Study on the Color Characteristics of Reproduced Oil Paintings Using a Machine Learning Algorithm

Hyeong Rok Song [1], Young Hoon Jo [1]

[1] Dept. of Cultural Heritage Conservation Sciences, Kongju National University, Kongju, Republic of Korea – skv9483@naver.com, joyh@kongju.ac.kr

Keywords: Color Characteristics, Digital Color Reproduction, Machine Learning, Pixel-based Statistics, Oil Painting.

Abstract

Accurate recording of the colors in cultural heritage is essential, where color data is crucial for various applications, including conservation, restoration, research, analysis, and archiving. Recently, advancements in digital color reproduction techniques have emerged, enabling precise documentation of colors using digital photography and image processing. This approach reproduces colors that closely match those of the cultural heritage object by correcting the image and profiling the camera. Notably, the correction process utilizes the device-independent CIE $L^*a^*b^*$ color space to ensure that the reproduced colors are consistent across different devices. Moreover, digital images consist of pixels, which facilitate data-driven statistical analysis. This study focused on digital color reproduction for Korean modern oil paintings, following a systematic process that included photography, digital color correction, and digital color space configuration. To enhance the reliability of color reproduction, it compared the spectral color measurement results of a color chart with the color differences observed in the reproduced images. The study then plotted the CIE $L^*a^*b^*$ color distribution of the images in a three-dimensional graph, where approximately 30 million pixels were classified using the K-means machine learning algorithm. Based on these classification results, representative colors were extracted, along with various analytical outcomes, such as the number of pixels, representative CIE $L^*a^*b^*$ color coordinates, and the percentage composition of each representative color. This research enabled oil paintings to be documented with accurate colors, and the resultant image data were used to extract representative colors using a machine learning algorithm. This method, wherein representative colors are derived through color reproduction, offers valuable insights into the color usage patterns and chromatic painting techniques of an artist, and even the authenticity of artworks.

1. Introduction

Color is a crucial element in the visual arts and reflects the artist’s intentions and emotions. In particular, the use of warm or cool colors significantly influences viewers’ impression and interpretation of the work. In addition, the pigments and color-mixing techniques generally used during the period in which the artist was active contribute to the unique chromatic characteristics of a painting, enabling the identification of historical features and the artist’s individual style. However, color analysis of artworks has traditionally relied on visual observation, which is inherently subjective and influenced by the viewer’s perceptions, the condition of the work, and the environment in which it is exhibited. Therefore, an objective and quantitative method for analyzing color in artworks is needed(Cramer et al., 2020). While spectrophotometers can be used to accurately quantify color in the device-independent CIE $L^*a^*b^*$ color space, this method requires physical contact with the artwork’s surface, and its applicability is limited by the size of the device’s aperture, which constrains the measurable area(Sanmartín et al., 2014). In particular, contact-based methods are discouraged in art analysis due to the risks they pose such as secondary contamination and physical damage to the artwork. To address these limitations, this study employs photographs obtained through digital color reproduction imaging to analyze CIE $L^*a^*b^*$ color data with machine learning–based statistical methods. This non-contact approach makes it possible to interpret the overall chromatic tendencies of oil paintings comprehensively and objectively.

2. Materials and Methods

2.1 Oil Paintings and Un-soung Pai

This study focuses on two oil paintings by the artist Un-soung Pai, the first Korean artist to study in Europe, graduating from the Universität der Künste Berlin. He presented solo exhibitions at venues such as Charpentier in France and was active in the European art scene. Through his use of realist techniques, he significantly influenced modern Korean art. The two oil paintings analyzed in this study, *Tug of War* and *still painting*, were painted in the 1930s and involve a variety of colors in their composition. *Tug of War* depicts a rural Korean village scene, harmoniously blending Eastern landscapes with Western painting styles. Meanwhile, *still painting* centers on a vibrantly colored arrangement of flowers and a flower vase. The placement of a framed image in the background provides a sense of spatial depth.

Still Painting(1930) *Tug of War(1930)*

Figure 1. Subject oil paintings.

2.2 Methods

To analyze the chromatic characteristics of two oil paintings created by the same artist in the same period, a machine learning–based approach was adopted. Pixel-level color accuracy was enhanced with a digital color reproduction imaging technique, conducted in consideration of a color management system. Device profiling was performed for the monitor (32UD59, LG

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-1395-2025 |

Electronics Inc) and the camera (5Ds, Canon) that enabled color correction in the device-independent color space CIE $L^*a^*b^*$. Color correction was based on spectrophotometric data obtained from a 140-patch color chart (ColorChecker Digital SG, Calibrite), with reference to CIE $L^*a^*b^*$ values. Adobe Lightroom Classic was used to process the raw image data, and X-Rite i1Profiler was utilized to generate device profiles. The environmental variables obtained during image capture were recorded, with lighting conditions measured at a color rendering index (CRI) of Ra 97 and a correlated color temperature (CCT) of 5200 K. The spectral characteristics of the illumination closely matched those of the D50 standard illuminant, ensuring optimal conditions for color reproduction. To assess the accuracy of color reproduction, the ΔE00 color difference between the color chart and the rendered image was calculated. Subsequently, the images were converted to the CIE $L^*a^*b^*$ color space on a pixel-by-pixel basis, and the color data were classified into 10 representative clusters, using the K-means clustering algorithm(Papia and Kondi, 2025).

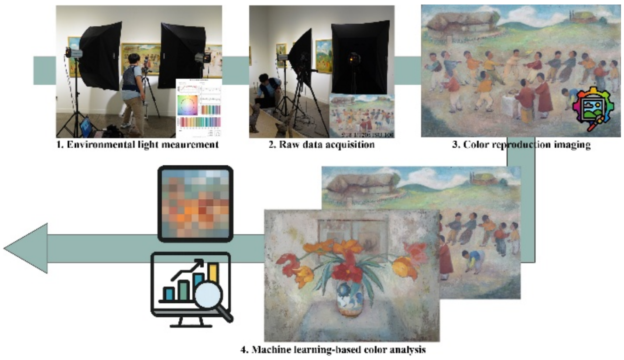


Figure 2. Workflow of Digital Color Reproduction and Machine Learning-Based Color Analysis of Oil Paintings.

3. Digital Color Reproduction Process and Results

3.1 Monitor Calibration and Color Chart Measurement

To accurately reproduce color, it is essential that the monitor provide stable and consistent color output. In this study, 461 color patches were displayed on the monitor and were measured to adjust the RGB signals and brightness levels, enabling color rendering in a way that is suitable for the observer's viewing environment. In addition, 140 reference colors from a color chart were measured using a spectrophotometer to obtain CIE $L^*a^*b^*$ values, which were then established as target reference values for the color correction.

3.2 Oil Painting Color Reproduction Imaging

RAW data from the digital camera contained the recorded response between light and the image sensor and required processing through digital development techniques. Color correction was performed with reference to the CIE $L^*a^*b^*$ values on a color chart measured with a spectrophotometer, adjusting the image to match corresponding colors. Numerical correction based on the CIE $L^*a^*b^*$ color space was conducted with processes such as lens distortion correction, exposure adjustment, white balance (color temperature), brightness, and chroma adjustment. As the resulting data do not inherently contain a defined color space or profile, it is essential to properly profile for an appropriate RGB color space. To achieve this, the colors within the camera's gamut were compared with the reference colors of the color chart in the CIE $L^*a^*b^*$ space, allowing for the selection and application of an optimal RGB color space for accurate image reproduction. The overall color reproduction was carried out using the same method used by Song and Jo (2021).

Figure 3. Color Reproduction Imaging Method.

3.3 Camera Color Accuracy Analysis

The CIE $L^*a^*b^*$ values extracted from the color-corrected chart image were compared with spectrophotometer measurements, showing average ΔE_{00} values of 1.1 (range: 0.1–4.0) for *still painting* and 1.0 (range: 0.1–5.7) for *Tug of War*. According to ISO 12647-8(ISO, 2021), these values represent very small color differences, indicating that the data are suitable for pixel color analysis.

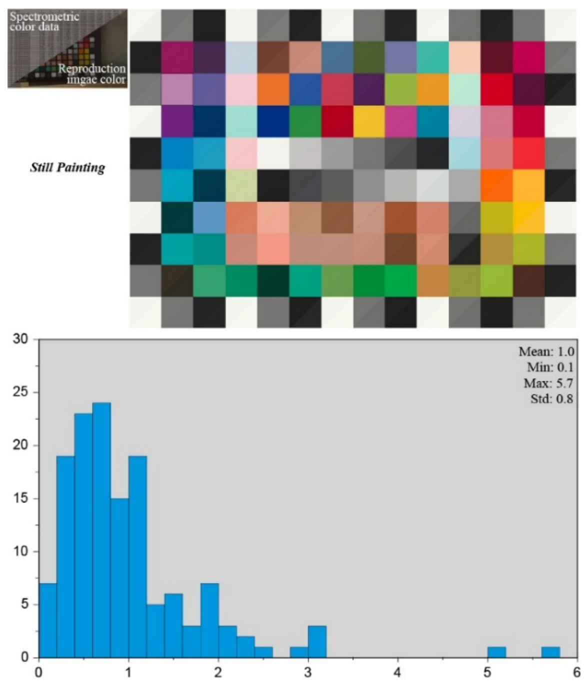


Figure 4. Color ΔE_{00} analysis result of *still painting*.

https://doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-1395-2025

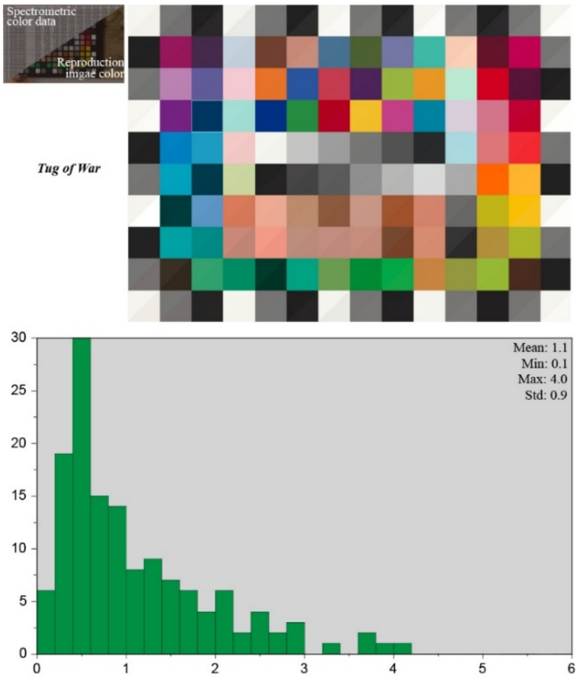


Figure 5. Color ΔE_{00} analysis result of *Tug of war*.

3.4 Color Reproduction Result

In the color reproductions, non-painting areas were removed using cropping. The *Still painting* image was 6,912 × 5,184 pixels, for a total of 35,831,808 pixels. The image of *Tug of War* image was 7,240 × 4,846 pixels, or 35,085,040 pixels in total. As the color reproduction process was based on the CIE $L^*a^*b^*$ color space, each pixel's $L^*a^*b^*$ values were consistent across all devices. This enabled the construction of objective image data that are suitable for analyzing oil paintings.

Figure 6. Color Reproduction Result Image of the Oil Paintings.

4. Machine Learning-Based Color Analysis of Oil Paintings

4.1 Still Painting

The color-reproduced image of *Still painting* was used as data accurately reflecting the original colors of the oil painting. Color analysis was conducted across the entire painting surface. This method involved calculating the CIE $L^*a^*b^*$ values for all the pixels and mapping them in a scatter plot in a three-dimensional color space. The analysis was performed using MATLAB R2024, a specialized software platform for machine learning. and the background color appeared predominantly in earth tones. The K-means algorithm was used to classify the color data into 10 distinct clusters. The classification identified that earth tones, greens, and reddish-browns were dominant. The reds and oranges were prominent in the flower petals. Green, found in the leaf areas and the patterns on the vase, was also clearly observed. In addition, the painting frame behind the vase

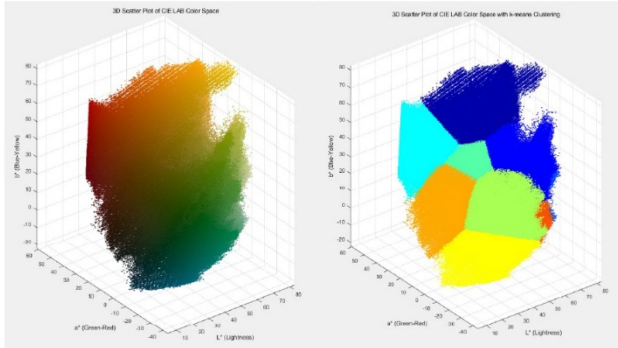


Figure 7. Machine Learning-Based Color Clustering of still *painting* in CIE $L^*a^*b^*$.

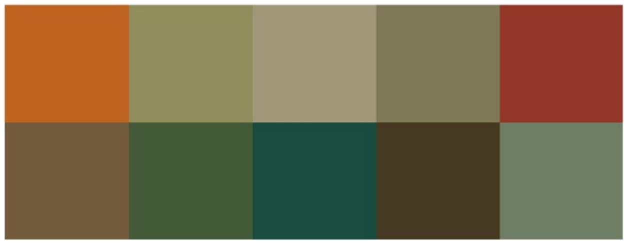
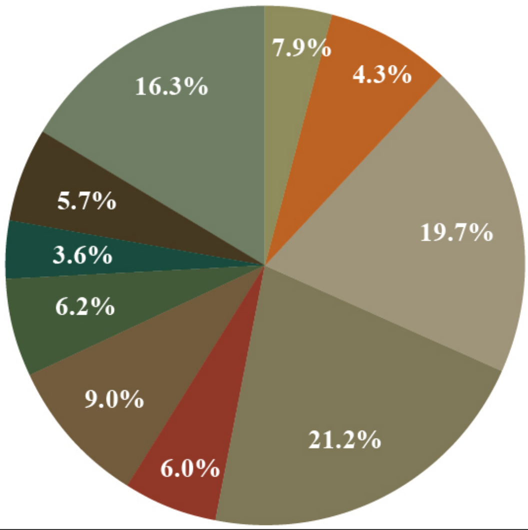


Figure 8. Results of the Top 10 Representative Colors Extracted from *still painting*.



Color	L*	a*	b*	%	Pixel
①	53.14	29.42	59.21	4.3	1,524,365
②	53.32	-6.80	30.20	7.9	2,846,535
③	63.69	-1.23	17.94	19.7	7,061,264
④	51.34	-2.93	21.47	21.2	7,608,575
⑤	36.86	35.97	34.05	6.0	2,155,314
⑥	41.18	4.09	24.43	9.0	3,209,206
⑦	35.98	15.66	19.52	6.2	2,228,396
⑧	28.60	-20.82	4.28	3.6	1,298,761
⑨	24.30	2.15	21.63	5.7	2,053,097
⑩	52.04	-10.34	14.26	16.3	5,846,295

Figure 9. Color distribution of *still painting* based on K-means clustering in CIE $L^*a^*b^*$.

Using K-means clustering, approximately 3.5 million pixels were grouped to extract representative colors. Among these, Color 4, a mid-tone beige, had the highest prevalence, at 21% (approximately 7.6 million pixels), followed by Color 3, a light beige tone, at 19.7% (approximately 7 million pixels). By contrast, Color 8, a teal color found in the leaves, showed the lowest proportion, at 3.6% (approximately 1.2 million pixels).

4.2 Tug of War

The color analysis for *Tug of War* was conducted using accurately color-reproduced data with an average color difference (ΔE_{00}) of 1.1. The CIE $L^*a^*b^*$ values were extracted for the entire area of the painting. The data were first visualized as a 3D scatter plot in the CIE $L^*a^*b^*$ color space, in which each point represented a single pixel from the image. The analysis covered approximately 35 million pixels, representing a naturally continuous distribution of colors across the space. Subsequently, the data were clustered using the K-means algorithm in the same color space, allowing for a clear visual understanding of the segmentation of the overall color distribution.

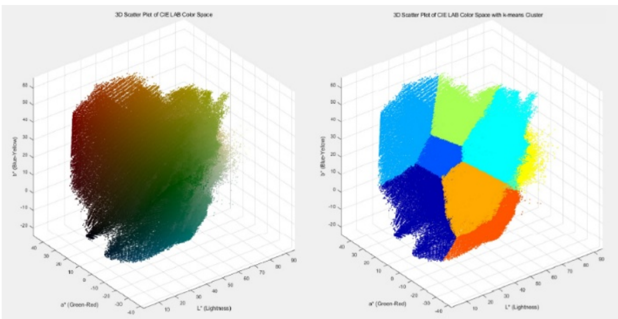
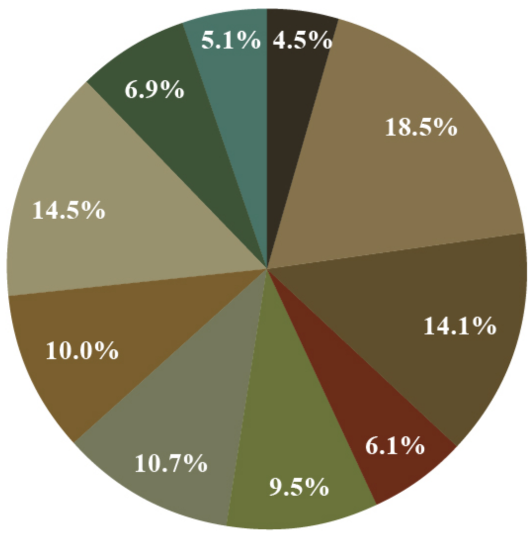


Figure 10. Machine Learning-Based Color Clustering of *Tug of War* in CIE $L^*a^*b^*$.

An analysis of the 10 representative colors extracted indicated that the red tones correspond to clothing elements, such as jackets and traditional garments. Dark brown tones appear in areas such as the ground and the hair of the figures, whereas mid-tone beige is commonly found in the background and as parts of the clothing. In particular, green tones are associated with the distant mountains and with grassy fields. Overall, the representative colors accurately reflect the actual color distribution used across the painting. Using the K-means algorithm, approximately 35 million pixels from *Tug of War* were clustered to extract representative colors. Among these, Color 2, a light beige tone, accounted for the largest share, at 18.5% (approximately 6.5 million pixels). This color was widely used in the background, in the faces of the figures, and in parts of the clothing, and it served as a foundational tone providing visual balance to the overall composition. Color 7, an olive green tone, represented 14.5% (approximately 5.1 million pixels) of the image and was primarily found in the fields, garments, and grassy areas of the painting. Other colors, such as the reddish brown of Color 4 (14.1%, approximately 5.0 million pixels), the mid-tone beige of Color 3 (10.0%, approximately 3.5 million pixels), and the dark olive of Color 6 (10.7%, approximately 3.8 million pixels) were relatively evenly distributed across the painting. By contrast, Color 1, a dark brown tone, featured the lowest proportion at 4.5% (approximately 1.6 million pixels), appearing mainly in the depiction of hair and shadowed areas of the figures.

Figure 11. Results of the Top 10 Representative Colors Extracted from *Tug of War*.



Color	L*	a*	b*	%	Pixel
①	17.57	0.24	11.56	4.5	1,588,973
②	50.51	0.96	27.81	18.5	6,480,360
③	34.76	1.85	28.08	14.1	4,958,106
④	26.64	26.85	32.59	6.1	2,136,457
⑤	47.80	-12.79	35.40	9.5	3,332,277
⑥	50.54	-6.55	17.40	10.7	3,753,034
⑦	42.86	4.13	38.14	10.0	3,515,604
⑧	61.55	-3.30	22.47	14.5	5,085,415
⑨	33.57	-14.91	17.27	6.9	2,432,961
⑩	46.19	-17.39	2.33	5.1	1,801,853

Figure 12. Color distribution of *Tug of War* based on K-means clustering in CIE $L^*a^*b^*$.

5. Discussion

Comparing the K-means clustering results for still *painting* and Tug of War shows that both analyses were conducted in a similar range, with approximately 35 million pixels in each image. In still *painting*, the dominant colors were light beige (CIE L*a*b*: 51.34, -2.93, 21.47) and mid-tone beige (CIE L*a*b*: 63.69, -1.23, 17.94). Meanwhile, in Tug of War, light beige and reddish brown were the primary colors. These major tones are interpreted as primarily used in the background areas. In both paintings, it was confirmed that Pai predominantly used beige-based tones. Additionally, red tones used in flowers and clothing in still *painting* (CIE L*a*b*: 36.86, 35.97, 34.05) and Tug of War (CIE L*a*b*: 26.64, 26.85, 32.59) had smaller proportions, serving as accent colors.

Item	*still painting*	*Tug of War*
Total Pixel	35,831,808	35,085,040
Main color	Light beige, mid beige	Light beige, reddish brown
Highest	Mid-tone beige(21.2%)	Light beige(18.5%)
Lowest	Green(3.6%)	Dark brown (4.5%)

Table 1. Comparison of Color Composition between *still painting* and *Tug of War.*

This method of extracting representative colors is an effective analytical approach for identifying an artist’s color usage patterns(Hao and Qi, 2022). This study focused on two oil paintings by Pai, but the artist’s overall color preferences and tendencies could be identified more precisely through the collection and statistical analysis of a larger dataset. In particular, this method enables an objective quantification of favored hues, color contrasts, and distributions of brightness, enhancing its applicability. Moreover, by training models on the colors used in an artist’s works, the extracted color data may serve as technical evidence for the authenticity of unknown pieces. As the demand for cultural products continues to grow, the use of an artist’s representative color palette presents opportunities for transforming their color identity into a branding asset in developing derivative goods and content. Finally, examining the chromatic patterns of contemporaneous or stylistically similar artists can provide objective data to supplement traditionally subjective aesthetic interpretations that are based largely on simple, unsupported visual observation. This study performed digital color reproduction of two oil paintings through a comprehensive process including photographic capture and various image processing techniques. Color data were extracted from the CIE $L^*a^*b^*$ values embedded in each pixel of the acquired images and were refined and statistically analyzed with K-means clustering. In this manner, 10 representative colors were identified for each painting. This approach is significantly faster, more efficient, and more accurate than conventional methods that require dozens to hundreds of manual measurements with a spectrophotometer followed by statistical processing. In particular, digital color reproduction imaging enables non-contact color analysis with the use of light, making it especially suitable for delicate objects such as cultural heritage artifacts. Furthermore, this method effectively overcomes the physical limitations of spectrophotometric devices, therefore showing high potential as an alternative. Future research could further validate the reliability and precision of this indirect color measurement method to expand its applicability to a broader range of painted artworks.

6. Conclusion

This study performed machine learning–based color analysis using images of two oil paintings by Pai Un-soung, *still painting* and *Tug of War*, acquired through digital color reproduction. Digital color reproduction, including image capture, color correction, and color space profiling, was performed based on the CIE $L^*a^*b^*$ color space. When compared with reference values measured using a spectrophotometer, the reproduced images showed a high degree of color accuracy, with average color differences of ΔE 1.1 for *still painting* and 1.0 for *Tug of War*, in accordance with ISO 12647-8 standards. Using machine learning software, K-means clustering (10 clusters) was performed on the entire set of image data for each painting. In *still painting*, the dominant colors were a medium beige (CIE $L^*a^*b^*$: 63.69, −1.23, 17.94), accounting for 21% of the pixels, and a light beige (CIE $L^*a^*b^*$: 51.34, −2.93, 21.47), at 19.7%. Bright red and orange hues (used in flower petals) and cyan-green tones (used for the leaves) appeared in smaller proportions, serving as accent colors. By contrast, *Tug of War* featured light beige (18.5%), olive green (14.5%), and reddish brown (14.1%) as the primary colors, with red tones used selectively in clothing to emphasize certain elements. Both paintings predominantly utilized beige tones as background hues, and a small portion of high-saturation colors was employed for visual emphasis, revealing a shared chromatic strategy. The analysis of these representative colors was effective for visually and quantitatively identifying the artist’s color usage patterns. Further data collection and statistical analysis of additional artworks can objectively document the artist’s preferred colors, lightness distribution, and contrast tendencies. These findings could serve as a technical reference for verifying the authenticity of artworks, providing a foundation for developing cultural products based on the artist’s unique color identity. Furthermore, comparing color patterns among contemporary artists of similar styles could provide a more objective complement to traditional visual aesthetic interpretation. Ultimately, this study proposes a novel methodological approach integrating digital color reproduction and machine learning–based color analysis, providing valuable contributions to the analysis, conservation, and utilization of artworks as part of cultural heritage.

References

Cramer, F., Shephard, G.E., Heron, P.J., 2020. The misuse of colour in science communication. *nature communication*, 11, 1-10.

Hao, Z, Qi, X., 2022. End-to-end concrete appearance analysis b ased on pixel-semantic segmentation and CIE Lab. *Cement and Concrete Research*, 161, 1-14.

ISO, 2021. ISO 12647-8:2021. Graphic technology: Process con trol for the production of half-tone colour separations, proof and production prints: Part 8: Validation print processes workingdir ectly from digital data. International Organization for Standardiz ation

Sanmartín, P., Chorro, E., Vázquez-Nion, D., Martínez-Verdú, F.M., Prieto, B., 2014. Conversion of a digital camera into a non -contact colorimenter for use in stone cultural heritage: The appl ication case to Spanish granites. *Measurement*, 56, 194-202.

Song, H. R. and Jo, Y. H, 2021. Digital color reproduction and documentation of oil painting using image processing, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVI-M-1-2021, 693–695. https://doi.org/10.5194/isprs-archives-XLVI-M-1-2021-693-2021, 2021.

Papia, E.M., Kondi, A., 2025. Quantifying subtle color transitions in Mark Rothko's abstract paintings through K-means clustering and Delta E analysis. *Journal of Culture Heritage*, 72, 194-204.