International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 3360-3371

https://doi.org/10.70517/ijhsa463284

The role of computer-assisted creation methods in promoting the innovation path of new media art and design

Xinyue Yuan[1,*]
[1] College of Design and Art, Wenzhou University of Technology, Wenzhou, Zhejiang, 325035, China
Corresponding authors: (e-mail: 19906711692@163.com).

Abstract The integration and application of computer technology and new media art can not only optimize the new media art presentation form, but also innovate the presentation form of art design. Focusing on the innovation of new media art design, this paper discusses and analyzes the promotion role of computer-aided creation methods in it. Based on the application of image generation technology in AI, taking the single-stage generative adversarial network as the basic model and introducing multiple intelligent modules, we design the text-generated image method based on DGF-GAN to stimulate and assist the design of new media art. The images generated by the DGF-GAN model are of excellent quality and diversity, and its FID and IS values perform optimally in the comparison experiments. Compared to the base model, the FID values of the DGF-GAN model are reduced by 6.34% and 5.83% and the IS values are improved by 1.46% and 6.60% on both datasets. In addition, the method has good training efficiency and image generation efficiency. The results show that the proposed method has a large development potential in new media art design, which can enhance the design flexibility and designer creativity, and promote the development of art creation in a more intelligent and precise way.

Index Terms computer-aided creation method, generative adversarial network, text-generated image, new media art design

I. Introduction

The centerpiece of new media is digital technology and message science, which is based on the theory of mass communication as a theoretical foundation, following the creation of modern art, and utilizing the comprehensive characteristics and functions of message technology to integrate different disciplines, involving the fields of science and art, culture and art, business, education and management [1]. New media includes image, text, audio, video and other forms, in which the communication of forms and the digitization of communication content are the characteristics of new media, in the process of digitization of message acquisition, access, processing and distribution are met in the production process of new media [2]-[4]. It has become a new message carrier in the post-linguistic era, after text and electronic technology [5], [6].

The history of new media shows that the creation of new media art relies on digital technology [7]. But just technical new media art is too mechanized is also no development prospects, there must be art within to have now slowly towards maturity of new media art, new media art is also in the combination of technology and art gradually independent [8]-[11]. It is different from traditional art forms, and has a contemporary and advanced nature [12]. In order to promote the development of new media art design towards a higher level, some new scientific and technological means and industrial technology can be integrated into new media art design [13], [14]. In particular, the application of digital technology and computer technology can allow new media art design to continuously form new design forms [15].

This paper discusses the application of computer technology in new media art design, and constructs an intelligent text-image generation model to inspire designers' creativity. Aiming at the problem that the single-stage DF-GAN model does not fully utilize text features, the conditional enhancement module and the affine fusion block of residual structure are introduced to enhance the fusion of coarse-grained text features and image features. Then the channel and pixel attention modules are used to promote the fusion of fine-grained text features and local image features to complete the design of the DGF-GAN-based text-to-image method. Comparison experiments are conducted on CUB dataset and Oxford-102 dataset to compare the FID values and IS values of multiple models to explore the image generation quality and text consistency of this paper's method. Finally, based on the analysis of metrics such as number of training times, training time, generation speed and memory occupation, the performance of the model in terms of training efficiency and image generation efficiency is evaluated.

II. Advancement of computer-assisted creative methods

At this stage, computer technology has become widely popularized and applied advanced technology, the technology and new media art design work to achieve the integration of application is to optimize the effect of art design is an important method. This paper is based on the application of computer-aided creative methods in new media art design, the following discussion of the role of computer technology in promoting new media art design. In terms of computer-aided creation methods, the main research content of this paper is computer programming, big data technology and artificial intelligence technology as the main technology of computer science.

II. A.Promote data collection and analysis

Utilizing big data automated data collection tools for external data collection and internal data mining, analyzing the development trend of new media art, predicting future art trends, and cleaning up and processing art maps in the context of new media. Use big data cleaning technology to quickly collect, process, store and analyze various art data, discover new art knowledge, create new value, and improve the expressive ability of new media art. The integration of information technology and art has led to the rapid growth of art-related data, which has become the basic resource of art, which in turn is categorized through the global mechanism of production, circulation, distribution and operation of art elements, and is applied to the modern new media art design, generating art maps and models for designers' reference. The data collection scope model is shown in Figure 1, which is derived from data collection, cleaning and storage operations through database, business system logs, Internet applications, container logs and operating system logs and network device logs.

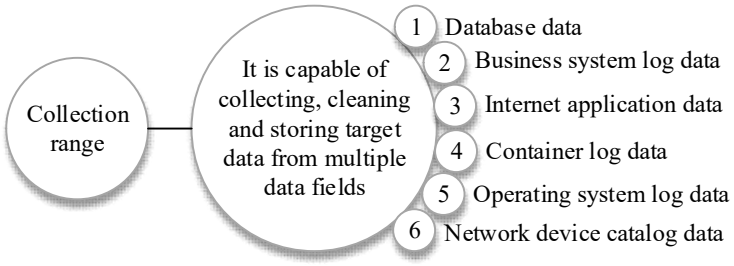

Figure 1: Data acquisition range model

II. B.Promotion of mobile terminal applications

The use of computer programs to create new media art that is interactive, interesting and practical for mobile terminal devices. The use of mobile terminal devices is becoming more and more common, coupled with the widespread use of wireless networks, the fusion media H5 animation mode is gradually replacing the traditional TC fixed mode. Mobile terminal devices are favored by most users for their easy operation, diversity and low threshold. Therefore, in order to adapt to the development of society, the new media art will also shift from the big screen to the cell phone screen.

II. C.Promoting user experience enhancement

New media art focuses on the public's experience and practicality, and computer technology can technically analyze the points of interest, practicality, convenience and effectiveness of new media art to find the user's concerns, and then summarize them through inductive reasoning and apply them to practical needs. Making full use of advanced computer technology, it perfectly integrates virtual and real art elements, creates new art forms beyond real life, and brings people different visual experiences and aesthetic feelings.

II. D.Promote the application of AI technology

The innovative paths of utilizing the artificial intelligence technology of the computer science branch to empower new media art design are mainly as follows:

(1) Based on image recognition and generation technology to stimulate creative inspiration

Image recognition and generation technology can accurately identify and generate a variety of artistic styles and design elements through deep learning algorithms and big data analysis, injecting new creative momentum for designers. On the one hand, image recognition technology can widely collect and analyze the characteristics of artworks of different styles and historical periods. By learning and categorizing the characteristics of a large number of art works, AI can quickly identify images with similar or identical styles or design elements, thus providing designers with ideas for reference and reference. On the other hand, image generation technology can

simulate and innovate existing art styles through advanced algorithms such as Generative Adversarial Networks (GANs) to create unique visual effects for art design.

(2) Color selection and creative assistance based on AI technology

Through advanced algorithms and deep learning models, the color management tools empowered by AI technology can achieve accurate color analysis and intelligent recommendations at the design stage, thus improving design quality and color matching efficiency.

(3) AI Intelligent Design Optimization

Through machine learning, deep learning and GANs technology, AI can not only optimize the design process, but also significantly enhance the artistic value of design works. On the one hand, AI technology assists the design of art forms, which can significantly enhance the innovation of art works. On the other hand, AI technology learns from a large number of design works, so as to quickly master the characteristics of different materials and textures, collocation methods and visual effects.

(4) Intelligent Interactive Exhibition

Optimizing the traditional exhibition mode through Augmented Reality (AR) and Virtual Reality (VR) technology will significantly enhance the audience's experience. The virtual exhibition hall will not be restricted by physical space, and can accommodate more exhibits and theme displays than the physical exhibition hall, so that the audience can enjoy the art and design works of various styles and cultural backgrounds in the virtual space, which will greatly improve the richness and interestingness of the exhibition.

III. DGF-GAN based text generation image method

Through the above analysis, the image generation technology in computer-aided creation can provide inspiration for new media art design, as well as real-time feedback and optimization of design solutions. In this paper, for the image generation technology, based on the single-stage generative adversarial network model (DF-GAN), the generative network is improved, and a text-generated image method based on double-granularity feature fusion, DGF-GAN, is proposed to improve the quality of the generated image.

III. A. Single-stage generative adversarial networks

Single-stage generative adversarial networks are usually GAN models that use only one generator and one discriminator in the generation process and generate target samples with only one forward propagation. Early single-stage generative adversarial networks were limited by the fact that their generation process was completed within a single stage, by the size of the model structure and the limitations of the model in capturing the intrinsic structure and features of the data, as well as by the pattern collapse problem that is prone to occur during the training process. However, in recent years, through the improvement of hardware level and resources, through the introduction of various improvements, a large number of excellent single-stage generative adversarial network models have emerged, which are capable of generating high-quality, high-resolution images only through a single-stage generator.

Compared to multi-stage GANs, single-stage GAN models are simpler and more computationally efficient in terms of the composition of the structure, but may be limited in terms of semantic consistency, image detail and diversity of the generated samples.

III. B. DGF-GAN modeling

The DGF-GAN model consists of a text encoder, a conditional enhancement module, a residual affine fusion block, a word-level attention module, and a one-way discriminator.

III. B. 1) Text Encoder

Simultaneous training of text encoder and GAN not only increases the number of parameters and training time of the model, but also obtains very limited improvement in objective metrics, so this study chooses to use a pre-trained text encoder to encode text. Although encoding text using the updated Bert model will improve the objective metrics of the generated images, the effectiveness of the text-generated image method depends more on the generative model itself. In addition, the utterances in the datasets commonly used for text-generated image tasks are usually short and the number of utterances is relatively small, which may be the reason why Bi-LSTM is chosen as the text encoder model for most of the current studies. In order to ensure the fairness of the comparison, this paper continues to use Bi-LSTM as the text encoder model. The text encoding in the text-to-generate-image task is vectors with visual differentiation ability, which requires having the text encoder and the image encoder trained at the same time.

All words in the dataset are first numbered, and then the embedding layer maps the corresponding words to dense vectors of fixed size, which are finally fed into Bi-LSTM for processing.In Bi-LSTM, the connection of two

directions of a hidden layer is encoded as a word, and the connection of two directions of the last hidden layer is encoded as the current utterance. All words in an utterance are encoded as $e \in R^{D} \times T$ and the utterance is encoded as $s \in R^{D}$. Then:

$$e, s = BiLSTM(text) \tag{1}$$

where T represents the number of words in the utterance and D represents the dimension of the word or utterance vector. The image encoder is built on the pre-trained Inception-V3 model of the ImageNet dataset. The input image is resized to 299×299 using bilinear interpolation and fed into the image encoder, the real image is encoded in the last average pooling layer, and the image sub-regions are encoded in the "Mixed_6e" layer. The final image code $\overline{f} \in R^{256}$ and the image sub-region code $f \in R^{256 \times 289}$ are obtained through the fully-connected layer and 1×1 convolutional layer. During training, the parameters of the Inception-V3 network are unchanged, and the gradient is updated in the fully connected layer of the image encoder and the 1×1 convolutional layer. Then:

$$\overline{v} \in R^{2048}, v \in R^{768 \times 289} = InceptionV3(img) \tag{2}$$

$$f = Wv, \overline{f} = \overline{W}\overline{v} \tag{3}$$

The number of sub-regions of an image is 289 and the dimension of the sub-regions is 256.The image encoder and text encoder are trained by minimizing the DAMSM loss. The inference process for L_{DAMSM} loss is as follows.

The similarity s_i between the i th word and the j th region is first calculated and then normalized to obtain $\overline{s}_{i,j}$. Based on $\overline{s}_{i,j}$ find the weighted sum of all regional visual vectors c_i, which represents the dynamic representation of the image sub-region associated with the i th word of the utterance. The correlation between the i th word e_i and the weighted sum of visual vectors c_i is denoted by the cosine similarity, which defines the image-text matching score between a single image (img) and a text description $(text)$ as $R(img, text)$:

$$s = e^{T} f, \overline{s}_{i,j} = \frac{\exp(s_{i,j})}{\sum_{k=0}^{T-1} \exp(s_{k,j})} \tag{4}$$

$$c_i = \sum_{j=0}^{288} \alpha_j v_j, \alpha_j = \frac{\exp(\gamma_1 \overline{s}_{i,j})}{\sum_{k=0}^{288} \exp(\gamma_1 \overline{s}_{i,k})} \tag{5}$$

$$R(img, text) = \log\left(\sum_{i=1}^{T-1} \exp(\gamma_2 R(c_i, e_i)) \right)^{\frac{1}{\gamma_2}}$$
$$R(c_i, e_i) = (c_i^{T} e_i) / (\| c_i \| \| e_i \|) \tag{6}$$

For a batch of image statement pairs, the a posteriori probability that an image img_k matches the statement description $text_k$ is $P(img_k \mid text_k)$, and the a posteriori probability that the statement description $text_k$ matches the image img_k is $P(text_k \mid img_k)$. The loss function L^{w} is defined as the negative log posterior probability of P:

$$P(text_k \mid img_k) = \frac{\exp(\gamma_3 R(img_k, text_k))}{\sum_{j=1}^{M} \exp(\gamma_3 R(img_k, text_j))} \tag{7}$$

$$L_1^{w} = -\sum_{k=1}^{M} \log P(text_k \mid img_k)$$
$$L_2^{w} = -\sum_{i=k}^{M} \log P(img_k \mid text_k) \tag{8}$$

If the image-text matching score between the whole image and the statement description is defined as follows $R(img,text)$, according to the above process, L_1^s and L_2^s are obtained. Then:

$$R(img,text) = (\overline{f}^T \overline{e}) / (\| \overline{f} \| \| \overline{e} \|) \tag{9}$$

Eventually, the loss values for word vs. image and utterance vs. image are summed to get the total L_{DAMSM}:

$$L_{DAMSM} = L_1^w + L_2^w + L_1^s + L_2^s \tag{10}$$

This is when the text encoding generated by the text encoder is visually distinguishable, i.e., able to distinguish between specific colors. When training the GAN, the text encoder and image encoder parameters are fixed, which reduces the number of parameters for a single training of the text-generated image task.

III. B. 2) Conditional Enhancement Module

The Conditional Augmentation (CA) module mainly copes with the context where image and text matching pairs have limited data and the text data is conditionally fixed. Given a text description s generate sentence feature vector φ_s by text encoder Bi-LSTM, input φ_s into the Embedding layer to get the mean value function $\mu(\varphi_s)$ that obeys the Gaussian distribution $N(\mu(\varphi_s),\Sigma(\varphi_s))$ of the mean function $\mu(\varphi_s)$ and the diagonal covariance matrix $\sigma(\varphi_s)$, which is computed to obtain the conditionally enhanced text vector $\hat{s}$. $\hat{s}$ is obtained from $\sigma(\varphi_s)$ and a random noise obeying a standard normal distribution $\varepsilon \sim N(0,1)$ by Hadamard product operation and spliced with $\mu(\varphi_s)$, and $\hat{s}$ is computed as follows:

$$\hat{s} = \mu(\varphi_s) + \sigma(\varphi_s) \odot \varepsilon \tag{11}$$

$\odot$ means multiply by elements.

III. B. 3) Residual affine fusion module

In affine transformation, the coarse-grained text feature $\hat{s}$ is predicted from two multilayer perceptrons, the channel scale parameter α and the channel scaling shift parameter β respectively. Firstly, the image features are linearly transformed according to the channel scale parameter α, and then they are translationally transformed according to the channel scaling shift parameter β. The computational process is as follows:

$$\begin{cases} \alpha = MLP_1(\hat{s}) \\ \beta = MLP_2(\hat{s}) \end{cases} \tag{12}$$

$$Affine(c_n \mid \hat{s}) = \alpha_n c_n + \beta_n \tag{13}$$

where Affine denotes the affine transformation, c_n is the nth channel of the input image feature, $\hat{s}$ is the text feature, and α_n and β_n are the parameters of the linear and translational transformations of the nth channel acting on the image feature, respectively.

III. B. 4) Word Attention Module

Two word-level attention modules: the channel attention module (CAM) and the pixel attention module (PAM) are introduced to explore the potential relationship between word features and image features.

(1) Channel Attention Module

Firstly, for the image feature $h_c \in R^{H\times W\times C}$, the global features and the most significant channel features are extracted using average pooling and maximum pooling, respectively. Then the visual features $Q_{ca} \in R^{C\times C}$ and $Q_{cm} \in R^{C\times C}$ are obtained by average pooling and maximum pooling through the two 1×1 convolutional layers after the matrix dimension transformation operation with the following equations:

$$\begin{cases} Q_{ca} = f_{conv}(Avg(h_c)) \\ Q_{cm} = f_{conv}(Max(h_c)) \end{cases} \tag{14}$$

Avg and Max denote average and maximum pooling, respectively. The f_{conv} denotes the l×1 convolutional layer. And for the word-level text feature $e \in R^{D\times T}$, the key $K_{ce} \in R^{C\times T}$ and the value $V_{ce} \in R^{C\times T}$ are obtained by two different 1×1 convolution operations, which share the same vector with the image feature. semantic space. Next, a

mean operation is performed on V_{ce} to extract the global representation of the word in each channel, and its transpose and V_{ce} do a dot-product operation to obtain the word's weight matrix $E_c \in R^{1\times T}$ in the sentence, which is used to identify the importance level of each word in the sentence:

$$\begin{cases} K_{ce} = f_{conv}(e) \\ V_{ce} = f_{conv}(e) \end{cases} \tag{15}$$

$$E_c = (f_{mean}(V_{ce}))^T V_{ce} \tag{16}$$

The f_{mean} represents the mean operation. Next the computation of channel attention weights is performed, using the query matrix and K_{ce} to do the product operation to get the word and channel similarity matrices γ_{ca} and γ_{cm}. Finally, it is multiplied with E_c and normalized by softmax to obtain the final average pooled and maximally pooled joint weights of channel attention W_{ca} and W_{cm}, with the following equations:

$$\begin{cases} \gamma_{ca} = Q_{ca} \cdot K_{ce} \\ \gamma_{cm} = Q_{cm} \cdot K_{ce} \end{cases} \tag{17}$$

$$\begin{cases} W_{ca} = softmax(\gamma_{ca}, E_c^T) \\ W_{cm} = softmax(\gamma_{cm}, E_c^T) \end{cases} \tag{18}$$

The final joint channel attention weights are multiplied element-by-element with the initial image features to update the image's features F_{ca} and F_{cm}, thus assigning more weights to the useful channels in the network:

$$\begin{cases} F_{ca} = W_{ca} \otimes h_c \\ F_{cm} = W_{cm} \otimes h_c \end{cases} \tag{19}$$

The $\otimes$ denotes element-by-element multiplication. Finally, the image features fused by maximum pooling and average pooling are fused using adaptive gating and residuals to obtain the adaptively fused image features F_c for image updating or generation with the following formula:

$$G_c = f_{sigmoid}(W_g \cdot (F_{ca} + F_{cm})) \tag{20}$$

$$F_c = G_c \cdot F_{ca} + (1 - G_c) \cdot F_{cm} \tag{21}$$

$$y_c = \eta_c \cdot F_c + h \tag{22}$$

G_c denotes the response gate for image feature fusion, $f_{sigmoid}$ denotes the sigmoid function, and W_g denotes the network parameters of the fully connected layer. η_c is a learnable scaling parameter initialized to 0. y_c denotes the final output feature.

(2) Pixel Attention Module

For the image feature map $h_p \in R^{H\times W\times C}$, average pooling and maximum pooling in spatial dimensions are performed first to extract the global features, and dimensional transformations are performed to obtain the average pooling query matrix $Q_{pa} \in R^{(H\times W)\times C}$ and the maximum pooling query matrix $Q_{pm} \in R^{(H\times W)\times C}$. For a given word-level text feature, it is processed in the same way as CAM to obtain the key $K_{pe} \in R^{C\times T}$, the value $V_{pe} \in R^{C\times T}$, and the weight matrix of the word in the sentence $E_p \in R^{1\times T}$. The semantic similarity matrices γ_{pa} and γ_{pm} between words and spatial pixels are obtained using the key K_{pe} and the query matrix dot product operation, followed by the dot product operation with the transpose of E_p, respectively, and the softmax function is used to obtain the transformation to $R^{H\times W\times 1}$ pixel-level attentional weights with the following formula:

$$\begin{cases} W_{pa} = softmax(\gamma_{pa}, E_p^T) \\ W_{pm} = softmax(\gamma_{pm}, E_p^T) \end{cases} \tag{23}$$

Element-by-element multiplication between pixel-level attention-weighted features and the original feature map is performed to update the image feature maps F_{pa} and F_{pm}. Subsequently, feature map fusion is performed by concatenating F_{pa} and F_{pm} according to the channel dimensions and placing the results into a 1×1 convolutional layer and a ReLU function to generate the merged feature map $F_p \in R^{H\times W\times C}$. Finally, adaptive residual concatenation is used so as to obtain the final output $y_p \in R^{H\times W\times C}$ with the following formula:

$$\begin{cases} F_{pa} = W_{pa} \square h_p \\ F_{pm} = W_{pm} \square h_p \end{cases} \tag{24}$$

$$F_p = f_{ReLU}(f_{conv}[F_{saw}; F_{smw}]) \tag{25}$$

$$y_p = \eta_p \cdot F_p + h_p \tag{26}$$

In the above equation, $\square$ denotes element-by-element multiplication, and F_{saw} and F_{smw} denote rescaled feature maps. $[;]$ denotes a tandem operation along the channel dimension, f_{conv} denotes a 1×1 convolution operation, and f_{ReLU} denotes the ReLU activation function. η_p is a learnable parameter initialized to 0 and y_p denotes the final output.

III. B. 5) One-way discriminators

The discriminator consists of six downsampled residual blocks and two convolutional layers, the generated image and the real image are passed through a series of downsampled residual blocks in the discriminator, respectively, which continuously reduces the resolution and extracts the features to get the image features with a resolution of 4×4 and splices them with the textual features, which are processed by the two convolutional layers to finally produce the feature output used to compute the adversarial loss. This single discriminative path provides the discriminator with a gradient that is realistic and matches the data points, thus helping the generator to converge faster and facilitating the generation of higher quality images.

III. B. 6) Loss function

To enhance the capability of the discriminator and to ensure the consistency of the image and text, MA-GP loss is used for real images that match the text with the following formula:

$$L_{MA-GP} = kE_{x\sim P_{(r)}}\left[\| \nabla_x D(x,\hat{s}) \| + \| \nabla_s D(x,\hat{s}) \|^p\right] \tag{27}$$

k and p are two hyperparameters for balancing the gradient effect, x denotes the real image, $x \sim P_r$ denotes obeying the real image data distribution, $\hat{s}$ denotes the textual features, ∇ denotes gradient descent, and $D(x,\hat{s})$ denotes the match between the discriminator corresponding to the real image data distribution and the textual features evaluation, and $\|*\|$ is the L_1 paradigm expression. From this, the adversarial loss formula for the discriminator can be obtained:

$$\begin{aligned} L_D = & -E_{x\sim P_{(r)}}[\min(0,-1+D(x,\hat{s}))] \\ & -\frac{1}{2}E_{G(z)\sim P_{(g)}}[\min(0,-1-D(G(z),\hat{s}))] \\ & -\frac{1}{2}E_{x\sim P_{(mis)}}[\min(0,-1-D(x,\hat{s}))] \\ & +L_{MA-GP} \end{aligned} \tag{28}$$

$P_{(r)}$ represents the distribution of real image data, $P_{(g)}$ represents the distribution of generated image data, $P_{(mis)}$ represents the distribution of unmatched data, $G(z)$ represents the generated image results, $D(G(z),\hat{s})$ represents the consistency evaluation results of the discriminator on the generated image and text features, and $D(x,\hat{s})$ represents the consistency evaluation results of the discriminator on the real image and text features.

In the conditional enhancement module, in order to avoid overfitting and help the generator produce more diverse and realistic images, it needs to be regularized. The L_{CA} is obtained by quantifying the difference between the conditional vector and the standard normal distribution through the KL scattering, with the following formula:

$$L_{CA} = D_{KL}(N(\mu_0(\hat{s}), \Sigma_0(\hat{s})) \,||\, N(0,1)) \tag{29}$$

From this, the generator loss function for the methods in this chapter can be obtained as follows:

$$L_G = -E_{G(z)\sim P_{(g)}}[D(G(z), \hat{s})] + L_{CA} \tag{30}$$

IV. Experimental results and analysis

IV. A. Experimental data set

In the field of text image generation, common datasets include CUB dataset, Oxford-102 dataset and COCO dataset. In this thesis, the CUB bird dataset and the Oxford-102 flower dataset are selected as the experimental datasets.The CUB bird dataset is a common dataset used for bird identification and localization, which contains about 11,788 images from 200 species of birds, with about 50 images of each species, and each image is annotated with a bird bounding box and label.The Oxford-102 flower The dataset is similar to the CUB dataset in that it covers 102 different flower classes, totaling 8189 images. Each image is equipped with 10 different text descriptions, providing rich material and annotation data for text image generation.

IV. B. Evaluation indicators

In this thesis, two evaluation metrics, IS and FID, are used to quantify the experimental results.IS is one of the important metrics for judging the performance of text-generated image models, which can objectively assess the quality of the generated images and the diversity of the images.IS is mainly computed by using the pre-trained Inception v3 model to assess the KL dispersion between the marginal distribution and the conditional distribution.FID is another important metric for assessing the performance of text-generated image field Another important metric for model performance. It aims to measure the similarity between two datasets, especially between the original image data distribution and the generated image data distribution.FID is also computed using the pre-trained Inception model to encode the features of the images, and subsequently the Fréchet distance between the two is computed.

IV. C. Experimental results

IV. C. 1) Comparison of FID and IS

The generated images are tested using the same test method as the model DF-GAN to calculate the FID and IS scores. Comparing the models MirrorGAN, DMGAN and DF-GAN, the FID comparison and IS comparison of different methods on different datasets are shown in Fig. 2 and Fig. 3. On the CUB dataset, compared to DF-GAN, the DGF-GAN algorithm in this paper reduces the FID from 13.41 to 12.56 and improves the IS from 4.78 to 4.85.On the Oxford-102 dataset, the DGF-GAN algorithm reduces the FID from 40.48 to 38.12, and improves the IS from 3.79 to 4.04.The value of the FID is reduced from 40.48 to 38.12 and from 3.79 to 4.04 on the two datasets decreased by 6.34% and 5.83%, and the IS values were improved by 1.46% and 6.60%. Thus, it can be seen that the images generated by the DGF-GAN model proposed in this paper are closer to the real images, the semantic consistency has been enhanced to some extent, and the image quality has been improved, which is conducive to the generation of new media art design images with better quality.

IV. C. 2) Training efficiency analysis

In order to test the training efficiency of the proposed DGF-GAN model, this paper tests the change of FID values under different training times on the CUB dataset and plots the corresponding curves.The curve of the change of FID metrics is shown in Fig. 4. Compared with the baseline model, the DGF-GAN model proposed in this paper performs better in terms of the speed of convergence, and the convergence is accomplished when the number of training times is up to 250 while the DF-GAN model completes convergence at 300 training times. Specifically, the FID values of the DGF-GAN model in this paper show a faster and stable decreasing trend as the number of training times increases. This result indicates that the model proposed in this paper can achieve good performance more quickly during training.

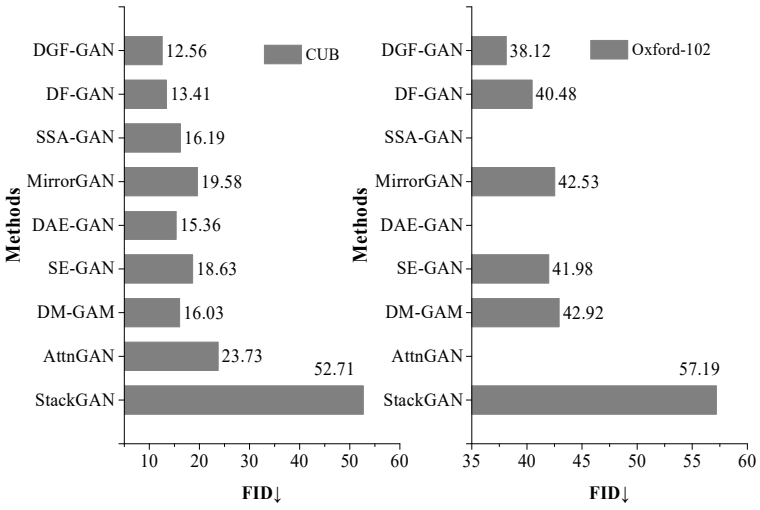

Figure 2: Comparison of FID between different datasets and methods

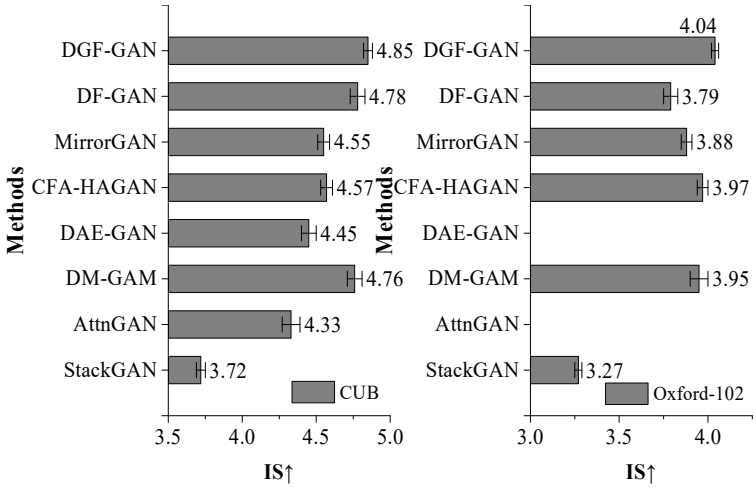

Figure 3: Comparison of IS between different datasets and methods

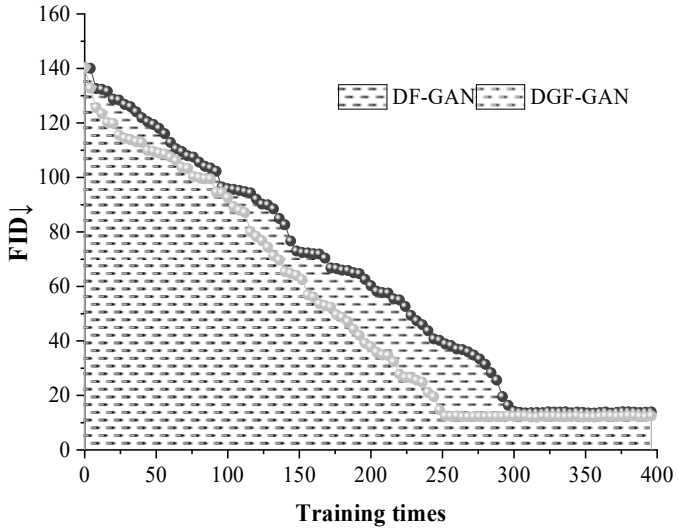

Figure 4: The change curve of the FID index

IV. C. 3) Model efficiency analysis

Meanwhile, this paper compares the model on CUB dataset in terms of training time, generation speed and memory occupancy, where the training time is the average of 10 training times, the generation speed is the time to generate 30,000 random images, and the memory occupancy is analyzed and compared from the process of training and generating the images, respectively, and all other parameters are the same. The results of the comparison of training time and generation speed of the model are shown in Fig. 5 and the memory occupancy of the model during training and generating images is shown in Fig. 6. The training time and generation speed of the DGF-GAN model in this paper are 140.24s and 645.79s, respectively, which are lower than the time required by other methods, and the memory occupancy at training and generating images are 77.68% and 51.31%, respectively, which performs optimally among the comparative methods, and is 3.01% and 2.64% lower than the DF-GAN model. This indicates that the model has better generation capability and efficiency to generate high quality images quickly in practical applications. Overall, the model in this paper has advantages in generating images from text, and has some practical value for new media art design.

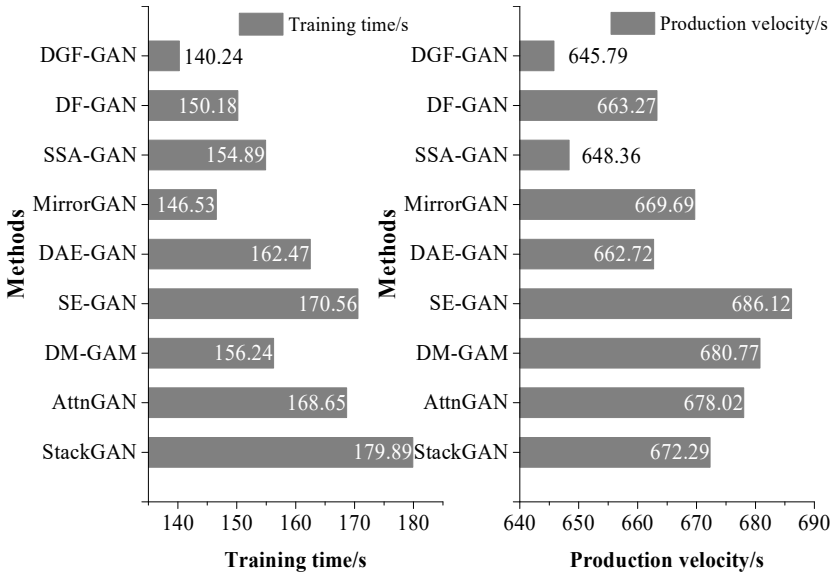

Figure 5: The comparison of the training time and the production velocity of different models

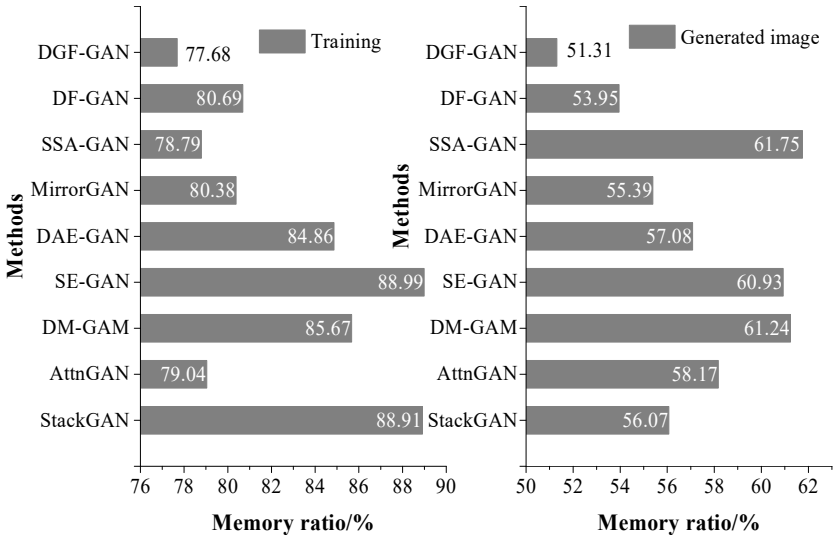

Figure 6: The memory ratio of different models in training and generating images

IV. C. 4) Analysis of ablation experiments

In this paper, a detailed ablation study of the proposed four key modules is carried out on the CUB dataset, and a comparison of the evaluation indexes of the ablation experiments is shown in Table 1, where CA refers to the

conditional enhancement module, RAB refers to the residual affine fusion module, CAM refers to the channel attention module, and PAM refers to the pixel attention module. After adding each module to the original DGF-GAN model, the FID values of the model decreased and the IS values increased, and the FID and IS values of the DGF-GAN model were 12.61 and 4.96, which were the best performances among all the collocations, indicating that the conditional augmentation module, the residual affine fusion module, the channel attention module, and the pixel attention module all have positive effect.

Table 1: Comparison of various evaluation indicators for the overall ablation experiment

Method	Evaluation index		
	FID ↓	IS ↑	
DF-GAN	15.54	4.57	0.07
DF-GAN+CA	15.21	4.64	0.05
DF-GAN+RAB	15.73	4.67	0.04
DF-GAN+CAM	15.49	4.69	0.06
DF-GAN+PAM	15.54	4.72	0.04
DF-GAN+CA+RAB	14.87	4.76	0.04
DF-GAN+CA+CAM	14.23	4.75	0.03
DF-GAN+CA+RAB	14.65	4.71	0.04
DF-GAN+RAB+CAM	14.22	4.79	0.05
DF-GAN+RAB+PAM	13.75	4.88	0.04
DF-GAN+CAM+PAM	13.78	4.85	0.03
DGF-GAN	**12.61**	**4.96**	**0.03**

V. Conclusion

At this stage, the integration and application of computer-aided technology and new media art design is an important method to optimize the effect of art design. This study first summarizes the application of computer-aided creation methods in new media art design, takes the promotion role of AI technology as the starting point, optimizes the single-stage generative adversarial network, and proposes the text-generated image method based on DGF-GAN, which can be used in the auxiliary creation of new media art design.

(1) In the experiments on two datasets, compared with other methods, the DGF-GAN model in this paper has the smallest FID value and the largest IS value. The former is reduced by 6.34% and 5.83% than the benchmark model, and the latter is improved by 1.46% and 6.60%. It shows that the DGF-GAN model can generate images of all categories more stably, with higher image resolution, and the semantic consistency of the generated images, and the diversity of the drawings are improved.

(2) The DGF-GAN model can realize fast convergence in the training process, and its training time and generation speed are the smallest among the compared methods, and the memory occupation in training and generating images is also lower than other models, which is 3.01% and 2.64% lower than the DF-GAN model. It reflects the training efficiency and image generation efficiency of DGF-GAN model, which is favorable for designers to use in the auxiliary creation of new media art design.

Computer technology provides innovative tools and creative resources for new media art design through data analysis, pattern recognition and generative algorithms, which not only gives design inspiration and optimizes the design process, but also makes the art design works more in line with the aesthetic needs of modern society. Computer-aided creation methods should be used in art design to promote the diversified development of art design and enhance the artistic value.

Reference

[1] McMullan, J. (2020). A new understanding of 'New Media': Online platforms as digital mediums. Convergence, 26(2), 287-301.
[2] Karidi, M. (2018). News media logic on the move? In search of commercial media logic in German news. Journalism Studies, 19(9), 1237-1256.
[3] O'Halloran, K. L., Pal, G., & Jin, M. (2021). Multimodal approach to analysing big social and news media data. Discourse, Context & Media, 40, 100467.
[4] Manganello, J., Bleakley, A., & Schumacher, P. (2020). Pandemics and PSAs: Rapidly changing information in a new media landscape. Health communication, 35(14), 1711-1714.
[5] Jagodzinski, J. (2021). A meditation on the post-digital and post-internet condition: screen culture, digitalization, and networked art. Post-Digital, Post-Internet Art and Education: The Future is All-Over, 61-80.

[6] Plantin, J. C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. New media & society, 20(1), 293-310.
[7] Waldfogel, J. (2015). Digitization and the quality of new media products. Economic analysis of the digital economy, 407.
[8] Tvrdišić, S. (2022). The impacts of digitalization on traditional forms of art. AM Časopis za studije umetnosti i medija, (27), 87-101.
[9] Ye, W., & Li, Y. (2022). Design and research of digital media art display based on virtual reality and augmented reality. Mobile Information Systems, 2022(1), 6606885.
[10] Kamposiori, C., Mahony, S., & Warwick, C. (2019). The impact of digitization and digital resource design on the scholarly workflow in art history. International Journal for Digital Art History, (4), 3-11.
[11] Orlova, A. V. (2021, July). Digitizing Art or How to Broaden the Viewer's Experience. In Proceedings of EVA London 2021 (pp. 35-38). BCS Learning & Development.
[12] Wang, R. (2021). Computer-aided interaction of visual communication technology and art in new media scenes. Computer-Aided Design and Applications, 19(S3), 75-84.
[13] Liu, F., Gao, Y., Yu, Y., Zhou, S., & Wu, Y. (2021). Computer aided design in the diversified forms of artistic design. Computer-Aided Design and Applications, 19(3).
[14] Xu, S. (2022). Research on Graphic Design of Digital Media Art Based on Computer Aided Algorithm. In 3D Imaging—Multidimensional Signal Processing and Deep Learning: 3D Images, Graphics and Information Technologies, Volume 1 (pp. 207-215). Singapore: Springer Nature Singapore.
[15] Ye, W., & Li, Y. (2022). Performance characteristics of digital media art design relying on computer technology. Mobile Information Systems, 2022(1), 2203259.