

<https://doi.org/10.1038/s40494-025-01595-0>

Color reconstruction of ancient paintings based on dual pooling attention and pixel adaptive convolution

Zengguo Sun^{1,2}✉, Zhiyuan Zhang² & Xiaojun Wu^{1,2}

In order to address inaccurate colors in reconstructed results, inconsistent color reconstruction for background and content, and poor local color reconstruction, a model is proposed for the color reconstruction of ancient paintings based on dual pooling attention and pixel adaptive convolution. Firstly, a Dual Pooling Channel Attention module is proposed to address inaccurate colors in reconstructed results. This module enhances the model's ability to extract features by assigning different weights to the image channels, thereby reducing inaccurate colors. Additionally, to solve the problem of inconsistent color reconstruction due to variations in background and content, a Content Adaptive Feature Extraction module is constructed. This module adaptively adjusts the convolutional parameters in terms of the differences in background and content, improving the overall effectiveness of color reconstruction. Lastly, a Contrastive Coherence Preserving Loss is introduced to solve the problem of poor reconstruction of local colors. The loss enhances the model's focus on image localization by constraining local features, thereby improving the local color reconstruction. Comparison experiments and ablation experiments are performed on various datasets. Experimental results show that, compared with the latest models, the proposed model effectively preserves the structural information and content details of ancient paintings. It produces reconstructed results with clear outlines and harmonious colors, achieving better color reconstructed results both globally and locally.

Chinese ancient paintings embody the national spirit and artistic achievements spanning five thousand years, owning immense artistic and historical value. Due to environmental factors or human activities, these paintings often suffer from fading and discoloration^{1–3}, which greatly diminishes their aesthetic appeal and historical value. Therefore, reconstructing the colors of ancient paintings is of great importance.

Color reconstruction of ancient paintings enhances colors in backgrounds, mountains, trees, rocks, and other elements with manual or digital methods. It addresses issues of fading and discoloration while preserving structural information and content details. Manual color reconstruction requires a deep analysis of the painting's style, content, and techniques, which is inefficient and can even damage the artwork⁴. As computer technology advances, the techniques for digital color reconstruction are widely used. Digital methods are easily adjustable, and they avoid direct manipulation of the original paintings, thus preventing potential damage. Digital color reconstruction techniques are based on traditional methods and deep learning methods. Traditional methods rely on model selection and

parameter settings. For example, Li et al.⁵ used multispectral imaging to identify pigments for mural color reconstruction. However, improper model selection or parameter settings can result in suboptimal reconstructed results.

In recent years, with the advancement of artificial intelligence, deep learning has been widely applied to color reconstruction tasks^{6–9}, enhancing the effectiveness of color reconstruction by deeply learning the characteristics of ancient paintings. Using deep learning techniques for the color reconstruction of ancient paintings typically involves constructing datasets for color reconstruction, training models to learn semantic and color features of the paintings, and ultimately applying the trained models to reconstruct colors. Color reconstruction models often use VGG as an encoder, which struggles to preserve image details. Meanwhile, inaccurate colors frequently appear in the reconstructed results due to the difficulties in accurately extracting semantic and color features. Additionally, models generally process images as a whole, resulting in suboptimal local color reconstruction. Therefore, it is crucial to preserve the structural information

¹Key Laboratory of Intelligent Computing and Service Technology for Folk Song, Ministry of Culture and Tourism, Xi'an, China. ²School of Computer Science, Shaanxi Normal University, Xi'an, China. ✉e-mail: sunzg@snnu.edu.cn

and content details of ancient paintings, accurately reconstruct colors, and enhance local reconstruction effects.

The CAP-VSTNet¹⁰ introduced a method by using flow-based models¹¹ for natural image style transfer, effectively preserving content information. Inspired by this, a color reconstruction model is proposed for ancient paintings based on Dual Pooling Attention and Pixel Adaptive Convolution. This model maintains the structural integrity and content details of ancient paintings by using the flow-based model, mitigates fading and discoloration issues, reduces inaccuracies in color reconstruction, and enhances the reconstruction effects both globally and locally. The main contributions of this paper are as follows:

- To address the issue of inaccurate colors in color reconstruction of ancient paintings, a Dual Pooling Channel Attention module (DPCA) is constructed. This module enhances the model's ability to extract features by assigning different weights to the image channels, reducing inaccurate colors and improving the color reconstruction of ancient paintings.
- To solve the problem of inconsistent color reconstruction due to variations in background and content, a Content Adaptive Feature Extraction module (CAFE) is constructed. This module adaptively adjusts the convolution parameters in terms of the differences in background and content, improving the overall effectiveness of color reconstruction.
- To improve the localized color reconstruction of ancient paintings, a Contrastive Coherence Preserving Loss (CCPL) is introduced. The loss enhances the model's focus on image localization by constraining local features, thereby improving the local color reconstruction.

Related work

The proposed method involves image style transfer, flow-based models, and color reconstruction. This section discusses related work in these three areas.

Image style transfer

Image style transfer is an important research area in computer vision. It transfers the style of a style image to a content image while preserving the content. The content of an image includes details and positions of objects, while style refers to the information such as color and brightness. Gatys⁶ used convolutional neural networks to extract image features, separating content and style information at the feature level. Following this, many deep learning-based methods for image style transfer emerged. To address the slow speed of style transfer, feed-forward neural networks were used to accelerate the process^{8,12}. Research then focused on using a single model for multi-style transfer^{13,14}, but these models only handle specific styles. Consequently, many studies^{15,16} aimed at universal style transfer. Li et al.¹⁷ proposed a linear transfer network for transferring any style to a content image. SANet¹⁸ added non-local self-attention to style transfer models but failed to align local features. AdaAttN⁷ combined SANet and AdaIN¹⁶, balancing local and global stylization effects. PhotoWCT¹⁹ replaced the upsampler operation in VGG decoders with uppooling, better preserving details in transfer results. WCT2²⁰ replaced pooling layers in VGG decoders with wavelet pooling, better preserving spatial information of content images. DPST²¹ and CAP-VSTNet¹⁰ introduced Matting Laplacian²² into networks to retain details in transfer results, achieving good results.

Flow-based models

Flow-based models are generative models that learn high-dimensional feature distributions with a series of reversible transformations. Dinh et al.¹¹ first proposed NICE to model high-dimensional complex data for image generation. Real NVP²³ introduced new transformation operations to reduce computation and improve image quality. Glow²⁴ simplified the structure of Real NVP, making the model more streamlined and standardized. Flow++²⁵ enhanced feature expression by introducing self-attention mechanisms and new coupling layers. Recently, flow-based models have been applied in image generation^{24,25}, speech synthesis²⁶, and other fields. BeautyGlow²⁷ designed a flow-based model for digital makeup. ArtFlow²⁸

used a flow-based model for artistic style transfer. CAP-VSTNet¹⁰ proposed a flow-based model for universal style transfer in images and videos.

Color reconstruction

Color reconstruction involves reconstructing color from faded or discolored images. Wan et al.²⁹ used variational autoencoders to reconstruct details and colors in photos. Pik-Fix³⁰ focused on reconstructing colors in degraded old photographs. JWA et al.³¹ used deep convolutional networks for image colorization of natural images. DDNM³² developed a zero-shot framework for colorizing grayscale natural images. Wang et al.³³ built a generative adversarial network model for color reconstruction of faded and discolored murals. DC-CycleGAN³⁴ extracted complex features of mural images to construct a model for the color reconstruction of murals. CR-ESRGAN³⁵ addressed the fading of cultural relics by developing a color reconstruction model, but suffers from inaccurate color reconstruction. The proposed model effectively preserves structural information and content details. It adapts to different contents of ancient painting, achieving better color reconstructed results both globally and locally.

Methodology

The model is proposed for color reconstruction. In this section, the proposed model is described in detail, including the overall structure, Dual Pooling Channel Attention module (DPCA), Content Adaptive Feature Extraction module (CAFE), and loss functions.

Overall structure

It is essential to preserve the structural and detailed information of degraded ancient paintings in color reconstruction. CAP-VSTNet¹⁰ uses a flow-based model for natural image style transfer, effectively retaining content information. Inspired by this, a method for the color reconstruction of ancient paintings is proposed with a flow-based model, considering the rich details and delicate colors of ancient paintings.

The structure of the constructed model is shown in Fig. 1, including the feature extraction module and the color reconstruction module. Firstly, the model uses the feature extraction module to obtain features from the degraded and reference images. Then, these features are fed into the color reconstruction module for color reconstruction, yielding reconstructed color features. Lastly, the reconstructed features are input back into the feature extraction module, and then the color-reconstructed image is obtained in terms of the reversibility of the flow-based model.

The feature extraction module consists of a channel expansion module, several reversible residual blocks, two Squeeze modules, and a channel refinement module. The channel expansion module increases the number of channels of the input image, thereby enhancing the model's ability to feature representation. The reversible residual blocks are stacked to achieve a larger receptive field, thereby extracting deeper features. The Squeeze module reduces the spatial dimensions of features while increasing the number of feature channels to capture larger-scale feature information. The channel refinement module reduces the number of feature channels to eliminate redundant information. The color reconstruction module integrates Cholesky decomposition³⁶ and WCT¹⁵ for color reconstruction, producing reconstructed features.

The forward and backward inference in reversible residual blocks are shown in Fig. 2. In the forward inference, let x be the input and y the output. The input x is split into x_1 and x_2 along the channel dimension. The transformations are formulated by (1) and (2):

$$y_1 = x_1 + F(x_2) \quad (1)$$

$$y_2 = x_2 \quad (2)$$

where y_1 and y_2 are the results obtained from x_1 and x_2 according to (1) and (2), respectively. These are concatenated along the channel dimension to form the output y of the reversible residual block. F is the sequential concatenation of the Dual Pooling Channel Attention module (DPCA) and the

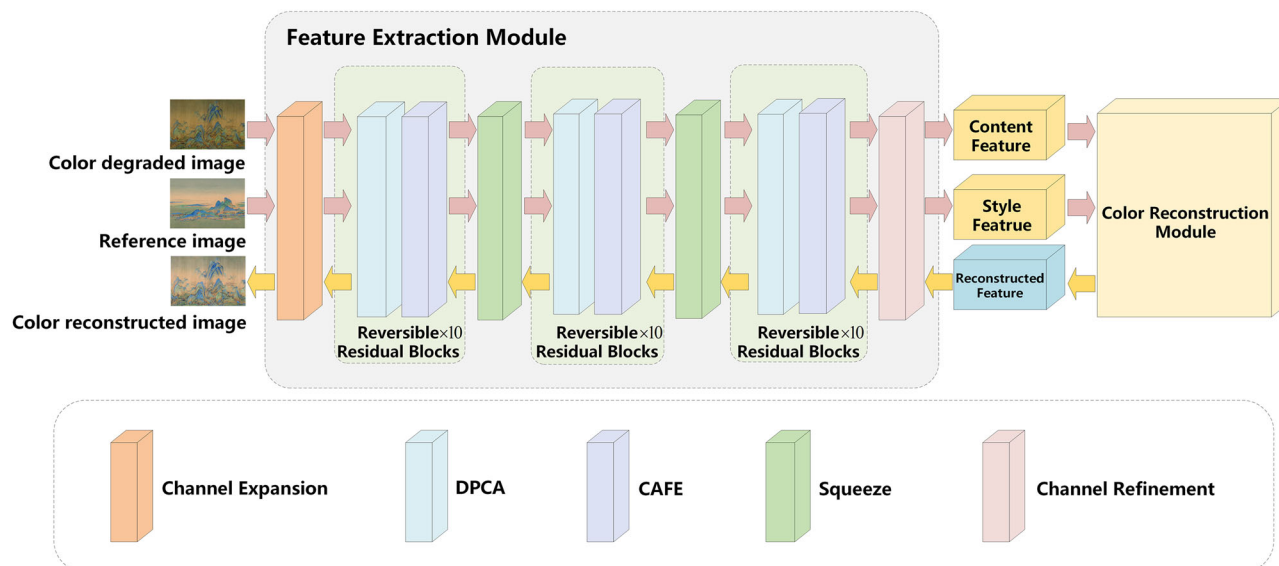
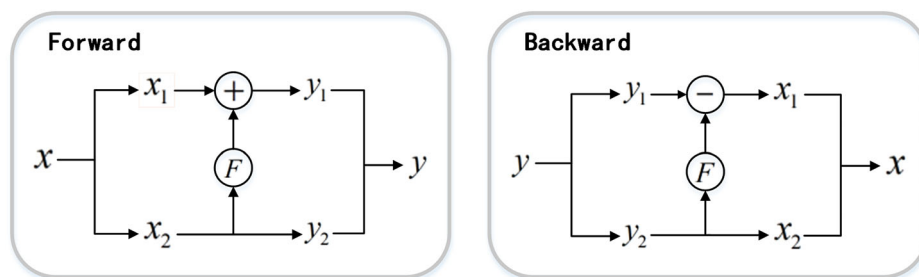


Fig. 1 | Structure of the proposed model.

Fig. 2 | Forward and backward inference in reversible residual blocks.



Content Adaptive Feature Extraction module (CAFE). In the backward inference, let y be the input and x the output. The input y is split into y_1 and y_2 along the channel dimension. The inverse process is formulated by (3) and (4).

$$x_2 = y_2 \quad (3)$$

$$x_1 = y_1 - F(x_2) \quad (4)$$

Information loss during data propagation in the model is avoided with these reversible transformations, effectively preserving the structure and details of degraded ancient paintings^{10,11}. The DPCA is introduced into the reversible residual block, which enhances the abilities to extract features and reduces inaccurate colors in the reconstructed results. The CAFE is incorporated into the reversible residual block, dynamically adjusting convolutional parameters to ensure color consistency between the background and the content. The Contrastive Coherence Preserving Loss is introduced to constrain the local features of degraded and reconstructed paintings, which improves the local color reconstruction.

Dual pooling channel attention module

Chinese ancient paintings exhibit rich color layers. It is essential to accurately extract features of ancient paintings for color reconstruction. The Efficient Channel Attention Network (ECANet)^{37,38} enhances the model's performance by calculating channel weights through local cross-channel interaction. Inspired by this, to address the issue of inaccurate colors in reconstructed results, the Dual Pooling Channel Attention (DPCA) module is constructed. This module assigns different weights to image channels,

enhancing the model's ability to extract and match image features, thereby reducing inaccurate colors, as shown in Fig. 3.

Firstly, an average pooling (AP) is performed on the input features of size $H \times W \times C$ to obtain features of size $50 \times 50 \times C$. Then, global average pooling (GAP) is applied to these features, producing features of size $1 \times 1 \times C$. After that, a convolution operation with a kernel size $k = 5$ and a Sigmoid activation function is performed to obtain the attention weights. Finally, the weights are element-wise multiplied with the input features to produce the output features. The features of images are better extracted and matched by assigning different weights to the image channels, reducing inaccurate colors in reconstructed results.

Content adaptive feature extraction module

Elements in Chinese ancient paintings, such as mountains, trees, and rocks, are often painted with different pigments and techniques from those used for backgrounds. This results in varying degrees of color degradation and makes it challenging to adaptively reconstruct colors based on the background and compositional elements. Pixel Adaptive Convolution (PAC)³⁹ modifies the convolution kernel according to image features, allowing adaptive feature extraction based on image content. Inspired by this, to address the aforementioned problem, the Content Adaptive Feature Extraction (CAFE) module is constructed. CAFE adjusts convolution kernel parameters based on image content, better adapting to the backgrounds and compositional elements of ancient paintings, as shown in Fig. 4.

CAFE consists of two convolution layers and one pixel adaptive convolution layer. When features are input into CAFE, convolution is first performed to obtain features f and v . Then, f is input into the kernel function K to obtain adaptive weights. After that, the weights are multiplied by the convolution kernel W to obtain pixel-adaptive convolution

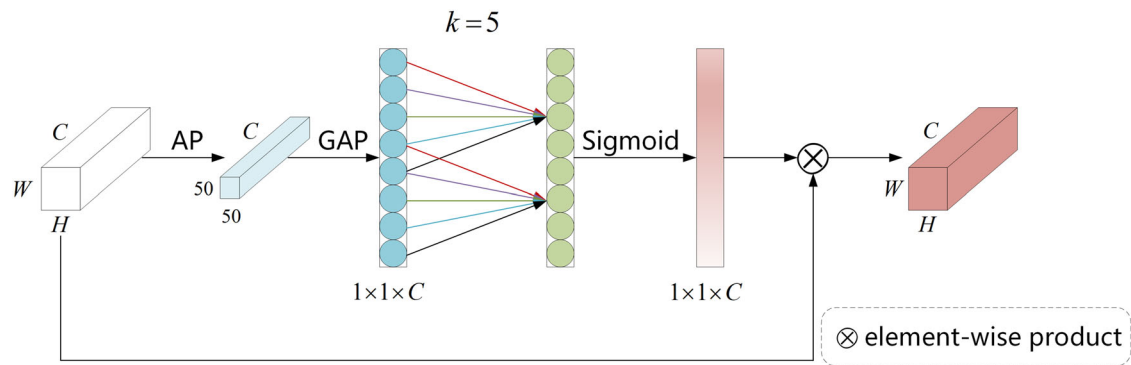


Fig. 3 | Dual Pooling Channel Attention module.

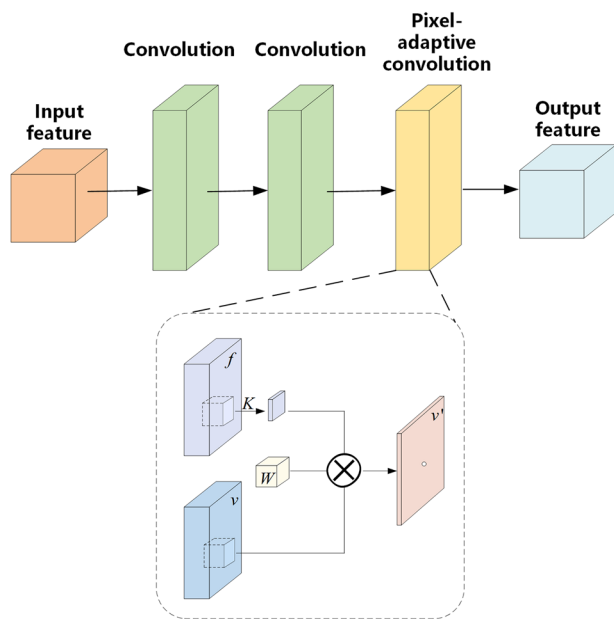


Fig. 4 | Content Adaptive Feature Extraction Module.

parameters. Simultaneously, the convolution kernel parameters are multiplied by the feature v to obtain the output feature v' . The kernel function K is formulated by (5):

$$K(f_{(x_i, y_i)}, f_{(x_j, y_j)}) = \exp\left(-\frac{1}{2}(f_{(x_i, y_i)} - f_{(x_j, y_j)})^T(f_{(x_i, y_i)} - f_{(x_j, y_j)})\right) \quad (5)$$

where $f_{(x_i, y_i)}$ represents the feature value at the position (x_i, y_i) in the feature map, T denotes the matrix transpose. The module obtains adaptive weights by inputting features into the kernel function K . The weights are then multiplied by the convolution kernel W , enabling the adjustment of convolution kernel parameters based on image content. This obtains better color reconstruction adapted to the backgrounds and compositional elements of ancient paintings.

Loss function

The local colors of Chinese ancient paintings are rich and detailed, making it challenging to reconstruct the color of paintings accurately. The Contrastive Coherence Preserving Loss⁴⁰ is introduced to enhance the model's focus on image localization, thereby improving the reconstruction of local colors.

Total loss. The total loss of the proposed model is formulated by (6):

$$L_{total} = \lambda_{ccp}L_{ccp} + \lambda_sL_s + \lambda_mL_m + \lambda_{cyc}L_{cyc} \quad (6)$$

where L_{ccp} is the contrastive coherence preserving loss, L_s is the style loss, L_m is the Matting Laplacian loss, L_{cyc} is the cycle consistency loss, and λ_{ccp} , λ_s , λ_m , and λ_{cyc} are the weights for L_{ccp} , L_s , L_m and L_{cyc} , respectively.

Contrastive coherence preserving loss. The Contrastive Coherence Preserving Loss (CCPL)⁴⁰ is introduced to enhance the model's focus on image localization by constraining local features, thereby improving the reconstruction of local colors. To apply CCPL, initially, the degraded and reconstructed images are input into an encoder to obtain feature maps. Feature vectors are then sampled from these maps, along with their eight nearest neighboring vectors for each vector. Then the difference vectors are obtained by vector subtraction between a vector and its neighboring vectors. CCPL constrains local features by using positive and negative pairs, positive pairs consist of difference vectors from the same spatial locations in the degraded and reconstructed paintings, while negative pairs consist of vectors from different locations. CCPL operates by maximizing the mutual similarity of positive pairs while minimizing the similarity of negative pairs, thereby enhancing local coherence. The contrastive coherence preserving loss is formulated by (7):

$$L_{ccp} = \sum_{m=1}^N \left[-\log \left[\frac{\exp(d_g^m \cdot d_c^m / \tau)}{\exp(d_g^m \cdot d_c^m / \tau) + \sum_{n=1, n \neq m}^N \exp(d_g^m \cdot d_c^n / \tau)} \right] \right] \quad (7)$$

where N is the total number of difference vectors. d_g^m and d_c^m are the m -th difference vectors in reconstructed and degraded paintings, respectively. d_c^n is the n -th difference vector in degraded paintings. τ is a hyperparameter set to 0.07.

Style loss. The style loss¹⁰ is formulated by (8):

$$L_s = \sum_{i=1}^4 \|\mu(\phi_i(I_{cs})) - \mu(\phi_i(I_s))\|_2 + \sum_{i=1}^4 \|\sigma(\phi_i(I_{cs})) - \sigma(\phi_i(I_s))\|_2 \quad (8)$$

where I_s is the reference painting, and I_{cs} is the color-reconstructed painting. Features of ancient paintings are extracted using the first four layers of the pre-trained VGG19 network. ϕ_i is the output feature map of the i -th layer of the VGG19 network. μ and σ are the mean and standard deviation, respectively. $\|\cdot\|_2$ is the L_2 -norm. The style loss constrains the color differences between the reconstructed and reference paintings, making their color styles similar.

Fig. 5 | Dataset construction process.

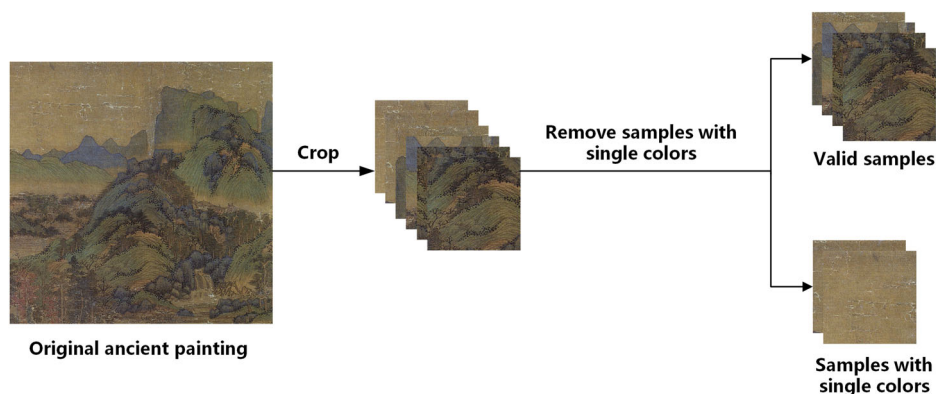


Table 1 | Number of samples of constructed dataset

Sample style	Training set	Test set	Total
ink wash	839	20	859
blue and green	1649	20	1669
light crimson	1340	20	1360
Total	3828	60	3888

Matting Laplacian loss. The Matting Laplacian loss^{10,22} is formulated by (9):

$$L_m = \frac{1}{N} \sum_{c=1}^3 V_c[I_{cs}]^T M V_c[I_{cs}] \quad (9)$$

where N is the number of image pixels. $V_c[I_{cs}]$ is the vectorized form of the color-reconstructed painting I_{cs} in channel c . T is the transpose of the obtained vector. M represents the Matting Laplacian matrix of the color-degraded painting. The Matting Laplacian loss constrains the pixel differences between the color-degraded and color-reconstructed paintings. This ensures better preservation of lines, brushstrokes, and other details in the color-reconstructed painting.

Cycle Consistency loss. The cycle consistency loss¹⁰ is formulated by (10):

$$L_{cyc} = \|\tilde{I}_C - I_C\|_1 \quad (10)$$

where \tilde{I}_C is the result obtained by inputting the reconstructed painting into the model, using the degraded painting as a reference. I_C is the degraded painting. $\|\cdot\|_1$ is the L_1 -norm. The cycle consistency loss allows the model to reconstruct the input image. This constrains the features of the degraded and reference paintings, ensuring accurate color matching between them.

Experiments and results

This section will introduce the dataset, training process, comparison experiments, and ablation study, respectively.

Construction of dataset

A dataset for the color reconstruction of ancient paintings is established to train and test the proposed model. Firstly, 214 paintings are selected, including the styles of ink wash, blue and green, and light crimson from different dynasties. These paintings are then cropped into 512×512 , excluding those with overly uniform colors like pure white or black. This leads to 3888 sample images. The construction process is shown in Fig. 5. 3828 samples are used as the training set, and 60 samples are used as the test set. The number of samples is detailed in Table 1, and some samples are shown in Fig. 6.

Training process

After testing, the weights for the model's loss function are set to $\lambda_{cp} = 5$, $\lambda_s = 1$, $\lambda_m = 1500$, and $\lambda_{cyc} = 10$. Adam optimizer is used to update the weights. The initial learning rate is set to 0.00001. The model is trained for 80,000 iterations with a batch size of 2. Training and testing are performed on a single NVIDIA RTX 3090 GPU.

The curve of the loss function of the proposed model during training is displayed in Fig. 7. The lighter line represents the raw loss function curve, while the darker line shows the smoothed curve obtained using TensorBoard. A logarithmic scale is used for the vertical axis to clearly depict the convergence trend of the loss function. Obviously, the model's loss steadily decreases and eventually stabilizes as iterations increase. This indicates that the model converges stably, leading to effective color reconstruction of ancient paintings.

Experimental results

The color reconstructed results on the constructed dataset are shown in Fig. 8. By comparing Fig. 8a with Fig. 8c, the proposed model effectively preserves the details of compositional elements like mountains, trees, and rocks of the degraded paintings. For instance, in Fig. 8c, the boundaries of the tree are distinct in the first row, the original lines of the river are preserved in the second row, and the contours of the mountain are maintained in the third row.

By comparing Fig. 8b with Fig. 8c, the model successfully transfers the color style of the reference paintings. For example, in Fig. 8c, the color of the rocks in the first row is similar to that of the reference painting, the overall greenish tint in the second row matches the reference, and the color of the mountains in the third row is similar to that of the reference painting.

In summary, the proposed model is demonstrated to be effective for color reconstruction of ancient paintings in ink wash, blue and green, and light crimson styles.

Comparison experiments

To verify the effectiveness of the proposed model, ArtFlow²⁸, PCA-KD⁴¹, Photo-WCT2⁴², DTP⁴³, CAP-VSTNet¹⁰, and CCPL⁴⁰ are selected for comparison. ArtFlow develops a flow-based method for natural image style transfer. PCA-KD employs knowledge distillation to construct models for natural image style transfer. PhotoWCT2 uses blockwise training and high-frequency residual skip connections for style transfer. DTP creates a model incorporating correlation and generation modules to perform style transfer on input images. CAP-VSTNet introduces a flow-based model for natural image style transfer. CCPL devises a versatile method for style transfer.

Qualitative comparison. To comprehensively evaluate the proposed model, faded ancient paintings are selected with ink wash, blue and green, and light crimson styles, respectively. Reference paintings are chosen based on the same styles, similar compositional elements, and good color preservation.

Fig. 6 | Samples of the constructed dataset. a The Ink wash samples. **b** The blue and green samples. **c** The light crimson samples (The first row is the degraded paintings, and the second row is the reference paintings).

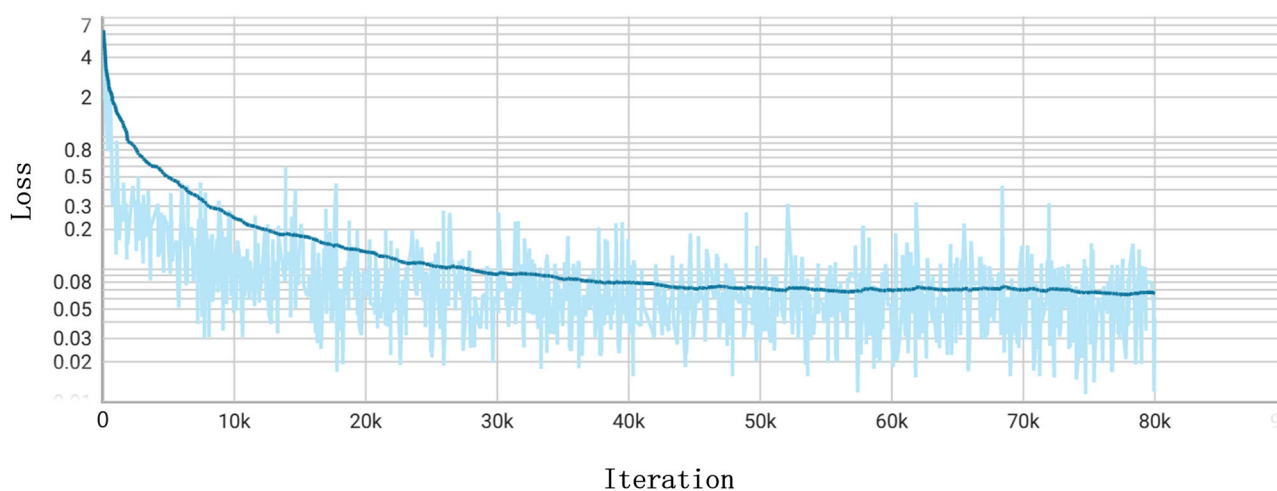
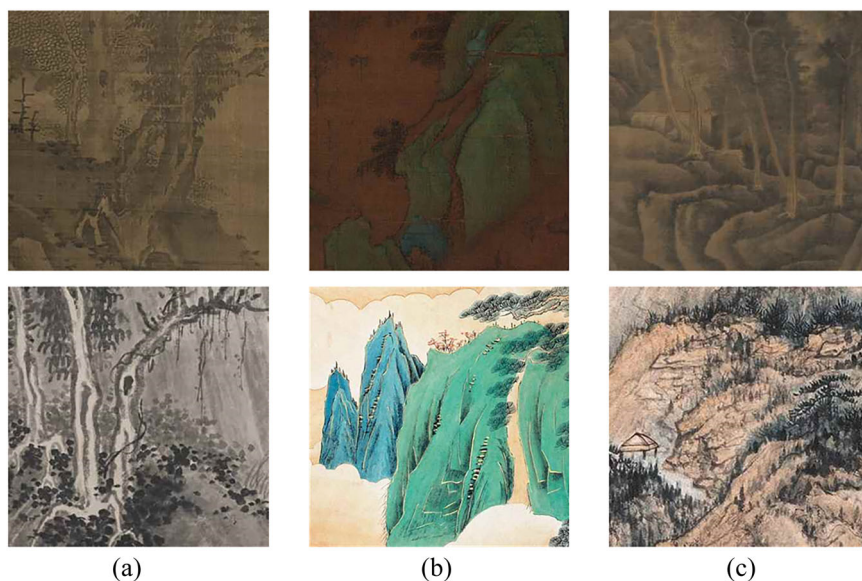


Fig. 7 | Curve of loss function changes.

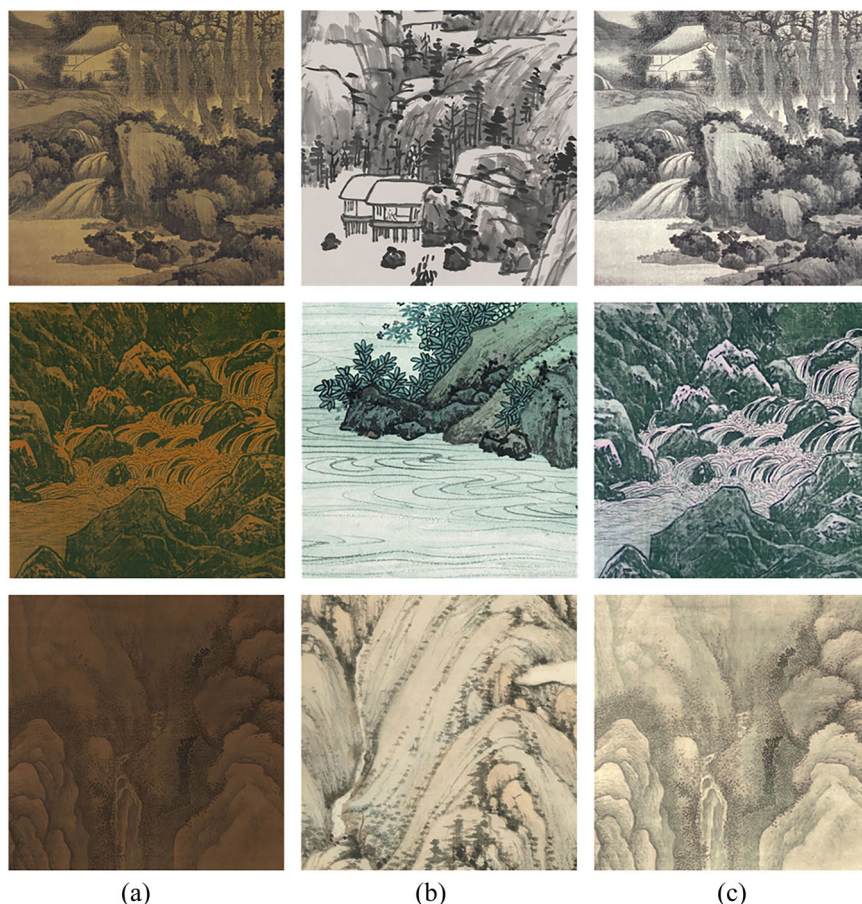
The comparative results of different models are shown in Fig. 9. In Fig. 9c, CAP-VSTNet preserves the content details well but fails to reconstruct color accurately. The color of the mountains in the third row differs significantly from the reference painting, and the color of the mountains in the upper part of the sixth row is not accurately reconstructed. In Fig. 9d, CCPL produces unsatisfactory color reconstructions. The color of the mountains in the third row deviates significantly from the reference painting, and the color of the lower-right mountain in the fifth row is not accurately reconstructed. The results of ArtFlow are similar in style to the reference paintings in Fig. 9b. However, noticeable color inconsistencies appear due to the lack of matching color and content features between input images. Details of ancient paintings are not well preserved compared to the degraded paintings, resulting in blurry details. Specifically, in Fig. 9e, significant color inconsistency appears in the upper part of the third row, and detail loss is observed in the mountains in the upper right of the fifth row. PCA-KD preserves content details in Fig. 9f, but the color still shows some differences from that of the reference paintings. For example, the first and fifth rows are overall blurry, and the colors of the mountain in the third row are not accurately reconstructed. PhotoWCT2 preserves content details but fails to reconstruct colors accurately in Fig. 9g. Incorrect colors are seen around the branches of the first row and the trees on the right side of the fifth row. The

results of DTP show color deviations and unclear details in Fig. 9h. The colors of the second and third rows differ from those of the reference paintings, and the details of the tree on the right side of the fifth row are blurred.

In Fig. 9i, the proposed model first incorporates a Dual Pooling Channel Attention module to correct color inaccuracies. Then, a Content Adaptive Feature Extraction module is built to improve color reconstruction for different backgrounds and elements in ancient paintings. Lastly, the Contrastive Coherence Preserving Loss is introduced to enhance local color reconstruction. In conclusion, the proposed model effectively preserves painting details and improves color reconstruction, both globally and locally.

Quantitative comparison. The experimental results in Fig. 9 correspond to the quantitative metrics presented in Tables 2–4, which represent the results for ink wash, blue and green, and light crimson styles, respectively. Structural Similarity (SSIM)⁴⁴ and Gram Loss¹⁰ are used as evaluation metrics. SSIM calculates the preservation of content details, while Gram Loss evaluates the color reconstructed effectiveness. A higher SSIM indicates greater structural similarity between the degraded and reconstructed paintings, while a lower Gram Loss indicates closer color style similarity between the reconstructed and reference paintings.

Fig. 8 | Color reconstructed results of the proposed model. a Degraded ancient paintings; **b** Reference paintings; **c** Reconstructed results (Rows 1, 2, and 3 show the reconstructed results in ink wash, blue and green, and light crimson styles, respectively).



Taking the results for the ink wash style as an example, it is evident that CAP-VSTNet has relatively poor color reconstruction, resulting in a high Gram Loss. CCPL shows unsatisfactory color reconstruction and fails to effectively preserve content details, resulting in lower SSIM and higher Gram Loss compared to the proposed model. ArtFlow performs poorly in preserving the structural details of the degraded paintings, resulting in a low SSIM. PCA-KD has relatively poor color reconstruction, resulting in a high Gram Loss. PhotoWCT2 fails to reconstruct the color of the ancient paintings effectively, resulting in a high Gram Loss. DTP shows poor color reconstruction and detail loss, thus having lower SSIM and higher Gram Loss compared to the proposed model. The same conclusion can be drawn for the other two styles.

The proposed model firstly incorporates a Dual Pooling Channel Attention module to correct color inaccuracies. Secondly, a Content Adaptive Feature Extraction module is constructed to handle different backgrounds and compositional elements in ancient paintings. Lastly, the Contrastive Coherence Preserving Loss is introduced to improve local reconstruction. In conclusion, the proposed model achieves the highest SSIM and the lowest Gram Loss, achieving the best performance in color reconstruction.

Ablation study

An ablation experiment is conducted on the constructed dataset to verify the effect of the Dual-Pooling Channel Attention module (DPCA), Content-Adaptive Feature Extraction module (CAFE), and Contrastive Coherence Preserving Loss (CCPL) on the reconstructed ability of the proposed model, including both qualitative and quantitative analysis.

The results of the ablation experiments for different modules are shown in Fig. 10, Fig. 11, and Fig. 12, respectively. DPCA improves the model's ability to match and extract image features by assigning different weights to the image channels, thereby reducing inaccurate colors. By

comparing Fig. 10c with Fig. 10d, after adding the DPCA, inaccurate colors around the leaves in the reconstructed paintings are reduced. CAFE adjusts convolution kernel parameters based on the differences in background and content, enhancing color reconstruction for the background and various elements. Figure 11c and d show that, when the CAFE is added, the background color of the reconstructed painting is more similar to that of the reference painting, improving the overall color reconstruction. CCPL enhances the model's focus on image localization by constraining local features, thereby improving the local color reconstruction. Figure 12c and d show that, after adding CCPL, the color of the stones on the right side of the painting is closer to that of the reference painting, improving local reconstruction. These comparisons demonstrate that each proposed module positively impacts the model's performance in color reconstruction.

The experimental results in Figs. 10–12 correspond to the quantitative metrics presented in Table 5. The DPCA focuses on prominent features by assigning weights to image channels, enhancing the model's ability to match and extract colors. Without DPCA, the model produces noticeable incorrect colors. The CAFE improves the color reconstruction of backgrounds and various elements by adjusting convolutional kernel parameters. In the absence of CAFE, color reconstruction of the backgrounds is less accurate. The CCPL enhances the model's ability to reconstruct local colors by constraining local features. Without CCPL, the model performs poorly in reconstructing local colors. The above analyses indicate that removing the proposed modules results in lower SSIM and higher Gram Loss compared to the proposed model, demonstrating the significance of the modules in color reconstruction.

The proposed model firstly incorporates a Dual Pooling Channel Attention module to correct color inaccuracies. Secondly, a Content Adaptive Feature Extraction module is constructed to improve the color reconstruction of backgrounds and elements. Lastly, the Contrastive Coherence Preserving Loss is introduced to enhance local color

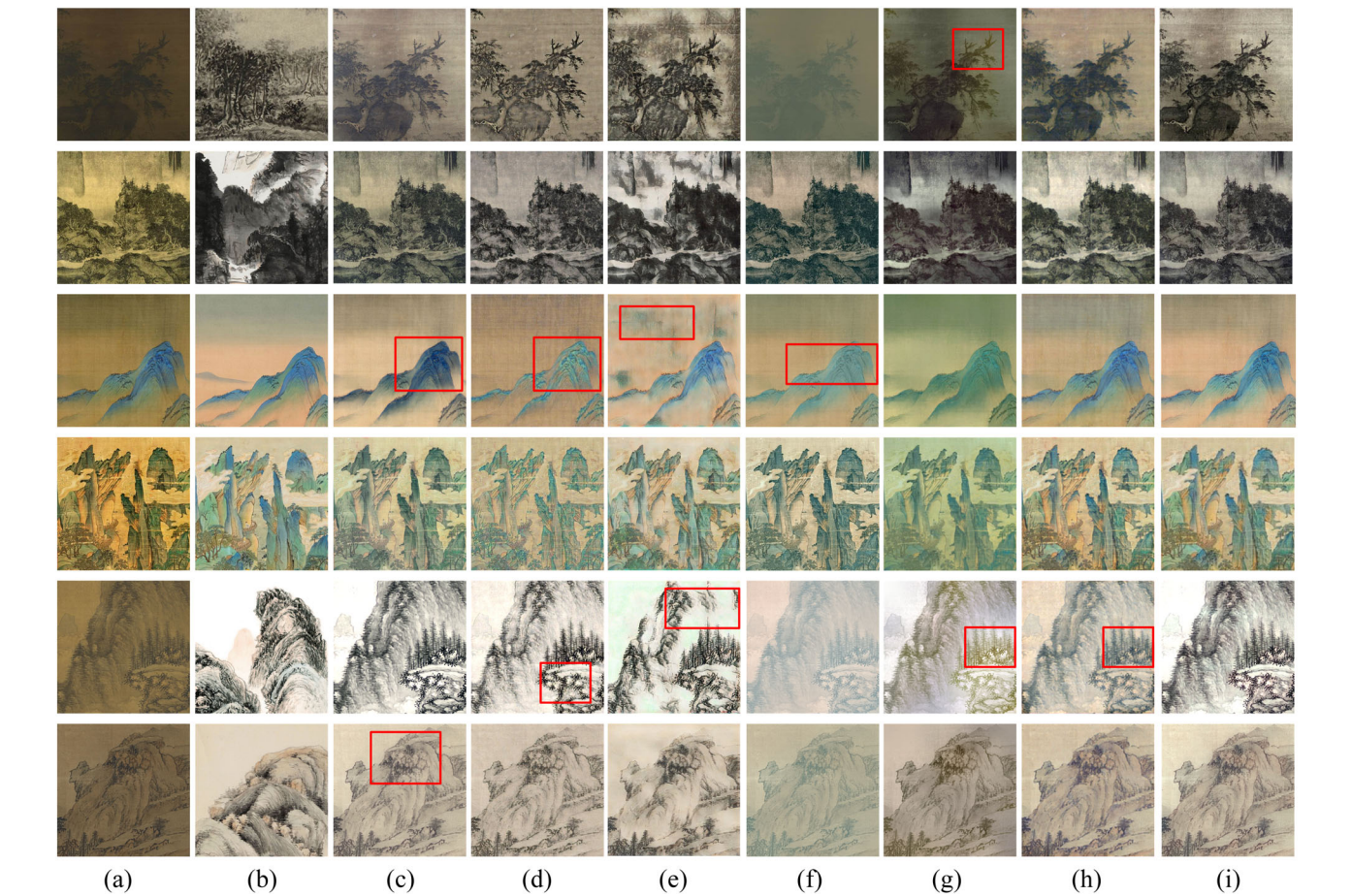


Fig. 9 | Comparison of reconstructed results using different models (Red boxes are used for comparison of details). **a** Degraded ancient paintings. **b** Reference paintings. **c** CAP-VSTNet. **d** CCPL. **e** ArtFlow. **f** PCA-KD. **g** PhotoWCT2. **h** DTP. **i** Ours (Rows 1 and 2 correspond to the style of ink wash, rows 3 and 4 correspond to the style of blue and green, and rows 5 and 6 correspond to the style of light crimson).

Table 2 | Comparison of evaluation metrics (ink wash style)

Model	SSIM ↑	Gram Loss ↓
CAP-VSTNet	0.940	0.03917
CCPL	0.705	0.05946
ArtFlow	0.689	0.00658
PCA-KD	0.949	0.05046
PhotoWCT2	0.763	0.03365
DTP	0.918	0.00836
Ours	0.973	0.00242

The bold values indicate the best performance in each table.

Table 3 | Comparison of evaluation metrics (blue and green style)

Model	SSIM ↑	Gram Loss ↓
CAP-VSTNet	0.885	0.00125
CCPL	0.668	0.02165
ArtFlow	0.673	0.00249
PCA-KD	0.903	0.01528
PhotoWCT2	0.706	0.02154
DTP	0.893	0.01079
Ours	0.938	0.00034

The bold values indicate the best performance in each table.

Table 4 | Comparison of evaluation metrics (light crimson style)

Model	SSIM ↑	Gram Loss ↓
CAP-VSTNet	0.784	0.01220
CCPL	0.681	0.01639
ArtFlow	0.650	0.00706
PCA-KD	0.724	0.05648
PhotoWCT2	0.741	0.10012
DTP	0.778	0.04165
Ours	0.822	0.00685

The bold values indicate the best performance in each table.

reconstruction. In conclusion, the proposed model effectively preserves the details of ancient paintings and addresses the issues of fading and color changes. It reconstructs paintings with consistent colors, achieving better results both globally and locally.

Experiments on landscape painting dataset

To demonstrate the generalization ability of the proposed model, it is trained and tested on a landscape painting dataset⁴⁵, and compared with the contrast models. This dataset contains 2192 landscape painting samples with size of 512 × 512, which are obtained from open-source museums. 2152 samples are used as the training set, and 40 samples are used as the test set.

The experimental results are shown in Fig. 13, with the corresponding quantitative metrics shown in Tables 6–8, which represent the

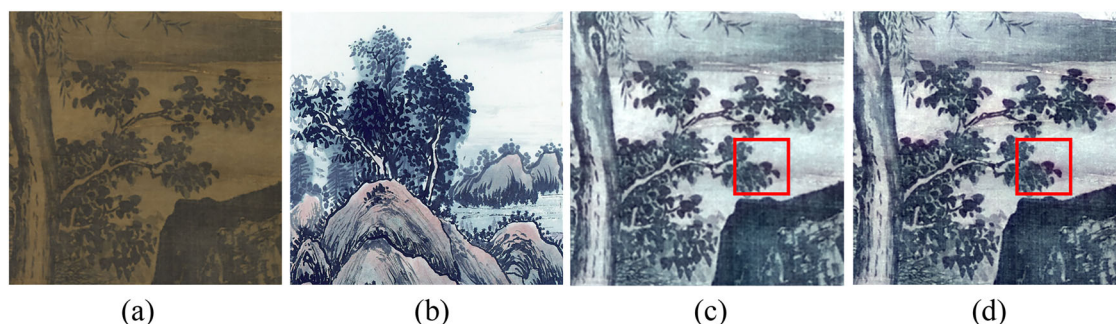


Fig. 10 | Comparison before and after using DPCA (Red boxes are used for comparison of details). a Degraded ancient painting. **b** Reference painting. **c** With DPCA. **d** Without DPCA.

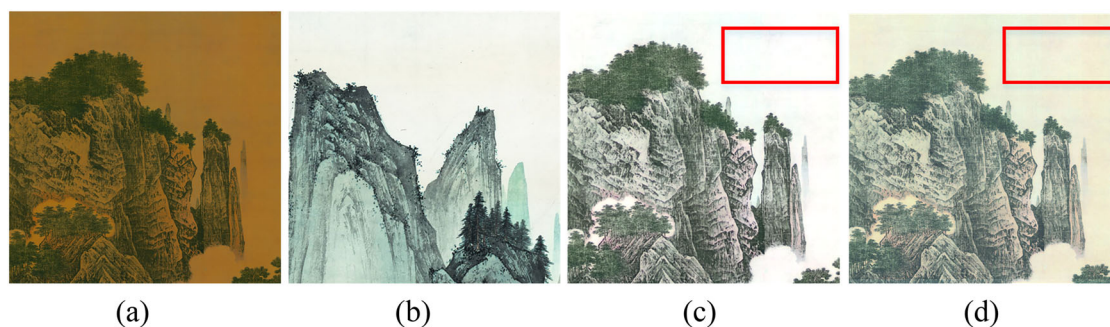


Fig. 11 | Comparison before and after using CAFE (Red boxes are used for comparison of details). a Degraded ancient painting. **b** Reference painting. **c** With CAFE. **d** Without CAFE.

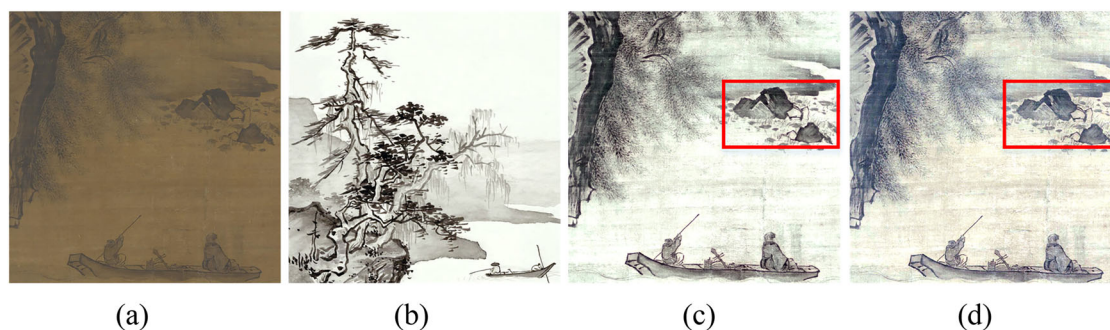


Fig. 12 | Comparison before and after using CCPL (Red boxes are used for comparison of details). a Degraded ancient painting. **b** Reference painting. **c** With CCPL. **d** Without CCPL.

Table 5 | Evaluation metrics for the ablation study

Model	SSIM \uparrow	Gram Loss \downarrow
Without DPCA	0.788	0.00872
Without CAFE	0.802	0.01528
Without CCPL	0.775	0.01070
Ours	0.805	0.00794

The bold values indicate the best performance in each table.

results for ink wash, blue and green, and light crimson styles, respectively. Taking the results for the blue and green style as an example, CAP-VSTNet has relatively poor color reconstruction in Fig. 13c, resulting in a high Gram Loss. The results of CCPL in Fig. 13d show unsatisfactory color reconstruction and fail to effectively preserve content details, resulting in lower SSIM and higher Gram Loss compared to the proposed model.

ArtFlow fails to preserve the details of the degraded paintings in Fig. 13e, resulting in blurred reconstructions with the lowest SSIM. The reconstructed colors of PCA-KD in Fig. 13f still differ from the reference paintings, resulting in a high Gram Loss. PhotoWCT2 produces unsatisfactory color reconstructions in Fig. 13g, resulting in the highest Gram Loss. The results of DTP in Fig. 13h show unclear details and poor color reconstruction, resulting in lower SSIM and higher Gram Loss compared to the proposed model. The same conclusion can be drawn for the other two styles.

In Fig. 13i, the proposed model effectively preserves the details of ancient paintings and produces consistent colors in the reconstructed results, leading to the highest SSIM and the lowest Gram Loss.

In conclusion, the results from both the constructed ancient painting dataset and the existing landscape painting dataset demonstrate that, the proposed model performs well in color reconstruction of paintings, proving its applicability in this domain.

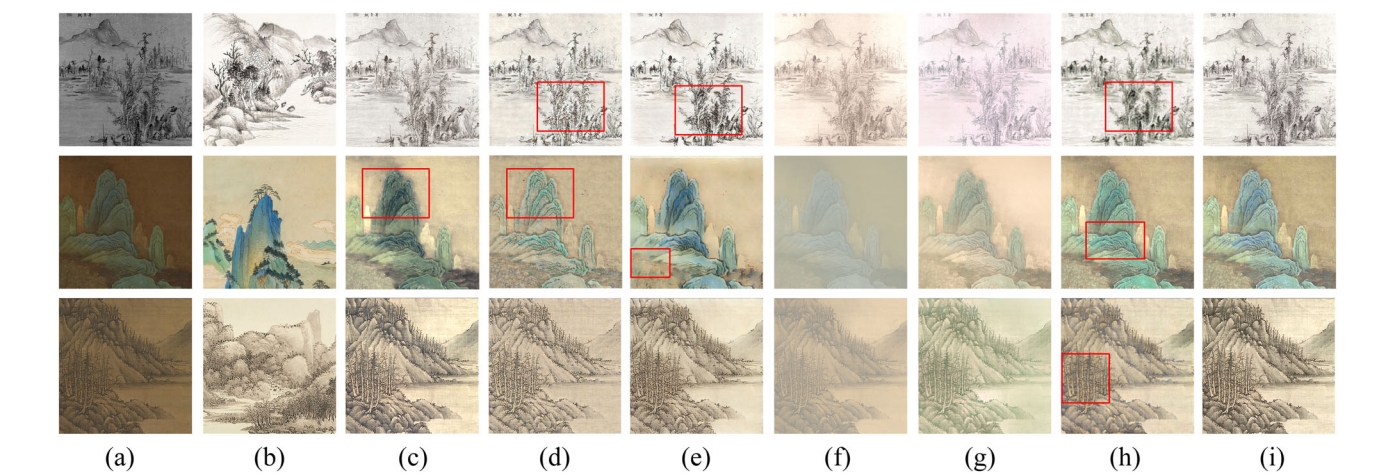


Fig. 13 | Reconstructed results of different models using the landscape painting dataset (Red boxes are used for comparison of details). a Degraded ancient paintings. b Reference paintings. c CAP-VSTNet. d CCPL. e ArtFlow. f PCA-KD. g PhotoWCT2. h DTP. i Ours (Rows 1, 2, and 3 correspond to ink wash, blue and green, and light crimson styles, respectively).

Table 6 | Comparison of evaluation metrics on the landscape painting dataset (ink wash style)

Model	SSIM ↑	Gram Loss ↓
CAP-VSTNet	0.793	0.00363
CCPL	0.686	0.01038
ArtFlow	0.664	0.00212
PCA-KD	0.729	0.02956
PhotoWCT2	0.715	0.06572
DTP	0.742	0.00543
Ours	0.806	0.00203

The bold values indicate the best performance in each table.

Table 8 | Comparison of evaluation metrics on the landscape painting dataset (light crimson style)

Model	SSIM ↑	Gram Loss ↓
CAP-VSTNet	0.644	0.00382
CCPL	0.518	0.01455
ArtFlow	0.502	0.00377
PCA-KD	0.747	0.01216
PhotoWCT2	0.715	0.01525
DTP	0.728	0.01423
Ours	0.753	0.00307

The bold values indicate the best performance in each table.

Table 7 | Comparison of evaluation metrics on the landscape painting dataset (blue and green style)

Model	SSIM ↑	Gram Loss ↓
CAP-VSTNet	0.588	0.01698
CCPL	0.528	0.01038
ArtFlow	0.524	0.00031
PCA-KD	0.584	0.00784
PhotoWCT2	0.568	0.09472
DTP	0.612	0.00241
Ours	0.677	0.00012

The bold values indicate the best performance in each table.

Conclusion

A color reconstruction model for ancient paintings is proposed based on dual pooling attention and pixel adaptive convolution. Firstly, a Dual Pooling Channel Attention module is proposed to address inaccurate colors in reconstructed results. This module enhances the model’s ability to extract features by assigning different weights to the image channels, thereby reducing inaccurate colors. Secondly, to solve the problem of inconsistent color reconstruction due to variations in background and content, a Content Adaptive Feature Extraction module is constructed. This module dynamically adjusts the convolutional parameters based on predefined features to adaptively extract features of ancient paintings, thereby improving the

overall color reconstruction of ancient paintings. Lastly, a Contrastive Coherence Preserving Loss is introduced to solve the problem of poor local color reconstruction. The loss enhances the model’s focus on image localization by constraining local features, thereby improving the local color reconstruction. Experiments are conducted on the constructed color reconstruction dataset of ancient paintings, including comparison experiments and ablation studies. The results show that the proposed model effectively preserves the structural information and content details of ancient paintings, while adapting to different backgrounds and content, achieving better color reconstruction both globally and locally. Additionally, color reconstruction experiments are performed on a landscape painting dataset, proving the effectiveness and applicability of the proposed model.

Data availability

The datasets used and analyzed in the current study are available from the corresponding author by reasonable request.

Received: 17 August 2024; Accepted: 9 December 2024;
Published online: 07 June 2025

References

- Hua, Z., Lu, D. & Pan, Y. Research on virtual color restoration and gradual changing simulation of Dunhuang Frasco. *J. Image Graphics* **7**, 79–84 (2002).
- Pan, Y. & Lu, D. Digital protection and restoration of Dunhuang mural. *J. Syst. Simul.* **15**, 310–314 (2003).

3. Wei, B., Pan, Y. & Hua, Z. An analogy-based virtual approach for color restoration of wall painting. *J. Comput. Res. Dev.* **36**, 1364–1368 (1999).
4. Huang, Y. Restoration and protection of silk calligraphy and painting cultural relics. *Identif. Apprec. Cult. Relics* **231**, 32–35 (2022).
5. Li, J. et al. Pigment identification of ancient wall paintings based on a visible spectral Image. *J. Spectrosc.* **2020**, 1–8 (2020).
6. Gatys, L. A., Ecker, A. S. & Bethge, M. Image style transfer using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2414–2423 (2016).
7. Liu, S. et al. AdaAttN: Revisit attention mechanism in arbitrary neural style transfer. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6649–6658 (2021).
8. Xu, Z. et al. Stylization-based architecture for fast deep exemplar colorization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9363–9372 (2020).
9. Hu, R. et al. Self-supervised color-concept association via image colorization. *IEEE Trans. Vis. Comput. Graph.* **29**, 247–256 (2022).
10. Wen, L., Gao, C. & Zou, C. CAP-VSTNet: Content affinity preserved versatile style transfer. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18300–18309 (2023).
11. Dinh, L., Krueger, D. & Bengio, Y. Nice: Non-linear independent components estimation. *arXiv Preprint arXiv: 1410.8516* (2014).
12. Ulyanov, D., Lebedev, V., Vedaldi, A. & Lempitsky, V. Texture networks: Feed-forward synthesis of textures and stylized images. *Proceedings of the 33rd International Conference on Machine Learning*. 1349–1357 (2016).
13. Chen, D. et al. Stylebank: An explicit representation for neural image style transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1897–1906 (2017).
14. Gong, X. et al. Neural stereoscopic image style transfer. *Proceedings of the European Conference on Computer Vision*. 54–69 (2018).
15. Li, Y. et al. Universal style transfer via feature transforms. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: ACM. 385–395 (2017).
16. Huang, X. & Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. *Proceedings of the IEEE International Conference on Computer Vision*. 1501–1510 (2017).
17. Li, X. et al. Learning linear transformations for fast image and video style transfer. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3809–3817 (2019).
18. Park, D. Y. & Lee, K. H. Arbitrary style transfer with style-attentional networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5880–5888 (2019).
19. Li, Y. et al. A closed-form solution to photorealistic image stylization. *Proceedings of the European Conference on Computer Vision*. 453–468 (2018).
20. Yoo, J. et al. Photorealistic style transfer via wavelet transforms. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9036–9045 (2019).
21. Luan, F. et al. Deep photo style transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4990–4998 (2017).
22. Levin, A., Lischinski, D. & Weiss, Y. A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 228–242 (2007).
23. Dinh, L., Sohl-Dickstein, J. & Bengio, S. Density estimation using Real NVP. *Proceedings of the International Conference on Learning Representations*. 1–32 (2017).
24. Kingma, D. P. & Dhariwal, P. Glow: Generative flow with invertible 1×1 convolutions. *Proceedings of the 32nd International Conference on Neural Information Processing Systems 2018*, 10236–10245 (2018).
25. Ho, J. et al. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *Proceedings of the International Conference on Machine Learning*. 2722–2730 (2019).
26. Prenger, R., Valle, R. & Catanzaro, B. Waveglow: A flow-based generative network for speech synthesis. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 3617–3621 (2019).
27. Chen, H. et al. Beautyglow: On-demand makeup transfer framework with reversible generative network. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10042–10050 (2019).
28. An, J. et al. Artflow: Unbiased image style transfer via reversible neural flows. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 862–871 (2021).
29. Wan, Z. et al. Old photo restoration via deep latent space translation. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 2071–2087 (2023).
30. Xu, R. et al. Pik-Fix: Restoring and colorizing old photos. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1724–1734 (2023).
31. Jwa, M. & Kang, M. Grayscale image colorization using a convolutional neural network. *J. Korean Soc. Ind. Appl. Math.* **25**, 26–38 (2021).
32. Wang, Y., Yu, J. & Zhang, J. Zero-shot image restoration using denoising diffusion null-space model. *Proceedings of the 11th International Conference on Learning Representations*. 1–31 (2023).
33. Wang, H. et al. Dunhuang mural restoration using deep learning. *Proceedings of the SIGGRAPH Asia 2018 Technical Briefs*. 1–4 (2018).
34. Xu, Z., Zhang, C. & Wu, Y. Digital inpainting of mural images based on DC-CycleGAN. *Heritage. Science* **11**, 169–181 (2023).
35. Zhou, X. et al. Color reconstruction of cultural relics image based on enhanced super-resolution generative adversarial network. *Radio Eng.* **53**, 220–229 (2023).
36. Kessy, A. & Lewin, A. Strimmer K. Optimal whitening and decorrelation. *Am. Stat.* **72**, 309–314 (2018).
37. Wang, Q. et al. ECA-Net: Efficient channel attention for deep convolutional neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11534–11542 (2020).
38. Liu, Y. et al. ABNet: Adaptive balanced network for multiscale object detection in remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–14 (2022).
39. Su, H. et al. Pixel-adaptive convolutional neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11166–11175 (2019).
40. Wu, Z. et al. CCPL: Contrastive coherence preserving loss for versatile style transfer. *Proceedings of the European Conference on Computer Vision*. 189–206 (2022).
41. Chiu, T. & Gurari, D. PCA-based knowledge distillation towards lightweight and content-style balanced photorealistic style transfer models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7844–7853 (2022).
42. Chiu, T. & Gurari, D. Photowct2: Compact autoencoder for photorealistic style transfer resulting from blockwise training and skip connections of high-frequency residuals. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2868–2877 (2022).
43. Kim, S., Kim, S. & Kim, S. Deep translation prior: Test-time training for photorealistic style transfer. *Proc. AAAI Conf. Artif. Intell.* **36**, 1183–1191 (2022).
44. Hong, K. et al. Domain-aware universal style transfer. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14609–14617 (2021).
45. Xue, A. End-to-end Chinese landscape painting creation using generative adversarial networks. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3863–3871 (2021).

Acknowledgements

We gratefully acknowledge Shaanxi Normal University's high-performance computing infrastructure SNNUGrid (HPC Centers: QILin AI LAB) for providing computer facilities. This work was supported by the National Key Research and Development Program of China (No.2017YFB1402102), the National Natural Science Foundation of China (No. 62377033), the Shaanxi Key Science and Technology Innovation Team Project (No. 2022TD-26), the Xi'an Science and Technology Plan Project (No. 23ZDCYJSGG0010-2022), and the Fundamental Research Funds for the Central Universities (No. GK202205036, GK202407007, GK202101004).

Author contributions

Conceptualization, Z.S.; methodology, Z.S.; validation, Z.Z. and X.W.; investigation, Z.Z.; data curation, Z.Z.; writing-original draft preparation, Z.S.; writing-review and editing, Z.S., Z.Z. and X.W.; supervision, Z.S.; funding acquisition, Z.S.; All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Zengguo Sun.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025