https://doi.org/10.2352/CIC.2023.31.1.8


# Development and Testing of Vividness and Depth Model Based on Different Scale Combination

***Molin Li, Ming Ronnier Luo; State Key Laboratory of Extreme Photonics and Instrumentation, Zhejiang University; Hang Zhou, China;***


## Abstract

*Vividness and Depth are widely used in image and textile industry. And these scales were derived from one-dimensional scales of CIELAB L* and $C^*_{ab}$. However, these scales are limited to relative scales with a reference white, which makes it difficult for them to adapt to the variation in the world. This paper has introduced an experiment that focuses on assessing the wide range of luminance levels in a visual context, ranging from 10 cd/m$^2$ to 10000 cd/m$^2$. The experiment employed a method called magnitude estimation to gauge the perception of Vividness, Depth, Whiteness, and Blackness scales. The judgments were obtained through 10 observer x 40 NCS Sample x 4 lux level x 4 scales x 1.1 (10% repeat set), resulting in a total of 7040 assessments. This article mainly introduces the development of vividness and depth scales using combinations of relative scales, absolute scales, and the mixture scales like CAM16-UCS.*


## Introduction

Vividness and Depth are two important perceptual attributes that play a significant role in visual perception and image quality assessment. They capture different aspects of visual experience, with Vividness representing the richness and intensity of color appearance, and Depth, in the context of color perception, refers to the perceived darkness or lightness of a color, specifically its concentration or intensity on a white background. Understanding and quantifying these attributes are crucial for various applications, such as display technology, image processing, and virtual reality systems.

Berns [1] introduced the concept of Vividness, Depth, and Clarity as perceptual dimensions based on CIELAB color space. Vividness captured the transition of a color from shadow to highlight, representing the progression from low to ample illumination. Depth, on the other hand, involved incrementally adding a fixed color against a white background. Clarity quantified the distance from the background to a color. However, Berns did not conduct experiments to validate the specific processes underlying these scales.

Cho et al. [2] further conducted interval scale experiments on Vividness and Depth, separately for Korean and British participants, comparing their cognitive perceptions. They also developed their own mathematical models [3] for Vividness and Depth using Hue-based and Ellipsoid-based approaches. However, their experiments faced two limitations: the absence of an origin point in the interval scale, making it challenging to pass through the origin, and the testing environment being limited to an SDR condition with a maximum luminance of only 500 cd/m$^2$.

The main contributions of this study were twofold: 1) Designing and conducting a magnitude estimation experiment based on a ratio scale with an origin point, covering a luminance range of 10 cd/m$^2$ to 10000 cd/m$^2$, in accordance with HDR conditions. 2) Determining the optimal combinations of absolute and relative scales from CIECAM16 and CAM16-UCS [4] for fitting Vividness and Depth.

## Experiment

### *Experiment Setup*

An experiment was carried out to scale colours in terms of magnitude estimation. The samples were selected from the NCS ATLAS II [5], including 11 Hues, Blackness (s) from 6 to 70, Chromaticness (c) from 00 to 80 chromaticness. The size of the samples was 3 by 3 square inches, with a mid-grey background (L* of 65.5). The reflectance of each sample was measured by a Konika Minolta CM700d spectrophotometer. The experiment was conducted in a Thouslite LEDView cabinet. Four luminance levels from 10 to 10000 cd/m$^2$ were set in this experiment, with a fixed color temperature of 6000 K. Participants sit in front of a lightbox with an anchoring rest and rating the samples according to 4 scales (whiteness, blackness, vividness and saturation). Only the results from the latter two scales will be introduced in this article. The viewing geometry is 0°/0°.

### *Experiment Procedure*

Before the beginning of the experiment, participants were required to complete a color blindness test and an experimental training session.

Upon the onset of the experiment, participants were provided with an A4-sized instruction sheet delineated in Simplified Chinese. This instruction explicitly outlined the definition of the scales under study, their characteristic traits, and incorporated graphical references that emphasized the orientation of the two-dimensional scales on the Munsell color page, further elucidating their practical applications in day-to-day scenarios. While participants immersed themselves in the instruction reading, a concurrent training on one-dimensional scales was carried out. This included pivotal scales like 'brightness'—defined as the amount of light contained in a color, and 'colorfulness'—defined as the quantity of color within a specific hue. Once participants demonstrated an understanding of these one-dimensional scales, they were introduced to the complex two-dimensional scales used in the experiment. A notable example would be 'vividness', which signifies the distance of a color from neutral black. It was elucidated that an increment in the quantity of light augments both brightness and chroma, indicating that the direction of this color shift essentially mirrors the direction of vividness.

Post the elucidation, and once the participants indicated their foundational grasp on the scale definitions and the associated directions of color variations, the experimenter initiated the training validation phase. Herein, participants were presented with an array of differently hued NCS color blocks. Initially, they were tasked to arrange these blocks with the Munsell color page representation from the instruction sheet, emphasizing a horizontal chroma and vertical Value configuration. This step was pivotal to ascertain participants' comprehension of the one-dimensional scales. Upon their successful subjective arrangement, they were further instructed to organize them based on specific two-dimensional scale directions, coupled with a numerical scale judgment. This phase was integral to confirm their grasp on the two-dimensional scale definitions and familiarize them with the magnitude estimation process. Post-training and validation, the experimenter proceeded with the main experiment, wherein each training session was dedicated to a specific two-dimensional scale. The training session costs about 15 mins. The specific training process was illustrated in Figure 1.

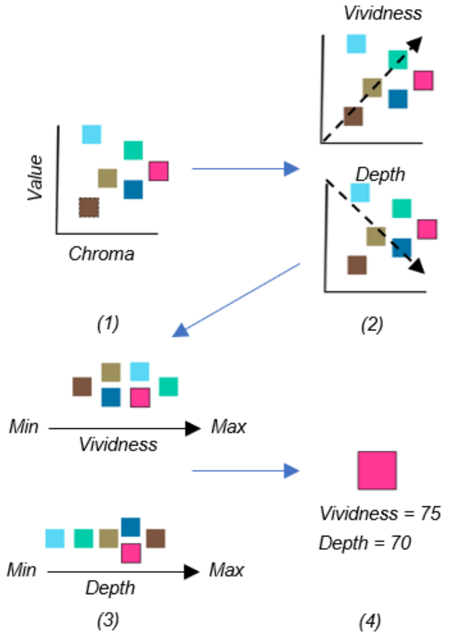



***Figure 1.*** *The schematic diagram illustrating the training process in the experiment was presented: (1) Arrange the color patched by the Munsell color page sequence;(2) Find the specified scales direction;(3) Sort the color patched by the specified scales; (4) grade each color patch by the origin point and the anchoring color.*

During the formal experiment, the experimenter randomly selected one luminance level from the four available options and determined the adapting time (ranging from 15 seconds to 5 minutes) based on the chosen luminance. Each luminance level required the completion of 44 judgments, consisting of 40 color blocks and 4 randomly repeated blocks, with the order of the color blocks being completely randomized. The observers were instructed to provide fixed answers after observing the color blocks for 15 seconds, without any opportunity for modification. Upon completing all estimations for a given luminance level, the participants adjusted the luminance level and repeated the same procedure until all luminance levels were completed. The formal experiment typically took around 70 minutes to finish. The specific samples used in the experiment, along with the training performance and formal experimental results, were illustrated in Figure 2. The configuration of the experimental platform was depicted in Figure 3.

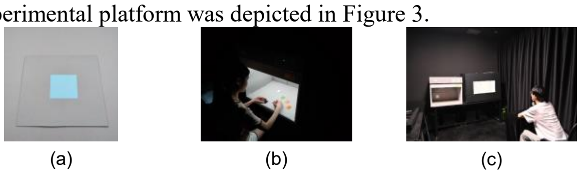



***Figure 2.*** *The demonstration of experimental setting:(a) NCS ATLAS Sample, (b) The real scene of training session (arrange the small color chip by the arrangement of Munsell colour page), (c) The formal experimental setup (Lights on for ease of photography).*

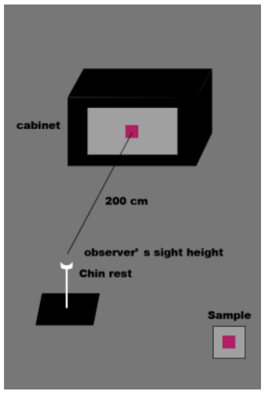



***Figure 3.*** *The schematic diagram depicting the experimental observation conditions.*

## *Observer*

A total of ten visually normal individuals, ranging in age from 18 to 25 years old (avg = 22 and std = 2), including five males and five females, participated in the experiment. Each participant engaged in the experiment only once for a specific scale. The experiment encompassed four luminance levels, four scales, ten participants, 40 samples, and 10% repeated sets, resulting in a total of 7040 estimations. The duration of each experiment was approximately 90 minutes, with the overall experiment requiring a total of 3600 minutes (4 scales x 10 participants x 90 minutes).

## Results

The experiment recorded the reflectance of 40 NCS ATLAS samples and the spectral power distribution (SPD) for four luminance levels. The XYZ values for each color block were calculated using the CIE1964 color matching functions. The participants' evaluations of the color blocks were documented under specific conditions. The Vividness data were calibrated based on an imagined black reference and an R20B3030 color block set at a fixed value of 50, while the Depth data were calibrated using a reference white from the actual environment and an R20B3030 color block set at a fixed value of 50.

## *Observer variation*

The consistency of observer differences in this experiment was assessed using the Standard Residual Sum of Squares (STRESS [6], as Equation (1)), which quantifies the consistency between data points. A smaller STRESS value indicates higher consistency among the data.

$$STRESS = 100 \times (\Sigma (F \times E_i - V_i)^2 / \Sigma V_i^2)^{0.5} \quad (1)$$

The calculation process for F is ($\Sigma E_i * V_i / \Sigma E_i^2$). $E_i$ represents the predicted data from the scale, while $V_i$ represents the raw data obtained from the experiment.

Conversely, the linearity of observer differences was evaluated using the Pearson correlation coefficient (R-value, as Equation 2).

$$R = \Sigma ((E_i - E_{avg}) \times (V_i - V_{avg})) / (\Sigma (E_i - E_{avg})^{0.5} \times (Vi - V_{avg})^{0.5}) (2)$$

The $R$ represents the R-value between the predicted data E and the observed data V. $V_i$ and $E_i$ represent specific observed data and predicted data, respectively. $V_{avg}$ and $E_{avg}$ denote the average values of the observed data and predicted data. The RE and V represent the linear relationship between the two, where a R value of 1 indicates a strong positive correlation, a value of 0 indicates no relationship, and a value of -1 indicates a strong negative correlation.

Typically, a lower STRESS value indicates a lower R-value between data points. When comparing each observer to the average data: For Vividness, the average STRESS was 19 and the R-value was 0.89. Observer 1 had the lowest R-value of 0.74 among the ten observers and a STRESS of 25. For Depth, the average STRESS was 20 with an R-value of 0.90. Observer 7 also had a low R-value at 0.86 with a STRESS of 22.

These data indicate a high level of consistency among the observers, primarily attributed to the anchoring points provided throughout the experimental process and the extensive training phase, which allowed the participants ample time to comprehend the definitions of the various scales.

## *Testing the performance of existing scales*

When testing the existing 2D scale data, we had three main questions to address:

1) The extent of correlation between 2D scale data and absolute scales (Brightness, Colourfulness) versus relative scales (Lightness, Chroma) in samples with a wide range of luminance variations.

2) The degree of correlation between 2D scale data and luminance versus chromaticity in samples with a wide range of luminance variations.

3) The effectiveness of the existing 2D scales in predicting the data obtained in this experiment.

To address question 1, we selected CIECAM16 Q, CAM16-UCS J', and Hellwig's optimized Brightness [7] (denoted as $Q_r$). For question 2, we chose CAM16-UCS M', Hellwig's improved Colorfulness (denoted as $M_r$), and CAM16-UCS J'. To tackle question 3, we utilized Berns' $V_{ab}$* and $D_{ab}$*. Subsequently, we calculated the R-values between these eight existing scales and the experimental data, yielding the results presented in Table. 1.

**Table. 1 The comparison of the R-value between the experimental data and the various scales.**

| Model | Vividness | Depth |
|---|---|---|
| CIECAM16 Q | 0.76 | -0.36 |
| CAM16-UCS J' | 0.65 | -0.91 |
| CAM16-UCS M' | 0.55 | 0.24 |
| CIECAM16 s | -0.16 | 0.50 |
| $V^*_{ab}$ | 0.73 | -0.69 |
| $D^*_{ab}$ | -0.26 | 0.86 |
| Hellwig $Q_r$ | 0.85 | -0.67 |
| Hellwig $M_r$ | 0.64 | 0.11 |

Regarding Vividness, the comparison of R-values (0.76 for Q and 0.65 for J') between CIECAM16 Q and CAM16-UCS J' suggests that it tends towards an absolute scale while still maintaining some degree of relative agreement. Additionally, Hellwig's optimization proved to be significantly effective for Vividness data. For Depth, it exhibited a strong linear relationship with Lightness (R-value of -0.91 for J'), while showing some linearity with $Q_r$ as well. Furthermore, both scales leaned more towards brightness than chroma. Finally, Berns' $V^*_{ab}$ and $D^*_{ab}$, which incorporate equal weighting of chroma and lightness based on relative scales, demonstrated good linear relationships with the experimental data (0.73 for $V^*_{ab}$ and 0.86 for $D^*_{ab}$), confirming their validity in this experiment. Additionally, the R-value between CIECAM16 s (saturation) and Depth indicated a similar trend between the two (both deviating from neutral colors with higher scale values), but without a strong linear relationship.

## Developing 2D Scales with Various Combinations

To further explore the three questions regarding 2D scales, we performed fitting analyses for the absolute scale combination, relative scale combination, and a mix-style combination similar to CAM16-UCS, using both the data from this experiment and Cho et al.'s data. Our main objectives were: (1) determining the most suitable combination for Vividness and Depth, and (2) investigating whether the HDR experimental data exhibits similar trends to Cho et al.'s data.

$$V_{relative} = k_o \times (J^2 + k_r \times C^2)^{0.5} \quad (3)$$

$$V_{absolute} = k_o \times (Q^2 + k_r \times M^2)^{0.5} \quad (4)$$

$$D_{relative} = k_o \times ((100 - J)^2 + k_r \times C^2)^{0.5} \quad (5)$$

$$D_{absolute} = k_o / Q_0 \times ((Q - Q_0)^2 + k_r \times M^2)^{0.5} \quad (6)$$

$$\text{Where } Q_0 = k_1 \times L_a^2 + k_2 \times L_a + k_3 \quad (7)$$

For the Vividness data, we employed Equation 3 and Equation 4 to fit the absolute scale combination and relative scale combination, respectively. These equations share similar structures, with the ratio coefficient, $k_o$, representing the influence of lightness/brightness, and $k_r$ representing the influence of chroma/colourfulness. When $k_r/k_o^2$ is greater than 1, it indicates a stronger impact of chroma/colourfulness on Vividness, while the reverse holds true for lightness/brightness.

Subsequently, for the Depth data, we utilized Equation 5 and Equation 6 for fitting. Equation 5 corresponds to the relative scale fitting, where the neutral white point is fixed at a value of 100. Thus, a structure similar to Equation 3 can be employed, with similar interpretations. Regarding the absolute scale, the adaptation of the white point is closely related to the environmental luminance. We introduced an adjustment factor, $Q_0$, to scale the entire luminance range. The parameter $Q_0$ is defined by Equation 7, representing its dependence on the background luminance adaptation La.

As for the mixing style, given the limited variation in the relative scales, we directly employed the formula for relative scale fitting.

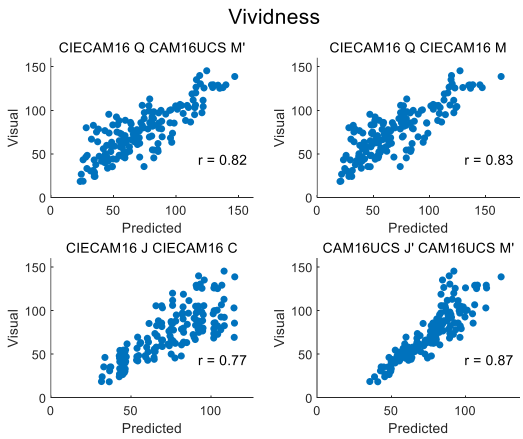
The fitting results are presented in Table 2 and Table 3, where the performance of the formulas is evaluated based on STRESS and R-value. Figure 4 displayed the scatter plots of different visual data against the predicted data from various models, with Cho's data transformed using the equation Visual = (Raw Data + 3) / 6 × 100.

**Table. 2 Illustration of the performance of 2D models with different scale combinations, where Q' represents CAM16-UCS Q', Q represents CIECAM16 Q, M' represents CAM16-UCS M', M represents CIECAM16 M, J represents CIECAM16 J, and C represents CIECAM16 C.**

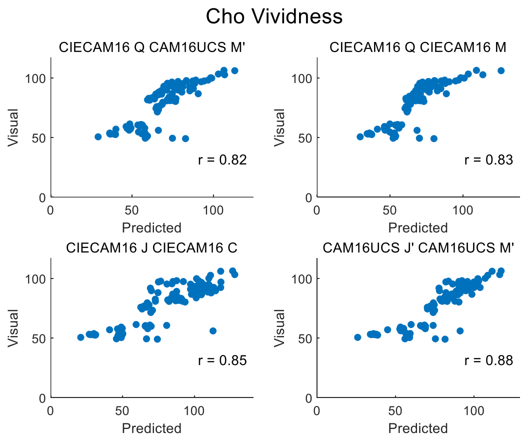
| Models | Scales | Eqs. | $k_o$ | $k_r$ | STRESS |
|---|---|---|---|---|---|
| Q, M' | V | 4 | 0.28 | 68.93 | 22 |
| Q, M | V | 4 | 0.29 | 26.84 | 23 |
| J, C | V | 3 | 1.13 | 1.33 | 23 |
| J', M' | V | 3 | 0.91 | 5.9 | 19 |
| Q, M' | D | 7 | 178.34 | 1.05 | 16 |
| Q, M | D | 7 | 177.87 | 0.6 | 15 |
| J, C | D | 5 | 0.99 | 0.34 | 15 |
| J', M' | D | 5 | 1.24 | 0.61 | 13 |

**Table. 3 Specific parameter settings table for equation 4**

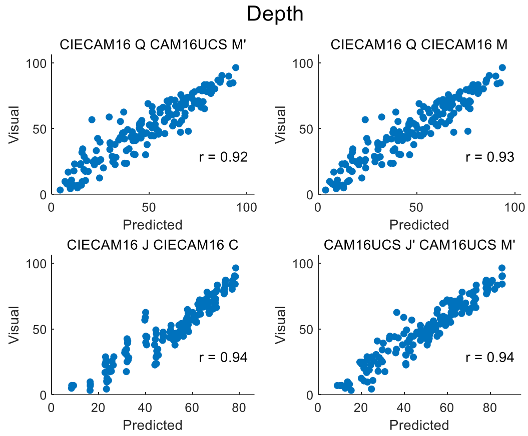
| Parameter | $k_1$ | $k_2$ | $k_3$ |
|---|---|---|---|
| Value | 4.12 | 9.63 | 93.32 |



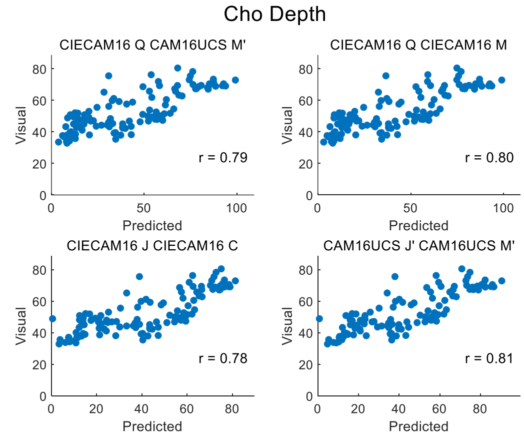

(a)




(b)




(c)




(d)

***Figure 4.*** *Scatterplot illustrations of different scale combinations against various visual data: (a) Vividness visual data in this study, (b) Vividness visual data by Cho et al., (c) Depth visual data in this study, (d) Depth visual data by Cho et al.*

From table 2, it can be observed that the trend of STRESS values is consistent with that of R-values, indicating that in this study, the F-test based on STRESS values can be used to compare the model performance among different scale combinations. The formula for the F-test, as indicated in Equation 8, was presented in the prompt.

$$F = STRESS_A^2 / STRESS_B^2 \quad (8)$$

The F-test is a statistical test used to compare the variances of two or more groups, aiming to determine if there are significant differences between them. The formula for the F-test is shown in Equation 8. The threshold for significant advantage is set at 1.3. Table 4 presents the average F-test results for different combinations, with the horizontal axis representing different scale combinations and the vertical axis representing different datasets. Each value represents the average F-test value of the corresponding scale combination compared to other combinations in the dataset. A higher value indicates a more significant advantage, and when the average value exceeds 1.3, it can be concluded that this combination is significantly superior to others.

**Table. 4 The average F-test table depicted the combinations of different scales on the horizontal axis, while the vertical axis represented the V and D datasets of the current experiment, along with Cho V and Cho D datasets from Cho**

**et al. When the data in the table exceeded 1.3, it indicated a significant superiority of that combination over others.**

| F-test | Q, M' | Q, M | J, C | J', M' |
|---|---|---|---|---|
| V | 1.05 | 1.02 | 0.68 | 1.44 |
| D | 0.77 | 0.90 | 1.16 | 1.29 |
| Cho V | 0.81 | 0.88 | 1.02 | 1.43 |
| Cho D | 0.95 | 1.05 | 0.94 | 1.08 |

From Table 4, two main findings can be derived: (1) For Vividness, the mixing style of CAM16-UCS demonstrated a significant advantage, as evident from Figure 4(a) and Figure 4(b), showing stronger linear relationships and smaller deviations. (2) For Depth, no specific combination of scales exhibited a significant advantage, although all CAM16-UCS combinations remained above 1. Notably, for the current experiment's data, this combination approached a significant advantage. In conclusion, the mixed combination of CAM16-UCS yielded predicted values that better aligned with the actual observations in both experiments.

## Discussion

The human eye's perception of brightness is related to the hue and colorfulness of a color, a relationship known as the HK effect. "Vividness" in our study is characterized as the gap between a color and neutral black, essentially representing the combined perceptions of brightness and colorfulness in the color. However, does this definition align with the phenomenon described by the HK effect? In CIECAM16, the factor of the HK effect was not taken into consideration for lightness and brightness. Hellwig [8] crafted a lightness and brightness equation encapsulating the HK effect by leveraging the foundational lightness of CIECAM16 and applying a cosine polynomial method with historical hue data. The scatter plot in Figure 5 presents the correlation between Hellwig $Q_{hk}$ and the experimental vividness data.

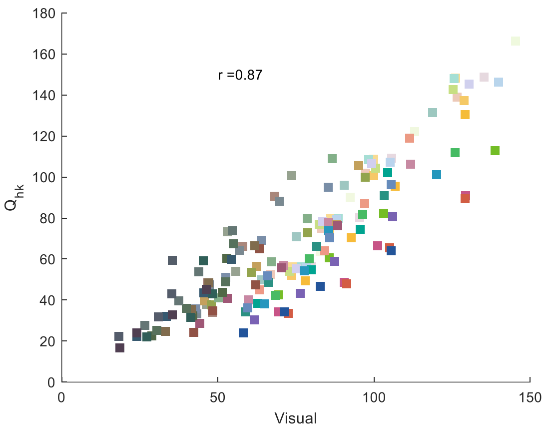



***Figure 5.** Scatterplot illustrations of Vividness against Hellwig $Q_{hk}$*

From this, it can be observed that the Vividness experimental data, when combined with brightness and the HK effect, has a higher r-value compared to CIECAM16 Q. Furthermore, the overall trend aligns with the direction of vividness.

However, for samples with high colorfulness, the predicted value of $Q_{hk}$ tends to be lower. In contrast, the combination of the CIECAM16 Q formula that defines distance from neutral black and the distance formula of CAM16-UCS M' doesn't present this issue. From the above, $Q_{hk}$ and vividness generally align in terms of the broader trend, but there are certain differences in their specific definitions.

## Conclusion

This study presented an experimental investigation conducted under HDR conditions, collecting Visual data of 2D scales such as Vividness and Depth across a luminance range of 10 cd/m² to 10,000 cd/m². A comparison with existing scales revealed that Vividness and Depth tended to be more inclined towards changes in brightness in high dynamic range. Furthermore, combinations of various scales from CIECAM16 and CAM16-UCS were formulated and fitted using the available experimental data. The obtained results indicated: 1) a similarity in trends between the HDR data obtained in this study and the LDR data obtained by Cho et al., and 2) through F-test analysis based on STRESS values, the combinations of CAM16-UCS J' and CAM16-UCS M' showed the best fit among the various scale combinations for the 2D experimental data. Future work may focus on comparing the existing Whiteness and Blackness data with NCS's W and S dimensions to obtain similar data, thus establishing a simple and efficient mixing-style-based 2D scales model applicable to both HDR and SDR scenarios.

## References


[1] Berns, R. S. (2014). Extending CIELAB: Vividness, depth, and clarity. Color Research & Application, 39(4), 322-330.

[2] Cho, Y. J., Ou, L. C., & Luo, R. (2017). A cross-cultural comparison of saturation, vividness, blackness, and whiteness scales. Color Research & Application, 42(2), 203-215.

[3] Cho, Y. J., Ou, L. C., Cui, G., & Luo, R. (2017). New colour appearance scales for describing saturation, vividness, blackness, and whiteness. Color Research & Application, 42(5), 552-563.

[4] Li, C., Li, Z., Wang, Z., Xu, Y., Luo, M. R., Cui, G., ... & Pointer, M. (2017). Comprehensive color solutions: CAM16, CAT16, and CAM16- UCS. Color Research & Application, 42(6), 703-718.

[5] Hård, A., & Sivik, L. (1981). NCS—Natural Color System: a Swedish standard for color notation. Color Research & Application, 6(3), 129-138.

[6] CIE 217:2016, "Recommended Method for Evaluating the Performance of Colour-Difference Formulae," (2016).

[7] Hellwig, L., & Fairchild, M. D. (2022). Brightness, lightness, colorfulness, and chroma in CIECAM02 and CAM16. Color Research & Application.

[8] Hellwig, L., Stolitzka, D., & Fairchild, M. D. (2022). Extending CIECAM02 and CAM16 for the Helmholtz–Kohlrausch effect. Color Research & Application, 47(5), 1096-1104.