CAnnotator: Photo-Guided Color Annotation for Degraded Ancient Paintings

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

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

Yihan Gao
yihangao@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

Junming Gao
junminggao@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

Songela Nur
songelanur@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

Buwei Zhou
zhoubuwei@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

Yuchen Lin
yuchenlin@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

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

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

Xiaojiao Chen*
chenxiaojiao@zju.edu.cn
Zhejiang University
Hangzhou, Zhejiang, China

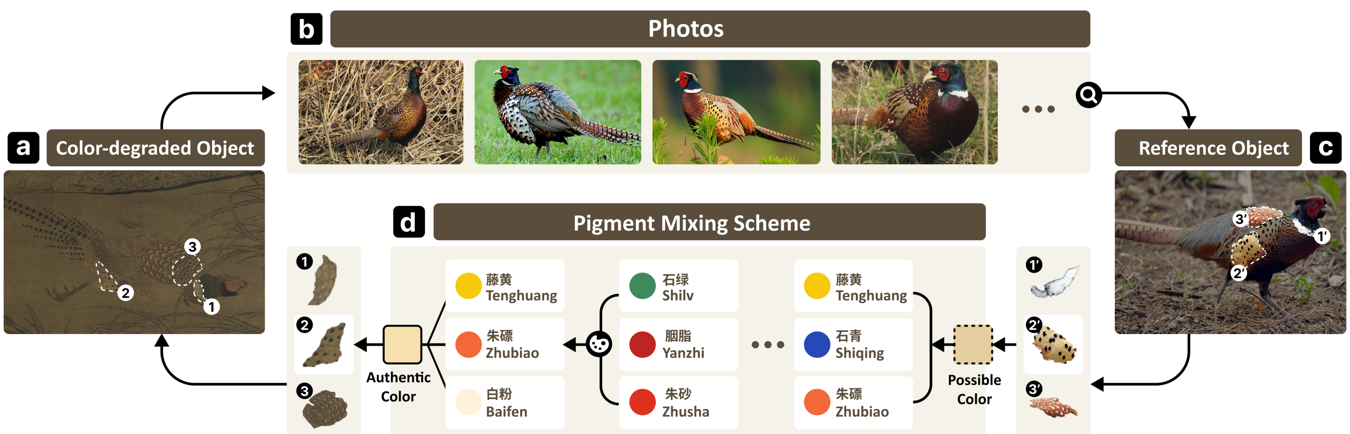

Figure 1: The workflow of CAnnotator: (a) extracting textures from a color-degraded object (labeled as *1, 2, 3*); (b) searching for a reference object with similar textures and postures from photo collection; (c) obtaining a possible color according to the textures of the reference object (labeled as *1', 2', 3'*); (d) determining an authentic color that can be reproduced by mixing traditional pigments whose Chinese names and associated translations are attached. Theoretically, a color can be decomposed into a set of pigments with different quantities.

*Corresponding author

UIST '25, Busan, Republic of Korea

Abstract

Ancient paintings suffer irreversible color degradation due to aging and improper conservation. Labeling degraded paintings with authentic colors becomes vital to protect these valuable cultural heritages, which is challenging due to missing color information.

ACM ISBN 979-8-4007-2037-6/2025/09
https://doi.org/10.1145/3746059.3747764

Users typically need to investigate relevant photos to infer authentic colors and then validate these colors by mixing traditional pigments. However, such a task could be exhausting. To ease the difficulty, we propose an interactive visualization tool, namely CAnnotator, that streamlines efficient human-AI collaboration for the color annotation of degraded ancient paintings. CAnnotator consists of three views: a paint-annotation view, a photo-reference view, and a pigment-mixing view. Given an ancient painting, the paint-annotation view is developed to help users extract its color-degraded object textures that would be propagated to the relevant photos using a texture tracking model. Based on the tracking results, the photo-reference view provides texture-color and object-posture filters to explore the photos that include the given texture colors and object postures. We train a deep learning model to simulate the mixing of physical pigments and employ the chain rule to support progressive pigment mixture using a novel flow-based color visualization. We demonstrate the usage of CAnnotator through a use case and evaluate its effectiveness through model experiments and an in-lab user study. Compared to the baseline, CAnnotator could improve user confidence of labeled colors and foster user engagement at the cost of additional time.

CCS Concepts

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

Keywords

cultural heritage, ancient artwork, fine art, information visualization, color annotation

ACM Reference Format:

Tan Tang, Yanhong Wu, Yihan Gao, Junming Gao, Songela Nur, Buwei Zhou, Yuchen Lin, Yingcai Wu, Xiaosong Wang, and Xiaojiao Chen. 2025. CAnnotator: Photo-Guided Color Annotation for Degraded Ancient Paintings. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25), September 28-October 1, 2025, Busan, Republic of Korea.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3746059.3747764

1 Introduction

Painting pigments are susceptible to degradation when exposed to light and oxidizing environments, leading to irreversible changes in perceived colors [30, 33]. The degradation of colors can diminish the visual appearance of paintings, which significantly impacts their aesthetic quality [22]. Unfortunately, most ancient paintings have suffered varying degrees of color degradation. For example, many vibrant red and yellow pigments in Van Gogh's paintings are gradually losing their original color [25]. Before restoring these paintings [9, 13, 29], conservators need to label the degraded objects in ancient paintings with authentic colors (see Fig. 2). This process is referred to as color annotation, which is challenging due to the absence of the original colors.

Color annotation consists of image exploration and pigment mixing. When original colors are unavailable, conservators first analyze extensive collections of reference images to infer possible colors (see Fig. 1b). For example, realistic paintings often aim to replicate observed objects as they exist today and are documented in photographs [8, 19]. These visual records of natural forms provide critical clues for narrowing down potential color choices [15]. However, photographic reproductions alone are insufficient as they may distort or misrepresent the original colors in the paintings. Typically, conservators must conduct pigment-mixing experiments using historical materials where only colors reproducible through traditional pigments are deemed authentic (see Fig. 1d). Both tasks require meticulous effort and expertise, making the process inherently time-consuming. This highlights the need for a novel color annotation tool, which comes with two obstacles:

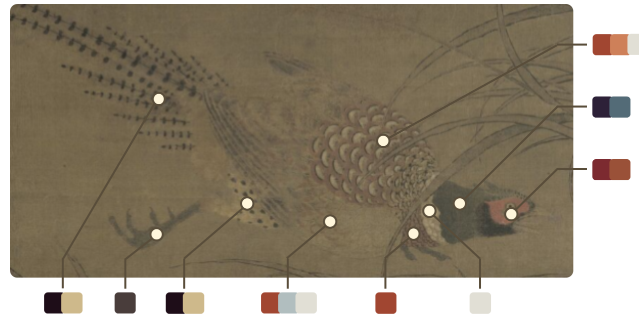

Figure 2: An example about color annotation where degraded textures are labeled by authentic colors.

Obtaining possible colors for degraded objects from photo collection. There are multiple variants of an object, each exhibiting distinct textures. For instance, the pheasant has nineteen subspecies in East Asia [49]. To obtain possible colors, conservators must select objects with consistent textures as references. Additionally, the references must match the postures depicted in the painting to avoid missing color details. This process requires searching through extensive photo collections to identify suitable candidates. Most online datasets lack detailed labels (e.g., texture-level information), making it difficult to achieve such a goal. Thus, a novel approach is required to associate photos and paintings by identifying shared textures and poses in the depicted objects.

Predicting authentic colors through mixing multiple pigments. To determine authentic colors, conservators must predict the mixture of pigments that align with the historical context of the original painting. Several machine learning models [14, 65] have been proposed to predict the mixing color of pairwise pigments. However, it is still difficult to describe the mixture of multiple pigments, which is a complicated physical process without a precise mathematical formulation. Moreover, end-to-end prediction operates as a black box, potentially overlooking the nuances of pigment mixture within conservators' expertise. Therefore, a new interactive tool is needed to predict the mixture color of multiple pigments while minimizing discrepancies from the possible colors.

In this study, we adopt the four-level nested model [44] to develop a visualization tool, namely CAnnotator, that supports easy color annotation for color-degraded paintings. The four levels are domain situation (Section 2), data/task abstraction (Section 3), visual encoding/interaction idiom (Section 6-7), and algorithm (Section 5), which ensures rigorous translation of conservators' needs into technical solutions. Collaborating with professional conservators, we characterize specific user requirements and extract corresponding design tasks. To address the first challenge, we employ a state-of-the-art interactive segmentation model [32] to extract its distinct textures through the paint-annotation view. We adopt an automatic tracking model [17] to propagate the extracted textures to photos, which are visualized using the photo-reference view. Additionally, we provide two interactive widgets, namely a texture-color filter and an object-posture filter, to facilitate the interactive exploration of photo collection. To overcome the second challenge, we model pigment mixture according to conservators' behaviors in real practices (Eq. 1 and Eq. 2) and train a DNN model to predict pigment mixture. Based on that, we design a novel flow-based color visualization that presents predicted pigments and enables progressive, interactive refinement of mixtures. Finally, we demonstrate the usage of CAnnotator through a use case and validate its effectiveness using both model and user evaluations.

The main contributions are as follows:

- We collaborate with professional conservators to abstract domain requirements and streamline the workflow of color annotation for ancient paintings.
- We develop a DNN model to predict pigment mixture and design a novel color palette to support the progressive mixing of multiple pigments.
- We implement a proof-of-concept prototype, namely CAnnotator, and validate its effectiveness and usability through comprehensive evaluations.

2 Related Work

This section introduces relevant studies from three aspects: painting annotation, image exploration, and pigment mixing.

2.1 Painting Annotation

To protect cultural heritage, the domain of ancient paintings has entered various annotation tasks, which can be broadly categorized into three types: painting techniques [36, 68, 69], painting content [31, 40], and emotions [5, 53].

The annotations of the painting techniques include meta-level concepts (e.g., author and school) and visual-level concepts (e.g., brush and color scheme). These concepts imbue low-level image features with understandable semantics, which is crucial for applications such as painting retrieval and discrimination [68]. To annotate paintings with these concepts, Leslie et al. [36] proposed a three-step methodology in which the painting image was first segmented into blocks. Then, each image block was assigned visual concepts through a transductive inference framework. Subsequently, the visual concepts of all image blocks were integrated to generate the meta concepts of the painting. The annotations of the painting content refer to descriptions of relationships among different objects within the painting [1]. A typical example is viewers' comments about the painting's content [27]. Such annotations can serve applications that require visual understanding, such as content-based painting retrieval [27] and painting accessibility for visually impaired people [43, 51]. The annotations of the conveyed emotions depict viewers' feelings from different groups after carefully watching the painting. These annotations highlight the connections between a painting's content and its visual impact [53] but do not address how to label color-degraded objects in ancient paintings.

Unlike the above annotation tasks, this study focuses on color annotation for degraded ancient paintings. Our task requires additional knowledge to bridge the color gap between degraded and original objects, which is challenging for conservators. Thus, we propose a novel tool that integrates deep learning models and interactive visualizations to help conservators determine authentic colors from photos. As with other endeavors in analyzing the colors of antiquities [18, 37], the generated color labels can be used to build a color database of ancient paintings, thereby aiding downstream tasks such as color analysis [16] or restoration [61].

2.2 Image Exploration

The remarkable advancements in deep learning has catalyzed the emergence of methods for exploring large-scale image collections [2]. These methods [6, 7, 11, 12, 38, 39, 58, 60] leverage deep learning models to extract visual and semantic features from images and employ hierarchical visualizations to represent their interrelations. For example, Xie et al. [64] proposed a galaxy metaphor-based method to reveal the semantic similarity between images. They first obtained semantic keywords for images using an image caption model and then simultaneously projected the images and the keywords into a 2D space. In this space, images were represented as planets and keywords as stars. The distance between the planet and the star encoded the relevance between the image and the keyword, while the distance between the stars encoded the co-occurrence frequency of the keywords in the images. To flexibly support various analysis tasks that require image search and exploration, Jan et al. [70] introduced an agent model that learned users' intentions in real time based on their interactions in browsing images. Under the framework they established, users were initially required to define specific image categories related to the analysis task and subsequently manually annotate a few images. The annotated images were used to train the agent model, which assisted users in image search and exploration according to user preferences.

Compared to previous image exploration tasks, we focus on bridging the gap between ancient paintings and photographic images, which are difficult to capture by a separate image classification or caption model. To solve this issue, we adopt a "segment first and then track" strategy to align object textures and postures in different images and provides interactive widgets to navigate users for efficient exploration.

2.3 Pigment Mixing

Paintings contain a wide range of colors drawn by artists using various pigments. Modeling the mapping between pigments and colors plays a crucial role in constructing digital paintings [54]. Fortunately, fruitful achievements have already been made in this domain. For example, Kubelka and Munk [34, 35] first proposed a mathematical equation to calculate the final reflectance of the pigment applied to a substrate based on the substrate's reflectance, the pigment's absorption and scattering coefficients, and the thickness of the pigment layer on the substrate. Duncan [24] further pointed out that the homogeneous mixing of pigments can be regarded as a linear combination of their absorption and scattering coefficients based on constants. These achievements have been continuously refined and widely employed in subsequent research [3, 20, 28, 59]. Additionally, another category of methods [14, 56, 65] utilizes deep learning techniques to learn non-linear color transformations from the original and mixed pigments. They take the physical parameters of the original pigments, such as reflectance and quantity, as input and directly predict their mixed reflectance. Such deep learning methods have shown superior ability of learning color-pigment mapping than those based on the Kubelka-Munk equation [14].

Although a few works provide pairwise pigment mixing tools [3, 14], they are inadequate for meeting conservators' requirements in color annotation, particularly in their inability to mix multiple pigments effectively. To bridge the gap, we design a novel flow-based color visualization that enables interactive mixing of multiple pigments and develop a DNN model to predict the mixed color.

3 Domain Abstraction

To better understand the domain requirements (**DR**), we closely collaborated with two professional conservators (EA and EB) who possess extensive experience in color annotation. They both come from the same authoritative Chinese painting and calligraphy restoration institution. We maintained weekly online meetings with these two conservators throughout a year-long collaboration. During such a period, we also conducted an on-site visit to observe their color annotation process and had multiple in-depth interviews.

3.1 Design Tasks

We follow the general workflow of color annotation and identify design tasks according to the domain requirements.

DR1. Gathering and analyzing reference photos. Given an ancient painting, conservators first need to determine the approximate identity of color-degraded objects (e.g., pheasant in Fig. 1). They often explore the objects' variants and search for relevant photos from the Internet. Due to the various conditions of the photography environment and cameras, conservators have to collect as many reference photos as possible in case of ignoring possible color choices. They also need to find objects with postures similar to the objects in the painting, as varying postures afford distinct observations. For example, EB noted that the pheasant possesses a tuft of feathers on its lower abdomen that is visible only when running. With sufficient photos, conservators need to analyze and filter irrelevant photos according to the textures of color-degraded objects. This process inspires the following design tasks:

- **T1 Support the extraction of object textures.** Object textures are the fundamental units for color annotation and important clues for identifying reference objects. Users should be able to efficiently extract the textures from the painting and associate them with reference objects.
- **T2 Provide an overview of object variants.** Object identification requires users to distinguish different object variants and select the most relevant one for the color-degraded object. Thus, users are required to understand the differences of the variants and quickly navigate photo collection.
- **T3 Filter photos based on objects' textures and postures.** Redundant textures and postures hinder the effective searching of reference objects in photos, which requires both texture and posture filters for efficient analysis.

DR2. Reproducing colors through mixing pigments. After inferring possible colors, conservators need to reproduce them using the pigments available when the painting was created. Given a target color, they will establish an initial mixing scheme based on their expertise and then pursue the target color by continuously adjusting the scheme. The changes of the mixing scheme primarily involve two ways: 1) altering the mixing ratios of different pigments (*See More Quantities*) and 2) modifying the types of pigments to be mixed (*See More Mixtures*). Additionally, conservators will apply pigments of varying quantities on the canvas to observe their actual colors. This action may be challenging for junior conservators as they lack extensive experience in color mixing. When encountering unfamiliar colors, they often need considerable trial-and-error experiments to determine the final mixing scheme. This process requires significant manual efforts to perform repetitive pigment mixings and entails high costs to support the consumption of certain precious pigments, which inspires the design tasks:

- **T4 Recommend pigment mixing schemes.** Due to the vast space of color choices, it is difficult for users to master all pigment mixing schemes. Thus, color recommendations are required to help them label unfamiliar colors.
- **T5 Enable progressive changes of pigment mixing schemes.** Active exploration of pigment mixing schemes is necessary in case color recommendations fail to provide optimal schemes. Users should be able to progressively change pigment mixing schemes according to their preferences.

3.2 Data Description

To support the tasks, we collect three types of data, namely ancient paintings, photos, and traditional pigments (see Fig. 3). The ancient painting data encompasses digital images of ancient paintings and meta-information detailing their creation background, sourced

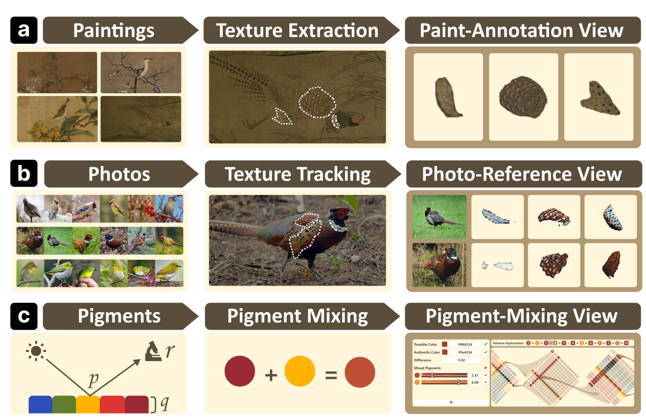

Figure 3: System overview: (a) extracting textures from paintings, (b) tracking photos according to textures, and (c) pigment mixing to obtain color labels.

from the publicly accessible collections of various art museums [45]. The photo data are sourced from publicly available image datasets, such as ImageNet [23], which contains over 14 million images. The traditional pigment data records the pigments' physical properties that are significant in pigment mixing simulation. According to colorimetry, the color of a digital image is influenced by both the light source and the object's reflectance spectrum. Typically, the light source is assumed to be natural daylight; therefore, the perceived color depends on the pigment's reflectance spectrum. Following previous pigment simulation works [34, 35, 65], we model changes in reflectance as a consequence of mixing pigments. Given pigments $\{p_1, p_2, ..., p_n\}$ and their quantity $\{q_1, q_2, ..., q_n\}$, which are real values, we define the reflectance spectrum as $r = f(p_1, q_1, ..., p_n, q_n)$ where n is the number of mixed pigments. Notice that the mapping f refers to a pigment-mixing model, which is grounded in physics and lacks a precise mathematical formulation. For simplicity, we define $r_i = f(p_i, q_i)$ for the i-th pigment and $r_{i,j} = f(p_i, q_i, p_j, q_j)$ for the mixture of the i-th pigment and the j-th pigment.

We used a public pigment dataset [14] containing 1,956 training samples and 488 test samples. The dataset contains 13 pigments present in varying quantities, resulting in distinct reflectance spectra. Each sample represents a pair of pigments, including their individual reflectance spectra $\{r_i, r_j\}$ and the reflectance spectrum of their mixture $r_{i,j}$. Each reflectance r is a 41-dimensional vector, corresponding to spectral measurements at wavelengths from 380 nm to 780 nm, sampled every 10 nm. Specifically, physical pigments were applied to a unit-area substrate, and their spectra were measured under the K1 light source using an SD1220 spectrometer [14]. The sub-dimension value of r is between 0 and 1. The reflectance spectra can be mapped to CIE XYZ tristimulus values using color-matching functions and then converted into Lab or RGB color spaces for accurate color rendering in visualization design.

4 System Overview

To satisfy the design tasks, we propose CAnnotator, a novel visualization tool that facilitates easy color annotation in three steps. We first extract object textures from color-degraded ancient paintings (Fig. 3a), then explore reference photos using a tracking model (Fig. 3b), and finally annotate textures according to determined pigment mixing schemes (Fig. 3c).

To develop CAnnotator, we first designed a paint-annotation view that combines a painting board and an interactive segmentation tool to decompose textures from a given painting (**T1**). The texture extraction model can accept user-input points to segment textures from color-degraded objects, thereby enabling flexible image segmentation. Second, we design a reference view that provides texture-color and object-posture filters to locate objects with specific textures (**T2**) and postures (**T3**) swiftly. The texture tracking model can propagate the extracted textures onto photos and automatically extract the corresponding textures. Third, we design a pigment-mixing view that incorporates a novel flow-based color visualization to enable flexible exploration of the pigment mixing schemes (**T5**). The pigment mixing model can predict the colors of different mixed pigments, which is the basis of the recommendation (**T4**) and adjustment of pigment mixing schemes.

CAnnotator is comprised of a database, a backend, and a frontend. The database is built on MinIO [42], which stores relevant data on ancient paintings, photographic images, and traditional pigments. The backend is implemented using the FastAPI in Python [62], which runs three deep learning models: a texture extraction model, a texture tracking model, and a pigment mixing model. The frontend is developed using React.js [57] and D3.js [46].

5 Model Design

This section introduces three deep learning models integrated in CAnnotator and explains how they satisfy the tasks.

5.1 Texture Extraction and Tracking

Texture is the essential characteristic for distinguishing object variants. Typically, texture extraction is supported by a supervised segmentation model, which is limited by insufficient domain datasets. Thus, we employ the state-of-the-art segment anything model (SAM) [32] that has demonstrated impressive zero-shot segmentation performance. Moreover, SAM can accept two types of user-input prompt points for interactive segmentation: positive points generated by users clicking on regions to be segmented and negative points generated by users clicking on regions to be excluded. Based on prompt points, SAM can accurately capture users' segmentation intentions and achieve high-quality segmentation results. These advantages make SAM a feasible solution for **T1**.

To satisfy **T2**, we need to obtain object textures in each photo that are semantically consistent with the extracted textures. Inspired by object tracking in videos [17, 50], we employ an advanced multi-object tracking model, namely DeAOT [67], to propagate the extracted textures onto photos. We chose this model because it has outstanding performance in simultaneously tracking multiple objects. Specifically, DeAOT involves Long Short-Term Transformer (LSTT) blocks [66] to propagate object information from reference and neighboring frames to the current frame. The reference frames refer to past frames sampled at specific intervals, including at least one manually annotated by the user (typically the first frame), and the neighboring frames refer to several consecutive frames before the current frame. Since photos do not have a temporal relationship similar to that between consecutive frames in a video, we cease the automatic sampling of reference frames, which is applied to acquire a long-term memory of object instances [66]. Moreover, due to the potential significant changes in object posture and camera perspective between adjacent photos, we deactivate the short-term attention mechanism in LSTT, which is applied to ensure smooth mask transitions between adjacent frames [66].

5.2 Pigment Mixing Simulation

The reflectance spectra of objects provide a more nuanced description of color compared to human-designed color spaces [47]. To support user-guided pigment mixing (**T4**), we develop a deep neural network (DNN) to learn the changes in pigment reflectance before and after mixing. We use reflectance to represent pigments in order to avoid the ambiguity caused by metamerism [26], a phenomenon where different pigments may appear identical in color under a given lighting condition. Concretely, we define the pigment mixing scheme as $s = \{(p_1, q_1), (p_2, q_2), ..., (p_n, q_n)\}$, where n

denotes the number of mixed pigments. The reflectance spectrum of mixed pigments under s can be described as $r_s = f(p_1, q_1, ..., p_n, q_n)$. The varying number of mixed pigments n and the limited training data [14] make it impossible to model the general pigment-mixing function f. Notice that a mixture of multiple pigments could be processed as a series mixing of paired pigments. Instead of directly modeling f, we introduce a DNN $\hat{f_\theta}$ that receives two reflectance spectra (r_i, r_j) and their quantity multipliers m_i, m_j, to predict the mixed spectrum $\hat{f_\theta}(r_i, m_i, r_j, m_j)$ where θ refers to the model weights. The multiplier m_i is a scalar that indicates how many times the pigment p_i with a fixed quantity q_i would be mixed. Then, the reflectance spectrum r_s can be re-formulated as

$$f(p_1, q_1, ..., p_n, q_n) = \hat{f_\theta}(r_n, m_n, f(p_1, q_1, ..., p_{n-1}, q_{n-1})) \quad (1)$$

Equation 1 follows the chain rule derived from our stepwise pigment-mixing strategy, which calculates the reflectance spectrum of scheme s recursively. In the public dataset [14], training samples could be extended as $\{(r_i, m_i, r_j, m_j, r_{i,j})\}_{i,j}$ where m_i and m_j are equal to 1 by default. The predction error could be calculated as $err_{i,j} = |r_{i,j} - \hat{f_\theta}(r_i, m_i, r_j, m_j)|$. Follow the previous study [14], we define the training loss as:

$$\mathcal{L}_\theta = (\frac{1}{|r_{i,j}|} + \alpha) err_{i,j} + \beta \left(\hat{f_\theta}(r_i, m_i, r_j, m_j) - \overline{r_{i,j}}\right)^2 \quad (2)$$

where α and β are two hyper-parameters that are set to 3 and 2 [14], and $\overline{r_{i,j}}$ denotes the average reflectance spectrum of the data samples computed during a training batch. The training goal is to minimize Eq. 2 by calculating the optimum model weights θ. Theoretically, a fully connected deep neural network could be used to approximate any smooth function. Thus, we adopt a multilayer perceptron to model $\hat{f_\theta}$. To avoid over-fitting, we conducted experiments to determine the appropriate number of layers for our task. The results suggest that a four-layer perceptron achieves the lowest validation loss, with input and output dimensions of 84 and 41 and hidden layers of sizes 80, 65, and 50.

6 Visualization Design

This section presents the visualization and interaction designs of CAnnotator to explain how they satisfy the design tasks.

6.1 Paint-Annotation View

The paint-annotation view consists of a painting board and an annotation list. The painting board (Fig. 4-a2) shows an ancient painting that involves several color-degraded objects. To support texture extraction (**T1**), we provide a floating toolbar integrating segmentation tools that have been widely used in commercial segmentation systems [4]. These tools include, from left to right, annotating positive points, annotating negative points, undoing the previous state, redoing to the next state, deleting all annotated points, and cutting the currently displayed segmentation. Figure 4-a2 illustrates the segmentation results for extracting color-degraded textures. The extracted textures are displayed in the annotation list, each paired with its corresponding real-photo counterpart and color labels (Fig. 4-a1). When a texture is selected for annotation, it appears in the pigment-mixing view (see Fig. 4-c1).

6.2 Photo-Reference View

The photo-reference view comprises a toolbar and a grid-based photo library. The toolbar (Fig. 4-b1) includes several tools for exploring photo collection, including a search input with a navigation switch and three buttons for tracking ( ), segmentation ( ), and confirmation ( ). Users can enter the object's name into the search input to retrieve relevant photos from a large photo database. By default, the retrieved photos are presented in a grid layout in the photo library, which can be changed using the navigation switch. After extracting the color-degraded textures from the painting, users can click the tracking button to automatically permutate photos with matched textures or poses. Moreover, users can click the segmentation button to extract additional textures from photos, which serve as references to improve tracking results. They can also click the photo to view its detailed content and the visual context of the tracked textures (Fig. 4-b4). After selecting a reference photo, users can tap the confirmation button to expand it to full-view mode.

In the photo library, we develop texture-color and object-posture filters to assist users in flexibly exploring objects with different textures (**T2**) and postures (**T3**). The texture-color filter (Fig. 4-b3) displays color squares representing the dominant colors in tracked textures. To generate these squares, we first cluster the pixels of the tracked textures to identify representative colors, then refine the results by removing outliers through additional clustering. Users can click a color square to filter the photo library, displaying only photos containing that color while temporarily hiding non-matching photos. We also develop an object-posture filter (Fig. 4-b2) using a scatter plot where each scatter represents a photo, and the distance between scatters indicates the objects' posture similarity. To calculate posture similarity, we construct posture vectors for each object based on their tracked textures (see Fig. 5). The geometric centers of each texture would serve as the values of posture vectors. We then utilize the widespread t-SNE algorithm [63] to project the posture vectors on a 2D scatter plot. The selected photos will be shown at the top, and their corresponding scatters will be highlighted with higher opacity and stroke-based outlines. To avoid visual occlusion, we provide drag-and-zoom interactions to allow users to change the observation range and scale.

6.3 Pigment-Mixing View

The pigment-mixing view consists of a labeling window (Fig. 4-c1), an annotation panel (Fig. 4-c2), a scheme bar (Fig. 4-c3), and a flow-based color visualization (Fig. 4-c4). The labeling window presents the extracted texture selected from the annotation list (Fig. 4-a1), below which are its editable color labels. Users can click a color label to examine its corresponding pigment mixing scheme. We also enable them to extract possible colors ( ) from photos. To determine authentic colors, users can manually add a pigment p ( ) and adjust its quantity q using a slider. Each slider consists of a color circle and two horizontal gradient bars. The circle presents the color of the pigment p at a unit quantity. We designed two

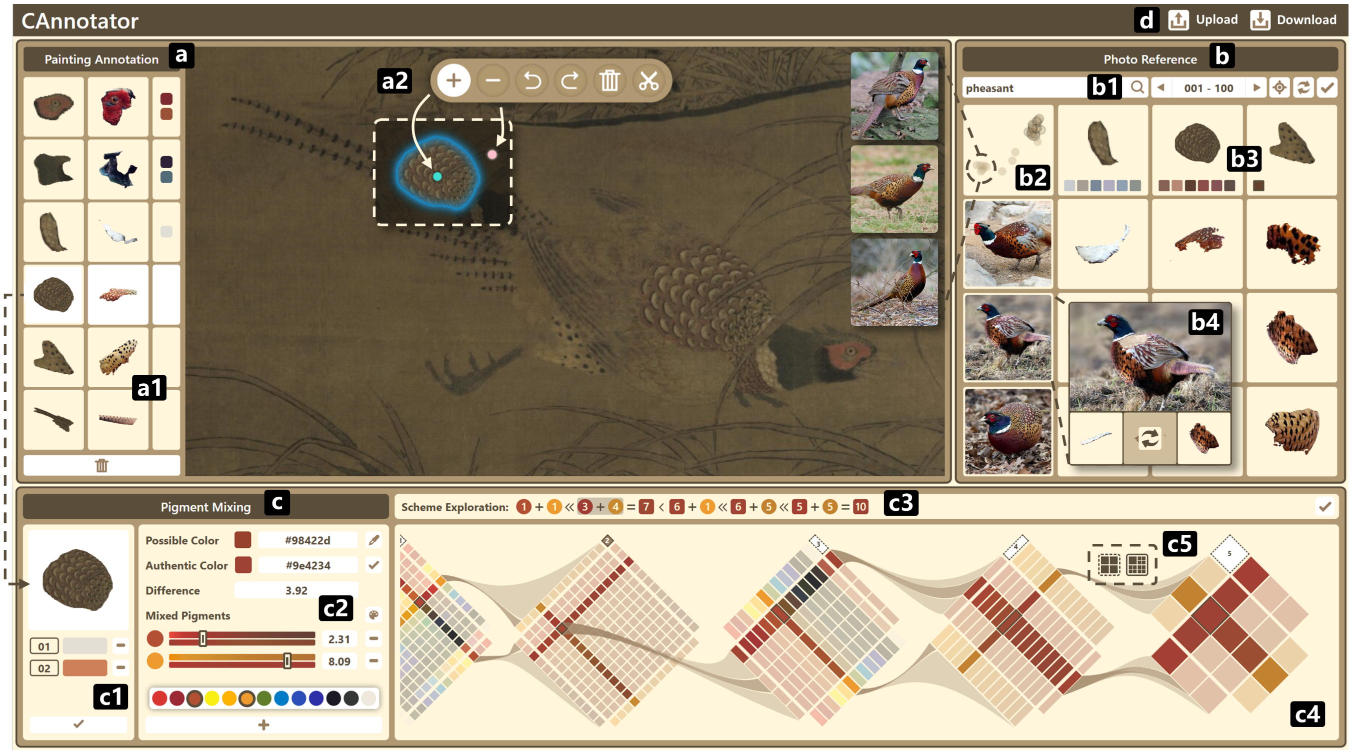

Figure 4: CAnnotator interface: (a) the paint-annnotation view consists of (a1) an annotation list and (a2) a painting board with interactive segmentation tools for extracting color-degraded textures; (b) The photo-reference view contains (b1) a toolbar with five tools for photo searching, page navigation, texture tracking, photo segmentation, and photo confirmation (from left to right). It also provides (b2) an object-posture filter and (b3) texture-color filters to efficiently navigate various object variants; The pigment-mixing view includes (c1) a labeling window, (c2) an annotation panel, (c3) a scheme bar, (c4) a flow-based color visualization (ColorFlow) with (c5) two sample controllers for exploring pigment-mixing schemes. (d) CAnnotator enables users to upload a painting and download its color annotations.

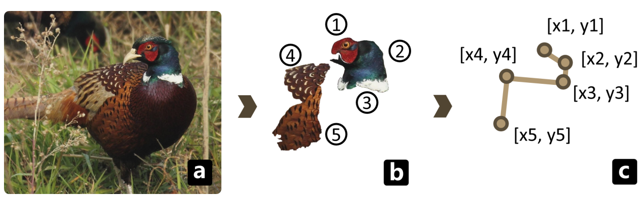

Figure 5: Given (a) a photo, we utilize its (b) tracked textures to construct (c) the posture representation vector by combining the center coordinate of each texture.

bars for a slider because the upper one presents the change of color while the bottom one presents its corresponding change of pigment. Such a pre-visualization could reduce manual efforts when users interact with sliders according to a previous study [21]. To satisfy **T4**, we also provide users with a recommendation button (🎨) that generates pigment mixing schemes based on a greedy strategy. Specifically, we iteratively adopt the two mixing methods (see Sec. 3.1), namely, *See More Quantities* and *See More Mixtures*, to generate a list of pigment mixing candidates. The candidate with the smallest color difference to the possible color is selected as the new scheme, and this searching process ends when the color difference no longer decreases. Users can complete the labeling task for the given texture using the confirm button (✔), which will update the color labels in Fig. 4-c1.

To support **T5**, we design a flow-based color visualization to support the interactive refinement of mixing schemes. We refer to it as *ColorFlow* in the following sections. We display the pigment's color rather than its reflectance because conservators, drawing on their practical painting experience, evaluate the success of a mixture chiefly by its final color. ColorFlow features a series of color matrices designed to visualize candidate mixing schemes generated by our greedy-based recommendation algorithm. Each matrix represents pairwise pigment mixing results: the first column and row display two sets of input pigments, while the internal grids show the resulting color when these pigments are combined. The recommended pigment mixture is highlighted in each matrix. In contrast to common color wheel designs (Fig. 6a), the matrix visualization enables more efficient use of display space for representing diverse schemes, owing to the absence of circular layout constraints. However, it remains difficult for users to perceive the relationship between adjacent matrices due to abrupt color transitions. One intuitive design is to arrange matrices horizontally, which allows for

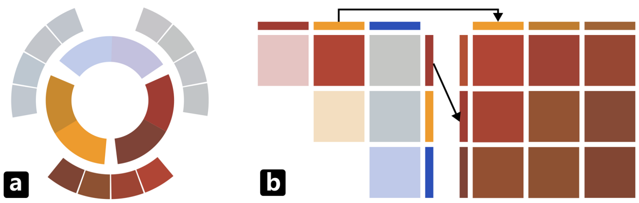

Figure 6: Two alternative designs for ColorFlow: (a) a color wheel where the outer palette is formed by mixing each pair of adjacent pigments from the inner palette; (b) adjacent matrices in a horizontal layout.

easy comparison of adjacent ones. This layout was implemented in the initial version of CAnnotator (Fig. 6b). While this simple layout is easy to read, we observed that users are more interested in understanding the progression of pigment mixtures rather than merely comparing adjacent matrices. Tracking these progression sequences is essential for accurately reconstructing pigment mixing schemes, which is a critical task in color annotation. To address this need, we propose a new layout that connects adjacent matrices using a Sankey-based flow (Fig. 4-c4). This flow visually traces the progression of pigment combinations, showing how the authentic color is approximated through a series of pairwise mixtures. We chose Sankey flows over simple lines because they effectively visualize one-to-many mappings and use darker flows to emphasize direct connections between consecutive pigments. We rotated the matrices to maintain a consistent left-to-right direction for the two Sankey flows. While this rotation reduces the readability of individual matrices, it preserves the overall interpretability of the pigment mixing process. Notice that ColorFlow may be extended to be longer than the visible window after iteratively mixing multiple pigments. Thus, we adopt a focus+context method that provides a scheme bar (Fig. 4-c3) to overview the mixing scheme, where color circles and squares refer to the base pigments and mixed pigments, respectively. Each "+" sign, along with the adjacent pigments, is linked to a color matrix, and the "«" and "<" signs represent the two actions of *See More Quantities* and *See More Mixtures*, respectively.

The following interactions support users in fully exploring pigment mixings. First, users can click the "+" sign to visually highlight the corresponding color matrix. Users can view detailed information for each mixture result by hovering over a grid cell within the color matrix. In the pop-up panel, users can freely explore other schemes by clicking the *See More Quantities* and *See More Mixtures* buttons, which will add a new color matrix and update the scheme bar accordingly. To enable more flexible exploration, we provide two sample controllers, namely, the uniform sampler and the subdivision sampler (Fig. 4-c5). The uniform sampler linearly adjusts pigment quantities in the latest mixing scheme and is set as the default option. The subdivision sampler, in contrast, uses smaller adjustment steps to support finer exploration. Users can switch between the two samplers by clicking the Sankey flows.

7 Usage Scenario

This section presents a use case to demonstrate how CAnnotator can help Bob, a painting conservator, to facilitate color annotation.

After loading an ancient painting (Fig. 4d), Bob uses the **paint-annotation view** to check the degradation degrees of its depicted objects through drag-and-zoom interactions. He enters "pheasant" into the search box (Fig. 4-b1), and the system retrieves 2,805 relevant photos from publicly available image datasets [23]. He then utilizes the segmentation tool (Fig. 4-a2) to extract color-degraded textures from the painting and clicks the texture-tracking button. To improve tracking results, he switches a photo to the painting board by clicking the segmentation button to extract extra tracking references. After clicking the texture-tracking button again, he sees a significant improvement in the new tracking results. Bob next explores the **photo-reference view** with the object-posture and texture-color filters to infer possible colors. He notices two major clusters and several minor clusters in the object-posture filter (Fig. 4-b2). Through the lasso interaction, he quickly grasps the posture features of each cluster. Moreover, Bob swiftly examines images of different variants using the texture-color filter. For example, in the variant shown in Fig. 4-b3, he narrows down the exploration scope to 20 photos using the two filters. Among all possible color choices, he believes that the original one is most likely either light yellow or white. Thus, he picks a photo with a similar posture that meets this color constraint (Fig. 1c) and clicks the confirmation button. To label the textures in the annotation list, Bob obtains their authentic colors using the **pigment-mixing view**. Bob encounters a color that he has never reproduced previously, and he clicks the recommendation button to seek assistance from our algorithm. Once the ColorFlow is updated (Fig. 4-c4), Bob reviews the mixing steps of the recommended scheme and confirms that there are no obvious errors. Therefore, he clicks the confirm button in the scheme bar (Fig. 4-c3) and continues to fine-tune the scheme by using the pigment sliders. After labeling all authentic colors, Bob clicks the download button (Fig. 4d) to end the annotation task.

8 Evaluation

This section first presents two experiments to evaluate the texture-tracking model and the pigment-mixing model. A comparative user study follows to evaluate the proposed visualization designs.

8.1 Model Experiments

To evaluate the performance of texture tracking, we conducted a comparative analysis between the original state-of-the-art model, DeAOT, and an enhanced version of DeAOT augmented with our interaction design (see Sec. 6.2). We selected two common objects in ancient Chinese paintings, namely pheasant and peony. For each object, we first collected a hundred photos that involved various variants and postures from the photo database. We used a background removal model [48] to preprocess all the photos to avoid the interference of backgrounds. Then, we randomly selected three photos and used the texture extraction model to obtain the initial textures for tracking. Each extracted texture was tracked only once across the photo set. To quantitatively evaluate the tracking results, we manually annotated the corresponding textures in the photos as ground truth. Given that textures are composed of repeating visual elements, we consider a tracked mask to be a successful sample if more than half of its pixels belong to the ground truth. Figure 7a presents the average precision and recall of the tracking results,

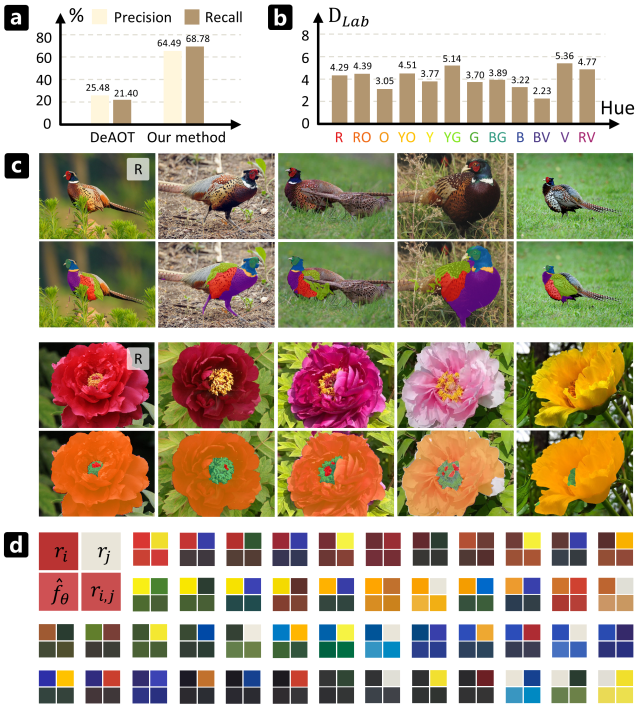

Figure 7: Model evaluation: (a) the quantitative results of the texture-tracking model; (b) the color differences of the test samples used in pigment mixing simulation; (c) the texture-tracking results of two common objects in ancient Chinese paintings; (d) The mixing results of pairwise pigments.

while Fig. 7c illustrates some examples, indicating that our method tracks textures more effectively than the original model.

To evaluate the pigment-mixing model, we performed a supervised learning framework on the public pigment dataset [14], optimizing the loss function defined by Eq. 2. The trained model was then evaluated on the corresponding test samples using the color difference metric D_{Lab} in the *Lab* color space. Figure 7b presents the results by grouping the test samples into twelve clusters based on the hue of their mixed pigments. The average color difference between the predictions and the ground truth across all test samples was 3.87, which is below the threshold of 5, indicating that the two colors are nearly indistinguishable to the naked eye [14]. This result demonstrates that our model effectively simulates the mixing of pigments in the dataset. Additionally, Fig. 7d provides a visual comparison of the predicted and ground-truth pigment mixtures. Each square consists of four sub-squares, representing a mixing experiment of two pigments. The top-left and top-right sub-squares display the original pigment colors. The bottom-left and bottom-right sub-squares show the mixed pigment color and the measured ground-truth color, respectively. The pigment pairs are selected to ensure uniform representation of all pigment types, while the quantity multipliers are randomly sampled for each square. Both quantitative and qualitative results validate that our simulated colors are close to the colors of physically mixed pigments.

8.2 User Study

We conducted a user study comparing CAnnotator with a baseline tool. Twelve participants were invited to complete color annotation for four paintings. Both quantitative and qualitative analyses were used to evaluate the effectiveness and usability of CAnnotator.

Baseline. To the best of our knowledge, there is no existing system that supports a fair comparison with CAnnotator. Therefore, we constructed an ablation version of CAnnotator to serve as the baseline. In this version, the two photo filters in the photo-reference view are disabled, and the ColorFlow in the pigment-mixing view is also removed. The painting-annotation view remains unchanged, as texture extraction is a prerequisite for labeling. The baseline setting is deemed reasonable and sufficient because a) conservators can complete color annotation following a traditional workflow with the baseline tool, and b) there is a one-to-one correspondence between the two challenges and the core components, allowing for fair comparison of CAnnotator and the baseline.

Participants. We invited 12 painting conservators (P1-P12, 11 females and 1 male) to participate in the user study. All participants came from the same institution for studying ancient Chinese paintings, including 9 junior and 3 senior conservators. The junior conservators had at least one year of experience in color annotation, while the senior conservators, with over ten years of experience, could perform color annotation independently. None of them had attended any prior discussions on CAnnotator. All participants had experience with professional design tools featuring rich UI components (e.g., Adobe Photoshop). Moreover, eight participants were familiar with basic visualizations, and five had prior exposure to AI technologies. The junior and senior conservators received ¥50 and ¥100, respectively, as compensation for the 1.5-hour study.

Data and Tasks. We collected 5 bird-and-flower paintings from the Chinese Song dynasty as experimental materials because a) the bird-and-flower painting plays a significant role in realistic paintings, and most of its depicted objects can be found in photos, and b) the Song dynasty represents the pinnacle of realistic paintings in China [15], and most surviving Song dynasty paintings exhibit color degradation. We used one painting to demonstrate the usage of CAnnotator and provided it for user practice. The remaining four paintings were scheduled for formal testing, labeled T1 to T4, respectively. To ensure task independence, the five paintings involve different species. Since the paintings contain multiple objects, participants must browse thousands of photos and assign dozens of labels to restore their degraded textures. Given the limited duration of the study, it is impossible for them to complete full color annotation. Therefore, we simplify the task by preselecting one color-degraded bird in each painting and only requiring participants to identify its reference objects from 100 predefined photos and annotate three predefined texture colors in the formal test. We chose birds for the tasks because they have richer textures than flowers, making color annotation more challenging. These materials were curated in collaboration with two experts, introduced in Section 3, to ensure that participants could complete our tasks.

Procedure. The study consisted of four sessions: a 30-minute training session, a 45-minute testing session, and a 15-minute interview session. In the training session, we first introduced the study objective and the tasks (5 min) to the participants. We then

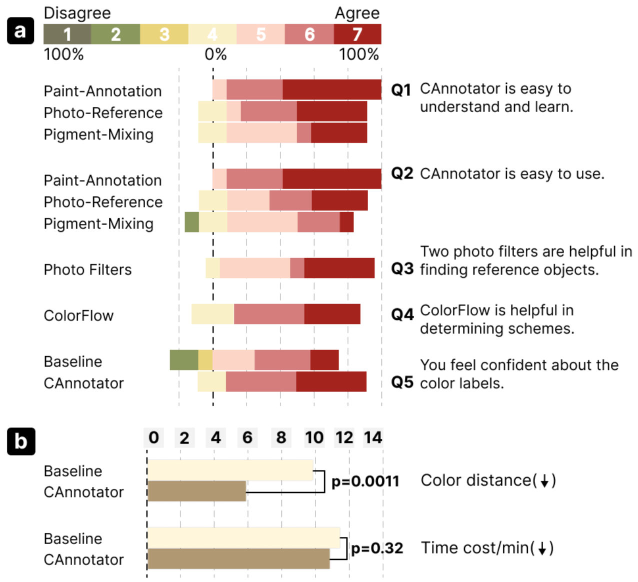

Figure 8: User study: (a) The stacked bar chart displays the distribution of questionnaire results. The length of each stacked sub-bar indicates the proportion of participants who gave the corresponding score. All the bars are horizontally shifted to the left and right, with the sub-bar for a score of 4 at the center. These results indicate that CAnnotator is useful, effective and that it increases participants' confidence in color annotation compared to the baseline. (b) The results of objective metrics also show that CAnnotator can improve the accuracy of labeled colors.

demonstrated the usage of CAnnotator (10 min) and trained them to reproduce the sample case to ensure they understood all system designs (15 min). In the testing session, the participants were asked to complete two formal testing tasks using CAnnotator and the baseline separately. They were encouraged to express their thoughts at any time during the training and testing sessions. The system and task order were counterbalanced to reduce the carryover effect. Finally, the participants were asked to complete a subjective questionnaire and a post-interview.

Quantitative results. We collected the responses of subjective questionnaires and analyzed objective performance (Fig. 8).

Subjective Questionnaire. The questionnaire used 7-point Likert scales to evaluate CAnnotator from three perspectives: *usability* (Q1, Q2), *effectiveness* (Q3, Q4), and *result confidence* (Q5). We did not include Q1-Q4 ratings for the baseline because a) the baseline is an ablated version of CAnnotator whose ratings on ease of learning (Q1) and ease of use (Q2) are expected to be higher or at least the same as CAnnotator, and b) Q3 and Q4 are designed to evaluate the effectiveness of the two key components that were removed in the baseline. Thus, these questions are not applicable to the baseline.

As shown in Fig. 8a, CAnnotator achieved an average score of over 5 across all three views in Q1 (*Paint-Annotation View*: M = 6.5, SD = 0.7; *Photo-Reference View*: M = 6.0, SD = 1.1; *Pigment-Mixing View*: M = 5.6, SD = 1.2) and Q2 (*Paint-Annotation View*: M = 6.5, SD = 0.7; *Photo-Reference View*: M = 5.8, SD = 1.1; *Pigment-Mixing View*: M = 5.0, SD = 1.3), indicating that users could easily understand and use CAnnotator to perform color annotation. We attribute these results to participants' daily painting experiences. During the painting process, they often first mentally break down an object, such as a bird, into meaningful parts like the beak, wings, and feet, and then paint each part individually. This process helps them understand the paint-annotation and photo-reference views, which are organized based on texture units. Furthermore, in real practices, conservators adjust the pigment mixing methods through the two actions described in Section 3.1, which is reflected in the designs of *See More Quantities* and *See More Mixtures*. Such experiences provide conservators with prior knowledge to learn the pigment-mixing view. As for effectiveness, CAnnotator received high ratings in Q3 (M = 5.8, SD = 1.1) and Q4 (M = 5.8, SD = 1.2), demonstrating that users are satisfied with the photo filters and ColorFlow.

Result confidence is a metric that measures users' trust for labeled colors produced by manual annotation (Baseline) and AI-assisted tools (CAnnotator). The rating results show CAnnotator (M = 6.1, SD = 1.1) outperforms the baseline (M = 5.0, SD = 1.8) in Q5 (p = 0.01). Statistically significant results demonstrate that CAnnotator effectively enhances user confidence in color labeling decisions. This indicates that users place a level of trust in CAnnotator comparable to or exceeding that of manual annotation (Baseline). Furthermore, our findings suggest that the integration of ColorFlow visualization—providing real-time, data-driven feedback on color accuracy and blending consistency—plays a critical role in supporting informed decision-making and achieving visually harmonious outcomes in pigment mixing.

Objective metrics. To further validate effectiveness, we calculated, for each participant, the mean color distance between the labeled colors and the target colors predefined by our domain experts. A Wilcoxon Signed-Rank Test conducted over the 12 participants showed that CAnnotator achieved significantly lower color distances than the baseline, with median distances of 5.87 versus 9.88 (W = 0, Z = -3.06, p < 0.05, r = 0.88). Additionally, we analyzed the total time each participant spent completing the four tasks (T1-T4). Although CAnnotator showed marginally better performance than the baseline with median time costs of 10.87 and 11.46, this difference was not statistically significant (p = 0.32). We found that participants had different behaviors when using the two systems. First, participants spent less time finding reference objects with the baseline, as some simply skimmed the photos and made quick selections without careful examination. In contrast, CAnnotator's two photo filters encouraged broader exploration of the photos, leading to more reliable choices. Second, participants spent less time annotating colors with CAnnotator. While CAnnotator required extra time to review the ColorFlow, participants adjusted pigment mixing methods less frequently. These observed behaviors indicate that while incorporating visualizations into data exploration inevitably introduces additional time and cognitive costs, it effectively fosters users' engagement. In tasks that pursue high accuracy, such as color annotation, the additional cost of visualization is a worthwhile trade-off.

Qualitative feedback. Post-experiment interviews provided qualitative feedback and user suggestions for further improvement. During the interview, participants explained their ratings (Q1–Q5) and suggested system enhancements. We categorized the feedback into study-relevant insights and engineering suggestions, and then summarized the main points in each. Two authors independently analyzed the data and resolved the differences through discussion.

ColorFlow has facilitated the collaboration between users and the recommendation algorithm. ColorFlow demonstrates the recommended pigment mixing process step by step, enabling users to identify anomalies and leverage their domain expertise to refine the mixing scheme. We observed that participants typically first examined the Sankey flow to gain an overview of pigment transitions, and then inspected specific matrices. Despite the rotation, participants could still read matrices at a glance since the mixed result is visually highlighted. As P7 stated: *"ColorFlow intuitively demonstrates the process of gradually adjusting pigment mixing schemes in the real world, which helped me understand how these pigments interact to produce the final outcome."* Some participants noted that ColorFlow fostered their trust in the recommendation algorithm. For example, P5 mentioned that *"ColorFlow enabled me to easily review potential anomalies in the pigment mixing process. By observing the color evolution in the matrix visualizations, I could confidently decide whether to accept the recommended schemes."* We surprisingly found that P4 achieved more accurate color mixtures for "ced8e4" using ColorFlow compared to the model's recommendations alone. ColorFlow also allows users to start exploring from any step and refine the schemes that meet their actual needs. P4 said: *"I like the design of the scheme bar, which is simple and clear. With its guidance, I can easily retrace my previous steps during the exploration process. Furthermore, the combination of ColorFlow and the pigment sliders empowers me to control the scheme more precisely."* We conclude that ColorFlow was most positively received, thanks to its automatic recommendation algorithm and transparent visualization of the pigment mixing process.

AI-empowered filters play a crucial role in exploring photo collection. Texture extraction and tracking enable a systematic analysis of object morphology and visual characteristics, allowing users to capture fine-grained information from complex photo content. For the two photo filters in CAnnotator, most participants praised their effectiveness in quickly locating the reference object. As P9 noted: *"I would first use the texture-color filter to eliminate photo objects that are clearly different from the color-degraded object in terms of texture. For example, in the waxwing task, I noticed that the color-degraded object's tail retained distinct traces of red pigment. So, I used the filter to quickly hide photo objects with tails of other colors."* For the object-posture filter, P6 commented that *"I tend to use this filter first to eliminate photo objects with inconsistent poses because I believe that it is easier to distinguish different texture colors in objects with consistent poses."* Specifically, the object-posture filter allows users to bypass sequential browsing by directly accessing structurally aligned photos (e.g., pheasants with matching "head-left" orientation), thereby narrowing the collection to just a few relevant photos for detailed analysis. After the post-analysis, we found no significant difference in the order users applied the two filters. Several participants also highlighted the overview function that the two filters serve in exploring photos. P2 said: *"The two filters provided a clear perspective on the distribution of photos across different dimensions, which informed my subsequent exploration."* For future improvements, P10 suggested adding a candidate reference window to the photo-reference view, as users often encounter similar objects during exploration.

9 Discussion

This section delves into the implications, generalizability, limitations, and potential future directions of CAnnotator. Furthermore, we highlight the design lessons learned that can inform the development of future human-AI collaboration systems.

Implication. Ancient paintings are precious treasures that offer profound glimpses into past human civilizations. To preserve these treasures, conservators have exerted tremendous efforts to combat color degradation. CAnnotator offers a novel visualization tool for reviving degraded colors in ancient paintings by enabling easy color annotation. It employs several AI technologies to automate many of the manual processes currently required by conservators, such as locating texture-consistent reference objects and determining pigment mixing schemes. Moreover, it leverages interactive visualization components to support trusted collaboration between humans and machines, which is meaningful for the HCI community to launch further research. The labeled colors serve as essential references for guiding the restoration of ancient paintings, ensuring that their historical context and authenticity are preserved. Moreover, reliable color analysis of ancient artworks inherently depends on these authentic colors to produce more accurate and trustworthy results. This underscores the critical role of documented, original color information in both conservation [61] and scholarly research [16].

Design Lessons. Recent advancements in deep learning models have led to the widespread adoption of end-to-end approaches to solve real-world problems. However, our research reveals that deep learning alone is insufficient for facilitating accurate color annotation. One key challenge is that color creation can involve complex processes. For example, brown can be obtained by mixing blue and red pigments or ochre and ink pigments, but the production cost of high-purity red and blue pigments is significantly higher than that of ochre and ink pigments. Conservators must consider the painting's historical context when choosing the most suitable method, highlighting the importance of domain expertise in accurate color annotation. This observation underscores the need for human-AI collaboration to develop effective visual analytic systems. To address this gap, CAnnotator proposes a flow-based color visualization that enables the progressive simulation of pigments, bridging the knowledge gap between deep learning models and human professionals.

For data exploration, users often adapt their navigation granularity based on the relevance of different regions within the search space. Irrelevant regions are typically traversed with coarse steps to accelerate scanning, while relevant ones require fine-grained, incremental exploration to hone in on specific targets. Traditional color palettes (e.g., color wheel) often fail to support such fluid transitions between exploration modes, as they typically enforce fixed step sizes that hinder adaptability. CAnnotator provides an example that integrates two sample controllers that let users dynamically adjust both the scope and granularity of their exploration (see

Fig. 4-c4). This flexibility reduces manual efforts when distinguishing between similar pigment mixing schemes, thereby enhancing both the efficiency and effectiveness of the exploration process.

Generalizability. CAnnotator is built on well-established deep learning models that have presented satisfactory extensibility across various image datasets. Although we only demonstrate the application of CAnnotator on Chinese flower-and-bird paintings, the high-level flexibility of our models allows CAnnotator to extend to other realistic themes. For instance, it can be applied to paintings featuring other plants and animals because the models have been trained on large-scale image datasets encompassing these themes. We argue that CAnnotator is most effective for color annotation involving realistic paintings with well-defined object textures. Compared to realistic paintings, abstract paintings (e.g., Wassily Kandinsky's paintings) may not be fully supported by CAnnotator because they predominantly express the painter's inner world. However, it does not exclude the possibility of extending our approach to these paintings or incorporating additional features that can help address their unique challenges. Although CAnnotator was built for realistic paintings, its workflow can also handle recognizable figures in abstract artworks. For instance, the distorted visual elements can be first restored into the underlying geometry and texture. Then, CAnnotator can be applied to annotate color labels for degraded elements. Furthermore, the integrated features of CAnnotator have the potential to inspire tasks in other domains. For example, the texture tracking model and the two photo filters can be utilized for non-expert image annotation [10]. The ColorFlow can be employed for color reproduction studies in other forms of artwork. By leveraging the power of human-AI collaboration, CAnnotator could contribute to preserve other types of cultural heritage.

Limitations and Future Work. Artists often employ diverse drawing techniques in painting scenarios, such as layering multiple pigments within the same area to achieve specific visual effects [52]. Additionally, the choice of binder in paints can influence perceived colors on the canvas [55]. Future research could integrate the extra physical variables to predict colors with more precision. In addition, future work could incorporate evaluation methods, such as CSO [41], to better compare the impact of different design choices on the color annotation of ancient paintings. Moreover, due to limited prior knowledge, users may not be aware of potential prediction errors in the pigment mixing model. Expanding the dataset and adopting more advanced model architectures may further enhance model performance. Finally, the efficiency of the texture tracking model may be compromised when processing large-scale datasets, such as those containing millions of photos. To address this limitation, parallel computing frameworks can be employed to accelerate texture-tracking workflows, ensuring scalability.

10 Conclusion

In this study, we propose CAnnotator, a novel color annotation tool for degraded ancient paintings. The system has been tailored to meet the domain requirements and employs several deep-learning models to tackle challenges related to texture tracking and pigment mixing. The painting-annotation, photo-reference, and pigment-mixing views have been developed to support the general workflow of color annotation, which includes gathering reference images and reproducing colors through traditional pigments. We evaluate CAnnotator by introducing a use case, conducting model experiments and an in-lab user study. The results indicate that CAnnotator can improve conservators' efficiency and increase their confidence in the annotations. We plan to apply CAnnotator to more types of fine art or other types of artworks in the future.

Acknowledgments

This work was supported by NSFC (62302440, U22A2032) and Zhejiang Provincial Natural Science Foundation of China under Grant No. (LD25F020003). The work was also supported by the Fundamental Research Funds for the Central Universities.

References

[1] Panos Achlioptas, Maks Ovsjanikov, Kilichbek Haydarov, Mohamed Elhoseiny, and Leonidas Guibas. 2021. Artemis: Affective Language for Visual Art. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, 11569–11579.

[2] Shehzad Afzal, Sohaib Ghani, Mohamad M. Hittawe, Sheikh F. Rashid, Omar M. Knio, Markus Hadwiger, and Ibrahim Hoteit. 2023. Visualization and Visual Analytics Approaches for Image and Video Datasets: A Survey. *ACM Transactions on Interactive Intelligent Systems* 13, 1 (2023), 1–41.

[3] Elad Aharoni-Mack, Yakov Shambik, and Dani Lischinski. 2017. Pigment-Based Recoloring of Watercolor Paintings. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering.* ACM, 1–11.

[4] Meta AI. 2023. Segment Anything Demo. Website. https://segment-anything.com/demo.

[5] Seyed A. Amirshahi, Gregor U. Hayn-Leichsenring, Joachim Denzler, and Christoph Redies. 2014. Jenaesthetics Subjective Dataset: Analyzing Paintings by Subjective Scores. In *Proceedings of European Conference on Computer Vision Workshop.* Springer, 3–19.

[6] Donald Bertucci, Md Montaser Hamid, Yashwanthi Anand, Anita Ruangrotsakun, Delyar Tabatabai, Melissa Perez, and Minsuk Kahng. 2022. DendroMap: Visual Exploration of Large-Scale Image Datasets for Machine Learning with Treemaps. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 320–330.

[7] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology.* 1–14.

[8] Rosina Buckland. 2013. *Painting Nature for the Nation, 1886–1901.* Brill, 126–163.

[9] Berta Carrión-Ruiz, Gabriel Riutort-Mayol, Adolfo Molada-Tebar, José L. Lerma, and Valentín Villaverde. 2021. Color Degradation Mapping of Rock Art Paintings using Microfading Spectrometry. *Journal of Cultural Heritage* 47 (2021), 100–108.

[10] Chia-Ming Chang, Chia-Hsien Lee, and Takeo Igarashi. 2021. Spatial Labeling: Leveraging Spatial Layout for Improving Label Quality in Non-Expert Image Annotation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, 1–12.

[11] Changjian Chen, Zhaowei Wang, Jing Wu, Xiting Wang, Lan-Zhe Guo, Yu-Feng Li, and Shixia Liu. 2021. Interactive Graph Construction for Graph-Based Semi-Supervised Learning. *IEEE Transactions on Visualization and Computer Graphics* 27, 9 (2021), 3701–3716.

[12] Changjian Chen, Jing Wu, Xiaohan Wang, Shouxing Xiang, Song-Hai Zhang, Qifeng Tang, and Shixia Liu. 2021. Towards Better Caption Supervision for Object Detection. *IEEE Transactions on Visualization and Computer Graphics* 28, 4 (2021), 1941–1954.

[13] Lieu-Hen Chen, Meng-Feng Tsai, Chien-Hui Hsu, and Yu-Sheng Chen. 2013. Aging and Reverse-aging Traditional Chinese Painting Images Based on Web-Mining. *New Generation Computing* 31, 4 (2013), 285–309.

[14] Meiyun Chen, Yabo Huang, Shengping Chang, and Ming Ouhyoung. 2019. Prediction Model for Semitransparent Watercolor Pigment Mixtures Using Deep Learning with a Dataset of Transmittance and Reflectance.

[15] Shuihua Chen. 2024. *Xingli Liangquan: Songhua zhong de Niaolei [Harmony of Form and Essence: Birds in Song Dynasty Paintings].* Zhejiang Ancient Books Publishing House.

[16] Xiaojiao Chen, Qinghua Liu, Yonghao Chen, Ruihan Wang, Yang You, Wanxin Deng, Wei Chen, and Xiaosong Wang. 2024. ColorNetVis: An Interactive Color Network Analysis System for Exploring the Color Composition of Traditional Chinese Painting. *IEEE Transactions on Visualization and Computer Graphics* 30, 6 (2024), 2916–2928.

[17] Yangming Cheng, Liulei Li, Yuanyou Xu, Xiaodi Li, Zongxin Yang, Wenguan Wang, and Yi Yang. 2023. Segment and Track Anything. *arXiv preprint arXiv:2305.06558* (2023).

[18] Yifan Cheng, Xiaoming Yang, and Wei Zhong. 2019. Hanxiu Secai Tezheng Biaozhu de Tanxi [Analysis on the Color Features of Han Embroidery]. *Fushi Daokan* 8, 6 (2019), 68–78.
[19] Yee Chiang. 1964. *The Chinese Eye: An Interpretation of Chinese Painting.* Midland Books.
[20] Cassidy J. Curtis, Sean E. Anderson, Joshua E. Seims, Kurt W. Fleischer, and David H. Salesin. 1997. Computer-Generated Watercolor. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques.* ACM, 421–430.
[21] Hai Dang, Lukas Mecke, and Daniel Buschek. 2022. GANSlider: How Users Control Generative Models for Images using Multiple Sliders with and without Feedforward Information. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, 1–15.
[22] Nouchka De Keyser, Fréderique Broers, Frederik Vanmeert, Steven De Meyer, Francesca Gabrieli, Erma Hermens, Geert Van der Snickt, Koen Janssens, and Katrien Keune. 2022. Reviving degraded colors of yellow flowers in 17th century still life paintings with macro- and microscale chemical imaging. *Science Advances* 8, 23 (2022), eabn6344.
[23] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, 248–255.
[24] D. R. Duncan. 1940. The Colour of Pigment Mixtures. *Proceedings of the Physical Society* 52, 3 (1940), 390.
[25] Sarah Everts. 2016. Van Gogh's Fading Colors Inspire Scientific Inquiry. *Chemical & Engineering News* 94, 5 (2016), 32–33.
[26] David H. Foster, Kinjiro Amano, Sérgio MC. Nascimento, and Michael J. Foster. 2006. Frequency of Metamerism in Natural Scenes. *Journal of the Optical Society of America A* 23, 10 (2006), 2359–2372.
[27] Noa Garcia, Benjamin Renoust, and Yuta Nakashima. 2020. ContextNet: Representation and Exploration for Painting Classification and Retrieval in Context. *International Journal of Multimedia Information Retrieval* 9 (2020), 17–30.
[28] Chet S. Haase and Gary W. Meyer. 1992. Modeling Pigmented Materials for Realistic Image Synthesis. *ACM Transactions on Graphics* 11, 4 (1992), 305–335.
[29] Devin Haslam, Soad Ibrahim, and Ayman Elmesalami. 2020. Color Restoration Survey and an Overdetermined System for Color Retrieval from Faded Images. In *Recent Advances in Engineering Mathematics and Physics.* 291–321.
[30] C. Hogan and F. Da Pieve. 2015. Colour degradation of artworks: an ab-initio approach to X-ray, electronic and optical spectroscopy analyses of vermilion photodarkening. *Journal of Analytical Atomic Spectrometry* 30, 3 (2015), 588–598.
[31] Laura Hollink, Guus Schreiber, Jan Wielemaker, and Bob Wielinga. 2003. Semantic Annotation of Image Collections. In *Proceedings of Knowledge Markup and Semantic Annotation Workshop.* Vrije Universiteit Amsterdam, 41–48.
[32] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. In *Proceedings of IEEE International Conference on Computer Vision.* IEEE, 4015–4026.
[33] M. Kotsidi, G. Gorgolis, M. G. Pastore Carbone, G. Anagnostopoulos, G. Paterakis, G. Poggi, A. Manikas, G. Trakakis, P. Baglioni, and C. Galiotis. 2021. Preventing colour fading in artworks with graphene veils. *Nature Nanotechnology* 16, 9 (2021), 1004–1010.
[34] Paul Kubelka. 1948. New Contributions to the Optics of Intensely Light-Scattering Materials. Part I. *Josa* 38, 5 (1948), 448–457.
[35] Paul Kubelka and Franz Munk. 1931. An Article on Optics of Paint Layers. *Zeitschrift für Technische Physik* 12, 593–601 (1931), 259–274.
[36] Liza Leslie, Tat-Seng Chua, and Jain Ramesh. 2007. Annotation of Paintings with High-Level Semantic Concepts Using Transductive Inference and Ontology-Based Concept Disambiguation. In *Proceedings of the 15th ACM International Conference on Multimedia.* ACM, 443–452.
[37] Hong Li. 2012. Tiannv Lai XiangShi, Jianghua Yu Ranyi [Color Analysis of Attire in Tang Dynasty Noblewomen's Portraiture]. *Dazhong Wenyi* 24 (2012), 20–21.
[38] Shixia Liu, Changjian Chen, Yafeng Lu, Fangxin Ouyang, and Bin Wang. 2018. An Interactive Method to Improve Crowdsourced Annotations. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 235–245.
[39] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology.* ACM, 1–17.
[40] Yue Lu, Chao Guo, Xingyuan Dai, and Feiyue Wang. 2022. ArtCap: A Dataset for Image Captioning of Fine Art Paintings. *IEEE Transactions on Computational Social Systems* (2022), 1–12.
[41] Wendy E. Mackay and Joanna McGrenere. 2025. Comparative Structured Observation. *ACM Transactions on Computer-Human Interaction* 32, 2 (2025), 1–27.
[42] MinIO. 2023. High Performance Object Storage for Modern Data Lakes and Data LakeHouses. Website. https://min.io.
[43] Meredith R. Morris, Jazette Johnson, Cynthia L. Bennett, and Edward Cutrell. 2018. Rich Representations of Visual Content for Screen Reader Users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, 1–11.
[44] Tamara Munzner. 2009. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 921–928.
[45] National Palace Museum. 2024. National Palace Museum. Website. https://www.npm.gov.tw.
[46] Observable. 2023. The JavaScript Library for Bespoke Data Visualization. Website. https://d3js.org.
[47] Noboru Ohta and Alan Robertson. 2006. *Colorimetry: Fundamentals and Applications.* Wiley.
[48] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar Zaiane, and Martin Jagersand. 2020. U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection. *Pattern Recognition* 106 (2020), 107404.
[49] Jiangyong Qu, Naifa Liu, Xinkang Bao, and Xiaoli Wang. 2009. Phylogeography of the Ring-necked Pheasant (Phasianus Colchicus) in China. *Molecular Phylogenetics and Evolution* 52, 1 (2009), 125–132.
[50] Frano Rajič, Lei Ke, Yu-Wing Tai, Chi-Keung Tang, Martin Danelljan, and Fisher Yu. 2023. Segment Anything Meets Point Tracking. *arXiv:2307.01197* (2023).
[51] Craig Rodkin. 2020. Describing Figures for ACM Publications. Website. https://authors.acm.org/proceedings/production-information/describing-figures.
[52] Ashok Roy. 1999. The National Gallery Van Dycks: Technique and Development. *National Gallery Technical Bulletin* 20 (1999), 50–83.
[53] Andreza Sartori, Dubravko Culibrk, Yan Yan, and Nicu Sebe. 2015. Who's Afraid of Itten: Using the Art Theory of Color Combination to Analyze Emotions in Abstract Paintings. In *Proceedings of the 23rd ACM International Conference on Multimedia.* ACM, 311–320.
[54] Šárka Sochorová and Ondřej Jamriška. 2021. Practical Pigment Mixing for Digital Painting. *ACM Transactions on Graphics* 40, 6 (2021), 1–11.
[55] Wazeer H. Solangi, Zulfiqar A. Noonari, Asghar A. Channa, Muhammad Q. Khan, and Abdul B. Siyal. 2014. Influence of Binders and Thickeners of Pigment Printing Paste on Light Fastness and Crocking Fastness of the Fabric. *International Journal of Science and Research* 3, 5 (2014), 1024–1033.
[56] Tomas Souper, Ana C Morgado, Ana Marques, Inês Silva, and Luís Rosado. 2023. Improving Color Mixture Predictions in Ceramics using Data-centric Deep Learning. In *Proceedings of the 8th International Conference on Machine Learning Technologies.* 221–229.
[57] Meat Open Source. 2023. React. Website. https://react.dev.
[58] Melissa E. Swift, Wyatt Ayers, Sophie Pallanck, and Scott Wehrwein. 2022. Visualizing the Passage of Time with Video Temporal Pyramids. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 171–181.
[59] Jianchao Tan, Stephen DiVerdi, Jingwan Lu, and Yotam Gingold. 2018. Pigmento: Pigment-Based Image Analysis and Editing. *IEEE Transactions on Visualization and Computer Graphics* 25, 9 (2018), 2791–2803.
[60] Tan Tang, Yanhong Wu, Yingcai Wu, Lingyun Yu, and Yuhong Li. 2021. VideoModerator: A Risk-aware Framework for Multimodal Video Moderation in E-Commerce. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 846–856.
[61] 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 *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology.* 1–15.
[62] tiangolo. 2023. FastAPI. Website. https://github.com/tiangolo/fastapi.
[63] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008).
[64] Xiao Xie, Xiwen Cai, Junpei Zhou, Nan Cao, and Yingcai Wu. 2018. A Semantic-Based Method for Visualizing Large Image Collections. *IEEE Transactions on Visualization and Computer Graphics* 25, 7 (2018), 2362–2377.
[65] Songhua Xu, Haisheng Tan, Xiantao Jiao, Francis C. M. Lau, and Yunhe Pan. 2007. A Generic Pigment Model for Digital Painting. In *Proceedings of Computer Graphics Forum.* Wiley, 609–618.
[66] Zongxin Yang, Yunchao Wei, and Yi Yang. 2021. Associating Objects with Transformers for Video Object Segmentation. In *Advances in Neural Information Processing Systems.* Curran Associates, Inc., 2491–2502.
[67] Zongxin Yang and Yi Yang. 2022. Decoupling Features in Hierarchical Propagation for Video Object Segmentation. In *Advances in Neural Information Processing Systems.* Curran Associates, Inc., 36324–36336.
[68] Marchenko Yelizaveta, Chua Tat-Seng, and Aristarkhova Irina. 2005. Analysis and Retrieval of Paintings Using Artistic Color Concepts. In *IEEE International Conference on Multimedia and Expo.* IEEE, 1246–1249.
[69] Marchenko Yelizaveta, Chua Tat-Seng, and Jain Ramesh. 2006. Semi-Supervised Annotation of Brushwork in Paintings Domain Using Serial Combinations of Multiple Experts. In *Proceedings of the 14th ACM International Conference on Multimedia.* ACM, 529–538.
[70] Jan Zahálka, Marcel Worring, and Jarke J. Van Wijk. 2020. II-20: Intelligent and Pragmatic Analytic Categorization of Image Collections. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 422–431.