Contrastive Learning for Large-scale Color-Name Dataset: Tackling Sparsity with Negative Sampling

Kecheng Lu
Renmin University of China
Beijing, China
lukecheng0407@gmail.com

Yue He
School of Information
Renmin University of China
Beijing, Beijing, China
heyue25@ruc.edu.cn

Yunhai Wang*
Renmin University of China
Beijing, China
cloudseawang@gmail.com

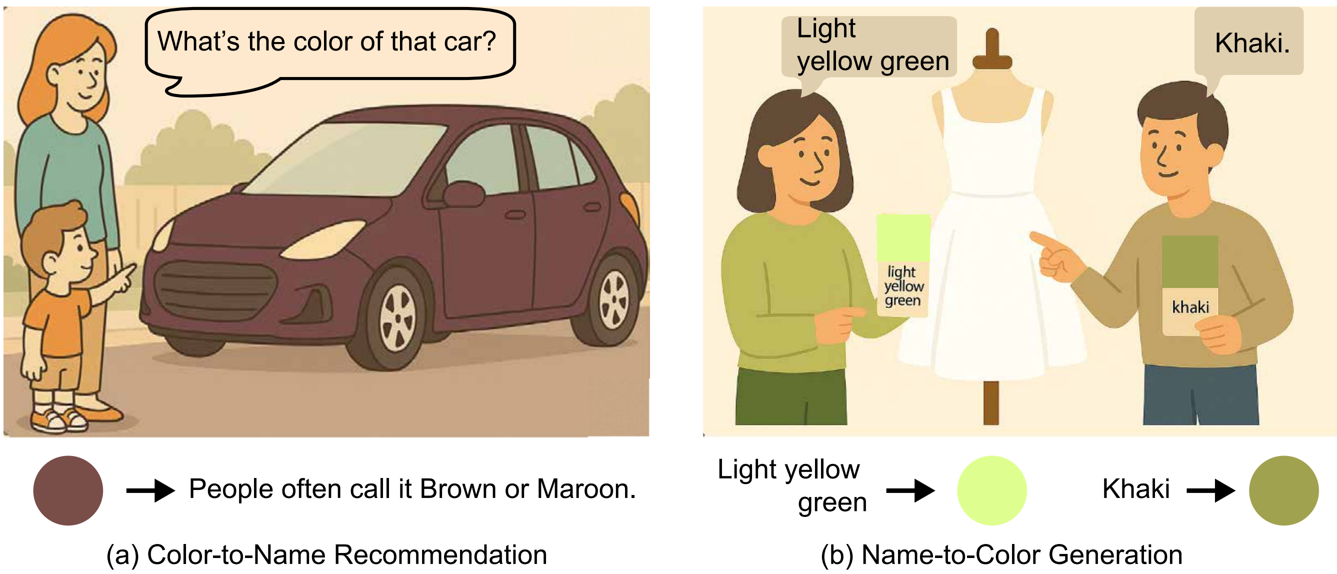

Figure 1: Qualitative summary of color–name bidirectional modeling. (a) Color-to-Name (C2N) recommendation: given an uncommon car color, our model retrieves plausible human-used names (e.g., *brown*, *maroon*) from a unified color–name embedding. (b) Name-to-Color (N2C) recommendation: given textual intents (e.g., *light yellow green* vs. *khaki*) for a white dress, the model returns a RGB swatches that match each description. Trained on large-scale crowdsourced data with severe sparsity and imbalance, our contrastive framework (negative sampling + multi-task losses) supports both directions in a single representation, yielding substantially higher C2N accuracy and lower N2C perceptual error than prior methods.

Abstract

Large-scale color datasets exhibit significant sparsity in name-color correspondences, substantially impeding the effectiveness of conventional methodologies. We propose a contrastive learning-based framework for color name generation and recommendation that addresses sparsity through negative sampling, supporting two core tasks: color-to-name recommendation and name-to-color generation. Our framework employs a multi-task contrastive learning architecture comprising three key components: (1) a pre-trained Transformer-based name encoder, (2) an RGB encoder, and (3) an RGB generator. The framework utilizes negative sampling to construct positive-negative pairs, contrasting RGB encoder outputs with positive and negative name embeddings. We adopt a multi-objective optimization strategy incorporating binary cross-entropy loss for neural collaborative filtering, and mean squared error loss for name-to-RGB mapping. Experimental results demonstrate substantial improvements over baseline methods, achieving 71.26% Top-10 accuracy in color-to-name recommendation and reducing CIELAB distance error to 26.61 in name-to-color generation.

*corresponding author

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791278

CCS Concepts

• **Human-centered computing** → **Information visualization**.

Keywords

Contrastive learning, Color name generation, Multi-task learning, Sparse data, Transformer

ACM Reference Format:
Kecheng Lu, Yue He, and Yunhai Wang. 2026. Contrastive Learning for Large-scale Color-Name Dataset: Tackling Sparsity with Negative Sampling. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3772318.3791278

1 Introduction

Color naming, the process of mapping the perceptual color stimuli to the linguistic expressions, provides a fundamental perspective into how humans perceive, categorize, and communicate the visual world. The discovery of name–color relationships not only advances our understanding of perception and language but also supports practical applications in visualization, design, and human–computer interaction [7, 8, 18, 19]. Color naming functions as a natural interface for human–machine communication: users often specify or interpret color through language (e.g., "a slightly warmer blue"), and systems that can reason about such mappings enable more intuitive and semantically consistent interaction [9, 29]. Accurate color–name modeling thus supports a range of HCI applications, including interactive color recommendation and visualization design [20, 28], accessibility enhancement for color vision deficiencies [6], and the creation of adaptive and inclusive design tools [37]. As a result, computational color naming serves not only as a linguistic modeling task but also as an enabling technology for human-centered and perceptually grounded interaction design.

From Berlin and Kay's seminal work on basic color terms [3] to large-scale cross-cultural surveys such as the World Color Survey [5], researchers have demonstrated both the universality and diversity of color naming. The release of crowdsourced datasets like XKCD [24] further reflects the growing attention to name–color relationships, showing that the lexicon of color extends far beyond a fixed set of basic terms and involves compositional modifiers, semantic nuance, and contextual variation. Driven by the heterogeneity of human color perception, the mapping between names and colors is inherently many-to-many: a single color may elicit several names, and one name may refer to a wide range of colors. Modeling this bidirectional relationship can thus be formalized as estimating the conditional probability distribution between names and colors. Earlier studies relied on small-scale experiments or handcrafted statistical models, such as fuzzy membership functions [1], perceptual boundary triangulation [21], and entropy-based probabilistic models [4]. Heer and Stone [8] extended this line by introducing a non-parametric model trained on the large-scale XKCD dataset, though their vocabulary reduction to 153 terms limited fine-grained color diversity. More recent neural approaches, e.g., Text2Color [13], have supported compositional descriptions but remain one-directional and data-constrained. Consequently, these methods capture only partial structure of the color–name space, motivating large-scale, contrastive formulations that learn from sparse and imbalanced correspondences.

The necessity for large-scale relationship mining has become increasingly clear, as datasets such as the World Color Survey and XKCD now provide millions of name–color pairs that enable the study of long-tail descriptors, cultural variation, and modifier composition. Recent works leveraging such data [8, 10] have shown its potential for visualization and generative design, yet two major challenges remain. First, the distribution of color names is highly imbalanced: while common color terms are associated with a broad spectrum of perceptual colors, a long tail of names appear infrequently. This results in sparse name-color connections for these names in data, which introduces high variance and ambiguity into capturing the name-color relationships. The second challenge is that the RGB representations of different colors are very close. The RGB color space is insufficient to fully reflect the differences in human perception of various colors, making it difficult to model subtle color differences when conducting fine-grained relationship discovery for large-scale data.

To address and embrace these challenges, we propose a contrastive learning framework for joint color–name embedding. As shown in Fig. 1, functionally, our framework is designed to support two dual tasks: **color-to-name recommendation** and **name-to-color generation**, within a unified representation. From a performance perspective, it introduces a multi-task architecture that integrates (1) a pre-trained Transformer-based name encoder, (2) an RGB encoder, and (3) an RGB generator. The key innovation lies in leveraging negative sampling and contrastive learning to construct positive–negative pairs, aligning name and color embeddings even under sparse supervision. Auxiliary binary cross-entropy loss (for collaborative filtering) and mean squared error loss (for RGB regression) further regularize the space, ensuring both semantic fidelity and perceptual accuracy. By explicitly addressing sparsity and imbalance, the framework enables robust representation learning at scale.

Extensive experiments demonstrate the effectiveness of our approach. Compared with baseline methods, our framework achieves a 33.6% improvement in Top-10 accuracy for color-to-name retrieval and a 7.0% reduction in CIELAB error for name-to-color prediction.

In summary, our contributions advance both the methodological and practical aspects of large-scale computational color naming, as follows:

- We propose a contrastive learning framework that unifies color naming and generation, explicitly designed to overcome the sparsity and imbalance challenges of large-scale datasets.
- We validate our method through both ablation studies and quantitative experiments, showing substantial gains in both functional performance and user-perceived quality.
- To facilitate reproducibility and foster future research, we have open-sourced the complete training and inference code of our model at https://github.com/IAMkecheng/contrastive-learning-color-name-model.
- Finally, we provide an project page with an *interactive* design tool implementing our method, supporting color-to-name recommendation and name-to-color generation, available at http://47.88.56.173:5000/.

2 Related Work

Understanding how humans perceive, name, and generate colors has long been a central topic at the intersection of vision science, linguistics, and human–computer interaction [2, 3, 5, 12, 14, 23, 33, 34]. Early studies examined universal tendencies in color naming [3, 5, 34], while more recent work has focused on computational modeling of perceptual–linguistic associations [4, 8, 21], neural text-to-color generation [13], and generative color-concept association [10]. We review prior work in three complementary directions: (1) color naming and perceptual reasoning, (2) computational models for color naming, and (3) color–concept association learning.

2.1 Color Naming & Semantic Understanding

Color naming provides an essential lens into how perceptual signals are mapped to linguistic categories. Berlin and Kay [3] identified eleven basic color terms that appear consistently across languages, inspiring subsequent studies on cross-linguistic variation in perceptual boundaries [5, 34]. Wierzbicka [33] emphasized that color semantics are conceptually grounded in environmental referents (e.g., sky, earth, fire) rather than purely neurophysiological mechanisms. Building on this foundation, the World Color Survey [5] enabled systematic comparisons of color naming across more than 110 languages.

To model these distributions, Benavente et al. [1] proposed a sigmoid–Gaussian function for fuzzy membership estimation, while Menegaz et al. [21] applied Delaunay triangulation in perceptual color space to approximate category boundaries. Benavente et al. [2] later released a fuzzy dataset anchored in the eleven basic categories, providing a resource for graded modeling. In parallel, Moroney [23] and Mylonas et al. [25] demonstrated the value of web-based platforms for scalable collection of naming data. The largest such effort, Munroe's XKCD survey [24], gathered more than three million name–color pairs, establishing a benchmark dataset for studying fine-grained perceptual naming.

Beyond linguistic and perceptual modeling, recent HCI research has increasingly explored how color semantics inform *inclusive design* and context-aware communication. Geddes et al. [6] proposed color–pattern design guidelines to improve accessibility for people with color vision deficiencies, while Zhu et al. [37] developed domain-specific color dictionaries that support consistent and interpretable naming across professional contexts. These efforts highlight how color semantics extend beyond cognitive categorization toward practical, human-centered applications in visualization, accessibility, and co-creative design.

Together, these works span controlled experiments, statistical models, and large-scale data collection. They reveal important regularities in human naming but remain primarily descriptive, leaving open the challenge of building computational models that generalize under data sparsity and compositional descriptors.

2.2 Computational Models for Color Naming

Building on these foundations, subsequent studies have developed computational systems to map names to colors. Kaufman [14] designed a rule-based interface for navigating HSV space with syntactic descriptors, while Mojsilovic [22] introduced a perceptually grounded naming scheme for region-based image analysis. Chuang et al. [4] modeled naming uncertainty using entropy-based measures, and Heer and Stone [8] applied non-parametric methods to the XKCD dataset, introducing saliency and name–distance metrics that have since been used in interactive tools [7, 18–20, 28]. Wang et al. [32] further developed an affective color transformation system that adjusts image tones according to emotion words.

Other work has integrated color naming into retrieval and recognition tasks. Liu et al. [17] incorporated high-level names into region-based image retrieval, bridging the semantic gap between low-level perception and user intent. Van de Weijer et al. [31] learned color names from weakly labeled real-world images using PLSA, explicitly highlighting sparsity as a bottleneck. Yu et al. [36] expanded the vocabulary with 28 non-basic descriptors to enhance discriminative resolution.

Inverse and categorical mappings have also been explored. Lindner et al. [16] predicted representative colors from abstract semantic expressions without embedding alignment. Setlur and Stone [29] exploited n-gram frequency and WordNet structure to assign semantically resonant colors for categorical data visualization. More recently, Jyothi and Okade [13] proposed Text2Color, an LSTM-based model that maps compositional descriptors (e.g., "light moss green") to RGB values and supports palette recommendation. However, their regression-based method lacks contrastive supervision and does not support bidirectional retrieval. Xu et al. [35] introduced *Color2Vec*, a web-based embedding model that learns word–color associations from large-scale image search data. It captures semantic relations and sociocultural variations in color meaning, but focuses on palette-level generation rather than fine-grained bidirectional color–name alignment.

Despite their diversity, most computational models treat naming as a one-way mapping task optimized with regression or cross-entropy losses. As a result, they struggle to generalize under sparse name–color correspondences and fail to align names and colors within a shared embedding space.

2.3 Color–Concept Association Learning

A complementary body of work links colors to abstract concepts. Jahanian et al. [12] used LDA-dual topic models to extract color–concept themes from design materials, while Ng and Chan [26] empirically compared associations made by designers and non-designers. Tham et al. [30] extended this line of inquiry to a cross-cultural framework, mapping abstract concepts to colors across languages.

Other studies have leveraged statistical and learning-based methods. Lin et al. [15] analyzed image distributions to assign semantically resonant colors to categories. Rathore et al. [27] estimated color–concept relations directly from image statistics, reducing reliance on manual annotation. Hu et al. [11] applied self-supervised colorization to capture conceptual priors, while Hegemann and Oulasvirta [9] proposed a cognitive framework—purpose, palette, prototype—to model the reasoning behind professional color choices.

Most recently, Hou et al. [10] introduced GenColor, a generative framework that uses diffusion models to synthesize images from abstract prompts and extract primary–accent palettes via segmentation. While effective for contextual palette design, such systems lack a unified embedding space and remain unsuited for bidirectional retrieval or learning under sparse supervision.

To overcome these limitations, we propose a contrastive learning framework for joint color–name embedding. Our method integrates a pre-trained Transformer-based name encoder, an RGB encoder, and an RGB generator within a multi-task architecture. By

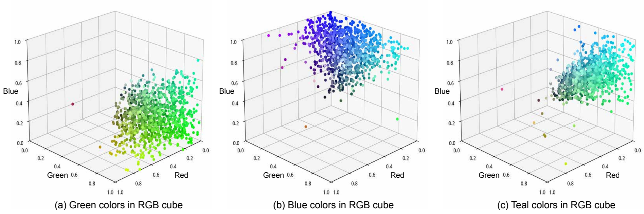

Figure 2: Color distributions in the RGB unit cube for three color names: (a) *green*, (b) *blue*, and (c) *teal*. Each panel plots 1,000 unique RGB samples randomly drawn from the XKCD dataset [24].

leveraging negative sampling and InfoNCE contrastive objectives, we construct a shared embedding space that generalizes to rare names, enables both name-to-color and color-to-name retrieval, and substantially improves accuracy and reconstruction fidelity under sparse supervision.

3 Method

3.1 Dataset Selection & Preparation

While the Heer and Stone model's vocabulary reduction to only 153 names simplifies the color name association problem [8], it inevitably discards valuable information from the long tail of infrequent color names. These rare names, though appearing only a few times in the dataset, play a crucial role in establishing semantic relationships between colors and the more commonly used names. For instance, a color that appears only with rare descriptors like "moss green" or "sage green" helps establish the semantic connection between that specific RGB value and the broader concept of "green." This semantic bridging effect is essential for learning robust color-name associations, as it provides additional context and constraints that guide the model toward better understanding of color semantics.

To this end, we base our experiments on the original large-scale crowdsourced dataset, namely the XKCD color survey [24], which provides millions of name–color correspondences. More specifically, it contains 3,252,135 color–name pair responses spanning 2,956,183 colors and 132,460 distinct names. This survey intentionally captured *free-form* naming behavior by asking participants to "type a name (word or phrase) you might use for that color,"[1] thereby reflecting spontaneous linguistic variation rather than constrained category selection. However, the raw survey data include noise, inconsistencies, and formatting heterogeneity. To address this, we implement a systematic cleaning pipeline with the following steps:

- **Normalization:** We lowercase all color names, replace any non-alphabetic character (punctuation, digits, hyphens, slashes, underscores, etc.) with a single space, collapse consecutive spaces, and trim. Consequently, hyphenated or slashed forms are tokenized as separate words (e.g., "blue-green" → "blue green", "blue/green" → "blue green") rather than merged.
- **Spelling correction:** Following Heer and Stone [8], we normalize common misspellings using a curated mapping; for example, "fuchsia" frequently appears as "fuscia", "fuschia", or "fushia", all mapped to "fuchsia".
- **Token filtering:** We retain only alphabetic characters and whitespace prior to tokenization, remove stopwords and semantically void tokens, and discard names that become empty after filtering.
- **Length control:** We require at least three characters and cap the name length at 32 tokens to ensure semantic compactness.

The final dataset consists of 3,245,135 cleaned name–color pairs, including 2,950,457 unique RGB values and 114,199 unique names, preserving more than 99.78% of the original data. We then randomly sample 10,000 responses as the test set and use the remainder for training.

3.2 Problem Formulation

Given a large-scale color dataset $\mathcal{D} = \{(r_i, g_i, b_i, n_i)\}_{i=1}^{N}$ where $(r_i, g_i, b_i) \in [0, 1]^3$ represents RGB color values and n_i represents the corresponding color name, our goal is to learn a unified embedding space that enables bidirectional mapping between colors and names.

A crucial characteristic of the XKCD dataset is the *many-to-many* relationship between colors and names: a single RGB color can be associated with multiple names (e.g., a specific RGB value might be called both "forest green" and "dark green"), while a single name can correspond to many different RGB values (e.g., "green" encompasses over 300,000 unique RGB values). Fig. 2 visualizes this phenomenon: for each name, the associated colors span a broad region of the color space, and, conversely, near-identical colors appear under different names. This many-to-many mapping poses substantial challenges for traditional approaches that implicitly assume a one-to-one correspondence, leading to biased estimates and brittle generalization.

The key challenges include: (1) severe sparsity of name-color correspondences, where many color names appear infrequently; (2) the many-to-many nature of the mapping, which requires the model to learn robust representations that can handle multiple valid associations; and (3) the need to capture the semantic similarity between different names for the same color and different colors for the same name.

3.3 Model Architecture

As illustrated in Fig. 3, our framework comprises three core components: (i) a pre-trained Transformer–based name encoder, (ii) an RGB encoder, and (iii) an RGB generator. All components are jointly trained to learn a shared embedding space of dimension $d = 64$.

3.3.1 Name Encoder. The name encoder $\mathcal{E}_n : \mathcal{N} \to \mathbb{R}^d$ maps color names to semantic embeddings using a pre-trained BERT model. Given a color name n, we first tokenize it using the BERT tokenizer and obtain token embeddings through the pre-trained BERT model:

$$\mathbf{h} = \text{BERT}(\text{tokenize}(n)) \quad (1)$$

[1] https://blog.xkcd.com/2010/03/01/color-name-survey/

where $\mathbf{h} \in \mathbb{R}^{L \times d_{bert}}$ is the sequence of hidden states from BERT, L is the sequence length, and $d_{bert} = 768$ is BERT's hidden dimension.

To obtain a fixed-size representation, we apply a projection layer that maps the BERT output to our target embedding dimension:

$$\mathbf{z}_n = \text{Normalize}(\text{MLP}(\mathbf{h})) \quad (2)$$

where the MLP consists of two linear layers with ReLU activation and dropout, mapping from d_{bert} to $2d$ and then to d, followed by L2 normalization.

3.3.2 RGB Encoder. The RGB encoder $\mathcal{E}_r : \mathbb{R}^3 \to \mathbb{R}^d$ maps RGB color values to the same embedding space. While RGB values alone provide limited information about color properties, we enhance the input representation by incorporating multiple color space representations to capture richer perceptual and semantic information.

Multi-Color Space Representation: For each RGB triplet (r, g, b), we convert it to a 16-dimensional feature vector by concatenating representations from different color spaces:

$$\mathbf{f} = [r, g, b, h, s, l, l_{lab}, a_{lab}, b_{lab}, h_{hcl}, c_{hcl}, l_{hcl}, c, m, y, k] \quad (3)$$

where the feature vector includes:

- **RGB (3D):** Original red, green, blue values
- **HSL (3D):** Hue, saturation, and lightness information
- **LAB (3D):** Perceptually uniform lightness and color-opponent dimensions
- **HCL (3D):** Hue, chroma, and lightness for better perceptual correspondence
- **CMYK (4D):** Cyan, magenta, yellow, and black for print color representation

This multi-color space representation provides rich information that is crucial for semantic understanding:

- **Brightness Information:** Lightness values from HSL, LAB, and HCL spaces capture how bright or dark a color appears, which directly corresponds to linguistic descriptors like "light blue" or "dark green"
- **Saturation Information:** Saturation and chroma values indicate color intensity, corresponding to terms like "vivid red" or "muted grey"
- **Hue Information:** Hue values across different spaces provide robust color category information, enabling better alignment with semantic color concepts

The RGB encoder then processes this enhanced feature vector through a multi-layer perceptron:

$$\mathbf{z}_r = \text{Normalize}(\text{MLP}(\mathbf{f})) \quad (4)$$

where the MLP has three linear layers: $d_{input} \to 128 \to 256 \to d$, with ReLU activations and L2 normalization. This multi-color space approach enables the encoder to learn richer representations that better align with the semantic space of color names, facilitating more accurate color-name associations.

3.3.3 Neural Collaborative Filtering (NCF) Module. To learn the similarity between RGB and name embeddings, we employ a Neural Collaborative Filtering module that takes concatenated embeddings and outputs a similarity score:

$$s = \text{NCF}(\mathbf{z}_r, \mathbf{z}_n) = \sigma(\text{MLP}([\mathbf{z}_r; \mathbf{z}_n])) \quad (5)$$

where $[\mathbf{z}_r; \mathbf{z}_n]$ denotes concatenation, and the MLP consists of three linear layers: $2d \to 256 \to 128 \to 1$, with ReLU activations and sigmoid output.

The NCF module enables efficient color-to-name recommendation by computing similarity scores between RGB embeddings and all name embeddings in the vocabulary. However, for name-to-color recommendation, this approach requires computing similarity scores against all RGB values in the dataset (approximately 3 million), which is computationally expensive and slow for real-time applications.

3.3.4 RGB Generator for Fast Inference. To address the computational bottleneck in name-to-color recommendation, we introduce an RGB generator $\mathcal{G} : \mathbb{R}^d \to \mathbb{R}^3$ that directly maps name embeddings to RGB values:

$$\hat{\mathbf{r}} = \sigma(\mathbf{W}_g \mathbf{z}_n + \mathbf{b}_g) \quad (6)$$

where $\mathbf{W}_g \in \mathbb{R}^{3 \times d}$ and $\mathbf{b}_g \in \mathbb{R}^3$ are learnable parameters, and σ is the sigmoid activation function ensuring output values are in $[0, 1]$.

The RGB generator is a simple linear layer that trades some accuracy for significant speed improvement. While the NCF-based approach provides more accurate recommendations by considering all possible RGB values, the generator enables real-time inference by directly producing RGB values from name embeddings without the need to compute similarity scores against the entire RGB database.

3.4 Loss Functions

Our training objective combines two loss functions to jointly optimize the embedding learning and RGB generation tasks:

3.4.1 Binary Classification Loss. For each RGB-color name pair (r, n), we generate a negative sample n^- by randomly sampling from the vocabulary. The NCF module is trained to distinguish positive pairs from negative ones using binary cross-entropy loss:

$$\mathcal{L}_{binary} = -\log(s^+) - \log(1 - s^-) \quad (7)$$

where $s^+ = \text{NCF}(\mathbf{z}_r, \mathbf{z}_n)$ and $s^- = \text{NCF}(\mathbf{z}_r, \mathbf{z}_{n^-})$ are the similarity scores for positive and negative pairs, respectively.

3.4.2 RGB Generation Loss. The RGB generator is trained to reconstruct the original RGB values from name embeddings using mean squared error loss:

$$\mathcal{L}_{mse} = \|\hat{\mathbf{r}} - \mathbf{r}\|_2^2 \quad (8)$$

where $\hat{\mathbf{r}} = \mathcal{G}(\mathbf{z}_n)$ is the generated RGB value and $\mathbf{r}$ is the ground truth. This loss function enables the generator to learn a direct mapping from name embeddings to RGB values, providing a fast alternative to the computationally expensive NCF-based recommendation approach.

3.4.3 Combined Objective. The final training objective is a weighted combination of the two losses:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{binary} + \lambda_2 \mathcal{L}_{mse} \quad (9)$$

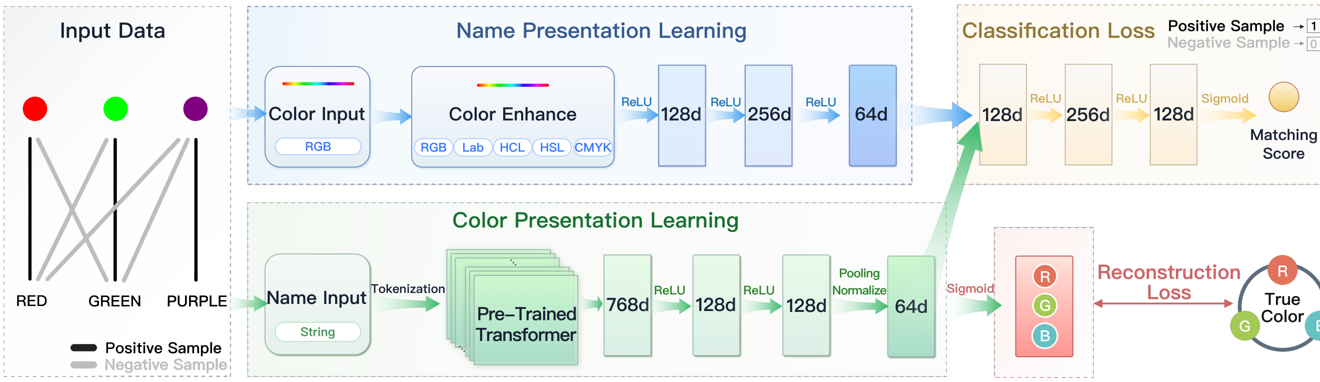

Figure 3: Framework of our proposed approach.

where $\lambda_1 = 0.01$ and $\lambda_2 = 1.0$ are the loss weights. The binary classification loss focuses on learning discriminative embeddings for accurate similarity matching, while the MSE loss enables the RGB generator to provide fast name-to-color mapping. This multi-task learning approach allows the model to simultaneously support both high-accuracy NCF-based recommendations and fast generator-based inference, offering flexibility in the trade-off between accuracy and speed.

3.5 Training Strategy

During training, we employ a comprehensive strategy to address the three key challenges identified in our problem formulation: data sparsity, many-to-many relationships, and semantic similarity learning.

3.5.1 *Addressing Data Sparsity (Challenge 1).*

- **Negative Sampling Strategy:** To overcome the severe sparsity of name-color correspondences, we employ random negative sampling for each positive RGB-name pair (r_i, g_i, b_i, n_i) by generating a negative sample n_j from the vocabulary such that $n_j \neq n_i$. This creates contrastive learning signals that help the model distinguish between valid and invalid color-name associations even with limited positive examples.
- **Uniform Sampling Distribution:** We employ uniform random sampling over all name–color pairs to mitigate data imbalance and ensure that rare names contribute equally to contrastive pair construction. This strategy allows the model to learn meaningful distinctions across the entire vocabulary despite the long-tail distribution of color names.

3.5.2 *Handling Many-to-Many Relationships (Challenge 2).*

- **Contrastive Learning Design:** Our contrastive learning approach naturally handles the many-to-many relationship by learning embeddings that cluster similar colors and names together in the shared space. The NCF module learns to assign high similarity scores to all valid color-name pairs, regardless of whether a color has multiple names or a name has multiple colors.
- **Soft Similarity Learning:** Instead of binary classification, we use continuous similarity scores that allow the model to learn nuanced relationships between colors and names, capturing the inherent ambiguity in color naming.

3.5.3 *Learning Semantic Similarity (Challenge 3).*

- **Shared Embedding Space:** By mapping both RGB values and color names to the same embedding space through the RGB encoder and name encoder, we enable the model to learn semantic relationships between colors and names. Similar colors and names are naturally clustered together in this shared space. The contrastive learning mechanism also ensures that semantically similar color-name pairs have high similarity scores, while dissimilar pairs have low scores. This enables the model to learn that different names for the same color are semantically related, and different colors for the same name share semantic properties.
- **Pre-trained Language Understanding:** The BERT-based name encoder provides rich semantic understanding capabilities, allowing the model to capture subtle semantic similarities between different color names (e.g., "forest green" and "dark green" for similar RGB values).

This comprehensive training strategy enables our model to learn a unified embedding space where similar colors and names are close together, while dissimilar ones are far apart, facilitating the bidirectional color-name transformation tasks and effectively addressing all three key challenges in color-name learning.

3.6 Training Procedure

The model is implemented in PyTorch and optimized using Adam with a learning rate of 1×10^{-4} for up to 60 epochs. Transformer weights are frozen during training to prevent overfitting on limited naming data, while projection layers are fine-tuned. Negative samples are generated on-the-fly to maximize coverage of rare descriptors. The model was trained and tested on an NVIDIA RTX 4090, ensuring high efficiency in both training and inference tasks.

4 Ablation Study

To validate the effectiveness of each component in our framework, we conduct comprehensive ablation studies by systematically removing individual components and evaluating their impact on performance. We evaluate three key tasks: Color-to-Name (C2N) recommendation using Top-K accuracy, Name-to-Color (N2C) recommendation using CIELAB distance, and Name-to-Color generation using CIELAB distance. All evaluations are performed on a held-out test set with 10,000 samples that is never used during training.

4.1 Experimental Setup

As shown in Tab. 1, we compare five different configurations by selectively enabling or disabling key components:

- **Configuration A:** Uses only 3D RGB input without enhanced multi-color space representation, but includes negative sampling, NCF, and MSE loss.
- **Configuration B:** Enables enhanced 16D RGB input but disables negative sampling, using only positive samples for training.
- **Configuration C:** Includes enhanced RGB and negative sampling but disables the NCF module, using simple inner product for similarity computation.
- **Configuration D:** Enables all components except the MSE loss in the RGB generator, focusing only on the contrastive learning objective.
- **Configuration F (Full Model):** Our complete framework with all components enabled.

4.2 Tasks & Metrics Design

To comprehensively evaluate our framework's performance across different tasks, we design three complementary evaluation metrics that capture both accuracy and perceptual quality:

Color-to-Name (C2N) Recommendation: We use Top-K accuracy metrics (Top-1 and Top-10) to measure how often the ground truth color name appears in the top-K recommendations. This metric evaluates the model's ability to understand color semantics and map RGB values to appropriate linguistic descriptions. Top-1 accuracy measures exact matches, while Top-10 accuracy captures near-miss performance, which is particularly important given the many-to-many nature of color naming.

Name-to-Color (N2C) recommendation. Following Jyothi et al. [13], we assess perceptual fidelity in CIE $L^*a^*b^*$ using the color difference ΔE between a recommended color and the ground truth. Because CIE $L^*a^*b^*$ is approximately perceptually uniform, smaller ΔE indicates a better perceptual match.

To enforce diversity among recommendations, we apply a JND-based separation constraint before scoring. Let τ denote one just noticeable difference (JND), with $\tau \approx 2.3$ in CIE $L^*a^*b^*$ [8]. Given the model's ranked candidate list, we construct a subset $S = \{c_1, \dots, c_m\}$ (greedy selection in rank order) such that

$$|S| \leq 10 \quad \text{and} \quad \Delta E(c_i, c_j) \geq \tau \;\; \forall\, i \neq j,$$

i.e., any two selected colors are mutually separated by at least one JND. We then report the evaluation score as the minimum perceptual error over this diverse set,

$$\min_{c \in S} \Delta E\big(c,\ c^\star\big),$$

where $c^\star$ is the ground-truth color. This criterion captures success if *any* of the top−10 mutually JND-separated recommendations is perceptually close to the target while avoiding redundancy among near-duplicates.

Name-to-Color (N2C) Generation: Similar to recommendation, we use CIELAB distance to evaluate the RGB generator's output quality. This metric assesses the trade-off between speed and accuracy in our generation approach, measuring how well the generator can produce perceptually reasonable colors from semantic descriptions.

The choice of these metrics is motivated by the need to evaluate both semantic understanding (C2N) and perceptual quality (N2C), while considering the practical requirements of real-world applications where both accuracy and speed are important.

4.3 Results and Analysis

Table 1 summarizes the ablation results across the three tasks; detailed findings are provided below.

Results.

- **Enhanced RGB representation (A vs. F).** Replacing raw 3-D RGB with the 16-D enhanced representation yields consistent (though modest) gains across tasks: C2N Top-1 improves from **34.89%** to **35.34%** (+0.45 pp; Top-10: 66.13% → 67.11%, +0.98 pp), N2C retrieval ΔE drops from **26.73** to **25.76** (−0.97, −3.6%), and N2C generation slightly improves (27.09 → 27.08). This indicates that augmenting brightness/saturation/hue cues helps learn more discriminative embeddings.
- **Negative sampling (B vs. F).** Removing negative sampling leads to a collapse in learning: C2N Top-1 falls to **0.00%** (Top-10: **0.00%**), and N2C retrieval ΔE explodes from **25.76** to **73.02** (+47.26, +64.7%). Negative sampling is therefore *critical* for contrastive alignment under sparse, imbalanced supervision.
- **NCF module (C vs. F).** Replacing NCF with a plain inner product severely harms recommendation: C2N Top-1 drops from **35.34%** to **6.58%** (−28.76 pp; Top-10: 67.11% → 31.91%, −35.20 pp), and N2C retrieval ΔE rises from **25.76** to **32.04** (+6.28, +19.6%). NCF's non-linear interaction learning is thus important for capturing complex name–color relations.
- **MSE loss in the RGB generator (D vs. F).** The regression loss is *essential* for N2C generation quality: without it, ΔE degrades sharply from **27.08** to **62.99** (+35.91, +57.0%). It also benefits retrieval and C2N (N2C-R ΔE: 28.19 → 25.76, −2.43; C2N Top-1: 34.49% → 35.34%, +0.85 pp).

4.4 Component Synergy

The ablations show strong complementarity. *Negative sampling* and *NCF* are the main drivers of representation quality under sparsity and many-to-many mappings; the *MSE loss* stabilizes and greatly improves the generator while modestly helping retrieval; the *enhanced RGB* encoding provides additional, consistent gains by injecting multi−color-space cues. Together, these components address data sparsity, many-to-many associations, and perceptual similarity learning to achieve the best overall performance (row F).

ID	Components				Results		
	Enhanced RGB	Neg. Sampling	NCF	MSE	C2N Top-1 / Top-10	N2C (R) ΔE (CIELAB)	N2C (G) ΔE (CIELAB)
A	✗	✓	✓	✓	34.89% / 66.13%	26.73	27.09
B	✓	✗	✓	✓	0.00% / 0.00%	73.02	27.13
C	✓	✓	✗	✓	6.58% / 31.91%	32.04	27.15
D	✓	✓	✓	✗	34.49% / 68.06%	28.19	62.99
F	✓	✓	✓	✓	**35.34% / 67.11%**	**25.76**	**27.08**

Table 1: Ablation of framework components. "Enhanced RGB" uses a 16-dimensional input in place of 3-D RGB; "Neg. Sampling" denotes the negative sampling procedure; "NCF" is the neural collaborative filtering module (when disabled, similarity is the inner product between aligned embeddings); "MSE" is the regression loss in the RGB generator. "C2N" is color-to-name recommendation; "N2C (R/G)" are name-to-color recommendation/generation, reported as ΔE in CIELAB (lower is better).

5 Quantitative Evaluation

We evaluate our framework on three core tasks using the same evaluation metrics as in our ablation study: Color-to-Name (C2N) recommendation, Name-to-Color (N2C) recommendation, and Name-to-Color (N2C) generation. All experiments are conducted on the same test dataset from XKCD, and we compare our approach against state-of-the-art baseline methods.

5.1 Test Data Preparation

To simulate real-world user scenarios and ensure comprehensive evaluation, we randomly sample 10,000 color-name pairs from our cleaned XKCD dataset. This approach provides a representative subset that reflects the diversity and complexity of actual user interactions with color naming systems.

However, it is important to note that this randomly sampled dataset is not suitable for evaluating the Heer and Stone baseline model [8] on the Name-to-Color (N2C) task. The Heer and Stone model operates on a fixed vocabulary of only 153 predefined color names, while our sampled dataset contains color names from the full XKCD vocabulary of 114,199 unique names. Since the Heer and Stone model cannot generate colors for names outside its fixed vocabulary, we exclude it from N2C evaluation on this dataset.

5.2 Color-to-Name Recommendation

The color-to-name recommendation task evaluates the model's ability to suggest appropriate color names for given RGB values. This task leverages the NCF module to compute similarity scores between RGB embeddings and all name embeddings in the vocabulary.

Evaluation Metrics: For each RGB value in the test set, we compute similarity scores with all color names and rank them by score. We evaluate using Top-K accuracy metrics (Top-1, Top-10) to measure how often the ground truth color name appears in the top-K recommendations.

Baseline Comparison: We compare our approach against Heer and Stone's probabilistic color naming model [8], which represents the state-of-the-art in computational color naming. Their model uses non-parametric methods to encode relationships in color naming datasets and has been widely used in interactive tools [7, 18–20].

Results: As shown in Tab. 2, our method attains 34.95% Top-1 and 71.26% Top-10, improving over Heer & Stone (34.67% / 53.34%). The Top-10 gain is +17.92 pp (≈ +33.6% relative), indicating substantially better name retrieval under ambiguity, while retaining sub-millisecond latency (0.43 ms). The substantial improvement demonstrates the effectiveness of our contrastive learning approach with enhanced RGB representation in learning discriminative embeddings that better capture the semantic relationship between colors and names. To further examine the model's behavior on infrequent color names, we analyzed the role of long-tail descriptors in the learned embedding space. Quantitatively, in 10,000 test cases, 15.67% of the Top-10 predictions contain compositional color names with three or more words (e.g., *"light light pink"*, *"royal navy blue"*, *"very light soft purple"*, *"slightly less light blue"*), indicating that the model successfully generates nuanced and compositionally rich descriptors. This suggests that while rare names have low frequency, they play an essential role in shaping the semantic continuity and fine-grained distinctiveness of the learned color–name embedding space.

5.3 Name-to-Color Recommendation

The name-to-color recommendation task evaluates the model's ability to find RGB values that match a given color name using the NCF-based recommendation approach. This task faces the computational challenge of evaluating against approximately 3 million RGB values in our dataset.

Evaluation Metrics: For each color name in the test set, we compute similarity scores with all RGB values using the NCF module and rank them by score. We evaluate using CIELAB distance metrics to measure perceptual similarity between recommended and ground truth RGB values.

Results: Our retrieval model achieves an average ΔE (CIELAB) of 25.34 with a per-query time of 3.75 ms. This demonstrates effective handling of the many-to-many mapping from a name (e.g., "green") to a large set of plausible RGB values. The results demonstrate our model's superior ability to handle the many-to-many

Task	Method	Top-1 ↑	Top-10 ↑	ΔE (LAB) ↓	Time (ms) ↓
C2N (Recommendation)	Heer & Stone [8]	34.67%	53.34%	–	0.01
	Ours	**34.95%**	**71.26%**	–	0.43
N2C (Recommendation)	**Ours**	–	–	**25.34**	3.75
N2C (Generation)	Text2Color [13]	–	–	27.24	0.7
	Ours	–	–	**26.61**	0.8

Table 2: Quantitative results on the XKCD test set. C2N = color-to-name recommendation; N2C = name-to-color. For C2N we report Top-k accuracy (%), for N2C we report the perceptual color difference ΔE (CIELAB, lower is better). Generation time is per query.

relationship, where a single name like "green" corresponds to hundreds of thousands of valid RGB values.

5.4 Name-to-Color Generation

The name-to-color generation task evaluates the RGB generator's ability to directly produce RGB values from color names, providing a fast alternative to the computationally expensive retrieval approach.

Evaluation Metrics: For each color name in the test set, we use the RGB generator to directly produce RGB values and compare them with ground truth RGB values using CIELAB distance metrics. We also measure inference time to demonstrate the speed-accuracy compromise.

Baseline Comparison: We compare against Jyothi and Okade's Text2Color model [13], which uses an LSTM-based approach to map compositional color descriptions to RGB values. This represents the current state-of-the-art in neural text-to-color generation.

Results: Our RGB generator attains an average $\Delta E = 26.61$ (CIELAB), outperforming Text2Color (27.24; -2% error). Inference is comparable to the baseline (0.8 ms vs. 0.7 ms per query) and is substantially faster than N2C retrieval (3.75 ms), providing a lightweight, low-latency alternative for name-to-color prediction.

5.5 Summary

Combining the quantitative results in Tab. 2 with the qualitative evidence in Fig. 4, our contrastive, dual-task framework with enhanced RGB encoding consistently improves both recommendation and generation. For **C2N**, we obtain **34.95%** Top-1 and **71.26%** Top-10 accuracy (vs. 34.67% / 53.34% for Heer & Stone), a **+17.92** pp Top-10 gain at 0.43 ms/query. Our method demonstrates consistent improvements across both Top-1 and Top-10 metrics. This indicates that our approach more accurately models the correlation between color and name, thereby enhancing the overall quality of name recommendations for any given color. Furthermore, due to the inherent uncertainty in the color-name mapping—where a single color can correspond to multiple names—recommending only a single name for a given color is often insufficient. Consequently, there is a need to recommend a set of relevant names. The greater improvement we achieve at Top-10 effectively addresses this requirement. In Fig. 4(a), the ground-truth label appears in the top-k list (**bold**), illustrating the typical retrieval behavior behind these numbers.

For **N2C (recommendation)**, our unified embedding achieves an average $\Delta E = \mathbf{25.34}$ (CIELAB) with 3.75 ms/query, while the exemplars in Fig. 4(b) show that the returned swatches are mutually JND-separated ($\tau \approx 2.3$), providing diverse, non-redundant matches to each name. For **N2C (generation)**, the RGB generator reaches $\Delta E = \mathbf{26.61}$ (vs. 27.24 for Text2Color) with comparable latency (0.8 ms vs. 0.7 ms). Despite employing a lightweight regression architecture optimized for real-time inference, our model achieves superior perceptual accuracy, benefiting from the discriminative representations learned through the contrastive learning framework.

Overall, the qualitative panels mirror the quantitative gains: contrastive alignment plus enhanced RGB encoding yields more semantically faithful name–color associations, stronger retrieval under ambiguity, and a lightweight generator for fast, accurate name-to-color prediction.

6 Interactive System

To promote the reproducibility and practical utility of our work, we have made both the model and the training code publicly available. The contrastive learning-based color-name model is hosted on GitHub at https://github.com/IAMkecheng/contrastive-learning-color-name-model. The model can be trained on an NVIDIA RTX 4070 Ti, with each epoch requiring only 0.35 hours, and the model typically converges after 30 epochs, ensuring efficient training.

In addition to the model code, we have also released a web-based interactive system[2] at http://47.88.56.173:5000/, which allows users to directly interact with the trained model for three key tasks:

- **Color-to-Name Recommendation**: Given an RGB color value, the system suggests a list of plausible color names that best describe the color, providing an intuitive mapping from perceptual colors to linguistic descriptions.
- **Name-to-Color Recommendation**: For a given color name, the system retrieves RGB values that are closest in perceptual space, offering a practical tool for designers to match specific color names with their visual counterparts.
- **Name-to-Color Generation (Quick Generation)**: This feature directly generates RGB values from color name embeddings, offering faster real-time inference by sacrificing

[2]The deployed system currently runs on CPU only, which results in slower response times compared to the GPU version.

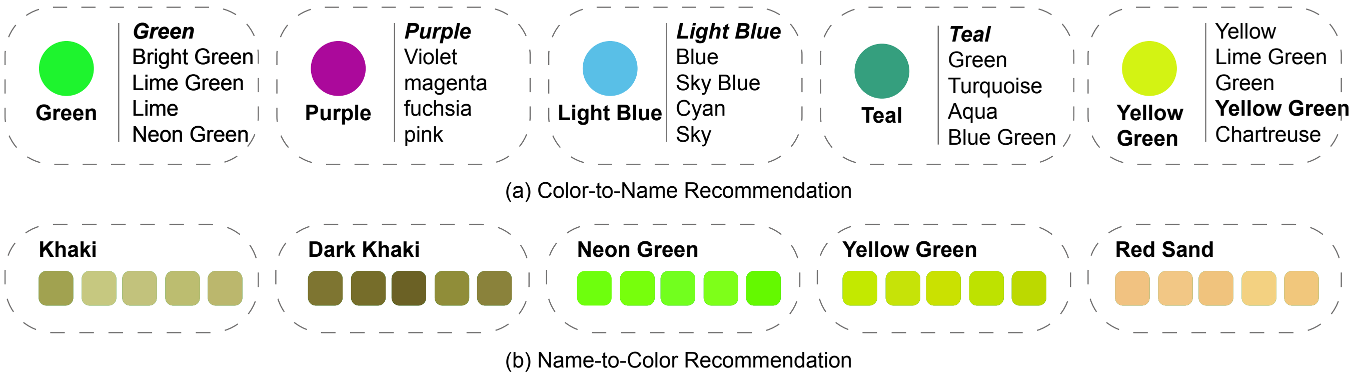

Figure 4: Qualitative results of the proposed framework. (a) Color-to-Name (C2N) recommendation: each panel shows the query color on the left (ground-truth swatch with its label) and our top-k predicted names on the right; within the recommendation list, the bold entry is the ground-truth label, indicating a correct hit. (b) Name-to-Color (N2C) recommendation: for each query name (title), we display the top-k RGB swatches predicted by our model from left to right; recommended colors are mutually JND-separated in CIE L*a*b* ($\tau \approx 2.3$).

some perceptual accuracy in favor of computational efficiency.

The system is designed to facilitate quick and easy use by designers and developers, enabling them to generate color names from RGB values or vice versa.

For ease of integration into custom projects, the system can be used as a Python library. Users can easily call the pre-trained model to generate color names, making it suitable for integration into various design, visualization, and human-computer interaction workflows.

By releasing both the model and the interactive web interface, we aim to ensure that our work is accessible to the research community and practical for real-world applications. This open-source approach eliminates the need for reproducing the model from scratch and facilitates benchmarking against competing models, supporting future work in this domain.

7 Conclusion, Discussion, and Future Work

7.1 Conclusion

In this work, we have presented a novel contrastive learning framework for color name generation and recommendation that addresses the fundamental challenges of large-scale color-name relationship mining. Our approach tackles three critical issues that have limited previous methods: severe data sparsity, complex many-to-many relationships, and the need for semantic similarity learning.

Key Contributions: Our framework introduces several key innovations that enable effective learning from sparse, large-scale color data. First, we propose a multi-task architecture that unifies color-to-name recommendation and name-to-color generation within a shared embedding space. Second, we demonstrate that enhanced RGB representation using multiple color spaces significantly improves semantic understanding by capturing brightness, saturation, and hue information that directly corresponds to linguistic descriptors. Third, we show that negative sampling and contrastive learning can effectively address data sparsity by creating learning signals from limited positive examples.

Experimental Validation: Our comprehensive evaluation demonstrates substantial improvements over existing methods. Compared to Heer and Stone's probabilistic model, we achieve 71.26% Top-10 accuracy in color-to-name recommendation (vs. 53.34%) and achieve $\Delta E = \mathbf{25.34}$ in perceptual accuracy for name-to-color retrieval. Against Jyothi and Okade's Text2Color model, our RGB generator achieves a better generation quality while providing comparable inference time. These results validate the effectiveness of our contrastive learning approach and multi-color space representation.

Practical Impact: The framework's ability to handle the many-to-many nature of color naming, where a single color can have multiple names and a single name can correspond to hundreds of thousands of RGB values, makes it particularly valuable for real-world applications in visualization, design, and human-computer interaction. The dual-task design provides flexibility in choosing between high-accuracy NCF-based recommendations and fast generator-based inference based on application requirements.

7.2 Discussion

Beyond algorithmic contributions, this work also advances the intersection between computational color modeling and HCI research by redefining color naming as a perceptual–linguistic interaction. Color naming represents a key channel through which humans communicate aesthetic and perceptual intent [7, 8, 18]. By constructing a shared embedding space between linguistic and perceptual modalities, our model lays the foundation for more intuitive and semantically grounded color interaction.

From an HCI perspective, the framework enables new forms of human–AI collaboration in color-related design tasks. Designers can express preferences through natural language (e.g., "make it slightly warmer" or "use a muted teal") and receive perceptually meaningful recommendations. These capabilities align with recent developments in interactive and co-creative visual systems [9, 19]. Moreover, the semantic embedding learned by our model supports applications such as adaptive visualization, explainable color mapping, and intelligent design assistance.

The framework also contributes to inclusive and accessible design. By explicitly modeling the semantic associations of color names, it can support users with color vision deficiencies through language-based feedback or perceptually distinguishable alternatives [6, 37]. Future extensions may incorporate user- or culture-specific adaptations, building personalized and context-aware color communication interfaces that better reflect diverse perceptual and linguistic experiences.

In summary, this work bridges computational modeling and human-centered color interaction. It demonstrates how large-scale perceptual–semantic alignment can inform the development of explainable, inclusive, and adaptive color systems for both research and industrial HCI applications.

7.3 Future Work

While our framework demonstrates significant improvements, several directions for future research remain promising:

Evaluation and User Studies: While our quantitative evaluation demonstrates substantial improvements over prior models in both color naming and generation accuracy, further user-centered validation is essential to assess the practical value of these gains. In particular, evaluating how designers and end-users perceive, interpret, and interact with the generated color recommendations remains an important direction. In future work, we plan to conduct perceptual and longitudinal studies using the released interactive tool to examine user satisfaction, creative utility, and alignment with real-world design workflows.

Cross-Cultural and Linguistic Extensions: Our current work focuses on English color names from the XKCD dataset. Future research could extend the framework to support multiple languages and cross-cultural color naming patterns, leveraging the rich cross-linguistic data from the World Color Survey and other multilingual resources. This would enable more inclusive and globally applicable color naming systems.

Compositional and Contextual Understanding: While our BERT-based name encoder captures some compositional aspects, there is room for improvement in handling complex color descriptions with modifiers, context, and emotional associations. Future work could explore more sophisticated language models or incorporate contextual information from surrounding text or visual context.

Interactive and Adaptive Learning: The current framework learns from static datasets. Future research could explore interactive learning approaches that adapt to user preferences and feedback, enabling personalized color naming systems that evolve with user needs and design contexts.

These future directions would further advance the field of computational color naming and create more powerful tools for color understanding, generation, and recommendation in various applications.

Acknowledgments

This work is supported by the grants of the NSFC (No.62502523, No.U2436209), the National Key R&D Program of China under Grant 2022ZD0160805, the Beijing Natural Science Foundation (L247027), the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China.

References

[1] Robert Benavente, Francesc Tous, Ramon Baldrich, and Maria Vanrell. 2002. Statistical modelling of a colour naming space. In *Conference on Colour in Graphics, Imaging, and Vision*, Vol. 1. Society of Imaging Science and Technology, 406–411.

[2] Robert Benavente, Maria Vanrell, and Ramon Baldrich. 2006. A data set for fuzzy colour naming. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur* 31, 1 (2006), 48–56.

[3] Brent Berlin and Paul Kay. 1991. *Basic color terms: Their universality and evolution.* Univ of California Press.

[4] Jason Chuang, Maureen Stone, and Pat Hanrahan. 2008. A probabilistic model of the categorical association between colors. In *Color and Imaging Conference*, Vol. 16. Society of Imaging Science and Technology, 6–11.

[5] Richard S Cook, Paul Kay, and Terry Regier. 2005. The world color survey database. In *Handbook of categorization in cognitive science.* Elsevier, 223–241.

[6] Connor Geddes, David R Flatla, Garreth W Tigwell, and Roshan L Peiris. 2022. Improving colour patterns to assist people with colour vision deficiency. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 1–17.

[7] C. C. Gramazio, D. H. Laidlaw, and K. B. Schloss. 2017. Colorgorical: creating discriminable and preferable color palettes for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 521–530. doi:10.1109/TVCG.2016.2598918

[8] Jeffrey Heer and Maureen Stone. 2012. Color naming models for color selection, image editing and palette design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 1007–1016.

[9] Lena Hegemann and Antti Oulasvirta. 2024. Palette, purpose, prototype: The three ps of color design and how designers navigate them. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems.* 1–19.

[10] Yihan Hou, Xingchen Zeng, Yusong Wang, Manling Yang, Xiaojiao Chen, and Wei Zeng. 2025. GenColor: Generative Color-Concept Association in Visual Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems.* 1–19.

[11] Ruizhen Hu, Ziqi Ye, Bin Chen, Oliver van Kaick, and Hui Huang. 2022. Self-supervised color-concept association via image colorization. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 247–256.

[12] Ali Jahanian, Shaiyan Keshvari, SVN Vishwanathan, and Jan P Allebach. 2017. Colors–Messengers of Concepts: Visual Design Mining for Learning Color Semantics. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 1 (2017), 1–39.

[13] Kondalarao Jyothi and Manish Okade. 2023. Text2Color Networks: Deep Learning Models for Color Generation from Compositional Color Descriptions. *International Journal on Artificial Intelligence Tools* 32, 06 (2023), 2350026.

[14] Arie Kaufman. 1986. Computer artist's color naming system. *The Visual Computer* 2, 4 (1986), 255–260.

[15] Sharon Lin, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. 2013. Selecting semantically-resonant colors for data visualization. In *Computer graphics forum*, Vol. 32. Wiley Online Library, 401–410.

[16] Albrecht Lindner, Nicolas Bonnier, and Sabine Süsstrunk. 2012. What is the color of chocolate?–extracting color values of semantic expressions. In *Conference on Colour in Graphics, Imaging, and Vision*, Vol. 6. Society of Imaging Science and Technology, 355–361.

[17] Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. 2005. Region-based image retrieval with high-level semantic color names. In *11th International Multimedia Modelling Conference.* IEEE, 180–187.

[18] K. Lu, M. Feng, X. Chen, M. Sedlmair, O. Deussen, D. Lischinski, Z. Cheng, and Y. Wang. 2021. Palettailor: discriminable colorization for categorical data. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 475–484. doi:10.1109/TVCG.2020.3030406

[19] Kecheng Lu, Khairi Reda, Oliver Deussen, and Yunhai Wang. 2023. Interactive context-preserving color highlighting for multiclass scatterplots. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–15.

[20] Kecheng Lu, Lihang Zhu, Yunhai Wang, Qiong Zeng, Weitao Song, and Khairi Reda. 2024. Color-Name Aware Optimization to Enhance the Perception of Transparent Overlapped Charts. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[21] Gloria Menegaz, Arnaud Le Troter, Jean Sequeira, and Jean-Marc Boi. 2006. A discrete model for color naming. *EURASIP Journal on Advances in Signal Processing* 2007, 1 (2006), 029125.

[22] Aleksandra Mojsilovic. 2005. A computational model for color naming and describing color composition of images. *IEEE Transactions on Image processing*

14, 5 (2005), 690–699.
[23] Nathan Moroney. 2003. Unconstrained web-based color naming experiment. In *Color imaging VIII: Processing, hardcopy, and applications*, Vol. 5008. SPIE, 36–46.
[24] Randall Munroe. 2010. Color Survey Results. https://blog.xkcd.com/2010/05/03/color-survey-results/. Accessed September 8, 2025.
[25] Dimitris Mylonas, Lindsay MacDonald, and Sophie Wuerger. 2010. Towards an online color naming model. In *Color and imaging conference*, Vol. 18. Society of Imaging Science and Technology, 140–144.
[26] Annie WY Ng and Alan HS Chan. 2018. Color associations among designers and non-designers for common warning and operation concepts. *Applied ergonomics* 70 (2018), 18–25.
[27] Ragini Rathore, Zachary Leggon, Laurent Lessard, and Karen B Schloss. 2019. Estimating color-concept associations from image statistics. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 1226–1235.
[28] Amey Salvi, Kecheng Lu, Michael E Papka, Yunhai Wang, and Khairi Reda. 2024. Color maker: a mixed-initiative approach to creating accessible color maps. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.
[29] Vidya Setlur and Maureen C Stone. 2015. A linguistic approach to categorical color assignment for data visualization. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 698–707.
[30] Diana Su Yun Tham, Paul T Sowden, Alexandra Grandison, Anna Franklin, Anna Kai Win Lee, Michelle Ng, Juhyun Park, Weiguo Pang, and Jingwen Zhao. 2020. A systematic investigation of conceptual color associations. *Journal of Experimental Psychology: General* 149, 7 (2020), 1311.
[31] Joost Van De Weijer, Cordelia Schmid, Jakob Verbeek, and Diane Larlus. 2009. Learning color names for real-world applications. *IEEE Transactions on Image Processing* 18, 7 (2009), 1512–1523.
[32] Xiaohui Wang, Jia Jia, and Lianhong Cai. 2013. Affective image adjustment with a single word. *The Visual Computer* 29, 11 (2013), 1121–1133.
[33] Anna Wierzbicka. 1990. The meaning of color terms: semantics, culture, and cognition. (1990).
[34] Jonathan Winawer, Nathan Witthoft, Michael C Frank, Lisa Wu, Alex R Wade, and Lera Boroditsky. 2007. Russian blues reveal effects of language on color discrimination. *Proceedings of the national academy of sciences* 104, 19 (2007), 7780–7785.
[35] Long Xu, Su Jin Park, and Sangwon Lee. 2023. Color2Vec: web-based modeling of word-color association with sociocultural contexts. *ACM Transactions on Computer-Human Interaction* 30, 4 (2023), 1–29.
[36] Lu Yu, Lichao Zhang, Joost van de Weijer, Fahad Shahbaz Khan, Yongmei Cheng, and C Alejandro Parraga. 2018. Beyond eleven color names for image understanding. *Machine Vision and Applications* 29, 2 (2018), 361–373.
[37] Zhenyang Zhu, Jiyi Li, Ying Tang, Kentaro Go, Masahiro Toyoura, Kenji Kashiwagi, Issei Fujishiro, and Xiaoyang Mao. 2023. Cc-glasses: Color communication support for people with color vision deficiency using augmented reality and deep learning. In *Proceedings of the Augmented Humans International Conference 2023*. 190–199.