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

Combinatorial Press

Combining Artificial Intelligence Algorithms to Optimize the Digital Innovation Design of Dong Brocade Patterns Research

Haiyun Yang[1], Jincheng Yang[1,✉]

[1] *School of Fine Arts and Design, Hunan First Normal University, Changsha, Hunan, 410000, China*

ABSTRACT

AI technology in the development and application of traditional texture recovery and reproduction, deep learning models for traditional texture information and color information consistency migration is still deficient, this paper by using the visual Transformer network advantage and visual Transformer network Transformer encoder structure optimization. That is to say, in the Transformer encoder, the multi-head self-attention module and feed-forward network module are called to process the input data and extract the image features, and then join the edge preservation smoothing technology to remove the strong edge information, preserve some weak edges and local colors, and generate the image texture information with the input texture. The color interpolation method is used to achieve the consistency of texture color texture and image texture migration. The result images of Dong brocade texture style migration show that the image texture migration model based on visual transformer is more capable of generating images with the best style loss value and the best content loss value, and is able to obtain more than 70% of user preference.

Keywords: visual transformer, multi-head self-attention, image texture, Dong brocade pattern

1. Introduction

Dong brocade has become one of the famous brocades in China for its exquisite weaving techniques, rich ethnic patterns, rich and profound cultural connotations and bright and harmonious colors, which witnesses and records the long history and profound cultural connotations of the Dong ethnic group [1]. The production process of traditional ethnic brocade is complex and delicate, including spinning, dyeing, weaving and other aspects. For spinning, hand-spun cotton or silk threads are used, which require great skill and patience to spin an even and tough thread [2-3]. Dyeing is usually done with vegetable dyes and has a unique color and permanence [4-5]. Weaving is the most complicated part of the whole production process, which needs to be done on a loom according to a pre-designed pattern

✉ Corresponding author.
E-mail address: 15974010808@163.com (J. Yang).
Received 03 March 2024; Revised 20 April 2024; Accepted 10 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-347

[6-7]. With the continuous development of industrial civilization, Dong brocade art is facing the problem of shrinking development space, which will cause irreparable and priceless losses, and the protection and inheritance of Dong brocade and innovative development have become a new trend of the times [8-10].

In today's highly developed material and spiritual civilization, people put forward higher requirements on the functionality and artistry of Dong brocade, so the inheritance of Dong brocade art should not only be limited to the dissemination of skills and the preservation of data, but also should be improved and innovated in line with the development of the times. In the era of informationization, the digital media industry is rapidly rising, and the use of digital technology provides the design concept and technical support for the inheritance and innovation of brocade art [11-14]. Through digital technology innovation brocade pattern design, more and more people accept and identify with the Dong brocade art, which can effectively promote the inheritance of intangible cultural heritage [15-16]. The integration of traditional Dong brocade art into modern design with the help of digital technology not only meets the pursuit of modern aesthetics and personalization, but also realizes the modern inheritance of Dong brocade art.

This paper analyzes the characteristics of Dong brocade craft and pattern, and the migration of Dong brocade pattern style needs to be considered from the three aspects of modeling, composition and color. Thus, a visual transformer network is proposed on the basis of deep learning model, focusing on analyzing the composition of Embedding layer and Transformer encoder. The visual Transformer network structure is designed to adjust the multi-head self-attention module and the feedforward network respectively. Realize texture migration of traditional textures by applying edge preserving smoothing method with fast global smoother. The color consistency module is set up to generate a new image containing texture information and color information using color interpolation to complete the traditional texture migration, and the steps together form a visual transformer-based image texture migration model. The advantages of using visual transformer network are analyzed, and the results of Dong brocade texture style migration are quantitatively analyzed.

2. Application of artificial intelligence technology to traditional patterns

2.1. Application of AI technology in pattern restoration and reproduction

With the continuous progress of machine learning and deep learning technology, AI technology is able to recognize and classify complex tattoo patterns [17]. It can accurately identify and learn the lines, colors, shapes and other features of the patterns, revealing the cultural logic and aesthetic laws behind the patterns. And by learning a large amount of historical data and images, thus maintaining the original cultural significance and aesthetic value in the reproduction process of patterns. And it can recover those ancient patterns that have been partially lost or damaged, improve the accuracy and efficiency of the recovery of non-heritage patterns, provide technical support for the protection and inheritance of traditional patterns, and realize the iteration and reconstruction.

With the help of deep learning models that can deeply learn the structure, color and form of traditional patterns, AI is able to generate pattern patterns that match the style of the original pattern and have novel creativity. This innovative pattern design not only retains the flavor of traditional elements, but also incorporates modern aesthetic innovation.

With the diversification and personalization of consumer aesthetics, AI technology is able to customize patterns according to different application scenarios and user needs, realizing the integration of traditional culture and modern design.

With the continuous development and improvement of AI technology, its application in the digital reproduction of non-heritage clothing will be more extensive and in-depth, and the in-depth fusion of science and technology and culture will bring new vitality and possibilities for the inheritance and development of non-heritage.

2.2. Dong brocade craft and pattern

Through the changes of the times, Dong women have accumulated rich experience in the process of continuous practice from planting cotton to dyeing and weaving, and the Dong brocade weaving technique has gradually formed a highly technical form of artistic activity. In order to ensure the quality of Dong brocade, its technique is unique and strictly characterized by complicated procedures and exquisite skills.

The production process of Dong brocade is very complicated and hand-operated, and the general process can be divided into four categories: raw material processing, yarn treatment, yarn finishing and loom weaving. The quality of yarn is a key factor in determining the quality of Dong brocade, so the Dong people have spent a lot of thought on the treatment of yarn, and the specific processes are cotton popping, cotton rolls, spinning, spinning, spinning, sizing, dyeing, coiling, etc. The Dong brocade has a long history of production, and it is all operated by hand.

Dong brocade has a long history of exquisite skills, Dong brocade pattern is complex and varied, through the changes of the times it has gradually formed a cultural symbols generally recognized within the nation.

1) Modeling. According to the different weaving methods, the pattern modeling expression can be divided into two types: "pictographic ideograph" and "image generated by image". Among them, the "pictographic ideogram" is manifested as the realism of the pattern modeling, which is derived from life, and is very similar to the image form in reality and has a high degree of recognition. The expression of "imagery from imagery" is also widely used in Dong brocade, which is mainly manifested in the abstraction of modeling.

2) Composition. Dong brocade patterns are rich in shape, with varied composition styles, commonly symmetrical and central, with freedom in regularity, and with a strong Dong family flavor. Dong brocade patterns use basic dots, lines and surfaces to form the overall picture, and most of them use geometric shapes composed of lines as the skeleton, and fill the patterns composed of dots and surfaces into them, so that the brocade surfaces are harmonized and symmetrical in structure.

3) Color. Dong brocade has plain brocade and colored brocade. Due to the influence of regional conditions and the limitation of production conditions, the traditional Dong brocade is mainly woven with plain brocade. Plain brocade is simple and generous, mostly with a single color or two colors, with simple colors such as black, white, blue and other cold tones. In the later period, due to changes in social concepts and aesthetic preferences, the color of brocade has become more brilliant, bold and unrestrained. Dong brocade color matching often use contrasting and complementary colors, the contrast of color changes in degrees, pay attention to balance, often use a color as the main color, other colors as a color accent.

2.3. The regeneration of modern language of Dong brocade art

Dong brocade has long been loved by people and become one of the symbols of Dong spirit and faith worship. The cultural spirit, folk customs and artistic aesthetics it contains have brought rich inspiration to contemporary design. From the point of view of the stylistic characteristics of modern art design, not all the contents of Dong brocade art can be borrowed, but must go through a process of selection and abandonment in the context of modern art design. In order to realize the innovation of selection and abandonment of Dong brocade art in modern design with ease, it is necessary to grasp the inner essence and cultural temperament of Dong brocade art, and then combine with the principles of modern design composition and technology to regenerate the Dong brocade art and culture through modern design language.

1) Modern variation of graphic structure. Designers should infiltrate modern composition principles into Dong brocade patterns, applying basic variation, decomposition and reconstruction, heterogeneous isomorphism and other methods to continuously expand new visual expression space, help Dong brocade and contemporary design to perfectly integrate, so as to make it conform to the

aesthetic judgment standards of the modern market.

2) New interpretation under color collision. The color of Dong brocade is a concentrated expression of the aesthetic and cultural nature of the Dong people, absorbing the essence of the Dong brocade color system. Inheriting and developing the color aesthetics of Dong brocade in contemporary design can help enrich the color expression of contemporary design. For the reuse of Dong brocade colors, firstly, we can combine the symbolic meaning, decorative aesthetics and contrasting and coordinating beauty of Dong brocade colors into the color design. Secondly, the color of Dong brocade can be summarized and refined, and the color of Dong brocade can be reconstructed by using modern color aesthetics.

3) Breakthrough of new materials and technology. The source of Dong brocade is mostly at the decorative level of traditional crafts, which makes most designers more inclined to two-dimensional pattern composition research when applying Dong brocade to contemporary design, often ignoring the innovation of materials and the reform of new media technology. Therefore, applying certain features of Dong brocade to modern design media materials or special materials will have innovative visual effects.

4) Innovation of integrating fashion elements. Injecting the blood of modern fashion on the basis of Dong brocade and integrating the market-recognized fashion elements into it, the design works will be more vivid and meet the aesthetic standard of modern people. This kind of thinking is an innovative expression of Dong brocade in contemporary design. Currently designers and brands are beginning to pay attention to the fashionable development of Dong brocade elements. Liang Yun's series of works, “Jinqi”, integrates Dong brocade into daily wearable accessories, expressing the life of a Dong woman through different choices of color schemes, patterns and materials.

3. Traditional tattoo style migration techniques

3.1. Demand Analysis of Traditional Cultural Image Reconstruction

The style information of traditional patterns (Dong brocade patterns) can be divided into texture information and color information. In the process of style migration, in addition to the original texture information and graphic patterns are migrated, the color style is also migrated to the target image, which will conflict with the original image color style, destroy the original color semantics of the target image, and contradict the user's intention. At the same time, the texture and pattern of traditional cultural images are very different and the color style is relatively single, the overall color of the pattern images drawn by the same author in the same period has great similarity and overlap, so it is a worthwhile research problem to extract a complete and clear texture with correct semantics from the images.

With the successful application of transformer in the field of Natural Language Processing (NLP), the image field also ushered in a wave of architectural combination with transformer and proposed visual transformer.Based on the above analysis, this paper proposes an image texture migration model (TexTr) based on visual transformer.

3.2. Visual Transformer

Vision Transformer, or ViT, is the first network model that discards convolution purely using Transformer for image classification [18-19].

The structure of the ViT model is shown in Fig. 1.ViT refers to the structure of Transformer and designs a vision Transformer as standard as possible with as few changes as possible to the Transformer.

ViT can be divided into an Embedding layer that chunks the image, a Transformer encoder and a classification header. The construction of the classification header is simple and mainly consists of two fully connected layers as well as an activation function. This section next focuses on Embedding layer

and Transformer encoder in detail.

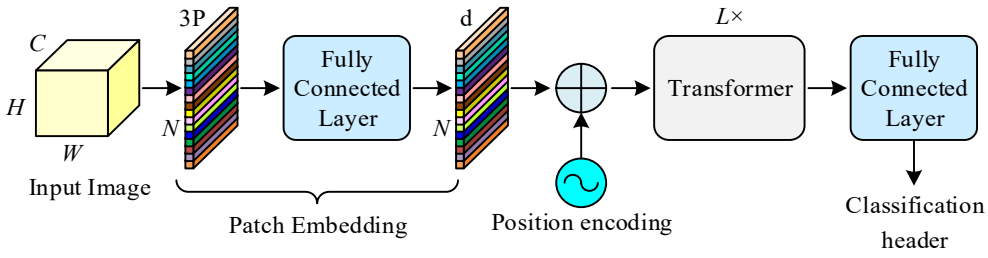

Fig. 1. ViT model structure

1) Embedding layer. In natural language processing Transformer model usually inputs a two-dimensional sequence of word vectors, but in the visual domain inputs are three-dimensional images. In order to input the image information into the Transformer, ViT slices the input image into multiple image blocks of the same size, which are called Patch in ViT, and then inputs each Patch into the Embedding layer to get the corresponding vector, which is called Token in ViT.

The details are as follows: the original idea of ViT was to use a standard Transformer for image classification tasks. In the original standard Transformer the input is a two-dimensional matrix of size N × D. N is the length of the input sequence and D is the dimension of each word vector in the input sequence. For the input image $I \in R$H × W × C, H, W, and C are the length, width, and number of channels of the image, respectively. In order to transform the 3D image into the dimensions required by the standard Transformer, ViT first divides the input image into a number of Patch of size P × P. Then the input sequence N can be computed as N = H × W/P2 .Meanwhile, the input dimension D of the image can be computed as D = P2 × C. The process of compression of a vector of size P2×C into D is known as Patch Embedding.To do the final classification task, ViT refers to the class labeling in BERT and adds a learnable vector of dimension 1 × D for classification. In addition, like Transformer, ViT also adds position encoding to identify the positional information of each Patch as a way to ensure the positional relationship of the images in the input information. The final image I_0 after the Embedding layer is shown below:

$$I_0 = \left[I_{cls}; I_P^1 E; I_P^2 E; \ldots; I_P^N E \right] + E_{pos} \tag{1}$$

Where I_{cls} is the learnable vector used for classification, I_P^1 to I_P^N are the N $Patch$ sliced into $P \times P$ sizes. E is the fully-connected layer parameter matrix used for compression of size $P^2 \times C$, and the output is of size D. E_{pos} is the position coding matrix of the same size as E.

2) Transformer encoder. The Transformer for machine translation uses the encoder-decoder architecture common in deep learning. The encoder encodes long sequences into word vectors, and the decoder generates words and statements from the word vectors to accomplish the machine translation task. The most important task in the vision task is to extract the image features, so only the structure of the encoder used in Transformer to extract the features is used in ViT.

The encoder in ViT mainly consists of Norm normalization layer, Multi-Head Attention Multi-Head Attention layer and MLP fully connected layer, which uses residual connection to stack the Encoder modules L times. The mechanism of Multi-Head Self-Attention is to do multiple attention operations, and then the obtained Q, K, and V are spliced to get the final attention result respectively. I.e:

$$Head_i = Attention\left(QW_i^q, KW_i^k, VW_i^v\right) \tag{2}$$

$$MultiHead\left(Q, K, V\right) = Concat\left(Head_1, \cdots, Head_n\right) W^m \tag{3}$$

Equation (2) and Equation (3) show the process of Multi-Head attention computation, where W_i^q,

W_i^k, and W_i^v are the learnable parameter matrices corresponding to Q, K, and V, respectively. $Head_i$ is the result of a set of self-attention, and the final Multi-Head is the result obtained after splicing multiple sets of self-attention mechanisms and then going through a fully-connected layer, where W^m is the learnable parameter matrix produced after the fully-connected layer.

3.3. Migration model based on visual transformer

3.3.1. Network structure. TexTr obtains the corresponding image features which are used in the next stage to transform the sequence from one texture feature through to another texture feature. First the given content intermediate map I_{CS} is segmented into a sequence of input content blocks:

$$I_{cs} = \{\varepsilon_{c1}, \varepsilon_{c2}, \ldots, \varepsilon_{c\kappa}\} \tag{4}$$

where the block embedding ε is of shape $L \times C$, $L = \frac{H \times W}{8 \times 8}$ is of length ε, and C is of dimension ε. The algorithm first sends it to the Transformer encoder of the content. In the Transformer model, the encoder is responsible for processing the input data and extracting features, and each of its layers consists of two main parts: the Multihead Self-Attention Module (MSA) and the Feedforward Network (FFN). Each pixel of the block sequence is fed into the self-attention module, which generates the query (Q), key (K) and value (V) matrices through three different linear transformations. Namely:

$$Q = I_{cs}W_q, K = I_{cs}W_k, V = I_{cs}W_v \tag{5}$$

Where, W_q, W_k, $W_v \in \square^{C \times d_{bead}}$, $d_{head} = \frac{c}{N}$, C are the vector dimensions and N is the number of heads of the multi-head attention mechanism.

Then the multi-head attention is calculated by the following equation:

$$\begin{aligned} &\mathrm{F}_{MSA}(Q,K,V) = Concat \\ &\left(Attention_1(Q,K,V),\ldots,Attention_N(Q,K,V)\right)W_o \end{aligned} \tag{6}$$

where $W_o \in \square^{C \times C}$ is the matrix of weights that can be trained and N is the number of heads of the multi-head attention mechanism. i.e:

$$\begin{aligned} Y_c' &= \mathrm{F}_{MSA}(Q,K,V) + Q \\ Y_c &= \mathrm{F}_{FFN}\left(Y_c'\right) + Y_c' \end{aligned} \tag{7}$$

where $\mathrm{F}_{FFN}\left(Y_c'\right) = \max\left(0, Y_c'W_1 + b_1\right)W_2 + b_2$. A layer normalization is applied after each self-attentive block.

Similarly, the block sequence $I_{td} = \{\varepsilon_{t1}, \varepsilon_{t2}, \ldots, \varepsilon_{tL}\}$ of the texture reference image after it has been partitioned into blocks is encoded in a similar process as Y_t. The Transformer decoder decodes the encoded content sequence Y_c based on the encoded texture image sequence Y_t.

In this paper, we use the content block sequence to generate the query Q and the texture block sequence to generate the key K and value V:

$$Q = Y_cW_q, K = Y_tW_k, V = Y_tW_V \tag{8}$$

After layer normalization, the resultant sequence X of the algorithm can be shown in the following equation:

$$\begin{aligned} X'' &= \mathrm{F}_{MSA}(Q,K,V) + Q \\ X' &= \mathrm{F}_{MSA}\left(X'',K,V\right) + X'' \\ X &= \mathrm{F}_{FFN}\left(X'\right) + X' \end{aligned} \tag{9}$$

$\mathrm{F}_{MSA}(\cdot)$ represents the multi-head self-attention function, which is then added to the query matrix Q via residual concatenation. X'' is the output from the previous step. The feedforward network $\mathrm{F}_{FFN}(\cdot)$ handles the input X', which is then added to the output of the previous step via the residual connection X'. The final output X of the algorithmic decoder module is shaped as $\frac{HW}{64}\times C$.

3.3.2. Loss function. First, the texture image I_t is decomposed into structural layer I_t' and detail layer I_{td} by the edge preserving smoothing method of Fast Global Smoother (FGS). I_t' is obtained from the smoother while $I_{t\alpha}$ is the difference between I_t and I_t'. Here FGS is achieved by minimizing the energy function, also called weighted least squares (WLS). The following definition is obtained as described above:

$$J\left(I_t'\right)=\sum_p\left(\left(l_{tp}-I_{tp}'\right)^2+\lambda\sum_{q\in\mathrm{N}(p)}\omega_{p,q}\left(l_t'\right)\left(l_{tq}'-I_{tp}'\right)^2\right) \tag{10}$$

Where p denotes the position of a pixel in the image and $\mathrm{N}(p)$ denotes the neighboring pixel at position p. I_{tp} and I_{tp}' are the pixel values at a location in the texture image and the structural layer, respectively. The formula consists of two parts and the smoothing process is done by the latter part. Parameter λ is positively correlated with the smoothing degree. $\omega_{p,q}\left(I_s'\right)$ is a spatially varying weighting function which is defined as follows:

$$\omega_{p,q}\left(I_t'\right)=\exp\left(-\left\|I_{tp}-I_{tq}\right\|/\sigma_c\right) \tag{11}$$

where σ_c is a range parameter that is negatively correlated with the degree of edge preservation. The algorithm obtains I_t' by minimizing the energy function $J\left(I_t'\right)$. The detail layer energy function I_{td} of the texture image can be obtained by:

$$I_{td}=I_t-I_t' \tag{12}$$

Edge Preserving Smoothing (EPS) removes overall color and strong edge information from a textured image. While retaining some weak edges and local colors, which can be considered as real image texture information.

In this paper, the content difference between the resultant image I_0 and the content image I_{cs}, and the style difference between I_o and the input texture image I_{td} are represented by two different loss terms, as well as the full-variance regularization loss. The content loss and the texture loss are constructed according to the method using the feature maps extracted from the pre-trained VGG model. The texture feature loss $\mathrm{L}_{texture}$ is defined as:

$$\begin{aligned}\mathrm{L}_{texture}=&\frac{1}{N_l}\sum_{i=0}^{N_l}\left\|\mu\left(\phi_i\left(I_o\right)\right)-\mu\left(\phi_i\left(I_{td}\right)\right)\right\|_2\\&+\left\|\sigma\left(\phi_i\left(I_o\right)\right)-\sigma\left(\phi_i\left(I_{td}\right)\right)\right\|_2\end{aligned} \tag{13}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ denote the mean and variance of the extracted features, respectively, $\phi_i(\cdot)$ denotes the features extracted from layer i of the pre-trained VGG19, and N_l is the number of layers. The content-aware loss $\mathrm{L}_{content}$ is defined as:

$$\mathrm{L}_{content}=\frac{1}{N_l}\sum_{i=0}^{N_l}\left\|\phi_i\left(I_o\right)-\phi_i\left(I_{cs}\right)\right\|_2 \tag{14}$$

where $\phi_i(\cdot)$ denotes the feature extracted from the ith layer of the pre-trained VGG19 and N_l is the number of layers. In addition, the texture migration results are denoised using the full variance regularization loss. It is defined as follows:

$$\mathrm{L}_{tv}(I)=\sum_{i,j}\sqrt{\left(\left|I_{i+1}-I_{i,j}\right|^2+\left|I_{i,j+1}-I_{i,j}\right|^2\right)} \tag{15}$$

where i and j denote the i,j -pixel positions of image I_o. The final objective function to be optimized is:

$$\mathrm{L}=\lambda_c\mathrm{L}_{content}+\lambda_t\mathrm{L}_{texture}+\lambda_{tv}\mathrm{L}_{tv} \tag{16}$$

Set λ_c, λ_t, and λ_{tv} to 10, 7, and 1 to mitigate the effects of magnitude differences.

3.3.3. Color consistency. First, the content image is converted to the YUV color space, from which the image can be divided into a color channel and a luminance channel. Then, the structural information of the content image is extracted using the FGS method. Finally, the structural layer and the color channel will be converted into a new image by a color interpolation method. The interpolation method is defined as follows:

$$C_i=\begin{cases}\dfrac{A_i\times B_i}{128} & A_i\le 128\\ 255-\dfrac{\left(255-A_i\right)\times\left(255-B_i\right)}{128} & A_i>128\end{cases} \tag{17}$$

where A denotes the pixel value of the upper layer and B denotes the pixel value of the lower layer.

4. Stylized migration and customized realization of Dong brocade pattern

4.1. Algorithm Training and Performance Analysis

4.1.1. Experimental training process. The experiment was trained using pytorch under python 3.9 conditions.

Convolutional neural networks do not have strict requirements on the input image specifications. But the Transformer series network has strict requirements on the input image, it needs 224×224 or 256×256 specification image to be able to input into the network, so the image input into the Transformer needs to be preprocessed accordingly. First use transforms.RandomResizedCrop() function to crop the input image with random aspect ratio and scale it to 224×224 specifications, which meets the input specifications of the Transformer network. Next the image information is converted into tensor (Tensor) information using transforms.ToTensor() function. Finally this information is normalized using transforms.Normalize() function.

Normalization is a common data processing operation that scales data so that it falls into an interval. The normalization operation makes the effect caused by the direct difference of different data less, and also speeds up the training, and improves the effect of training, as shown in equation (18):

$$x=\frac{x-\mu}{\delta^2} \tag{18}$$

where μ is the mean of the corresponding dimension of the input image and δ is the variance of the corresponding dimension of the input image.

After the image data is input, the Vi T-base is taken as an example with dimensions [224,224,3]. The input image data is then further processed using a convolutional layer with 16×16, step size of 16, and the number of convolutional kernels is 768, at which time the dimensionality of the image data is [14,14,768], so that the whole image is segmented into 14×14 small image blocks. Then the information of each image block is spread and processed, and the dimension of the data obtained is [196,768]. In this way the image information basically meets the requirements. Then a trainable category label is spliced on these image information, the dimension of the category label is [1,768],

and the dimension is [197,768] after the completion of splicing. Next, it is added to the position encoding vector, which has a dimension of [197,768], and the dimension after addition remains unchanged, so that the total dimension is still [197,768]. Finally, this information is fed into the Transformer Encoder module for training.

After the training is completed, the output of the last Transformer Encoder layer is sliced in the output of the Transformer Encoder to extract the vector representing the category labels, which is obtained with a vector dimension of [1,768]. Then the vectors of category labels are processed using a fully connected layer and the probability of image classification is obtained using softmax classifier.

4.1.2. Experimental setup. The experimental environment here is shown in Table 1.

Table 1. Experimental environment

Experimental environment	Name
CPU	Inter(R)Core(TM)i7-6300HQ
Running memory	8GB
Operating system	Windows 11
Compiled language	Python 3.9
Code compiler	Py Charm Community Edition 2021.3.2
Depth learning framework	Pytorch 1.10
CUDA version	8.0.0

The experiments in this paper use training loss (Loss), classification accuracy, and single sample training time as experimental evaluation metrics.

Here the representative networks of convolutional neural networks MobileNet, ResNet, EfficientNet and the representative networks of Transformer model ViT, Swin Transformer are used for classification experiments on the classification dataset. Among them MobileNet uses v2 version, ResNet uses ResNet50. Vit uses the model as Vit-base, Swin Transformer uses Swin-base, and all networks are trained using the pre-trained weights of ImageNet, and the experimental hyper-parameters are shown in Table 2.

Table 2. Experimental superparameter setting

Parameter	Parameter value
Input image height	224
Input image width	224
Category number	8
Training wheel number	100
Training set sample number	3000
Test set sample number	600
Initial learning rate	0.0003
Dropout/Dropath	0.2
Whether to use pre-training weights	YES

4.1.3. Experimental results and analysis. Experiment 1: Obtain the accuracy rate of the representative network of convolutional neural network as well as the representative network of Transformer model on the dataset.

The accuracy rates of different algorithms are shown in Figure 2. The number of training rounds for each algorithm is 100.The accuracy rate of MobileNet model always fluctuates in the range of 0.88 to 0.92 during the training process. All other models have accuracy rates above 0.92, and the ViT model has the highest accuracy rate.

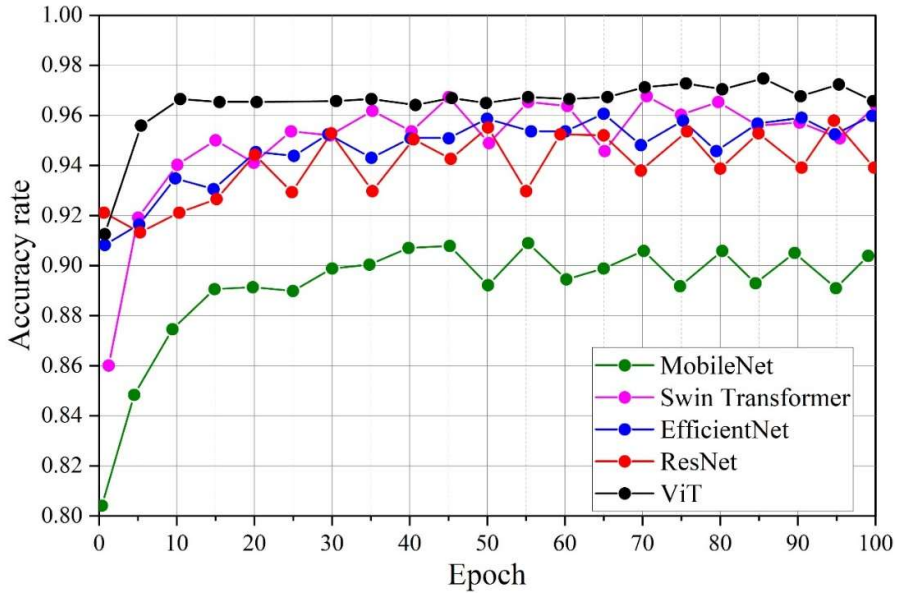

Fig. 2. Accuracy of different algorithms

Experiment 2: Obtaining the loss values of different representative algorithms, the loss curves of each algorithm are shown in Fig. 3. The ViT model has the largest loss function at the initial training, which is about 2.3, and when the number of training times comes to 20, the loss value of the ViT model is under control.

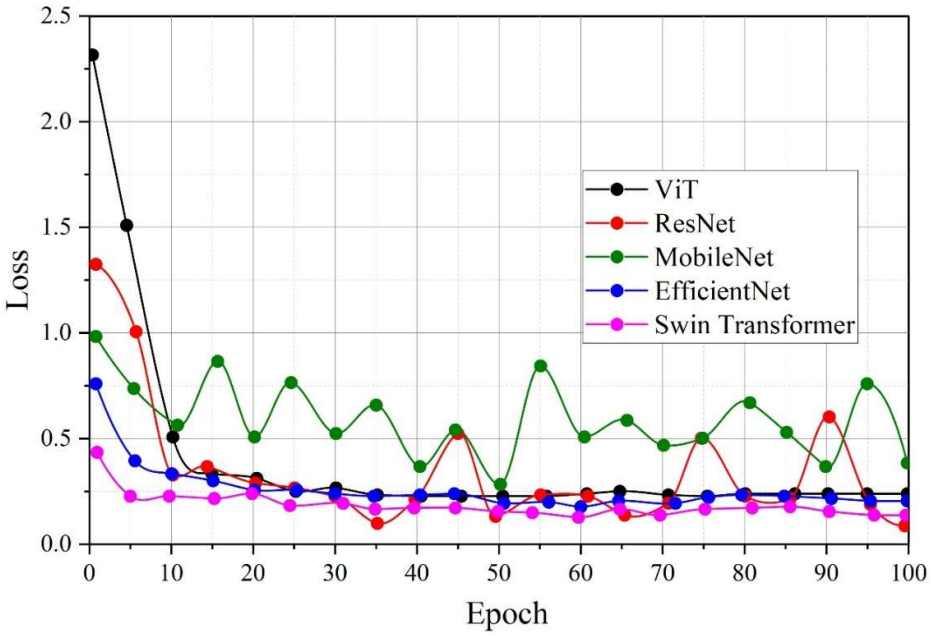

Fig. 3. The loss curve of each algorithm

Experiment 3: Obtaining the training time of different algorithms. The training length and the number of training times for each algorithm are shown in Fig. 4.Swin Transformer requires the longest training time for a single sample, ResNet also requires longer training time, followed by EfficientNet and ViT, and MobileNet requires the shortest training time.The MobileNet model requires about 0.3s of training time, and the model responds rapidly.

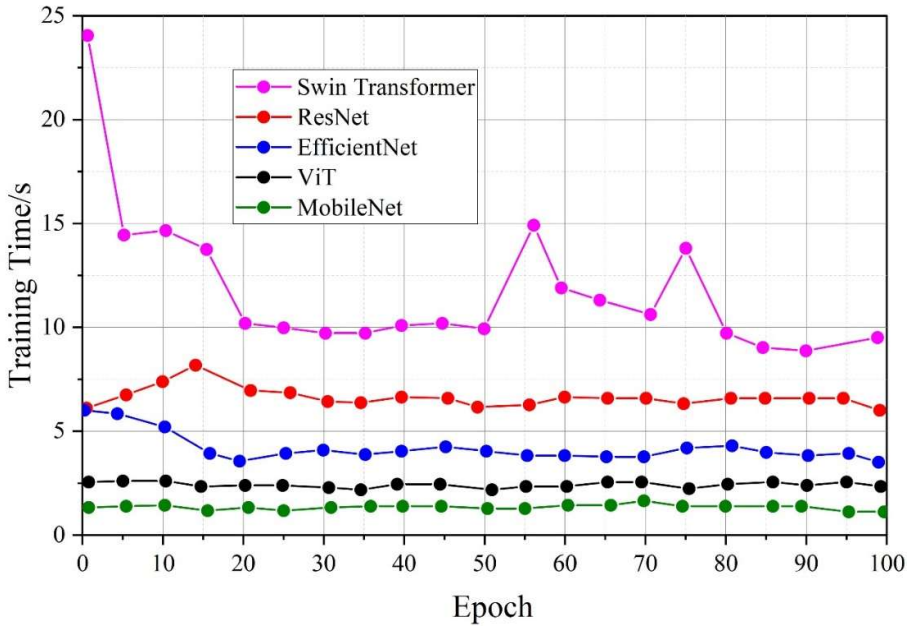

Fig. 4. The training length and training number of each algorithm

To select the training parameters of the Epoch with the highest accuracy rate among them from the 100 epochs, which are used as the comparison criteria for the subsequent experiments.

The training parameters of the Epoch with the highest accuracy rate are shown in Table 3.The classification accuracy rates of the Transformer model representative networks ViT and Swin Transformer on the dataset are 0.975 and 0.971 respectively.Comprehensively, it can be seen that the classification accuracy rate of the ViT class of networks is slightly better than the convolutional neural network in this dataset.

Table 3. The training parameters of the highest accuracy epoch

Network name	Maximum accuracy Epoch	Acc	Loss	Training Time/s
EfficientNet	65	0.960	0.391	4.60
Vi T	85	0.975	0.415	3.13
Swin Transformer	70	0.971	0.359	15.2
ResNet	95	0.958	0.533	8.01
MobileNet	55	0.910	0.876	0.30

4.2. Quantitative analysis of the results of the migration of Dong brocade pattern styles

The quantitative comparison experiment was divided into 3 parts. Under the condition of the same experimental environment and the same parameters, first, the style loss and content loss of TexTr and the remaining 5 baseline methods were compared. Under the same experimental environment, 5 existing methods were selected as baseline methods for performance comparison and analysis. Among them, the classical CNN-based style migration methods are Avatar, AdaIn, and MCC-Net. The engineered flow network-based style migration method is ArtFlow, and the visual Transformer model-based style migration method is $StyTr^2$

Secondly, the training time of each method and the time taken by each method to process a single image migration task with a size of 256×256 are compared. For each method 10 style images and 10 content images were selected from external data for 100 times of the style migration task, and the mean of style loss and content loss were calculated to avoid chance.

The results of style loss comparison under the same environment and parameters are shown in Fig. 5. The style loss and content loss are adopted as the quantitative evaluation index of the effect. Under the same conditions, the smaller the content loss and style loss, the more complete the image content and the closer the style features are to the style image in the style migration task.

The style loss values of TexTr, Avatar, AdaIn, MCCNet, ArtFlow, and $StyTr^2$ are 1.24, 3.38, 1.96, 1.77, 3.72, and 1.55, respectively.TexTr adopts a hierarchically structured visual Transformer model, which has the best style loss value in the comparison, i.e., it can generate style closest to Dong brocade pattern style image of the resultant image.

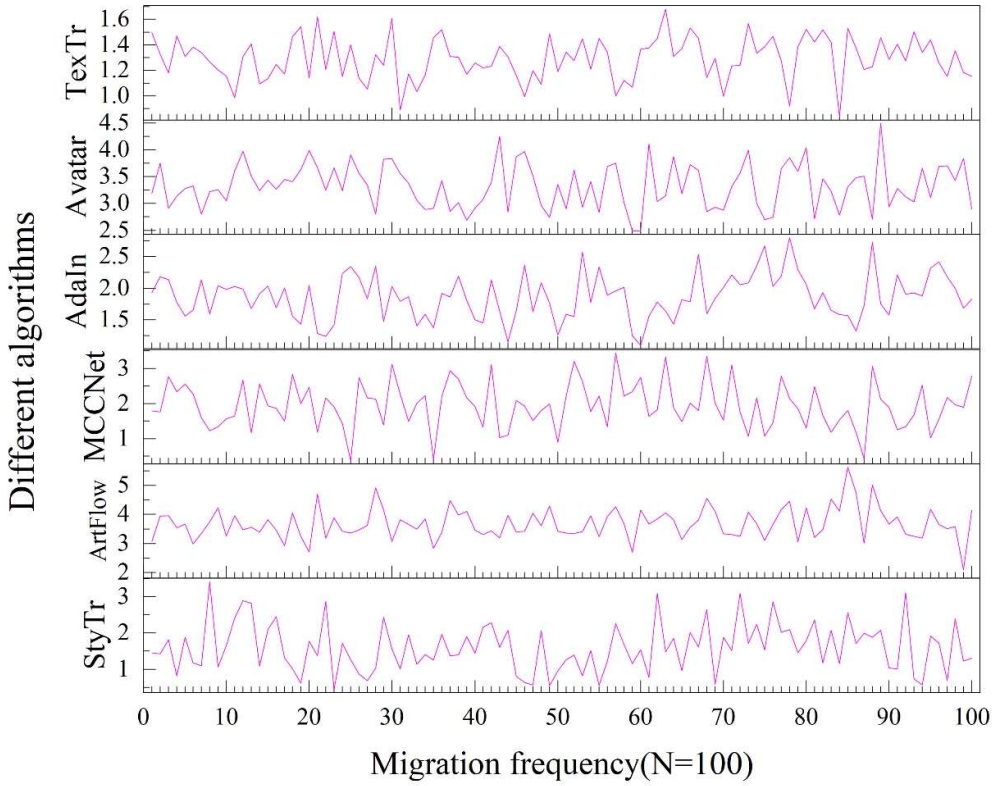

Fig. 5. The same environment and parameter lower wind loss comparison results

The results of content loss comparison under the same environment and parameters are shown in Fig. 6. The content loss values for each algorithm are 2.01, 3.42, 3.08, 3.29, 2.27, and 1.96, respectively.

CNN-based methods (e.g., Ada In, Avatar, MCCNet, etc.) may cause problems such as missing content image semantics and difficulty in maintaining the overall image information because of the spatial invariance of convolutional neural networks, resulting in the content loss of convolutional neural networks are higher than the other methods.

ArtFlow is an engineered flow network based method that minimizes the image reconstruction error and recovery bias, making its content loss better than other convolutional neural networks. However, ArtFlow cannot get better style features in the style migration task due to its spatial localization, making all its style loss values higher.

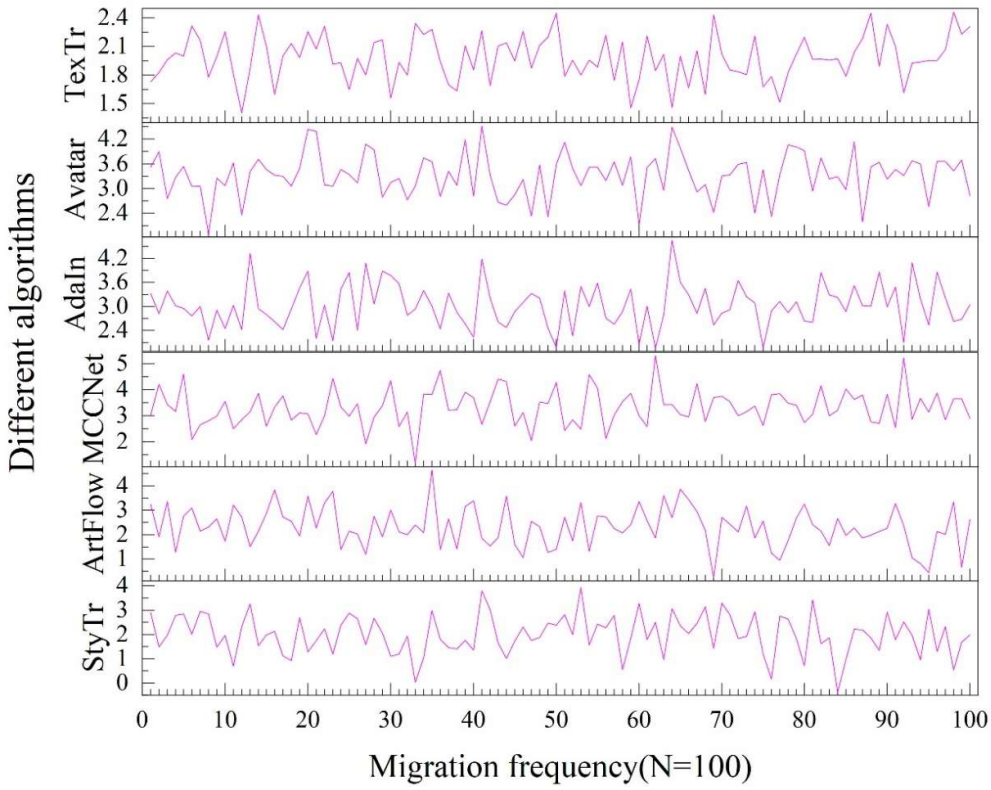

Fig. 6. The same environment, the comparison of content loss under parameters

Comparison of training time under the same environment and parameters is shown in Table 4.The training time of the TexTr method is 8.2 h. Under the condition of the same environment and parameters, the CNN method iterates the model with complex floating-point operations, and the fitting model time is longer. $StyTr^2$ The attention mechanism is computationally complex when the input image aspect is large.Some parameters of the Transformer similarly affect the training time, such as the number of heads of the multi-head attention mechanism, the number of Transformer layers, and so on.

The visual Transformer-based method is not as fast as the CNN-based method in the migration task of a single image due to its complex computation. However, the TexTr method proposed in this paper still outperforms the $StyTr^2$ method.

Table 4. The same environment, the parameter is compared

Different methods	Training hours/h	The length of the 256*256 image style migration length/s
TexTr	8.2	0.251
Avatar	14.3	0.296
AdaIn	11.9	0.143
MCCNet	13.1	0.027
ArtFlow	11.5	0.339
$StyTr^2$	14.5	0.273

4.3. AI Dong brocade pattern customization implementation

To further compare the advantages and disadvantages of this paper's method with other methods, a user study was conducted and 80 participants were invited to evaluate the results of the different methods. The other methods were Avatar, AdaIn, MCCNet, ArtFlow, and $StyTr^2$ models.

Prior to the study, participants were introduced to the purpose and details of the study. The participants were shown the results generated by the method of this paper and another randomly selected method and asked to choose under three criteria:

1) Which result has a better stylized Dong brocade pattern?
2) Which Dong brocade pattern stylization result better preserves the content structure?
3) Which Dong brocade pattern stylization result conveys stylistic information more consistently?

Each participant conducted 10 rounds of comparisons and collected 800 ballots. The statistics of user preferences for different methods are shown in Table 5. Among the comparison methods $StyTr^2$ models perform better in terms of overall quality, content preservation and style consistency, but the TexTr model proposed in this paper is above 70% in all three metrics. The results show that this paper's method outperforms the other methods in the three criteria of overall quality, content preservation and style consistency.

Table 5. User preferences statistics for different methods

Different methods	Overall	Contnet	Style
TexTr	0.7809	0.8224	0.7734
Avatar	0.5693	0.5617	0.5213
AdaIn	0.4071	0.3986	0.4339
MCCNet	0.3629	0.5244	0.4608
ArtFlow	0.5125	0.5580	0.5963
$StyTr^2$	0.6635	0.7263	0.6745

5. Conclusion

This paper summarizes the application of AI technology in the development of traditional pattern recovery and reproduction by optimizing the visual Transformer network and adding a color consistency module to meet the needs of style migration for color and texture consistency of Dong brocade patterns.

The normalization processing technique is applied to specification of the image input to the Transformer series network. The Loss, Acc and Training Time of the representative network of convolutional neural network and the representative network of Transformer model are compared respectively. the Loss, Acc and Training Time of the representative network of Transformer model, ViT, are higher than those of the representative network of convolutional neural network in the dataset, and the algorithm has the advantage, which can be used as the Dong brocade pattern color and texture information migration model with Acc=0.975.

The image texture migration model based on visual transformer has smaller style loss value and content loss value in Dong brocade pattern style migration, with style loss value=1.24 and content loss value=2.01, which can generate the resultant image closest to the Dong brocade pattern style. In the user preference survey of the generated images of Dong brocade pattern style, the visual transformer-based image texture migration model gets the best recognition in the three criteria of overall quality, content preservation and style consistency.

Funding

This work was supported by Research on educational inheritance and digital innovation design of Dong brocade in Hunan (Subject Number: 22YBA271).

References

[1] Yan, J., Mayusoh, C., Inkuer, A., & Puntien, P. (2024, June). Comparative study on pattern characteristics and cultural connotation of Zhuang, Miao, and Dong Ethnic Brocade in Guangxi, China. In INTERNATIONAL ACADEMIC MULTIDISCIPLINARY RESEARCH CONFERENCE IN PARIS 2024 (pp. 296-306).

[2] Zhang, F., & KROTOVA, T. (2024). The influence of Silk Road culture on modern design: artistic features of Chinese brocade patterns. Art and Design, (1), 56-67.

[3] Wenji, Z., Rongrong, C., & Li, N. (2022). Design and Cultural Aspects of 20th Century Chinese Xiangjin Brocade. Fibres & Textiles in Eastern Europe, (3 (151), 116-129.

[4] Feng, X., Wang, F., & Heffernan, K. (2024). A Preliminary Study of Chinese Brocade Names. Names, 72(3), 1-13.

[5] Cheng, X., Yang, J., Jiang, L., & Hu, A. (2019). Interpreting and semantically describing Chinese traditional brocade: Xilankapu. The Electronic Library, 37(3), 474-489.

[6] Gupta, M., & Arora, V. (2017). Ashavali brocades from traditional to modern times. International Journal of Home Science, 3(2), 353-358.

[7] Qin, Z., Ji, T., & Cao, Y. (2020, December). Understanding, Cooperation, Recreation and Application—A Teaching Practice towards the Design of Ethnic Patterns. In 2020 6th International Conference on Social Science and Higher Education (ICSSHE 2020) (pp. 239-245). Atlantis Press.

[8] Liu, Q., & Sirisuk, M. (2024). Dong brocade in Hunan, China: Literacy and re-invention of tradition in the perspective of intangible cultural heritage protection. International Journal of Education and Literacy Studies, 12(3), 65-73.

[9] Jiang, Y., & Yang, Y. (2021). Endogenous Power of Rural Economic Development in China: A Case Study of Cultural Capital of Dong Brocade. Contemporary Chinese Political Economy and Strategic Relations, 7(3), 1809-1840.

[10] YUAN, X., & CHUPRINA, N. (2024). ECOLOGICAL DESIGN CONCEPT OF TEXTILE PRODUCTS: A STUDY ON BROCADE WEAVING MATERIALS OF THE CHINESE ETHNIC MINORITY YAO. Art and Design, (4), 51-61.

[11] Hu, Y. (2023). Application research of artificial intelligence in the innovation of Zhuang brocade digital art: taking "the speech of Zhuang brocade" as an example. Journal of Computer and Communications, 11(5), 42-51.

[12] Fan, Z., Zhou, B., Zhang, S., & Liu, M. (2017, July). Computer Simulation of Brocade Based on Modeling Structures. In 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER) (pp. 990-995). IEEE.

[13] Lin, W., Mengyue, M., Wang, J., Li, M., & Chen, J. (2022). Digitizing and recovering the knowledge of traditional Chinese colour of the Nanjing brocade. SCIRES-IT-SCIentific RESearch and Information Technology, 12(1), 95-110.

[14] Feng, H., Sheng, X., Zhang, L., Liu, Y., & Gu, B. (2024). Color Analysis of Brocade from the 4th to 8th Centuries Driven by Image-Based Matching Network Modeling. Applied Sciences, 14(17), 7802.

[15] Xiaoling, C. H. E. N., Xiaoqin, P. E. N. G., & Xiandong, X. I. E. (2021). Dong brocade pattern artistic features and innovative application in modern apparel. Wool Textile Journal, 49(1).

[16] Nhan, H. T. T. Challenges in Integrating Traditional Brocade into Modern Design: An Interdisciplinary Approach to Sustainable Preservation and Innovation in Vietnam. INTERNATIONAL JOURNAL OF ASIAN CULTURE, 36.

[17] Leonardo Fontoura,Daniel Luiz de Mattos Nascimento,Julio Vieira Neto & Rodrigo Goyannes Gusmão Caiado. (2025). Energy Gen-AI technology framework: A perspective of energy efficiency and business ethics in operation management. Technology in Society,81,102847-102847.

[18] Huiyao Dong,Igor Kotenko & Shimin Dong. (2025). A lightweight vision transformer with weighted global average pooling: implications for IoMT applications. Complex & Intelligent Systems,11(5),215-215.

[19] Folasade Olubusola Isinkaye,Michael Olusoji Olusanya & Ayobami Andronicus Akinyelu. (2025). A multi-class hybrid variational autoencoder and vision transformer model for enhanced plant disease identification. Intelligent Systems with Applications,26,200490-200490.