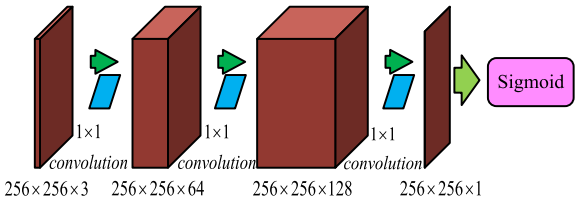


FIGURE 7. Optimize discriminator structure.

As Figure 7, the optimized discriminator consists of three layers of convolution, with each layer having a consistent kernel size of 1 × 1 and a sliding step size of 1. There is no Padding structure, so the size of the feature map will not change after passing through each layer of convolution. The number of channels in the first layer is 3, the second layer is 64, and the third layer is 128, resulting in a feature layer with 1 channel. The pixels in the feature map are judged by the sigmoid layer for true or false probability, and then BCEloss is used to calculate the final loss. The consistency loss in the loop comes from the consistency of the loop, as defined in Equation (11).

$$L_{cyc}(G,F)=E_x\left[\|F(G(x))-x\|_1\right]+E_y\left[\|G(F(y))-y\|_1\right] \tag{11}$$

In equation (11), L_{cyc} represents the cyclic consistent loss. Re-introduce identity loss as defined in Equation (12).

$$L_{idt}(G,F)=E_y\left[\|G(y)-y\|_1\right]+E_x\left[\|F(x)-x\|_1\right] \tag{12}$$

L_{idt} in Equation (12) represents identity loss. The adversarial loss in forward mapping is Equation (13).

$$L_{gan}(G,D_Y)=E_y\left[\log D_Y(y)\right]+E_x\left[\log(1-D_Y(G(x)))\right] \tag{13}$$

In equation (13), L_{gan} represents the adversarial loss in the forward mapping; D_Y is the discriminant in forward mapping. The constructed total Loss function is Equation (14).

$$L(G,F,D_X,D_Y)=L_{gan}(G,D_Y)+L_{gan}(F,D_X)+\lambda L_{idt}(G,F) \tag{14}$$

D_X in equation (14) represents discrimination in reverse mapping. The Loss function is used to train the model, and a mature unsupervised image conversion model is obtained. To ensure the performance of the model, it is necessary to calculate the structural similarity between the reference image and the processed image after completing the design, and then push back to optimize the model. The calculation of structural similarity is Equation (15).

$$SSIM(f,g)=\frac{(2u_fu_g+C_2)(2\sigma_f\sigma_g+C_2)}{(u_f^2+u_g^2+C_1)(\sigma_f^2+\sigma_g^2+C_2)} \tag{15}$$

In equation (15), $SSIM$ is structural similarity. g is the original image. f represents the processed image. u is the average brightness of the image. σ is the standard deviation of the image itself. C is a constant that avoids the denominator of the Equation being 0. The process of painting image conversion algorithm constructed by the paper is shown in Figure 8.

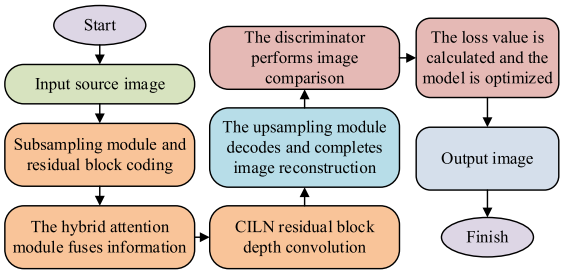


FIGURE 8. Painting image conversion algorithm flow.

From Figure 8, in the process of painting image conversion, the source image to be converted is first input and encoded by the downsampling module and residual block in the generator. Afterwards, the hybrid attention module is used to enhance control over the generator, focusing attention on the objects and edges of the fused image. This reduces the loss of texture details during the conversion process. Then the CILN residuals are used for deep convolution, and the image is decoded and reconstructed with the upsampling module. The reconstructed image input discriminator is compared with the original image, and the loss value is calculated to optimize the model. Finally, the converted painting image is output.

IV. PERFORMANCE TESTING AND APPLICATION ANALYSIS OF MAM-GAN'S PIC ALGORITHM

PIC is an important technical means in image secondary processing. This section will test the performance of the research method in PIC and analyze the application effect using real images.

A. PERFORMANCE TESTING OF MAM-GAN'S PIC ALGORITHM

To analyze the effectiveness of MAM-GAN's PIC algorithm during PIC, this chapter conducts performance testing and application analysis on the algorithm. Table 1 shows the basic software and hardware data for constructing the experiment.

The experiment used Lightroom CC to perform hazy processing on the Middlebury Stereo dataset, forming 608 pairs of images with different hazy degrees to form a hazy dataset. The dataset was then used for testing, along with the SAR image conversion dataset. The input image information in the constructed data set is shown in Table 2.

TABLE 1. Basic environmental parameters of the experiment.

Parameter variables	Parameter selection
Operating system	Windows10
Software environment	Pytorch
System PC side memory	16G
System running memory	64GB
CPU main frequency	3.30GHz
Graphics card model	NVIDIA GeForce Titan X
CPU	Intel（R）Core（TM）i7-10400K

TABLE 2. Input image information.

Use	Image type	Resolution (dpi)	Quantity (sheet)
Training set	Real image	1024×768	152
		2048×1536	147
	Painting image	1024×768	154
		2048×1536	153
Test set	Real image	1024×768	151
		2048×1536	150
	Painting image	1024×768	162
		2048×1536	148

Firstly, the loss curve of the research method was tested and compared with the Pix2pix algorithm and Digital Radiograph (DR) algorithm, as shown in Figure 9.

From Figure 9, the loss values of the three methods all decrease with the increase of iteration times, showing a rapid decline in the early stage, and gradually stabilizing after slowing down in the later stage. Pix2pix entered a slow decline stage in the 56th iteration, with the previous loss value decreasing from 0.092 to 0.028. At the 400th iteration, the loss value decreased to 0.023, and there were many fluctuations during the overall decline process. At the 54th iteration of the DR algorithm, it began to slowly decline. Previously, the loss value decreased from 0.076 to 0.022, and at the 400th iteration, the loss value decreased to 0.017, with fewer fluctuations during the overall decline process. At the 32nd iteration of the research method, the loss value slowly decreased from 0.054 to 0.012. At the 400th iteration, the loss value decreased to 0.008. There were few fluctuations during the overall decline process. The above data shows that the research method has a faster Rate of convergence, and the convergence process maintains good model stability. The test results of the peak signal-to-noise ratio (PSNR) of the research method are Figure 10.

In Figure 10, in both datasets, the overall PSNR of all methods increases with increasing bit rates. In the hazy dataset, the initial PSNR of Pix2pix is 16.4dB. After increasing the bit rate to 1000kb/s, the PSNR rises to 24.1dB, with significant fluctuations in the rate of rise. The initial PSNR of DR is 17.7dB. After increasing the bit rate to 1000kb/s, the PSNR rises to 27.6dB, with significant fluctuations. The initial PSNR of the research method is 24.5dB, and after increasing the bit rate to 1000kb/s, the PSNR rises to 34.9dB with relatively small fluctuations. In the SAR dataset, the initial PSNR of Pix2pix is 17.1dB. After increasing the bit rate to 1000kb/s, the PSNR rises to 24.4dB with little fluctuation. The initial PSNR of DR is 21.2dB. After increasing

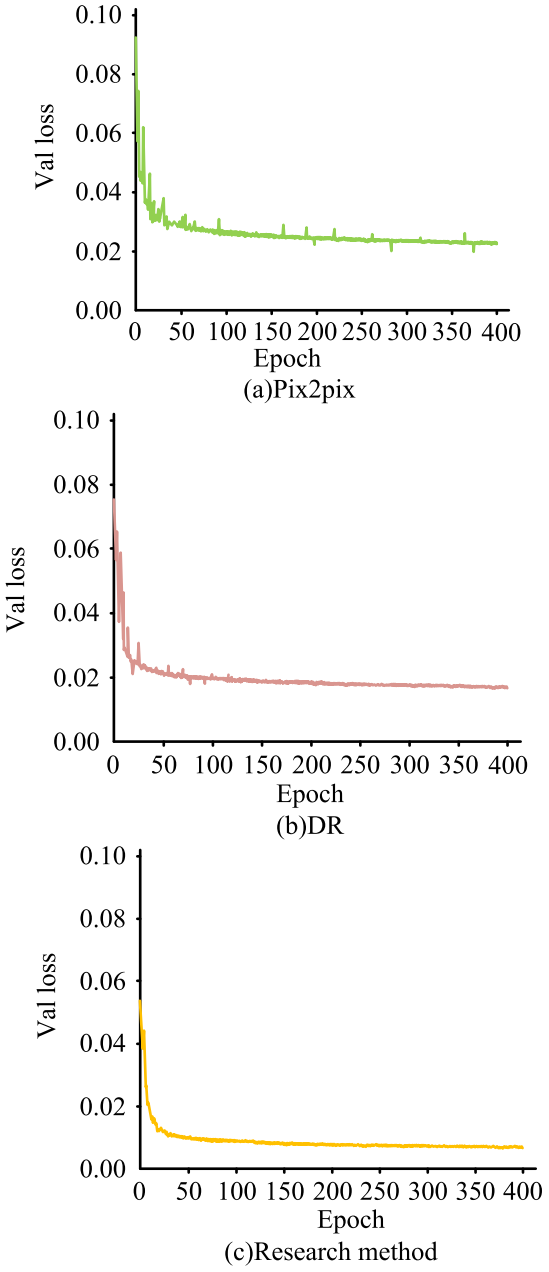


FIGURE 9. Loss curve.

to a bit rate of 1000kb/s, the PSNR rises to 27.6dB with little fluctuation. The initial PSNR of the research method is 27.1dB. After increasing the bit rate to 1000kb/s, the PSNR rises to 33.6dB with little fluctuation. Given this, the PSNR of the research method is higher and relatively stable, indicating that it processes less image distortion. Figure 11 shows the ROC rectangular curve for drawing the research method.

From Figure 11, the Receiver operating characteristics of Pix2pix and DR intersect with each other during the rising process, and the enclosed area is relatively close. It indicates that some functional performances of the two types of images are different during image conversion, but the overall performance is relatively close. The Receiver working feature of the research method is closest to the upper left corner of the ROC

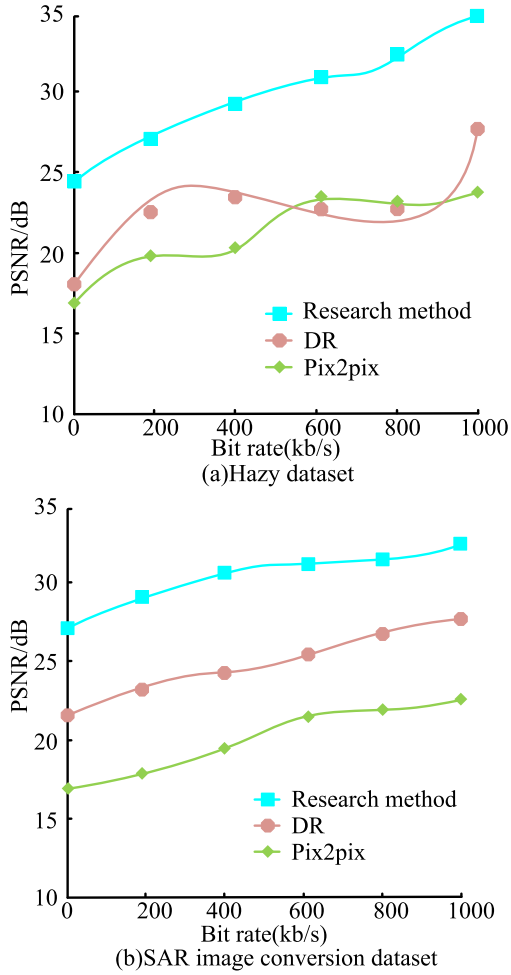


FIGURE 10. Peak signal-to-noise ratio test.

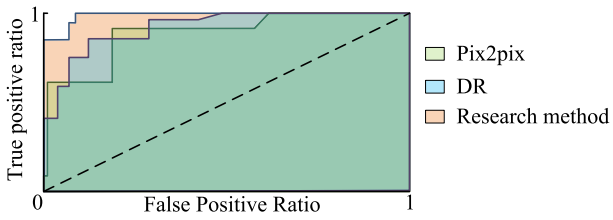


FIGURE 11. ROC rectangular curve.

rectangle during the ascending process, and the area is larger than Pix2pix and DR. This indicates that the various functions of the research method perform well in image conversion. The F1 value of the research method was tested and Figure 12 was obtained.

In Figure 12, in both datasets, the F1 values of each method iteratively increase and continue to rise. In the hazy dataset, the initial F1 values of Pix2pix, DR, and research methods are 63.7, 69.4, and 78.6, respectively. After 200 iterations, they rise to 74.3, 85.7, and 95.2, respectively. In the SAR image conversion dataset, the initial F1 values of Pix2pix, DR, and research methods are 63.5, 67.8, and 75.3, which rise to 77.1, 85.2, and 97.4 after 200 iterations. Therefore, the research

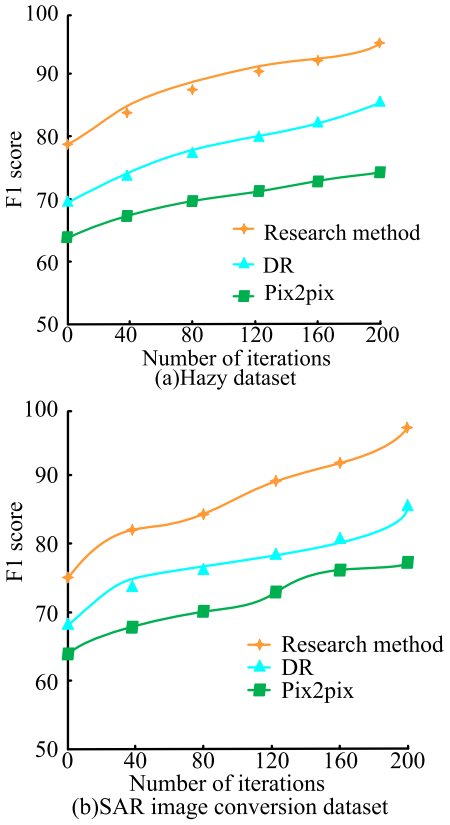


FIGURE 12. F1 value test.

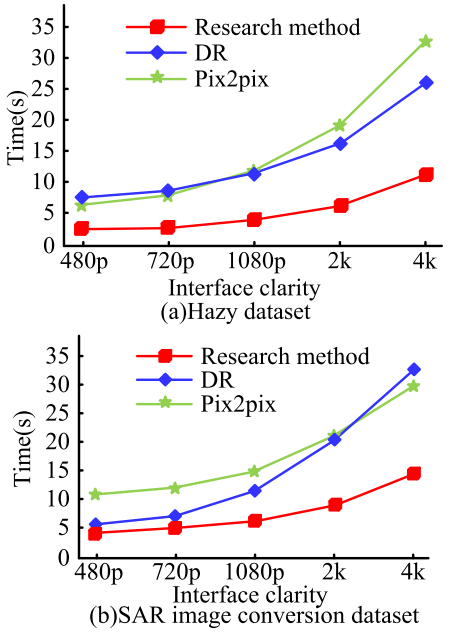


FIGURE 13. Run time.

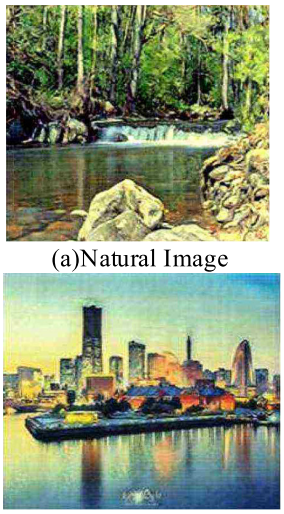
method has a higher F1 value, indicating better accuracy and recall. The test results of the running time are Figure 13.

In Figure 13, during runtime, the runtime of all three methods increases with the increase of image clarity, and the speed of runtime increase continuously with the increase of

clarity. In the hazy dataset, the running time of Pix2pix, DR, and research methods is 6.2s, 7.5s, and 2.7s when the image clarity is 480p. When the clarity increases to 4k, the running time increases to 32.4s, 25.8s, and 11.1s. In SAR, the running time of Pix2pix, DR, and research methods is 10.8s, 5.8s, and 4.2s when the image clarity is 480p. When the image clarity increases to 4k, the running time increases to 29.8s, 32.7s, and 14.6s. The data indicates that the research method runs faster in different types and sharpness of images.

B. APPLICATION ANALYSIS OF PIC ALGORITHM IN MAM-GAN

When analyzing the PIC application of the research method, 60 different scene images were taken with the camera for application analysis. The input image for example painting image conversion is shown in Figure 14.



(a)Natural Image

(b)City image

FIGURE 14. Instance input image.

Firstly, the pixel loss of the image during the PIC process is tested using an indoor and an outdoor scene image, as shown in Figure 15.

In Figure 15, the pixel loss of the image increases with the deepening of image conversion. In indoor images, the pixel loss of DR, Pix2pix, and research algorithms at 100% conversion is 9.67k, 11.34k, and 2.97k. In outdoor images, the pixel loss amounts corresponding to DR, Pix2pix, and research algorithms are 10.46k, 13.55k, and 5.38k, respectively. This indicates that the research method can better maintain pixel integrity and reduce information loss in image conversion. The test results of the structural similarity of the processed oil painting style conversion using the research method are displayed in Table 3.

From Table 3, when converting oil painting style images, the optimal structural similarity value for images with a 20% conversion degree is 0.9324. The similarity value of 40% is 0.9148. The value of 60% is 0.9334; 80% is 0.9295; 100% is 0.8975. The optimal structural similarity value of the research method did not show any deviation, indicating that it

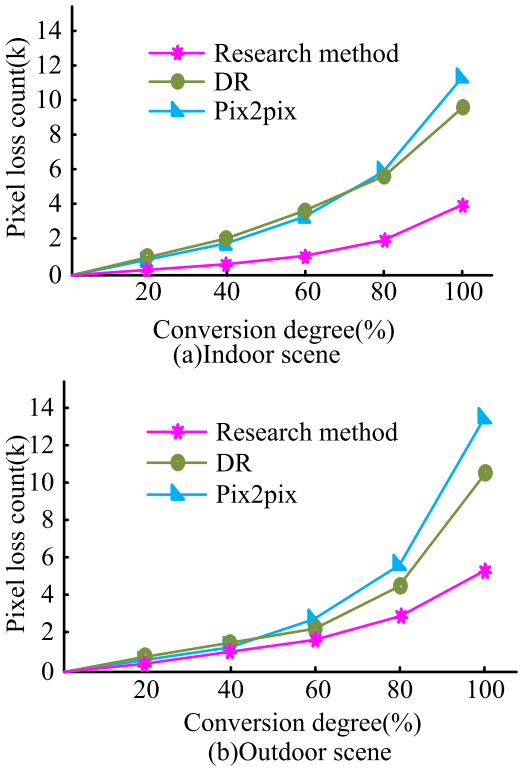


FIGURE 15. Pixel loss.

can accurately call parameters to obtain the optimal converted image during PIC. Figure 16 shows the PIC generation effect of the research method in natural and urban images.

From Figure 16, both DR Method and research method realize the painting conversion of nature image and city image. These two methods have achieved separate conversions of four painting styles: Van Gogh, Ukiyoe, Monet, and Cezanne, as well as various style conversions in different proportions. From the perspective of human naked eye observation, the conversion effect of DR Method on natural images is poor in the brightness balance of the bright parts of the picture, resulting in insufficient contrast of the whole picture. But the overall sharpening is higher, translating the image into more realistic in Van Gogh style. The conversion effect of DR Method on urban images is poor in terms of picture saturation, and the overall permeability of the picture is insufficient. The research method has good brightness balance and color balance in natural images, but the sharpening degree is not enough when dealing with complex image areas such as leaves. This research method has better clarity and saturation in the conversion results of urban images. And it performs better defogging on the image, giving it a better sense of transparency. The results show that the research method can effectively convert painting images and has good generalization performance.

V. DISCUSS

In the digital age today, remarkable progress has been made in the fields of computer graphics and artificial intelligence.

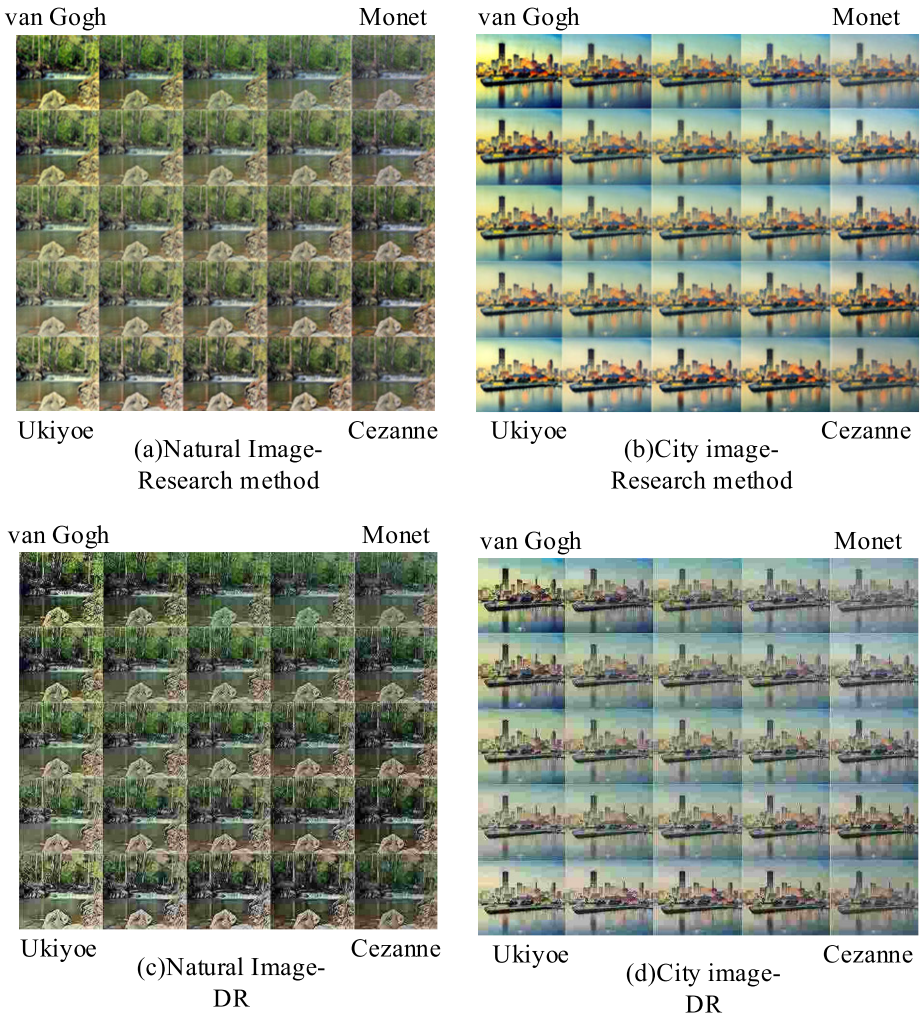


FIGURE 16. PIC effect.

TABLE 3. Similarity matrix of painting image conversion structure.

/	Clear image	20% Reduction	40% Reduction	60% Reduction	80% Reduction	100% Reduction
Clear image	0.9426	0.9308	0.9091	0.8890	0.8671	0.8332
20% Oil painting	0.9078	0.9324	0.9160	0.9093	0.8983	0.8705
40% Oil painting	0.8644	0.8925	0.9148	0.9066	0.9058	0.8875
60% Oil painting	0.8216	0.8588	0.8789	0.9334	0.9016	0.8920
80% Oil painting	0.7828	0.8252	0.8522	0.8738	0.9295	0.8716
100% Oil painting	0.7489	0.7938	0.8523	0.8517	0.8728	0.8975

Significant breakthroughs have also been made in the fields of image processing and synthesis. This paper takes painting image conversion as the goal of the method, and uses computer technology to construct painting image conversion algorithm. The experimental results indicated that the convergence speed of the loss curve in the iterative process of this research method is relatively fast. This indicated that the loss function composed of L1 regularization loss, generative adversarial loss, perception loss, and feature matching loss can play a role in model optimization. The peak signal-to-noise ratio of the research method was high and stable. This indicated that the sampling module introduced by the mixed attention mechanism in the research method has good sampling quality and can effectively retain key information of the image during the encoding and decoding process. By observing the effects of continuous image conversion and mixed multi style conversion, integrating deep network interpolation into the generative adversarial network can effectively achieve parameter balance and continuous calculation of the algorithm in different style conversions. It is proved that the research method is feasible in the conversion of painting image.

VI. CONCLUSION

ICT is an important technical component in image processing. This study proposes a MAM-GAN conversion algorithm for the PIC problem. A three stage generator structure has been designed and mixed attention modules have been added to the encoder and decoder. Afterwards, an attention feature map is generated by multiplying coefficients, and a multi-scale decision maker is introduced to obtain adversarial loss. DNI is fused into the model to achieve smooth image conversion. Finally, the effectiveness test results of the research method indicated that its loss curve rapidly decreases at the end of the 32nd iteration, reaching 0.012. Its initial PSNR in the SAR image conversion dataset was 27.1dB. The area enclosed by the curve in the ROC rectangular curve test was larger than that of other methods. After 200 iterations, the F1 value test reached a maximum of 97.4. When the image clarity in the hazy dataset was 4k, the running time was 11.1 seconds, which is lower than other methods. The optimal structural similarity value during the oil painting style image conversion test was maintained above 0.8975, without any deviation, effectively completing the mixed conversion of different painting styles in the image. The above results indicate that the research method has better model computing power and can effectively improve the quality of image conversion. However, the number of image samples used in the experiment is relatively small. In the future, the number of experimental samples will be increased, the experimental results will be enriched, and the plan will be optimized.

REFERENCES

[1] M. Ružička, M. Vološin, J. Gazda, T. Maksymyuk, L. Han, and M. Dohler, "Fast and computationally efficient generative adversarial network algorithm for unmanned aerial vehicle–based network coverage optimization," *Int. J. Distrib. Sensor Netw.*, vol. 18, no. 3, Mar. 2022, Art. no. 155014772210755, doi: 10.1177/15501477221075544.

[2] L. Wang, L. Wang, and S. Chen, "ESA-CycleGAN: Edge feature and self-attention based cycle-consistent generative adversarial network for style transfer," *IET Image. Process.*, vol. 16, no. 1, pp. 176–190, Oct. 2022, doi: 10.1049/ipr2.12342.

[3] Y. Li, S. Tang, R. Zhang, Y. Zhang, J. Li, and S. Yan, "Asymmetric GAN for unpaired image-to-image translation," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5881–5896, Dec. 2019, doi: 10.1109/TIP.2019.2922854.

[4] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 14–34, Sep. 2021.

[5] Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, "Generative adversarial networks: A survey toward private and secure applications," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–38, Jul. 2021, doi: 10.1145/3459992.

[6] M. R. Elmezayen and S. U. Ay, "A new blind image conversion complexity metric for intelligent CMOS image sensors," *IET Image Process.*, vol. 15, no. 3, pp. 683–695, Feb. 2021, doi: 10.1049/ipr2.12053.

[7] M. Xie, C. Li, X. Liu, and T.-T. Wong, "Manga filling style conversion with screentone variational autoencoder," *ACM Trans. Graph.*, vol. 39, no. 6, pp. 1–15, Nov. 2020, doi: 10.1145/3414685.3417873.

[8] Z. Xingyu, W. Yong, and L. Xiaofei, "Approach for ISAR imaging of near-field targets based on coordinate conversion and image interpolation," *J. Syst. Eng. Electron.*, vol. 32, no. 2, pp. 425–436, Apr. 2021, doi: 10.23919/JSEE.2021.000036.

[9] T. P. Deepa and A. N. Rao, "Estimation of a point along overlapping cervical cell nuclei in pap smear image using colour space conversion," *Int. J. Biomed. Eng. Technol.*, vol. 33, no. 1, p. 77, Aug. 2020, doi: 10.1504/ijbet.2020.10029770.

[10] M. Kłosowski, "Hybrid-mode single-slope ADC with improved linearity and reduced conversion time for CMOS image sensors," *Int. J. Circuit Theory Appl.*, vol. 48, no. 1, pp. 28–41, Jan. 2020, doi: 10.1002/cta.2713.

[11] C. Li and Z. Mao, "Generative adversarial network–based real-time temperature prediction model for heating stage of electric arc furnace," *Trans. Inst. Meas. Control*, vol. 44, no. 8, pp. 1669–1684, Oct. 2021, doi: 10.1177/01423312211052213.

[12] J. Niu, Z. Liu, Q. Pan, Y. Yang, and Y. Li, "Conditional self-attention generative adversarial network with differential evolution algorithm for imbalanced data classification," *Chin. J. Aeronaut.*, vol. 36, no. 3, pp. 303–315, Mar. 2023, doi: 10.1016/j.cja.2022.09.014.

[13] M. Sajjad, F. Ramzan, M. U. G. Khan, A. Rehman, M. Kolivand, S. M. Fati, and S. A. Bahaj, "Deep convolutional generative adversarial network for Alzheimer's disease classification using positron emission tomography (PET) and synthetic data augmentation," *Microsc. Res. Technique*, vol. 84, no. 12, pp. 3023–3034, Jul. 2021, doi: 10.1002/jemt.23861.

[14] Z.-Y. Yu and T.-J. Luo, "Research on clothing patterns generation based on multi-scales self-attention improved generative adversarial network," *Int. J. Intell. Comput. Cybern.*, vol. 14, no. 4, pp. 647–663, Oct. 2021, doi: 10.1108/ijicc-04-2021-0065.

[15] Y. Fan, G. Wen, F. Xiao, S. Qiu, and D. Li, "Detecting anomalies in videos using perception generative adversarial network," *Circuits, Syst., Signal Process.*, vol. 41, no. 2, pp. 994–1018, Feb. 2022, doi: 10.1007/s00034-021-01820-8.

[16] B. Fang, M. Jiang, J. Shen, and B. Stenger, "Deep generative inpainting with comparative sample augmentation," *J. Comput. Cognit. Eng.*, vol. 1, no. 4, pp. 174–180, Sep. 2022, doi: 10.47852/bonviewjcce2202319.

[17] S. Yang, E. Y. Kim, and J. C. Ye, "Continuous conversion of CT kernel using switchable CycleGAN with AdaIN," *IEEE Trans. Med. Imag.*, vol. 40, no. 11, pp. 3015–3029, Nov. 2021, doi: 10.1109/TMI.2021.3077615.

[18] Y. Yang and X. Song, "Research on face intelligent perception technology integrating deep learning under different illumination intensities," *J. Comput. Cognit. Eng.*, vol. 1, no. 1, pp. 32–36, Jan. 2022, doi: 10.47852/bonviewjcce19919.

[19] A. Jabbar, X. Li, and B. Omar, "A survey on generative adversarial networks: Variants, applications, and training," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–49, Oct. 2021, doi: 10.1145/3463475.

[20] D. Saxena and J. Cao, "Generative adversarial networks (GANs): Challenges, solutions, and future directions," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–42, May 2021, doi: 10.1145/3446374.

[21] A. Kammoun, R. Slama, H. Tabia, T. Ouni, and M. Abid, "Generative adversarial networks for face generation: A survey," *ACM Comput. Surv.*, vol. 55, no. 5, pp. 1–37, Dec. 2022, doi: 10.1145/3527850.

XINXIN NIE was born in Shandong, China, in 1980. She received the Bachelor of Arts and Master of Arts degrees from the Shaanxi University of Science and Technology, Shaanxi, China, in 2004 and 2007, respectively.

From 2007 to 2014, she was a University Teacher with the Engineering and Technical College, Chengdu University of Technology, Sichuan, China. She is currently with the Chengdu Jincheng College, Chengdu, China. She has participated in the writing of two professional books, published over ten academic papers, applied for two school level projects, and received one industry university research project from the Ministry of Education.

JING PU was born in Nanchong, Sichuan, China, in 1985. She received the bachelor's degree in animation and the master's degree in design and art from Sichuan University, China, in 2008 and 2011, respectively, and the Ph.D. degree in design from Sangmyung University, South Korea, in 2023. Since 2011, she has been a Lecturer in product design and digital media art with Sichuan Agricultural University. She has completed 14 related articles.