PColorizor: Re-coloring Ancient Chinese Paintings with Ideorealm-congruent Poems

Tan Tang
tangtan@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

Yanhong Wu
yanhongwu@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

Peiquan Xia
3210101568@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

Wange Wu
wuwange@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

Xiaosong Wang
wang64@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

Yingcai Wu*
ycwu@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

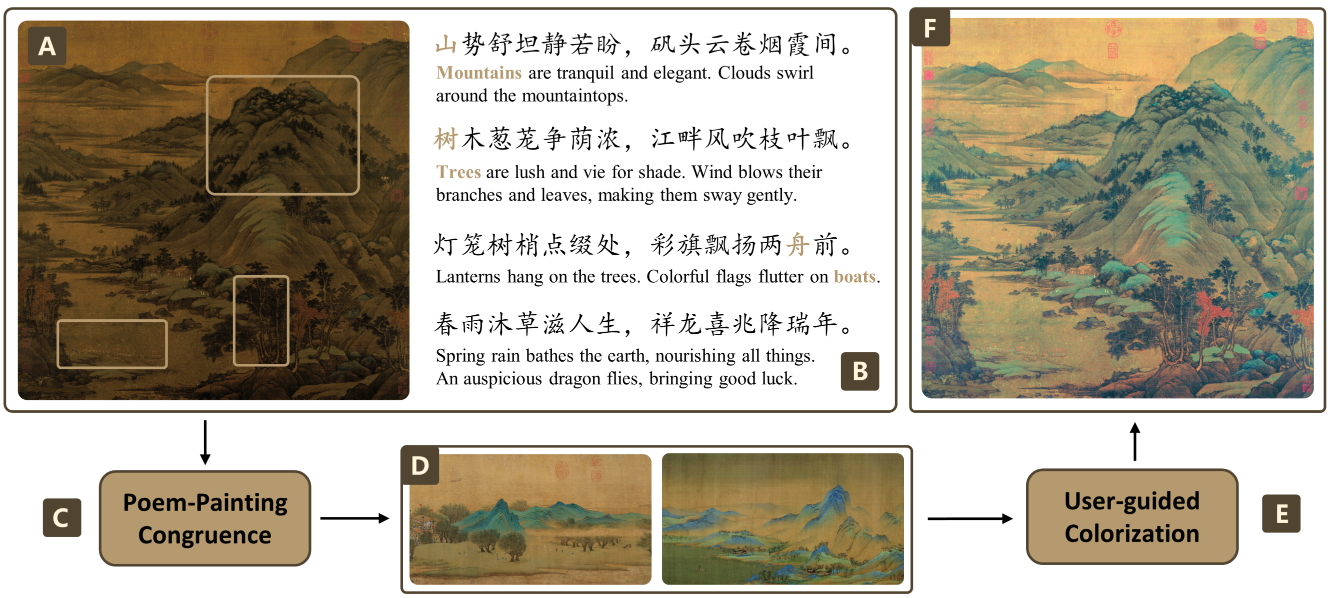

Figure 1: The illustration of color restoration using PColorizor: For a color-faded painting [84] (A), an ideorealm-congruent poem (B) is first provided as input into a poem-painting congruence module (C) to search for relevant reference paintings. Then, restorers utilize a color analysis module to find reliable color schemes (D) that have been applied to depict similar ideorealm in the reference paintings [74, 82]. Moreover, a user-guided colorization module (E) is employed to assist restorers in trial-and-error colorization. Finally, the system outputs a colorful revision (F) of the painting that can be used to direct physical restoration.

ABSTRACT

Color restoration of ancient Chinese paintings plays a significant role in Chinese culture protection and inheritance. However, traditional color restoration is challenging and time-consuming because it requires professional restorers to conduct detailed literature reviews on numerous paintings for reference colors. After that, they have to fill in the inferred colors on the painting manually. In this paper, we present PColorizor, an interactive system that integrates advanced deep-learning models and novel visualizations to ease the difficulties of color restoration. PColorizor is established on the principle of poem-painting congruence. Given a color-faded painting, we employ both explicit and implicit color guidance implied by ideorealm-congruent poems to associate reference paintings. We

*Corresponding author: Yingcai Wu, ycwu@zju.edu.cn

UIST '23, October 29-November 1, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0132-0/23/10…$15.00
https://doi.org/10.1145/3586183.3606814

propose a mountain-like visualization to facilitate efficient navigation of the color schemes extracted from the reference paintings. This visual representation allows users to easily see the color distribution over time at both the ideorealm and imagery levels. Moreover, we demonstrate the ideorealm understood by deep learning models through visualizations to bridge the communication gap between human restorers and deep learning models. We also adopt intelligent color-filling techniques to accelerate manual color restoration further. To evaluate PColorizor, we collaborate with domain experts to conduct two case studies to collect their feedback. The results suggest that PColorizor could be beneficial in enabling the effective restoration of color-faded paintings.

CCS CONCEPTS

• **Human-centered computing → Visualization systems and tools**.

KEYWORDS

Ancient Chinese Art; Color Restoration; Visual Analytic

ACM Reference Format:
Tan Tang, Yanhong Wu, Peiquan Xia, Wange Wu, Xiaosong Wang, and Yingcai Wu. 2023. PColorizor: Re-coloring Ancient Chinese Paintings with Ideorealm-congruent Poems. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), October 29-November 1, 2023, San Francisco, CA, USA*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3586183.3606814

1 INTRODUCTION

Ancient Chinese paintings are both witnesses of Chinese history and inheritors of Chinese culture. Similar to other Chinese antiquities, most ancient Chinese paintings have shown varying degrees of color fading due to improper preservation. This problem makes it difficult for viewers to resonate with the conceptual feelings conveyed by the paintings. These conceptual feelings are referred to as ideorealm. Color restoration aims to extend the life of ancient Chinese paintings, which has more than 1,600 years of history. For a color-faded painting, restorers need to sort through a wealth of documents and images with the intention of inferring a faithful interpretation of the original colors. In addition, modern restorers perform numerous trial-and-error attempts to achieve satisfactory coloring results on the digital paintings. This challenging and tedious process usually takes months, even years, to run.

The recent advances in deep learning have put a spotlight on re-coloring ancient Chinese paintings. A plethora of automatic methods have been adopted for various image-to-image translation tasks, such as colorization and style transfer. Moreover, automatic methods with limited interactions are also proposed to enable controllable image translations. However, the "black box" problem of deep learning models produces an obstacle to understanding color inference, which is essential for human restorers to evaluate the reliability of coloring results.

To infer faithful colors, we focus on associating ancient Chinese paintings with ideorealm-congruent poems. We argue that the ideorealm implied by poems are important cues that could guide the color restoration for ancient paintings. First, Chinese poems and paintings are regarded as two congruent forms of expression for conveying the emotions and ideas of their authors. Generally, Chinese painters are enthusiastic about inscribing poems on paintings to highlight the congruent ideorealm indicated by visual and textual imageries. Such cross-modal congruence is revealed by the principle of "congruence between poems and paintings," proposed by a famous ancient Chinese poet (also painter) Dongpo Su. Second, color fading leads to severe information loss, which produces gaps for human restorers to find relevant paintings for reference. Unlike visual information, textual poems have been better preserved, which can serve as essential tools to mitigate information loss and to identify reference paintings.

However, developing such a mixed-initiative method comes with two major challenges: *(1) Associating poems and paintings using congruent ideorealm.* Ancient Chinese poems are composed of high-level abstract words that differ from the natural language used in modern Chinese, which produces obstacles for employing existing cross-modal models to associate textual poems and paintings. Learning such hidden representations across modalities constitutes the first challenge, which is still under-explored. *(2) Re-coloring paintings with large-scale color references.* Ancient artists created a lot of paintings during the long history of Chinese art. There can be many associated paintings to consider when recoloring an ancient painting. Interactive visualizations play an important role in bridging the gaps in human-AI collaboration. To facilitate the co-coloring of faded paintings, it is essential to overcome the difficulties of visualizing large-scale color collections, which poses the second challenge.

To overcome the above challenges, we propose PColorizor, an interactive system that enables deep collaboration between restorers and deep learning models. We first leverage CLIP (Contrastive Language-Image Pre-training) [45] to learn poem-painting congruence due to its powerful strength in cross-modal alignment learning. We then propose a visual analytic approach to assist restorers in obtaining accurate color schemes. Based on this approach, we develop an interface that consists of three views: (a) a board view that supports restorers to construct flexible queries of different imageries; (b) a timeline view that allows restorers to understand the color distribution and evolution of reference paintings; (c) a reference view that explores the ideorealm understood by CLIP and provides a window to display the reference painting. Moreover, we adopt an auto-coloring model to support colorization previews with easy-to-use interactions. We finally conducted two case studies to evaluate PColorizer, which suggests that restorers can improve the efficiency of color restoration and discover new color insights.

The major contributions of our study are summarized as follows:

- We propose PColorizor, a novel **interactive system** that leverages restorers with deep learning models to facilitate easy color restoration for ancient Chinese paintings.
- We propose a novel **visual analytic approach** that extracts implicit color references from large-scale paintings based on the principle of congruence between paintings and poems.
- We conduct two **case studies** to demonstrate the usage of PColorizer and collaborate with domain experts to evaluate its effectiveness.

2 RELATED WORK

This section discusses three relevant topics to our study: image colorization, color perception, and color design tools.

2.1 Image Colorization

As one of the most classic image-to-image translation tasks, image colorization, which aims at adding color to grayscale images, has received considerable attention from researchers. In general, existing image colorization methods can be divided into two categories: automatic colorization [9, 23, 50, 68, 86] and user-guided semi-automatic colorization [17, 29, 87].

Automatic colorization methods utilize deep learning models to learn color inference from large training data. Cheng et al. [9] first employed a fully connected neural network, which receives three-level feature descriptors for each pixel in a grayscale image and outputs the predicted chrominance values. To overcome the desaturated colorization caused by standard regression loss, Zhang et al. [86] formulated image colorization as a multinomial classification problem with a convolutional neural network. However, these pixel-independent colorization methods may fail to consider spatial relations of pixels because they overlook the context of the pixel in the image. Therefore, autoregression models such as PixColor [15] and ColTran [23] have been proposed to model both spatial and semantic interrelationships between pixels at low and high resolution. Meanwhile, Su et al. [61] further extracted instances from the image and fused their colorization results to reduce color bleeding artifacts. In addition, advanced generative models such as GAN [68] and diffusion models [50] have also been employed to achieve plausible image colorization.

Compared to the described end-to-end methods, user-guided semi-automatic colorization methods assume that users play a significant role in reducing coloring uncertainty and artifacts. According to the type of user interaction, these methods can be classified as scribble- [21, 25, 43, 73, 87], example- [16, 17, 49, 64, 72], and text-based colorization [1, 5, 29, 71, 90].

Scribble-based colorization methods propagate user-input color hints from scribbles. To determine the propagation direction, early methods mainly utilize the similarity measures of low-level image features such as luminance [25, 78] and texture [28, 43]. However, these methods often require intensive and reliable user intervention to achieve remarkable colorization results. To reduce the user burden, Zhang et al. [87] proposed a deep learning approach that allows users to control the colorization through sparse color hints. Kim et al. [21] further used scribbles to solve the color bleeding problem at weak grayscale contrast or along the edges inside objects. Considering that the color hints at different positions may require different colorization efforts, Xia et al. [73] located the most representative color anchors and offered them to users as hint position guidance.

Example-based colorization methods aim to transfer the color of a reference image to a target image. Pioneering methods [49, 70] employed pixel-to-pixel matching based on low-level image features, which is time-consuming and inaccurate. To optimize the transfer process, subsequent methods have used semantic segmentation [20, 64], multiple references [33, 89], and superpixel computation [16]. However, the results of these methods still rely heavily on the quality of the reference image, which should have similar objects, illumination, and viewpoint to the target image. To alleviate this problem, He et al. [17] first presented a deep learning-based similarity subnetwork to detect semantically congruent regions in the reference/target image for direct color transfer. In addition, they also developed another colorization subnetwork to predict the pixel color of the regions without appropriate matches. Xu et al. [75] and Wu et al. [72] further facilitated such two stages to produce more robust and vivid colorization results.

Text-based colorization methods refer to extracting explicit [5, 22, 29, 71, 90] or implicit [1] color guidance from natural language and applying it to the input image. Concretely, explicit color guidance typically includes apparent objects and their corresponding colors, such as *a green bear toy* and *a purple umbrella*. However, these methods are limited to specific textual descriptions of color, which cannot cover the entire color space. Instead, implicit color guidance comes from understanding abstract language such as *sunny* and *frozen*. Specifically, Bahng et al. [1] proposed Text2Colors, which first predicts multiple color palettes for the text and then colors the image with a palette chosen by users.

Recently, Huang et al. [19] presented UniColor, a unified framework that first integrates the above three types of user interaction by transforming them into color hints represented by mathematical vectors. In this study, we follow the basic idea of UniColor for its comprehensive support of user interaction. In addition, we also utilize several traditional image processing techniques [16, 28] to deal with the colorization problem caused by the lack of paintings with unusual styles.

2.2 Color Perception and Design Tools

Color is the fundamental element that makes up the world as perceived by humans. Previous psychological work has extensively explored areas such as color preference [35, 36, 53] and color-concept association [18, 46, 52]. Maule et al. [30] pointed out that human perception and cognition of color are influenced by factors such as biology, experience, culture, and environment. Moreover, the use of color has also been extensively discussed in various design-related fields [55, 63, 66]. For example, the colormap problem has been well studied in the past 20 years [3, 27, 47, 48, 51], which contributes various approaches for visualizing scalar data. Specifically, Borland et al. [3] first summarized three problems with common rainbow colormaps: confusing the viewer, obscuring the data, and actively misleading interpretation. To obtain more reasonable design principles for colormaps, Schloss et al. [51] conducted a comprehensive comparison of user performance on a color-quantity mapping task using colormaps with different background colors. They found that the effect of the background color on this task depends on whether the colormap uses a color scheme with different opacities. Furthermore, Reda et al. [48] demonstrated that colormaps with high color categorization can improve user performance in cognitive tasks even though they are considered unsuitable for perceptual tasks.

In addition to the theoretical research on color, a large number of color design tools are developed for the efficient extraction and application of color schemes [6, 41, 57, 58, 60, 76]. For example, Phan et al. [41] constructed a continuous design space based on art collections and presented an interpolation method based on density estimation to generate infinite color schemes. Shugrina et al. [58] unified all tasks involved in creating color schemes through an interactive interface to improve the efficiency of color designs. Yan et al. [76] developed FlatMagic to assist professionals in fast flat colorization for digital comics. De-Stijl [57] was developed by considering the important role of color proportions and placement in designing color schemes.

For artworks such as ancient Chinese paintings, viewers pay attention to both the intuitive feelings conveyed by colors and the conceptual feelings implied by painters. The scientific color theory based on intuitive feelings is sometimes not applicable to the perception of art [36]. For example, ancient Chinese painters may deliberately employ bright colors to hide negative emotions. Such contrast is one of the ancient Chinese art styles that could emphasize painters' emotions (see Case 1 in Section 7.1). In this study, we leverage the congruence between ancient Chinese paintings and poems to learn the deep alignment of their conceptual feelings (see Section 3.1). We also build an interactive system that seamlessly integrates visualization and image colorization techniques to improve the color restoration workflow. This system allows restorers to flexibly explore the relevant color space and efficiently preview different color schemes.

3 DOMAIN ABSTRACT

In this section, we first introduce the congruence principle of ancient Chinese paintings and poems. We then provide a summary of the three stages involved in the traditional color restoration workflow through interviewing two domain experts. They are highly experienced senior artists with over a decade of expertise in studying ancient Chinese art. Finally, we propose a set of user requirements to guide the design and development of PColorizor.

3.1 Congruence between Paintings and Poems

The congruence principle of ancient Chinese paintings and poems was proposed by Shi Su, one of the most famous painters and poets of the North Song Dynasty of China [56]. This principle suggests that both painting and poem are artistic forms for literati to borrow scenery to express emotions [85]. The congruence principle has been extensively studied in various fields [4, 40, 67], yet its application to ancient Chinese art remains underexplored. A specific manifestation of this principle is that paintings and poems often use the same objects to depict the scenery and convey congruent conceptual feelings. In Chinese culture, these objects and conceptual feelings are called *imagery* and *ideorealm*, respectively [54]. For example, the painting and the poem in Fig. 1 contains three categories of imagery (Fig. 1A): tranquil mountains, lush trees, and boats decorated with colorful flags. These imageries create a scenery of people celebrating festivals on the outskirts, which suggests the painter's wish for a prosperous and peaceful world. Western artists have also expressed similar views. For example, the 5th-century ancient Greek poet Simonides said [24], *"Painting is mute poetry, poetry is a speaking picture."* Influenced by the congruence principle, activities such as "writing poems for paintings" and "drawing paintings according to poems" were widespread among later Chinese literati groups [39]. The poems produced by these activities, called *ideorealm-congruent poems*, are either inscribed directly on the paintings or recorded in books such as Xuanhe Huapu [26]. Even though some paintings have disappeared, the associated poems have survived, such as Shi Su's "Hui Chong Chunjiang Wanjing" [83]. To learn poem-painting congruence, we develop a cross-modal model that supports color restoration by easing the difficulty of searching for reference paintings.

3.2 Virtual Color Restoration and Expert Interview

Color reduction refers to the phenomenon of color changes (e.g., hue, saturation, or lightness) that have occurred to the painting over time [2]. To solve this problem, traditional color restoration relies on professional painters to manually restore the color-faded paintings [12]. However, such a process is not only time-consuming and laborious but also causes physical damage to the original paintings [14]. Therefore, virtual color restoration based on digital images has rapidly developed over the past 20 years [38, 69, 91]. The digitally restored version of the paintings can serve as a reliable reference for physical restoration. To gain a comprehensive understanding of the current state of the color restoration domain, we organized weekly one-hour offline meetings with these two restorers for a period of two months. Based on the meeting content, we abstracted the three stages of the color restoration workflow and summarized the difficulties.

Collection of reference paintings. The core of color restoration is to find faded colors from other well-preserved paintings. However, this work can be difficult even for experienced restorers because they can not remember all the color schemes used in ancient Chinese paintings. Therefore, restorers first need to analyze the color-faded painting, such as imagery categories, painting techniques, and the painter's background. Then, they need to collect relevant paintings based on these characteristics for reference. The collection stage is often time-consuming due to the huge search space and complicated painting analysis. Content-based image search methods have been applied to improve the efficiency of collection. However, due to the incomplete color information, the result of these search methods usually involves plenty of irrelevant paintings, which increases the burden of reviewing work.

Selection of color schemes. Different color schemes may exist in the collected reference paintings. Restorers need to filter out the color schemes that disobey the basic principles of color restoration. First, the color schemes should not conflict with the ideorealm presented by paintings. Second, the color schemes should follow the historical evolution of coloring skills adopted by painters across different periods. The newly-founded color schemes should not appear in the paintings created during older periods. Restorers have to judge whether a color scheme satisfies the two principles on their own knowledge, which is also difficult and challenging.

Colorization of faded paintings. During the colorization stage, restorers need to outline unclear lines first and then colorize the imageries according to the verified color schemes. As each imagery in a painting may have multiple verified color schemes, restorers must conduct extensive trial-and-error experiments to find the optimal combination of these color schemes to achieve the overall harmony of the painting. This process is repetitive and laborious for restorers. Advanced image editing software (e.g., Adobe Photoshop and Affinity Photo) has offered powerful tools (e.g., the Photoshop color match tool [42]) to modify colors efficiently. However, these software programs are designed for general image editing tasks and do not cater to virtual color restoration.

3.3 User Requirements

Based on the interviews with our experts, we further summarized a list of user requirements, where **R1**-**R3**, **R4**-**R5**, and **R6** correspond to the three stages of virtual color restoration in sequence. The requirements outline what users expect from our system when interacting with it. To collect relevant paintings, users first need to formulate queries that clearly communicate what they are looking for in order for the system to search accordingly (**R1**). Moreover, the search results should be directly visualized (**R2**), and the poem-painting congruence should be revealed for users (**R3**). To select color schemes, users require a comprehensive understanding of how color schemes have evolved over time and investigate the system-generated insights for deep color analysis (**R5**). Based on the chosen color schemes, users need to preview colorization results through trial-and-error experimentation (**R6**).

- **R1 Construction of search queries.** Each restorer has their own knowledge about the color restoration of ancient Chinese paintings, which leads to various intentions for the reference painting search. Thus, the interface should provide restorers with tools to construct the search queries flexibly.
- **R2 Awareness of search results.** The matching score of reference paintings directly reflects their congruence to the queried imageries of the color-faded paintings, which is necessary for restorers to find valuable paintings efficiently. Moreover, for the huge color space involved in the search results, an overview of these colors is beneficial for restorers to select paintings that contain specific colors quickly.
- **R3 Alignment of painting imageries.** Since the color schemes used for restoration are extracted based on imagery units, the reference paintings that match more queried imageries tend to contain more intensive and effective color schemes. Besides, reference paintings may share the same color schemes for some imageries. Such relationships should be effectively visualized to demonstrate the frequency of different color schemes.
- **R4 Understanding of color evolution.** In addition to following the congruence principle, the color schemes used for restoration must conform to the historical evolution of colors (as mentioned in Section 3.2). Restorers believe that demonstrating the temporal changes of color schemes can not only assist in narrowing down the search range for reference colors but also offer reliable guidance based on color statistics.
- **R5 Investigation of model insights.** The insights from deep learning models are not always consistent with user recognition. Restorers are interested in the reasoning behind the behavior of deep learning models, especially when the models hold insights contrary to their expertise.
- **R6 Trial-and-error colorization.** In the general workflow of colorization, restorers always re-color faded paintings while reviewing the reference. The determination of the final color schemes requires restorers to constantly try various color schemes and adjust the detail of local color transfer. Therefore, the interface needs to integrate semi-automatic colorization methods to accelerate the colorization process.

4 SYSTEM OVERVIEW

In response to the user requirements (Section 3.3), we propose PColorizor, a novel interactive system to assist restorers in the efficient restoration of color-faded paintings. Figure 2 demonstrates the modules involved in the system: *Query Construction* module provides image and text selection tools in a painting board (**R1**). *Poem-Painting Congruence* module utilizes the congruence between poems and paintings to retrieve reference paintings from a painting database (**R1**). *Color Analysis* module designs mountain metaphors and a hierarchical bar chart to visualize the congruence scores (**R2**) and temporal distribution (**R4**) of the reference paintings. It also integrates a glyph-based scatterplot to show the relationships between the paintings (**R3**). *Ideorealm Analysis* module shows how the deep learning model understands the congruence principle of poems and paintings (**R5**). *User-guided Colorization* module employs a semi-automatic colorization model to facilitate the colorization process (**R6**). We implemented the system as a web application using React as the front-end framework, Flask in Python as the back-end framework, and MySQL as the database platform.

5 ALGORITHM DESIGN

In this section, we introduce the models used in the poem-painting congruence module and the user-guided colorization module. For the poem-painting congruence module, we employ CLIP [45] for its powerful learning ability of image-text representation alignment (**R1**). For the user-guided colorization module, we employ a combination of UniColor [19] and general image colorization methods [16, 28] that do not require any data training to overcome the lack of paintings with unusual styles (**R6**). To train or fine-tune these two models, we constructed a dataset containing more than 11,000 pairs of paintings and poems from the Internet. Specifically, we first defined a set of keywords such as "ancient Chinese paintings with inscribed poems" and "paintings in Xuanhe Huapu". Based on these keywords, we wrote Python scripts to automatically collect painting images and text descriptions in their source web pages via the Baidu search engine. For the text descriptions of each painting, we then used NLP techniques such as text classification [62] to extract poems or other sentences that represent the ideorealm of the painting. Finally, we manually checked partial pairs of painting images and texts to improve the quality of this dataset further.

5.1 Poem-Painting Congruence

CLIP [45] is a famous vision-language model pre-trained on 400 million image-text pairs on the web. This model has been widely applied in image classification [45] and generation tasks [10, 37, 88] due to its excellent performance on the similarity calculation between images and texts. In this work, we employ a Chinese version of CLIP [77] to measure the congruence between paintings and poems in accordance with the above reason. Considering that the original training data of CLIP rarely contains our post-processing painting-poem pairs, we fine-tune CLIP using our dataset. For convenience, we denote a painting as a and a poem

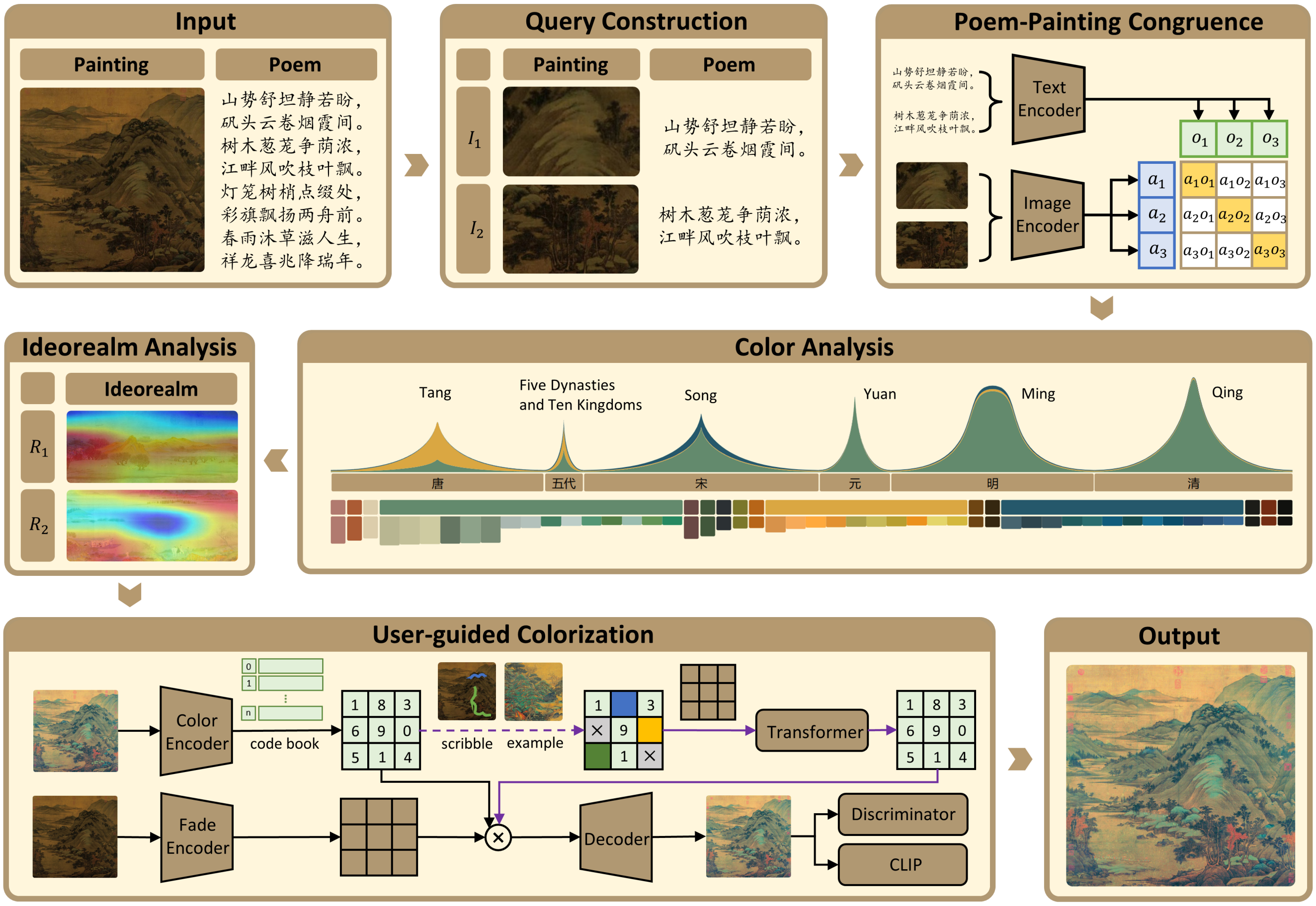

Figure 2: The system overview of PColorizor: For a color-faded painting and its ideorealm-congruent poem, the query construction module provides users with image and text selection tools to construct flexible queries. The poem-painting congruence module is then applied to search reference paintings. For the recalled reference paintings, users can identify reliable color schemes and verify their applicability using the color analysis module. The ideorealm analysis module further demonstrates how the model understands the congruence principle. For the verified color schemes [59], users can conduct trial-and-error colorization with the user-guided colorization module to restore the color-faded painting.

as o. Given a batch of paired paintings and poems $PP_{batch} = \{\{a_1, o_1\}, \{a_2, o_2\}, ..., \{a_n, o_n\}\}$, CLIP uses an image encoder E_{image} and a text encoder E_{text} to project them into a joint embedding space, in which the projected paintings and poems are represented as $PP_{embed} = \{\{e_1^a, e_1^o\}, \{e_2^a, e_2^o\}, ..., \{e_n^a, e_n^o\}\}$. Specially, $\{e_i^a | i \in \{0, n\}\}$ and $\{e_j^o | j \in \{0, n\}\}$ are both embedding vectors with same dimension d. To construct the self-supervised contrastive loss in CLIP, paired paintings and poems $\{e_i^a, e_j^o\}_{i=j}$ are labeled as positive samples, while unpaired paintings and poems in the batch $\{e_i^a, e_j^o\}_{i \neq j}$ are labeled as negative samples. Then, such contrastive loss can be defined as [45]:

$$L_{CLIP-train} = min(\sum_i^n \sum_j^n (e_i^a \cdot e_j^o)_{i \neq j} - \sum_i^n (e_i^a \cdot e_i^o)) \quad (1)$$

where $\cdot$ refers to cosine similarity. Fig. 2A shows a schematic diagram of the model architecture of CLIP. For the fine-tuned model, we show how it understands three different types of ideorealm in Case 1 and the Appendix to demonstrate its performance.

5.2 User-guided Colorization

As the main component of the user-guided colorization module, UniColor [19] supports both unconditional and conditional colorization, which facilitates the collaboration between restorers and deep learning models. As shown in Fig. 2C, there are two subnetworks in UniColor, namely *Chroma-VQGAN* (black) and *Hybrid-Transformer* (purple). The Chroma-VQGAN is a variant of VQGAN [13] that quantizes the latent representation of images according to a learnable discrete codebook to obtain a robust image reconstruction capability. Apart from the basic architecture of VQGAN, the Chroma-VQGAN involves a grey encoder to provide the reconstruction decoder with the structure information of the input image. Considering that the color-fading paintings may contain partial valuable color inference, we replace the grey encoder with a fade encoder that shares the same architecture with the color encoder to integrate such information in reconstruction. Furthermore, we adopt an extra CLIP loss [10, 88] to encourage the reconstructed image to have congruent ideorealm with the poem. Therefore, the total loss we used for training Chroma-VQGAN is defined as:

$$L_{Chroma-VQGAN} = L_{rec} + L_{VQ} + L_{GAN} + L_{CLIP} \tag{2}$$

where L_{rec}, L_{VQ}, L_{GAN} represent reconstruction loss, vector quantized loss, and adversarial loss respectively [13, 19]. Owing to the fact that the Chroma-VQGAN requires paired color-fading and color-preserved paintings as training data, we utilize several automatic colorization models [23, 68] that are trained on the color-fading paintings to generate the color-fading version for each color-preserved painting in our dataset. For the Hybrid-Transformer that receives masked color tokens with color hints and color-fading painting features, we follow the basic user interaction conversion and training strategy presented in [19]. Such a subnetwork can finally output a complete quantized image representation demanded by the reconstruction decoder in the Chroma-VQGAN.

6 INTERFACE DESIGN

In this section, we introduce three views that constitute the visual interface of PColorizor: the board view (Sec. 6.1, Fig. 3A), the timeline view (Sec. 6.2, Fig. 3B), and the reference view (Sec. 6.3, Fig. 3C). The board view integrates the query construction module (**R1**) and the user-guided colorization module (**R6**). The timeline view visualizes the results of the color analysis module (**R2, R3, R4**). The reference view integrates the ideorealm analysis module (**R5**) and the user-guided colorization module (**R6**).

6.1 Board View

The board view (Fig. 3A) comprises a painting board and two collapsible panels at the right and bottom.

The painting board shows the input painting and provides the following four tools in the left toolbox. The lasso selection tool helps users draw and select imageries in the painting, which will be used for searching for reference paintings (**R1**). The color transfer tool automatically colorizes the selected imageries based on the reference painting with the algorithm detailed in subsection 5.2. The pencil and color extractor tools assist users in drawing scribbles on the painting, which provide additional guidance for the colorization algorithm, facilitating user-model collaboration (**R6**).

The right collapsible panel shows the input poem and allows users to click on multiple Chinese characters in the poem to assemble a textual description for the selected imagery. The right panel also includes a table that displays all imageries in the current query, allowing users to add or remove imageries. The bottom collapsible panel comprises a list of paintings that users save during the colorization task. Users can select two paintings to compare them side-by-side on the painting board.

6.2 Timeline View

To help users find accurate color schemes from the reference paintings, we design the timeline view (Fig. 3B) using three visualizations. First, mountain metaphors summarize the count and congruence score distribution of the paintings under each dynasty (**R2**) and the proportion of the color schemes used in them (**R4**). Second, glyph-based scatterplots demonstrate the similarity of the paintings about the imageries and the color schemes they contain (**R3**). Third, a hierarchical bar chart allows users to find a subset of the paintings with specific color schemes (**R2**).

Mountain metaphors (Fig. 3B-1). To reveal the evolution of color schemes over time (**R4**), we split the reference paintings in the search results by dynasty and visualize the paintings from each dynasty based on a mountain metaphor. Each mountain represents a dynasty. The width of a mountain encodes the length of the dynasty, and the height indicates the number of reference paintings belonging to that dynasty. The shape of the mountain is drawn with symmetrical cubic Bézier curves, the smoothness of which is controlled by the highest congruence score of the reference paintings (**R2**). Moreover, to visualize the color scheme for each dynasty (**R4**), we split the mountain vertically based on the distribution of the colors used in the paintings.

We employ the mountain metaphor for two reasons. First, a mountain has characteristics such as altitude, shape, and multiple layers, which are suitable for encoding various variables simultaneously. Second, restorers are familiar with these characteristics of the mountains because these characteristics are frequently involved in Chinese landscape paintings. Using this metaphor may reduce the cognitive burden of restorers when viewing the visualization.

Glyph-based scatterplots (Fig. 3B-3). To explore the relationships among the reference paintings (**R3**), users can click on a dynasty and turn the mountain metaphor into a glyph-based scatterplot. The scatterplot shows all reference paintings and each is represented by a glyph visualization. The glyph visualization comprises a stack of horizontal strips. Each strip represents an imagery in the painting, the congruence score of which is encoded with the strip's opacity. To lay out the glyph visualizations in the scatterplot with respect to the imagery and color scheme similarities of the paintings, we first construct an imagery vector for each painting describing the co-occurrence of imageries based on their congruence scores and a color scheme vector based on the color extraction method detailed in the following explanation of the hierarchical bar chart. Then, we employ t-SNE to project the imagery and color scheme vectors into the x and y dimensions of the glyph visualizations inside the scatterplot, respectively, such that the paintings that share the similar imageries are closer horizontally, and the paintings that share the similar color schemes are closer vertically.

Hierarchical bar chart (Fig. 3B-2). To explore different color schemes (**R2**) flexibly, we employ the hierarchical bar chart to provide users with an overview of the color space extracted from the reference paintings. To build this overview, we discretize the distribution of the original colors in the CIELAB space as follows. First, we divide the CIELAB space into 8,000 color subspaces evenly. We choose the CIELAB space because its color changes are most consistent with humans' color perception. Then, we extract the representative colors in each reference painting using the superpixel computation [16] to reduce the number of colors. Next, we count these representative colors in different color subspaces and construct a bar chart, where each bar represents a color subspace with at least one representative color, and the number of the representative colors in this subspace is encoded by the bar's height. Finally, considering the number of bars could be too large to fit in

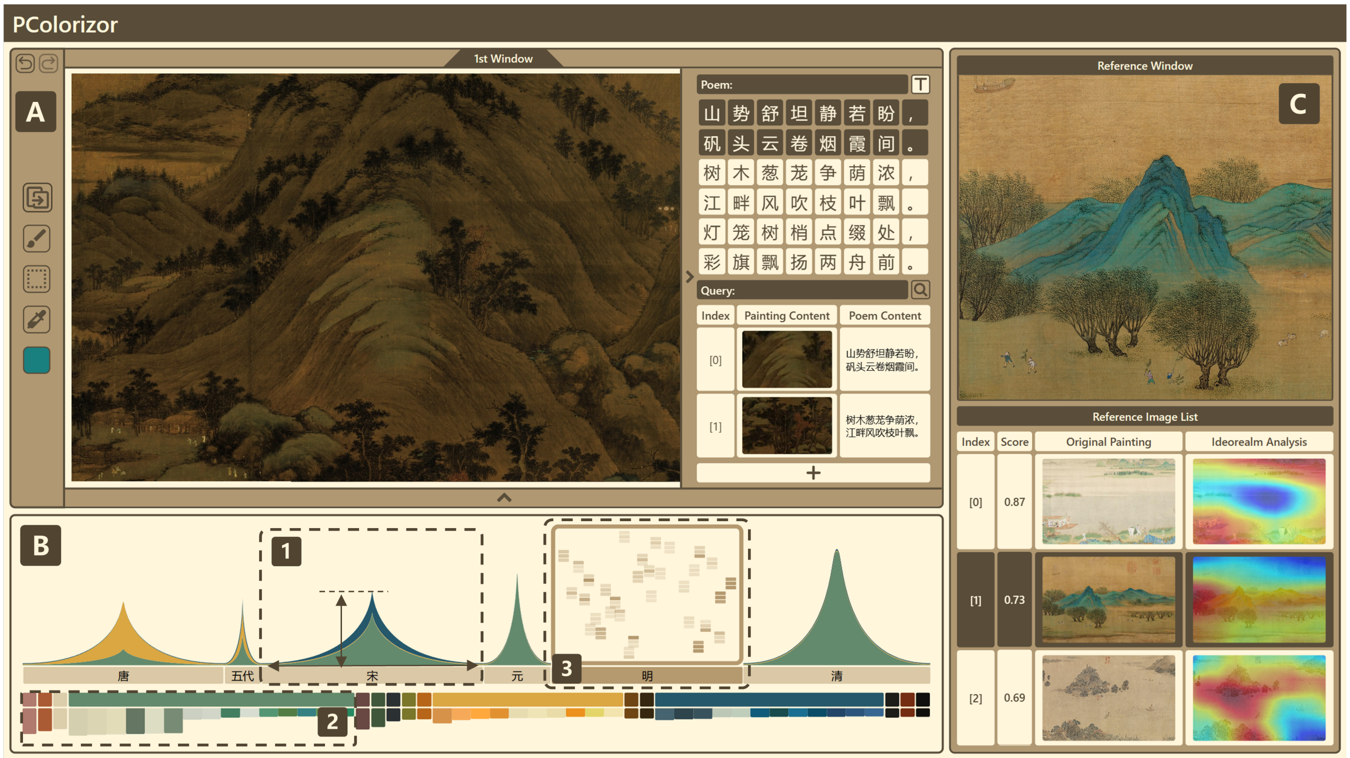

Figure 3: The interface of PColorizor: (A) The board view contains a painting board that provides four tools (from top to bottom are a color transfer tool, a pencil tool, a lasso tool, and a color extractor tool) and a right collapsible panel that includes a grid-based text component to help users construct the search query and modify the colorization results. (B) The timeline view is comprised of mountain metaphors and a hierarchical bar chart. The mountain metaphors (B-1) visualize the congruence scores and temporal distribution of color schemes. The hierarchical bar chart (B-2) summarizes the color space extracted from the reference paintings. Users can use the hierarchical bar chart to control the color scheme displayed in the mountain metaphors. The mountain metaphor can be switched to a glyph-based scatterplot (B-3) that reveals the similarity of the paintings. (C) The reference view illustrates the model insights about the congruence principle and provides users with a preview window to assist them in fast colorization [31, 74, 79].

the timeline view, we further employ k-means clustering on these bars with k set to 16 in this study. Based on the clustering results, we add k toggle buttons to the bar chart, which are used to expand or close the detailed bars in the clusters (Fig. 4B). It is worth noting that only the expanded clusters will be displayed in the mountain metaphors due to the limited display space of the timeline view.

Interactions. There are three interactions between the visualizations in the timeline view. First, users can click the dynasty button to switch between the mountain metaphor and the glyph-based scatterplot. Specifically, we set a minimum width for the scatterplot. When the width of the selected mountain metaphor is less than the minimum width, other mountain metaphors will be proportionally reduced to make space for the scatterplot. Second, users can click on the glyph in the scatterplot to quickly view the color schemes of its corresponding painting in the hierarchical bar chart and the color evolution in the mountain metaphors. Third, users can use the mouse's left button to click on the bar to overlook its corresponding color subspace (Fig. 4D). Users can also use the mouse's right button to click on the bar to focus on its corresponding color subspace, meaning that all other color subspaces are overlooked. The overlooked colors will not be included in generating the mountain metaphors and the glyph-based scatterplot.

6.3 Reference View

The reference view (Fig. 3C) includes a painting list and a preview window that allows users to inspect the details of the painting. To demonstrate the ideorealm understood by CLIP to users (**R5**), we provide a heatmap-based visualization for each painting in the list. The computation method of the heatmap follows the approach used in [19]. In the heatmap-based visualization, red covers the region with high congruence between the input painting and poem, while blue covers the region with low congruence. Based on the heatmap, users can learn which parts of the painting are considered by the model to be congruent with the ideorealm in the poem. Moreover, users can click on the painting to view its high-resolution details through interactions such as drag and zoom in the preview window. For the imagery that has verified color schemes, users can use the

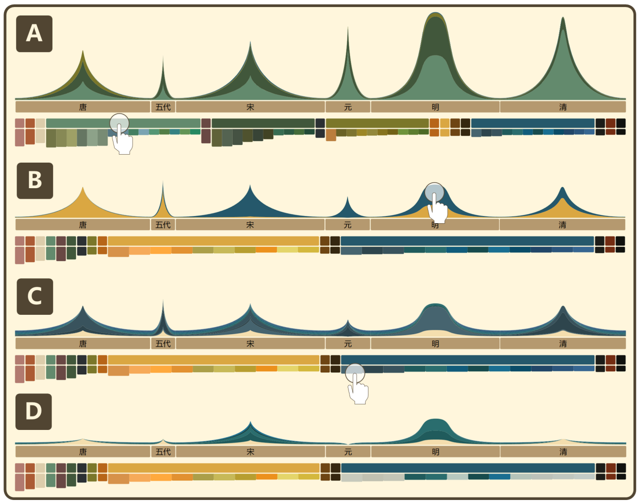

Figure 4: The illustration of the interactions between the hierarchical bar chart and the mountain metaphors: For an initial state of the mountain metaphors (A), users can click the toggle buttons above the bar chart to control the color subspace clusters displayed in the mountain metaphors (B). Users can also click the color regions in the mountain metaphors to view the detailed distribution of each color subspace in the cluster (C). Furthermore, users can click the bars to filter the color subspaces that they are not interested in (D).

color transfer tool in the painting board to select its corresponding image region for color transfer. Additionally, users can also use the color extractor tool to sample the color in the painting for subsequent scribble drawings.

7 CASE STUDIES

In this section, we present two case studies conducted by the two collaborative restorers (E1, E2). They are the same experts who were also interviewed in Section 3. In the first case, E1 used the poem-painting congruence module to search the reference paintings of *Dongting East Hill*. In the second case, E2 utilized PColorizor to restore *Residents on the Outskirts of Dragon Abode*. We also interviewed the two restorers to collect their feedback for subsequent system iteration.

7.1 Dongting East Hill

E1 was interested in the effectiveness of the poem-painting congruence module in PColorizor. To explore whether the module has learned the deep alignment between ancient Chinese paintings and poems, he chose a painting he greatly appreciates, namely "Dongting East Hill" [32] (Fig. 5A).

Background. The author of this painting is Mengfu Zhao, a descendant of the royal family of the Song Dynasty but serves as a senior official in the Yuan Dynasty due to pressure from the ruler of the Yuan Dynasty. This painting was created after the travel to Dongting East Hill, which depicts the beautiful scenery of that place using turquoise color schemes. However, according to the poem embedded in the painting (Fig. 5A), the author did not use

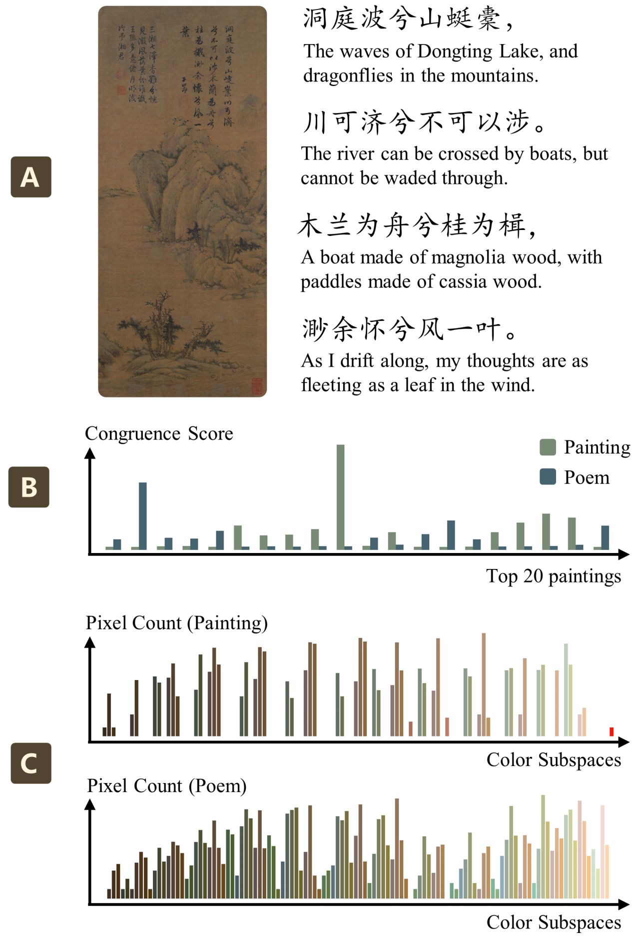

Figure 5: The illustration of Case 1 (Dongting East Hill): (A) The painting with an inscribed poem. (B) The score comparison of the reference paintings recalled by the painting and the poem, respectively. (C) The comparison of color distribution based on the top ten scoring paintings.

this scenery to express positive emotions. The true intention was to express his yearning for the Song Dynasty and his shame for serving as an official in the Yuan Dynasty. This implicit expression is known as "borrowing beautiful scenery to express negative emotions," which is an essential characteristic of Chinese culture. For example, the imagery "A boat made of magnolia wood" was first used by Yuan Qu, a famous patriotic poet from the Warring States Period, to remember his lost homeland.

Case Insights. Due to the color-fading problem, only faint traces of color can now be seen in this painting. Thus, E1 wanted to know if the poem-painting congruence module could learn the ideorealm to assist him in searching reference paintings. We compared the reference paintings searched by the color-faded painting and its ideorealm-congruent poem separately through a grouped bar chart (Fig. 5B). Concretely, we integrated the top ten scoring paintings from each modality and utilized the height of two adjacent bars to encode the congruence scores of each painting. We then placed them sequentially on the x-axis, where the order holds no significance.

Based on the grouped bar chart, E1 noticed that the two modalities exhibited apparent discrepancies in search preferences. E1 stated, *"It is interesting that none of the reference paintings gets high matching scores on both the content of the color-faded painting and the ideorealm of the poem."* To explore the reasons for the search preferences, we further visualized the color histograms of each modality's top ten scoring paintings (Fig. 5C) to reveal the distribution difference of colors involved in the search results. After viewing the details of the color histograms, E1 found that the color histogram constructed on the search results of the poem involves more color subspaces than that of the painting. Most of these additional color subspaces are similar to the turquoise color schemes, which means that the poem-painting congruence module appears to understand the ideorealm in the inscribed poem of "Dongting East Hill". It is also unsurprising that the painting modality cannot accurately convey this ideorealm because of the color-fading problem. After E1 confirmed the accurate color schemes contained in the search results of the poem, he concluded the exploration of the poem-painting congruence module.

To further demonstrate the ability of the poem-painting congruence module, we selected two other paintings and poems whose ideorealm expression methods are "borrowing beautiful scenery to express positive emotions" and "borrowing chaos scenery to express negative emotions," respectively. Like "Dongting East Hill," we utilized the color-faded painting and its ideorealm-congruent poem to search for reference paintings separately. We also visualized the color histograms of the search results to show the color preferences of the two modalities (see Appendix). The results suggest that the poem-painting congruence module can understand and distinguish the different types of ideorealm. After further discussion with E1, we considered that the poem-painting congruence module seems to have learned the different expression methods in poems for different ideorealm. For example, poems that use the "borrowing beautiful scenery to express negative emotions" method usually depict beautiful scenery first and then describe specific imageries to convey implicit emotions. This characteristic does not frequently appear in poems that use the "borrowing beautiful scenery to express positive emotions" method.

7.2 Residents on the Outskirts of Dragon Abode

In this case, E2 used PColorizer to restore one of the masterpieces created by the famous Chinese painter Yuan Dong from the Five Dynasties and Ten Kingdoms, titled "Residents on the Outskirts of Dragon Abode" [84] (Fig. 1A).

Background. This painting depicts a scene of people celebrating festivals on the outskirts of the capital city, created to celebrate the virtues and achievements of the country's ruler. As one of Yuan Dong's rare and vibrant landscape paintings and the pioneer of the Southern School of Chinese landscape painting, this artwork has been highly coveted by subsequent painters. This painting has been carefully collected and admired by celebrities such as the Qianlong Emperor, who also wrote numerous poems to praise its depicted ideorealm. However, due to over 1,000 years of age, it has suffered from severe color fading.

Case Insights. To conduct color restoration, E2 selected one poem with the most relevant descriptions of the painting's ideorealm as the input to PColorizor. After uploading the painting and poem, E2 first browsed the entire painting to understand the fading of different imageries. Then, E2 utilized the lasso tool on the painting board to crop the image regions related to the color-faded imageries. For each imagery with a rectangle region, E2 further clicked on the words and sentences describing the imagery in the grid-based text component. Finally, E2 constructed a query that consists of three imageries: a tranquil mountain, a lush tree, and a boat adorned with colorful flags. After E2 clicked the search button, the poem-painting congruence module returned reference paintings. A new visualization appeared in the timeline view (Fig. 6A). E2 quickly closed several unfolded clusters in the hierarchical bar chart and retained the clusters that included blue and green colors. He explained, *" I am only interested in the color schemes associated with blue and green colors because Yuan Dong is a landscape painter skilled in using shades of these colors."* Through viewing the mountain metaphors, E2 found that the mountain metaphor of the Ming Dynasty had the highest altitude and a relatively smooth peak, which implied that plenty of reference paintings existed in the Ming Dynasty, including some with high poem-painting congruence. Hence, E2 clicked on the dynasty button to further inspect these paintings.

In the glyph-based scatterplot, E2 discovered that three glyphs contained three opaque strips, which indicated their corresponding paintings have high congruence on the queried imageries (Fig. 6B). Furthermore, these glyphs were divided into two clusters whose vertical distance could not be ignored, which meant a noticeable color difference existed between the two clusters. By hovering over the glyphs, E2 learned the meta information of the painting and found that they were created by the same author, namely Ying Qiu, a painter from the Ming Dynasty. Considering Ying Qiu was renowned for his mastery of various painting techniques in past dynasties, E2 stated, "It is likely that Qiu Ying's paintings contain the correct colors, as he often used earlier coloring techniques."

To further check whether the color schemes used in these paintings conformed to the historical context of color evolution, E2 clicked on the glyph of the painting (see RP1 in Fig. 6B). For the color scheme displayed in the hierarchical bar chart, E2 then sequentially checked the temporal distribution of each specific color by utilizing the bar chart to control the color regions in the mountain metaphors. E2 wanted to ensure that these colors appeared before the Five Dynasties and Ten Kingdoms. After a period of exploration, E2 discovered that most of the colors had been employed in the Five Dynasties and Ten Kingdoms, except for a blue color that began to be used in the Song Dynasty (Fig. 6C). Based on this finding, E2 stated, *"Apparently, this color should be a variant developed in later dynasties. In the subsequent color restoration, I must avoid referencing the content involving this color."* Following the process above, E2 finally identified five high-quality reference paintings and added them to the list in the reference view.

E2 wondered how the deep learning model aligned the ideorealm between the painting and the poem. Therefore, E2 viewed the heatmap-based visualization for each reference painting and found that although the model had a relatively accurate perception of

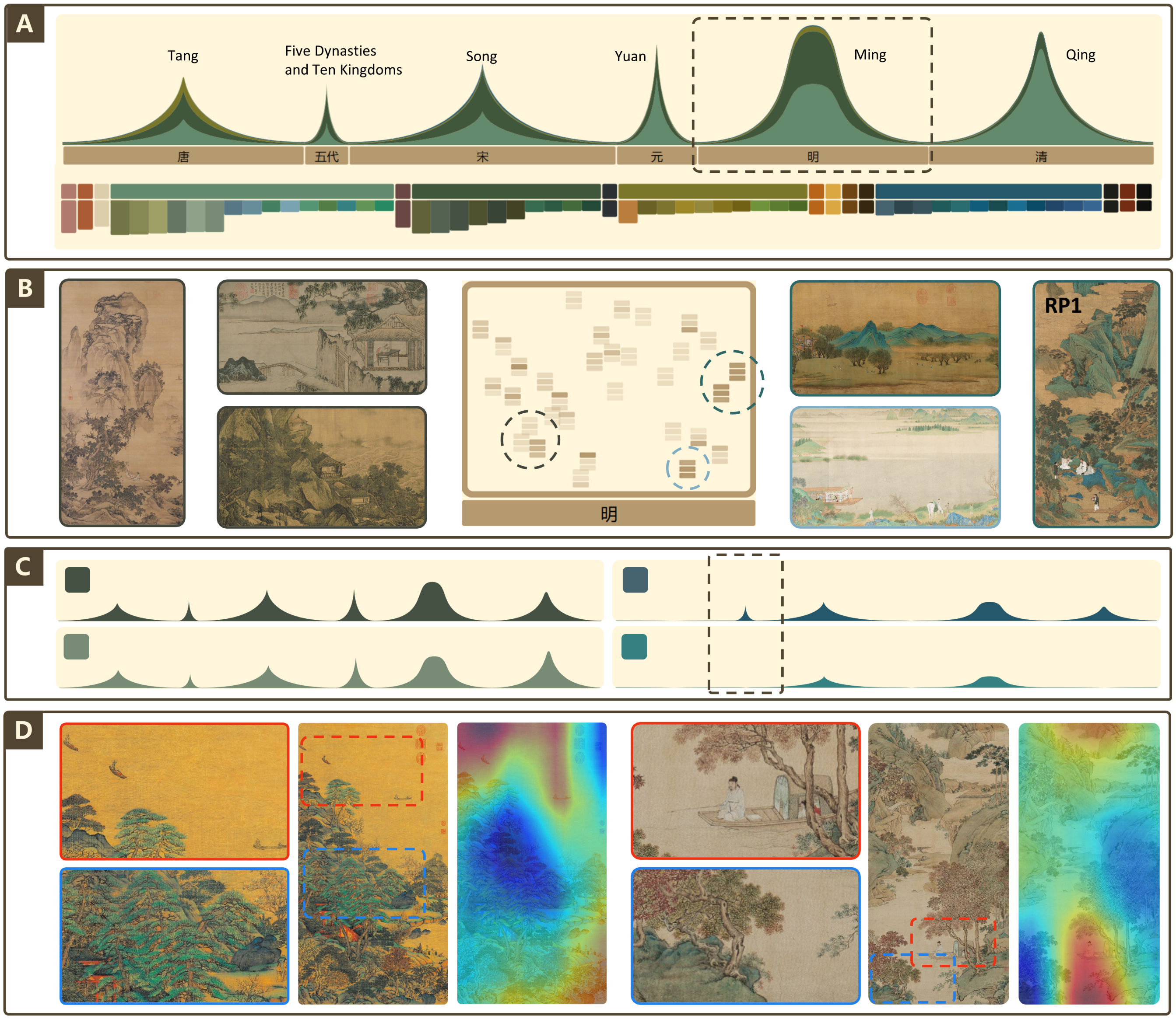

Figure 6: The illustration of Case 2 (Residents on the Outskirts of Dragon Abode): (A) The temporal distribution of the reference paintings that contain green and blue colors. The mountain metaphor that represents the Ming Dynasty has the highest altitude and a relatively smooth peak, which means the Ming Dynasty has the largest number of reference paintings that also obtained high congruence scores. (B) The imagery and color scheme similarity of the reference paintings [7, 8, 11, 74, 79, 81] in the Ming Dynasty. Two painting clusters on the right show the close horizontal position and distant vertical position, which means that the two clusters share similar imageries with different color schemes. (C) The evolution of the primary colors in RP1 [81]. Only one blue color has not appeared in the Five Dynasties and Ten Kingdoms. (D) Two examples of the model biases [59, 80]. The image regions covered by red and blue are respectively relevant to the imagery "people sitting on the boat" and the imagery "green mountains and trees," which means that the former imagery is more congruent to the ideorealm than the latter imagery.

the overall ideorealm, it exhibited obvious biases on specific imageries. For example, E2 discovered that the model believed that "people sitting on the boat" were more in line with the congruence of the ideorealm than "green mountains and trees" (Fig. 6D). E2 stated, *"This discovery is very interesting. It can explain some puzzles I encountered when finding reference paintings and improve my exploration efficiency."* E2 then clicked on the index button of one painting to display it on the preview window. After identifying the reliable imageries in this painting, E2 utilized the color transfer tool on the painting board to select the target and source image regions for fast colorization. Moreover, E2 also used the color extractor and pencil tools on the painting board to further adjust the colorization results. After several iterations, E2 exported the current restoration version for subsequent restoration work.

We interviewed E2 to collect expert opinions about the system. Overall, E2 had positive attitudes on PColorizor,

> *"It plays a positive role in finding reference paintings and completing colorization. First, it opened a new idea for me to search for reference paintings based on the congruence between the poem and the painting. Second, the timeline view also provides a meaningful summarization of the color insights of the paintings, which shortens the time of manually reviewing. Moreover, I appreciate the working experience of restoring a painting with the automatic colorization model, which allows me to avoid the high cost of trial-and-error colorization."*

Furthermore, he also provided some improvement suggestions for PColorizor. 1) The dynasties in the timeline view can be divided into more fine-grained periods. For instance, the Song Dynasty can be subdivided into the Northern and Southern Song Dynasties. Such fine-grained dynasty presentations can help restorers judge historical facts more accurately. 2) The heatmap-based visualization in the reference view should also be displayed in the preview window for viewing more details of the model insights.

8 DISCUSSION

This section presents the benefits, generalizability, and limitations of PColorizor, which is also guidance for future work.

8.1 Benefits

Ancient paintings are valuable "photos" that depict human histories vividly, which are also vulnerable with the passage of time. Virtual color restoration extends the lives of ancient paintings but is hindered by the low efficiency of manual colorization. PColorizor aims to mitigate the race against time by leveraging AI assistance to support human restorers. On the one hand, we employ the CLIP model to establish a bridge between paintings and poems, which supports a visualization and exploration of possible color references extracted from the cross-modal information. On the other hand, we adopt the auto-coloring model and develop a set of interactions to involve restorers in the trial-and-error process of colorization previews. PColorizor not only support painting colorization but also provides color knowledge extracted from ancient Chinese paintings, which may serve as an educational tool for junior restorers.

The missing of high-fidelity references and lack of high-quality restored examples are common obstacles to restoring ancient paintings. To overcome these obstacles, we bridge the gaps between Chinese poems and paintings based on the principle of poem-painting congruence. PColorizor put a spotlight on cross-modal models (e.g., CLIP) that implement the congruence principle, which offers a promising avenue for addressing other issues in ancient paintings, such as content missing.

8.2 Generalizability

Our research primarily focuses on examining the congruence between ancient Chinese paintings and poems. However, we also suspect that this principle may be applicable to other forms of Chinese art, such as porcelain and architecture, due to the central role that poetry plays in traditional Chinese culture. Additionally, Western art has explored the concept of congruence between paintings and poetry [24]. Therefore, we believe that this principle could have significant implications for fields such as cultural heritage restoration and recreation. PColorizor's modules can also be expanded to more general domains. For instance, during cross-modal alignment training, the poem-painting congruence module can accept text inputs in various formats. Therefore, for paintings that lack ideorealm-congruent poems, alternative textual descriptions can be used to execute this module. Additionally, the color analysis module can also be applied to visualize color information from other sources.

8.3 Limitations and Future Work

We list the limitation and future work of our study as follows.

Data. Data serves as a crucial foundation for artificial intelligence. Recent impressive models, such as ChatGPT [34] and CLIP, are trained using vast amounts of data. Nevertheless, the limited scale of available data restricts the types of automation that can occur. For example, we want to further improve the poem-painting module to support more diverse types of idearealm, but encountered difficulties due to overfitting caused by a limited dataset. A promising solution lies in using color information from other ancient artifacts, such as porcelain, architecture, and clothing, which are all concrete artistic expressions deeply rooted in Chinese culture.

Model performance. Ancient Chinese paintings and poems contain diverse ideorealm. In addition to the three primary categories of ideorealm mentioned in Section 7.1, the poem-painting module needs to be further evaluated on other unusual ideorealm. Furthermore, even for the three primary categories of ideorealm, the poem-painting module can not guarantee that every high-scoring reference painting is correct. In this regard, more advanced cross-modal models and higher-quality datasets may improve the performance of the poem-painting module. Unlike the photos commonly used in image colorization, Chinese paintings often have ultra-high resolution and contain rich local color details. Learning from this image data type remains a challenge for current colorization models. In this work, we adopted a data augmentation strategy by capturing multiple local images through a sliding window. However, such a strategy is still unable to fundamentally solve the problem of the color detail loss caused by down-sampling. For this problem, Wang et al. [69] employed a crop-and-combine method, which may provide a feasible solution.

Visualization scalability. Although we have set two display scales for the mountain metaphors and the bar chart in the timeline view, it is still challenging to overcome perception limitations as the number of reference paintings increases. For example, visual occlusion may appear in the glyph-based scatterplot when too many glyphs are contained in the limited display space. To alleviate this problem, more scalable visualization methods [44, 65] can be employed in future work.

LLM-based color restoration. The revolution triggered by LLMs (Large Language Models) such as ChatGPT [34] has affected numerous industries, and we believe that the color restoration of ancient paintings can also benefit from it. In addition to providing color inference and logical explanation, LLMs can also improve the quality of painting restoration by accepting verbal descriptions from restorers, which is not supported by the current framework. Thus, the integration of restorers and LLMs in re-coloring ancient paintings has become a promising research direction. To foster effective human-AI collaboration, users should review LLM results to address ethical concerns raised by automation tasks.

9 CONCLUSION

We propose PColorizor, a novel visual analytic approach to facilitate the easy restoration of ancient Chinese paintings. To ease the difficulties of manual restoration, we adopt deep learning-based models to provide effective color references and support efficient trial-and-error colorization. We especially employ the CLIP model to associate poems and paintings using the congruence principle. To leverage AI insights, we further visualize poem ideorealm using heatmaps that predict the visual attentions of the CLIP model and develop a mountain-based visual metaphor that exhibits the temporal evolution of color schemes. Finally, we collaborate with two domain experts and conduct real-world case studies to demonstrate the usage of PColorizor.

ACKNOWLEDGMENTS

The work was supported by the National Social Science Fund of China (19ZDA046), the National Key R&D Program of China (2022YFE0137800), NSFC (U22A2032), and the Collaborative Innovation Center of Artificial Intelligence by MOE and Zhejiang Provincial Government (ZJU). The work was also supported by the Fundamental Research Funds for the Central Universities.

REFERENCES

[1] Hyojin Bahng, Seungjoo Yoo, Wonwoong Cho, David Keetae Park, Ziming Wu, Xiaojuan Ma, and Jaegul Choo. 2018. Coloring with Words: Guiding Image Colorization Through Text-Based Palette Generation. In *Proceedings of European Conference on Computer Vision*. Springer, 431–447.

[2] Baogang Wei, Yonghuai Liu, and Yunhe Pan. 2003. Using hybrid knowledge engineering and image processing in color virtual restoration of ancient murals. *IEEE Transactions on Knowledge and Data Engineering* 15, 5 (2003), 1338–1343.

[3] David Borland and Russell M. Taylor Ii. 2007. Rainbow Color Map (Still) Considered Harmful. *IEEE Computer Graphics and Applications* 27, 2 (2007), 14–17.

[4] Yining Cao, Jane L. E, Zhutian Chen, and Haijun Xia. 2023. DataParticles: Block-Based and Language-Oriented Authoring of Animated Unit Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA.

[5] Jianbo Chen, Yelong Shen, Jianfeng Gao, Jingjing Liu, and Xiaodong Liu. 2018. Language-Based Image Editing with Recurrent Attentive Models. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 8721–8729.

[6] Shenghui Cheng, Wei Xu, and Klaus Mueller. 2019. ColorMap ND : A Data-Driven Approach and Tool for Mapping Multivariate Data to Color. *IEEE Transactions on Visualization and Computer Graphics* 25, 2 (2019), 1361–1377.

[7] Zhou Cheng. 1368—1644. Northern Seas. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/61bb72ec33115819641cf354.

[8] Zhou Cheng. 1368—1644. Seclusion beyond the Spring Waters in Spring. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/608a6c1fe11ca96100860fd0.

[9] Zezhou Cheng, Qingxiong Yang, and Bin Sheng. 2015. Deep Colorization. In *Proceedings of IEEE International Conference on Computer Vision*. IEEE, 415–423.

[10] Katherine Crowson, Stella Biderman, Daniel Kornis, Dashiell Stander, Eric Hallahan, Louis Castricato, and Edward Raff. 2022. VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance. In *Proceedings of European Conference on Computer Vision*. Springer, 88–105.

[11] Zhu Duan. 1368—1644. Gazing at the Distant River in the Mist. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/608a623142b4355c93c903e7.

[12] Zu E and Yu Xinchi. 2022. Shenxian Yijing, Mogu Chuanzhen [Interview with Master Zuo E, Representative Inheritor of the National Intangible Cultural Heritage Project - Traditional Art of Copying and Replicating Ancient Calligraphies and Paintings]. *Zijin Cheng* 12, 54–75 (2022).

[13] Patrick Esser, Robin Rombach, and Bjorn Ommer. 2021. Taming Transformers for High-Resolution Image Synthesis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 12868–12878.

[14] XY Fu, Y Han, ZJ Sun, XJ Ma, and YQ Xu. 2017. Line-drawing enhanced interactive mural restoration for Dunhuang Mogao Grottoes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 4 (2017), 99.

[15] Sergio Guadarrama, Ryan Dahl, David Bieber, Jonathon Shlens, Mohammad Norouzi, and Kevin Murphy. 2017. PixColor: Pixel Recursive Colorization. In *Proceedings of British Machine Vision Conference*. British Machine Vision Association, 112–124.

[16] Raj Kumar Gupta, Alex Yong-Sang Chia, Deepu Rajan, Ee Sin Ng, and Huang Zhiyong. 2012. Image Colorization using Similar Images. In *Proceedings of ACM International Conference on Multimedia*. ACM, 369–378.

[17] Mingming He, Dongdong Chen, Jing Liao, Pedro V. Sander, and Lu Yuan. 2018. Deep Exemplar-based Colorization. *ACM Transactions on Graphics* 37, 4 (2018), 1–16.

[18] Ruizhen Hu, Ziqi Ye, Bin Chen, Oliver van Kaick, and Hui Huang. 2023. Self-Supervised Color-Concept Association via Image Colorization. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 247–256.

[19] Zhitong Huang, Nanxuan Zhao, and Jing Liao. 2022. UniColor: A Unified Framework for Multi-Modal Colorization with Transformer. *ACM Transactions on Graphics* 41, 6 (2022), 1–16.

[20] Revital Irony, Daniel Cohen-Or, and Dani Lischinski. 2005. Colorization by Example. In *Proceedings of Eurographics Conference on Rendering Techniques*. Eurographics Association, 201–210.

[21] Eungyeup Kim, Sanghyeon Lee, Jeonghoon Park, Somi Choi, Choonghyun Seo, and Jaegul Choo. 2021. Deep Edge-Aware Interactive Colorization against Color-Bleeding Effects. In *Proceedings of IEEE International Conference on Computer Vision*. IEEE, 14647–14656.

[22] Hyunsu Kim, Ho Young Jhoo, Eunhyeok Park, and Sungjoo Yoo. 2019. Tag2Pix: Line Art Colorization Using Text Tag With SECat and Changing Loss. In *Proceedings of IEEE International Conference on Computer Vision*. IEEE, 9055–9064.

[23] Manoj Kumar, Dirk Weissenborn, and Nal Kalchbrenner. 2021. Colorization Transformer. In *Proceedings of International Conference on Learning Representations*.

[24] Rensselaer W. Lee. 1940. Ut Pictura Poesis: The Humanistic Theory of Painting. *The Art Bulletin* 22, 4 (1940), 197–269.

[25] Anat Levin, Dani Lischinski, and Yair Weiss. 2004. Colorization using Optimization. In *Proceedings of Special Interest Group on Computer Graphics and Interactive Techniques*. ACM, 689–694.

[26] Duan Ling. 1996. "Xuanhe Huapu" Tanwei [A Probe into Xuanhe Huapu]. *Meishu Yanjiu* 4, 31–35 (1996).

[27] Yang Liu and Jeffrey Heer. 2018. Somewhere Over the Rainbow: An Empirical Assessment of Quantitative Colormaps. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[28] Qing Luan, Fang Wen, Daniel Cohen-Or, Lin Liang, Ying-Qing Xu, and Heung-Yeung Shum. 2007. Natural Image Colorization. In *Proceedings of Eurographics Conference on Rendering Techniques*. Eurographics Association, 309–320.

[29] Varun Manjunatha, Mohit Iyyer, Jordan Boyd-Graber, and Larry Davis. 2018. Learning to Color from Language. In *Proceedings of the Conference of North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 764–769.

[30] John Maule, Alice E. Skelton, and Anna Franklin. 2023. The Development of Color Perception and Cognition. *Annual Review of Psychology* 74 (2023), 87–111.

[31] Wang Meng. 1690-1722. Ten Thousand Miles of the Yangtze River. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/6102e0eb5c31e364adf7e065.

[32] Zhao Mengfu. 1254—1322. Dongting East Hill. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/635f574709b9265e9e1c075c.

[33] Yuji Morimoto, Yuichi Taguchi, and Takeshi Naemura. 2009. Automatic Colorization of Grayscale Images using Multiple Images on the Web. In *SIGGRAPH 2009: Talks*.

[34] OpenAI. 2022. Introducing ChatGPT. https://openai.com/blog/chatgpt.

[35] Stephen E. Palmer and Karen B. Schloss. 2010. An Ecological Valence Theory of Human Color Preference. *Proceedings of the National Academy of Sciences* 107, 19 (2010), 8877–8882.

[36] Stephen E. Palmer, Karen B. Schloss, and Jonathan Sammartino. 2013. Visual Aesthetics and Human Preference. *Annual Review of Psychology* 64 (2013), 77–107.

[37] Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. 2021. StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery. In *Proceedings of IEEE International Conference on Computer Vision*. IEEE, 2065–2074.

[38] Soo-Chang Pei, Yi-Chong Zeng, and Ching-Hua Chang. 2004. Virtual Restoration of Ancient Chinese Paintings Using Color Contrast Enhancement and Lacuna Texture Synthesis. *IEEE Transactions on Image Processing* 13, 3 (2004), 416–429.
[39] Li Peng. 2014. On the Commonalities of Artistic Principles of Landscape Poetry and Painting in Liu Song Dynasty. *Canadian Social Science* 10, 6 (2014), 217.
[40] Yihao Peng, JiWoong Jang, Jeffrey P. Bigham, and Amy Pavel. 2021. Say It All: Feedback for Improving Non-Visual Presentation Accessibility. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA.
[41] Huy Q. Phan, Hongbo Fu, and Antoni B. Chan. 2018. Color Orchestra: Ordering Color Palettes for Interpolation and Prediction. *IEEE Transactions on Visualization and Computer Graphics* 24, 6 (2018), 1942–1955.
[42] Adobe Photoshop. 2023. Match Colors in Your Image. Website. https://helpx.adobe.com/photoshop/using/matching-replacing-mixing-colors.html.
[43] Yingge Qu, Tien-Tsin Wong, and Pheng-Ann Heng. 2006. Manga colorization. *ACM Transactions on Graphics* 25, 3 (2006), 1214–1220.
[44] Ghulam J. Quadri, Jennifer A. Nieves, Brenton M. Wiernik, and Paul Rosen. 2022. Automatic Scatterplot Design Optimization for Clustering Identification. *IEEE Transactions on Visualization and Computer Graphics* (2022).
[45] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of International Conference on Machine Learning*. PMLR, 8748–8763.
[46] Ragini Rathore, Zachary Leggon, Laurent Lessard, and Karen B. Schloss. 2019. Estimating Color-Concept Associations from Image Statistics. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 1226–1235.
[47] Khairi Reda, Pratik Nalawade, and Kate Ansah-Koi. 2018. Graphical Perception of Continuous Quantitative Maps: The Effects of Spatial Frequency and Colormap Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.
[48] Khairi Reda and Danielle Albers Szafir. 2021. Rainbows Revisited: Modeling Effective Colormap Design for Graphical Inference. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1032–1042.
[49] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley. 2001. Color Transfer between Images. *IEEE Computer Graphics and Applications* 21, 5 (2001), 34–41.
[50] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. 2022. Palette: Image-to-Image Diffusion Models. In *Proceedings of Special Interest Group on Computer Graphics and Interactive Techniques*. ACM, 1–10.
[51] Karen B. Schloss, Connor C. Gramazio, Allison T. Silverman, Madeline L. Parker, and Audrey S. Wang. 2018. Mapping Color to Meaning in Colormap Data Visualizations. *IEEE Transactions on Visualization and Computer Graphic* 25, 1 (2018), 810–819.
[52] Karen B. Schloss, Laurent Lessard, Charlotte S. Walmsley, and Kathleen Foley. 2018. Color Inference in Visual Communication: the Meaning of Colors in Recycling. *Cognitive Research: Principles and Implications* 3, 1 (2018), 1–17.
[53] Karen B. Schloss and Stephen E. Palmer. 2011. Aesthetic Response to Color Combinations: Preference, Harmony, and Similarity. *Attention, Perception, & Psychophysics* 73 (2011), 551–571.
[54] Rainer Schulte. 2014. *Translating China for Western Readers: Reflective, Critical, and Practical Essays*. SUNY Press.
[55] Vidya Setlur and Maureen C. Stone. 2015. A Linguistic Approach to Categorical Color Assignment for Data Visualization. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 698–707.
[56] Wang shaohua. 2001. Su Shi "Shihua Yilv" de Neihan [The Connotation of Shi Su's "Congruence between Paintings and Poems"]. *Wenyi Lilun Yanjiu* 77–83 (2001).
[57] Xinyu Shi, Ziqi Zhou, Jingwen Zhang, Ali Neshati, Anjul K. Tyagi, Ryan Rossi, Shunan Guo, Fan Du, and Jian Zhao. 2023. De-Stijl: Facilitating Graphics Design with Interactive 2D Color Palette Recommendation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA.
[58] Maria Shugrina, Wenjia Zhang, Fanny Chevalier, Sanja Fidler, and Karan Singh. 2019. Color Builder: A Direct Manipulation Interface for Versatile Color Theme Authoring. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.
[59] Li Sixun. 1500. Emperor Taizong Arriving at the Jiucheng Palace. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/6229cf4dc0a0c85ae709de6a.
[60] Stephen Smart, Keke Wu, and Danielle A. Szafir. 2019. Color Crafting: Automating the Construction of Designer Quality Color Ramps. *IEEE Transactions on Visualization and Computer Graphic* 26, 1 (2019), 1215–1225.
[61] Jheng-Wei Su, Hung-Kuo Chu, and Jia-Bin Huang. 2020. Instance-Aware Image Colorization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 7968–7977.
[62] Yu Sun, Shuohuan Wang, Shikun Feng, Siyu Ding, Chao Pang, Junyuan Shang, Jiaxiang Liu, Xuyi Chen, Yanbin Zhao, Yuxiang Lu, et al. 2021. Ernie 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation. *arXiv preprint arXiv:2107.02137* (2021).
[63] Danielle A. Szafir. 2017. Modeling Color Difference for Visualization Design. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2017), 392–401.
[64] Yu-Wing Tai, Jiaya Jia, and Chi-Keung Tang. 2005. Local Color Transfer via Probabilistic Segmentation by Expectation-Maximization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 747–754.
[65] Wenbo Tao, Xinli Hou, Adam Sah, Leilani Battle, Remco Chang, and Michael Stonebraker. 2020. Kyrix-s: Authoring Scalable Scatterplot Visualizations of Big Data. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 401–411.
[66] Chin Tseng, Ghulam J. Quadri, Zeyu Wang, and Danielle A. Szafir. 2023. Measuring Categorical Perception in Color-Coded Scatterplots. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA.
[67] Barbara Tversky, Maneesh Agrawala, Julie Heiser, Paul Lee, Pat Hanrahan, Phan Doantam, Chris Stolte, and Marie P. Daniel. 2020. Cognitive Design Principles for Automated Generation of Visualizations. In *Applied Spatial Cognition*. Psychology Press, 53–74.
[68] Patricia Vitoria, Lara Raad, and Coloma Ballester. 2020. ChromaGAN: Adversarial Picture Colorization with Semantic Class Distribution. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2434–2443.
[69] Han-Lei Wang, Ping-Hsuan Han, Yu-Mu Chen, Kuan-Wen Chen, XinYi Lin, Ming-Sui Lee, and Yi-Ping Hung. 2018. Dunhuang mural restoration using deep learning. In *SIGGRAPH Asia Technical Briefs*. ACM, 1–4.
[70] Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller. 2002. Transferring Color to Greyscale Images. *ACM Transactions on Graphics* 21, 3 (2002), 277–280.
[71] Shuchen Weng, Hao Wu, Zheng Chang, Jiajun Tang, Si Li, and Boxin Shi. 2022. L-CoDe:Language-Based Colorization Using Color-Object Decoupled Conditions. In *Proceedings of AAAI Conference on Artificial Intelligence*. AAAI, 2677–2684.
[72] Yanze Wu, Xintao Wang, Yu Li, Honglun Zhang, Xun Zhao, and Ying Shan. 2021. Towards Vivid and Diverse Image Colorization with Generative Color Prior. In *Proceedings of IEEE International Conference on Computer Vision*. IEEE, 14357–14366.
[73] Menghan Xia, Wenbo Hu, Tien-Tsin Wong, and Jue Wang. 2022. Disentangled Image Colorization via Global Anchors. *ACM Transactions on Graphics* 41, 6 (2022), 1–13.
[74] Wang Ximeng. 1113. A Thousand Li of Rivers and Mountains. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/62429b3b43eebd107d0c2ee8.
[75] Zhongyou Xu, Tingting Wang, Faming Fang, Yun Sheng, and Guixu Zhang. 2020. Stylization-Based Architecture for Fast Deep Exemplar Colorization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 9360–9369.
[76] Chuan Yan, John J. Young Chung, Yoon Kiheon, Yotam Gingold, Eytan Adar, and Sungsoo R. Hong. 2022. FlatMagic: Improving Flat Colorization through AI-Driven Design for Digital Comic Professionals. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA.
[77] An Yang, Junshu Pan, Junyang Lin, Rui Men, Yichang Zhang, Jingren Zhou, and Chang Zhou. 2022. Chinese CLIP: Contrastive Vision-Language Pretraining in Chinese. *arXiv preprint arXiv:2211.01335* (2022).
[78] Liron Yatziv and Guillermo Sapiro. 2006. Fast Image and Video Colorization using Chrominance Blending. *IEEE Transactions on Image Processing* 15, 5 (2006), 1120–1129.
[79] Chou Ying. 1368-1644. Saying Farewell at Xunyang. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/608a619eaa7c385c8d942ecd.
[80] Chou Ying. 1368—1644. Maple Creek Fishing. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/608a61a0aa7c385c8d943112.
[81] Chou Ying. 1368—1644. Peach Orchard Wonderland. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/608a61a0aa7c385c8d943115.
[82] Chou Ying. 1542-1545. Along the River During the Qingming Festival. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/609672f536bd622e5d1ee9bf.
[83] Wei Yishi. 1983. Su Shi "Hui Chong Chunjiang Wanjing" Shangxi [Appreciation of Su Shi's "Hui Chong Chunjiang Wanjing"]. *Beijing Shifan Daxue Xuebao* 3, 95–96 (1983).
[84] Dong Yuan. 907-960. Residents on the Outskirts of Dragon Abode. ltfc.net (Website). https://g2.ltfc.net/view/SUHA/621dbd0d5495b316e05a3051.
[85] Jun Zhang. 2021. Exploring the Application of Traditional Elements in Cultural and Creative Product Design. *Art and Design Review* 9, 332–340 (2021).
[86] Richard Zhang, Phillip Isola, and Alexei A. Efros. 2016. Colorful Image Colorization. In *Proceedings of European Conference on Computer Vision*. Springer, 649–666.
[87] Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S. Lin, Tianhe Yu, and Alexei A. Efros. 2017. Real-Time User-Guided Image Colorization with Learned Deep Prior. *ACM Transactions on Graphics* 36, 4 (2017), 1–11.
[88] Yufan Zhou, Ruiyi Zhang, Jiuxiang Gu, Chris Tensmeyer, Tong Yu, Changyou Chen, Jinhui Xu, and Tong Sun. 2022. TiGAN: Text-Based Interactive Image

Generation and Manipulation. In *Proceedings of AAAI Conference on Artificial Intelligence*. AAAI, 3580–3588.

[89] Zhe Zhu, Haozhi Huang, Zhipeng Tan, Kun Xu, and Shimin Hu. 2016. Faithful Completion of Images of Scenic Landmarks Using Internet Images. *IEEE Transactions on Visualization and Computer Graphics* 22, 8 (2016), 1945–1958.

[90] Changqing Zou, Haoran Mo, Chengying Gao, Ruofei Du, and Hongbo Fu. 2019. Language-based Colorization of Scene Sketches. *ACM Transactions on Graphics* 38, 6 (2019), 1–16.

[91] Zheng Zou, Peng Zhao, and Xuefeng Zhao. 2021. Virtual Restoration of the Colored Paintings on Weathered Beams in the Forbidden City Using Multiple Deep Learning Algorithms. *Advanced Engineering Informatics* 50 (2021), 101421.