



Custom perceptual color distance for visually impaired people

J. Liu, D. Muselet, E. Dinet, Y. Mizokami, P. Katemake, P. Colantoni, A. Trémeau

► To cite this version:

J. Liu, D. Muselet, E. Dinet, Y. Mizokami, P. Katemake, et al.. Custom perceptual color distance for visually impaired people. *Computer Vision and Image Understanding*, 2026, 273, pp.104942. <10.1016/j.cviu.2026.104942>. <hal-05747252>

HAL Id: hal-05747252

<https://hal.science/hal-05747252v1>

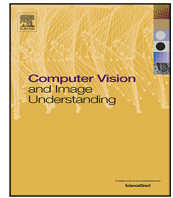
Submitted on 11 Sep 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



Custom perceptual color distance for visually impaired people[☆]

J. Liu^a, D. Muselet^{b,*}, E. Dinet^c, Y. Mizokami^a, P. Katemake^d, P. Colantoni^c, A. Trémeau^c

^a Graduate School of Science and Engineering, Chiba University, Japan

^b La Rochelle Université, Laboratoire L3i, La Rochelle, France

^c Université Jean Monnet Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France

^d Department of Imaging and Printing Technology, Faculty of Science, Chulalongkorn University, Thailand

ARTICLE INFO

Keywords:

Low-vision
Perceptual color distance
Siamese architecture
Metric learning

ABSTRACT

Evaluating color differences is a crucial step in many computer vision applications. Most existing methods rely on perceptual color difference equations designed for observers with normal vision. However, these metrics are not tailored to visually impaired observers, whose color perception can differ substantially from that of the standard observer. In this paper, we propose a learning-based approach to estimate perceptual color differences adapted to observers with a specific visual impairment. We collect psychophysical data to train and evaluate our model and formulate the problem as a metric learning task. Since collecting large-scale psychophysical datasets is challenging, we introduce a lightweight Siamese neural network that combines limited human judgments with prior knowledge from conventional perceptual color distances. Experiments show that our architecture can predict perceptual color differences for unseen color pairs and generalizes across illumination conditions and to an independent dataset collected with new observers and stimuli. The proposed learned metric provides a flexible alternative for adapting color-based computer vision applications to specific visual impairments.

1. Introduction

Perceptual color differences play a fundamental role in many computer vision and image processing applications, including image enhancement (Shi et al., 2024), image segmentation (Sabaneh and Sabha, 2024), image classification (Swamy et al., 2024), image compression (Prativadibhayankaram et al., 2024), object detection (X. Chen et al., 2024), image colorization (Hu et al., 2024), remote sensing (Wang et al., 2024), medical image analysis (W.-W. Chen et al., 2024), camera calibration (Suominen and Egiastian, 2024), tone mapping (Kim et al., 2024), image watermarking (Kumar et al., 2024), and data augmentation (Asdikian et al., 2024). In all these applications, color differences should ideally reflect human perception rather than simple numerical distances computed in a color space.

For this reason, numerous perceptual color difference equations have been proposed over the past decades, including ΔE_{ab} and ΔE_{00} , which remain widely used in both academia and industry (Mokrzycki and Tatol, 2011). These metrics were designed to approximate the perception of an average human observer represented by the CIE standard observer (Commission Internationale de l'Eclairage, 2019). Although natural inter-observer variability exists even among people

with normal vision (Chris Yi-Ho Bai, 2021), these perceptual models implicitly assume a healthy visual system.

This assumption becomes questionable when considering observers with visual impairments. According to the World Health Organization, more than two billion people worldwide live with some form of visual impairment. Diseases such as cataracts can modify visual perception by reducing contrast sensitivity, altering color discrimination, and introducing blur. Consequently, the color differences perceived by visually impaired observers may substantially differ from those predicted by conventional perceptual color metrics. Despite the growing interest in accessible computer vision systems, dedicated perceptual color distances adapted to these populations remain largely unexplored.

A possible solution would consist in conducting extensive psychophysical experiments for each visual impairment and deriving dedicated perceptual models from these observations. However, such experiments are costly, time-consuming, and difficult to reproduce for every pathology and severity level. Moreover, collecting sufficiently large and homogeneous groups of participants remains challenging.

In this work, we investigate whether a perceptual color distance adapted to visually impaired observers can be learned from a limited amount of psychophysical data. Instead of manually designing a

[☆] This article is part of a Special issue entitled: 'ACVR 2024' published in Computer Vision and Image Understanding.

* Corresponding author.

E-mail address: damien.muselet@univ-lr.fr (D. Muselet).

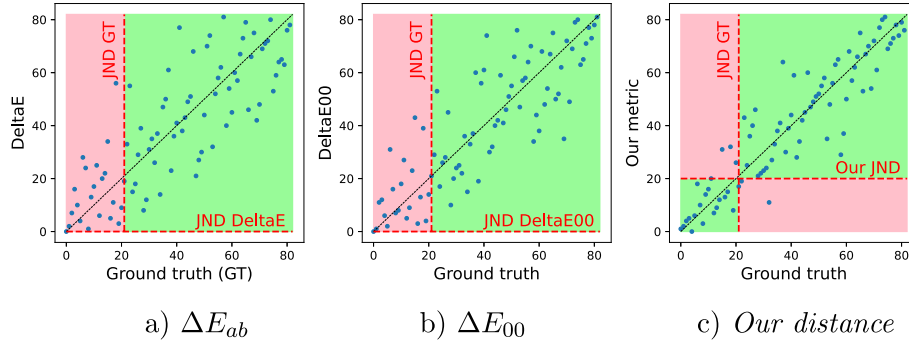


Fig. 1. The unsuitability of the default color distance equations (ΔE_{ab} in (a) and ΔE_{00} in (b)) for observers suffering from cataracts, and the correlation of our learned distance with the ground-truth perceptual judgments in (c).

new color difference equation, we formulate the problem as a metric learning task and propose a lightweight Siamese neural network that combines sparse perceptual judgments with prior information from existing perceptual color difference equations.

More specifically, this paper addresses the following research questions:

- Can a learned perceptual metric better predict the color differences perceived by visually impaired observers than conventional perceptual color difference equations?
- Can such a metric be learned from a limited amount of psychophysical data?
- Does incorporating prior knowledge from classical perceptual color distances improve the generalization ability of the learned metric?

The main contributions of this paper are fourfold:

- We experimentally analyze the limitations of conventional perceptual color difference equations when predicting color judgments from observers with cataract-simulated vision.
- We formulate customized perceptual color difference estimation as a metric learning problem that can be addressed using a limited amount of psychophysical observations.
- We propose a lightweight Siamese neural network that combines sparse perceptual data with prior knowledge extracted from classical perceptual color difference equations.
- We validate the proposed approach on both a public psychophysical dataset and a newly collected independent dataset. The experiments show that the proposed metric achieves a better agreement with human perceptual judgments than the considered conventional color difference metrics and baselines.

2. Motivations

Before introducing the proposed learning framework, it is first necessary to assess whether conventional perceptual color difference equations remain valid for observers with visual impairments. If these metrics already provide an accurate description of impaired color perception, learning a dedicated metric would bring limited benefit. Conversely, demonstrating a clear discrepancy between existing perceptual distances and human judgments would justify the development of a customized perceptual model.

To investigate this question, we use the psychophysical dataset introduced by Radsamrong et al. (2023). This dataset contains perceptual evaluations collected from twenty observers wearing cataract-simulating goggles corresponding to a visual acuity of approximately 0.06. Participants compared pairs of color patches, allowing the authors to establish a perceptual ranking ranging from very similar to very

Table 1

Rank correlations between the ground truth color difference perception and 3 color distances: ΔE_{ab} , ΔE_{00} and the distance provided by our approach.

Criteria	ΔE_{ab}	ΔE_{00}	Our distance
Kendall	0.5958	0.6422	0.7478
Spearman	0.7936	0.8303	0.9075

different. This ranking is considered as the perceptual ground truth throughout this paper. Based on the observers' responses, the color pairs can also be divided into two categories depending on whether their perceived difference is below or above the Just Noticeable Difference (JND) threshold. We refer to this classification as the JND Ground Truth (JND GT).

Fig. 1 compares this perceptual ground truth with the rankings obtained using two widely adopted perceptual color difference equations, namely ΔE_{ab} and ΔE_{00} . Since the CIELAB coordinates of all color patches are available, both distances can be computed for every color pair. The 82 color pairs considered in this analysis are sorted according to each distance, allowing a direct comparison with the perceptual ranking established from the psychophysical experiment.

In Fig. 1, the horizontal axis represents the perceptual ranking obtained from the observers, whereas the vertical axis corresponds to the ranking induced by each color difference equation. The first 21 color pairs are below the perceptual JND threshold. The horizontal red line indicates the conventional JND value ($\Delta E = 1$), while the black dashed diagonal represents perfect agreement between the predicted and perceptual rankings.

Two important observations can be drawn from these results. First, both ΔE_{ab} and ΔE_{00} exhibit a large dispersion around the diagonal, indicating that the rankings produced by these metrics do not fully match human perception under cataract-impaired vision. Second, the conventional JND threshold appears to underestimate the perceptual threshold observed during the psychophysical experiment. As a consequence, some color pairs perceived as identical by participants may be classified as distinguishable by classical color difference equations.

These observations suggest that existing perceptual color difference equations are not sufficient to accurately model the color perception of observers suffering from cataracts. Consequently, relying only on these metrics may introduce perceptual inconsistencies in vision systems intended for visually impaired users.

To further quantify this discrepancy, we compute the Kendall and Spearman rank correlations between each tested distance and the perceptual ground truth. The results are reported in Table 1. Both conventional color difference equations exhibit moderate agreement with human perception, with a Kendall correlation of 0.5958 for ΔE_{ab} and 0.6422 for ΔE_{00} .

These findings motivate the development of a dedicated perceptual color distance learned directly from psychophysical observations. The

right-hand plot of Fig. 1 illustrates the behavior of the proposed metric on the same benchmark. The learned distance produces a ranking closer to the perceptual ground truth while providing a more accurate estimation of the JND threshold. As confirmed by Table 1, the proposed approach achieves the highest agreement with human perception, demonstrating the potential of learning customized perceptual color distances from sparse psychophysical data.

3. Related work

3.1. Color perception under visual impairment

Understanding color perception under visual impairment has been widely studied through psychophysical experiments. Different pathologies, such as cataracts, glaucoma, and age-related macular degeneration, affect visual perception through distinct mechanisms, making it difficult to derive unified perceptual models.

Cataracts are among the most prevalent visual impairments worldwide. They are caused by a progressive opacification of the eye's lens, leading to reduced light transmission, increased scattering, and spectral attenuation, particularly affecting short- and medium-wavelength light. These effects result in blurred vision, reduced contrast sensitivity, and altered color perception, where colors appear desaturated and less discriminable (Kitakawa et al., 2009; Brøndsted et al., 2013; Boyd, 2014).

Several works have investigated ways to improve visual performance for low-vision users, notably through adaptive illumination systems. It has been shown that both illuminance and spectral composition significantly influence visual perception in individuals with low vision (Evans et al., 2010; Henry et al., 2020). Katemake et al. (2019) demonstrated that tuning the spectral power distribution can enhance color contrast perception for cataract patients, while Yang et al. (2023) proposed multichannel LED systems to dynamically adapt illumination conditions.

Despite these advances, most existing approaches focus on modifying the viewing conditions rather than modeling how visually impaired observers perceive color differences. In particular, perceptual color metrics specifically adapted to visually impaired populations remain largely unexplored.

3.2. Simulation of visual impairment

Due to the difficulty of conducting large-scale psychophysical experiments with visually impaired participants, several approaches have been proposed to simulate visual impairments for normally sighted observers.

Image-based simulation methods typically rely on image processing operators to mimic specific visual deficits. For instance, Gaussian blurring can approximate the reduced sharpness associated with cataracts (Lane et al., 2019), while spatial masking has been used to simulate scotomas in glaucoma and age-related macular degeneration. These methods are simple and scalable but cannot fully reproduce the subjective variability and complexity of real visual impairments.

Physical simulation devices, such as impairment goggles, provide a more realistic alternative and are widely used in psychophysical studies (Zagar and Baggarly, 2010). They have been employed in studies such as Radsamrong et al. (2023), Yang et al. (2023) to evaluate color discrimination under simulated low-vision conditions. However, such devices remain limited in their ability to reproduce all characteristics of real impairments, especially those involving dynamic effects such as moving scotomas (Copolillo et al., 2017).

More recently, extended reality (XR) systems combined with eye tracking have been proposed to improve the realism of visual impairment simulation (Wu et al., 2018; Macnamara et al., 2021). While promising, these systems are costly, complex to deploy, and still limited

in their ability to faithfully reproduce all spectral and photometric distortions induced by visual impairments (Acevedo et al., 2022).

Overall, simulation methods mainly facilitate the generation of experimental stimuli but do not address the fundamental problem of modeling perceptual color differences under visual impairment.

3.3. Perceptual color difference metrics

Perceptual color difference metrics aim at quantifying perceived differences between colors in a way that better matches human vision than purely numerical distances. The classical ΔE_{ab} metric was defined in the CIELAB color space to approximate perceptual uniformity. Its widely used extension, ΔE_{00} (Luo et al., 2001), improves perceptual accuracy by incorporating corrections for lightness, chroma, and hue interactions.

More recent color appearance models, such as CIECAM02-UCS and CAM16-UCS (Li et al., 2017; Fairchild, 2013), further refine perceptual uniformity by explicitly modeling additional aspects of human color vision and viewing conditions.

However, these models were developed using measurements from observers with normal vision and rely on the CIE standard observer. Consequently, they implicitly assume a healthy visual system and their ability to predict color differences perceived by visually impaired observers remains largely unknown.

This limitation motivates the development of perceptual color metrics that can adapt to specific visual conditions by directly learning from human perceptual judgments.

3.4. Metric learning and siamese networks

Metric learning aims at learning distance functions that reflect semantic or perceptual similarities between data samples (Kaya and Bilge, 2019). Classical formulations include Mahalanobis distance learning, where a positive semi-definite transformation matrix is optimized to enforce small distances between similar samples and large distances between dissimilar ones (Xing et al., 2002). Extensions such as Large Margin Nearest Neighbor (LMNN) and triplet-based losses further improve flexibility by introducing relative constraints between samples (Weinberger and Saul, 2009).

However, linear metric learning approaches remain limited in their ability to capture complex nonlinear relationships. To overcome this limitation, kernel-based and local metric learning approaches have been proposed (Perrot et al., 2014; Wang et al., 2012).

With the development of deep learning, Siamese networks have become a widely used framework for deep metric learning. These architectures consist of two identical encoders with shared weights that project input samples into a common embedding space, where a distance-based objective is optimized (Koch et al., 2015) (see Fig. 2). Siamese networks have been successfully applied to several similarity learning problems, including face recognition (Schroff et al., 2015), image matching (Melekhov et al., 2016), visual tracking (He et al., 2018), and 3D matching (Dai et al., 2017; Lim et al., 2016).

Recent studies have also highlighted the importance of appropriate evaluation protocols, sampling strategies, and baselines when assessing deep metric learning approaches (Musgrave et al., 2020).

In this work, we adapt the metric learning paradigm to the specific problem of perceptual color difference estimation under visual impairment. Unlike many deep metric learning applications that rely on large-scale annotated datasets, our objective is to learn meaningful perceptual representations from a limited amount of psychophysical data.

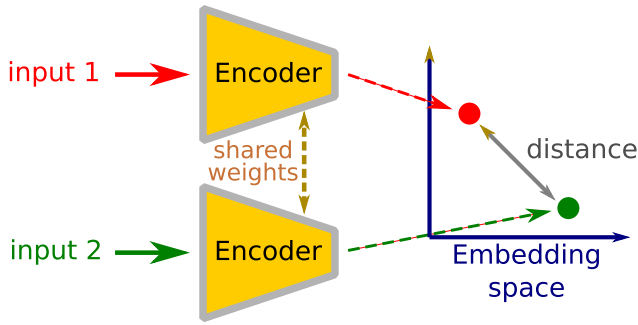


Fig. 2. A Siamese architecture with a single encoder projecting data in a non linear embedding space.

4. Experimental data

Visual impairments can significantly alter color perception and reduce the ability to discriminate between chromatic stimuli. In particular, cataracts induce blur, light scattering, and spectral attenuation, leading to reduced contrast sensitivity and altered color appearance (Kitakawa et al., 2009). These effects have been investigated through psychophysical experiments, including the dataset introduced by Radsamrong et al. (2023), which is used as the main training source in this work.

To evaluate the generalization ability of the proposed model, we additionally collected a new independent dataset under similar experimental conditions. This dataset is not used for training and is exclusively dedicated to testing. It allows us to evaluate whether the learned perceptual metric generalizes to new color stimuli and new observers.

The dataset of Radsamrong et al. (2023) contains 800 color pairs acquired under four different illumination conditions (200 pairs per illumination). In our experiments, this dataset is used either with a random split (75% for training and 25% for testing) or using a leave-one-illumination-out protocol, where three illuminations (600 pairs) are used for training and the remaining illumination (200 pairs) is used for testing.

The newly collected dataset contains 102 color pairs evaluated by six independent observers. It is exclusively used as an external test set, with the model trained on the complete dataset of Radsamrong et al. (2023). This protocol enables a cross-dataset evaluation involving unseen stimuli and different observers.

In the following, we describe the experimental protocol used to collect this dataset, designed to closely reproduce the setup of Radsamrong et al. (2023).

4.1. Experimental setup

The experiment was conducted in a controlled darkroom environment with a neutral gray background and controlled illumination conditions, as illustrated in Fig. 3.

Illumination was provided using 5 W LED light sources with a correlated color temperature of 6000 K. The spatial configuration of the lights was adjustable to ensure uniform coverage. The illuminance at the stimulus plane was set to 720 lux, exceeding the ISO 8995-1:2002 recommendation for indoor workplaces (de l'Eclairage, 2001). This higher level was chosen to compensate for the reduced visibility induced by the low-vision simulation goggles used during the experiment.

The viewing distance was fixed at approximately 70 cm, corresponding to a viewing angle of about 45 degrees.

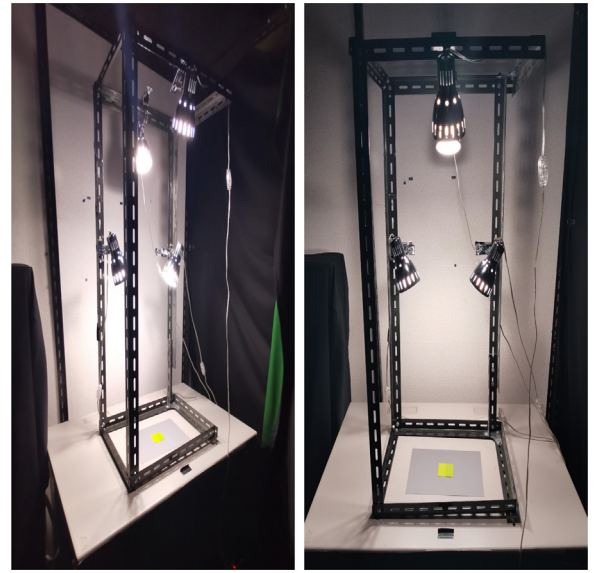


Fig. 3. Experimental setup used for data collection.

4.2. Color stimuli

The stimuli consist of color patches defined in the Practical Color Coordinate System (PCCS), developed by the Japan Color Research Institute to support color harmony studies (Japan Color Research Institute, 2007). This system is widely used in industrial design and applied color research (Wakata and Saito, 2013).

Unlike the Munsell-based stimuli used in Radsamrong et al. (2023), the present dataset is constructed from a different set of PCCS color patches. This difference ensures that the evaluation samples are not identical to the training samples and allows the assessment of generalization beyond the original experimental conditions (see Fig. 4).

From the PCCS palette, we selected multiple tone categories covering a wide range of perceptual attributes:

- dark greyish (dkg) and greyish (g): low lightness and low saturation,
- vivid (v) and strong (s): medium lightness and high saturation,
- pale (p) and light (lt): high lightness and low saturation.

Each patch is denoted using the format “tone + hue” (e.g., “dkg2” refers to hue 2 in the dark greyish tone group), as illustrated in Fig. 5.

The PCCS system defines 24 hues per tone category. In our experiment, we consider a hue step of 2, allowing comparisons such as dkg2–dkg4 or dkg4–dkg6. In addition, comparisons are performed within identical hues across different tones (e.g., dkg2 vs g2, or v4 vs s4).

Overall, 72 distinct color patches were selected, forming 102 color pairs. Each patch measures 1.5 cm × 4.5 cm.

The (L^* , a^* , b^*) coordinates of all stimuli were measured using a colorimeter under controlled illumination conditions. These measurements are used to compute classical color difference metrics such as ΔE_{ab} and ΔE_{00} . Fig. 6 compares the distribution of our dataset (blue) with that of Radsamrong et al. (2023) (red), highlighting differences in color coverage and supporting the complementarity of both datasets for evaluating generalization.

4.3. Participants

Six participants (four males and two females), aged 21 to 28, took part in the experiment. The objective of this dataset is not to train

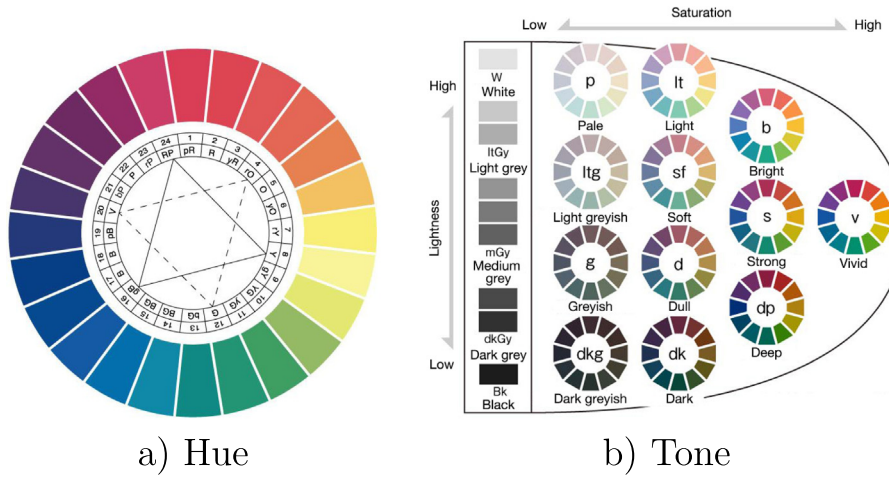


Fig. 4. The Practical Color Coordinate System (PCCS), consisting of (a) hue and (b) tone dimensions (Japan Color Research Institute, 2007). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

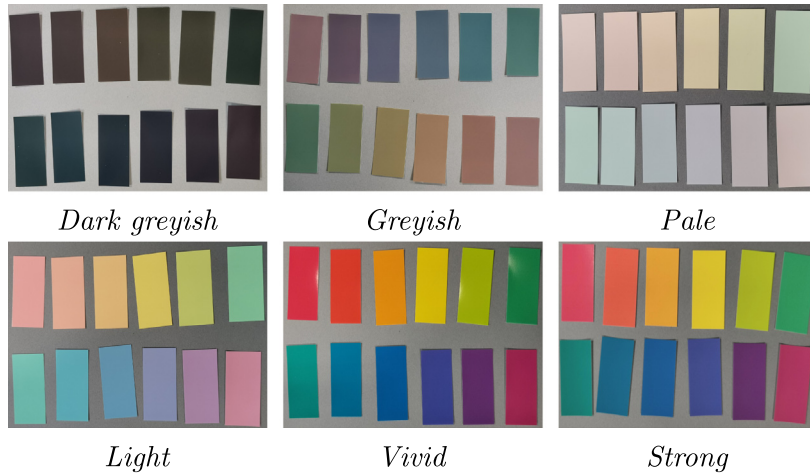


Fig. 5. Examples of color patches used in the experiment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

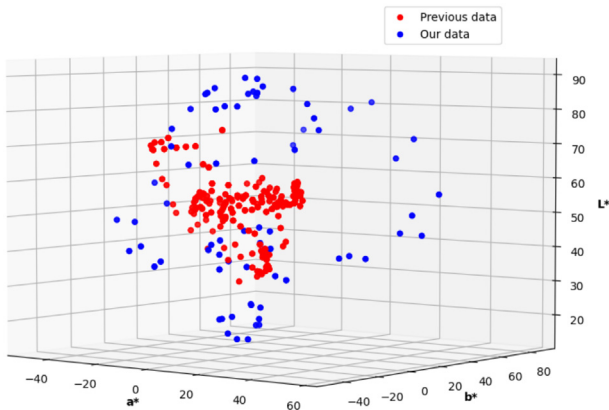


Fig. 6. Distribution of (L^*, a^*, b^*) values for the dataset of Radsamrong et al. (2023) (red) and the proposed dataset (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Criteria for color difference assessment.

Answer	Color difference assessment
"Cannot see difference"	0
"Maybe see slight difference"	1
"Able to see clear difference"	2

the model, but to provide an independent evaluation set for assessing cross-dataset generalization when training on Radsamrong et al. (2023).

All participants reported normal or corrected-to-normal vision and passed the Ishihara test (Ishihara, 1924), ensuring no color vision deficiencies. To reproduce the visual impairment studied in Radsamrong et al. (2023), participants wore low-vision goggles simulating moderate cataract-like blur with a visual acuity of approximately 0.08.

While wearing these goggles, participants evaluated 102 color pairs and selected one of three responses, as defined in Table 2.

5. Custom perceptual distance

Starting from sparse psychophysical annotations of color differences collected for observers with simulated visual impairment (Radsamrong

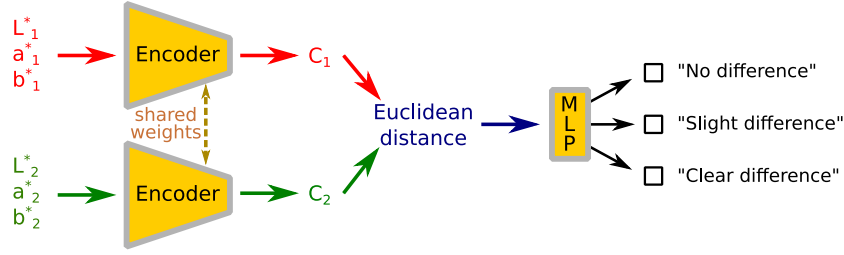


Fig. 7. A Siamese architecture with a pair of L^* , a^* , b^* coordinates as input and a 3D vector as output. This 3D vector is a probability distribution over the 3 possible decisions.

et al., 2023), we propose a learning framework to estimate a perceptual color distance from data. The objective is to learn a function that generalizes beyond the limited set of annotated color pairs and predicts perceptual judgments for unseen stimuli.

Once learned, this function can be used as a task-specific perceptual distance, in a similar spirit to classical metrics such as ΔE_{ab} or ΔE_{00} , enabling downstream computer vision applications to be adapted to non-standard visual perception.

5.1. Siamese embedding formulation

We consider pairs of colors represented in the CIELAB space. Each color is processed by a shared encoder network f_θ , producing embeddings in a latent space:

$$C_1 = f_\theta(x_1), \quad C_2 = f_\theta(x_2), \quad (1)$$

where $x_1, x_2 \in \mathbb{R}^3$ correspond to the (L^*, a^*, b^*) coordinates of the two colors.

The encoder is trained in a Siamese configuration with shared weights, as illustrated in Fig. 7. A Euclidean distance is then computed between the two learned embeddings:

$$d_\theta(x_1, x_2) = \|C_1 - C_2\|_2. \quad (2)$$

This distance is not directly interpreted as a perceptual score, but serves as a learned representation of color similarity that is further processed by a prediction network.

The shared encoder consists of five fully connected layers:

$$3 \rightarrow 32 \rightarrow 64 \rightarrow 64 \rightarrow 64 \rightarrow 64,$$

where ReLU activations are applied after each layer except the final one, which outputs the 64-dimensional embedding without activation.

5.2. Prediction of perceptual judgments

Human perceptual responses are modeled as a discrete probability distribution over the three perceptual classes defined in Table 2. Given a pair of colors (x_1, x_2) , the model predicts:

$$p(y \mid x_1, x_2), \quad (3)$$

where $y \in \{0, 1, 2\}$ corresponds to the three possible perceptual judgments.

The learned embedding distance $d_\theta(x_1, x_2)$ is provided as input to a multi-layer perceptron (MLP), which predicts the three output logits associated with the perceptual categories (see Fig. 7). The predicted probability distribution is obtained by applying a softmax operation to these logits.

This formulation directly matches the experimental protocol, where multiple observers provide categorical judgments that can be aggregated into empirical probability distributions.

5.3. Incorporating spatial information in color space

The relationship between learned embedding distance and perceptual similarity may vary across different regions of the color space. To account for this effect, we augment the MLP input with the original CIELAB coordinates of the two colors:

$$z = [d_\theta(x_1, x_2), x_1, x_2]. \quad (4)$$

This allows the model to adapt the mapping between the learned distance and perceptual judgment depending on the location in color space, instead of assuming a globally uniform relationship.

The corresponding architecture is illustrated in Fig. 8.

5.4. Incorporating classical color differences as prior

Although classical perceptual distances such as ΔE_{ab} are not optimal for observers with visual impairment (as shown in Fig. 1), they still encode useful information about color relationships established for normal vision.

We therefore introduce an additional variant where the input to the MLP is augmented with the classical color difference value (see Fig. 9):

$$z = [d_\theta(x_1, x_2), \Delta E_{ab}(x_1, x_2)]. \quad (5)$$

This design allows the network to exploit complementary information from:

- a learned embedding adapted to the target population,
- a classical perceptual prior derived from standard color difference equations.

5.5. Final architecture

The final proposed architecture (see Fig. 10) combines all previous components such that the input of the MLP is:

$$z = [d_\theta(x_1, x_2), x_1, x_2, \Delta E_{ab}(x_1, x_2)]. \quad (6)$$

This architecture jointly leverages:

- a learned Siamese embedding,
- spatial context in the CIELAB space,
- the classical ΔE_{ab} perceptual distance as complementary prior information.

For all four model variants evaluated in the ablation study, the Siamese encoder and the prediction network have exactly the same architecture. Only the input features provided to the prediction network are modified, allowing us to isolate the contribution of each additional component.

The prediction network is a fully connected MLP composed of:

$$d_{in} \rightarrow 32 \rightarrow 32 \rightarrow 16 \rightarrow 8 \rightarrow 3,$$

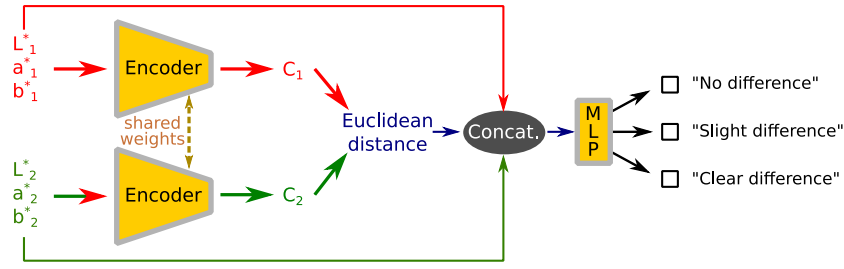


Fig. 8. The Siamese architecture with addition of the considered L^* , a^* , b^* coordinates as input to the MLP that predicts the final decisions.

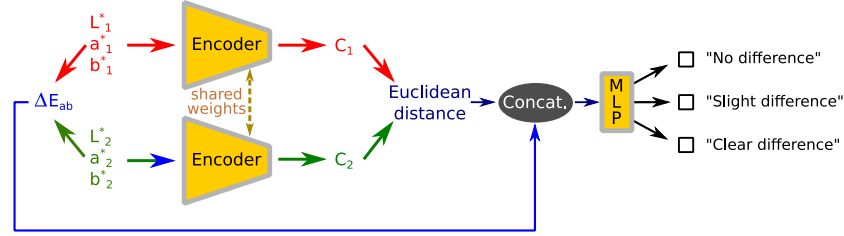


Fig. 9. The Siamese architecture with addition of the considered L^* , a^* , b^* coordinates and their associate color distance ΔE_{ab} as input to the MLP that predicts the final decisions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

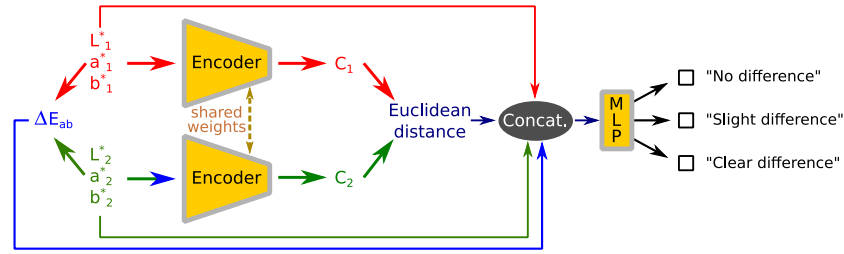


Fig. 10. Our proposed architecture. The Euclidean distance in the learned embedding space is concatenated with the two input CIELAB colors and their associated ΔE_{ab} value before being processed by the prediction network.

where d_{in} depends on the considered input configuration. ReLU activations are applied after each hidden layer. The final layer outputs three logits corresponding to the perceptual categories defined in Table 2. The softmax operation converts these logits into the predicted probability distribution.

Depending on the considered architecture variant, the input vector of the prediction network contains:

- only the learned embedding distance $d_{\theta}(x_1, x_2)$,
- the learned distance concatenated with the two CIELAB coordinates,
- the learned distance concatenated with ΔE_{ab} ,
- the learned distance concatenated with both CIELAB coordinates and ΔE_{ab} .

Using the same encoder and prediction network across all variants ensures that the ablation study measures only the contribution of the additional input features and not differences in model capacity.

This formulation is evaluated experimentally through the ablation studies presented in the next section, allowing us to quantify the contribution of each component of the proposed framework.

6. Experiments

This section presents an extensive evaluation of the proposed learned color distance. We aim to assess (i) its ability to generalize across standard splits, (ii) its robustness to illumination changes, and (iii) its transfer to an independent psychophysical dataset collected under different conditions.

6.1. Datasets

We consider two datasets in our experiments.

The first dataset, denoted D_1 , is the dataset introduced by Radsamrong et al. (2023). It contains psychophysical color difference annotations collected under controlled conditions using cataract-simulating goggles. In total, 800 color pairs have been collected for this dataset.

The second dataset, denoted D_2 , is our independently collected dataset described in Section 4. It was designed specifically as an out-of-distribution test set, with new observers, new color patches, and different stimulus distributions. This dataset contains 102 color pairs in total.

Importantly, D_1 is used for training and validation, while D_2 is used exclusively for final testing, ensuring a strict evaluation protocol for generalization.

6.2. Evaluation protocols

We evaluate the proposed method under three complementary protocols.

(1) *random split on D_1* . We first consider a standard random split of D_1 , where 75% of the samples are used for training and 25% for testing. This setting evaluates interpolation performance within the same dataset distribution.

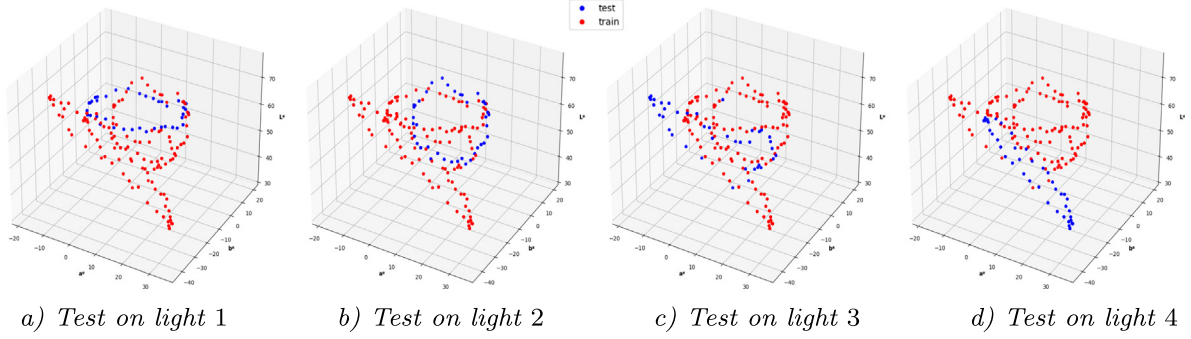


Fig. 11. Cross-illumination evaluation protocol. The model is trained on three lighting conditions (red) and tested on the remaining one (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2) *cross-illumination generalization on D_1* . To assess robustness to acquisition conditions, we exploit the fact that D_1 was collected under four different illumination settings. We construct four leave-one-light-out experiments:

- train on lights 2–4, test on light 1,
- train on lights 1,3,4, test on light 2,
- train on lights 1,2,4, test on light 3,
- train on lights 1–3, test on light 4.

This protocol evaluates the ability of the model to generalize across lighting variations (see Fig. 11).

(3) *cross-dataset generalization*. Finally, we evaluate transfer performance by training on the full D_1 dataset and testing on D_2 , which contains new observers and new color stimuli.

6.3. Compared methods

We compare the proposed model against several baselines.

Naive baseline. We compute the mean empirical distribution of labels on the training set and use it as a constant predictor. This provides a reference performance level that does not depend on the input color pair.

ED regression baseline. We also consider a parametric baseline that maps classical ΔE_{ab} distances to the output decision distribution using an exponential decay function. This baseline evaluates whether a simple monotonic transformation of a standard perceptual metric is sufficient to model low-vision perception.

Ablation variants of the proposed architecture. To analyze the contribution of each input component, we evaluate four variants of the same Siamese architecture. The encoder and prediction network architectures are kept identical in all cases, while only the additional features provided to the prediction network are modified:

- Siamese: only the learned embedding distance is provided to the prediction network,
- Siamese + L^*, a^*, b^* : the learned distance is concatenated with the original CIELAB coordinates,
- Siamese + ΔE_{ab} : the learned distance is concatenated with the classical color difference,
- Full model: the learned distance is concatenated with both CIELAB coordinates and ΔE_{ab} .

6.4. Training setup

All models are trained for 300 epochs with a batch size of 8. The optimization is performed using the Adam optimizer with an initial learning rate of 5×10^{-4} . The input CIELAB values are augmented

Table 3

Cross-entropy on random split (75% train/25% test) of D_1 .

Model	Cross-entropy
Naive prediction	0.517
ED regression	0.449
Siamese	0.376
Siamese + L^*, a^*, b^*	0.373
Siamese + ΔE_{ab}	0.371
Full model	0.368

during training by adding Gaussian noise with a standard deviation of 0.01. No weight decay is applied. Since the ground-truth annotations are represented as probability distributions over the three perceptual categories, we use a soft-label categorical cross-entropy loss between the predicted and reference distributions.

6.5. Evaluation metric

We report cross-entropy as the main quantitative metric, since the task consists in predicting a probability distribution over three perceptual decisions.

To improve interpretability, we also report:

- a naive baseline cross-entropy (constant predictor),
- qualitative visualizations of predicted vs. ground-truth distributions on representative samples.

6.6. Results on random split

Table 3 reports results on the random split of D_1 . All learned models outperform the naive baseline, confirming that the networks successfully capture meaningful structure in the data. Furthermore, incorporating both learned representations and classical color information consistently improves performance.

6.7. Results under illumination changes

Table 4 reports results for the leave-one-light-out experiments. The learned approaches consistently outperform the naive predictor, confirming that the models capture meaningful perceptual structure beyond a constant prior distribution.

The full model achieves the best performance for three out of the four illumination conditions and remains competitive in the remaining case. These results suggest that combining learned representations with classical perceptual information generally improves robustness to illumination changes.

Table 4

Cross-entropy under cross-illumination generalization (train on three lighting conditions and test on the remaining one).

Model	Light 1	Light 2	Light 3	Light 4
Naive prediction	0.864	0.810	0.780	0.751
ED regression	0.653	0.515	0.397	0.337
Siamese	0.607	0.462	0.500	0.399
Siamese + L^*, a^*, b^*	0.590	0.483	0.412	0.369
Siamese + ΔE_{ab}	0.610	0.468	0.425	0.334
Full model	0.586	0.473	0.358	0.300

Table 5

Cross-entropy on independent dataset D_2 (trained on D_1).

Model	Cross-entropy
Naive prediction	0.740
ED regression	0.321
Siamese	0.313
Siamese + L^*, a^*, b^*	0.301
Siamese + ΔE_{ab}	0.300
Full model	0.285

6.8. Results on independent dataset D_2

Table 5 shows results when training on D_1 and testing on the independent dataset D_2 . This evaluation is the most challenging experiment of the paper, as it measures out-of-distribution generalization across both observers and stimuli.

The proposed model significantly outperforms both the naive baseline and the classical ΔE_{ab} -based regression. This confirms that the learned representation transfers beyond the training distribution and can model perceptual judgments for previously unseen experimental conditions.

6.9. Qualitative analysis

To complement the quantitative evaluations, Fig. 12 illustrates qualitative predictions for 50 randomly selected color pairs from the test split of dataset D_1 . This visualization is intended to provide an intuitive understanding of the behavior of the proposed model within the experimental conditions used for training and evaluation. The independent generalization capability of the model is assessed separately through the quantitative evaluation on dataset D_2 presented in Section 6.

Each example shows the ground-truth distribution (blue) and the predicted distribution (orange).

These visualizations confirm that the proposed method captures the overall structure of perceptual judgments, including both agreement and uncertainty patterns across observers.

Although most predictions closely follow the human judgments, some discrepancies remain. These differences may originate from the limited number of psychophysical annotations available for training and from the intrinsic variability of human perception, where identical color pairs may receive different judgments across observers.

7. Conclusion

In this paper, we introduced a learning-based approach to estimate perceptual color differences tailored to observers with a specific visual impairment. The proposed method formulates perceptual color difference estimation as a metric learning problem and relies on a lightweight Siamese neural architecture operating on CIELAB inputs. By combining sparse psychophysical annotations with prior knowledge from classical color difference metrics, the proposed framework can learn a customized perceptual distance from a limited amount of data.

A key objective of this work was to investigate whether a learned perceptual metric could better capture low-vision color perception than

conventional color difference equations. Experimental results demonstrate that the proposed approach improves agreement with human judgments across several evaluation protocols, including unseen color pairs within the same dataset, cross-illumination conditions, and a fully independent dataset collected with new observers and stimuli. These results suggest that the learned representation captures relevant perceptual structures beyond the original training conditions.

Unlike classical perceptual color difference formulas, which are designed for the standard CIE observer, the proposed model directly learns from human judgments under low-vision conditions. This makes it a flexible alternative for adapting perceptual distances to non-standard visual conditions, where handcrafted metrics may become inadequate.

Beyond the methodological contribution, the proposed approach may benefit several application domains where perceptual color similarity plays an important role, such as lighting design for improved visual discrimination in indoor environments or assistive visualization systems for low-vision users. In particular, integrating learned perceptual distances into rendering or extended reality (XR) pipelines could contribute to improving the accessibility of visual content.

The present work nevertheless has several limitations. First, the model is currently trained for a single type of visual impairment, namely simulated moderate cataracts, and its ability to generalize to other visual disorders remains to be investigated. Second, the current datasets rely on simulated visual impairments using optical devices rather than measurements from large cohorts of patients with diagnosed pathologies. Third, although the proposed framework is specifically designed to learn from sparse psychophysical data, its performance remains dependent on the availability and quality of perceptual annotations. Finally, the evaluation has been conducted on controlled color-patch experiments, which provide a rigorous benchmark but do not fully capture the complexity of natural scenes and real-world visual tasks.

These limitations also define several promising directions for future work. Extending the framework to other types of visual impairments, collecting larger and more diverse psychophysical datasets, validating the approach with observers affected by real visual disorders, evaluating the learned metric in real-world imaging applications, and investigating more explicit uncertainty modeling in perceptual judgments are natural next steps toward more general perceptually adapted color metrics.

CRedit authorship contribution statement

J. Liu: Writing – original draft, Software, Resources, Investigation, Formal analysis, Data curation. **D. Muselet:** Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **E. Dinet:** Writing – review & editing, Visualization, Resources, Formal analysis, Conceptualization. **Y. Mizokami:** Writing – review & editing, Project administration, Methodology, Investigation, Data curation, Conceptualization. **P. Katemake:** Writing – review & editing, Project administration, Methodology, Data curation, Conceptualization. **P. Colantoni:** Writing – review & editing, Project administration, Methodology, Conceptualization. **A. Trémeau:** Writing – review & editing, Supervision, Project administration, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

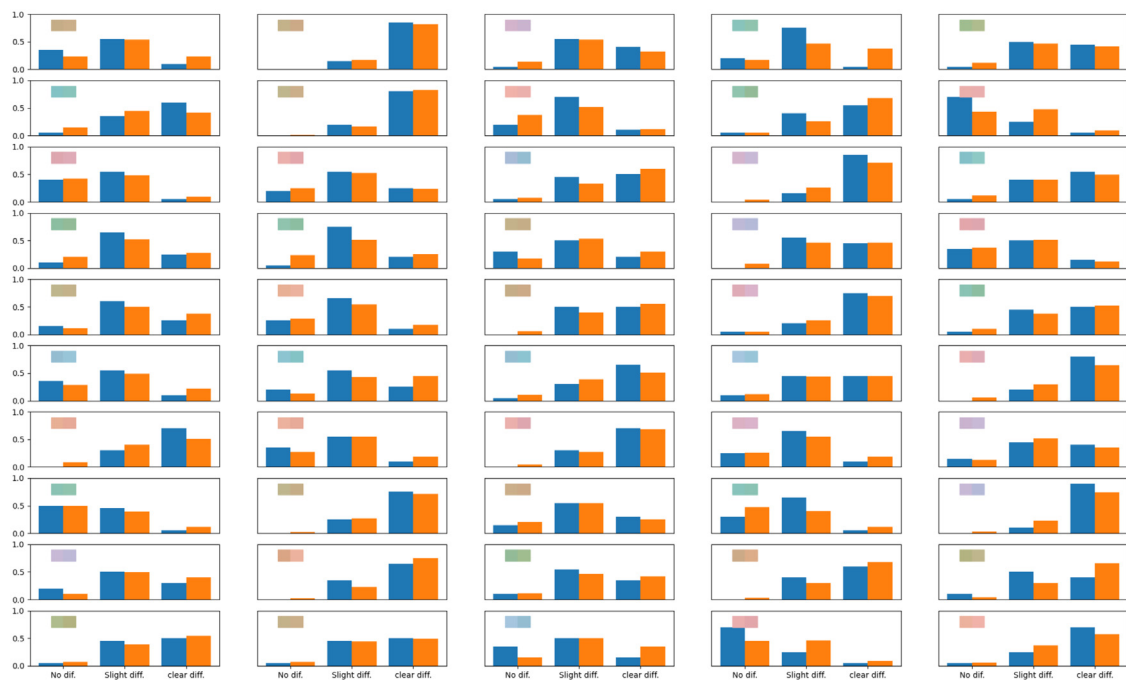


Fig. 12. 50 examples of prediction from our learned distance, along with the corresponding ground-truth. Ground-truth in blue and prediction in orange. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

- Acevedo, V., Colantoni, P., Dinet, E., Trémeau, A., 2022. Real-time low vision simulation in mixed reality. In: 16th International IEEE Conference on Signal-Image Technologies and Internet-Based System.
- Asdikian, J.P.H., Li, M., Maier, G., 2024. Performance evaluation of YOLOv8 and YOLOv9 on custom dataset with color space augmentation for real-time wildlife detection at the edge. In: 2024 IEEE 10th International Conference on Network Softwarization (NetSoft). IEEE, pp. 55–60.
- Boyd, K., 2014. What are cataracts. *Am. Acad. Ophthalmol.*
- Brøndsted, A.E., Lundeman, J.H., Kessel, L., 2013. Short wavelength light filtering by the natural human lens and IOLs—implications for entrainment of circadian rhythm. *Acta Ophthalmol.* 91 (1), 52–57.
- Chen, X., Fang, Y., Feng, M., 2024. Research on fruit detection algorithms for picking robots in complex environments. In: 2024 IEEE 2nd International Conference on Control, Electronics and Computer Technology. ICCECT, pp. 887–895.
- Chen, W.-W., Tseng, C.-C., Huang, C.-C., Lu, H.H.-S., 2024. Improving deep-learning electrocardiogram classification with an effective coloring method. *Artif. Intell. Med.* 149, 102809.
- Chris Yi-Ho Bai, L.-C.O., 2021. Observer variability study and method to implement observer categories for novel light source projection system. *Color Res. Appl.* 46 (5), 1019–1033.
- Commission Internationale de l'Eclairage, 2019. Colorimetry - part 1: CIE standard colorimetric observers. In: *IEC ISO 11664-1:2019, Colorimetry*.
- Copolillo, A., Christopher, A., Lyons, A., 2017. Effects of simulated low vision on postural adjustment to changes in center of mass in older adults. *Occup. Ther. Health Care* 31 (2), 115–125.
- Dai, A., Ritchie, D., Bokeloh, M., Reed, S., Sturm, J., Nießner, M., 2017. Shape completion using 3D-encoder-predictor CNNs and shape synthesis. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 5868–5877.
- Evans, B., Sawyerr, H., Jessa, Z., Brodrick, S., Slater, A., 2010. A pilot study of lighting and low vision in older people. *Light. Res. Technol.* 42 (1), 103–119.
- Fairchild, M.D., 2013. *Color Appearance Models*. John Wiley & Sons.
- He, Z., Balaban, S., et al., 2018. Twofold siamese network for real-time object tracking. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*.
- Henry, R., Duquette, J., Wittich, W., 2020. Comparison of two lighting assessment methods when reading with low vision. *Optom. Vis. Sci.* 97 (4), 257–264.
- Hu, Z., Shkurat, O., Kasner, M., 2024. Grayscale image colorization method based on U-net network. *Int. J. Image Graph. Signal Process.* 16, 70–82.
- Ishihara, S., 1924. *Tests for colour-blindness*. Tokyo: Handaya.
- Japan Color Research Institute, 2007. *Practical color coordinate system*. URL <https://www.siliken.co.jp/pccs/pccs02.html>.
- Katame, P., Radsamrong, A., Dinet, E., Heng, C.W., Kuang, Y., Kalavally, V., Trémeau, A., 2019. Influence of LED-based assistive lighting solutions on the autonomous mobility of low vision people. *Build. Environ.* 157, 172–184.

- Kaya, M., Bilge, H.S., 2019. Deep metric learning: A survey. *Symmetry* 11 (9), 1066.
- Kim, Y.-J., Son, D.-M., Lee, S.-H., 2024. Retinex jointed multiscale CLAHE model for HDR image tone compression. *Mathematics* 12 (10), 1541.
- Kitakawa, T., Nakadomari, S., Kuriki, I., Kitahara, K., 2009. Evaluation of early state of cyanopsia with subjective color settings immediately after cataract removal surgery. *J. Opt. Soc. Amer. A* 26 (6), 1375–1381.
- Koch, G., Zemel, R., Salakhutdinov, R., 2015. Siamese neural networks for one-shot image recognition. In: *Proceedings of the ICML Deep Learning Workshop*.
- Kumar, S., Verma, S., Singh, B.K., Kumar, V., Chandra, S., Barde, C., 2024. Entropy based adaptive color image watermarking technique in YC b C r color space. *Multimedia Tools Appl.* 83 (5), 13725–13751.
- Lane, J., Robbins, R.A., Rohan, E.M., Crookes, K., Essex, R.W., Maddess, T., Sabeti, F., Mazlin, J.-L., Irons, J., Gradden, T., et al., 2019. Caricaturing can improve facial expression recognition in low-resolution images and age-related macular degeneration. *J. Vis.* 19 (6), 18–18.
- de l'Eclairage, C.I., 2001. Guide on interior lighting. In: *ISO 8995-1:2002(E)/CIE S 008/E:2001*.
- Li, C., Li, Z., Wang, Z., 2017. Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS. *Color Res. Appl.* 42 (6), 703–718.
- Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M., 2016. Siamese networks for image super-resolution. In: *European Conference on Computer Vision Workshops*.
- Luo, M.R., Cui, G., Rigg, B., 2001. The development of the CIE 2000 colour-difference formula. *Color Res. Appl.* 26 (5), 340–350.
- Macnamara, A., Chen, C., Schinazi, V.R., Saredakis, D., Loetscher, T., 2021. Simulating macular degeneration to investigate activities of daily living: A systematic review. *Front. Neurosci.* 15, 663062.
- Melekhov, I., Kannala, J., Rahtu, E., 2016. Siamese network for image matching. In: *Proceedings of the European Conference on Computer Vision Workshops*.
- Mokrzycki, W., Tatol, M., 2011. Color difference delta e - a survey. *Mach. Graph. Vis.* 20, 383–411.
- Musgrave, K., Belongie, S., Lim, S.-N., 2020. Metric learning reality check. *arXiv preprint arXiv:2003.08505*.
- Perrot, M., Habrard, A., Lechelle, W., Ralaivola, L., 2014. Metric learning in the Gaussian mixture model framework. In: *International Conference on Machine Learning*.
- Prativadibhayankaram, S., Panda, M.P., Seiler, J., Richter, T., Sparenberg, H., Fossel, S., Kaup, A., 2024. A study on the effect of color spaces in learned image compression. *arXiv preprint arXiv:2406.13709*.
- Radsamrong, A., Katame, P., Dinet, E., Kalavally, V., Trémeau, A., 2023. Determination of narrow-band LEDs for luminance contrast enhancement. In: *Imaging Conference Japan Proceedings*. Tokyo: Imaging Society of Japan, pp. 11–14.
- Sabaneh, K., Sabha, M., 2024. Improving SLIC superpixel by color difference-based region merging. *Multimedia Tools Appl.* 83.
- Schroff, F., Kalenichenko, D., Philbin, J., 2015. FaceNet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 815–823.

- Shi, J., Yu, S., Li, H., Zhang, X., Liu, C., 2024. Underwater image enhancement based on adaptive color correction and multi-scale fusion. *Multimedia Tools Appl.* 83.
- Suominen, J., Egiazarian, K., 2024. Camera color correction using splines. *Electron. Imaging* 36, 1–6.
- Swamy, K.V., Radhika, V., Kadiyala, R.D., 2024. Skin disease classification using image preprocessing and machine learning. In: *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation. IATMSI, 2*, IEEE, pp. 1–4.
- Wakata, T., Saito, M., 2013. The impression of hue in practical color co-ordinate system (PCCS).
- Wang, Q., Chi, K., Jing, W., Yuan, Y., 2024. Recreating brightness from remote sensing shadow appearance. *IEEE Trans. Geosci. Remote Sens.* 62.
- Wang, J., Kalousis, A., Woznica, A., 2012. Parametric local metric learning for nearest neighbor classification. *Adv. Neural Inf. Process. Syst.* 25.
- Weinberger, K.Q., Saul, L.K., 2009. Distance metric learning for large margin nearest neighbor classification. In: *Advances in Neural Information Processing Systems*.
- Wu, H., Ashmead, D.H., Adams, H., Bodenheimer, B., 2018. Using virtual reality to assess the street crossing behavior of pedestrians with simulated macular degeneration at a roundabout. *Front. ICT* 5, 27.
- Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S., 2002. Distance metric learning with application to clustering with side-information. In: *Advances in Neural Information Processing Systems*.
- Yang, L., Dinet, É., Katemake, P., Trémeau, A., Colantoni, P., 2023. A multichannel LED-based lighting approach to improve color discrimination for low vision people. In: *Human Vision and Electronic Imaging*.
- Zagar, M., Baggarly, S., 2010. Low vision simulator goggles in pharmacy education. *Am. J. Pharm. Educ.* 74 (5).