International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-19

https://doi.org/10.70917/ijcisim-2026-0022
Article

Discussion on the Optimization Method of Modern Ink Figure Painting Composition Structure Based on Geometric Algorithm

Momo Feng and Ying Bai *

Shanghai Academy of Fine Arts, University of Shanghai, Shanghai, 200072, China; FengDKfdk@163.com

Abstract: Composition is an important factor in determining the aesthetic quality of modern ink figure painting, and the existing optimization algorithms in this field still have problems such as insufficient accuracy and rotational consistency, this paper proposes a composition optimization algorithm based on geometric algorithm. First of all, the color variety of the subject area and the color variety of the background area are chunked, and the color harmony feature is extracted by partition. Then, using the rule of thirds and the principle of visual balance as the principle, a deep convolutional neural network with VGG-16 as the model backbone is used to detect the saliency of the modern ink figure painting. In this paper, the accuracy of the composition structure optimization enrichment of modern ink figure paintings stays above 75%, and the attention and sensitivity to the three influencing factors of composition, ink color and texture, which are more important in the composition of modern ink figure painting style and aesthetics, are better, which not only provides a guarantee for maintaining the original painting style and aesthetics of the drafts after the optimization of the composition structure of the modern ink figure paintings, but also provides a theoretical basis for the modernization and transformation of the traditional paintings.

Keywords: color harmony characteristics; rule of thirds; VGG-16; modern ink figure painting composition

1. Introduction

Chinese ink painting is a unique form of artistic expression in China, distinguished by its distinctive artistic characteristics and cultural spirit [1-2]. In Chinese culture, the form, meaning, and aesthetic pursuit of ink painting reflect the uniqueness and profound historical heritage of traditional Chinese culture [3-4].

Since the last century, the continuous collision of Chinese and Western cultures has led to the diversification of ink painting figure painting. Modern ink figure painting is a product of the combination of traditional ink painting language and Western painting language. It is different from the traditional ink figure painting's “similar but not identical” imagery and also different from the rigorous scientific modeling of Western painting [5-6]. Modern Ink figure painting has become a growing trend in contemporary society and holds significant potential for further development. First and foremost, there have been significant changes in composition. While the landscapes in reality are infinite, the canvas is limited. Within this limited space, artists must first establish the composition to express their artistic objectives and ideas. Composition is the first step and plays a crucial role not only in traditional painting but also in contemporary composition [7-8]. Composition holds great importance in contemporary ink painting, particularly in large-scale creations where it takes on innovative forms. Ink painting places a strong emphasis on composition, focusing on the arrangement of visual elements. Artists rearrange compositions based on their subjective perceptions in a vibrant manner to achieve the desired artistic ambiance [9]. Therefore, the form of composition plays a significant role in modern ink figure painting, allowing artists to experiment with new compositional approaches without constraints during the painting process. This has elevated aesthetic standards and driven societal progress.

Reference [10] uses directional gradient histograms to capture the edges and texture features of figure

paintings, and combines a smart seagull optimization algorithm with an adaptive recurrent neural network to predict and analyze the composition of modern ink figure paintings, achieving high-precision reproduction of the composition. Reference [11] uses deep neural networks to extract low-level features from painting images, calculates their variance to form a two-dimensional display, and completes the composition of the artwork image. Literature [12] confirms that the spatial composition of traditional paintings—the shape of the frame—influences viewers' aesthetic preferences, with viewers preferring to place two figures at a certain horizontal distance in the composition. Literature [13] analyzes the compositional structure of the painting "The Creation and the Exodus from the Garden of Eden," using both intangible and tangible circular geometric structures to form the completed work under precise scaling proportions. It is evident that geometric structures remain firmly established as the foundation of painting composition. However, in contemporary ink figure paintings, the geometric forms of compositional structures lack a sense of visual impact, and the structural layers are insufficiently defined. Geometric algorithms, a computational method based on geometric concepts, utilize the properties and transformations of geometric shapes to provide efficient solutions for various practical problems. They represent one strategy for optimizing painting compositional structures [14].

In this study, the subject of modern ink figure painting is extracted and optimized for composition, and the composition of figure painting is classified using a structure optimization network for modern ink figure painting composition consisting of three modules: feature extraction module, saliency regression module and image size recovery module. Combining the rule of thirds and the principle of visual balance, the VGG-16 network is utilized to realize the saliency monitoring of modern ink figure paintings. A series of experiments, such as comparative analysis and sensitivity analysis, are designed to verify the excellent performance of this paper's model in the optimization of the composition layout of modern ink figure paintings and the sensitivity of disturbing factors, and to test the effective balance between the need to maintain the artistic style of ink figure paintings and the optimization of composition structure.

2. Optimization of Compositional Structure of Modern Ink Figure Painting Based on Geometric Algorithm

2.1. Image Segmentation and Image Synthesis

2.1.1. Characteristics of Image Search Use

Color, in keyword-based image recommendation, is the first semantic information to be introduced. After studying the results of text segmentation, it is found that the adjectives in the color category are all relatively simple. This requires a method to constrain and simplify multiple colors in a single image so that the corresponding image of the same or similar color can be searched by simple color adjectives.

In the search process, we use a color feature called primary color. Given a picture, the color histogram of the picture is calculated, and then a set of constraints and simplification algorithms are established by analyzing the information of the color histogram, and then an RGB value is obtained to replace the color histogram of the whole picture as the main color, which describes the color characteristics of the whole picture.

Another thing that needs to be considered is the aggregation degree of the image color. It is empirically found that for a given keyword, it is easier to select an image with higher color concentration for subsequent operations such as object segmentation extraction.

Saliency, again, is semantic information that needs to be considered in keyword-based image recommendation. Compared with color, saliency information cannot be obtained directly from textual information, but for a given keyword, the degree of saliency of the keyword object in the picture of the picture recommendation result is very important for the user to be able to quickly select a satisfactory keyword object.

The saliency ratio of the subject object needs to be tested for the saliency of the candidate image first, and then based on the results of the saliency detection, the ratio of the saliency region of the candidate image in the overall image is calculated.

After solving for the color and saliency features of the image, we can calculate the score of each image based on these feature values and the corresponding weight, the formula is shown below:

$$S = \alpha * MC + \beta * Sa + \gamma CA \tag{1}$$

where MC denotes the main color of the image, Sa denotes the image saliency eigenvalue, CA denotes the image color aggregation, α, β, and γ are the weights of the main color of the image, the saliency, and the color aggregation, respectively, and it is permissible to allow one of them to be zero and satisfy that $\alpha + \beta + \gamma = 1$ holds.

2.1.2. Image Segmentation and Image Synthesis Methods

Image segmentation [15] is a very common and widely used image processing technique. Firstly, according to the part of the image that can be clearly distinguished as belonging to the object or the background, the color distribution of the object and the background in the image is established, that is, the color model of the object and the background, and then for the part of the image that is not labeled as the object or the background, it is decided whether each pixel belongs to the object or the background. or background.

Usually, to determine whether a pixel belongs to the object or the background, the following two factors need to be taken into account:

1) The color of the pixel itself. If its color is closer to the object's color relative to the background color, then it is more likely to belong to the object part, and vice versa.

2) Which part its neighboring pixels belong to.

In our experiments, we define "adjacent" as the 4 pixels above, below, left and right of a pixel, not the 8 pixels around it. If all the pixels adjacent to the pixel belong to the object, then it is more likely that it belongs to the object than to the background, and the closer its color is to the color of the pixels adjacent to it, the more likely it is, and vice versa. Considering these two factors together, the energy function is built:

$$E(X)=\lambda_1\sum\nolimits_{p\in P}E_1(p,x_p)+\lambda_2\sum\nolimits_{<p,q>\in N}E_2(p,q,x_p,x_q) \tag{2}$$

where P represents all pixels in the image, p denotes a particular pixel point, N is the set of all pairs of neighboring pixels in the image, and $<p,q>\in N$ denotes the number of pixels p, q in the pixels that are adjacent to each other in the image. X is a kind of segmentation result of the image, i.e., it has been determined whether each pixel belongs to the background or to an object, and x_p is the sign of the pixel p, which can take only two values:

$$x_p=\begin{cases}0 & P \text{ belongs to the object in the segmentation result}\\ 1 & P \text{ belongs to the background in the segmentation result}\end{cases} \tag{3}$$

In the above formula, E_1 represents the role that the color of the pixel itself plays in determining whether a pixel belongs to an object or to the background, reflecting the first factor in determining the attribution of a pixel. The E_2 represents the role that a pixel's neighboring pixels play in determining the pixel's attribution, reflecting the second factor. λ_1 and λ_2 denote the weights of E_1 and E_2 respectively.

Firstly, the color model of the background and object needs to be built based on the already divided background part and object part. Divide the pixels in the object into m blocks, and for each segmented pixel block, calculate the mean of all the pixel RGB vectors in it and their covariance matrices with the following formula:

$$\mu_i=\frac{1}{l}\sum\nolimits_{p\in p_i}C_p \tag{4}$$

$$\sigma_i=\frac{1}{l}\sum\nolimits_{p\in p_i}(C_p-\mu_i)(C_p-\mu_i)^T \tag{5}$$

where C_p denotes the RGB column vector of pixel p. Thus, the color model of the object is obtained: m ternary normal distribution function $\left\{N(\mu_1,\sigma_1),N(\mu_2,\sigma_2),\ldots,N(\mu_m,\sigma_m)\right\}$.

The same method can be used to obtain the color model of the background: n a ternary normal distribution function $\left\{N(\mu_1^*,\sigma_1^*),N(\mu_2^*,\sigma_2^*),\ldots,N(\mu_n^*,\sigma_n^*)\right\}$.

Build the energy function. For those pixels p that need to be judged as belonging to the object or the background, the probability that its RGB vector is in the i th normal distribution of the object color model is:

$$p = \frac{1}{(2\pi)^{\frac{3}{2}} \det(\sigma_i)^{\frac{1}{2}}} e^{-\frac{1}{2}(C_p-\mu_i)^T \sigma_i^{-1}(C_p-\mu_i)} \tag{6}$$

The largest of these m probabilities is chosen as the degree of similarity between the color of the pixel p and the color of the object, denoted as:

$$K_p^F = \max_{1\le i\le m} \frac{1}{(2\pi)^{\frac{3}{2}} \det(\sigma_i)^{\frac{1}{2}}} e^{-\frac{1}{2}(C_p-\mu_i)^T \sigma_i^{-1}(C_p-\mu_i)} \tag{7}$$

Similarly, the color of pixel p is similar to the background color:

$$K_p^B = \max_{1\le i\le n} \frac{1}{(2\pi)^{\frac{3}{2}} \det(\sigma_i)^{\frac{1}{2}}} e^{-\frac{1}{2}(C_p-\mu_i)^T \sigma_i^{-1}(C_p-\mu_i)} \tag{8}$$

From each of the above formulas and the meaning indicated by E_1, we define E_1 as follows:

$$\begin{aligned} E_1(p, x_p = 1) &= \frac{\log K_p^F}{\log K_p^F + \log K_p^B} \\ E_1(p, x_p = 0) &= \frac{\log K_p^B}{\log K_p^F + \log K_p^B} \end{aligned} \tag{9}$$

From the above equation, it can be seen that for the pixel p to be determined if, the closer its color is to the color of the object, the larger the value of K_p^F will be, and thus the larger the value of $\log K_p^F$ will be, and then the result computed according to the above equation $E_1(p, x_p = 1)$ will be smaller, i.e., for E_1, the closer the color of the pixel is to the color of the object, $E_1(p, x_p = 1)$ will have a smaller value, and vice versa.

The E_2 represents the role that a pixel, and the pixels adjacent to it, play in determining the attribution of that pixel, reflecting the second factor. In general, the more similar the color of a pixel and its neighboring pixels are, the more likely they are to both belong to the object or to the background at the same time, and thus the less likely they are to be separated in the segmentation result, so if this is the case where $x_p \neq x_q$ then the larger the value of E_2 value should be larger and vice versa. Thus, E_2 is defined as:

$$E_2 = \begin{cases} 0 & x_p = x_q \\ \left|x_p - x_q\right| * g(C_p, q) & x_p \neq x_q \end{cases} \tag{10}$$

Based on each of the above formulas, the image segmentation result can be solved when the energy function takes the minimum value.

Image synthesis refers to the process of forming a new image by superimposing or combining two or more images, Poisson synthesis method is to synthesize the target object and the background by using Poisson equation, which can make the synthesized image look more real to a certain extent, and this paper adopts Poisson synthesis method to realize image synthesis.

2.2. Color Harmony Feature Extraction

Color harmony [16] is one of the most important features that determine the aesthetic quality of an image. Currently, the more classical color harmony models are: the Matsuda color harmony template and the Moon-Spencer [17] color harmony model. The Matsuda color harmony template only considers hue and does not consider saturation and brightness. When selecting a template, it is more difficult to determine the applicable template because an image may conform to more than one template, while the

Moon-Spencer color-toning model overcomes this drawback. It measures the harmony by calculating the relative values of the main color and other colors on Munsell color space, and when the hue values fall in the same, similar or contrasting region, the two hues are considered to be in harmony. When the hue value falls in the fuzzy region, the two hues are considered to be discordant.

There are more color harmony models used for image aesthetics evaluation, but they do not work well for images with complex colors, and are only applicable to cases where the color combinations are relatively simple, thus giving rise to the idea of chunking, where an image is regarded as a collection of multiple simple color blocks. However, this method does not take into account the difference between the color variety of the subject area and the background area. To compensate for this shortcoming, this paper proposes an improved method for extracting color harmony features. For an image, the color variety of the subject region is often more than that of the background region. Therefore, this paper treats them separately and chunks the main region and the background region respectively.

The color harmony feature extraction process is shown in Figure 1.

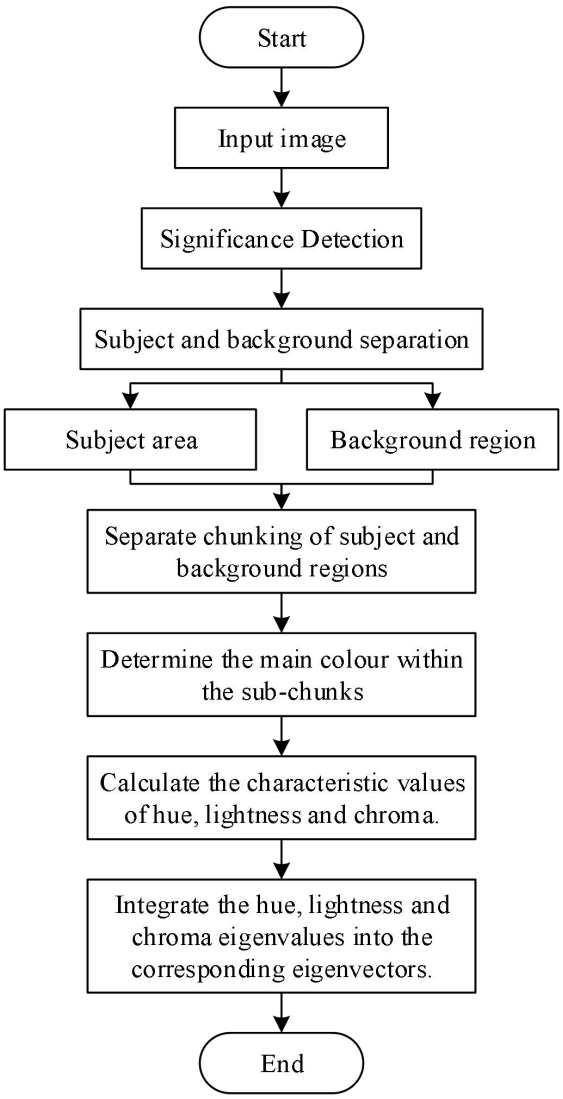

Figure 1. Color harmony characteristic extraction process.

First, the hue value of the main color is subtracted from the hue value of each pixel in the image sub-block to obtain H_s . Then, the absolute value of H_s is determined to be harmonious or not according to the Moon-Spencer color harmony theory, and the symbolic function sgn for determining whether the hue is harmonious or not is shown in Equation (11):

$$\mathrm{sgn}(H_s)=\begin{cases}1 & (7<|H_s|<12)\|(28<|H_s|<50)\||H_s|<0.05\\ 0 & Other\end{cases} \tag{11}$$

where 1 stands for harmony and 0 for dissonance. The sgn is just a flag, when H_s satisfies $(7<|H_s|<12)\|(28<|H_s|<50)\|(|H_s|<0.05)$ when it represents that the hue value of this pixel is in harmony with the hue value of the main color, otherwise, it is considered discordant.

Finally, the number of pixels within the sub-block that are in harmony with the main color is counted and the ratio of the number of harmonized pixels to the total number of pixels in the current sub-block is used as the hue eigenvalue of this sub-block as shown in Equation (12):

$$h_i=\frac{np\_h_i}{NP_i}, i=1,\cdots,z \tag{12}$$

where np_h_i is the number of pixel points in the i th image sub-block with harmonious color tone, NP_i is the total number of pixel points in the sub-block, and z is the total number of chunks in the subject area or background area.

To determine whether the brightness and color harmony of the sub-block is shown in Equation (13) and Equation (14):

$$\mathrm{sgn}(VC)=\begin{cases}1 & (fir\_R>1\&\&\mathrm{sec}\_R<1)\|thr\_R>1\|fir\_R<0.05\\ 0 & Other\end{cases} \tag{13}$$

$$\begin{cases}fir\_R=\dfrac{(oth\_C-avg\_CMain)^2}{3^2}+\dfrac{(oth\_V-avg\_VMain)^2}{0.5^2}\\ \mathrm{sec}\_R=\dfrac{(oth\_C-avg\_CMain)^2}{5^2}+\dfrac{(oth\_V-avg\_VMain)^2}{1.5^2}\\ thr\_R=\dfrac{(oth\_C-avg\_CMain)^2}{7^2}+\dfrac{(oth\_V-avg\_VMain)^2}{2.5^2}\end{cases} \tag{14}$$

where avg_CMain and avg_VMain are the average value of the chroma and the average value of the luminance of the main color area within the sub-block, respectively, and oth_C and oth_V are the values of the chroma and luminance of other pixels in the sub-block excluding the main color area, respectively. The luminance and colorimetric eigenvalues are counted according to Equation (15):

$$vc_i=\frac{np\_vc_i}{NP_i}, i=1,\cdots,z \tag{15}$$

where np_vc_i is the number of pixel points with harmonious luminance and chroma within the i th image sub-block.

(5) Integrate the hue, luminance and colorimetric eigenvalues into the corresponding eigenvectors. After calculating the hue eigenvalues, luminance and chroma eigenvalues of all sub-blocks, normalize the obtained values to the $[0,1]$ interval respectively, and divide the interval into R equal parts, and according to the number of the hue eigenvalues, luminance and chroma eigenvalues that fall in each sub-interval, obtain the hue eigenvectors and the luminance and chroma eigenvectors as shown in Eqn. (16), Eqn. (17) are shown:

$$F_h=\{Num\_h_1,\cdots,Num\_h_i,\cdots,Num\_h_R\} \tag{16}$$

$$F_{vc}=\left\{Num\_vc_1,\cdots,Num\_vc_i,\cdots,Num\_vc_R\right\} \tag{17}$$

where F_b is the eigenvector of hue, F_{cu} is the eigenvector of luminance and chroma, Num_h_i and Num_vc_i are the number of i copies of the hue value, luminance, and chroma value, respectively, and R is taken as 10.

2.3. Optimization of the Compositional Structure of Ink Figure Painting

2.3.1. Principles of Composition Optimization Based on Geometric Algorithms

Composition refers to the organization of the objects to be represented in an appropriate form according to the elements such as the subject matter of the image and the main body of the picture, through a certain relationship, to form a coordinated whole. Composition is widely used in the fields of photography, painting, design and aesthetic evaluation of images. In photography, in order to pursue the aesthetic effect, you need to follow a certain composition method. There are more than ten commonly used classical composition methods, including symmetry, framing, center composition, triangle composition, leading line composition, diagonal composition, golden spiral composition and so on. These methods are complicated and lined up, which brings some troubles to the automatic optimization of image aesthetics using computer algorithms. Therefore, in this paper, based on Occam's razor law, the rule of thirds and the principle of visual balance are selected for composition optimization.

The rule of thirds utilizes the golden section ratio to set the length of a straight line segment as $h\left(h=h_1+h_2\right)$, when the ratio relationship that satisfies $h_1/h_2=h_2/\left(h_1+h_2\right)$ The best visual balance can be obtained when the proportionality relationship of] is used, so that the image screen can achieve a more organized and stable state. By utilizing this relationship, the image screen can be divided into 9 regions, forming a network layout of 3×3 and 4 intersections of the division lines called anchor points. When composing a picture, the aesthetic effect of the image can be enhanced by placing the prominent subject to be emphasized at the anchor point position.

2.3.2. Composition Optimization Methods

In order to extract the subject of the image and optimize the composition, this paper is based on deep convolutional neural network architecture for saliency detection of the image. The network in this paper uses VGG-16 [18] as the backbone architecture of the model. The network consists of three modules, which are feature extraction module, saliency regression module and image size recovery module. After training, the network can achieve full-resolution saliency regression end-to-end without prior knowledge of the scene in question.

The feature extraction module contains 5 sets of convolutions that use a hierarchical architecture to extract semantic feature information of an image. Specifically, these 5 groups of convolutions contain 2, 2, 3, 3, and 3 convolutional layers, respectively, and the convolutional kernels are all 3×3 in size. In this paper, the ReLU rectified linear activation unit is used as the activation function instead of the traditional Sigmoid. Compared with the traditional activation function, ReLU can make the network converge faster. Meanwhile, in order to retain more edge information and expand the sensory field of the network, this paper modifies the convolution size of the maximum pooling layer from 2×2 to 3×3. In terms of iteration step size, the first 3 groups are set to 2 and the subsequent ones to 1. The output of this module is a feature mapping of 1/8 size of the input image.

The significance regression module contains 3 sets of convolutions, each with a full convolutional layer followed by a ReLU activation function layer and a Dropout layer. This module obtains the saliency score of each pixel from the regression of the forward input feature mapping. Since deep learning requires a large number of training samples to obtain satisfactory results, and the current data used for the saliency task is relatively small. In order to better train the network, this paper uses two loss functions weighted and averaged by the number of pixels, in the form of:

$$\mathrm{L}=\frac{1}{N}\left[\mathrm{L}_g(I_m,J_m,\varphi)+\lambda\times\mathrm{L}_r(I_m,J_m,\varphi)\right] \tag{18}$$

where I_m is the input image of the network. J_m is the target salient mapping image corresponding to the output. φ is the network parameters. λ is the network hyperparameter, which is set to 0.5 in this paper. N is the number of pixels in the whole image input to the network. The loss function consists of

two terms, where L_g is the global loss function, which represents the basic loss of all pixels of the image. L_r is the salient region loss function, which represents the additional loss of pixels belonging to the salient objects. The pixel number value averaging operation allows the loss function to be independent of the input image size. Specifically, L_g is denoted as:

$$\mathrm{L}_g(I_m, J_m, \varphi) = \sum_{i=1}^{N} \psi\left(F(I_{mi}, \varphi) - J_{mi}\right) \tag{19}$$

where I_{mi} denotes a pixel in the input image. J_{mi} denotes the pixel in the target significantly mapped image corresponding to the output. $F(\cdot)$ denotes the abstraction process of the network, and $F(I_{mi}, \varphi)$ denotes the significance regression score value of pixel I_{mi}. $\psi(\cdot)$ is the robust loss function. L_r is denoted as:

$$\mathrm{L}_r(I_m, J_m, \varphi) = \frac{N^-}{N} \sum_{j=i}^{N^+} \psi\left(F(I_{mj}, \varphi) - J_{mj}\right) \tag{20}$$

where N^+ is the number of pixels occupied by significant objects and N^- is the number of pixels occupied by non-significant objects. $\psi(\cdot)$ is defined as:

$$\psi(x) = \begin{cases} \frac{1}{2}x^2 & if\ |x| \le 1 \\ |x| - \frac{1}{2} & otherwise \end{cases} \tag{21}$$

It follows that $\psi(x)$ is derivable and its derivative is:

$$\psi'(x) = \begin{cases} x & if\ |x| \le 1 \\ sign(x) & otherwise \end{cases} \tag{22}$$

Compared with the traditional ℓ_1 loss function, the loss function in this paper is continuously derivable at the zero point, and thus it allows the network to converge more stably near the zero point. During the training process, this paper utilizes the standard stochastic gradient descent method to minimize the loss function.

The image size recovery module consists of a size recovery layer, which restores the forward input to the original input size and finally outputs a full resolution salient image. In this paper, the network is trained on the MSRA10K dataset. Compared with the traditional saliency detection methods based on the underlying pixel information and utilizing a priori features, the network model in this paper can better extract the high-level semantic information of the image, and the results are more relevant to the human visual perception.

After obtaining the salient feature image, this paper calculates the weighted average of the image pixels based on the pixel values from 0 to 255 to further obtain the pixel center of gravity point. Subsequently, the algorithm matches the pixel center of gravity with the anchor position in the rule of thirds based on the principle of visual balance, finds the anchor point at the closest Euclidean distance from the center of gravity, and moves the image along with the pixel center of gravity of the salient image to the position of this anchor point, which then crops the redundant part of the image to realize the reconstruction of the image that conforms to the aesthetics of the image and visual balance.

3. Experiment on Optimizing the Compositional Structure of Modern Ink Figure Painting

3.1. Experimental Environment and Evaluation Criteria

The experimental environment involved in composition prediction is as follows: the operating system

is ubuntu 16.04, the graphics card is a Tesla V100-DGXS, and the deep learning framework is PyTorch. To train the model we use the ResNet50-blurpooling pre-trained model (intercepted to the last convolutional layer) trained on ImageNet for fine-tuning. The dataset is KU_CPC with 3000 images in the training set and 1000 images in the test set. A total of 200 epochs are trained, and the optimizer uses Stochastic Gradient Descent (SGD) with a momentum of 0.8, an initial learning rate of 0.002, a weight decay of 0.0002, and a Batchsize set to 128.

Accuracy discrimination: because of the existence of images with the specificity of multiple compositions, the accuracy of the composition model is discriminated as follows, for images with multiple truth value labels, we randomly select one of the truth values as the training label at each epoch of training, and because of a total of 200 epochs of training, the method allows multiple truth values of the image to participate in the training. During testing, when the prediction result satisfies one of the multiple truth values, then we consider the classification is correct. The accuracy is defined as follows:

$$Accuracy = \frac{N_{correct}}{N_{total}} \tag{23}$$

Rotation consistency discrimination: in addition, we propose a criterion for evaluating the rotational consistency of the model by doing two rotations of the image at a random angle $\theta\left(-8^{\circ} \le \theta \le 8^{\circ}\right)$ to discriminate whether the two skewed images of their evaluation results are consistent with the original image or not, and the formula is defined as follows:

$$\tilde{\alpha}_{X,\theta_1,\theta_2} = \begin{cases} \arg\max P\left(Rotate_{\theta_1}(X)\right) = \arg\max P\left(Rotate_{\theta_2}(X)\right) \\ = \arg\max P(X) \end{cases} \tag{24}$$

where $\tilde{\alpha}_{X,\theta_1,\theta_2}$ is the frequency of the model outputting the same compositional result, θ_1, θ_2 are the angle of the two rotations respectively, $Rotate_{\theta_i}(X)$ is the image after rotation, and $\arg\max$ represents the selection of the maximum softmax probability value as the result of the composition category. In order to avoid the unfairness caused by the random angle, we conducted $n = 5$ experiments for each image finally averaging the results. The final rotation consistency rate is defined as follows:

$$Consistency = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{\alpha}_{X_i,\theta_1,\theta_2}}{n * N_{total}} \tag{25}$$

3.2. Comparison of Model Accuracy and Rotational Consistency

To validate the effectiveness of our proposed models for the compositional classification task, all models are trained in the same environment using the hyperparameters recommended in the corresponding papers, and the same dataset is used to train and validate the accuracy and rotational consistency of the models. To verify the effectiveness of our proposed model for compositional optimization, we added an adaptive pooling layer on top of the AlexNet network so that it can be trained using images of multiple scales, and the model obtained from the multi-scale training is AlexNet_multi. In addition, we added the STN structure to the Res50-blur backbone model to train to obtain the Res50-blur+STN. The experimental results are shown in Table 1, where all models except AlexNet are trained using the multi-scale training approach. The models in this paper reach the optimal accuracy and rotational consistency whether or not data enhancement is used, and the accuracy and rotational consistency are kept above 90%. It shows that the composition optimization method in this paper can adjust the image composition well, which makes the accuracy and rotation consistency of the model improved.

Table 1. Accuracy and rotation consistency.

Net	Train without augmentation		Train with augmentation	
	Accuracy (%)	Consistency (%)	Accuracy (%)	Consistency (%)
AlexNet	87.7304	69.0769	87.8229	69.1330
AlexNet_multi	88.1920	69.0214	88.1921	73.0442
Res50	90.4987	78.9855	90.6820	80.9966
Res50-blur	90.6829	83.3581	90.4055	84.7973
Res50-blur+STN	90.3141	83.1730	89.9446	84.1876
(Our)	90.8671	83.7080	90.9143	90.1481

In order to analyze the applicability of the model in this paper for the task of image composition classification in detail, we compare the prediction results of AlexNet and the model in this paper for various types of compositions, and calculate the ROC curves and AUC values for each of the five types of ink figure painting compositions: central composition, full-width composition, open-close composition, layered composition and triangular composition, where AUC value is the area under the ROC curve, and the closer the value is to 1, the better the model works. AUC value is the area under the ROC curve, and the closer the value is to 1, the better the model is. The ROC curves of each composition category are shown in Figures 2 and 3, where the horizontal axis represents the pseudo-positive rate and the vertical axis represents the true-positive rate. In order to compare the fairness of the experiments, both AlexNet and the model in this paper are trained using data enhancement.

The AUC values of this paper's model for all types of compositions are higher than those of AlexNet, and the ROC curves for all types of compositions are closer to the (0,1) point of the coordinate axis than those of AlexNet. It is worth noting that the AUC value of the model in this paper reaches 1 for the centered composition, and the AUC values for several other types of compositions are all greater than 0.9.

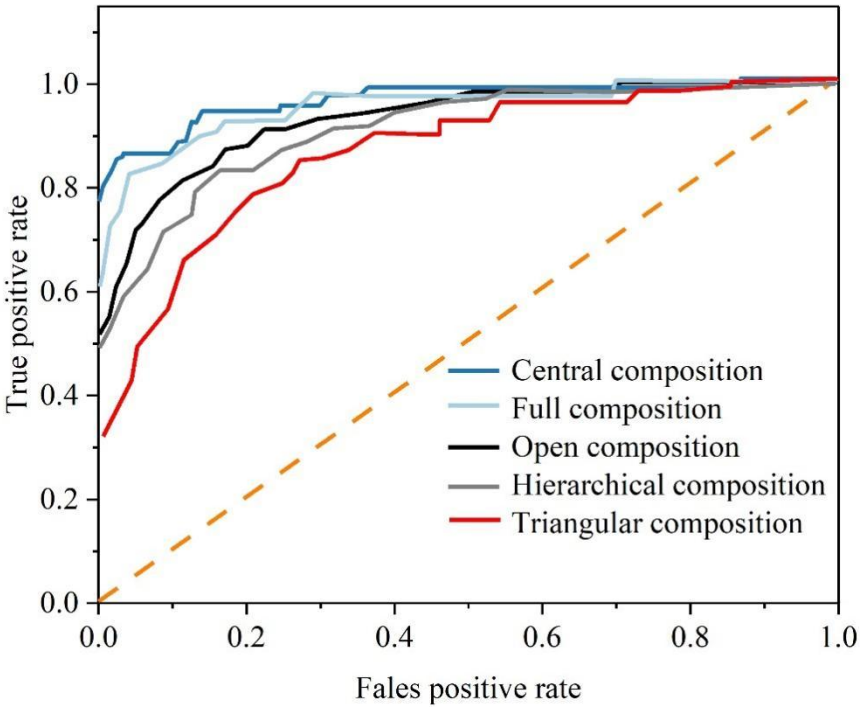

Figure 2. The corresponding curve of the AlexNet model.

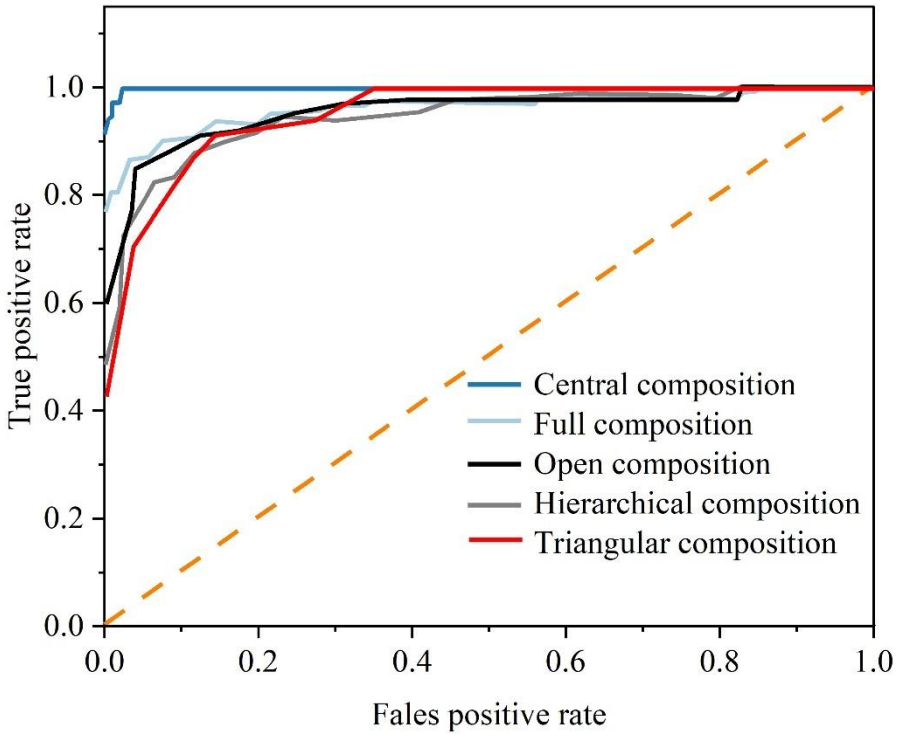

Figure 3. The corresponding curve of the model.

Table 2 demonstrates the accuracy of each composition category, and it is observed that the accuracy of this paper's model is higher than AlexNet in predicting all compositions except the trichotomous composition, and the prediction accuracies of this paper's model are above 0.9. It shows the advanced classification effect of this paper's model for each composition. It can be found that the accuracy of the model prediction is lower for hierarchical and triangular compositions, and after our observation and analysis, this is due to the fact that these two types of compositions are more subjective and have a certain degree of ambiguity in the data annotation, and at the same time are prone to exist with other compositions in the image at the same time.

Table 2. Comparison of the classification accuracy of each composition.

Net	Central composition	Full composition	Open composition	Hierarchical composition	Triangular composition
AlexNet	0.8978	0.8758	0.8645	0.7894	0.7685
Ours	0.9456	0.9345	0.9245	0.9221	0.9203

3.3. Analysis of Model Sensitivity to Rotation

To investigate the effect of skewed images on the compositional classification model, we tested the model for rotational sensitivity. We defined a set of continuous rotation angles in the range of (-15°,15°), and rotated the image continuously according to the above angles with the center as the origin, and as the image rotated, we recorded the model's prediction accuracy for each compositional image after rotating at different angles, thus reflecting the model's sensitivity to rotation, and the sensitivity to rotation of the AlexNet model and the model in this paper are shown in Figures 4-Figures 7. The sensitivity of the AlexNet model and the model in this paper to rotation is shown in Fig. 4-Fig. 7, Fig. 4 and Fig. 6 show the accuracy of the model without data enhancement, and Fig. 5 and Fig. 7 show the accuracy of the model with data enhancement.

The AlexNet model is highly sensitive to rotation, and a slight angular rotation causes the accuracy of the model to fluctuate drastically, plummeting to 20% for both vertical and horizontal compositions in the interval range of ±5 to ±15 degrees, and the accuracy of diagonal compositions also varies with successive changes in angle, proving that AlexNet is not rotationally invariant for most compositions. The AlexNet model trained with data augmentation is less sensitive to rotation, and the accuracy is stable and concentrated for all types of compositions except triangular compositions, but the sensitivity to

triangular compositions is still high, proving that the data augmentation method is effective in decreasing the rotation sensitivity of the model.

Without data enhancement, our model is significantly less sensitive to rotation than AlexNet, and the stability of the model is greatly improved, but it is still sensitive to the rotation of triangular composition images. When the model in this paper is trained with data augmentation, we can find that the accuracy is stable at more than 75% for all composition types as the angle is changed, even for triangular compositions, which are extremely sensitive to rotation operations. The model in this paper trained with data augmentation greatly suppresses the interference caused by rotation, reduces the sensitivity to rotation, and improves the spatial invariance of the model substantially. The above analysis proves that the data enhancement approach and the model in this paper have significant effects in the composition classification task, and can be fully applied to the optimization of the composition structure of ink figure paintings, while AlexNet cannot be generalized to general images.

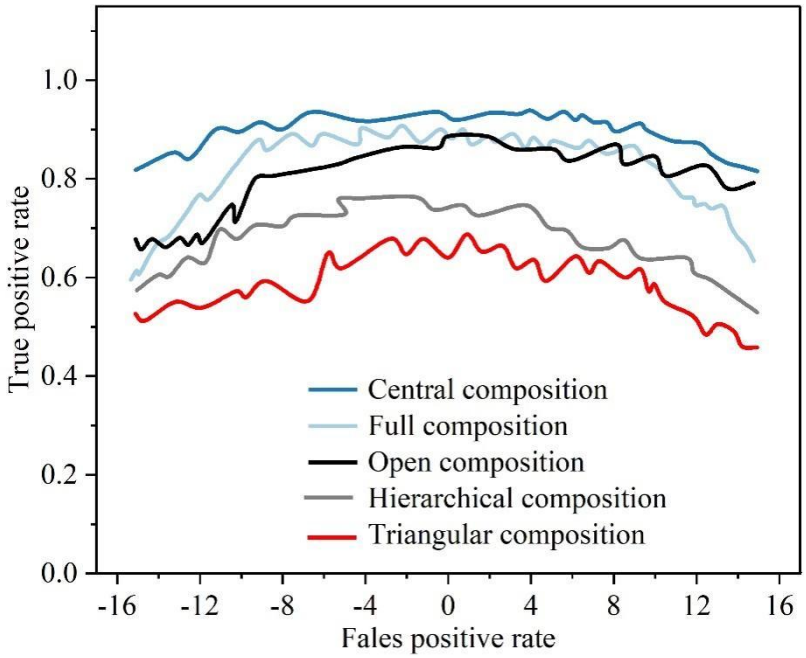

Figure 4. Alexnet model accuracy without data enhancement.

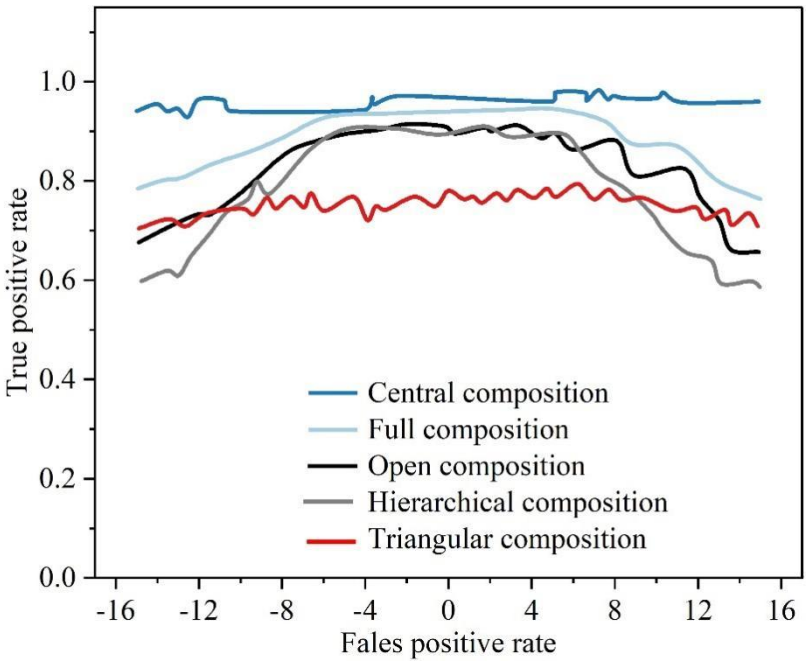

Figure 5. Using data enhanced alexnet model accuracy.

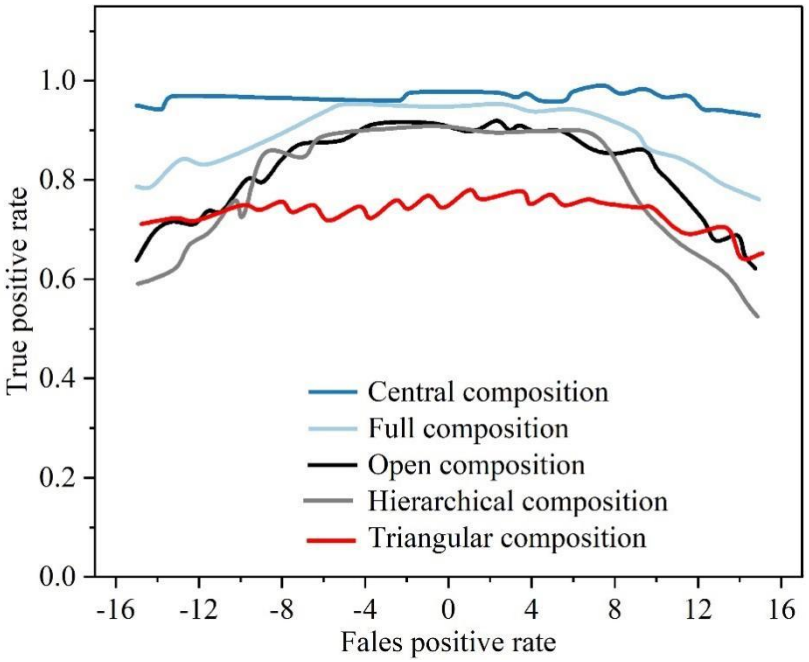

Figure 6. The accuracy of the model is not used in the paper.

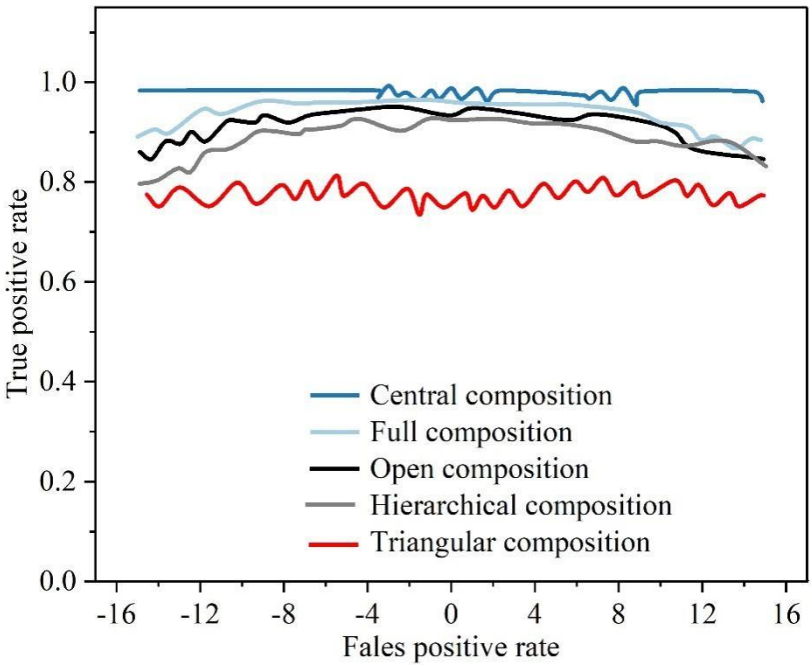

Figure 7. The accuracy of the model of this article with data enhancement.

3.4. Aesthetic Information Measurement of Image Contrast

Contrast refers to the different brightness levels between the brightest white and the darkest black in the light and dark areas of an image, i.e., the size of the gray-scale contrast of an image. A larger range of differences means a larger contrast, and a smaller range of differences means a smaller contrast. The gray level histogram is a function of the distribution of gray levels in an image. Gray scale histogram is the frequency of occurrence of all pixels in a digital image, according to the size of the gray value. The histogram equalization method in MATMLB is utilized to enhance the image contrast, the gray level

histogram of the original image and the equalized gray level histogram are shown in Fig. 8 and Fig. 9 respectively. The MATMLB custom function is further utilized to compute the image entropy for the comparison of entropy before and after image contrast enhancement.

From the experimental results, after histogram equalization, the contrast of the image is significantly enhanced, the visual effect is significantly improved and the image quality is enhanced. From the analysis of the experimental results, the range of gray value of the image histogram after histep equalization is obviously enlarged; the histogram of the image after histep equalization tends to be flat, and the gray level is reduced, which indicates that the gray level is merged; the entropy value of the image before contrast enhancement is 7.2512, and the entropy value of the image after contrast enhancement is 7.9654, which indicates that the entropy value of the image is able to reflect the difference in image contrast. In the calculation of entropy value of images with cross-reference, the entropy value shows the consistent rule of change with the aesthetic evaluation of contrast, i.e., the entropy value of contrast with better aesthetic evaluation is also higher. Therefore, image entropy can be used as a measure of universal aesthetic judgment of image contrast. It should be noted that, on the basis of obtaining a large number of image entropy, whether the entropy value and its distribution corresponding to the contrast with aesthetic universality can be inferred, so as to give the objective scale entropy in line with the universality of aesthetics, is subject to further research.

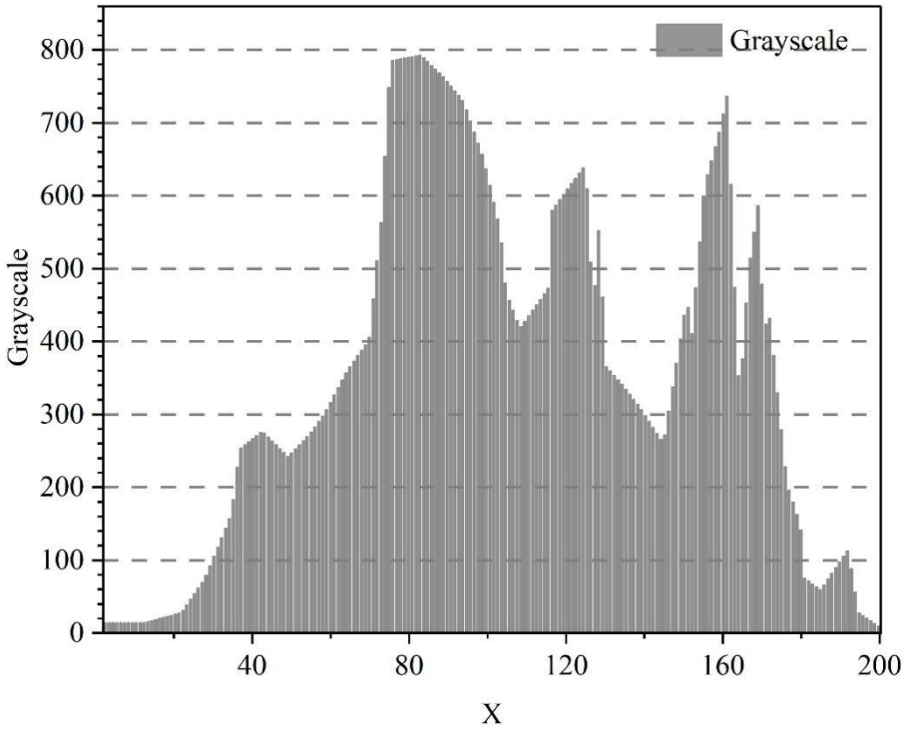

Figure 8. Gray histogram.

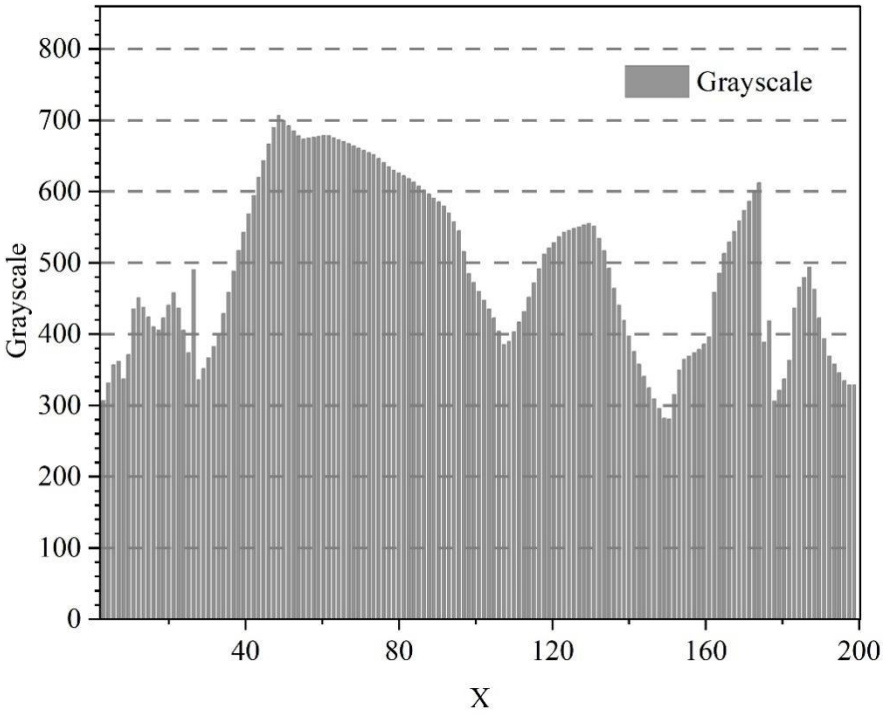

Figure 9. Equilibrium gray histogram.

3.5. Complexity and Semantic Clarity Calculations for Modern Ink Figure Drawings

3.5.1. Complexity of Ink Figure Drawing

Rich colors add to the complexity of a painting. We calculate the number of different colors used in each painting and examine the results by generating a hue histogram. First, we convert the image from RGB to the HSV color space, where H (x, y), S (x, y) and V (x, y) denote hue, saturation and luminance. We obtain the number and location of pixels with hue equal to h. Second, the saturation threshold determines whether each pixel with hue equal to h is a colored pixel or a non-colored pixel (grayscale, black, or white pixel) and eliminates the non-colored pixels. Third, we counted the number of pixels on each of the six color bands by dividing the 360-degree hue proportionally into six intervals representing red, yellow, green, blue, cyan, and magenta. If the number of pixels of a color was proportionally greater than 0.1% of the total number of pixels, the color was determined to be used in that painting. A threshold of 0.1 was determined to be the most appropriate through a series of comprehensive experiments.

In the second step, we set a saturation threshold because low-saturation colors can be represented by luminance-controlled grayscale values, while high-saturation colors can be represented by hue. The saturation threshold determines the conversion between hue and luminance. The threshold value depends on the luminance because low-luminance colors are always close to grayscale.

Figure 10 shows the hue histogram of In the Window, Out the Window after removing non-colored pixels.Hue in HSV color space is defined as an angle in the range 0 to 2π. Different angles represent different colors, and certain segments of consecutive angles in the hue bands represent this similar colors. Hues can thus consist of six color bands. The three primary colors red, green and blue, yellow, cyan and magenta. Six color bands are sufficient to describe the colors in a chosen painting.

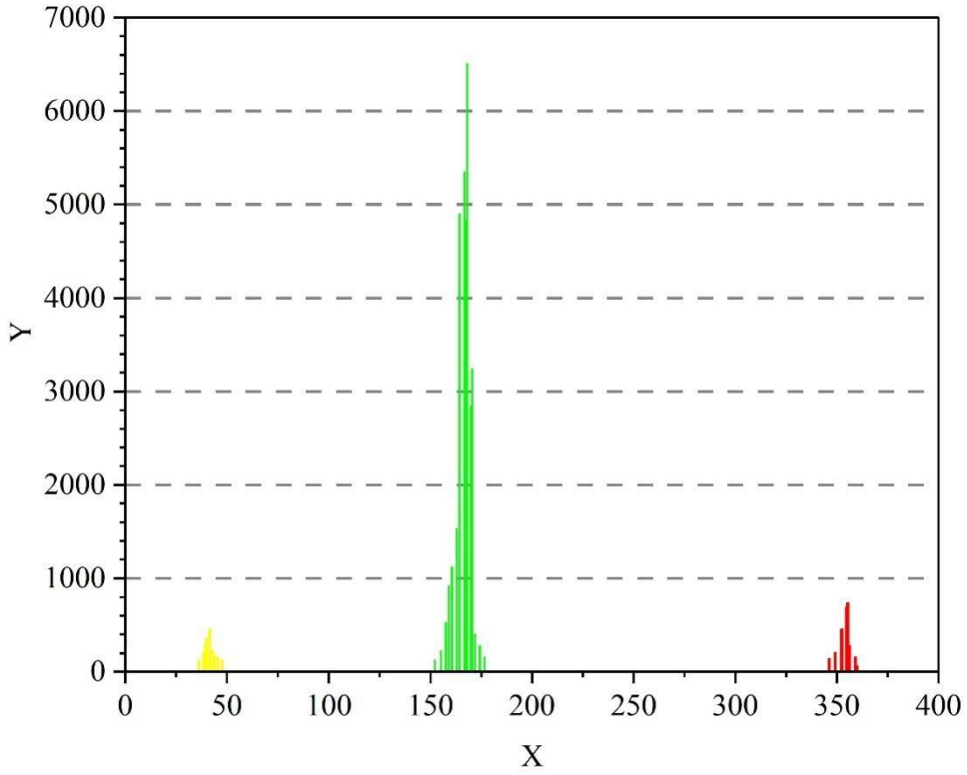

Figure 10. Remove non-colored pixels of "window, window" tonal histogram.

3.5.2. Semantic Clarity and Complexity in Modern Ink Figure Painting

We selected a sample of 20 ink and watercolor figure paintings by Wu Guanzhong, of which 15 were abstract and 5 were figurative. We invited 100 Chinese university students as volunteer participants to assess the complexity of the paintings using a 7-point Likert scale (1 = not simple, 7 = very complex), and no participant was color blind. Although some participants may have heard of Wu Guanzhong, none were professional artists or familiar with his paintings. Each participant was tested individually using a computer to view each painting and then rated for perceived complexity and semantic transparency. Their level of education and gender were also recorded. The reported data were used in the regression model we discuss below. Figure 11 shows the statistics of the average scores on visual complexity for the 20 paintings, where the yellow bars represent abstract paintings, while the green bars represent figurative paintings. All of the figurative paintings had lower complexity scores than the abstract paintings, except for this figurative painting numbered 18 which had a complexity score of only 1.65.

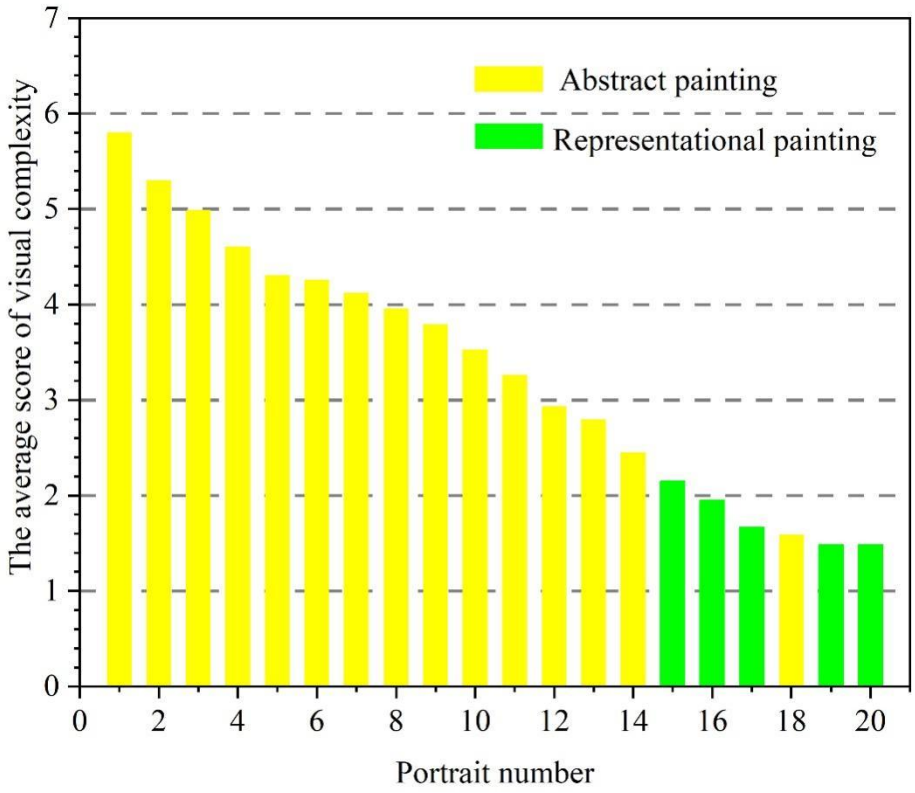

Figure 11. Visual complexity average statistics.

3.6. Compositional Optimization Model Sensitivity Analysis

In order to further reveal the interpretability of the composition optimization model, experimental interference with three important painting elements, namely composition, ink color, and texture, is carried out here, and the sensitivity of the depth model response to changes in these factors is tested, and the experimental results are shown in Fig. 12. Firstly, 120 small squares with sizes ranging from 10 × 10 to 60 × 60 pixels were selected and randomly masked at any position on the ink figure painting image, thus interfering with the overall layout, and it can be seen that the correlation coefficients between its model prediction results and the manual scores decay rapidly with the increase of the layout interference, which suggests that the composition optimization model is more sensitive to the spatial layout.

Secondly, different gray scale coefficients g are set to interfere with the overall color scale of the ink painting image, which ranges from 0 to 1. The closer to 0, the darker the image is, and the correlation coefficient of the gray scale coefficients gradually decreases with the increase of the interference of the color scale, which shows that the model is more sensitive to the hierarchical change of the ink color gray scale.

Finally, different levels of Gaussian noise are added to the ink painting image to interfere with the texture, and the larger the variance parameter s is, the rougher the image is. It can be seen that the correlation coefficient decreases with the increase of the variance and the increase of the noise, indicating that the composition optimization model is more sensitive to the texture of the brush strokes.

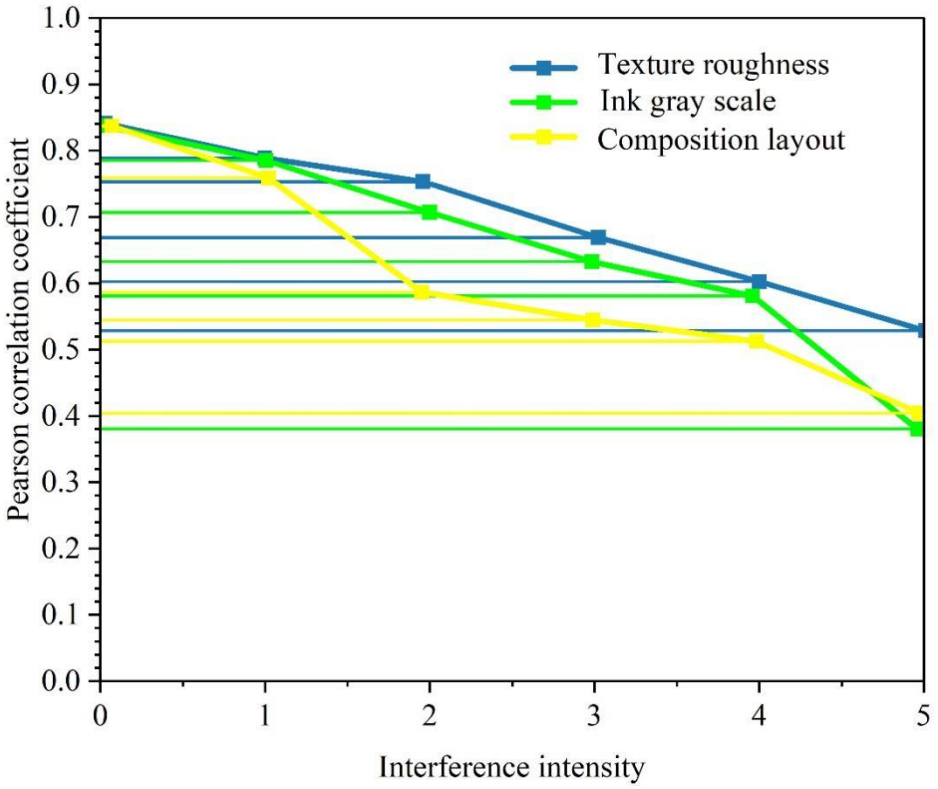

Figure 12. Experimental results.

4. Conclusion

In this paper, a composition optimization method for modern ink figure painting is proposed based on the characteristics of the main body of the picture, the principle of visual balance and the rule of thirds.

The accuracy and rotational consistency of the model are improved to a large extent compared with the comparison model AlexNet, and the accuracy and rotational consistency are always kept above 90%.

When the modern ink figure painting composition optimization model is not using data enhancement, the rotation sensitivity is reduced to a certain extent compared with the AlexNet model, and the stability of the model is further improved, but it is more sensitive to the rotation of triangular composition images. After training with data enhancement, the accuracy and stability of all composition types are improved for both the comparison model and the model in this paper, especially the accuracy of all composition types of the model in this paper stays above 75%. It shows that the data enhancement training can further optimize the performance of the composition structure optimization algorithm for modern ink figure painting in this paper.

The model is more sensitive to changes in composition, ink color and texture, so the model can accurately capture the detail changes in modern figure ink painting, thus better preserving the artistic style of traditional figure ink painting in the composition optimization process, providing an operable technical path for the optimization of the composition structure of traditional figure ink painting, and expanding the application of geometric algorithms in the field of art.

References

1. Dong, T., & Dechsubha, T. (2024). The Artistic Performance of Ink Painting In The Digital Era. Pakistan Journal of Life & Social Sciences, 22(2).
2. Kups, H. (2021). A Look at the Chinese Art of Abstract Ink Painting. Nowa Polityka Wschodnia, 203.
3. Wang, C., Su-Lynn, G., & Chen, A. (2020). From Chinese aesthetic to art and design: exploring Chinese aesthetics of Chinese ink painting to create contemporary art. The International Journal of Visual Design, 14(1), 11.
4. Leung, W. Y. (2021). Two sides of landscape in ink-wash painting: Chinese landscape painting in expressive arts practice. Creative Arts in Education and Therapy (CAET), 209-220.
5. Liu, S. (2022). Appreciation and Analysis of Liu Guohui's Realistic Ink and Wash Figure Paintings. International Journal of Frontiers in Sociology, 4(4).

6. Yang, S. (2024). Modelling and Brushwork: A Discussion on the Creation of Chinese Ink Figure Painting. International Journal of Education and Social Development, 1(2), 48-51.
7. Xue, B. (2022). A study of the influence of design composition on Chinese painting. Highlights in Art and Design.
8. Soxibov, R. (2023). Composition and Its Application in Painting. Science and innovation, 2(C5), 108-113.
9. Fan, Z. B., & Zhang, K. (2020). Visual order of Chinese ink paintings. Visual Computing for Industry, Biomedicine, and Art, 3, 1-9.
10. Huang, M. (2024). Analysis And Research on the Composition and Line Characteristics of Chinese Meticulous Figure Painting Based on Deep Learning. Pakistan Journal of Life & Social Sciences, 22(2).
11. Redies, C. (2020). The Way I Paint—How Image Composition Emerges During the Creation of Abstract Artworks. i-Perception, 11(3), 2041669520925099.
12. Zheng, L., Weidong, Z., & Xuchen, G. (2015). Aesthetic preference in the spatial composition of traditional Chinese paintings. Perception, 44(5), 556-568.
13. Ciobanu, G., & Ungureanu, C. (2015). Visible and invisible structures in renaissance paintings. Materials Today: Proceedings, 2(6), 3884-3888.
14. Ollivier, Y., Arnold, L., Auger, A., & Hansen, N. (2017). Information-geometric optimization algorithms: A unifying picture via invariance principles. Journal of Machine Learning Research, 18(18), 1-65.
15. Yanjie Zhou,Feng Zhou,Fengjun Xi,Yong Liu,Yun Peng,David E. Carlson & Liyun Tu. (2025). Efficient few-shot medical image segmentation via self-supervised variational autoencoder. Medical Image Analysis,104,103637-103637.
16. Mei Lyu,Ge Qu,Jiaxuan Shi,Dong Sun & Yi Tian. (2025). A Method for Studying Building Color Harmony in Coastal Historic and Cultural Districts: A Case Study of Mojiko, Japan. Buildings,15(9),1496-1496.
17. Ryu Sook-Hee. (2007). Analysis of Interior Color Status in the Welfare Facility for the Elderly in Cheongju - Focused on the Application of Moon-Spencer's Theory of Color Harmony -. JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Planning & Design,23(6),313-320.
18. Hala Mohammad,Jiawei Li,Bochao Li,Jamilu Tijjani Baraya,Sana Kone,Zhenlong Zhao... & Jingquan Lin. (2025). Extreme Ultraviolet Multilayer Defect Profile Parameters Reconstruction via Transfer Learning with Fine-Tuned VGG-16. Micromachines,16(5),541-541.