ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Image Creative Generation and Expressiveness Improvement Based on Deep Learning in Visual Communication Design

Ke Ma[1,*] *and Lin Qi*[2]

[1] School of Art and Design, Zhengzhou University of Economics and Business, Zhengzhou, Henan, 451191, China
[2] HeNan Province Information Consultation Designing Research Co., Ltd., zhengzhou, HeNan, 450003, China

SUMMARY: *In this paper, a progressive creative image intelligent generation method combined with contrast learning is proposed, based on progressive growth generative adversarial network, using progressive growth training method to generate high resolution creative images. The contrast learning loss function based on the target key corner point features is constructed, and the multi-scale image feature loss function is combined to improve the quality of image generation. Based on the multi-dimensional validation of CIFAR-10 and CelebA datasets, explore the applicability of this paper's method in visual communication design. The improved PGGAN algorithm Inception Score reaches 8.91±0.04, significantly outperforming all compared algorithms. When the optimal control parameter $\gamma = 0.5$ is adopted, the improved PGGAN algorithm performs better than the algorithms that introduce the style loss and the contrast learning loss alone. The FID minimum is reached about 40 steps earlier than the original model, and the generation quality and convergence speed are significantly improved. The subjective evaluation shows that this paper's method tops all three scores, with a composite score of 4.51±0.03, and the image generation results are more creative and expressive. The method in this paper has excellent performance in image generation and can provide new possibilities and directions for visual communication design.*

KEYWORDS: *visual communication design; generative adversarial network; progressive generation; image generation*

1 Introduction

Visual communication design is a cross-discipline that synthesizes the fields of art and science, where designers realize visual design through handmade creation and software tools, and communicate effectively through visual elements and information transfer methods, covering a wide range of fields, such as advertising, branding, packaging, user interface design, and so on. In the new media era, designers express their design works more through the transformation of material carriers to the form of immaterial carriers, thus breaking through the conventional media, and their design contents and styles are subconsciously changed, breaking the limitations of conventional material carriers, and bringing more possibilities for the form of design expression [1-3]. In this mode, with the increase in the amount of visual information of the media platform, as well as the integration of Photoshop, Illustrator, InDesign, After Effects and other related design software, the efficiency of the design process has been improved to a

*MK860303@163.com
https://doi.org/10.65102/is2026125 Published: 30April,2026

certain extent, but the design works need to be more rich in expressiveness, but also requires designers to master the tools of the new era on the basis of its traditional products [4-7]. In specific practice, the new art form semantics and performance style approach how to apply still need to be more perfect and iterative, in which most of the designers are affected by the traditional concept of the speed of innovation and development is slow, the effect of the degree of creativity is mainly dependent on the degree of their own understanding of the way to get transformed through communication between designers, brainstorming, etc., the existence of the works of the degree of creativity and lack of expressive power, design efficiency and other low Problems [8-12]. With the development of artificial intelligence technology, new media technology to the age of smart media and burst out more possibilities, deep learning (DL) of the “boom wind” blew into the new media era, bringing a new paradigm of visual communication design.

The most commonly used intelligence technique in the field of image idea generation and visual presentation improvement is generative artificial intelligence. Chopra and Thakur [13] concluded that Generative Adversarial Networks (GAN), Deep Neural Networks, Computer Aided Design of Multiple Generators (Diverse Image Generation), and Neural Style Migration are capable of generating realistic and visually appealing images. Chen et al [14] combined convolutional GAN to synthesize innovative images and introduced style migration to compensate for the shortcomings of generating innovative images that cannot be customized in terms of style, thus generating creative and flexible artworks. Liu [15] describes the capability of Variable Auto-Encoder (VAE) in the field of generating high quality and diverse images, image enhancement, and image denoising, however, VAE may also generate blurred images, which limits its application in the field of image creation. Han et al [16] added a creative concept catalyst to a stable diffusion model to reduce noise interference in the creative image generation process and improve feature extraction performance, allowing the model to generate more creative images and enabling user-friendly creative control.

DL's application of creative image generation in the field of visual communication design. Li and Tang [17] proposed a channel attention module based on Deep Convolutional Neural Networks (CNNs) and Squeeze-and-Excitation Networks for generating descriptive text into an image with more distinctive line features and which can be displayed as a dynamic graphic. Zhao et al [18] formulated a DL-based creative idea generation method that searches for creative images in a database of highly creative image stimuli through typical correlation analysis and CNN, thus providing designers with new creative design thinking. Kong [19] developed a DL model based on improved GAN and VAE, whereas the improved GAN generates more innovative and unique advertisement images and also retains the details and realism, the visual quality, creativity and user engagement of the advertisement images are improved. Gao [20] constructed a hybrid model based on GAN and CNN, which fully utilizes the advantages of GAN (image creativity) and CNN (image structural accuracy), and introduces a style transformation technique to generate a new image with the original style features. Wang and Hong [21] developed an image enhancement technique based on deep neural network and sparse CNN enhancement module, which outputs images with high fidelity, details and textures, and good subjective visualization.

The application of DL for expressiveness in the field of visual communication design. Zhang et al [22] used DL for digital art image style conversion, specifically integrating CNN, adaptive instance normalization, and Gram matrix-based style representation, to optimize the efficiency and quality of style conversion and to enhance creative expression. Hou et al [23] designed a DL-based algorithm for image processing, which can improve the efficiency and accuracy of image creativity in visual communication design, thus improving image quality and

artistic expression. Zheng [24] proposed a DL-based algorithm for artistic style transformation in visual communication, in which a long and short-term memory network classifies style images and content images, captures sequential data and characterizes creative style features to enhance the visual presentation of style-transformed images. Fan [25] extracted visual communication content and image features with the help of a deep CNN model and aesthetically scored the images, introducing migration learning and optimization algorithms to optimize the model, which improved the image aesthetic score (15%) and visual communication effect (20%). Luo and Zeng [26] proposed an enhanced DL-based automatic historical image colorization technique to convert grayscale images into color images, which can be applied to the design of application interfaces in visual communication design to optimize the interface presentation.

In this paper, we first explore the LayoutGAN model architecture in generative adversarial network GAN, specifying its principles. The basic framework of incremental growth generative adversarial network is adopted, and the contrast learning paradigm is introduced to construct the loss function. By measuring the style difference between the generated image and the real image, the consistency between the generated style of the image generation model and the real image style is improved. Based on CIFAR-10 and CelebA datasets, comparative experiments are carried out. Design ablation experiments to verify the effectiveness of the improvement scheme in this paper. Combine the results of visualization analysis and subjective evaluation analysis to assess the image generation performance of the method in this paper.

2 Image generation model construction based on generative adversarial network

2.1 Construction of Generative Adversarial Network Model

2.1.1 Model Architecture

The core element of visual communication design is layout design, which is not only the skillful use of space, but also the pursuit of aesthetics and attractiveness. In recent years, with the development of artificial intelligence technology, the automatic generation technology of images in visual communication design has gradually attracted attention. Generative Adversarial Network (GAN) is an advanced deep learning model, among which LayoutGAN is an attempt to apply GAN to automatic generation of planar layout. In LayoutGAN, the basic architecture of GAN is given a new meaning, which contains two parts: generator and discriminator. The generator is responsible for learning and modeling the features of real layouts to generate new layout designs, and the discriminator is responsible for evaluating the authenticity and quality of the generated layouts, and the two compete with each other during the training process to jointly improve the model performance. Specifically, the generator generates new layout samples by learning the distribution law of the real layout, these samples are sent to the discriminator for evaluation, and the discriminator determines whether these samples are real layouts or generated by the generator, through continuous iteration and optimization, the generator gradually generates more realistic and high-quality layout designs, and the discriminator becomes more sensitive and accurate.The goal of the LayoutGAN is to enable the The goal of LayoutGAN is to make the judgment probability of the discriminator on the generated layout 0.5, that is, to reach the Nash equilibrium state, which means that the quality of the layout generated by the generator and the real layout can no longer be distinguished, so as to realize the automatic generation of the layout design, and the application

of this technology will greatly promote the innovation and development of the field of visual communication design.The basic model architecture of GAN is shown in Fig. 1.

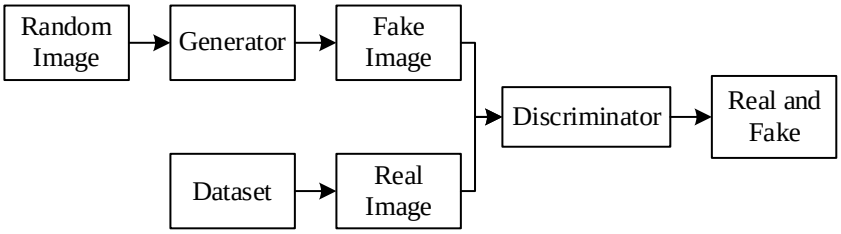


Figure 1: Architecture of generate adversarial network model

GAN shows significant advantages in processing unstructured data such as images, but the traditional GAN approach, i.e., directly taking layout images as input, has some limitations, which often makes it difficult to capture the complex dependencies between layout elements and unable to accurately recognize the patterns of the layouts, and thus the quality of the generated layouts is often unsatisfactory. To overcome these problems, LayoutGAN innovatively chooses to deal with structured data by introducing two unique discriminator mechanisms, namely the relational discriminator and the wireframe rendering discriminator. The relational discriminator focuses on evaluating whether the relationship between elements in the generated layout conforms to the laws of the real layout, while the wireframe rendering discriminator is responsible for transforming the generated layout into a visualized image for more intuitive comparison and evaluation, and the combination of these two discriminators further improves the fidelity and accuracy of the layout generated by LayoutGAN. In the LayoutGAN framework, the generator plays a central role in generating the corresponding layout map based on the input element class probabilities and geometric parameters, a process in which the generator needs to deeply understand and deal with the relationships between elements and accurately control the geometric parameters, to ensure that the generated layouts meet the design requirements as well as have a good visual effect. By introducing these innovative mechanisms, LayoutGAN not only improves the efficiency and effect of the automatic generation of graphic layout, but also provides new ideas and methods for the research in this field, so that designers can more efficiently and accurately generate graphic layouts that meet the requirements, and promote the development of graphic design.

2.1.2 Principles

At the heart of generative adversarial networks lies a dynamic game between two key components, the generator G and the discriminator D. In Gan's structure, the role of the generator G is to generate realistic data samples $G(Z)$ from the random noise Z input to G. Make the generated data samples $G(Z)$ as deceptive as possible to the discriminator D, making it difficult for it to distinguish the difference between the generated data and the real data. The generator G is designed to simulate the distribution of real data, and as training proceeds, the quality of the generated data gradually improves and keeps approaching the real data. On the other hand, the role of D is mainly used to differentiate between the generated data and the real data, receiving the input data x and outputting a probability value $D(x)$, which reflects the confidence level that it recognizes x as the real data. When $D(x)$ is close to 1, it indicates that the discriminator is highly confident that x is real; while when $D(x)$

is close to 0, it indicates that the discriminator believes that x is likely to be a forgery. The task of the discriminator D is to continuously improve its discriminative ability to more accurately recognize the fake data generated by the generator G. As the training iterates, the generator G gradually converges to the real one, while the discriminator D's discrimination accuracy also improves. This adversarial training mechanism drives both sides to optimize continuously until the ideal state is reached: the generator G is able to create data that is sufficiently fake to make it difficult for the discriminator D to make a clear judgment of authenticity in the face of such generated data, i.e., $D\left(G\left(Z\right)\right)$ is close to 0.5, which indicates that the discriminator's judgment of authenticity for the authenticity judgment of the generated data reaches the maximum uncertainty. After the training is completed, the resulting generative model G has a strong data generation capability, and is able to generate high-quality and diverse data samples based on the learned data distribution. Its structure is shown in Figure 2.

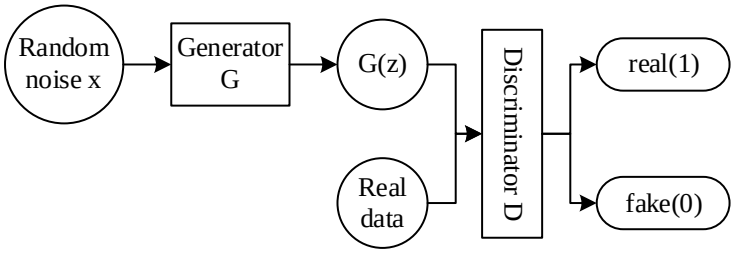


Figure 2: Generative adversarial network structure

During the training of the generative adversarial network, a unique phased strategy is used to ensure that the capabilities of the generator and the discriminator are improved simultaneously. It starts with the first stage of training, which focuses on optimizing the discriminator D. In this phase, each sample in the real dataset x is input into the discriminator D one by one, and then D will output a probability value between 0 and 1. This probability value intuitively reflects the discriminator's level of confidence in the authenticity of the input sample x, with a value close to 1 meaning that the discriminator D has a high probability of considering the sample to be real data. The core objective of this stage is to maximize the discriminator D's accuracy in recognizing real data through iterative training, that is, to make the output value of D as close as possible to the ideal value of 1. This is followed by the second stage of training, where the generator G and discriminator D are involved in the training together, forming a dynamic adversarial mechanism. In this phase, a random noise vector z is first input into the generator G. Based on the probability distribution learned from the real data set, G transforms this noise vector z into spurious samples $G\left(z\right)$. These generated spurious samples are then fed to the discriminator D for evaluation. At this point in time, the task of the discriminator D is to categorize these false samples as 0 as much as possible, i.e., to reduce the level of confidence in the truthfulness of these false samples. This interactive process actually motivates the discriminator D to perform a careful binary classification task in the framework of supervised learning: labeling the true data as 1 and the false data generated by the generator G as 0. The training process is shown in Figure 3.

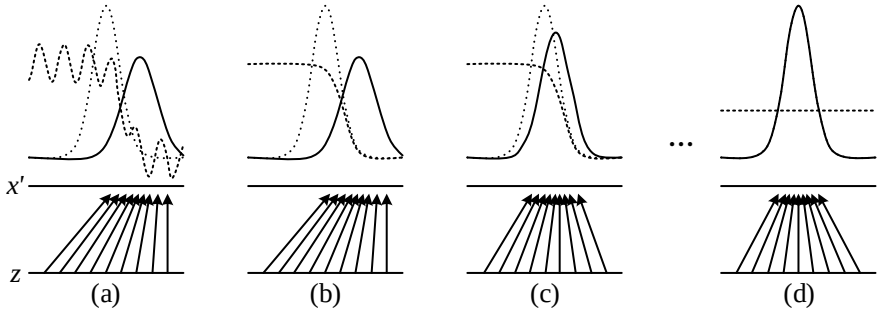


Figure 3: Training process

As can be visually recognized from the image, the black dashed line depicts the real data distribution P_{data}, while the green solid line reflects the distribution P_G simulated by the generative model G. Meanwhile, the blue dashed line corresponds to the distribution characterized by the discriminator D. The horizontal line at the bottom of the figure identifies the interval of uniformly sampled noise z, while the horizontal line at the top corresponds to the range of data x, and the upward arrows indicate the mapping of random noise z to data $x = G(z)$ through the generative model G. The convergence process of Gan is demonstrated from Fig. a to Fig. b, where the generating distribution P_G has shown some proximity to the real data distribution P_{data} at the stage of Fig. a. D is a partially accurate classifier, fixing G in Figure b while updating D converges to $D^*(x) = p_{data}(x)/(p_{data}(x) + p_G(x))$. In Figure c, the iterative update of the generator G prompts the gradient signal of the discriminator D to play a key role in guiding the evolution of $G(Z)$ towards a region closer to the real data. With the advancement of the training process, in Figure d, when the generator G and the discriminator D have been trained for many rounds and have a strong performance, they will gradually reach an equilibrium state. At this time, the distribution P_G simulated by the generative model G reaches a high degree of consistency with the real data distribution P_{data}, and it is difficult for the discriminator D to effectively distinguish the difference between the real data and the generated data anymore. The expression is shown in equation (1).

$$\min \max(D, G) = E_{x \sim P_{data}}[\log(D(x))] + E_{x \sim P_G}[\log(1 - D(G(z)))] \tag{1}$$

As a powerful generative model, the core mechanism of Generative Adversarial Network is to efficiently simulate complex data distributions and generate high-quality samples through adversarial training of generators and discriminators. Unlike traditional methods, Gan does not need to rely on Markov chains for tedious repeated sampling, and avoids the complex inference process and the calculation of variational lower bounds in the learning phase, effectively solving the problem of approximate computation of high-dimensional probability distributions, and showing significant theoretical and practical advantages. Although Gan has many advantages, it still faces challenges in practical applications. The first one is the convergence and stability of the model, and Gan training is essentially a dynamic game, which requires a highly synchronized balance between the generator (G) and the discriminator (D). However, this balance is difficult to maintain in actual training, and it is easy for D to converge quickly and G to diverge, or for G to converge but D to fail. In addition, missing patterns are also a common problem. The generator may be degraded during training, generating only a limited

number of repetitive samples, which cannot cover all the patterns and features of the data distribution, severely limiting the diversity and realism of the generated data, and hindering the performance of Gan in complex and diverse scenarios.

2.2 Progressive Target Creative Image Generation Method Combining Contrastive Learning

2.2.1 Asymptotic growth generating adversarial network structure

Progressive Growth Generative Adversarial Networks (PGGANs) are derived from classical generative adversarial networks and thus have a similar underlying framework, with the difference that the network structure of a PGGAN is updated synchronously with the stage of training: the generator progressively puts more and more networks into training through training, generating creative images ranging from low- to high-resolution; the discriminator similarly increases the number of network layers put into training with the resolution of the input image; the discriminator similarly increases the number of network layers put into training with the resolution of the input image. the discriminator similarly increases the number of network layers put into training as the resolution of the input image increases. This progressive training mechanism allows the model to first learn the characteristics of the data at low resolution, and then gradually turn its attention to finer sizes, which can effectively stabilize the training process.

(1) Generator Network Structure

The overall structure and training process of the PGGAN generator is shown in Figure 4. The overall structure of the generator is divided into two parts: encoder and decoder.The implementation of the equalized learning rate in PGGAN includes the following steps:

1) Learning rate initialization: the network adopts a simple $N(0,1)$ initialization method;

2) Dynamic scaling parameter calculation: the weight parameters of each layer are dynamically scaled during the training process, and the adjusted parameters are represented as:

$$\hat{w}_i = \frac{w_i}{c_i} \tag{2}$$

where w_i denotes the weight parameter and c_i denotes the normalization constant, which varies depending on the chosen initializer;

3) Gradient update: The backpropagation uses weight parameters that have been dynamically scaled to maintain a consistent rate of parameter update for each layer.

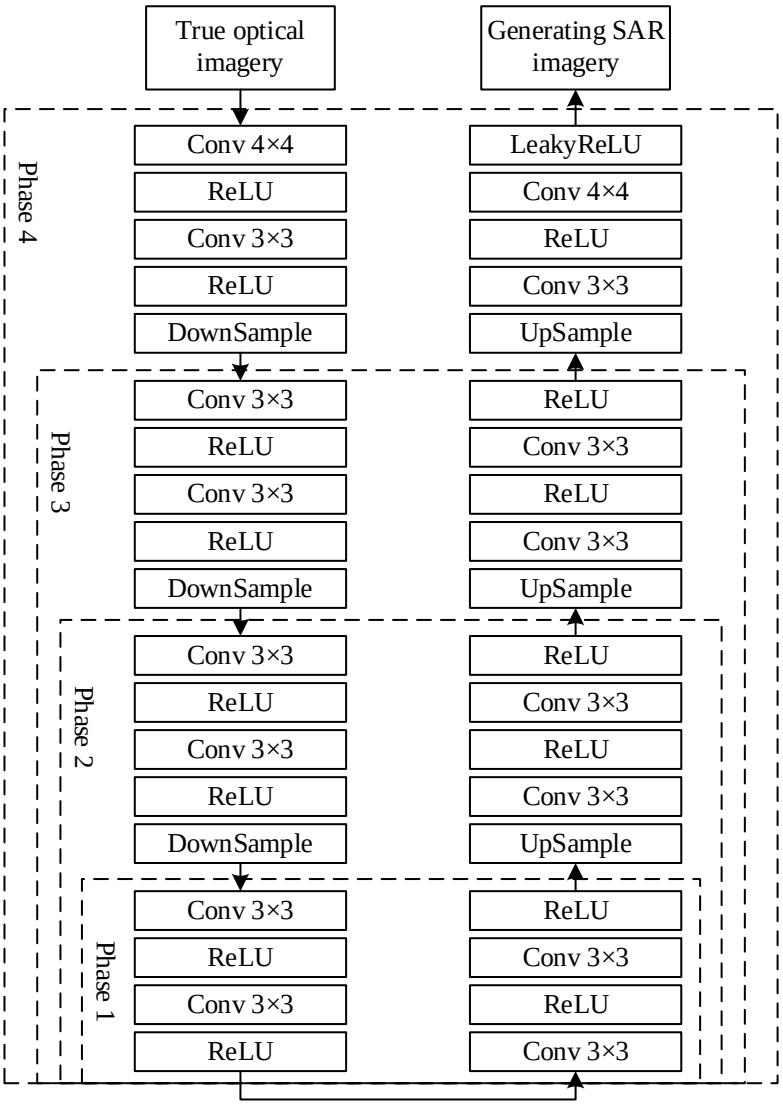


Figure 4: Structure of PGGAN generator

Let the feature vector be $v_i = (v_{i1}, v_{i2}, \ldots, v_{ic})$, where c denotes the number of channels, and the pixel-by-pixel normalized feature can be expressed as:

$$\hat{v}_i = \frac{v_i}{\sqrt{\frac{1}{c}\sum_{j=1}^{c} v_{ij}^2 + \delta}} \tag{3}$$

where $L2$ is normalized to the feature vector of each pixel point of the feature image, and δ represents the divide-by-zero constant, which is usually a very small value.

During the progressive network growth phase, PGGAN uses an alpha-smoothing transition mechanism in order to prevent the oscillations that occur when new networks are added directly to the old ones.An example of the PGGAN encoder and decoder module transition is shown in Fig. 5. The transition process constructs a structure similar to a residual network, treats the newly added network as a structure similar to a residual block, and introduces a weighted summation of alpha coefficients. The weighted output features can be expressed as:

$$x = (1-\alpha) \cdot x_{old} + \alpha \cdot x_{new} \tag{4}$$

x_{old} denotes the 32×32 feature maps after the output from the old network, and x_{new} denotes the 64×64 feature maps after the output from the newly added network after downsampling. The α denotes the excess factor, which gradually grows linearly from 0 to 1 with the training process.

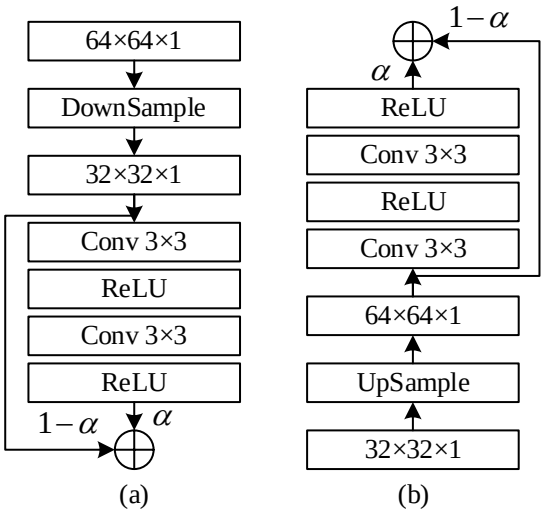


Figure 5: Example of encoder and decoder module transition

(2) Discriminator network structure

The structure of the network discriminator is shown in Figure 6. Equalized learning rate, pixel-by-pixel normalization, and network smoothing transition mechanism are also used in the construction and training process of the discriminator network.

In order to prevent the network from pattern collapse, a small number of standard deviations are used in the last module of the discriminator to regularize the data. This is implemented by first computing the pixel-by-pixel standard deviation of the feature map notation, and later broadcasting this standard deviation to the full feature map size and splicing it with the original feature map as a statistical feature of the data.

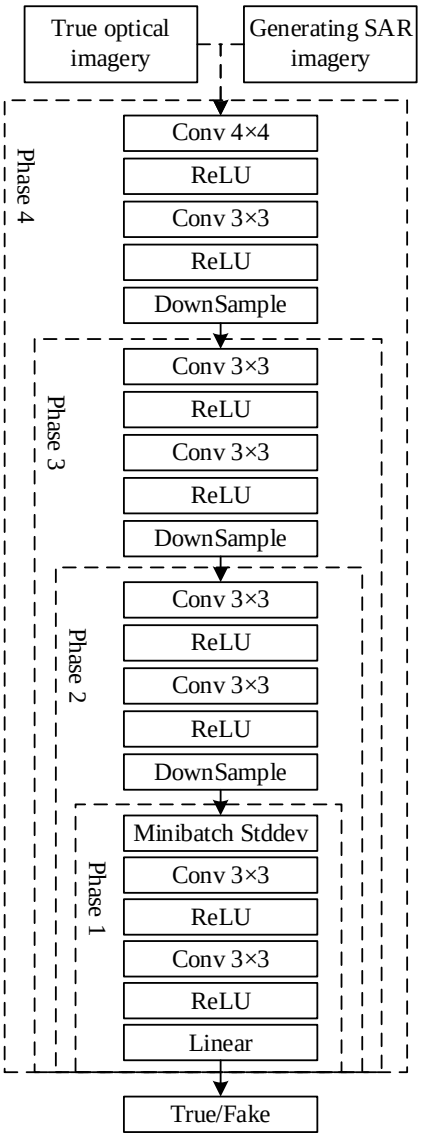


Figure 6: Structure of PGGAN discriminator

2.2.2 Loss function

(1) Contrastive learning loss function

Contrast learning can be used to learn the general features of the data without labeling, contrast learning loss function construction is mainly divided into the following steps: positive/negative sample pair construction, feature extraction and contrast function design.

1) Positive/negative sample pair construction: Assume that the real creative image dataset is $S=\{s_1,s_2,\ldots,s_n\}$ and the generated creative image dataset is $X=\{x_1,x_2,\ldots,x_n\}$. Construct positive-negative sample pairs for the generated image x_i: for the generated creative image x_i, the real creative image s_i is a positive sample and the other real creative images are negative samples. In other words, for x_i, the positive sample pair is (x_i,s_i) and the negative sample pair is $\{(x_i,s_j)\,|\,k\neq i\}$.

2) Feature extraction: Harris corner detection is performed on the image, and the Harris corner determination formula is expressed as:

$$\begin{aligned} R &= \det(M) - k \cdot trace^2(M) \\ M &= \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} \\ I_x &= \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y} \end{aligned} \tag{5}$$

where I_x and I_y denote the gradient of the image in the x and y directions respectively, M is the covariance matrix of the image, $\det(\cdot)$ denotes the matrix determinant computation, $\text{trace}(\cdot)$ denotes matrix trace computation, and k denotes the empirical constant.

After calculating to get R it is determined whether the point is a corner point or not according to the value of R. When a point has $R > 0$ it is considered to be a corner point, $R < 0$ is an edge and $R \approx 0$ is a flat region. Finally, the filtered corner points are suppressed by non-extremely large values, and the weak corner points are removed to get the final corner features of the image.

3) Comparison learning function design: the loss function for positive samples (x_i, x_j) is defined as follows:

$$L_{cl} = -\log \frac{\exp(d(x_i, s_i)/\tau)}{\sum_{k=1, k\neq i}^{N} \exp(d(x_i, s_j)/\tau)} \tag{6}$$

where $d(\cdot)$ calculates the similarity between features; τ denotes the temperature coefficient. From the loss function, it can be seen that the sum of similarity between the generated creative image x_i and all the mismatched real creative image feature points.

(2) Style loss function

In order to make the target binding structure of the generated image similar to the real image while the overall image style is closer to the creative image, the method in this paper adds a style loss function when constructing the loss function. The style loss function further guides the network to generate images that are closer in style to the real image by measuring the style difference between the generated image and the real image. The style loss usually contains three steps: feature extraction, Gram matrix construction and loss calculation.

1) Feature extraction: Assuming that the real creative image is S_r and the generated creative image is S_g, the feature $F(S_r)$ obtained from the real creative extraction and the feature $F(S_g)$ obtained from the generated creative image extraction;

2) Gram matrix calculation: the inner product is also called the dot product of vectors, for two vectors $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$, its inner product calculation formula is expressed as:

$$a \cdot b = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n \tag{7}$$

The matrix consisting of the two-by-two inner product between any k vectors in an n-dimensional Euclidean space is the Gram matrix of these k vectors, which can be expressed as:

$$\Delta(\alpha_1,\alpha_2,\ldots,\alpha_k)=\begin{pmatrix}(\alpha_1,\alpha_1) & (\alpha_1,\alpha_2) & \ldots & (\alpha_1,\alpha_k)\\ (\alpha_2,\alpha_1) & (\alpha_2,\alpha_2) & \ldots & (\alpha_2,\alpha_k)\\ \ldots & \ldots & \ldots & \ldots\\ (\alpha_k,\alpha_1) & (\alpha_k,\alpha_2) & \ldots & (\alpha_k,\alpha_k)\end{pmatrix} \tag{8}$$

For the feature map with shape $H\times W\times C$ output from the feature extraction network, the feature map is first reshaped into a two-dimensional feature map with shape $(H\times W)\times C$, and then the Gram matrix for each feature map of the real creative image and the generated creative image is obtained by calculating them separately according to Eq:

$$G=\tilde{F}\cdot\tilde{F}^{*}\in\square^{C\times C} \tag{9}$$

3) Loss calculation: the style loss function of the lth layer feature map is measured by calculating the $L2$ paradigm of the Gram matrix of the real and generated images, which is expressed as:

$$L_{sty}^{l}=\frac{1}{4C^{2}(HW)^{2}}\sum_{i,j}\left(G_{ij}^{l}(F^{l}(S_r))-G_{ij}^{l}(F^{l}(S_g))\right)^{2} \tag{10}$$

where $1/4C^{2}(HW)^{2}$ denotes the normalization factor and $F^{l}(S_r)$ is used to balance the contribution of different layers to the loss function. The multi-layer style losses are weighted to obtain the final style loss function:

$$L_{sty}=\sum_{l}\lambda_l\cdot L_{sty}^{l} \tag{11}$$

where λ_l represents the lth layer style loss function weight coefficient.

(3) Overall loss function

The overall loss function of the method in this paper mainly consists of two parts: the adversarial loss and the contrast learning loss. The contrast learning loss is shown in the previous section, while the adversarial loss can be divided into generator loss L_G and discriminator loss L_D. The two in the game and competition to optimize their respective capabilities, and ultimately reach the “Nash equilibrium” state, the model convergence. The GAN generator and discriminator adversarial loss can be expressed as follows:

$$\begin{aligned}L_D&=-E_{x\sim P_{data}(x)}[\log D(x)]-E_{y\sim P_y(y)}[\log(1-D(G(y)))]\\ L_G&=-E_{y\sim P_y(y)}[\log D(G(y))]\end{aligned} \tag{12}$$

The overall loss function of the method in this paper can be expressed as:

$$L=\lambda_{adv}(L_D+L_G)+\lambda_{cl}L_{cl}+\lambda_{sty}L_{sty} \tag{13}$$

where λ_{adv}, λ_{cl} and λ_{sty} are hyperparameters, respectively.

3 Research on the applicability of image generation model in visual communication design

3.1 Comparative Experimental Analysis

For the CIFAR-10 dataset, an Inception classification network is first constructed, and the Inception Score of multiple algorithms are tested using the distribution of this network, and the experimental results are shown in Table 1. DCGAN, BEGAN, WGAN, etc. are superior to the traditional generative adversarial algorithms, but the gap is obvious with the asymptotic generative algorithms. The algorithm in this paper significantly outperforms the comparison algorithms with an Inception Score of 8.91±0.04, which is about 23.24% better than the next best performing PGGAN, and the standard deviation is more minimized, i.e., the generation quality and stability are both optimal.

Table 1: Comparison of generation quality on the CIFAR-10 dataset

Network framework	Inception Score (64×64)
DualGAN	4.98±0.09
cGAN	5.32±0.07
DCGAN	6.93±0.05
BEGAN	6.56±0.08
WGAN	6.89±0.11
PGGAN	7.23±0.09
The proposed	8.91±0.04

For the CelebA dataset, an Inception classification network is constructed to measure the score of the generated images belonging to the expected categories. The control parameter γ is used to regulate the balance between image authenticity and diversity, and the experimental results are shown in Table 2. The Inception Score generally decreases with the increase of resolution, but this paper's algorithm outperforms PGGAN in all resolutions.The moderate control of γ can effectively optimize the generative performance, and the control parameter γ=0.5 achieves the best balance, and the Inception Score of this paper's algorithm achieves 12.43, 9.25, 256×256, 256×256, and 12.43, 9.25, respectively. The Inception Score of this paper's algorithm reaches 12.43 and 9.25 at resolutions of 128×128 and 256×256, respectively.

Table 2: Comparison of generation quality on the CelebA dataset

Network framework	Image generation resolution	Inception Score
PGGAN	128×128	10.12
The proposed(γ=0.5)	128×128	12.43
The proposed(γ=0.7)	128×128	11.75
The proposed(γ=0.9)	128×128	10.46
PGGAN	256×256	5.34
The proposed(γ=0.5)	256×256	9.25
The proposed(γ=0.7)	256×256	8.31
The proposed(γ=0.9)	256×256	7.58

3.2 Analysis of ablation experiments

The same Inception classification network is used to perform ablation experiments on the

CelebA dataset, and the improved PGGAN adopts the optimal control parameter $\gamma = 0.5$, and its experimental results are shown in Table 3. The ablation experiments show that both the introduced style loss and contrast learning loss can improve the performance of PGGAN, and the Inception Score increases from 10.12 to 12.43 when the resolution is 128×128, which is close to the 14.00 of the real image.The resolution of 256×256 still remains the leading position, which verifies that the contrast learning loss and the style loss introduced in the generation process are Effectiveness.

Table 3: Comparison of generation quality under different control parameters γ

Network framework	Image generation resolution	Inception Score
Real image	128×128	14.00
PGGAN	128×128	10.12
PGGAN+StyLoss	128×128	11.43
PGGAN+ContrastiveLoss	128×128	11.58
The proposed	128×128	12.43
Real image	256×256	10.00
PGGAN	256×256	5.34
PGGAN+StyLoss	256×256	7.82
PGGAN+ContrastiveLoss	256×256	8.04
The proposed	256×256	9.25

The network training loss function curve of the improved PGGAN on the CelebA dataset is shown in Fig. 7. The discriminator loss drops rapidly from 2.60, then gradually rises and stabilizes around 1.33, accompanied by slight fluctuations, indicating that the discriminator and generator reach a dynamic balance in the confrontation. The generator loss decreases from 2.31 to around 1.35 and finally stabilizes around 0.72, reflecting the continuous improvement of the generator's ability to generate realistic images. The gradient penalty loss decreases rapidly from 2.13 to 0.24 and finally converges to 0 and remains stable, indicating that the gradient paradigm is close to 1, the discriminator satisfies the Lipschitz constraint, and the training process is stable. The rotation classification loss for real samples starts from 1.43 and drops rapidly to around 1.08 (about 8k steps), indicating that the generator learns quickly on the self-supervised rotation task, and the generated images start to have some geometric consistency. The performance is smooth in the later stages, with the loss stabilizing around 1, indicating that the generator's performance on the rotation classification task is gradually stabilizing. There are some fluctuations during the descent, between 0.9 and 1.1, with a small range. In addition, there are no consistently high values or sharp fluctuations, indicating that the generator's training on the self-supervised task is stable. The rotational classification loss of the generated samples decreases rapidly from 1.43 to 0.5, indicating that the discriminator learns quickly on the self-supervised rotational task and the classification performance improves rapidly. The mid-term fluctuation occurs between 8k and 35k steps, where the loss fluctuates (0.4 to 0.6) because the quality of the images generated by the generator is improving and the discriminator needs to keep adjusting to more complex adversarial environments. Later on, starting at 35k steps, the loss gradually stabilizes around 0.5, indicating that the performance of the discriminator on the rotational classification task has stabilized and the classification accuracy is high.

Overall, the training process performs well on the gradient-constrained and self-supervised tasks, indicating that the model is stable on the CelebA dataset, and the self-supervised rotational task effectively improves the generator's learning ability without pattern collapse, which provides a guarantee for generating high-quality images.

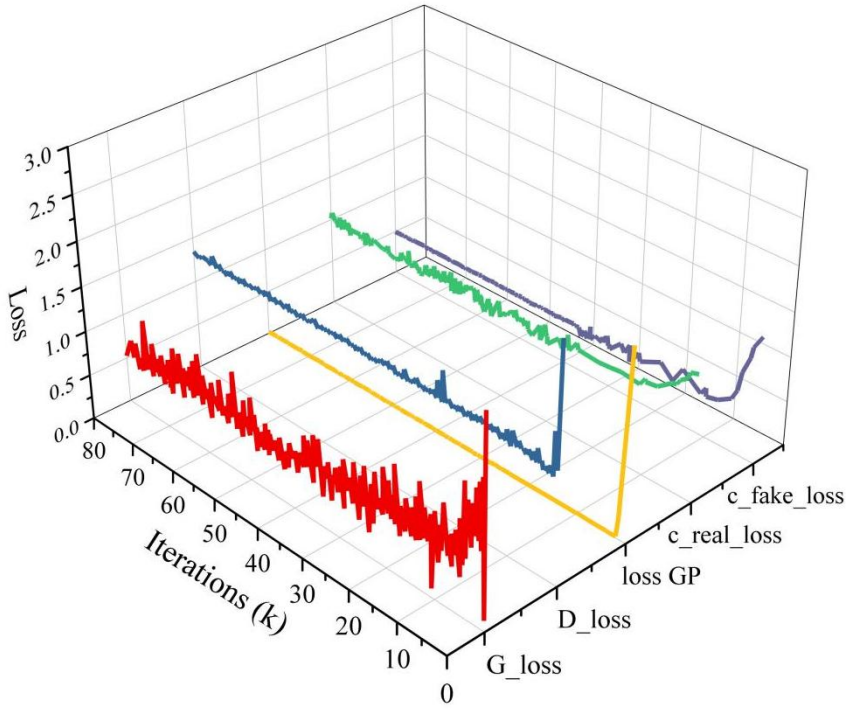


Figure 7: Network training loss function curve

Comparison of the loss function curves of the generator and discriminator before and after the improvement of PGGAN is shown in Fig. 8. The generator of PGGAN decreases rapidly in the early stage of training, but the fluctuation is large, and there is still an unstable jitter after convergence to 0.432, which indicates that the generator generates images with poor continuity. The loss curve of the improved PGGAN generator declines more smoothly, with significantly reduced oscillations, and converges around 0.378, indicating that the loss function effectively suppresses drastic pixel-level variations in the generated image and promotes the continuity and consistency of the image.Although PGGAN rapidly improves the discriminative ability at the initial stage, its loss oscillations are larger, and ultimately fluctuate around 1.256, which may indicate that the discriminator over fitted the local pseudo-differences in the image. The curve of the improved PGGAN discriminator rises slightly slower, but fluctuates less, and stabilizes around 1.147 after convergence, reflecting its ability to discriminate changes in the details of the generated image with more generalization. This verifies the key role of the loss function in this paper in improving the generation quality and structural similarity.

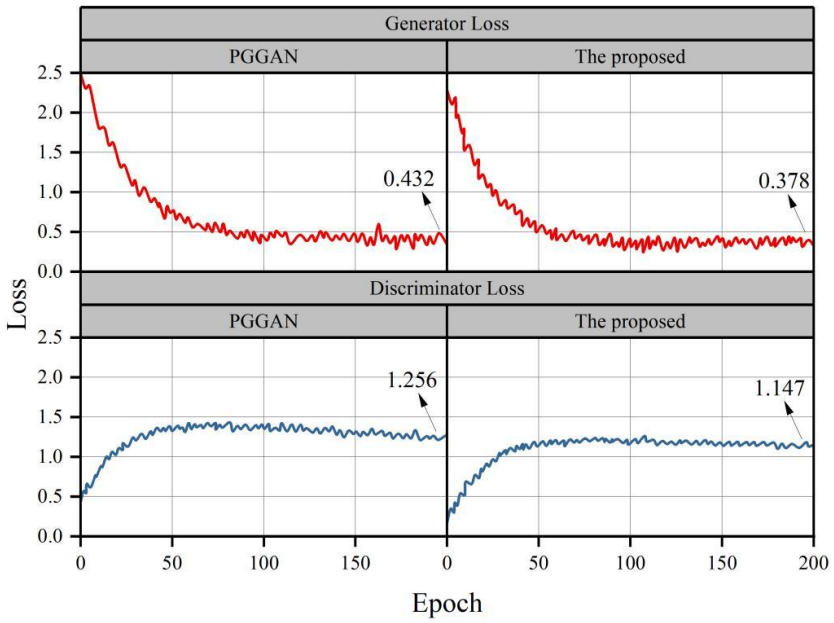


Figure 8: Comparison of loss function curves before and after PGGAN improvement

FID is computed to quantify the similarity of the generated image to the real image distribution as it comprehensively evaluates the quality of generation through the statistical distances (mean and covariance) in the feature space of Inception V3, which is particularly suitable for capturing distributional bias and consistency of details in complex scenes.The decreasing trend of FID before and after the improvement of PGGAN is shown in Fig. 9. At the beginning of training, both PGGAN and improved PGGAN have higher FID scores. As the number of iterations increases, the FID scores of both show a decreasing trend. However, the number of iterative steps required for the improved PGGAN to reach a lower FID value is significantly less than that of the PGGAN. Specifically, PGGAN takes about 120 iterations to reach its own lowest FID value, while the improved PGGAN reaches its lowest level in about 80 iterations. This indicates that the improved PGGAN converges faster during the optimization process and is able to obtain better generation results and lower FID values than PGGAN with fewer iteration steps.

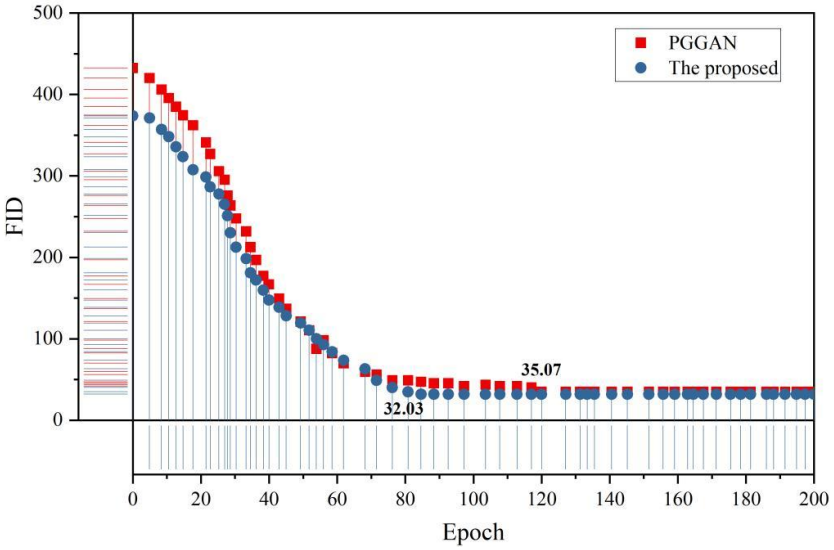
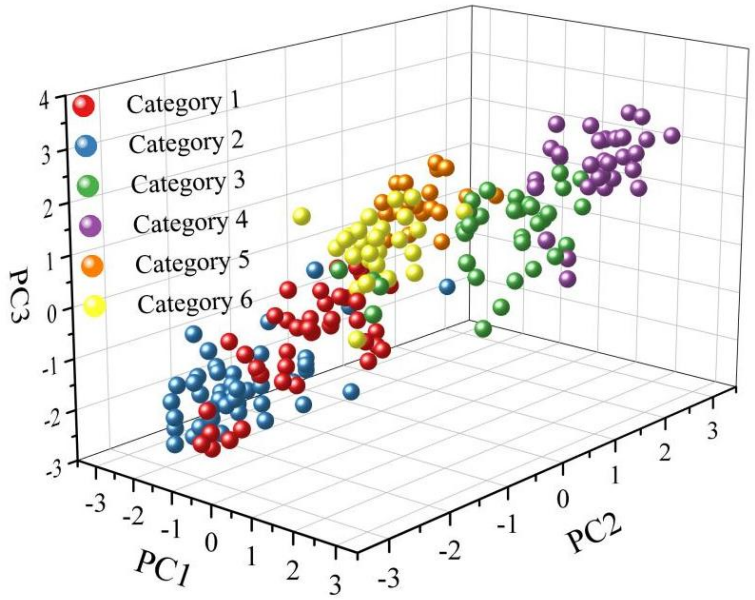


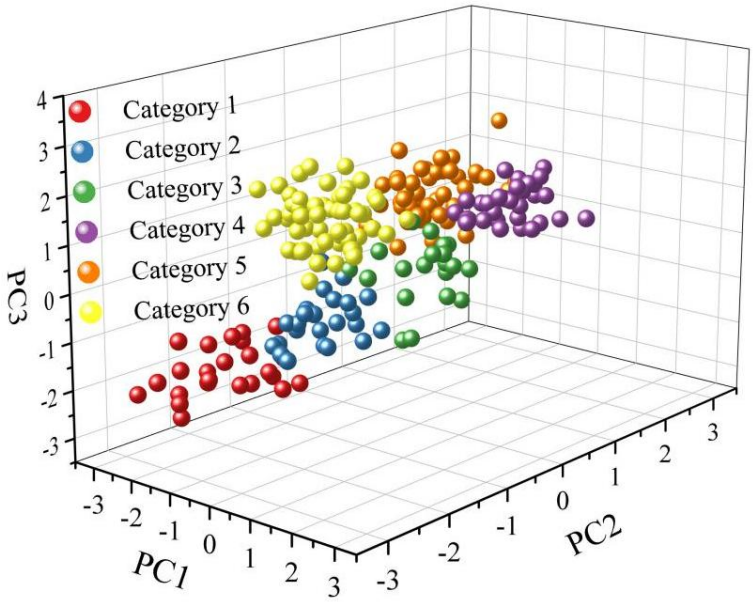
Figure 9: FID decline trend before and after PGGAN improvement

3.3 Visualization and analysis

The visualization analysis was performed on the CelebA dataset, and the results are shown in Fig. 10 (a~b). PGGAN and improved PGGAN are applied to 200 test images respectively, and different colors in the figure represent 6 classes, which visually demonstrates the comparison of the classification effects of PGGAN and improved PGGAN. The representations obtained by this paper's algorithm have clear inter-class boundaries and compact intra-class aggregation, while PGGAN suffers from feature dispersion and noise residuals.



(a)PGGAN



(b)The proposed

Figure 10: visualized analysis results

3.4 Subjective evaluation analysis

In order to obtain user preferences for different methods of generating image results, 100 college students majoring in visual communication were randomly invited to serve as subjects. These subjects had normal or corrected normal vision and were not informed in advance of the specific purpose of the experiment to minimize potential bias against the experiment. Subjects were asked to rate the images on a scale of 1 to 5. The results of the subjective ratings are shown in Table 4. Subjects rated the creativity (4.23 ± 0.05), expressiveness (4.34 ± 0.02), and composite score (4.51 ± 0.03) of the images generated by the method in this paper as the highest with minimal standard deviation, indicating that subjects were most satisfied with the quality of the images generated by the method in this paper and that there was little variance in the ratings.

Table 4: Results of subjective evaluation

	Creativity	Expressiveness	Overall score
DualGAN	2.51±0.17	2.45±0.12	2.38±0.09
cGAN	3.16±0.22	3.29±0.14	3.19±0.11
DCGAN	3.78±0.15	3.67±0.08	3.74±0.13
BEGAN	3.45±0.13	3.58±0.09	3.63±0.11
WGAN	3.62±0.14	3.76±0.11	3.71±0.12
PGGAN	3.81±0.09	3.98±0.06	4.04±0.07
The proposed	4.23±0.05	4.34±0.02	4.51±0.03

In addition to this, a preference experiment was set up to more visually compare the subjects' preferences between this paper's method and several other methods. The preference experiment is based on giving the specified reference image and content image, and letting the subjects select the image between this paper's method and some comparative methods to generate the result image with better effect. The results of the preference experiment between this paper's method and different methods are shown in Fig. 11. The selection rate of this paper's method in each comparison group is above 70%, which intuitively reflects that it is more preferred in the visual communication professional group.

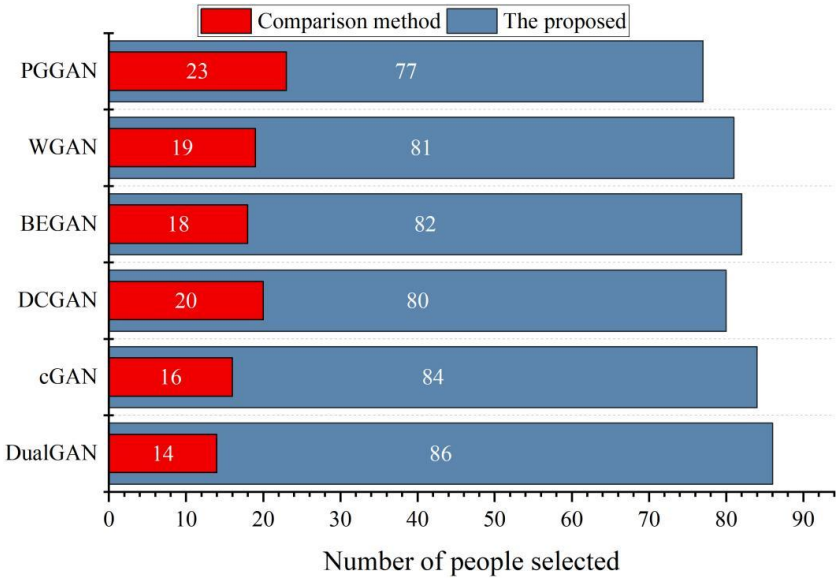


Figure 11: Results of the Preference Experiment

4 Conclusion

Based on generative adversarial network, this paper constructs an improved progressive target creative image generation model and explores its applicability in visual communication design through experiments.

(1) In the comparison experiments, the algorithm in this paper achieves an Inception Score of 8.91±0.04, which is ahead of the existing mainstream methods, and is about 23.24% higher than that before improvement. Under different resolutions and different control parameters, the algorithm in this paper maintains a higher Inception Score than PGGAN and reaches the optimization at $\gamma = 0.5$, which is 22.83% and 73.22% higher than PGGAN, respectively.

(2) In the ablation experiment, the introduction of style loss and contrast learning loss alone can improve the generation effect of PGGAN, and the Inception Score under 128×128 resolution is improved to 11.43 and 11.58, respectively, while the algorithm in this paper fuses the two Inception Score to reach 12.43, which is close to the level of the real image.256×256 The improvement is even greater at 256×256 resolution, which is only 7.5% lower than the real image gap. On the CelebA dataset, the improved PGGAN discriminator loss is reduced from 2.60 to about 1.33, the generator loss is reduced from 2.31 to 1.35, and the gradient penalty loss is reduced from 2.13 to 0.24 and stabilized, which satisfies the Lipschitz constraint. The rotational classification loss of real samples is reduced from 1.43 to 1, and the generator loss of generated samples is reduced from 1.43 to 0.5. The improved model has less fluctuation of generator loss, more stable convergence, and smoother discriminator curves than the original PGGAN, and the FID evaluation reveals that the original model needs about 120 steps to reach the minimum, while the improved model has only about 80 steps, which is faster convergence and higher quality, and verifies the validity of the loss function in improving the generator effect.

(3) In the visualization analysis, both PGGAN and improved PGGAN can classify 200 test images into 6 categories. The algorithm in this paper can better learn the discriminative representation of the categories, effectively extract relevant information and discard redundant information.

(4) In the subjective evaluation dimension, this paper's method has an advantage of 11.02%, 9.05%, and 11.63% in the three dimensions of creativity, expressiveness, and overall score, respectively, compared to the PGGAN with sub-optimal performance. The preference experiment further verified that subjects preferred the resultant images generated by this paper's method over the comparison methods.

About the Author

Ke Ma was born in Xuchang City, Henan Province in 1986. She graduated from Xi'an University of Architecture and Technology with a master's degree. Her main research directions include visual communication design and digital media art.

References

[1] Guo, H. (2022). Innovation and Development of Visual Communication Design in the New Media Era. Innovation, 3(3), 2022.

[2] Nong, L., & Long, B. (2023, January). Visual Communication in New Media Art Design. In International Conference on Innovative Computing (pp. 604-610). Singapore: Springer Nature Singapore.

[3] Soreanu, C. (2020). From media to mediums of expression. visual art communication and meaning from fine arts to advertising. Anastasis Research in Medieval Culture and Art, 7(2), 261-276.

[4] Guan-Chen, L., & Ko, C. H. (2021). Photoshop and illustrator use in instructional design: a case study of college visual communication course design. The International Journal of Design Education, 15(1), 119.

[5] Hernada, O. E., Sumarno, Jamaaluddin, J., & Sumartik. (2024, July). Enhancing design and print services: A web-based ordering system for indesign digital printing store. In AIP Conference Proceedings (Vol. 3167, No. 1, p. 040025). AIP Publishing LLC.

[6] YLove, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., ... & Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. Journal of Statistical Software, 88, 1-17.

[7] Yu, G., Akhter, S., Kumar, T., Ortiz, G. G. R., & Saddhono, K. (2022). Innovative application of new media in visual communication design and resistance to innovation. Frontiers in Psychology, 13, 940899.

[8] Liang, C., Chang, C. C., & Liu, Y. C. (2019). Comparison of the cerebral activities exhibited by expert and novice visual communication designers during idea incubation. International Journal of Design Creativity and Innovation, 7(4), 213-236.

[9] Saris, B. (2020). A review of engagement with creativity and creative design processes for visual communication design (VCD) learning in China. International Journal of Art & Design Education, 39(2), 306-318.

[10] Anggrianto, C. (2022, November). The Shifting Paradigm of Visual Communication Design Profession. In International Conference on Sustainability in Creative Industries (pp. 175-183). Cham: Springer Nature Switzerland.

[11] Roxburgh, M., & Irvin, J. (2018, July). The future of visual communication design is almost invisible or why skills in visual aesthetics are important to service design. In Proceedings of the Service Design Proof of Concept Conference (ServDes2018) (pp. 199-215). Milano Italy: Politecnico di Milano.

[12] Cao, Q. (2025). AI algorithms in visual communication design: enhancing design creativity and efficiency. International Journal for Simulation and Multidisciplinary Design Optimization, 16, 10.

[13] Chopra, M., & Thakur, A. (2025). Generative AI for the Creation of Images. Intelligent Solutions for Smart Adaptation in Digital Era: Select Proceedings of InCITe 2024, Volume 2, 1278, 193.

[14] Chen, H., Zhao, L., Qiu, L., Wang, Z., Zhang, H., Xing, W., & Lu, D. (2020). Creative and diverse artwork generation using adversarial networks. IET Computer Vision, 14(8), 650-657.

[15] Liu, J. (2025). Research on the Application of Variational Autoencoder in Image Generation. In ITM Web of Conferences (Vol. 70, p. 02001). EDP Sciences.

[16] Han, J., Kwon, D., Lee, G., Kim, J., & Choi, J. (2025). Enhancing creative generation on stable diffusion-based models. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 28609-28618).

[17] Li, Y., & Tang, Y. (2023). Novel creation method of feature graphics for image generation based on deep learning algorithms. Mathematics, 11(7), 1644.

[18] Zhao, T., Yang, J., Zhang, H., & Siu, K. W. M. (2021). Creative idea generation method based on deep learning technology. International Journal of Technology and Design Education, 31(2), 421-440.

[19] Kong, M. (2025). Deep learning model optimization in creative generation for new media animated ads. Discover Artificial Intelligence, 5(1), 94.

[20] Gao, S. (2025). Creative generation and evaluation system of art design based on artificial intelligence. Discover Artificial Intelligence, 5(1), 118.

[21] Wang, Z., & Hong, D. (2024). Visual Communication Design Based on Sparsity-Enhanced Image Processing Models. International Journal of Advanced Computer Science & Applications, 15(5).

[22] Zhang, S., Qi, Y., & Wu, J. (2025). Applying deep learning for style transfer in digital art: enhancing creative expression through neural networks. Scientific Reports, 15(1), 11744.

[23] Hou, X., Liu, Q., & Zhang, X. (2025). Image processing algorithm of visual communication design based on deep learning in digital background. Discover Artificial Intelligence, 5(1), 180.

[24] Zheng, S. (2024). Deep learning-based artistic style transformation algorithm in visual communication. International Journal of System Assurance Engineering and Management, 1-10.

[25] Fan, S. (2025). Image Aesthetics and Visual Communication Optimization Based on Deep Convolutional Neural Network. Procedia Computer Science, 262, 236-243.

[26] Luo, H., & Zeng, Q. (2022). Study on the application of visual communication design in APP interface design in the context of deep learning. Computational intelligence and neuroscience, 2022(1), 9262676.