Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Design of a neural network-based automated style migration technique in digital media art

Jiayi Yu[1, ✉]

[1] The Shanghai Conservatory of Music, The Institute of Digital Media Art, Shanghai, 200031, China

ABSTRACT

The continuous development of neural network makes the automated style migration technology also rise to a new height. This paper selects digital media art as the research field, constructs Cycle GAN, a cyclic consistent generative adversarial network structure applied to digital media art, on the basic framework of GAN, and optimizes it by adding bilinear interpolation and attention mechanism, so as to build up a style migration model for digital media art. In the style migration simulation experiment, the IS test values of this paper's model on the photo2vangogh and photo2monet datasets are 5.32 and 6.03, and the FID test values are 97.52 and 75.55, which are better than the other comparative models. Similarly, the optimized performance of FID, SSIM and PSNR values on the dataset is also better than other comparative models, and the style migration performance of the model is verified. Using the model of this paper to design a digital topography with Chinese traditional ink painting as the content, we explore the correlation between the design attributes of the style migration design works in digital media art and the audience's cognitive evaluation and overall perception. Among the design attributes, "plot relevance" (4.375) and "atmosphere rendering" (3.38) have the highest T-value, which is the most important influence on audience perception.

Keywords: Generative Adversarial Networks, Bilinear Interpolation, Attention Mechanism, Style Migration, Digital Media Art

1. Introduction

With the development of the times and the emergence of social media, images play an increasingly important role in people's communication. Compared with the direct use of text for communication, sharing through images is more in line with everyone's reading habits and preferences, and also more able to satisfy everyone's desire to share. Especially in the popularization of hardware such as cell phones and computers and the extremely developed social platforms today, people are keen to share their own image works on social platforms in order to attract the attention of others. However, a single

✉ Corresponding author.
E-mail address: Aabc18321858305@126.com (J. Yu).
Received 01 March 2024; Revised 25 May 2024; Accepted 10 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-527

image processing technology will only bring about aesthetic fatigue, and it is difficult to rely on one's own aesthetics and technology to make good-looking and unique images. Therefore, the style migration technique was born, and people began to explore how to simulate the unique style of art images through computer algorithms, and transfer it to their own images, so that they also have a unique artistic style [1-3].

Artistic style migration, as one of the popular research topics in the field of computer vision, aims to skillfully integrate the unique style of artistic images into content images to create new images full of artistic flavor [4-5]. Extracting the deep information of images with the help of deep learning techniques and realizing the complex transformation of artistic features between images are the key factors that have made significant progress in art style migration [6-9]. Deep neural networks have superior results in creating artistic images, and Neural Style Migration (NST) is beginning to receive more and more attention, which has become an important milestone in style migration research [10-13]. Since NST technique can simulate a wide range of art styles without the need for separate design or tedious operations to generate a specific style, it is highly favored by users and developers [14-16]. This has made it possible to use style migration techniques in image processing, video production, game development, and other areas.

Gatys, L. A. et al. optimized the image target recognition function using convolutional neural networks to clarify the high-level semantic information in the image, based on which they designed an artistic style neural algorithm to achieve the generation of new images with high perceptual quality by separating and restructuring the content and style in natural images [17]. Chen, H. et al. introduced two contrast losses to construct an internal-external style migration method to address the artifacts caused by color disharmony and repetition in the original art style migration method, which not only enhances the visual effect of the generated art images, but also improves the stability and consistency of the rendered video clips [18]. Semmo, A. et al. showed that image-based rendering of example art relies on image pre-training and low-level image feature texture migration, leading to limitations in neural style migration for modeling high-fidelity media art, and proposed activation of a neural network layer to match content and style statistics to make generalized style migration possible [19]. Luan, F. et al. introduced a deep-learning photographic style migration method to constrain the output of an image by local affine in the color space, which effectively suppresses the picture distortion in order to produce a realistic photographic style migration effect [20]. Kumar, T. S. et al. investigated the application of neural style migration on video sequences by introducing the VGG16 model to construct a content function to compute the style information of the video frames, which is combined with a balanced loss function to compute the style and content loss to optimize the generation process of the subsequent images, and ultimately to achieve an immersive visual experience [21]. Ruder, M. et al. proposed a neural style migration network initialization and loss function suitable for video, which avoids the need for manual mapping during the style migration process by generating consistent and stable stylized video sequences that migrate the style of a single image to the entire video sequence [22]. Kadish, D. et al. described the problem of crossdrawing in neural network object detection with limited recognition of texture rather than shape, so a process of training neural networks was designed to generate a large dataset for training and validation to improve the original image processing process [23]. Chen, H. et al. constructed a Dual Style-Learning Art Style Migration (DualAST) framework that learns the overall art style of a group of artworks to determine the image tone while learning the style of a specific artwork to determine the image generation details, which is combined with a style control module to achieve effective art style migration [24].

In the beginning of this paper, the network structure and mathematical principles of Generative Adversarial Network (GAN), a classical neural network model, are discussed in detail, and Cycle GAN, a cycle-consistent generative adversarial network, and Disco GAN, another representative unsupervised method, are introduced as two kinds of GAN models. In the subsequent design of the digital media art style migration model, the Cycle GAN model is selected as the basis for the improvement of the network model structure. A multi-head self-attention mechanism is added between the encoder and the converter to enhance the model's understanding of the global structure of the image, and the inverse convolution in the decoder's decoding network is replaced with bilinear interpolation to reduce the artifacts generated by the inverse convolution operation, thus improving the structure of the original CycleGAN network model and enhancing the effect of style migration generation in digital media art. The style migration performance of this paper's model is examined through quantitative comparison experiments and ablation experiments. The model of this paper is used to design a digital topography with Chinese ink painting as its content, which is used as a representative of style migration design works in digital media art, to explore the correlation between digital media art design and audience's cognition and perception, and to analyze the influence of the audience's cultural level on their own cognitive impact on the design attributes of digital media art works.

2. Generating Adversarial Network Models

Generative Adversarial Network (GAN) is a neural network model based on adversarial learning [25]. The GAN model is based on the idea of zero-sum game and aims to learn the distributional patterns of data through the adversarial process between generators and discriminators.

2.1. Network structure of generative adversarial networks

The network structure of Generative Adversarial Network (GAN) consists of two parts: generator network and discriminator network.

1) Generator. The main task of the generator is to transform a random input vector (usually Gaussian distributed random noise) into sample data similar to the training data. The network structure of the generator is usually a neural network consisting of multiple fully connected layers and convolutional layers, where the convolutional layers are used to process image data and the fully connected layers are used to process other types of data. The output of the generator is a sample data similar to the training data.

2) Discriminator. The main task of the discriminator is to distinguish the sample data generated by the generator from the real training data. The network structure of the discriminator is also usually a neural network consisting of multiple fully connected and convolutional layers, where the convolutional layer is used to process image data and the fully connected layer is used to process other types of data. The output of the discriminator is a scalar value that indicates whether the input data is real training data or sample data generated by the generator.

2.2. Mathematical principles of generative adversarial networks

Maximum likelihood is used to determine the optimal parameters θ. First, we take m samples from a given real dataset $P_{data(x)}$ to form a collection $\{x_1, x_2, x_3, \cdots, x_m\}$ as a training set to compute the corresponding generative model distribution probabilities $P_{G(x;\theta)}$ and construct the likelihood function as follows:

$$L = \prod_{i=1}^{m} P_G\left(x^i;\theta\right) \tag{1}$$

To maximize the probability of generating a distribution P_G close to the data distribution P_{data} of the real dataset, that is, the probability of sampling xi each real distribution needs to be maximized so that the product of the probabilities of sampling xi all real distributions is maximized, i.e., we just need to find a suitable θ to make L as large as possible, and the problem becomes finding the optimal set of parameters θ^*:

$$\theta^* = \arg\max_\theta \prod_{i=1}^{m} P_G\left(x^i;\theta\right) \tag{2}$$

In order to simplify the computational process, we take logarithms of the likelihood function and convert multiplication to addition, as shown in equation (3):

$$\begin{aligned}\theta^* &= \arg\max_\theta \prod_{i=1}^{m} P_G\left(x^i;\theta\right)\\ &= \arg\max_\theta \log \prod_{i=1}^{m} P_G\left(x^i;\theta\right)\\ &= \arg\max_\theta \sum_{i=1}^{m} \log P_G\left(x^i;\theta\right)\\ &\approx \arg\max_\theta E_{x\sim p_{data}}\left[\log p_G(X;\theta)\right]\end{aligned} \tag{3}$$

To minimize the difference between $P_{G(x;\theta)}$ and $P_{data(x)}$, we use the KL scatter as a metric [26]. By applying some deformations and simplifications to the KL scatter, we obtain a formula (4) that contains an integral term. To further simplify, derive this formula and ensure computational efficiency, we add an integral term that does not affect the optimal solution of θ^*, which is finally transformed into a minimization to find the optimal parameter set θ^*, thus minimizing the KL scatter between P and $P_{data(x)}$:

$$\begin{aligned}\theta^* &= \arg\max_\theta \left(\int P_{data}(x)\log P_G(x;\theta)dx - \int P_{data}(x)\log P_{data}(x)dx\right)\\ &= \arg\max_\theta \int P_{data}(x)\log\frac{P_G(x;\theta)}{P_{data}(x)}dx\\ &= \arg\min_\theta \int P_{data}(x)\log\frac{P_{data}(x)}{P_G(x;\theta)}dx\\ &= \arg\min_\theta \int P_{data}(x)\log\frac{P_{data}(x)}{P_G(x;\theta)}dx\end{aligned} \tag{4}$$

Understanding the above equation from the GAN perspective is that the generator generator defines a distribution P_G and the discriminator acts as an evaluator to compute L. Our goal is also to train an optimal generator G_{best} so that P_G is as close as possible to P_{data}, which translates into the formula:

$$G_{best} = \arg\min_G Div\left(P_G, P_{data}\right) \tag{5}$$

where Div is expressed as a dispersion or a distance. At this point, if it is known P_{data} distribution, it can be a good solution to the above problem (for example, the data obey the normal distribution, then set the parameters of the normal distribution is unknown, the use of BP can be a good fit), but now the distribution is unknown. Previously said discriminator is similar to an evaluator, the purpose is to separate real samples or generated samples - can be regarded as a binary classification task. Then, its objective function can be formulated:

$$V\left(G,D\right) = E_{x\sim P_{data}}\left[\log D(x)\right] + E_{x\sim P_G}\left[\log(1-D(x)\right] \tag{6}$$

And the goal of the discriminator is to maximize this equation, i.e.

$$D_{best} = \arg\max_D V\left(D,G\right) \tag{7}$$

Theoretically $D(x)$ could be any function, rewriting the above equation further:

$$V(D,G)=\int_x \left[P_{data}(x)\log(D(x))+P_G(x)\log(1-D(x))\right]dx \tag{8}$$

When training the discriminator the generator is usually fixed, so P_G can be regarded as a constant and P_{data} is clearly a constant, rewriting the above equation as

$$V(D,G)=\int_x a\times\arg(D(x))+b\times\log(1-D(x))\,dx \tag{9}$$

To maximize the integral, the product function must be maximized at each position. Derive the product function so that the derivative is equal to zero:

$$\begin{gathered} f=a\times\log(D(x))+b\times\log(1-D(x)) \\ \frac{\partial f}{\partial x}=0 \\ \frac{a}{D}=\frac{b}{1-D} \end{gathered} \tag{10}$$

You can get D_{best}:

$$D_{best}(x)=\frac{P_{data}(x)}{P_{data}(x)+P_G(x)} \tag{11}$$

Substituting into V gives:

$$\begin{aligned} V(D,G)&=-2\log 2+KL\left(P_{data}\left\|\frac{P_{data}+P_G}{2}\right.\right)+KL\left(P_G\left\|\frac{P_{data}+P_G}{2}\right.\right) \\ &=-2\log 2+2JSD\left(P_{data}\|P_G\right) \end{aligned} \tag{12}$$

where JSD denotes JS scatter. Therefore, the task of the discriminator is to recognize the difference between the distribution of the generated data and the original training data, and the specific measure is the JS scatter [27]. Going back to Eq. (6) to solve the problem of incomputable scatter:

$$G_{best}=\arg\min_G\max_D V(G,D) \tag{13}$$

2.3. Generative Adversarial Network Architecture Applied to Digital Media Arts

Generally, according to the different data sources, the style transfer methods based on generative adversarial networks can be roughly divided into two categories, one is unsupervised generative adversarial networks, and the other is supervised generative adversarial networks. Among the supervised methods, the more well-known is the conditional generative adversarial network, which requires paired dataset training, i.e., paired raw and stylistic images. Among the unsupervised methods, the representative is the cyclic consistency generative adversarial network (Cycle GAN), which does not require paired image datasets, and since the cyclic consistency generative adversarial network will be described in detail later, this chapter will introduce another representative unsupervised method, also known as the "twin brother" of Cycle GAN, Disco GAN [28].

1) CGAN. For the purpose of conditional GAN, additional information is added to both discriminator and generator y, y which can be aliased labels or other types of data. During training CGAN adds auxiliary information y to the input z of the generator network and also adds auxiliary information y to the input x of the discriminator network.The loss function of CGAN is shown in equation (14):

$$\min_G \max_D V(G,D) = E_{x\sim P_{data}}[\log D(x|y)] + E_{z\sim P_G}[\log(1-D(G(z|y)))] \tag{14}$$

From the formula, CGAN is equivalent to adding a condition to both the generator and discriminator parts of the original GAN, and this loss function is closely related to the input information y, which means that for different labels, it can be understood that there are different objective functions, which better control the category of the outputs, and make the directionality of the generated images improved.

2) Disco GAN. Disco GAN is a model designed to achieve image migration between different domains, such as converting sketches to real photos, converting daytime photos to nighttime photos, and so on. Image conversion is achieved by training two complementary generator networks and two complementary discriminator networks.

The Disco GAN model is represented in a mathematical expression as shown in equation (15):

$$\begin{aligned} x_{AB} &= G_{AB}(x_A) \\ x_{ABA} &= G_{BA}(x_{AB}) = G_{BA} \circ G_{AB}(x_A) \end{aligned} \tag{15}$$

The difference between the image processed by the two generators G_{AB} and G_{BA} and the original image is L_{CONSTA} and its expression is shown in equation (16):

$$L_{CONST_A} = d\left(G_{BA} \circ G_{AB}(x_A), x_A\right) \tag{16}$$

And the standard GAN generator loss function is:

$$L_{GAN_B} = -E_{x_A\sim P_A}\left[\log D_B\left(G_{AB}(x_A)\right)\right] \tag{17}$$

Therefore, the loss function of generator G_{AB} is expressed as:

$$L_{G_{AB}} = L_{G_{AN_B}} + L_{CONST_A} \tag{18}$$

The loss function of discriminator B is denoted as:

$$L_{D_B} = -E_{x_{x_B\sim P_B}}\left[\log D_B(x_B)\right] - E_{x_{A_A}\sim P_A}\left[\log\left(1 - D_B\left(G_{AB}(x_A)\right)\right)\right] \tag{19}$$

The real samples X_A in domain A are processed through G_{AB} are all going to be inside domain B, similarly the real samples X_B in domain B are processed through G_{BA} are all going to be inside domain A. When training the model, the samples in domains A and B will both be processed, so the generator loss function of the Disco GAN model can be expressed as:

$$L_G = L_{G_{AB}} + L_{G_{BA}} \tag{20}$$

The discriminator loss function is:

$$L_D = L_{D_A} + L_{D_B} \tag{21}$$

3. Digital Media Art Style Migration Model

With the continuous development of science and technology and economy, people's experience requirements for said digital media art are also changing, and the emergence of style migration technology provides a solution to people's specific art needs. In the above paper, this paper proposes a kind of neural network model based on adversarial learning, Generative Adversarial Network (GAN), and correspondingly proposes the structure of Generative Adversarial Network applied to digital media art. In this chapter, based on its theoretical foundation, two optimization modules, namely, multi-head self-attention mechanism and bilinear interpolation method, are added to build a digital media art style migration model.

3.1. Cyclic Consistency Generating Adversarial Network Model

Cycle GAN belongs to an unsupervised cross-domain image style migration model. It consists of two generative adversarial networks that mirror each other to form a cyclic network. Cycle GAN uses two generators and two discriminators to realize the mutual mapping between X and Y. Generator G defines the mapping from X to Y and Generator F defines the mapping from Y to X. Discriminator D_X is used to determine whether the data is from X or from the generated $F(y)$, and discriminator D_Y is used to distinguish whether the data is from Y or from the generated $G(x)$. CycleGAN also introduces cyclic consistency loss, where the image x goes through the mapping to y and back to x, and calculates the loss between the original image x and the mapped image in order to make $F\left(G(x)\right)\approx x$ and $G\left(F(y)\right)\approx y$. CycleGAN is a widely applied widely used model for image style migration and is unique in that it does not require one-to-one training. This flexibility gives CycleGAN the potential for a wide range of applications in various fields.

The adversarial loss is consistent with the original GAN. For generator G and discriminator D_Y, the objective function is expressed as:

$$\begin{aligned} L_{GAN}\left(G, D_Y, X, Y\right) = {} & E_{y\sim P_{data}(y)}\left[\log D_Y(y)\right] \\ & + E_{X\sim P_{data}(x)}\left[\log\left(1-D_Y\left(G(X)\right)\right)\right] \end{aligned} \tag{22}$$

The objective function of generator F and discriminator D_X is $L_{GAN}\left(G, D_Y, X, Y\right)$ as follows:

$$\begin{aligned} L_{GAN}\left(F, D_X, Y, X\right) = {} & E_{x\sim P_{data}(x)}\left[\log D_X(x)\right] \\ & + E_{y\sim P_{data}(y)}\left[\log\left(1-D_X\left(G(Y)\right)\right)\right] \end{aligned} \tag{23}$$

The cyclic consistency loss, on the other hand, uses the L1 paradigm to directly compute the gap between the generated image and the real image after the input image goes from the X-domain to the Y-domain and then to the X-domain:

$$\begin{aligned} L_{cyc}\left(G, F\right) = {} & E_{x_x\sim P_{data}(x)}\left[\left\|F\left(G(X)-1\right)\right\|_1\right] \\ & + E_{y\sim P_{data}(y)}\left[\left\|G\left(F(Y)-1\right)\right\|_1\right] \end{aligned} \tag{24}$$

So the final loss is composed of these three components, then the complete objective function of Cycle GAN is:

$$\begin{aligned} L\left(G, F, D_X, D_Y\right) = {} & L_{GAN}\left(G, D_Y, X, Y\right) \\ & + L_{GAN}\left(F, D_X, Y, X\right) + \lambda L_{cyc}(G, F) \end{aligned} \tag{25}$$

where X and Y represent the data distributions in the source and target domains, respectively, and x and y are sample data from the two data domains, respectively. generator G learns the mapping from the source domain X to the target domain Y, while generator F learns the mapping from Y to X in the opposite direction. D_X, D_Y are denoted as discriminators, and the weight coefficients λ are used to balance the consistency of the mappings against the adversarial loss.

3.2. Image style migration model based on improved CycleGAN

3.2.1. Multiple self-attention mechanisms. The self-attention mechanism is a mechanism that can model global dependencies within a sequence. It obtains a new feature representation of a sequence by computing the correlations between positions in a sequence. Self-attention consists of three vector representations: query vector $Query$, key vector Key and value vector $Value$. The model computes the inner product correlation between the query vector and the key vector to obtain an attention distribution, i.e., the correlation weight of each position with respect to the other positions of the sequence. This attention weight is then used to do a weighted sum with the value vectors to obtain a new feature representation of the entire sequence, called the attention output. Compared with the traditional RNN, this global attention mechanism can model the dependencies of arbitrary distances in long sequences more efficiently and in parallel. The self-attention formula is as follows:

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{26}$$

A multi-head self-attention mechanism consists of multiple self-attention modules, which are essentially the integration of multiple self-attention computations [29]. Each self-attention computation is called a "head", and different heads are independent of each other. The multi-head mechanism can learn different subspace representations of queries, keys, and values in parallel.

The specific computation process of the multi-head self-attention layer is as follows.

1) Obtaining Q, K and V vectors: Suppose the input sequence is $X = [x_1, \ldots, x_L] \in \mathbb{R}^{L\times D}$, where L is the length of the sequence and D is the dimension of the input vector. The input sequence X is first linearly transformed to obtain for each x_i its corresponding query vector $q_i \in \mathbb{R}^{D}$, key vector $k_i \in \mathbb{R}^{D}$ and value vector $v_i \in \mathbb{R}^{D}$. The linear transformation process for the whole input sequence X can be expressed as:

$$Q = XW^Q \in \mathbb{R}^{L\times D} \tag{27}$$

$$K = XW^K \in \mathbb{R}^{L\times D} \tag{28}$$

$$V = XW^V \in \mathbb{R}^{L\times D} \tag{29}$$

where $W^Q \in \mathbb{R}^{D\times D}$, $W^K \in \mathbb{R}^{D\times D}$, $W^V \in \mathbb{R}^{D\times D}$ are learnable linear mapping matrices. Next, the vector subset $\left\{Q_i \in \mathbb{R}^{L\times Dh}, K_i \in \mathbb{R}^{L\times Dh}, V_i \in \mathbb{R}^{L\times Dh} \middle| i = 1,2,\ldots,H\right\}$ used for each head can be obtained by splitting it into H equal parts according to the number of heads along the direction of the last dimension of the mapped Q, K, V vectors, i.e., the dimension D of the vectors, which $D_h = \frac{D_v}{H}$ denotes the dimension of the vectors for each head pair. Then, the self-attention computation needs to be performed based on the Q, K, and V vectors of each head.

2) Self-attention calculation. Calculate the self-attention of each head separately, taking the j th head $head_j$ as an example, Eq:

$$head_i = Attention(Q_j, K_j, V_j) \in \mathbb{R}^{L\times Dh} \tag{30}$$

3) Multiple results fusion. Multiple self-attention results are fused to obtain the final output vector Z.

$$\begin{aligned} Z &= MultiHeadAttention \\ &\triangleq (head_1 \oplus head_2 \oplus \ldots \oplus head_H)W \end{aligned} \tag{31}$$

where $\oplus$ denotes a splice along the last dimension of the vector, and the dimension of the spliced vector may not be equal to the dimension of the original vector D, so the matrix W^O is multiplied here to map the vector to the original input dimension. That is

$$MultiHead\left(Q,K,V\right)=Concat\left(head_i,\cdots,head_h\right)W^O \tag{32}$$

$$head_i=Attention\left(QW_i^Q,KW_i^K,VW_i^V\right) \tag{33}$$

3.2.2. Bilinear interpolation method. Bilinear interpolation does not produce the problem of mismatch between convolution kernel and step size, and any scaling multiples can be accurately interpolated, thus effectively eliminating the checkerboard effect of anti-convolution. Bilinear interpolation has a small amount of computation, which improves the efficiency while guaranteeing the effect. The specific principle is:

Let the image to be up-sampled be I and its size be $H\times W$. Scale it to obtain image I' with size $rH\times rW$ (r being the scaling factor). For each pixel $\left(x,y\right)$ in I', first calculate its corresponding coordinate $\left(x',y'\right)$ in the original image I:

$$x'=\frac{x}{r} \tag{34}$$

$$y'=\frac{y}{r} \tag{35}$$

The pixel values are then calculated by the bilinear interpolation formula:

$$\begin{aligned}I'\left(x,y\right)=&(1-\lambda x)(1-\lambda y)I\left(x'',y''\right)+\lambda x(1-\lambda y)I\left(x'',y'+1\right)\\&+(1-\lambda x)\lambda yI\left(x'+1,y''\right)+\lambda x\lambda yI\left(x''+1,y''+1\right)\end{aligned} \tag{36}$$

where $\lambda x=x'-x''$, $\lambda y=y'-y''$, x'' are the coordinates obtained by rounding down x', and y'' is the coordinate obtained by rounding down y'.

3.2.3. Improving the network structure of the model. This chapter presents an improved CycleGAN modeling approach for style migration tasks in digital media art. The improved network is based on the generator structure of the original CycleGAN in the following two aspects: first, a multi-head self-attention mechanism is added between the encoder and the converter, and a new network structure connection from the encoding layer to the converter layer is designed to enable better output of style features to the converter layer. The self-attention module can model the global dependencies between different regions of the image and capture the intrinsic structural information of the image. The multi-head design can learn the feature representations of different subspaces, and different heads can focus on different global structural information of the input image, such as shape, texture, etc., and aggregate the information to enhance the model's understanding of the global structure of the image. This parallel multi-task learning approach can enhance the effect of style migration, thus generating resultant images with more natural style migration. Second, in the decoder's decoding network, the inverse convolution is replaced with bilinear interpolation for up-sampling. Bilinear interpolation can obtain more natural and smooth sampling results, effectively reducing artifacts such as the checkerboard effect that is easily produced by the anti-convolution operation. This further optimizes the visual quality of the style migration results. By incorporating the self-attention mechanism with the use of bilinear interpolation, the improved CycleGAN network proposed in this paper further enhances the generation of style migration in digital media art while maintaining the processivity of the original network.

4. Digital Media Art Style Migration Simulation Experiment

In order to verify the performance of the experimental results of the digital media art style migration model based on the generative adversarial network model constructed in this paper, two commonly used datasets for style migration, photo2vangogh and photo2monet, are used in this chapter to carry out the style migration simulation experiments.

4.1. Experimental environment and parameter settings

During the training phase, the specific setup of the experiment and the parameters of the network model are specified in Table 1.

Table 1. Experimental environment and parameter setting

Experimental environment		Parameter setting	
Environmental name	Specification	Parameter name	Parameter value
Operating system	Ubuntu 20.04	Beta	(0.5,0.999)
CPU	Intel(R)Xeon(R)Silver 4210R CPU@2.40GHz	Optimizer	Adam
GPU	4×Nvidia RTX 3090	Batch	1
Memory	256GB DDR4	Initial learning rate	0.0002
Programming language	Python 3.8	Decay period	100
Depth learning framework	PyTorch 1.11.0	Epoch	200

4.2. Quantitative comparison experiments

The fluctuation amplitude of the total loss function of the digital media art style migration model constructed in this paper on the photo2vangogh and photo2monet datasets is specifically shown in Fig. 1. It can be intuitively seen that although the fluctuation amplitude of the first 100 epochs is too large, the direction of the total loss function of the later 100 epochs tends to be smooth. This indicates that the proposed model compares the image generated by the generator with the input image under the constraint of the associated loss function, which makes it more difficult for the discriminator to determine the authenticity of the image. The content of the image generated by the model in this paper is consistent with the input image, and the style of the generated image is consistent with the style of the target image.

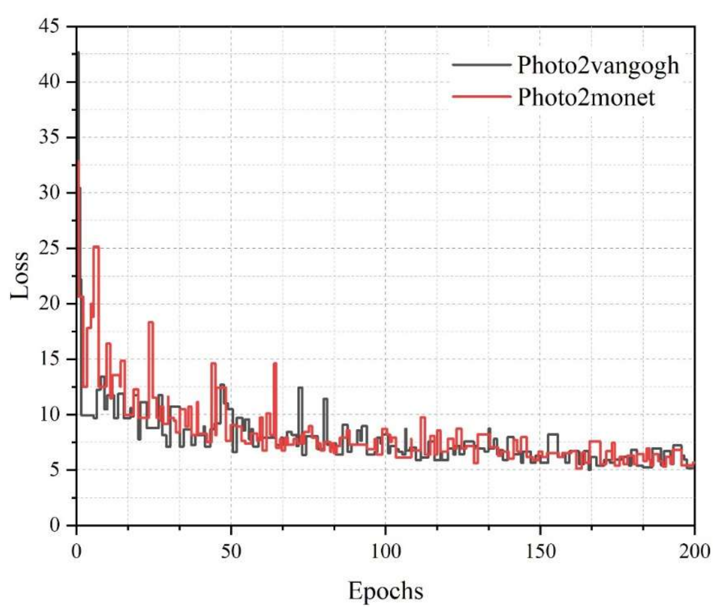

Fig. 1. Total loss function

To further validate the effectiveness of the model in this paper, the images generated by the model need to be evaluated for IS and FID quality. The quality evaluation results are specifically shown in Table 2. The IS test values of this paper's model in both photo2vangogh and photo2monet datasets are higher than those of other comparative models, reaching 5.32 and 6.03, respectively, while the FID test values are both lower than those of the other comparative models, only 97.52 and 75.55.This indicates that this paper's model is consistent with the quantitative results, i.e., it can better balance the style migration results in terms of style similarity and the degree of content preservation, and also reduces the occurrence of common problems in digital media art image generation at the same time.

Table 2. IS and FID

	Photo2vangogh		Photo2monet	
Model	IS	FID	IS	FID
MUNIT (Cyclic uniform loss is 10.00)	5.2±0.80	113.39	4.73±0.42	108.22
UNIT (Cyclic uniform loss is 10.00)	5.17±0.48	101.63	5.01±0.36	155.38
CycleGAN (Identity mapping loss=0.5)	4.15±0.33	151.27	5.08±0.37	93.72
UGATIT	4.42±0.34	187.63	3.72±0.43	128.74
Cycle-DPN-GAN	4.56±0.32	154.82	5.20±0.56	88.75
Model of this article	5.32±0.75	97.52	6.03±0.87	75.55

4.3. Ablation experiments

In order to test the validity of each part of the model in this paper, ablation experiments are carried out on the photo2vangogh and vangogh2photo datasets. The experimental results are specifically shown in Table 3. From the results, it can be observed that, comparing with the original CycleGAN model, the intermediate models using only bilinear interpolation or attention mechanism are both effective in improving the visual effect of style migration. When both bilinear interpolation and attention mechanism are used, i.e., the model in this paper, the style transformation of the generated image is the most natural and realistic, which proves the synergistic effect of the two modules. The values of FID, SSIM and PSNR of this paper's model are 153.718, 0.671 and 21.153 on the photo2vangogh dataset, and 149.669, 0.68 and 21.914 on the vangogh2photo dataset, respectively.Obviously, this paper's model outperforms the other models in terms of the three indexes of FID, SSIM and PSNR. Numerical optimization performance are better than other comparative models, which fully illustrates the positive effect of the two improved modules of bilinear interpolation and attention mechanism on optimizing the effect of CycleGAN style migration.

Table 3. The results of the ablation experiment

Model	Photo2vangogh			Vangogh2photo		
	FID	SSIM	PSNR	FID	SSIM	PSNR
CycleGAN	218.459	0.454	17.034	225.917	0.428	17.442
CycleGAN(bilinear)	169.166	0.599	19.532	176.702	0.505	18.35
CycleGAN(attention)	161.632	0.603	20.407	165.384	0.597	19.933
Model of this article(bilinear+attention)	153.718	0.671	21.153	149.669	0.68	21.914

5. Digital Media Art Style Migration Design Practices

With the help of the digital media art style migration model proposed in this paper, this paper designs a digital topiary with traditional Chinese ink painting as the core content through the automated style migration function with the purpose of cultural relics protection and cultural heritage. Through communication with experts, this paper concludes that the design perception of digital media art style migration design works belongs to six basic attributes in three dimensions. Specifically as follows.

1) Visual layer, i.e., visual attractiveness, novelty, and theme matching.

2) Psychological layer, i.e., the degree of atmosphere rendering and plot relevance.

3) Emotional layer, i.e., the degree of emotional mobilization.

In this chapter, using the Internet as a medium for the main audience groups of digital media art, the questionnaire is scored using a five-point Likert scale, i.e., 1 point represents the weakest result of perceptual evaluation, and 5 points is the strongest. Before the respondents formally evaluate and fill in the questionnaire, the digital topography of the ink painting designed in this paper will be displayed to avoid the respondents losing objectivity when evaluating the feedback. A total of 308 data results were collected and analyzed for credibility, and the results showed that the Cronbanch'sAlpha value was 0.894, which far exceeded the minimum acceptable value of 0.7. The gender, disciplinary background, and cultural level of the respondents in the 308 data are shown in Table 4.

Table 4. The information of the interviewees

<table>
<tr><th>Factor</th><th>Feature</th><th>Number</th><th>Total number</th></tr>
<tr><td rowspan="2">Gender</td><td>Man</td><td>160</td><td rowspan="2">308</td></tr>
<tr><td>Female</td><td>148</td></tr>
<tr><td rowspan="2">Subject background</td><td>Liberal arts</td><td>195</td><td rowspan="2">308</td></tr>
<tr><td>Science major</td><td>113</td></tr>
<tr><td rowspan="3">Cultural horizon</td><td>Undergraduate</td><td>155</td><td rowspan="3">308</td></tr>
<tr><td>Graduate student</td><td>122</td></tr>
<tr><td>Other</td><td>31</td></tr>
</table>

5.1. Correlation analysis between design attributes and audience perception

The study took “overall design perception” as the dependent variable and the other six factors as independent variables for regression analysis, and the results of the regression analysis are shown in Table 5. The results show that “Plot relevance” and “Atmosphere rendering” have reached a strong significance level of 3*, sig<0.001. “Design-Theme Match” also reached a significance level of two *'s, sig<0.01. In terms of T-value size, “Plot relevance” (4.375) and “Atmosphere rendering” (3.38) ranked first and second, respectively, which indicates that these two design attributes have an extremely important influence on the respondents' perception of digital media art and design works. The three attributes of “visual attractiveness”, “visual novelty” and “emotional mobilization” are not significant, which indicates that these three attributes are not the reasons for the overall perception and recognition of the digital topiary designed in this paper. This suggests that these three attributes are not the reason for the overall perception and recognition of the ink painting digital topography, i.e., digital media art design designed in this paper.

Table 5. Regression analysis

Dependent variable	Simple correlation		Normalized regression coefficient	T	Sig.
Overall design perception	Visual attraction	0.562**	0.076	1.165	0.244
	Design novelty	0.460**	-0.052	-0.891	0.372
	Theme matching	0.676**	0.204**	2.707	0.006

	Emotional mobility	0.660**	0.12	1.432	0.152
	Ambient rendering	0.644**	0.218***	3.38	0.000
	Plot correlation	0.701**	0.335***	4.735	0.000
*** indicates p<0.001,** indicates p<0.01,* indicates p<0.05					

5.2. Analysis of the influence of cultural level factors on the perception of design attributes

This study first makes the null hypothesis that there is no variability in the evaluation and overall perception of the six attributes of the ink painting digital topiaries designed in this paper due to the respondents' gender factors and disciplinary backgrounds, and from the results of the independent samples t-test, the significance (sig) of both the evaluation of the basic attributes of the design of the ink painting digital topiaries and the overall perception is >0.05, which means that there is no variability of the evaluation of the design attributes of the ink painting digital topiaries, i.e., digital media art design due to the differences in the gender and the disciplinary backgrounds. That is to say, there is no difference in the evaluation of the design attributes and overall perception of ink painting digital topiaries, i.e., digital media art design, due to differences in gender and academic background.

In this section, we will focus on the influence of the respondents' cultural level factor on the design attribute evaluation and overall perception of the ink painting digital topography designed in this paper. The analysis of the number of variables of cultural level and the perception of digital media art design attributes is specifically shown in Table 6. The sig values of the attributes “emotional mobilization” and “plot relevance” are >0.05, which indicates that the evaluation of these two scene attributes is more or less the same at different cultural levels. The significance (sig) of the remaining four attributes and the overall perception were all <0.05, which indicates that the different cultural levels led to many different perceptions in judging the design attributes of the ink digital topiaries. Combined with the comparison of Scheffe, this paper finds that the average number of undergraduates is greater than the average of graduate students, so it can be judged that among the four attributes except "emotional mobilization" and "plot relevance", undergraduates are higher than graduate students in the design attribute evaluation and overall perception evaluation of style transfer design of digital media art. The “other” cultural level is not significantly different from the other two groups. From the perspective of F-score, the cultural level has the greatest influence on the evaluation of digital media art design, and the overall relationship between the size is "visual attraction" (5.717**>), "theme matching" (5.614**), > "atmosphere rendering" (4.125*) > "design novelty" (3.684*).

Table 6. Variation analysis and scheffe comparison

-	-	Sum of squares	df	Mean square	F	significance	Group	Scheffe comparison
Visual attraction	Intergroup	8.563	2	4.288	5.717	0.005***	1.Undergraduate (N=155) 2.Graduate student (N=122) 3. Other (N=31)	1>2
	Within group	189.408	252	0.742	-	-		
	Total amount	197.871	254	-	-	-		
Design novelty	Intergroup	6.611	2	3.314	3.684	0.024*		1>2
	Within group	225.138	252	0.882	-	-		
	Total amount	231.738	254	-	-	-		
Theme matching	Intergroup	9.458	2	4.742	5.614	0.003**		1>2
	Within group	212.852	252	0.856	-	-		
	Total amount	222.318	254	-	-	-		
Ambient rendering	Intergroup	7.46	2	3.754	4.125	0.018*		1>2
	Within group	229.463	252	0.912	-	-		
	Total amount	236.952	254	-	-	-		
Overall	Intergroup	5.181	2	2.585	3.258	0.02*		1>2

design perception	Within group	200.254	252	0.748	-	-		
	Total amount	205.446	254	-	-	-		

On the whole, when the digital media art style migration model constructed in this paper is used as a means to carry out the automated style migration design, the “plot relevance and atmosphere rendering” in the design attributes also determine whether the design of the digital media artwork can bring the audience a good design perception. Meanwhile, the cultural level has an important influence on the cognitive evaluation and overall perception of the design attributes of digital media art works. Undergraduates have higher scores on the cognitive evaluation and overall perception of the design attributes, which also indicates that undergraduates are more in pursuit of the visual impact experience than postgraduates, while postgraduates can evaluate the design works more rationally and present a more conservative attitude towards the design perception of digital media art works. Therefore, the digital topography designed in this paper with traditional Chinese ink painting as the core content should focus more on creating rich and emotional design modifications, and refining the common aesthetic elements for audiences of different cultural levels.

6. Conclusion

Aiming at the problems of CycleGAN, a cyclic consistent generative adversarial network model, this paper proposes to add a multi-head self-attention mechanism between the encoder and decoder as well as an improved CycleGAN model using bilinear interpolation to construct a style migration model for digital media art. Digital media style migration simulation experiments are carried out to test the automated style migration performance of this paper's model. In the quantitative comparison experiments, the fluctuation of the total loss function of this paper's model on the photo2vangogh and photo2monet datasets is too large in the first 100 epochs, and the total loss function tends to be stable from 120 epochs. Meanwhile, on the photo2vangogh and photo2monet datasets, the IS test values of this paper's model reach 5.32 and 6.03, and the FID test values are 97.52 and 75.55, respectively, with the former being higher than all the contrasting models and the latter being lower than all the contrasting models. This indicates that the generated image style of this paper's model is consistent with the target image style, and a better balance can be achieved in terms of both style similarity and content preservation degree. The effectiveness of the bilinear interpolation and attention mechanism module of this paper's model is further examined through ablation experiments. The values of FID, SSIM and PSNR of this paper's model are 153.718, 0.671 and 21.153 on the photo2vangogh dataset, and 149.669, 0.68 and 21.914 on the vangogh2photo dataset, respectively, and the numerical optimization performances of the model on the three indexes of FID, SSIM and PSNR are all outperforms the original CycleGAN model, the intermediate model using only bilinear interpolation or the attention mechanism.

The digital media art style migration model constructed in this paper is used as a design tool to design a digital topography with traditional Chinese ink painting as its content, and it is used as a research subject to explore the correlation between the design attributes of the style migration design work and the audience's perception. In the regression analysis between audience perception and six attributes of the design work, “plot relevance” and “atmosphere rendering”, “design and theme matching” all reached a significant level (sig<0.01). The t-values of “plot relevance” and “atmosphere rendering” are the highest among all independent variables, with 4.375 and 3.38 respectively, indicating that these two design attributes have an important impact on audience perception. After excluding the significance of the audience's gender factor and disciplinary background (sig>0.05), the influence of the audience's cultural level factor on their own perceptions of design attributes was further explored. The design attributes “emotional mobilization” and “plot relevance” had sig values >0.05, while the remaining four attributes and the overall perceived significance (sig) were

<0.05. From the perspective of F-score, the cultural level has the greatest impact on the evaluation of digital media art design is "visual attraction" (5.717**), followed by theme matching (5.614**), atmosphere rendering (4.125*) and "design novelty" (3.684*). The ink painting digital topiary designed in this paper should pay more attention to creating a rich and emotionally expressive design, catering to the common aesthetic elements of the audience.

References

[1] Wang, X., Oxholm, G., Zhang, D., & Wang, Y. F. (2017). Multimodal transfer: A hierarchical deep convolutional neural network for fast artistic style transfer. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5239-5247).

[2] Sheng, J., Song, C., Wang, J., & Han, Y. (2019). Convolutional neural network style transfer towards Chinese paintings. IEEE Access, 7, 163719-163728.

[3] Dinesh Kumar, R., Golden Julie, E., Harold Robinson, Y., Vimal, S., Dhiman, G., & Veerasamy, M. (2022). Deep convolutional nets learning classification for artistic style transfer. Scientific Programming, 2022(1), 2038740.

[4] Cai, Q., Ma, M., Wang, C., & Li, H. (2023). Image neural style transfer: A review. Computers and Electrical Engineering, 108, 108723.

[5] Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2019). Neural style transfer: A review. IEEE transactions on visualization and computer graphics, 26(11), 3365-3385.

[6] Li, J., Wang, Q., Chen, H., An, J., & Li, S. (2020, November). A review on neural style transfer. In Journal of Physics: Conference Series (Vol. 1651, No. 1, p. 012156). IOP Publishing.

[7] Santos, I., Castro, L., Rodriguez-Fernandez, N., Torrente-Patino, A., & Carballal, A. (2021). Artificial neural networks and deep learning in the visual arts: A review. Neural Computing and Applications, 33, 121-157.

[8] Hien, N. L. H., Van Huy, L., & Van Hieu, N. (2021). Artwork style transfer model using deep learning approach. Cybernetics and Physics, 10(3), 127-137.

[9] An, J., Huang, S., Song, Y., Dou, D., Liu, W., & Luo, J. (2021). Artflow: Unbiased image style transfer via reversible neural flows. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 862-871).

[10] Liu, L., Xi, Z., Ji, R., & Ma, W. (2019). Advanced deep learning techniques for image style transfer: A survey. Signal Processing: Image Communication, 78, 465-470.

[11] Bhangale, K. B., Desai, P., Banne, S., & Rajput, U. (2022, May). Neural style transfer: reliving art through artificial intelligence. In 2022 3rd International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.

[12] Shibly, K. H., Rahman, S., Dey, S. K., & Shamim, S. H. (2020). Advanced artistic style transfer using deep neural network. In Cyber Security and Computer Science: Second EAI International Conference, ICONCS 2020, Dhaka, Bangladesh, February 15-16, 2020, Proceedings 2 (pp. 619-628). Springer International Publishing.

[13] Jing, Y., Mao, Y., Yang, Y., Zhan, Y., Song, M., Wang, X., & Tao, D. (2022, October). Learning graph neural networks for image style transfer. In European Conference on Computer Vision (pp. 111-128). Cham: Springer Nature Switzerland.

[14] Khan, A., Ahmad, M., Naqvi, N., Yousafzai, F., & Xiao, J. (2019). Photographic painting style transfer using convolutional neural networks. Multimedia Tools and Applications, 78, 19565-19586.

[15] Gupta, B., Govinda, K., Rajkumar, R., & Masih, J. (2022). Neural artistic style transfer using deep neural

networks. In Proceedings of Academia-Industry Consortium for Data Science: AICDS 2020 (pp. 1-12). Singapore: Springer Nature Singapore.

[16] Kotovenko, D., Sanakoyeu, A., Lang, S., & Ommer, B. (2019). Content and style disentanglement for artistic style transfer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4422-4431).

[17] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2414-2423).

[18] Chen, H., Wang, Z., Zhang, H., Zuo, Z., Li, A., Xing, W., & Lu, D. (2021). Artistic style transfer with internal-external learning and contrastive learning. Advances in Neural Information Processing Systems, 34, 26561-26573.

[19] Semmo, A., Isenberg, T., & Döllner, J. (2017, July). Neural style transfer: A paradigm shift for image-based artistic rendering?. In Proceedings of the symposium on non-photorealistic animation and rendering (pp. 1-13).

[20] Luan, F., Paris, S., Shechtman, E., & Bala, K. (2017). Deep photo style transfer. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4990-4998).

[21] Kumar, T. S., Nadeem, N., Menon, S. R., & Anjali, T. (2024). Convolutional Neural Network-Enhanced Video Artistry: Leveraging VGG16 for Dynamic Style Transfer. Procedia Computer Science, 233, 363-370.

[22] Ruder, M., Dosovitskiy, A., & Brox, T. (2016). Artistic style transfer for videos. In Pattern Recognition: 38th German Conference, GCPR 2016, Hannover, Germany, September 12-15, 2016, Proceedings 38 (pp. 26-36). Springer International Publishing.

[23] Kadish, D., Risi, S., & Løvlie, A. S. (2021, July). Improving object detection in art images using only style transfer. In 2021 international joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.

[24] Chen, H., Zhao, L., Wang, Z., Zhang, H., Zuo, Z., Li, A., ... & Lu, D. (2021). Dualast: Dual style-learning networks for artistic style transfer. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 872-881).

[25] Donghao Zhang,Zhengzheng Wang,Yu Tang,Shengshan Pan & Tianming Pan. (2024). Three-Dimensional Reconstruction of Retaining Structure Defects from Crosshole Ground Penetrating Radar Data Using a Generative Adversarial Network. Remote Sensing(21),3995-3995.

[26] Huipu Xu,Yongzhi Li,Meixiang Zhang & Pengfei Tong. (2024). Sonar image segmentation using a multi-spatial information constraint fuzzy C-means clustering algorithm based on KL divergence. International Journal of Machine Learning and Cybernetics(prepublish),1-18.

[27] Goel Parth & Ganatra Amit. (2023). Unsupervised Domain Adaptation for Image Classification and Object Detection Using Guided Transfer Learning Approach and JS Divergence. Sensors (Basel, Switzerland)(9).

[28] Wang Zilong,Yang Zhe,Song Xiangning,Zhang Hongzhe,Sun Biao,Zhai Jinglei... & Liang Pei. (2023). Raman spectrum model transfer method based on Cycle-GAN. Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy123416-123416.

[29] Zehai Gao,Dongzhe Yang,Baojun Li,Zijun Gao & Chengcheng Li. (2024). Reference Crop Evapotranspiration Prediction Based on Gated Recurrent Unit with Quantum Inspired Multi-head Self-attention Mechanism. Water Resources Management(prepublish),1-21.