Research Article

Research on the Creation of Chinese National Cultural Identity Symbols Based on Visual Images

LianShuQing Yang¹ and Jiahui Li²

¹Hebei Normal University for Nationalities, Chengde 067000, China
²School of Art and Design, Yanshan University, Qinhuangdao 066000, China

Correspondence should be addressed to Jiahui Li; ljhyw@stumail.ysu.edu.cn

Received 25 July 2022; Revised 23 August 2022; Accepted 29 August 2022; Published 13 September 2022

Academic Editor: Lianhui Li

Copyright © 2022 LianShuQing Yang and Jiahui Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cultural symbol generation has always been a challenging task to achieve symbols that can represent national culture and promote people’s identification with Chinese culture. In this paper, we combine generative adversarial network (GAN) to propose a symbolic generation model of Chinese national cultural identity based on visual images. First, combining pattern search regular terms with generator cross-loss functions based on GAN generative adversarial networks to improve the pattern collapse phenomenon of generative adversarial networks. Second, the normal convolutional layer of the generator in the network structure is replaced with a deep-space separable convolution to improve the real-time performance of the model by reducing the model parameters. Through extensive testing on real datasets, the results show that the model in this paper can generate higher performance ethnic culture symbols while maintaining better temporal performance, which has some practical application value.

1. Introduction

The identification of Chinese cultural symbols and the image of the Chinese nation is an important element in strengthening the theoretical connotation, narrative system, and discourse system of Chinese national community consciousness. Promote the heritage protection and innovative intermingling of the cultures of all ethnic groups, establish and highlight the Chinese cultural symbols shared by all ethnic groups and the image of the Chinese nation, and enhance the identification of all ethnic groups with Chinese culture [1]. Establishing and highlighting the Chinese cultural symbols shared by all ethnic groups and the image of the Chinese nation is of great significance in realizing the Chinese dream of the great rejuvenation of the Chinese nation, casting a firm sense of Chinese national community and enhancing the international image of China [2].

Chinese cultural symbols are those symbols that can reflect the characteristics of Chinese culture and the characteristics of different ethnic groups in different regions, which embody the cultural connotation and spirituality of China’s long and profound history. Chinese cultural symbols can generally be divided into two main categories, one being natural symbols, such as mountains and rivers, and stars and constellations; the other category is man-made symbols, such as architecture and clothing. And there are numerous classifications of theories of their origin, such as the totemic theory, in which the predecessors of symbols were images created by ancient people for the natural objects they worshiped; functional statement originated from practical utility; decoration originated from the role of beautification, etc. In addition, Chinese national cultural symbols contain cultural connotations such as Confucianism, political and ethical culture, and ethnic culture [3]. Chinese cultural symbols such as Confucius culture and chess culture are an inseparable part of Chinese national culture and have rich cultural symbol resources.

Nowadays, the study of cultural symbols and visual images is highly valued, but the understanding of the cultural symbols and visual images of the Chinese community should be based on theoretical sorting and discernment, and we should emphasize the dynamic and process nature, the
interconnection and transformation between figurative and conceptual, and the coherence between intuitive visual images and abstract visual ideas [4]. Visual image research should use the human body as a medium, focus on the practitioner’s active choice, and examine the relationship between the symbolic image and the intentional mental image. It emphasizes the multi-layered nature of cultural symbols and visual images in a holistic view of the Chinese national community and seeks to construct a mutually inclusive identity. Visual images should be recognized and understood within the multimodality of multisensory interventions, avoiding falling into the trap of visual centrism. In design practice, attention should be paid to discovering and grasping cultural symbols and visual image resources, respecting cultural knowledge and concepts in applied design, expanding participation and sharing, and implementing the theoretical understanding of casting a firm sense of Chinese national community as the main line [5].

Both in terms of national development evolution and national cultural psychology, the Chinese nation as a self-conscious national entity emerged from the confrontation between China and the Western powers in the last hundred years, but being a self-contained national entity is the result of thousands of years of the historical process. The pluralistic pattern of the Chinese nation is a product of history, but also a tangible existence. Chinese national community consciousness is expressed concretely as a state of unanimity, conceptual fit, and unity of will among members of the nation around important issues such as the historical lineage, survival intention, and direction of development of the Chinese nation. In the long process of development of Chinese civilization, Chinese people of all ethnic groups have contributed to the formation and development of Chinese culture, forming national images and cultural symbols with unique Chinese characteristics, these images and symbols are the concentrated embodiment of the living habits, spirituality, and values of the Chinese people of all ethnic groups, which have a universal cultural and political identity [6, 7].

Casting a firm sense of Chinese national community first comes from identifying with Chinese cultural symbols and the image of the Chinese nation in daily life. The image of the Chinese nation refers to the overall image of the Chinese community to the outside world, which is a relatively stable and generally recognized cultural symbol system formed through historical accumulation. Chinese cultural symbols and the image of the Chinese nation are pluralistic and integrated historical existences, both as objective realities and as spiritual and cultural existences. National identity is not an empty and abstract spiritual idea, but a concrete and sensual image rooted in real life [8]. These constructive imaginary realities are derived from everyday reality. The Chinese mountains, rivers, lakes, plants and flowers, festivals, cultural customs and activities, and “mascots” with national characteristics are typical representatives of the most Chinese image and cultural symbolism in daily life. From daily life reality, we analyze the national and cultural psychological structure behind these familiar national images and cultural symbols, from perceptual experience to rational thinking, and systematically explore Chinese images and cultural symbols with cultural geography, cultural psychology, cultural politics [9, 10], etc. It is of great theoretical and practical significance to cast a firm sense of Chinese national community.

Usually, cultural symbols are mainly composed of “shape,” “meaning,” “color,” and “movement” in their visual expression and application to grasp the four aspects. They do not exist as isolated symbols but are integrated to function together. The graphics will “speak,” the mood will “feel,” the colors will “tell,” and the dynamics will “tell.” In addition to giving the audience a sense of beauty, an excellent image film should be able to communicate directly with the audience. To this end, this paper proposes a symbol creation technique for visual images with the help of deep learning techniques. Firstly, the mode collapse of generative adversarial network (GAN) is improved by combining the pattern search regularization term with generator cross-loss function. Secondly, the common convolution layer of the generator in the network structure is replaced by deep-space separable convolution to improve the real-time performance of the model by reducing model parameters.

2. Current Status of Research

Chinese national culture has undergone thousands of years of development and evolution and has accumulated a wealth of symbolic material resources, and although our visual image design has made a lot of achievements in using these materials, some of them always have problems such as blindness and imitativeness in using local culture, favoring the form, and failing to accurately convey the main idea. Art designers [11] must have an overall clear knowledge of these national symbols in bridging modern visual communication design and national cultural symbols, and they should deeply understand the cultural background and spiritual connotation behind the symbols, be good at exploring, extending, and reconstructing them, reflecting the regional characteristics and national temperament they contain, and creating design work with strong vitality that meet the aesthetics of contemporary visual images [12].

2.1. Image Generation Model. In recent years, deep neural networks have made great progress in the field of image generation. Generally speaking, there are two types of commonly used generative models for image generation. One type of variational auto-encoder (VAE) [13] based generative model is a probabilistic statistical model. It mainly consists of an encoding network for inferring the statistical information of the input image and a decoding network for reconstructing the input image. This type of generative model has a series of advantages such as stable training and fast convergence. However, since the objective function of the VAE model optimization is the lower bound of the log-likelihood function, the generated image looks blurred overall. Another class of image generation models based on GANs [14] implicitly captures the probability distribution of real images through adversarial learning.
between generators and discriminators. Through adversarial training, GANs-like models can generate clearer and more realistic images. However, the original GANs models still have many significant drawbacks, such as pattern collapse and training instability. Researchers have proposed some effective methods to stabilize the training process and improve the quality of the generated images. Besides, to make the generated images have certain desired properties, the researchers propose to use conditional GANs to constrain the generated images. For example, the introduction of auxiliary information such as class labels to guide the generation of handwritten digital images. So far, CGANs-based models have been widely used in super-resolution image generation, image style conversion, image restoration, and other fields. Because of the excellent performance of the adversarial network in generating images, we also use the adversarial network to complete the generation of target symbols.

2.2. Symbol Generation Model. Research work on symbol generation using deep neural networks can be broadly classified into two categories: one category is that the researchers use discrete symbol properties for symbol generation. For example, literature [15] proposed to encode information such as gender, expression category, and hair color into the bottleneck layer of the conditional VAE model to generate a facial expression generation model with diverse appearances. Literature [16] classified different expression states such as angry and happy into different domains and proposed StarGAN to realize the interconversion between several typical expressions. The discriminator of StarGAN needs to determine not only the authenticity of the generated image but also the domain from which it comes. Although these methods can generate high-quality symbolic expression images, however, the encoded discrete symbolic attributes are not sufficient to describe the rich culture. To solve this problem, the researchers explored how to integrate continuous auxiliary information into the generative model. Literature [17] proposed the CDAAE model that can separate symbolic information from each other. In CDAAE, given a reference symbol image, multiple cultural representations of the same culture can be generated by changing the FAU (Facial Action Unit) labels that represent different cultural strengths. GAGAN combines symbol shapes and GANs to make the generated symbols realistic, natural, and with specified symbol shapes. However, due to the semi-supervised nature of GAGAN itself, it does not provide any control over the generated symbolic information. Concerning the generation of continuous symbols, literature [18] have proposed a direct linear interpolation of two different symbol shapes using symbol feature points, and then these shapes are compressed into a one-dimensional coding vector using a fully connected network, and finally, the coding vector is fed into the adversarial network to generate the continuous symbols. Literature [19] also proposed the G2GAN model for cultural symbol synthesis with symbolic feature points as the controllable condition. G2GAN achieves the removal and generation of cultural symbols through two generative networks, respectively, and then achieves the conversion of arbitrary symbols.

In this paper, we propose a simple and effective model for the generation of Chinese cultural identity symbols based on existing methods of facial expressions generation. To model different symbol shapes, we combine pattern search regular terms with generator crossover loss functions based on GAN generative adversarial networks and replace the normal convolutional layer of the generator in the network structure with a deep-space separable convolution improving the real-time performance of the model by reducing the model parameters.

3. Methodology

3.1. GAN. GAN consists of two parts: a generative network and a discriminative network. The discriminative network is to distinguish between the fake samples generated by the generator and the real samples. In contrast, generative networks are used to confuse the discriminator by generating fake samples. The generative and discriminative networks are trained simultaneously throughout the training process, constituting a dynamic two-player min-max game [20]. The GAN training process is shown in.

\[
\min_{G} \max_{D} V(D, G) = \mathbb{E}_{x \sim P_{data}(x)} \left[ \log D(x) \right] + \mathbb{E}_{z \sim P_{z}} \left[ \log (1 - D(G(z))) \right],
\]

where \(x\) is the true sample, \(\bar{x}\) is the generated sample, \(P_{data}\) is the true sample distribution, \(P_{G}\) is the generated sample distribution, \(D\) is the discriminator, and \(G\) is the generator. In the “game” process, when the \(D(\bar{x})\) value increases, the generated sample is largely close to the real sample; when the \(D(x)\) value is larger and the \(D(\bar{x})\) value is smaller, the generated samples are more easily distinguished from the real samples. During the continuous game between the generator and the discriminator, the discriminator and the generator seek to minimize the JS (Jensen-Shannon) scatter. When and only when \(P_{data} = P_{G}\), the global reach is optimal.

3.2. StarGAN. Compared to other GAN models, StarGAN solves the problem of interconversion between multiple cultural symbol categories. The model structure of StarGAN [21] is relatively simple and efficient, and its generator receives the input target domain \(c\) and the input sample \(x\). The false samples output by \(G\) are transmitted to \(D\), on the one hand, to determine the true and false samples and to perform domain classification. On the other hand, this fake sample is transmitted again to the generator with the target domain label \(c'\) of the input sample as input, to confuse the output sample with the original input sample and improve the similarity between them. The network structure of StarGAN is shown in Figure 1.

3.3. MS-StartGAN. In cultural symbol image generation, the quality of the samples obtained after text description transformation has an important impact on cultural symbol recognition. MSGAN [22] proposes to quantify the mode of
model collapse, i.e., distance ratio, to improve the mode collapse phenomenon by increasing the distance ratio, which in turn improves the sample quality. StarGAN can solve the problem of interconversion between multiple symbol classes, and the model structure is relatively simple and efficient. Thus, this paper proposes a pattern search StarGAN (MS-StarGAN) by combining the features of StarGAN and MSGAN models. MS-StarGAN adds a pattern search regular term to the generator objective function, which further solves the pattern collapse phenomenon by increasing the distance ratio to avoid input vectors with similar features appearing at the same mapping position all the time, resulting in improved quality and richness of symbolic images. The generator structure uses spatially separable convolution instead of convolutional layers, thus reducing the training parameters of the model and effectively improving the stability of the model training. The principle of MS-StarGAN is shown in Figure 2.

The dummy sample $G(x, c)$ is transmitted to the discriminator, which determines whether the generator generated the dummy sample $G(x, c)$ or the input sample $x$, and the domain classification result. And the fake sample $G(x, c)$ generated by the generator will be passed back to the generator again. At this time, this dummy sample $G(x, c)$ with the target domain label $c'$ of the input sample is used as input, to confuse the output sample with the original input sample and improve the feature similarity between them. Adding a pattern search regular term between the input sample and the generated sample further improves the pattern collapse phenomenon and makes the generated cultural representation symbol sample more natural and smoother.

3.3.1. Generating Networks. The structure of the MS-StarGAN generator is shown in Figure 3. The entire network depth of the MS-StarGAN generator is 18, which includes 3 different convolutional layers, 6 residual blocks (each of
which contains 2 layers of ordinary spatial convolutional layers), and 2 transposed convolutions. First, the first part receives samples and labels as input, with a convolution kernel size of $7 \times 7$, a step size of 1, and a fill size of 3, and adds the instance normalization layer and ReLu as the excitation function. The incentive function accelerates training and improves stability. The second and third layers are down sampled with a convolution kernel size of $4 \times 4$, a step size of 1, and a fill size of 2 to obtain a $4 \times 4 \times 256$ feature map. Secondly, a spatially separated convolution is used in the middle part to separate $3 \times 3$ into $3 \times 1$ and $1 \times 3$, to reduce the number of parameters for network training. The final up sampling is performed using transposed convolution and the output uses the inverse function Tanh.

3.3.2. Discriminant Network. For the discriminator network, the entire network depth is 7. It inputs either true or false samples and determines their truth or falsity and the target domain they belong to, with a convolution kernel size of $7 \times 7$, a step size of 1, and a fill size of 3, and adds the instance normalization layer and ReLu as the excitation function. The incentive function accelerates training and improves stability. The second and third layers are down sampled with a convolution kernel size of $4 \times 4$, a step size of 1, and a fill size of 2 to obtain a $4 \times 4 \times 256$ feature map. Secondly, a spatially separated convolution is used in the middle part to separate $3 \times 3$ into $3 \times 1$ and $1 \times 3$, to reduce the number of parameters for network training. The final up sampling is performed using transposed convolution and the output uses the inverse function Tanh.

Traditional standard convolution uses a convolution kernel of size $L \times L$ and an output channel of size $M$ for each convolution calculation; the deep convolution DW given in Figure 5 uses $M$ convolution kernels of size $L \times L$ for each convolution calculation during the operation, and the output is usually 1. The point-by-point convolutional PC uses $M$ convolutional kernels of size $1 \times 1$ for convolutional filtering in each operation. The depth convolution DW and point-by-point convolution PC can be spliced into a standard convolution with a convolution kernel of size $L \times L$ and channel $M$. Among them, the number of parameters involved in the standard convolutional CNN calculation process is shown in:

$$\Delta_{\text{CNN}} = L \times L \times M \times N. \tag{2}$$

The number of parameters involved in the computation of depth-separable convolution with the combination of depth convolution DW and point-by-point convolution PC is shown in:

$$\Delta_{\text{DW,PC}} = L \times L \times 1 \times M + 1 \times 1 \times M \times N, \tag{3}$$

$$R = \frac{\Delta_{\text{DW,PC}}}{\Delta_{\text{CNN}}} = \frac{L \times L \times 1 \times M + 1 \times 1 \times M \times N}{L \times L \times M \times N} = \frac{1}{L^2} + \frac{1}{N}. \tag{4}$$

From equation (4), it can be seen that when the convolution kernel $L \geq 2$, the number of parameters involved in the depth-separable convolution calculation process is significantly less than the number of parameters involved in the
standard convolution calculation process. Thus, using depth-separable convolution instead of standard convolution can reduce the time overhead during the convolution calculation.

In the deep convolution process, a channel of the feature map is convolved by only one convolution kernel, and the number of convolution kernels is equal to the number of channels. The depth convolution is calculated as shown in:

\[ F_{l,w,m} = \sum_{i,w,m} L_{l,w,m} \times X_{i+1, w+m} \times L \times L \times M \]

where \( F \) denotes the input feature map; \( L \) is a filter of length \( l \) and width \( w \), \( X \) denotes the input matrix, and \( m \) denotes the \( m \)th channel of \( M \).

3.3.3. Crossover Loss Function. To achieve end-to-end optimization of the model in this paper, the cross-entropy loss function is used to calculate the deviation of the target value from the actual output value. The optimal value intervals used for positive and negative samples were obtained through extensive experiments. This function causes the neural network to output in the form of a probability distribution. Thus, cross-entropy can calculate the distance between the predicted probability distribution and the actual output probability distribution. Meanwhile, adding the equilibrium parameter \( \theta \) can improve the prediction accuracy. The calculations are shown in equations (6)–(8).

\[ X_t = uh_t + b, \quad (6) \]

\[ P(y|x) = \frac{e^{h(x,y)}}{\sum_{y} e^{h(x,y)}}, \quad (7) \]

\[ L(\theta) = -\sum_{t=1}^{r} \log p(y^{t} | y_{t-1}^{*}), \quad (8) \]

where \( \theta \) denotes the model cross-entropy loss balance parameter.

4. Experimental Results and Analysis

4.1. Experimental Environment and Evaluation Index. The data set of this paper is mainly divided into two parts. The first part is self-built test data set, and the training data set mainly includes self-built data set and the existing open-source network data set. It mainly includes 13 kinds of symbols of national culture, including Chinese national heroes, festival representatives, and mascot copywriting.

The experimental environment in this paper is Windows 10 operating system. The experimental platform is an Intel(R) Core i7-7800X processor with six cores and six threads at 3.5 GHz. The experiments use the Pytorch deep learning framework to build the model, the development language is Python, and the V100 32G GPU with SGD optimizer to optimize the model parameters. The model has an initial learning rate of 0.0001, a Batch Size of 32, a weight decay rate of 0.0005, and a momentum of 0.9. Besides, to prevent overfitting of the model, Dropout is set to 0.5. The Loss and accuracy curves of the training and testing phases are shown in Figure 6. It can be seen that after the number of iterations reaches 240, the accuracy and Loss curve regions of the training and testing phases are smooth and the model reaches stable convergence.

To verify the effectiveness of the algorithm in this paper, a number of mainstream evaluation metrics are used to evaluate the model performance, specifically Accuracy, Precision, Recall, F1-score, and Time Overhead (TO) of action recognition for a single image, and the calculated expressions are shown in equations (9)–(12). The confusion matrix is shown in Table 1. In particular, the calculations between the Precision and Recall metrics are contradictory, and for this reason, the Precision-
Figure 6: Curves of the training and testing phases. (a) Loss curve. (b) Accuracy curve.

Table 1: Confusion matrix calculation.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>$T_P$</td>
<td>$F_P$</td>
</tr>
<tr>
<td>Negative</td>
<td>$F_N$</td>
<td>$T_N$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.1</th>
<th>No.2</th>
<th>No.3</th>
<th>No.4</th>
<th>No.5</th>
<th>No.6</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Confusion matrix No.1" /></td>
<td><img src="image2.png" alt="Confusion matrix No.2" /></td>
<td><img src="image3.png" alt="Confusion matrix No.3" /></td>
<td><img src="image4.png" alt="Confusion matrix No.4" /></td>
<td><img src="image5.png" alt="Confusion matrix No.5" /></td>
<td><img src="image6.png" alt="Confusion matrix No.6" /></td>
</tr>
</tbody>
</table>

Figure 7: Confusion matrix. (a) No. 1. (b) No. 2. (c) No. 3. (d) No. 4. (e) No. 5. (f) No. 6.
Recall curve is used for comparison in this paper. The larger the area enclosed under the curve, the better the classification performance of the model.

\[
\text{Accuracy} = \frac{Tp + Tn}{Tp + Fp + Tn + Fn} \tag{9}
\]

\[
\text{Precision} = \frac{Tp}{Tp + Fp} \tag{10}
\]

\[
\text{Recall} = \frac{Tp}{Tp + Fn} \tag{11}
\]

\[
F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}
\]

4.2. Analysis of Results. The confusion matrix generated by the method in this paper in six sets of experiments is given in Figure 7, where the rows of the matrix represent the real symbol labels and the columns represent the symbol labels generated by the algorithm. From the confusion matrix, it can be concluded that the number of successive generations for the six symbols in the six sets of experiments is 205, 205, 211, 206, 203, and 208, and the accuracy of their generation is 93.61%, 93.61%, 93.36%, 92.79%, 92.69%, and 94.98%, respectively. In addition, the model in this paper can achieve a generation rate of 8 ms/images. From the above data, it can be seen that the model in this paper tends to perform stably on the results of multiple experiments, and also has good real-time performance, which verifies that the model in this paper has good robustness.

4.3. Comparison of Related Work. In order to verify the validity of the model in this paper, comparison experiments are conducted with the current mainstream models A [17], B [25], C [18], and D [26], respectively, and analyzed under the same data and environment. The detailed data are shown in Table 2. Figure 8 gives a comparison of the time overhead of the different models.

From Table 2, it can be seen that the model in this paper can achieve 93.68% Accuracy, 93.02% Precision, 92.96% Recall, and 92.88% F1. In terms of Accuracy, compared with the two best performing models C and D among all the compared models, the model in this paper improves (92.36% → 93.68%) and (92.37% → 93.68%), respectively. In terms of Precision, the model in this paper improved (91.98% → 93.02%) and (92.56% → 93.68%). In Recall, the models in this paper improved (92.05% → 92.96%) and (92.37% → 93.68%). In terms of F1, the model in this paper improves (91.81% → 92.88%) and (91.94% → 92.88%). The above data further verify that the model in this paper has better performance for symbol generation of Chinese national culture. This greatly promotes the spread of Chinese culture and increases the sense of national cultural identity and belonging.

From Figure 8, we can see that the model in this paper can achieve a generation rate of 8 ms/images, model A can achieve a generation rate of 18.9 ms/images, model B can achieve a generation rate of ms/images, model C can achieve a generation rate of 12.86 ms/images, and model D can achieve a generation rate of 13.91 ms/images generation rate. The above data further show that the model in this paper can achieve a better generation rate, mainly because this paper uses depth-separable convolution instead of traditional convolution in the discriminator stage to achieve a lower model time overhead by reducing the model parameters.

5. Conclusion

In this paper, we propose a new national culture symbol generation model based on the GAN network, which can greatly promote human’s sense of identity and belonging to national culture and rapidly spread national culture. Specifically, this paper first proposes a pattern search network, MS-StartGAN, based on the GAN generative network by modeling based on the described data, and the pattern collapse phenomenon is further addressed by adding pattern search regular terms to the generator objective function to increase the distance ratio and alleviate input vectors with similar features from appearing at the same mapping positions all the time, which improves the quality and richness of the generated symbols. By testing on a large number of experiments, the results show that the model in this paper has good real-time performance while maintaining high generation accuracy.

Data Availability

The data used to support the findings of this study can be obtained from the corresponding author upon request.
Conflicts of Interest

The author declares that there are no conflicts of interest.

References


