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

Creative stimulation mechanism of human-computer collaborative interaction in art design creation based on generative artificial intelligence

Shiting Li[1,*]

[1] Jiangxi Institute of Applied Science and Technology, Nanchang, Jiangxi, 330100, China

SUMMARY: *The rise of generative artificial intelligence (AIGC) is observed in many fields. The key benefit of using the technology lies in the increase in efficiency and creativity. The current research aims to explore the mechanism of human-computer collaboration creativity stimulation in generative AI used for the generation of art design. Developing a single-scale generative adversarial network (GAN) called S_OpenGAN intended for human-computer cooperation, the intelligent generation of art creation images is achieved, and the SD method is applied to evaluate the art design creation results. Based on the experimental data, it is concluded that the developed algorithm generates high-quality images with a significant diversity even when there are insufficient amounts of data. Besides, the generated results outperform the ones generated using alternative approaches such as SNI, DAT, TransEditor, and StyleGAN-2. What is more, the inference and memory consumption of S_OpenGAN is less than that of competitive approaches on all platforms, allowing generating images with the help of fewer resources. In addition, the S_OpenGAN algorithm possesses the ability to change the preferences of observers through specific changes in image elements.*

KEYWORDS: *artificial intelligence; generative adversarial network; human-computer interaction; art design creation; semantic difference SD method*

1 Introduction

The very essence of Generative Artificial Intelligence (GAI) involves creation and generation of unique content via computer algorithms and modeling. Unlike conventional supervised learning or classification models, GAI does not just recognize and classify existing data but can also generate new content based on existing information and knowledge. It is precisely due to this unique feature of GAI that it has found a wide variety of applications in many areas [1-3]. At present, the core technologies that are used for GAI include Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Autoencoders (AEs). The most popular and commonly utilized among them is GAN [4]. Since the notion of GAN was coined by Goodfellow et al. in 2014, much improvement and progress has been made in the area of image and natural language generation [5]. However, efforts at improving the performance of GAI continue, with researchers working on increasing the quality and diversity of generated content by optimizing network architecture and training process and by inventing innovative GAI approaches [6-8]. Thus, GAI can be considered a promising technology capable of imitating the processes of creative human thought and behavior and facilitating innovation in

*13184575945@163.com
https://doi.org/10.65102/is2026174
Published: 30 April, 2026

the domain of art and design, among others [9].

Due to its unique innovative application features, GAI has demonstrated significant research value and development prospects in the field of painting art creation. Rani et al. used the "Obvious" art group in Paris that created the "Edmond de Bellamy" through GAN as an example to explain how GAN, as a deep learning architecture for generative modeling, can generate artworks through the competition between two neural networks in a zero-sum game framework [10]. Zhou and Lee discovered that GAI can effectively assist artists in exploring novel ideas, regardless of whether their previous works were original or not. Their works are more likely to be recognized by their peers. Moreover, the popularization of GAI has reduced the concentration of the value of works among users, indicating that in the process of generating images from text, the "conception" and "screening" abilities have become particularly important. This collaboration between human exploration and GAI technology has given rise to a new creative working mode called "generative synesthesia", where humans and machines achieve organic integration in the creative process [11]. Gregor explores the creative use of generative AI in visual arts and design, proposing a framework of guiding principles called ORCA/E, which encompasses maintaining an open perspective, continuous reflection and introspection, establishing a framework for commonsensical communication, designing based on technological availability, and focusing on ethical and legal issues; the framework provides practical guidelines for the synergistic collaboration of humans and AI in the creative field [12]. Xiao pointed out that GAN is able to efficiently synthesize images with high-fidelity details and clear edges through the adversarial training mechanism of generator and discriminator, and it has been widely used in the field of AI painting for style migration, generation of hyper-realistic artworks, and interactive digital tools, which has pushed forward the technological innovation and expansion of the boundary of the application of generative AI in art creation [13].

The application of GAI in graphic design has become an increasing trend, especially the application of GPT and Stable Diffusion, which provides designers with new perspectives and tools [14]. In addition, natural language understanding algorithms can quickly transform designers' ideas into clear and accurate textual descriptions, helping graphic designers quickly understand the needs of their clients and teams. Qiu et al. integrated innovation diffusion (IDT) theory and UTAUT2 model to construct a structural model influencing graphic designers' behavior of using GAI tools, most factors have a positive impact on the use of the behavior, in which GAI anxiety significantly affects the designers' adoption decisions, the study provides a theoretical basis and practical guidance for designers and GAI developers [15]. Chacón et al. argue that design tools relying on GAI are gradually emerging, but most graphic designers struggle to utilize their full potential due to a lack of knowledge of the relevant technologies, by developing a tool with a graphical user interface, GANSta, a system that supports designers to design and train GANs in a more intuitive way, and apply them to different stages of the design process [16]. Weiland and Miscavage explored public perceptions of GAI-created media and its impact on the creative arts field, finding that respondents were generally familiar with and used GAI for entertainment purposes, and that most found GAI-generated works to be visually appealing but did not view GAI as a life form [17].

In addition, the intervention of GAI in the field of 3D art provides designers with new creative tools and methods. In sculpture design, GAI can utilize algorithms and analysis of large databases to generate sculpture design solutions, which are not only fast, but also produce aesthetic effects beyond the design's preconceptions [18]. For example, Prabha et al. proposed a GAI framework to support formal analysis and innovative generation of sculpture art by combining deep residual convolutional neural networks (CNN) with GAN algorithms, which can learn key elements such as texture, features, composition and form from existing sculptures

to generate new designs while maintaining the authenticity and aesthetic value of the original art [19]. Ranjan et al. combed through the development of AI art from early experiments to current visual art creation, and then explored the specific roles of generative algorithms in the mediums of painting, sculpture, and digital art, respectively, and distilled four innovative features unique to AI art, which provide a comprehensive perspective for an in-depth understanding of the evolution of the field [20]. Ge et al. generated sculptural objects through machine learning, explored the cross-interaction of human-machine creativity, proposed two algorithms for generating 3D point clouds, and realized a more diverse creation of 3D objects, and these algorithmic results were eventually integrated into a complete art installation, which demonstrated the technical feasibility and at the same time triggered a reflection on the relationship between the AI as a creative subject and the audience and artist [21]. Hamid utilizes GAI tools to assist in the creation of preliminary drawings and designs in the field of fine art and sculpture, with the aim of helping artists and art students to expand their creative styles, improve efficiency and precision, and adapt to contemporary technological developments, and has found that Midjourney excels in generating sculptural designs of high aesthetic and professional value, and has been regarded as the best GAI tool for this purpose [22].

In the process of visual creation for film and television, designers can also use painting, photography, film and television subplotting, portraits, and other data related to the subject matter of the movie to train the GAI, and then the generative AI will create the design of the movie characters, scene layout, plot design, and storyboards [23]. Liu and Pan enhance the experience reliability of VR animation works by optimizing the human-computer interaction technology, based on the CNN-support vector machine (SVM) architecture and combined with the error correction strategy, which provides effective technical support for the interaction design of VR animation, and is of positive significance for the promotion of the development of VR animation and the enhancement of the quality of artistic experience [24]. Aoun explains the nature of AI, its evolutionary process, and its operational mechanism in creating visual discourse environments, focusing on the potential of AI technology in the construction of aesthetic environments in film and television, aiming to reveal the specific ways in which it generates visual narrative space [25]. In the link of character modeling and dynamic production, GAI can automatically model and animate virtual characters according to their design and action description. Abootorabi et al. pointed out that the application of GAI in character animation includes content covering a number of key technical directions such as facial animation, expression rendering, image compositing, virtual character generation, gesture and motion modeling, and object and texture compositing [26].

In addition to the innovative application in the traditional visual creation field of film and television, GAI in the field of interactive film and television production also brings more interactivity and participation for the audience.Zou believes that GAI is becoming a key driving force for content upgrading in the film and television industry by virtue of the big data technological innovation and changes in the audience's demand. From the viewpoint of the value logic of film and television production, GAI has become the key driving force for content upgrading in the film and television industry by means of the collaborative human-machine screenwriting, virtual actor substitution, and audience-oriented content production, etc., GAI is deeply integrated into the whole industry chain, promoting structural changes such as the integration and reconstruction of the production model and the reconfiguration of labor resources [27]. Konzack uses Baudrillard's theory of mimesis as a framework for exploring the transformative impact of generative AI on media production such as advertising, film, and television, and uses case studies such as Void, Eternity and The Producers to reveal how GAI has contributed to the proliferation of hyper-real media mimesis [28]. Yan et al. explored the evolution of Artificial Intelligence Generated Content (AIGC) in contemporary digital art, and

its impact on the art creation mode and production process, sorted out the development lineage of AIGC technology, focused on analyzing its wide application in the field of film and television character design, and pointed out the potential that AIGC has shown in the creation of characters [29]. To summarize, the development of GAI technology has a far-reaching impact on art design, which not only changes the current way of creating art works, but also brings new forms of expression. Designers need to pay close attention to the new technology, constantly learn and explore new ideas of art creation, embrace the development trend of technology, and develop together with technology.

This paper focuses on the process of creating inspiration for artistic design creation via generative artificial intelligence, proposing an approach called S_OpenGAN that is a single scale based generative network. S_OpenGAN works on the concept of human-computer interaction in which the generated image is reviewed by a human reviewer subjectively. The score generated during the review is used in the penalty term of the loss function of the generator, thus updating the parameters of the generator. For determining the effectiveness of S_OpenGAN model, SNI, DAT, TransEditor, and StyleGAN-2 are used as comparative experiments against it. The experiments include testing of image generation quality among others, including aesthetic evaluations performed on art creative images produced by S_OpenGAN.

2 Art design creation based on human-computer interaction generative adversarial network

In order to use generative artificial intelligence technology to promote creative stimulation in art design creation, this paper introduces a generative adversarial network based on human-computer collaborative interaction to realize intelligent generation of art creative images.

2.1 Generating Adversarial Networks

The GAN architecture involves two main parts: the generator and the discriminator. While the generator seeks to generate images with characteristics similar to the ones observed in real images, the discriminator acts in the capacity of identifying whether the input image is genuine or not. The interaction between these two parts is competitive in nature. The only way in which the generator can succeed in deceiving the discriminator is by improving itself and ensuring that the generated images mimic the real images to approximate the real data distribution. On the other hand, the discriminator must be able to differentiate between real images and the generated ones to prevent being fooled. When either party dominates in terms of ability, then the learning process of the other will be hindered. Therefore, the best situation is when both the generator and the discriminator achieve Nash equilibrium at 50%.

GAN uses two different network structures, while using unsupervised training, the discriminator itself to measure the generated data, the model has a strong learning ability. The trained GAN model has more powerful image generation capability compared to autoencoder (AE) and variational autocoding (VAE).

2.1.1 Principles of Generating Adversarial Networks

The generative adversarial network generator produces an image from the input random noise, while the discriminator is responsible for discriminating the generated image. The real image data distribution is denoted as p_{data}, but we can't know the expression of p_{data}, so we can't generate the image directly by sampling the real image. Assuming that the generated data

conforms to the distribution $P_G(x;\theta)$, θ is a parameter of the distribution, in order to make the generated data closer to the distribution of the real data, according to the theory of great likelihood, it is to solve for the appropriate θ to maximize the value of $P_G(x,\theta)$. The derivation process is as follows:

$$\begin{aligned}\theta^* &= \arg\max \prod_{i=1}^{m} P_G\left(x^i;\theta\right) = \arg\max \log \prod_{i=1}^{m} P_G\left(x^i;\theta\right) \\ &= \arg\max \sum_{i=1}^{m} \log P_G\left(x^i;\theta\right) \approx \arg\max E_{x\sim p_{data}}[\log P_G(x;\theta)] \\ &= \arg\max \int_x p_{data}(x)\log P_G\left(x;\theta\right)dx - \int_x p_{data}\log p_{data}(x)dx \\ &= \arg\max \int_x p_{data}(x)\left[\log P_G\left(x;\theta\right) - \log p_{data}(x)\right]dx \\ &= \arg\min \int_x p_{data}(x)\log \frac{p_{data}(x)}{P_G\left(x;\theta\right)}dx \\ &= \arg\min KL\left[p_{data}(x)\middle\| P_G\left(x;\theta\right)\right]\end{aligned} \tag{1}$$

From the derivation above, it is clear that the goal of great likelihood estimation is to minimize the KL scatter of the two distributions p_{data} and p_G, but without knowing the expressions for p_{data} and p_G, it is not possible to compute their KL scatters, and so a discriminant (D) is introduced to measure it. The discriminator gives higher scores to the true samples and lower scores to the generated samples. The discriminator output layer typically uses a Sigmoid activation function with an output range of $(0,1)$, where an output of 0 indicates a generated sample and an output of 1 indicates a true sample. The discriminator network objective function formula can be expressed as:

$$V(G,D) = E_{x\sim p_{data}}[\log D(x)] + E_{x\sim p_G}[\log(1-D(x))] \tag{2}$$

The optimal D is expressed as:

$$D^* = \arg\max V\left(G,D\right) \tag{3}$$

The optimal generator is the one that makes it hit the highest possible score while fixing the discriminator, so it can be expressed as:

$$G^* = \arg\min E_{x\sim p_G}[\log(1-D(x))] \tag{4}$$

To facilitate solving for the maximum value of $V\left(G,D\right)$, the discriminator is computed by first fixing the generator during training:

$$\begin{aligned}V\left(G,D\right) &= E_{x\sim p_{data}}[\log D(x)] + E_{x\sim p_G}[\log(1-D(x))] \\ &= \int_x p_{data}(x)\log D_x dx + \int_x p_G(x)\log[1-D(x)]dx\end{aligned} \tag{5}$$

The generator is constant, and the $V\left(G,D\right)$ maximum can be obtained when the inverse

is 0. The calculation procedure is:

$$\frac{\partial V(G,D)}{\partial D}=\frac{p_{data}(x)}{D(x)}+\frac{p_G(x)}{1-D(x)}=0$$
$$D^*=\frac{p_{data}(x)}{p_{data}(x)+p_G(x)} \tag{6}$$

D^* is the solution that maximizes $V(G,D)$, which is obtained by substituting $V(G,D)$:

$$\begin{aligned}V(G,D^*)&=E_{x\sim p_{data}}\log\frac{p_{data}(x)}{p_{data}(x)+p_G(x)}\\&+E_{x\sim p_G}\log\frac{p_{data}(x)}{p_{data}(x)+p_G(x)}\\&=\int_x p_{data}(x)\log\frac{p_{data}(x)}{p_{data}(x)+p_G(x)}dx\\&+\int_x p_G(x)\log\frac{p_{data}(x)}{p_{data}(x)+p_G(x)}dx\\&=-2\log 2+2JSD\left(p_{data}\|p_G\right)\end{aligned} \tag{7}$$

where JSD denotes JS scatter. It can be known from the above equation that the optimal discriminator D, measures the JS scatter between the true distribution p_{data} and the generating distribution p_G. After obtaining the optimal D, in order to get the optimal generator G, fix D, and then find the G that gets the highest score in D. The optimal discriminator can be expressed as:

$$\arg\min_G\max_D V(G,D) \tag{8}$$

2.1.2 Training for generating adversarial networks

From the principle of Generative Adversarial Network (GAN), the optimization training of GAN is the process of solving the maximum and minimum values alternately, and fixing the generator when training the discriminator such that $\max_D V(G,D)$. While training the generator then fix the discriminator such that $\min_G V(G,D)$. When trained to the Nash equilibrium state, it is possible to generate images that are false to reality.

Theoretically, no matter how complex the background of the image can be generated using GAN, the process of generator learning the distribution of real sample features is shown in Figure 1. Where the cyan dashed line represents the real sample (p_{data}) feature distribution, the red line represents the distribution of the generated data (p_G), and the blue dashed line is the discriminant probability output by the discriminator. The random noise z is mapped onto x so that x fits the distribution of the real data as closely as possible. Figure (a) represents the initial state, where there is still a large gap between p_G and p_{data} and the discriminator results are unstable. Figure (b) represents the fixed generator, which optimizes the discriminator to give high scores to the real samples and low scores to the generated samples. Figure (c)

shows the fixed discriminator, optimizing the generator so that the generated samples are closer to the distribution of the real samples so that the discriminator can not make a correct judgment on them. Figure (d) shows that after repeating the steps in Figure (b) and Figure (c) for many times, the model reaches the ideal state, where p_G and p_{data} completely overlap.

Although GAN has powerful image generation ability, the training of GAN has been a difficult problem. In order to improve GAN, researchers have made different attempts on loss function, activation function and optimizer respectively. Nowadays, the problems of GAN training difficulty and gradient vanishing have been better solved, and the common training techniques include using the loss function of WGAN to solve the model collapse, and using leaky-relu as the activation function to avoid gradient vanishing.GAN has now become the mainstream method of image generation.

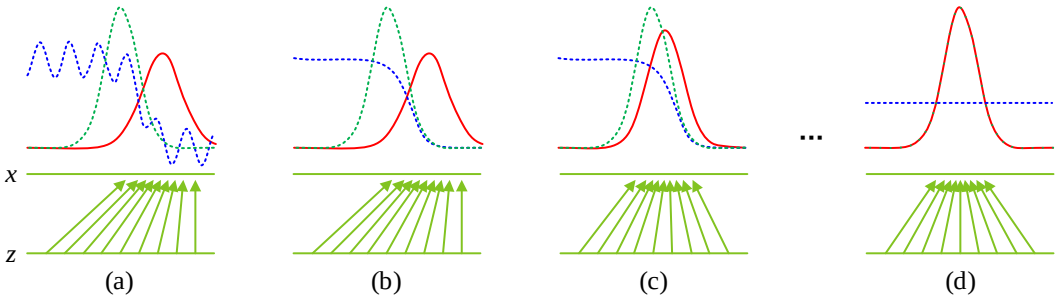


Figure 1: The distribution process of GAN learning features

2.2 Generative Adversarial Networks Based on Collaborative Human-Computer Interaction

This research brings about human-computer interaction in GAN, thus developing an open GAN architecture. The use of human-computer interaction helps to enable quick and quality accomplishment of image-generation tasks through GANs.

The realization process of human-computer interaction is shown in Fig. 2, D1 and D2 are discriminators with the same structure, and G1 and G2 are generators with the same structure.

The images X and Y enter into the generator network G1 and G2 respectively, and output the images $\hat{Y}$ and $\hat{X}$ after transforming, and then give the evaluation to the two generated images, and feed the evaluation results back to the generator network G1 and G2 through the generator loss function to complete the process of updating the generator parameters, and at the same time, the discriminator D1 and D2 are used to discriminate the The real images X and Y and the generated images $\hat{Y}$ and $\hat{X}$, and the discriminating results are also fed back to the generator network to complete the process of updating the discriminator parameters.

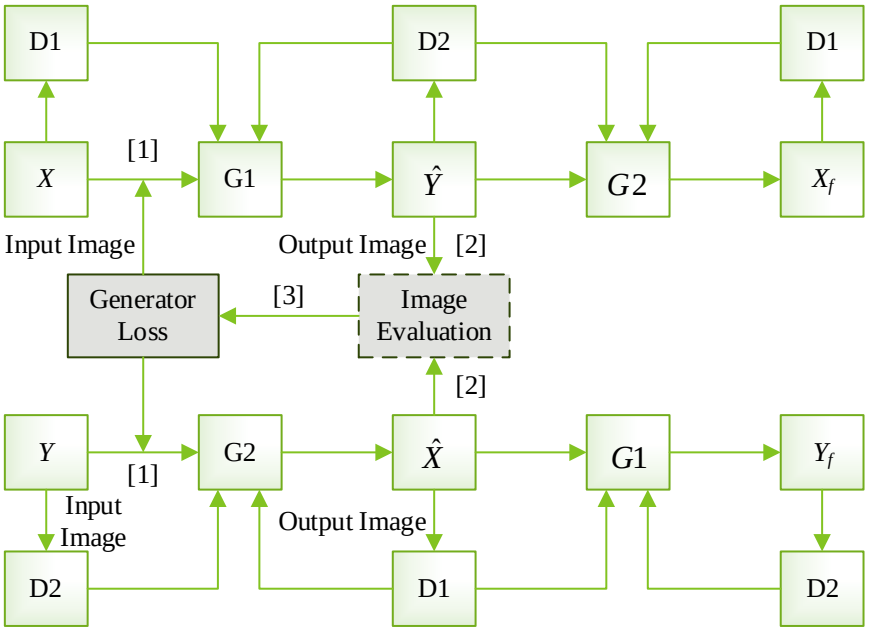


Figure 2: The learning process of interactive generative adversarial networks

Human-computer interaction as a new approach to open GAN, in order to explore its improvement on the effect of GAN generation, this chapter proposes the single scale model S_OpenGAN for small resolution image generation.

2.2.1 S_OpenGAN single-scale architecture

(1) S_OpenGAN architecture

For small resolution image architecture, this paper designs single scale generative architecture for fast HCI generation, S_OpenGAN architecture is shown in Fig. 3.

In this chapter, the entire single scale model is constructed using the structure of convolution and dense blocks, and the input and output dimensions of the architecture are the same.The generator network of S_OpenGAN has 10 large layers containing 6 convolutional layers, 3 dense blocks, and 1 smoothing layer. Among them, the first three convolutional layers are used for image downsampling with the aim of reducing the image size and the number of parameters, and the last three convolutional layers are used for image upsampling with the aim of restoring the original size of the image. Each dense block contains eight dimensional convolutions for dense connectivity, where four convolution kernels of 3×3 are used for dense connectivity and the other four convolution kernels of 1×1 are used to reduce the number of channels.

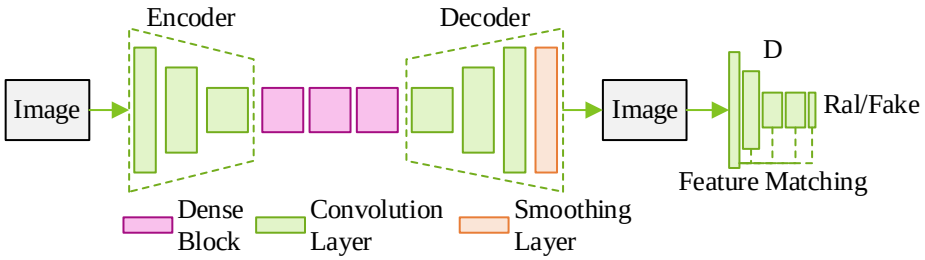


Figure 3: S_OpenGAN architecture

(2) Dense block improvement

For small resolution images, S-OpenGAN aims at fast generation, considering the number of parameters, this paper expands four sets of convolution of 1×1 size on the basis of the original

dense block, which is used to reduce the number of channels as a way to carry out an additional process of parameter reduction. In the original dense block, each contains five convolution blocks, each of size 3×3, with batch normalization and Relu activation, and the improved dense block is shown in Fig. 4.

In the original dense block, this paper adds a convolution of size 1×1 after each convolution block for further reducing the network parameters, and also uses IN normalization instead of BN normalization, the reason is that BN is dependent on the batch size, and when the batch size is large, there are good results, on the contrary, when the batch size is small, the BN will make the generated image has too many artifacts, while IN can get rid of this dependency.

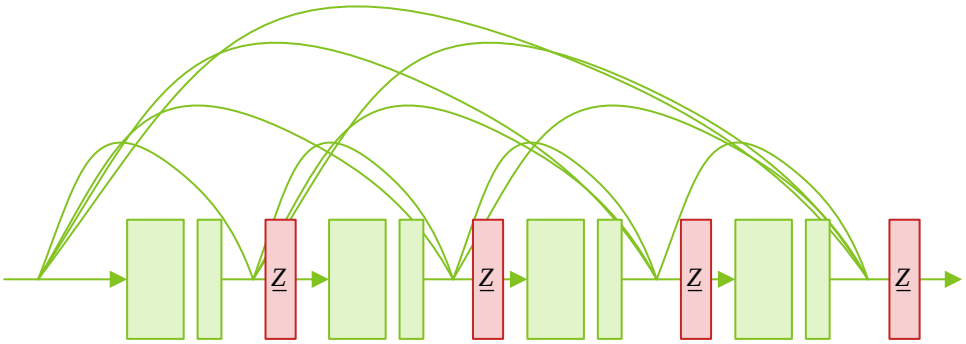


Figure 4: The improved dense blocks

Normalization has the purpose of keeping the size of the gradient passed to the input layer limited while performing backpropagation if the distribution of the input layer has significant variance. This is because a big gradient will make it necessary to keep the learning rate relatively low to avoid missing the optimal point. Therefore, when selecting the learning rate, the scaling of the input layer needs to be considered, and normalization of the data directly helps in making that decision. The formula for normalization is:

$$x_i = \left(x_i - \mu_i\right)/\sigma_i \tag{9}$$

where μ_i is the mean value of the input and σ_i is the standard deviation, the formulas are respectively:

$$\mu_i = \frac{1}{m}\sum_{k\in S_i} x_k \tag{10}$$

$$\sigma_i = \sqrt{\frac{1}{m}\sum_{k\in S_i}\left(x_k - \mu_i\right)^2 + \epsilon} \tag{11}$$

In BN $S_i = \left\{k \middle| k_c = i_c\right\}$, k_c and i_c denote the indexes along the C axis. S_i denotes that for each channel, BN computes μ_i and σ_i along the $\left(N,H,W\right)$ axis. $S_i = \left\{k \middle| k_c = i_c, k_N = i_N\right\}$ in IN denotes that IN computes each sample and each channel along the $\left(H,W\right)$ axis, for GAN, normalizing the full number of channels will lead to incomplete feature representation information, and IN can circumvent this drawback.

(3) Smoothing network layer

As a result of using a penalty term in the loss function, an erratic movement can be observed when the generation process fails to provide an optimal image. The effect that the mentioned effect has on the output image is reduced by adding a smoothing network layer, which is depicte in Fig. 5.

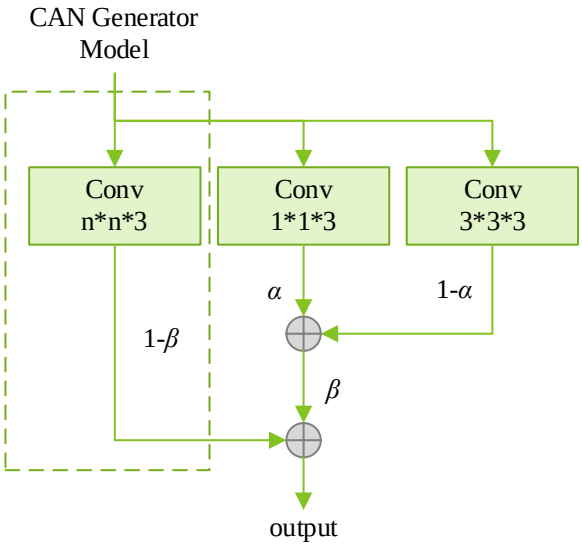


Figure 5: Smoothing the network layer structure

The Original Net part of the figure represents the original network, which outputs the generated image through a $1\times1\times3$ convolution. The smoothing model connects three parts after the Original Net. The first part is a convolution of $1\times1\times3$, the second part is a convolution of $1\times1\times3$, and the third part is a convolution of $3\times3\times3$. Both α, $1-\alpha$ and β, $1-\beta$ represent the output probability of the current channel. As the CNN is able to learn larger sensory field features, fewer subtle features are needed, but since the loss of the image changes instantaneously, a smaller 3×3 convolution is used as a smoothing term as a way to obtain better features. In this paper, we set α to increase with the number of iterations and β to decrease with the number of iterations:

$$\begin{aligned} \beta^{-1} &\propto \frac{1}{\sqrt{step+1}} \\ \alpha &\propto \sqrt{step+1} \end{aligned} \tag{12}$$

In Eq. (12) *step* denotes the step size and a proportional function is used to define the size of α as well as β.

2.2.2 Loss function

(1) Generative loss function

The overall HCI generation loss is:

$$L = L(G,D) + L_{recon}^{s2} + L_{recon}^{s1} + L1 \tag{13}$$

The generator loss function is formulated as follows in the HCI section:

$$L(G,D)=(1-\lambda)*E_{X\sim p_{data(x)}}\left[\max_{D}\log D(G(x))\right] +\lambda*Score*\min_{G}P(G(x)) \tag{14}$$

where $Score$ is the evaluation score of the quality of the generated image. The $P(G(x))$ is the penalty term with the following formula:

$$P(G(x))=E_{X\sim p_{data(x)}}\left[\min_{G}\left(A_{quality}(G(x))\right)\right] \tag{15}$$

$A_{quality}$ represents the determination of the quality of the current generated image, when the image quality is higher, the higher $A_{quality}$ is, then the smaller the penalty for $L(G,D)$ is. That is to say the objective of $L(G,D)$ is $\min P(G(x))$. In the experiments, in order to facilitate gradient descent, the loss function of Eq. (16) is used instead of the above penalty loss:

$$P(G(x))=E_{X\sim p_{data(x)}}\left[\max_{G}\frac{\sum_{i=1}^{N}A_{quality}(G(x_i))-A_{quality}(y_i)}{N}\right] \tag{16}$$

N is the value of batchsize. Maximizing the difference between $G(x_i)$ as well as y_i is used to get a larger penalty in the generator loss function.

Based on the original cyclic consistency, this paper provides an additional definition of style:

$$\begin{aligned} L_{recon}^{s2}&=E_{c1\sim p_{data(c1)}}E_{s2\sim p_{data(s2)}}\min_{D}\left(G(c1,s2)-G(c1,s1)\right)\\ L_{recon}^{s1}&=E_{c2\sim p_{data(c2)}}E_{s1\sim p_{data(s1)}}\min_{D}\left(G(c2,s1)-G(c2,s2)\right)\end{aligned} \tag{17}$$

where $c1,s1$ represents the vector of X domain images after encoding, and $c2,s2$ represents the vector of Y domain images after encoding. They characterize the content of the current image and the corresponding style, respectively. In this paper, we reconstruct this part by using $G(c1,s2)$ for generating an image with content $c1$ and style $s2$, and taking the difference with the original $G(c1,s1)$ for constructing the minimized difference of the contents of the two to satisfy the style change while the content remains unchanged for domain changes.

The L1 loss function is shown in Equation (18):

$$L_1=E_{y\sim p_{data(y)}}\left\|G(x)-y\right\|_1 \tag{18}$$

$G(x)$ denotes the process of transforming the x-domain image into the y-domain by the generator G. The purpose of L1 loss is to minimize the difference between the generated image and the target domain image, so that the distributions of the two domains are approximately similar.

In practical gradient descent, this paper uses Equation (19) to calculate the pixel difference between the generated image and the actual image:

$$L_1 = \frac{\sum_{i=1}^{n} \left| G(x_i^r) - y_i^r \right|}{n} \tag{19}$$

where r is the number of layers, $G(x_i^r)$ represents the generated image and y_i^r represents the real image.

(2) Discriminator loss function

In order to ensure that the generated image does not produce too much blurring, this paper simultaneously adds an input of a blurred image set in the discriminator loss function, which comes from the result of Gaussian filtering in the X domain. The new discriminator loss function is formulated as follows:

$$\begin{aligned} V(G,D) = & E_{X\sim P(x)}\left[\log\left(1-\min_G \max_D G(x)\right)\right] + E_{X\sim P(x)}\left[\max_D \log D(x)\right] \\ & + E_{X_{blurry}\sim P(x)}\left[\min_G \max_D \log\left(1-D\left(G\left(X\_blurry\right)\right)\right)\right] \end{aligned} \tag{20}$$

The X_blurry is a dataset of the X domain after Gaussian blurring, as a way for the discriminator to optimize its minimum value and generate a clearer, higher quality image.

3 S_OpenGAN-based Artistic Creative Image Generation Experiments

In order to verify the effectiveness of the proposed Generative Adversarial Network Model for Human-Computer Interaction (S_OpenGAN), this chapter utilizes the S_OpenGAN model for painting image generation and aesthetic evaluation of the generated images.

3.1 Experiments on S_OpenGAN model application

3.1.1 Experimental data set

In this paper, a dataset of Chinese traditional painting elements of five art types is collected and constructed in the experiments, namely, flower painting (TFP), ink landscape painting (ILP), face painting (FP), potted plant drawing (PPP), and brushstroke landscape painting (MLP) of Tangka art. Since the size of the traditional painting element image dataset is relatively small, this paper proposes a data enhancement method in the acquisition phase to enrich the training samples.

The proposed image data of traditional painting elements is shown in Table 1. In addition, considering the problem of blurred edges and fading of the works due to age in reality, the image quality is enhanced by image enhancement methods such as denoising, sharpening, and color histogram equalization during the production of the training set.

Table 1: Proposed traditional painting element image datasets

Dataset	Quantity	Enhanced quantity	Resolution
Thangka flower paintings (TFP)	2341	13125	128
Ink-wash landscape painting (ILP)	1426	-	256
Face painting (FP)	637	-	128
Potted plant painting (PPP)	642	3649	128
Meticulous landscape painting (MLP)	518	3526	128

3.1.2 Experimental results and analysis

(1) Comparison of generation results

In order to evaluate the performance of the suggested technique, it will be compared with three decoupling-based techniques that represent the state of the art in generation, namely SNI, DAT, and TransEditor, as well as StyleGAN-2 which is commonly used. In order to have an unbiased comparison, the four models have been trained using the same dataset under the same configuration.

To measure the performance quantitatively, two common measures, namely FID metric and PPL metric, have been used for evaluation purposes. The results of the quantitative comparison for the traditional painting element image dataset are shown in Table 2, where ↓ indicates that the smaller the value of the metric is better.

It can be seen that the FID scores of the S_OpenGAN method in this paper are better than the other compared methods on all other datasets, except StyleGAN-2 which has a slightly better FID score (43.6<47.0) on the Gongbi Landscape Painting (MLP). The number of original images in the Gongbi landscape painting dataset is the smallest among multiple datasets, and the S_OpenGAN method in this paper can produce diverse results with arbitrary matching of content and style, and the small number of Gongbi landscape paintings leads to arbitrary combinations of images beyond the representation space of the dataset, resulting in the FID of this paper's method being slightly higher than that of StyleGAN-2. In addition, the DAT is not able to successfully train the DAT method on the Ink and Landscape Painting (ILP) dataset for successful training. Based on the evaluation of the decoupling score of PPL, it can be concluded that the technique called S_OpenGAN developed in this paper produces better results regarding decoupling compared to other techniques used for the same purpose with respect to the traditional painting dataset..

Table 2: Objective index results of FID/PPL by different methods

Indicator	Method	Dataset				
		TFP	ILP	FP	PPP	MLP
FID↓	StyleGAN-2	24.6	41.3	32.1	29.5	43.6
	SNI	66.4	126.7	30.5	100.3	74.5
	DAT	67.2	-	125.4	60.5	146.9
	TransEditor	24.5	64.7	52.3	57.6	90.1
	S_OpenGAN	21.8	39.4	24.1	26.4	47.0
PPL↓	StyleGAN-2	104.6	65.9	60.7	116.8	55.9
	SNI	152.3	241.7	258.9	161.6	369.4
	DAT	140.5	-	172.7	83.8	75.9
	TransEditor	134.9	206.7	74.6	87.5	136.8
	S_OpenGAN	63.4	44.5	26.1	78.3	47.6

In order to evaluate further the decoupling capability of different generation techniques, we

calculate the decoupling metrics in Z Gaussian noise space, W_c content space, and W_s style space, aside from the standard potential space. Z-PPL represents the decoupling metric in the original Gaussian noise sampling space, and W_c-PPL and W_s-PPL represent the decoupling metrics in content and style spaces, respectively. The quantitative values showing the decoupling capability of the methods in each space using the traditional painting element dataset are shown in Table 3. Since TransEditor operates in both content and style spaces because of its dual-spaces characteristics, it was only evaluated in those two spaces. Because StyleGAN-2 uses a non-decoupled design, it was excluded from the comparisons. The results show that compared to other techniques, our proposed technique exhibits better decoupling capability.

Table 3: PPL values for different spaces of different methods

Dataset	Evaluation indicators	Method			
		SNI	DAT	TransEditor	S_OpenGAN
TFP	Z-PPL↓	279.8	289.6	-	83.9
	W_c-PPL↓	153.7	87.2	119.5	44.3
	W_s-PPL↓	53.8	50.2	42.7	7.9
ILP	Z-PPL↓	441.5	-	-	95.2
	W_c-PPL↓	238.7	-	142.6	38.5
	W_s-PPL↓	228.4	-	59.3	3.4
FP	Z-PPL↓	486.8	902.4	-	53.3
	W_c-PPL↓	211.9	92.7	69.4	24.9
	W_s-PPL↓	176.3	33.4	3.7	2.1
PPP	Z-PPL↓	226.7	131.2	-	123.8
	W_c-PPL↓	157.5	75.4	83.6	73.9
	W_s-PPL↓	87.5	4.5	6.4	2.7
MLP	Z-PPL↓	647.8	159.1	-	69.6
	W_c-PPL↓	421.9	69.5	92.4	41.1
	W_s-PPL↓	257.4	9.6	44.8	3.9

For testing the image generation quality of each method, subjective evaluation was conducted. Each technique generated 600 images. Since the rating system for traditional paintings is different from natural images in everyday life, twelve experts were recruited for the classification of images into two groups: "acceptable image quality" and "unacceptable image quality." The findings of the subjective evaluation are provided in Figure 6. Compared to other techniques, the proposed algorithm produces a greater percentage of images with good quality; that is, out of the 600 images, 67 percent were classified as acceptable images, which confirms the validity of the proposed generative model to some extent.

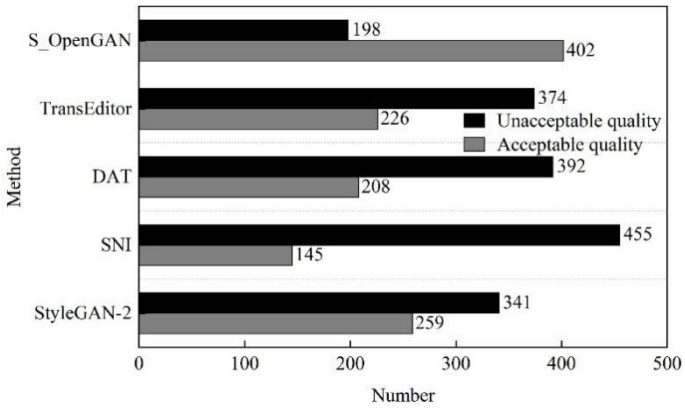


Figure 6: Subjective test results

(2) Extended Experiments on Natural Image Generation

In order to investigate the generality of the proposed S_OpenGAN model, the experiments were extended to some real-world datasets, which include the Oxford 102 Flower dataset of real flowers, the CUB-200-2011 bird dataset, the automobiles dataset, and the FFHQ high-resolution faces dataset. In order to have a fair comparison between S_OpenGAN and StyleGAN-2, both models were retrained on the same dataset under the same settings. The results are shown in Table 4.

It can be observed that S_OpenGAN outperforms StyleGAN-2 in terms of both measures for the remaining datasets except the case of the PPL measure for the bird dataset where the StyleGAN-2 model shows a slight superiority (162.4 vs. 179.3). It means that while S_OpenGAN was developed to address artistic images, it also proves its effectiveness in generating high-quality images for nature in terms of decoupling.

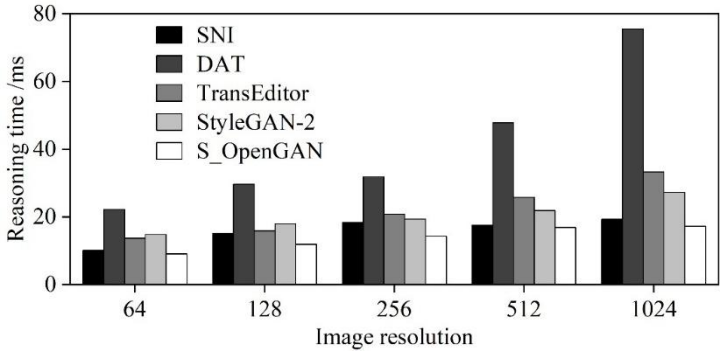
Table 4: FID/PPL results of StyleGAN-2 and DPGN on natural image datasets

Dataset	Resolution	Quantity	StyleGAN-2		S_OpenGAN	
			FID↓	PPL↓	FID↓	PPL↓
Real flowers	128	8190	17.4	56.5	15.7	53.2
Birds	128	8145	11.8	162.4	9.5	179.3
Automobile	128	8856	16.8	100.7	15.4	96.5
High-definition facial recognition	256	52000	10.6	45.7	9.8	44.2

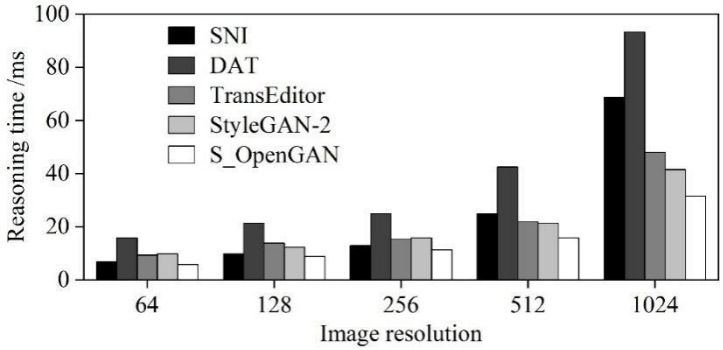
(3) Model performance experiment on each platform

The comparison results of the computational resources required by each method in reasoning to generate images of different resolutions are shown in Fig. 7. Among them, (a)~(c) represent the comparison of inference time of each method on three platforms, namely Intel i7-8700K CPU, GeForce GTX Titan XGPU and Snapdragon870 ARM, respectively, and (d) represents the comparison of memory occupation on GPU platform.

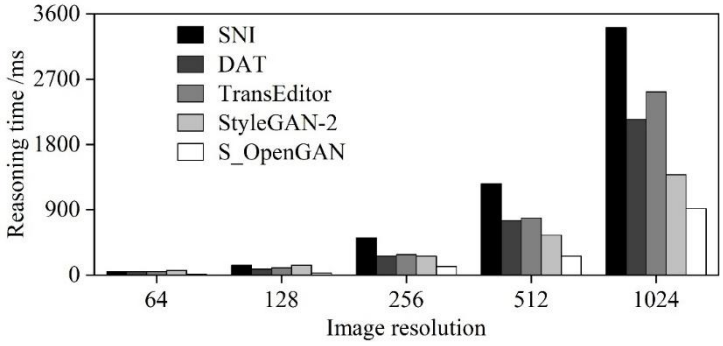
It can be seen that the inference time and memory consumption of the S_OpenGAN method in this paper are less than the other comparative methods on all platforms, and with the growth of the resolution of the generated images, the growth of S_OpenGAN's demand for computational resources is relatively smooth. SNI and DAT, on the other hand, grow rapidly. This indicates that the method in this paper requires less computational resources for image generation at each resolution, which can provide design ideas for future high-resolution image generation methods running on common devices.



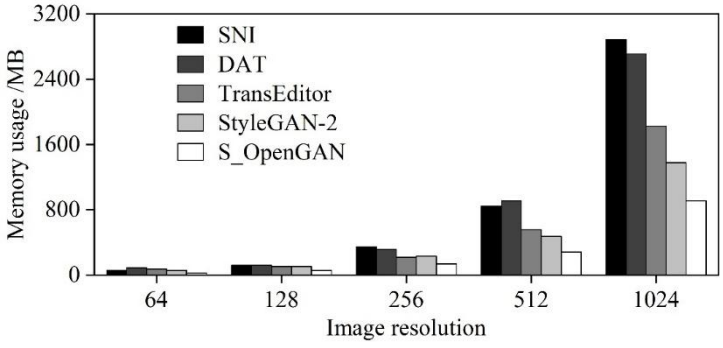
(a) Comparison of inference time on CPU platforms



(b) Comparison of inference time on GPU platforms



(c) Comparison of inference time on the ARM platform



(d) Comparison of memory usage for model inference

Figure 7: Comparison of inference time and memory usage of different methods

3.2 Subjective aesthetic evaluation of generated images

3.2.1 Aesthetic evaluation data

In this paper, we use the semantic difference method of behavioral psychology research for questionnaire development to investigate the aesthetic preference of experimental pictures, and to understand the evaluator's comprehensive evaluation and aesthetic preference of experimental pictures. With this method, we aim to determine the psychological aesthetic preference through the verbal scale, so as to obtain the quantitative data of the subject's feeling structure, and provide basic data for the subsequent analysis.

(1) Acquisition of Experimental Pictures

The experimental images in this paper are selected from the painting “Children and Mothers in the Garden”, which is creatively designed using the S_OpenGAN model to generate seven modified images. In this experiment, we modified the orientation of the child's face in 4 types: turning to the woman's face, turning to the picture appendages, turning to the woman's feet, and turning to other directions.

(2) Semantic Difference Scale Development

Currently, there are no clear rules regarding exactly how many adjectives should be used for the SD method to be measured. In this study, the art faculty was consulted to determine 10 pairs of positive and negative adjectives, which basically summarize the characteristics of people's aesthetic preferences for the work. In addition, psychologists experiment that the respondents can not confuse the processing of the sensory scale will not be more than seven, too high evaluation scale design will interfere with the evaluator's judgment, too low evaluation scale will result in a lower evaluation effect, the evaluation scale is generally the best for the 4-7 level, this study uses the 7-level evaluation scale semantic difference scale to quantify human aesthetic preferences.

(3) Experimental Steps

The aesthetic preference test was conducted through an electronic questionnaire. A total of 18 participants took part in the test, including 10 males and 8 females. 18 questionnaires were collected, with a 100% response rate. The questionnaire data can illustrate the aesthetic preferences of the 18 respondents for the 8 pictures. The evaluation of aesthetic preference selected 10 pairs of adjectives, using a 7-point evaluation scale with semantic difference. Participants chose the most appropriate adjective range to describe their subjective feelings towards the picture, and received corresponding scores. For example, "very cute", "a little cute", "a bit cute", "not bad or good", "a little annoying", "a little annoying", "very annoying" correspond to the scores: 3, 2, 1, 0, -1, -2, -3. Each participant completed the test independently, and no hints were given during the experiment.

3.2.2 Data analysis and results of aesthetic evaluations

Aesthetic preference survey data based on semantic differential scales are exported from electronic questionnaires to Excel format data, and the data analysis software uses the sociological statistical software PASW Statistics 18.In this paper, principal component analysis and clustering are used to do further analysis.

(1) Principal component extraction

Using PASE Statistics 18 maximum variance rotated factor analysis on 10 pairs of adjectives for principal component analysis to obtain two groups of principal components as shown in Table 5, tentatively named the two groups of principal components “component 1” and “component 2”.

It can be seen that component 1 includes adjectives such as "Lovely", "Faithful", "Warm", "Clever", "Desirable", and "Fiery". Component 2 includes adjectives like "Sensitive", "Stable",

and "Emotional". Based on the score distribution of the group of adjectives "Lovely-Unlovely", we believe that component 1 is a positive aesthetic component, while component 2 is considered a negative aesthetic component.

Table 5: Principal component matrix after orthogonal rotation with maximum variance

	Component	
	1	2
Lovely-Unlovely	0.897	0.184
Faithful-Faithfuless	0.869	0.395
Warm-Cold	0.855	0.347
Clever-Stupid	0.792	0.019
Eager-Halfhearted	0.785	0.441
Desirable-Undesirable	0.736	-0.354
Sensitive-Insensitive	-0.143	0.475
Fiery-Clam	0.786	-0.225
Stable-Unstable	-0.402	0.604
Emotional-Rational	0.018	0.336

(2) Picture categorization

According to the Semantic Difference Scale, each picture is described by 10 pairs of adjective intervals, and there are 18 sets of 10-dimensional vectors for each picture. We take the 10-dimensional vectors as the attribute values of the aesthetic evaluation of each picture, and use hierarchical clustering to analyze these 8 pictures by clustering, and the analysis tool also uses PASE Statisties 18 statistical software, and selects the intra-group connection and the squared Euclidean distance as the parameter of hierarchical clustering. The results of hierarchical clustering are shown in Figure 8.

It can be seen that the eight images are divided into two groups. The 1st, 4th, 6th, and 7th pictures where the child's face is facing the mother are clustered to group 1, and the 3rd, 5th, and 8th pictures where the child's face is not facing the mother are clustered to group 2. The second picture is more specific to be clustered to group 2. However, the 5th picture and the 2nd picture have the same character relationship in the composition, the difference is that in the 2nd picture there is a chicken in the direction of the child's and the woman's line of sight.

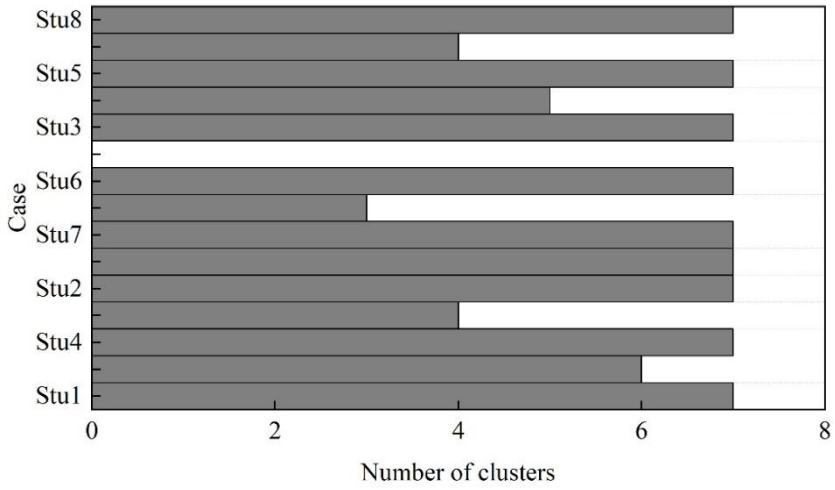


Figure 8: Results of hierarchical clustering

We standardized the scores of the semantic difference scale, and obtained through the effect analysis of principal components: the probability of the main effect of component 1's concomitant probability and the probability of the main effect of component 2's concomitant probability are both greater than 0.05, which indicates that component 1 and component 2 passed the ANOVA chi-square test. The results of the variance chi-square test indicate that the data of aesthetic evaluation satisfy the premise of ANOVA, so a one-way ANOVA based on the LSD method was conducted on Group 1, Group 2, and Figure 2 on Component 1 and Component 2 in two by two.

The results of the one-way ANOVA showed that there was a statistically significant difference between Group 1, Group 2 and Figure 2 on Component 1 ($P<0.05$). On component 2, there was a statistically significant difference between group 1 and group 2 ($F=41.53$, $P<0.001$), figure 2 and group 2 ($F=17.26$, $P<0.001$), while the difference between group 1 and figure 2 ($F=7.05$, $P=0.41$) was not significant. The results of ANOVA indicated that subjects had different aesthetic preferences for Figure 2 than for Groups 1 and 2 on the positive component, but subjects had the same aesthetic preferences for Figure 2 and Group 1 on the negative component.

In summary, it can be seen that the S_OpenGAN model in this paper realizes the creative design of art images, and the experimental pictures with certain elements changed have a significant effect on the subjects' aesthetic preference, and the subjects' judgment of the direction of the line of sight of the characters portrayed in the art work and the position of the appendages affect their aesthetic preference.

4 Conclusion

In this paper, S_OpenGAN, a generative adversarial network based on human-computer collaborative interaction, is designed, and S_OpenGAN is applied to realize the construction of creative stimulation mechanism in art design creation.

Apart from the marginally better FID score achieved by StyleGAN-2 using the Gongbi landscape painting data set, the FID scores obtained using the S_OpenGAN method are better than those of other comparison methods on the rest of the data sets. Moreover, the decoupling metric used (PPL) in S_OpenGAN is greater than the others, implying better decoupling performance and better control over generation of paintings with various combinations of painting styles and contents. Also, our method provides an acceptable percentage of good image quality by getting 67% out of 600 images generated. It also performs well in generation and decoupling metrics and is resource-efficient for generating various image resolutions.

Through the aesthetic evaluation experiments, this paper finds that people's aesthetic preferences in appreciating different artworks are affected by the judgment of the direction of view of the characters portrayed in the artworks and the position of the appendages. The S_OpenGAN model is able to influence the aesthetic preference of subjects by fine-tuning art images for creative design, which helps users to utilize it for art design creation.

References

[1] Banh, L., & Strobel, G. (2023). Generative artificial intelligence. Electronic Markets, 33(1), 63.

[2] Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., ... & Vaidya, P. (2025). Artificial intelligence in developing countries: The impact of generative

artificial intelligence (AI) technologies for development. Information development, 41(3), 1036-1054.

[3] Mogaji, E., Viglia, G., Srivastava, P., & Dwivedi, Y. K. (2024). Is it the end of the technology acceptance model in the era of generative artificial intelligence?. International Journal of Contemporary Hospitality Management, 36(10), 3324-3339.

[4] Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187-192.

[5] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[6] Choudhury, M. M., Eisenbart, B., & Kuys, B. (2025). Artificial intelligence (AI) in the design process–a review and analysis on generative AI perspectives. Proceedings of the Design Society, 5, 631-640.

[7] Wenjing, X., & Cai, Z. (2023). Assessing the best art design based on artificial intelligence and machine learning using GTMA. Soft Computing-A Fusion of Foundations, Methodologies & Applications, 27(1).

[8] Petrenko, A. (2025). Generative Artificial Intelligence (GAI) for Research and Creative Activities. In System Analysis and Data Mining (pp. 251-265). Cham: Springer Nature Switzerland.

[9] Zhu, W., Guo, R., Zhu, G., Li, C., Li, H., & Song, Y. (2025). GAI4DE: Harnessing the Design Process to Integrate GAI into Design Studios. International Journal of Human–Computer Interaction, 1-24.

[10] Rani, S., Jining, D., Shah, D., Xaba, S., & Singh, P. R. (2023, April). The role of artificial intelligence in art: a comprehensive review of a generative adversarial network portrait painting. In International Conference on Intelligent Computing & Optimization (pp. 126-135). Cham: Springer Nature Switzerland.

[11] Zhou, E., & Lee, D. (2024). Generative artificial intelligence, human creativity, and art. PNAS nexus, 3(3), pgae052.

[12] Gregor, S. (2025). Being Creative with a Non-Human: The Use of Generative Artificial Intelligence in Art. Australasian Journal of Information Systems, 29.

[13] Xiao, Y. (2025). Research on the Application of Generative Adversarial Networks in Artificial Intelligence Painting. Advances in Engineering Technology Research, 15(1), 1460-1460.

[14] Gai, Z., & Yang, T. (2024). Digital art creation and visual communication design driven by internet of things algorithm. Computer-Aided Design and Applications, 135-149.

[15] Qiu, T., Yang, D., Zeng, H., & Chen, X. (2024). Understanding graphic designers' usage behavior of generative artificial intelligence tools. Kybernetes.

[16] Chacón, J. C., Nimi, H. M., Kloss, B., & Kenta, O. (2020, December). Towards the development of AI based generative design tools and applications. In International Conference on Design, Learning, and Innovation (pp. 63-73). Cham: Springer International Publishing.

[17] Weiland, S. J., & Miscavage, J. (2025). Generative Artificial Intelligence: A Threat to the Graphic Design Industry?. Journal of Applied Professional Studies, 6(12).

[18] Radwan, D. R. M. (2025). Exploring the Effectiveness of Artificial Intelligence in Fostering Contemporary Sculpture Design. Journal of Art, Design and Music, 4(1), 12.

[19] Prabha, A. J., Sarkar, R., Mohanraj, B., & Kavin, B. P. (2025). Creation and Analysis of Sculptures Using the Principles of Generative AI. In Generative AI and Creativity (pp. 43-61). Auerbach Publications.

[20] Ranjan, R., Mishra, A., Maharana, S. P., & Kumari, S. (2024). AI in Visual Arts: Exploring Generative Algorithm. In The Pioneering Applications of Generative AI (pp. 41-60). IGI Global.

[21] Ge, S., Dill, A., Kang, E., Li, C. L., Zhang, L., Zaheer, M., & Poczos, B. (2019). Developing creative ai to generate sculptural objects. arXiv preprint arXiv:1908.07587.

[22] Hamid, H. S. (2024). Employing artificial intelligence to generate drawings and designs of sculptural forms. Al-Academy, 469-484.

[23] Zhao, X., & Zhao, X. (2024). Application of generative artificial intelligence in film image production. Comput Aided Des Appl, 21, 15-28.

[24] Liu, X., & Pan, H. (2022). The path of film and television animation creation using virtual reality technology under the artificial intelligence. Scientific Programming, 2022(1), 1712929.

[25] Aoun, M. A. (2025). Artificial Intelligence and Its Artistic Uses for Generating Environmental Aesthetics in Film and Television. Al-Academy, 297-314.

[26] Abootorabi, M. M., Ghahroodi, O., Zahraei, P. S., Behzadasl, H., Mirrokni, A., Salimipanah, M., ... & Asgari, E. (2025). Generative AI for Character Animation: A Comprehensive Survey of Techniques, Applications, and Future Directions. arXiv preprint arXiv:2504.19056.

[27] Zou, Y. (2025). The Value Logic, Challenges, and Development Strategies of Generative AI Empowering Film and Television Production. Advances in Education, Humanities and Social Science Research, 15(1), 662-662.

[28] Konzack, L. (2025). Generative AI, Simulacra, and the Transformation of Media Production. Athens Journal of Mass Media & Communications, 11(3).

[29] Yan, H., Tan, S., Zhang, K., & Wang, D. (2024). A New Paradigm for Contemporary Film and Television Characters Design in the Context of Artificial Intelligence. Advances in Education, Humanities and Social Science Research, 11(1), 194-194.