

Study on the Innovative Reproduction and Global Dissemination Mode of Non-Heritage Cultural Symbols Driven by Digital Technology

Mengshan Lin¹, Xiangyuan Zeng^{2,*} and Cheng Wang³

¹ School of Art and Design, Fuzhou University of International Studies and Trade, Fuzhou, Fujian, 350202, China

² School of Design, Fujian University of Technology, Fuzhou, Fujian, 350118, China

³ School of Art and Design, Huaiyin Institute of Technology, Huai'an, Jiangsu, 223003, China

Corresponding authors: (e-mail: 19959172134@163.com).

Abstract Under the background of globalization, cultural innovation has become a key factor and frontier field of national power competition, and it has gradually become a strategic choice for all countries in the world to promote their own culture to go out and improve international communication ability. In the article, CycleGAN network is improved to realize the intelligent generation of non-heritage cultural symbols, and the image style migration effect of the improved model is proved by the ablation experiment and the comparison experiment on different datasets. Then the fuzzy set qualitative comparative analysis is applied to analyze the influencing factors and grouping paths of popular videos related to non-heritage cultural symbols spreading in YouTube platform. It is found that there are two grouping paths for the spread of NRH cultural symbols: flexible spread of pan-life content and implicit spread of pan-knowledge content. The global dissemination of NRH cultural symbols can be carried out in terms of relevant departments leading, multi-party participation, digital recording and display, network platform dissemination and digital creative product development.

Index Terms CycleGAN, fuzzy set qualitative comparative analysis, non-heritage cultural symbols, grouping paths

I. Introduction

With the rapid development of global economic integration, contemporary art design, under the perspective of communication science, meets the aesthetic needs of the masses, conforms to the trend of the times, and forms an unprecedentedly prosperous art form [1]. Art symbols of intangible cultural heritage around the regional commercial economic industry and the people's growing spiritual and cultural needs, combined with art and technology in one, through cultural creative products to show the art symbols of intangible cultural heritage characteristics, become one of the important carriers for multicultural communication [2]-[4]. In recent years, in the face of the current active cultural atmosphere, trendy audience thinking and open communication environment, the dissemination of non-heritage culture has also followed the trend of the times, producing a lot of sought-after non-heritage cultural boutiques, and a variety of non-heritage cultural festivals, cultural and creative products, short videos are emerging in an endless stream, which has greatly enriched the market of non-heritage cultural dissemination [5]-[7]. However, the ensuing problem is how to break through the shackles of homogenization and aesthetic fatigue in many cultural excellence, and realize the effective connection between cultural communication and audience, which is a further exploration in the quality-heavy and efficiency-heavy communication following the weight-scattering communication of non-heritage culture [8].

Driven by digital technology, the communication development of non-heritage culture is no longer limited to offline and traditional media channels, and the application of the current strategy is gradually changing to the combination of multiple communication paths, in order to expand the scope of influence and realize the "new life" of non-heritage culture [9]. Yu, L utilized digital database and virtual reality technology in digital technology to realize the living conservation, inheritance innovation and sustainable development of weaving techniques of intangible cultural heritage in Suzhou, China, so as to solve the cognitive group limitation of intangible cultural heritage [10]. He, Y et al. identified three important values of digital technology assisting the promotion of Intangible Cultural Heritage (ICH), namely perceived utilitarian value, perceived hedonic value and perceived symbolic value [11].

The current research on the dissemination and protection of non-heritage culture at home and abroad mainly focuses on the dissemination in the context of the new media era, digital development and the combination of tourism economy. Kim, S et al. argued that ICH has threatened its own authenticity in the current process of commoditization, and that a sustainable tourism approach is needed to achieve the successful dissemination and

promotion of ICH as a sustainable tourism resource [12]. Rodil, K explored the interaction between digital technology and intangible cultural heritage and proposed participatory design as an effective way to address cultural design and build digital artifacts, which not only enhances the artistic taste of the product, but also endows it with a unique cultural appeal [13]. Aljaberi, S. M and Al-Ogaili, A. S fused traditional craft intangible cultural heritage with a physical carrier form of digital communication and found that this fusion was effective in increasing tourist satisfaction and contributing to the promotion and preservation of intangible cultural heritage [14]. Dimitropoulos, K et al. New information technology-based methodology for capturing and analyzing intangible cultural heritage (i-Treasures) to support the safeguarding and dissemination of intangible cultural heritage (ICH) [15].

The article firstly introduces the network model structure of CycleGAN, and then proposes an improvement method based on CycleGAN, which improves the original CycleGAN network model structure by designing the structure of the multi-head self-attention module between the encoder and the converter and by using the bilinear interpolation method of up-sampling. The effectiveness of this paper's method in generating migrated image quality is demonstrated through the presentation of detailed experimental results, ablation experiments and comparison experiments. Subsequently, based on the short videos of non-heritage cultural symbols on YouTube, a fuzzy set qualitative comparative analysis is applied to investigate the causal relationship between the eight conditional variables of short video title information, video duration, account type, video language, narration, character presentation, originality and innovation, and the outcome variable of "dissemination effect", which is the most important factor in the development of non-heritage cultural symbols. The causal relationship between the eight conditional variables and the outcome variable "dissemination effect" is analyzed, and the group path of non-heritage cultural symbols video dissemination is analyzed. Finally, based on the experimental results, a global communication strategy for non-heritage cultural symbols is proposed.

II. A Generative Adversarial Network-based Stylistic Migration Approach for Non-Heritage Cultural Symbols

II. A. Generating Adversarial Networks

II. A. 1) GAN

For a traditional GAN generative adversarial network, it contains a generative network G and a discriminative network D . The function of the generative network is to generate an image $G(z)$, z representing a random noise. The function of the discriminative network is to determine the authenticity of the image. Let the input picture be x , then $D(x)$ represents the probability that the picture is true. When the value of $D(x)$ is 1, then the truthfulness of this picture is 100%, and when the value of $D(x)$ is 0, then it means that this picture must be false. Throughout the process of network training, the generative network G generates as many pictures as possible that can fool the discriminative network D . The discriminative network D judges the real picture and the picture generated by G as much as possible. In this way, the network G and D form a dynamic "game" process, mutual promotion and mutual improvement. Ultimately the most ideal result is that the picture generated by G $G(z)$ basically close to the real picture, D can not identify the real picture, at this time $D(G(z)) = 0.5$. At this time the generative network G that can be used to generate images. Suppose the real training image is x , the distribution of sample images $p_{data}(x)$, and this distribution is unknown. Let a noise z distribution be $p_z(z)$ and $p_z(z)$ is known. Construct the cross-entropy loss function expression as follows:

$$V(D, G) = E_{x \sim p_{data}(x)}[\ln D(x)] + E_{z \sim p_z(z)}[\ln(1 - D(G(z)))] \quad (1)$$

First, some samples $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$ are extracted from the known $p_z(z)$. Then the same number of real images $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ are taken out of the trained data. Let θ_d be the parameter of the discriminator D . The

loss gradient is $\nabla \frac{1}{m} \sum_{i=1}^m [\ln D(x^{(i)}) + (\ln(1 - D(G(z^{(i)})))]$, adding the gradient when updating θ_d . Set the parameters

of the θ_g discriminator G with loss gradient $\nabla \frac{1}{m} \sum_{i=1}^m [\ln(1 - D(G(z^{(i)})))]$, and when updating θ_g Subtract the

gradient. Each update of the parameters in D is immediately followed by an update of G , or as appropriate D can be updated several times before updating G . The whole process can be completed with random noise detection in $p_z(z)$, after which G produces new samples that conform to the $p_{data}(x)$ distribution.

II. A. 2) CycleGAN

Cycle GAN Cyclic Consistent Generative Adversarial Network, CycleGAN does not require paired data, it contains two generators and discriminators to realize the spatial mapping from X to Y . X can be a large number of ordinary photographs, and Y can be a large number of ceramic style pictures, so that the mapping of ordinary photographs

to be converted into ceramic style pictures can be learned [16]. Assuming that F is a mapping from space X to F , i.e., a picture in X is converted to a picture in Y as $F(x)$, its loss during training is basically the same form as the loss of the original GAN. The expression is as follows:

$$L_{GAN}(F, D_Y, X, Y) = E_{y \sim p_{data}(y)} [\ln D_Y(y)] + E_{x \sim p_{data}(x)} [\ln(1 - D_Y(F(x)))] \quad (2)$$

The loss is nullified if only one of the losses is such that the mapping F maps all x to the same picture in Y . The “cyclic consistency loss”, defined as $G(y)$, is introduced in a recurrent generative network. A “cyclic consistency loss” is introduced in a cyclic generative network, which ensures that a picture from X is transformed into Y and the transformed picture is transformed back into X , defined as $G(y)$. CycleGAN satisfies both the F -mapping and the G -mapping and satisfies both $F(G(Y)) \approx y$ and $G(F(x)) \approx x$. Cyclic Losses Define the expression as follows:

$$L_{cyc}(F, G, X, Y) = E_{x \sim p_{data}(x)} [\|G(F(x)) - x\|_1] + E_{y \sim p_{data}(y)} [\|F(G(y)) - y\|_1] \quad (3)$$

The final three components make up the total loss, and the expression is as follows:

$$L = L_{GAN}(F, D_Y, X, Y) + L_{GAN}(G, D_X, X, Y) + \lambda L_{cyc}(F, G, X, Y) \quad (4)$$

This cleverly solves the problem that the training samples do not match each other.

II. B. Image style migration modeling method based on improved CycleGAN

II. B. 1) Multi-pronged 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: the query vector Query, the key vector Key and the 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 summation with the Value vector 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:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (5)$$

The structure of self-attention is shown in Figure 1.

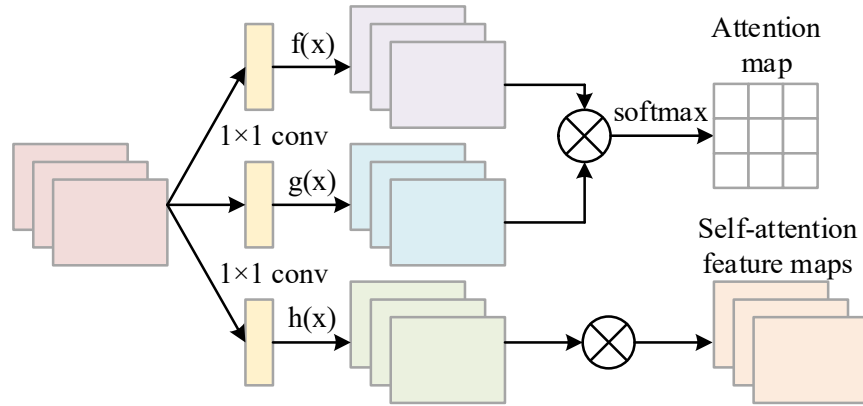


Figure 1: Self-attention structure

The multi-head self-attention mechanism consists of multiple self-attention modules, which are essentially integrations of multiple self-attention computations. 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 process of multi-head self-attention computation is as follows: first, the input sequence $X \in \mathbb{R}^{L \times D}$ is mapped to different query, key, and value vector spaces, and then the query, key, and value vectors are partitioned into H copies according to the number of heads, H , along the last dimension. Next, in each copy of the query, key, and value subspace, the self-attention computation is executed separately to obtain the output of H heads $[\text{head}_{\{1\}}, \dots, \text{head}_{\{H\}}]$. Finally, all head outputs are spliced together as the final multi-head self-attention output vector Z . The multi-head mechanism allows the model to jointly learn different subspace representations of the input and integrate features from different representation subspaces, thus enhancing the expressive power of the model.

Compared with single-head attention, the multi-head attention mechanism can obtain richer semantic information [17].

The specific computational process of the multi-head self-attention layer is:

(1) Obtain Q, K, V vectors: assume the input sequence is $X = [x_1, \dots, x_L] \in R^L \times D$, 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 $q_i \in \square^D$, the key vector $k_i \in \square^D$ and the value vector $v_i \in \square^D$. The linear transformation process for the whole input sequence X can be expressed as:

$$Q = XW^Q \in \square^{L \times D} \tag{6}$$

$$K = XW^K \in \square^{L \times D} \tag{7}$$

$$V = XW^V \in \square^{L \times D} \tag{8}$$

where $W^Q \in \square^{D \times D}$, $W^K \in \square^{D \times D}$, and $W^V \in \square^{D \times D}$ are the learnable linear mapping matrix. Next, the subset of vectors $\{Q_i \in \square^{L \times D_h}, K_i \in \square^{L \times D_h}, V_i \in \square^{L \times D_h} \mid i=1,2,\dots,H\}$, which $D_h = \frac{Dv}{H}$ denotes the vector dimension corresponding to each head. Then, the self-attention calculation 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 jth head $head_j$ as an example, the formula is:

$$head_i = Attention(Q_j, K_j, V_j) \in \square^{L \times D_h} \tag{9}$$

$$Z = MultiHeadAttention \square (head_1 \oplus head_2 \oplus \dots \oplus head_H)W \tag{10}$$

(3) Multi-head result fusion: multiple self-attention results are fused to get the final output vector Z.

Where \oplus denotes splicing along the last dimension of the vector, the spliced vector dimension may not be equal to the dimension D of the original vector, so the matrix W^0 is multiplied here to map the vector to the original input dimension. I.e:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^0 \tag{11}$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{12}$$

The structure of the multinomial self-attention is shown in Figure 2.

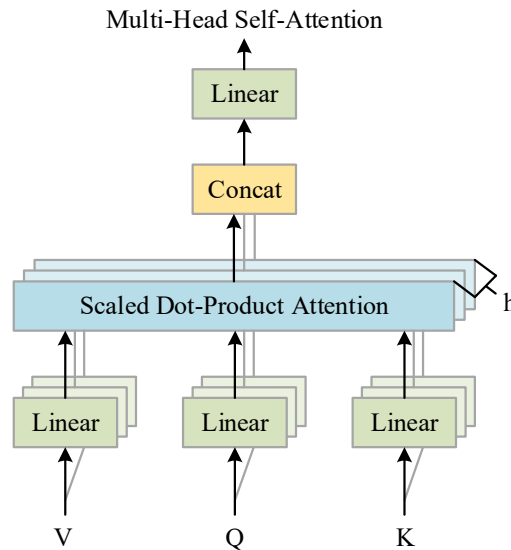


Figure 2: Multi-head self-attention structure

II. B. 2) Bilinear Interpolation

The sampling method of bilinear interpolation is used instead of the inverse convolution, which can provide smoother sampling and avoid the artifact problem of the inverse convolution because the bilinear interpolation is a linear interpolation using the surrounding 4 points, which can make full use of more peripheral pixel information, and the local convolution operation is smoother compared to the inverse convolution. Bilinear interpolation does not produce the problem of mismatch between the convolution kernel and the step size, and can accurately interpolate at any

scaling multiples, thus effectively eliminating the checkerboard effect of the anti-convolution [18]. Bilinear interpolation calculation is small, in order to ensure the effect at the same time improve the efficiency. The specific principle is to set the image to be up-sampled as I , whose size is $H \times W$. Scale it to get the image I' , whose size is $rH \times rW$ (r is the scaling factor). For each pixel (x, y) in I , first calculate its corresponding coordinates (x', y') in the original image I . Then the image I' is up-sampled to $rH \times rW$ (r is the scaling factor):

$$x' = \frac{x}{r} \quad (13)$$

$$y' = \frac{y}{r} \quad (14)$$

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

$$I'(x, y) = (1 - \lambda x)(1 - \lambda y)I(x'', y'') + \lambda x(1 - \lambda y)I(x'' + 1, y'') + (1 - \lambda x)\lambda yI(x'', y'' + 1) + \lambda x\lambda yI(x'' + 1, y'' + 1) \quad (15)$$

where $\lambda x = x' - x''$, $\lambda y = y' - y''$, x'' is the coordinate obtained by rounding down the x' is the coordinate obtained by rounding down and y'' is the coordinate obtained by rounding down y' [19].

II. B. 3) Improving the network structure of the model

In this section, an improved CycleGAN modeling approach is proposed for the image style migration task. 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 while maintaining the processivity of the original network.

In the overall network structure, the input images include X and Y domains, and a generator is G

$X \rightarrow Y$, which maps the image in the source domain X to generate the corresponding image in the target domain Y . Another generator is $F: Y \rightarrow X$, which maps the image in the target domain Y to generate the corresponding image in the source domain X in the opposite direction. The specific improved generator structure will be elaborated below. The discriminator D_X serves to distinguish between the real image in the source domain X and the fake image generated by the generator $F(G(x))$, while the discriminator D_Y needs to distinguish between the real image in the target domain Y and the fake image generated by the generator $G(F(y))$. Adversarial loss in generator G to discriminator D_Y and generator F to discriminator D_X . The cyclic consistency loss is used to compute the difference between an X -domain (source domain) image that has been transformed through the Y -domain (target domain) and then to the X -domain (source domain), and the original source domain image. The specific network structure flow is as follows: two domains X and Y are fixed training samples $\{x_i\}_{i=1}^N (x_i \in X)$ and $\{y_j\}_j = 1^N (y_j \in Y)$ respectively, and the data distributions are $x \sim p_{data}(x)$ and $y \sim p_{data}(y)$. When an image is converted from the X domain to the Y domain, given an image x (NRM style image) from the X domain and a style image y (NRM style image) from the Y domain as input, the generator G converts the NRM image from the X domain into the output image $G(x)$ of the NRM style from the Y domain, and the generator F converts the image $G(x)$ into the original image x , i.e., $F(G(x)) \approx x$. D_Y is used to distinguish the difference between the generated image $G(x)$ and the NRM style image y , and the adversarial loss is computed. Conversely, the same is true when the image is transformed from the Y -domain to the X -domain, i.e., $y \approx G(F(y))$, and D_X is used to distinguish the difference between the generated image $F(y)$ and the non-legacy cultural style image x .

III. Model performance testing and analysis of results

III. A. Experimental environment

The experimental environment used in this section is: using an Intel Core I7-12700 machine with 8 cores at 4.9 GHZ, 128G of host memory, and a GeForce RTX 3050 graphics card. The operating system is Ubuntu, and the deep learning framework is PyTorch 1.7.

III. B. Experimental data set

(1) Photo-paper-cut dataset

We used a one-to-one dataset of face photo images and paper cutout images for model training, which consists of 900 pairs of face photos and paper cutouts created by paper cutout artists based on the photos (in this paper, these paper cutout images are referred to as artist cutouts). Among them, the artist cutouts are all fully frontal, but the face photographs do not guarantee that the person's line of sight is flat, and may be taken obliquely sideways. We preprocessed the photographs and paper cutout images to form the face-paper cutout dataset, and then, randomly divided the dataset into 847 training data pairs and 53 test data pairs.

(2) CelebA dataset

In order to increase the model's focus on features, we design the fixed encoder, which is obtained from the pre-training of the classification task. The pre-training uses labeled face photos labeled with gender and whether or not they are wearing glasses. We filter them from the dataset CelebA. For each photo, the percentage of its left and right faces in the photo is calculated using feature point detection, and if the difference between the percentages exceeds a specified value, it is considered as a photo in which the complete features are not visible and is removed. After screening, the photos are preprocessed to adjust the face size and position. A total of 2174 of these images, together with the face photos in the face-paper cutout dataset, are divided into 8:2 ratios and used as the training set and test set for the classification model.

(3) Paper-cut dataset

The original image of paper cutouts is 787×1181 pixels. We randomly crop 512×512 pixel paper cut images from 1000 original images, from which we select images with complex paper cut edges to ensure that the data are representative. The cropped images were then subjected to a random rotation and flipping process. This step aims to introduce variations in various edge orientations, thus allowing the model to be exposed to and learn more diverse edge situations during the training process. This enhancement strategy helps to improve the model's ability to recognize and process complex edge features in paper-cut images.

III. C. Image quality evaluation index

In the field of digital image processing and computer vision, image quality assessment plays a crucial role. With the continuous advancement of image analysis techniques and the expansion of applications, objective and accurate assessment of image quality becomes particularly important. It is not only about the optimization of visual experience, but also a key aspect of performance evaluation for a variety of image processing tasks, such as image compression, enhancement, reconstruction and generation. In this context, a variety of image quality assessment metrics have been developed with the aim of measuring image quality from different perspectives and levels. The metrics used in this paper include NIMA, FID, KID, LPIPS, SSIM and PSNR.

III. D. Paper-cut image generation experiment

We adopt a network architecture and use SGD and Adam optimizers for the generator and discriminator, respectively. The learning rate starts at 0.0001 and decreases linearly to 0 after 200 epochs. The average speed of generating a paper cutout image is 0.05 seconds on this GPU.

III. D. 1) Baseline experiments and analysis of results

In this section we show the results of several image translation models for the paper cutting task and compare them qualitatively and quantitatively. The baseline models are Pix2pix, CycleGAN, U-GAT-IT, and MUNIT. We use NIMA to score the paper cut images generated by the models in this paper and the other four models to assess the visual quality of the images. At the same time, handmade paper cut images as labels (Ground truth) are also involved in the scoring. The NIMA scores of the paper cut images generated by different methods are shown in Fig. 3. The average scores in descending order are Ground truth, this paper's model, U-GAT-IT, Pix2pix, MUNIT, and CycleGAN, with scores of 5.19, 5.03, 4.87, 4.64, 4.5, and 4.32, respectively. This paper's model significantly outperforms the other four compared methods and has the smallest gap in quality when compared to handmade paper cut images. This is mainly due to the fact that the model in this paper focuses on the task of generating paper cutout images and employs migration learning, a two-branch generator, and an average face loss function to enhance the consistency between the generated paper cutout images and human perceptual preferences. In addition, the U-

GAT-IT over-attention module achieves unsupervised image-to-image translation and shows some adaptability to paper-cut images with changing shapes. As a result, UGAT-IT performs second best in terms of performance, after the model in this paper. On the other hand, the model in this paper lacks an explicit objective function to constrain the generation of paper-cut styles, leading to the potential generation of unreasonable details.

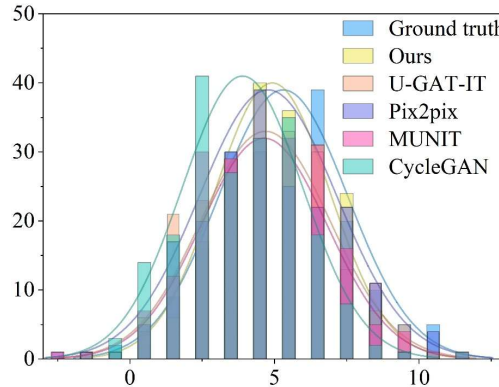


Figure 3: Nima scores for paper-cut images generated by different methods

The quality of the content of each generated cutout was assessed by comparing it to its corresponding hand cutout. The values of the five metrics for the five models in the test set are shown in Table 1. It is clear from this that our models achieved the best results for the five metrics. The five algorithms show a strong correlation with the NIMA scores when evaluated using the FID metrics. The FID values for CycleGAN are significantly higher than the other four algorithms. This is mainly due to the fact that the feature vectors of the paper cutout images generated by CycleGAN are very different from those of the real images, resulting in higher FID values. Compared to the other four algorithms, the KID value of MUNIT is much higher, indicating that there is a large gap between the image generated by MUNIT and the real image. The results generated by Pix2pix are second only to CutGAN in terms of FID, SSIM, PSNR and LPIPS metrics, indicating that the image generated by Pix2pix is more similar to the artist's paper cutout in overall terms.

Table 1: Test the five measurements of the five models

Model	FID ↓	KID ↓	SSIM ↑	PSNR ↑	LPIPS ↓
Pix2pix	54.0804	0.0285±0.0165	0.6882±0.0025	11.6847±1.8726	0.2737±0.0038
CycleGAN	106.0831	0.0246±0.0143	0.6009±0.0023	11.2949±1.4153	0.2948±0.0056
U-GAT-IT	60.7261	0.029±0.0103	0.6799±0.0018	11.479±2.0656	0.2941±0.0065
MUNIT	82.0306	0.0704±0.0181	0.6461±0.0016	11.3083±1.1534	0.2937±0.0042
Ours	50.7911	0.0208±0.0146	0.7031±0.0021	12.2032±1.9277	0.2388±0.0025

III. D. 2) Ablation experiments and analysis of results

In order to gain a deeper understanding of the role of each module in the model of this paper, we designed five ablation experiments. These experiments are designed to gradually strip away key elements in order to evaluate the contribution of each part to the overall performance. The five evaluation metrics for the six models on the test set are shown in Table 2. As can be seen from the table, our models achieve the best results in terms of the five metrics. CycleGAN1 and CycleGAN4 have significantly higher FID and KID counts than CutGAN, indicating that the use of a two-branch generator and the removal of jump connections in the U-Net structure have a significant impact on the quality of the generated paper cutout images. CycleGAN2 only utilizes the paper cutout dataset's 822 face images for pre-training, and CycleGAN merged an additional 2,174 images from the ClebeA dataset for pre-training. CycleGAN2 demonstrated comparable values for the five metrics as CycleGAN, which suggests that the number of images used in the pre-training has a relatively small effect on the quality of the generated paper cutout images. Although CycleGAN3 has lower values than CycleGAN for these five metrics, the difference between them is not significant. This suggests that the average facial loss has a slight positive effect on model performance. Except for the FID metric, CycleGAN5 performs very similarly to CycleGAN on the other four metrics, which suggests that the use of IN can slightly improve model performance.

Table 2 Five evaluation indicators of 6 models in the test set

Model	FID ↓	KID ↓	SSIM ↑	PSNR ↑	LPIPS ↓
CycleGAN0	74.0432	0.0433±0.0228	0.6943±0.0013	12.1959±1.3892	0.2419±0.003
CycleGAN1	54.8931	0.0193±0.0131	0.7006±0.0016	12.0107±1.3601	0.2405±0.0032
CycleGAN2	53.1438	0.0268±0.0161	0.6877±0.0014	12.1205±1.542	0.2489±0.0028
CycleGAN3	56.3372	0.0233±0.017	0.693±0.0013	12.1143±1.5095	0.2467±0.0033
CycleGAN4	71.0155	0.0393±0.0207	0.6923±0.0011	12.0658±1.3176	0.2611±0.0028
CycleGAN5	62.2073	0.0271±0.0196	0.7003±0.0012	12.1238±1.3474	0.2487±0.0024
CycleGAN	51.6569	0.0171±0.0133	0.7021±0.0017	12.1988±1.9369	0.2412±0.0033

III. E. Paper-cut image magnification experiment

In order to quantitatively assess and compare the amplification effects of these methods, we used SSIM, PSNR, and NIMA as evaluation metrics. The SSIM, PSNR and NIMA of different methods are shown in Table 3. In terms of reduction, SSIM and PSNR with bicubic interpolation are better than nearest neighbor interpolation, bilinear interpolation, EDSR, and slightly worse than the method in this paper, which is due to the fact that it can smooth the image better and reduce jaggedness. In terms of visual effect, nearest neighbor interpolation jaggedness is obvious, and bilinear interpolation and bicubic interpolation smooth the image causing edge blurring, which affects its aesthetics, and therefore scores lower. Compared with other methods, the SSIM, PSNR and NIMA of this paper's model are 0.9449, 31.3665 and 4.9247, respectively, which are higher than other methods. In summary, this paper's method performs better in terms of detail retention and visual quality.

Table 3: SSIM, PSNR and NIMA of different methods

Method	SSIM ↑	PSNR ↑	NIMA ↑
EDSR	0.911	28.6573	4.2875
Recent adjacent interpolation	0.9251	26.4425	4.212
Two-wire interpolation	0.922	28.0378	4.2474
Double interpolation	0.942	29.5482	4.2151
Ours	0.9449	31.3665	4.9247

IV. Factors affecting the enhancement of the effectiveness of the dissemination of cultural symbols of NRHs

IV. A. Research methodology and design

IV. A. 1) Research methodology

This section explores the influencing factors on the dissemination of non-heritage culture on YouTube platform, and the specific reasons and methods are as follows:

(1) Considering the limitations of the number of variables in the qualitative comparative analysis and the fact that too many conditions may lead to the difficulty of fully explaining the group results, regression analysis is applied in advance to clarify which factors have a more significant correlation with the effect of the dissemination of non-heritage culture.

(2) The factors affecting the dissemination effect are not independent of each other, but are the result of the joint action of many factors. Qualitative comparative analysis can generate multiple causal relationships and multiple logical condition combinations according to the multiple factors and results of the cases, and explore the group effect that affects the dissemination effect of non-heritage culture.

IV. A. 2) Case Selection

The videos related to the dissemination of non-heritage culture published on the YouTube platform were selected for the study. Specific sample acquisition: (1) Time range setting. In order to maximize the acquisition of videos related to the dissemination of non-heritage culture, the time range of video acquisition was not limited. (2) Video search and screening. Secondly, we filtered the videos with high popularity and representativeness. In social media platforms, videos with more than 1 million plays are usually considered as high popularity videos. Therefore, using this as a criterion, we further select the non-heritage cultural communication videos with high broadcasting volume. In the end, 100 popular videos related to the dissemination of non-heritage culture were obtained, which were used as research cases.

IV. A. 3) Variable Measurements and Results

(1) Conditional variable measurement. The number of plays, the number of likes and the number of comments on the videos of non-heritage cultural symbols were assigned weight ratios of 0.5, 0.3 and 0.2, respectively, in order to calculate and measure the overall communication effect.

(2) Measurement of outcome variables. According to the propagation characteristics of non-heritage cultural symbol videos on the YouTube platform and the measurement basis of the corresponding variables in the existing studies, the selection and coding of the conditional variables are shown in Table 4.

Table 4: Conditional variable selection and coding

Dimension	Variable	Measurement and coding instructions
Information source	Publisher identity	Media, organization and other (1) personal (1)
	Account activity	Monthly update number (continuous variable)
	Fan control	Person-time (continuous variable)
Propagation content	Title information	Statement (0) An exclamation sentence, a question (1)
	Video duration	Minute (continuous variable)
	Video form	Horizontal screen (0) Vertical (1)
	Video language	Chinese is the main (0) Foreign language is dominated (1)
	Interpretation mode	In the objective narrative (0) The main objective is the description (1)
	Character rendering	Based on Chinese people (0) Based on Chinese foreigners (1)
	Originality	Secondary works (0) Original works (1)
	Innovation	Similar themes and perspectives are more than (0) New and unique, unique (1)
acceptor	Affective expression	Emotional value (continuous variable)
	Mainstream media participation	No international mainstream media interaction (0) Interact with the international mainstream media (1)
	Opinion leadership	No means of meeting a leader (representative) Interested in meeting the leader at (1)

(1) On the information source dimension measurement setting. A total of three variables are involved: publisher identity, account activity, and fan base. The identity of publishers is divided into two categories: individuals, media and organizations. The number of monthly updates of the narrator is counted by combining the Weibo index related to the measurement of social media platform activity. The number of followers of the narrator is the number of followers of the narrator in YouTube.

(2) Measurement settings of communication content dimensions. Nine variables are involved, including title information, video duration, video format, and video language. The title information involves three types of sentences: declarative, exclamatory and interrogative, and most of the video samples are dominated by declarative titles, so they are categorized into declarative and non-declarative sentences. The video duration is the length of time the video is played. In terms of video format, the videos of non-heritage cultural symbols involve two types of horizontal and vertical screens, so they are categorized accordingly. In terms of video language, NLCS videos are generally presented in Chinese and English characters or in English, so they are adjusted to Chinese and English on the basis of the classification criteria of pure Chinese, Chinese with English subtitles, and foreign languages in the previous studies. In terms of narration mode, the existing studies set three criteria: objective presentation, subjective evaluation and strong recommendation, while the video communication of non-heritage cultural symbols is mainly based on objective analysis or subjective-objective combination of narrative mode, so this is used as the basis for measurement. In terms of character presentation, the classification criteria are mainly Chinese and foreigners. Originality is mainly based on whether the video is original or not. Innovativeness was measured by the novelty of the subject matter and the unique perspective of the video. In terms of emotion expression, the Pysentiment library, which is commonly used in English emotion calculation, was used to calculate the emotion of the video text, and the Harvard IV-4 English General Emotion Library, which is a classic in the psychosocial field, was used to analyze the video text to get the overall emotion expression value of the video.

(3) About the receiver dimension measurement setting. Two variables are involved: mainstream media and opinion leader participation. The mainstream media factor is based on whether the video has been reported by other mainstream international media, or whether the video has generated interactive behaviors such as retweeting or commenting in international social media platforms such as YouTube. The opinion leader factor is whether the video distribution interacts with influential users, in which the opinion leaders are identified by the user's professional

authentication, the number of followers greater than 10,000, and the number of video comments and likes exceeding 1,000 items.

IV. B. Analysis of results

IV. B. 1) Univariate necessity analysis

Before analyzing the group effect of the symbolic communication effect of non-heritage culture, it is first necessary to determine whether a single conditional variable is a necessary condition for the outcome variable. Generally speaking, when the consistency level of a single variable reaches 0.9 and above, it indicates that the conditional variable is necessary for the outcome variable. In this study, after importing the calibrated variable data into the fsQCA software, 18 data results were obtained by conducting the necessity condition test on the condition variables, and the results of the one-way necessity analysis are shown in Table 5 (“~” means “not”, i.e., the condition variables do not exist). As can be seen from the data in the table, the highest level of consistency of all single conditional variables was 0.85259, which did not reach the determination standard of 0.9, indicating that there was no significant causal relationship between the conditional variables and the outcome variables, and group effect analysis was needed.

Table 5: Analysis of single factor necessity analysis

Conditional variable	Consistency	Coverage
Title information	0.85259	0.53545
~ Title information	0.42081	0.6383
Video duration	0.84767	0.47225
~ Video duration	0.41226	0.7219
Video form	0.67462	0.56629
~ Video form	0.70712	0.6127
Video language	0.6641	0.58825
~ Video language	0.61196	0.56937
Interpretation mode	0.58609	0.63257
~ Interpretation mode	0.75702	0.54233
Character rendering	0.8758	0.47774
~ Character rendering	0.41129	0.63553
Originality	0.88312	0.44937
~ Originality	0.36102	0.71798
Innovation	0.69259	0.61494
~ Innovation	0.76508	0.56845
Affective expression	0.70433	0.54032
~ Affective expression	0.63342	0.55234

IV. B. 2) Analysis of configuration effects

The purpose of the group effect analysis is to determine the extent to which the combination between different condition variables explains the outcome variable. Based on the size of the selected sample cases, this study sets the threshold for the frequency of cases in fsQCA to 1 and the consistency threshold to 0.8, and calculates three types of solutions, including the complex solution, the intermediate solution, and the parsimonious solution. The intermediate solution was chosen for further analysis in this study because it circumvents the overly detailed causal delineation in the complex solution and the lack of the parsimonious solution in the expression of the group path mechanism. Based on the presentation form of the QCA analysis results, the conditions that exist in both the intermediate and parsimonious solutions are regarded as core conditions, while the conditions that appear only in the intermediate solution are regarded as edge conditions. The results of the grouped path analysis of propagation effects are shown in Table 6 (● represents the presence of the core condition, ⊗ represents the absence of the core condition, ● represents the presence of the edge condition, ⊗ represents the absence of the edge condition, and blank means that the presence or absence of this conditional variable has no relationship with the outcome variable.)

Table 6: The results of the propagation effect configuration path analysis

Conditional variable	Configuration path		
	1	2	3
Title information	●	●	⊙
Video duration			
Video form			●
Video language			
Interpretation mode		⊙	⊙
Character rendering			
Originality			
Innovation	●	●	⊙
Affective expression	⊙		⊙
Original coverage	0.27556	0.2374	0.28117
Unique coverage	0.04599	0.02328	0.15063
consistency	0.86608	0.85648	0.82768
The coverage of the solution	0.41034		
Consistency of solutions	0.84576		

As shown in the table, the consistency level of the overall solution and the individual solutions of each grouping is higher than 0.8, which meets the requirement of QCA that the consistency of all combinations is greater than 0.75, indicating that the three grouping paths all have strong explanatory power for the non-heritage cultural symbols video to obtain good dissemination effect. In addition, the coverage of the three paths is 0.41034, indicating that the result can explain about 41% of the sample cases, and the model effect is relatively good.

IV. B. 3) Robustness Tests

In order to avoid the chance and sensitivity of the histogram analysis results, this study chose to increase the consistency threshold as well as change the case frequency threshold to test the robustness of the results. First, the consistency threshold was set to 0.85, and the consistency level of the overall solution increased from the original 0.84576 to 0.89676, and there was no substantial change in the histogram paths, which indicates that the original histogram analysis results are somewhat robust. Second, adjusting the case frequency threshold to 2 produces the results of histogram 1 and histogram 2 in the original results, and although histogram 3 in the original results is eliminated as a logical remainder because of the threshold reduction, the histogram solution of the test can be used as a subset of the original results, and thus the original results can also be considered robust.

IV. B. 4) Conclusion of the analysis of configuration effects

Based on the analysis of the homogeneous and heterogeneous relationship between the results of the existing group effect, this study further summarizes the paths of Chinese culture short videos to achieve good international communication effects, and the typical path characteristics of the international communication of short videos of NRL cultural symbols are shown in Table 7. It can be found that pan-life Chinese culture short videos emphasize rich narratives and adopt low-context expressions, and enhance the communication effect of short videos by tapping into the “common currency” and communicating in a close to the audience, easy and interesting narrative way. Pan-knowledge Chinese culture videos emphasize simple narratives and convey high-context culture to the audience in a subtle way, so that they can be infected and inspired by it, thus enhancing the communication effect of the short videos.

Table 7: The typical path feature of non-relic symbol myopia is the typical path feature

Configuration type	Corresponding configuration	Configuration characteristics	Configuration connotation
The content of the pan-living content is spread	Configuration 1 Configuration 2	Rich narrative and tibetan culture	Dig "the common money to spread in a way that is close to the audience and easily and interesting
The knowledge content is recessive	Configuration 3	Narrative simplicity and high context culture	The high context culture is transmitted to the audience in an obscure way, allowing it to be infected and inspired

V. Strategies for the global dissemination of cultural symbols of NRHs

V. A. *Relevant sector-led and multi-party participation*

The State attaches great importance to the digitization and dissemination of cultural symbols of non-heritage, and the Ministry of Culture and Tourism and other departments have been actively promoting it, formulating guidelines and regulations, and guiding and supporting cultural units at all levels to carry out relevant projects. Relevant local departments have taken into account local characteristics, formulated plans and promoted the digitization of local non-heritage. Colleges and universities, research institutions, enterprises, NGOs and other parties actively participate, contribute wisdom and resources, and cooperate to jointly promote the flourishing development of digital dissemination of non-traditional cultural heritage.

V. B. *Digital Recording and Presentation*

Comprehensive and systematic digital records are the basis for the digital dissemination of non-heritage cultural symbols. Various regions have adopted digital technologies, such as text, pictures, audio and video, to record and archive non-heritage in a comprehensive manner and to establish databases and digital archives. A large number of valuable non-legacy materials have been digitized, providing a wealth of first-hand materials for dissemination and research. Some regions have also set up digital exhibition centers and online pavilions, using virtual reality, holographic projection and other technologies to vividly reproduce non-legacy items, giving audiences an immersive experience and opening up new spaces for dissemination.

V. C. *Dissemination through online platforms*

The Internet has created a broad platform for the dissemination of non-heritage cultural symbols. A number of websites and databases on intangible cultural heritage have been set up, such as the China Intangible Cultural Heritage Website, which centrally displays a huge amount of resources for the public to search and browse. Specialized websites have been set up for key projects, showing the historical status and inheritors, and providing online experience and interaction. Portal websites and cultural websites have opened special channels for intangible cultural heritage, and have increased dissemination efforts through news reports, special planning, and live video broadcasting. Microblogging, WeChat, short videos and other social platforms have also become important positions for the dissemination of non-legacy, with cultural institutions and inheritors using the platforms to release developments, interact with the public and expand the scope of dissemination.

V. D. *Digital Creative Product Development*

With the integration of digital technology with the cultural and creative industries, digital creative products of non-heritage cultural symbols have been emerging. Some non-legacy elements have been refined and transformed into digital visual symbols and creative content, which are widely used in animation, games, film and television, design and other fields. For example, animation images inspired by Dunhuang murals and Peking Opera faces have revitalized the ancient art. Handheld games based on non-heritage skills allow players to learn about non-heritage in entertainment. Applying non-heritage patterns to clothing and home design to develop fashion products with great cultural flavor. These cross-border integration of creative products have enhanced the contemporary expression of non-heritage, expanded the space of the communication market and attracted the attention of young consumers. Some non-heritage cultural symbols and digital creative products have gone abroad to participate in international cultural exhibitions, becoming a bright business card of Chinese culture to the world.

VI. Conclusion

In this paper, we propose an image style migration model to improve the Cycle Generative Adversarial Network, and experimentally prove that improving the network structure of CycleGAN achieves better style generation of non-heritage cultural symbols. It also analyzes the grouping path of the video dissemination of non-heritage cultural symbols. The article draws the following conclusions:

In the image magnification experiments, the SSIM, PSNR and NIMA values of this paper's model are higher than those of other comparative methods, which are 0.9449, 31.3665 and 4.9247, respectively. From this, we can conclude that this paper's method performs better in terms of detail retention and visual quality.

By analyzing the grouping path of the video communication of non-heritage cultural symbols, there are two grouping paths: flexible communication of pan-life content and implicit communication of pan-knowledge content.

Therefore, it is necessary to use digital technology, innovate communication and promotion methods, and strengthen international exchange and cooperation, so that the cultural symbols of non-heritage can be renewed and energized in the inheritance and innovation.

Funding

This work was supported by Research Project of Fujian University of Technology (GY-Z2/3036).

References

- [1] Vecco, M. (2010). A definition of cultural heritage: From the tangible to the intangible. *Journal of cultural heritage*, 11(3), 321-324.
- [2] Ionesov, V. I., & Kurulenko, E. A. (2015). Things as characters of culture: symbolic nature and meanings of material objects in changing world. *Scientific Culture*, 1(3), 47-50.
- [3] Xiong, Z. (2024). Research on the Application of Cultural Symbols and Local Elements in Graphic Design. *Frontiers in art research*, 6(10), 51-57.
- [4] Abdullah, A. H., & Abdillah, N. (2021). Heritage value of the malayness socio-cultural symbols in millennium artist series of arts: A research analysis in semiotics. *Journal of Educational and Social Research*.
- [5] Wang, Q. (2022). The digitisation of intangible cultural heritage oriented to inheritance and dissemination under the threshold of neural network vision. *Mobile Information Systems*, 2022(1), 6323811.
- [6] Liu, Z. (2025). The construction of a digital dissemination platform for the intangible cultural heritage using convolutional neural network models. *Heliyon*, 11(1).
- [7] YAO, G. Z. (2023). The Challenges and Solutions for the Digital Dissemination of Intangible Cultural Heritage in China. *Journal of Xihua University (Philosophy & Social Sciences)*, 42(4), 75-82.
- [8] Yuan, L., & Briel, H. (2016). The Dilemma of the Revitalization of Intangible Cultural Heritage and Global Homogenization: The Case of Techno Nezha in Taiwan. *IAFOR Journal of Cultural Studies*, 1(2), 53-70.
- [9] Lavikka, R., Kallio, J., Casey, T., & Airaksinen, M. (2018). Digital disruption of the AEC industry: Technology-oriented scenarios for possible future development paths. *Construction management and economics*, 36(11), 635-650.
- [10] Yu, L. (2023). Digital Sustainability of Intangible Cultural Heritage: The Example of the "Wu Leno" Weaving Technique in Suzhou, China. *Sustainability*, 15(12), 9803.
- [11] He, Y., Chen, X., & Wang, L. (2023). How Digital Events Promote Intangible Cultural Heritage? A User Experience Perspective. *Proceedings of the Association for Information Science and Technology*, 60(1), 974-976.
- [12] Kim, S., Whitford, M., & Arcodia, C. (2021). Development of intangible cultural heritage as a sustainable tourism resource: The intangible cultural heritage practitioners' perspectives. In *Authenticity and authentication of heritage* (pp. 34-47). Routledge.
- [13] Rodil, K. (2017). A perspective on systems design in the digitisation of intangible cultural heritage. *International Journal of Intangible Heritage*, 12, 190-198.
- [14] Aljaberi, S. M., & Al-Ogaili, A. S. (2021). Integration of cultural digital form and material carrier form of traditional handicraft intangible cultural heritage. *J Fusion Pract Appl*, 5(1), 21-30.
- [15] Dimitropoulos, K., Tsalakanidou, F., Nikolopoulos, S., Kompatsiaris, I., Grammalidis, N., Manitsaris, S., ... & Manitsaris, A. (2018). A multimodal approach for the safeguarding and transmission of intangible cultural heritage: The case of i-Treasures. *IEEE intelligent systems*, 33(6), 3-16.
- [16] Xiaotong Luo, Wenjin Yang, Yuan Xie & Yanyun Qu. (2025). Farewell to CycleGAN: Single GAN with decoupled constraint for unpaired image dehazing. *Neurocomputing*, 636, 129888-129888.
- [17] Mingyue Lu, Yi Xu, Wenke Chu, Jiaye Gu & Muyang Gao. (2025). Contrastive multi-view clustering based on multi-head attention mechanisms and three-way decision. *Knowledge and Information Systems*, (prepublish), 1-24.
- [18] Xiao Chen, Chengzhuo Xu, Minglin Zhang, Xi Li & Zhihao Liu. (2025). A bilinear interpolation scheme for polar coordinate quantum images. *Chinese Journal of Physics*, 95, 493-507.
- [19] Madhusmita Priyadarshini Sahoo & Rajeswari Sridhar. (2025). Bilinear Interpolation Augmented Deep Feature Extraction with an Improved Remora Optimization-Based Deep Convolutional Neural Network for Skin Lesion Classification. *International Journal of Image and Graphics*, (prepublish).