

Article

Research on a Color Emotion Perception Method Integrating Unsupervised Clustering and Deep Learning

Yinghan Lan ^{1,*}¹ School of Design, Royal College of Art, London, UK

* Correspondence: Yinghan Lan, School of Design, Royal College of Art, London, UK

Abstract: This paper addresses the problem of color emotion perception by proposing an analytical method that integrates unsupervised clustering and deep learning. The aim is to automatically extract emotional information from color features and achieve interpretable modeling. First, statistical features of single-color images are extracted based on the HSV color space, and a regression model is constructed to predict pleasure scores. Comparative analysis is performed using various models, including linear regression, decision trees, random forests, k-nearest neighbor (KNN), and fully connected neural networks. The results show that nonlinear models and local modeling methods based on similar samples perform best. In the classification task, by constructing a binary label for "stimulation," the basic classification model can accurately characterize the psychological feeling of color in the arousal dimension. The deep learning model further demonstrates powerful feature extraction capabilities. Its convolutional feature map visualization results show significant differences in the local feature responses of different color samples. Second, K-means clustering is used to group color samples in an unsupervised manner, and semantic interpretation is performed based on emotion proportions. Color clusters are divided into types such as "calm and pleasant," "stimulating and active," and "energetic and pleasant," revealing distribution patterns of color in the dimensions of pleasure, calmness, and stimulation. Comprehensive analysis shows that brightness and saturation features significantly influence perceived pleasure, while stimulation and overall emotional atmosphere can be intuitively expressed through clustering results. The method presented in this paper not only achieves quantifiable modeling of color emotion perception but also provides data support for emotion-oriented color design and application, laying the foundation for future research on emotion prediction in multi-color combinations and complex scenarios.

Keywords: color emotion perception; HSV color features; regression and classification modeling; deep learning; convolutional feature maps; unsupervised clustering; semantic interpretation of emotion

Published: 06 March 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Research Background and Significance

With the continuous development of artificial intelligence and data-driven methods, human-computer interaction systems are gradually shifting from a traditional function-oriented approach to one focused on emotion perception and emotional interaction. How computer systems understand human subjective feelings has become a key research focus in fields such as emotion computing, computer vision, and intelligent design. Visual information is the primary channel through which humans perceive the external environment, and color, with its intuitiveness and strong emotional induction capability, plays a fundamental and crucial role in emotional expression and psychological reactions.

Numerous studies in psychology and visual design have shown that different colors or color combinations can evoke varying degrees of emotional responses, such as pleasure,

displeasure, calmness, or stimulation, thereby influencing cognition, judgment, and behavioral decisions. Therefore, color is widely used in many design fields, including product design, user interface design, advertising design, and spatial environment design. Traditional color design is mostly based on designers' experience and intuition, lacking systematic emotion modeling methods and making it difficult to meet the demands of intelligent design and automated decision-making.

From a methodological perspective, existing research on color emotion largely relies on questionnaires and statistical analysis to examine emotional evaluations of participants under different conditions, thereby revealing relationships between color and emotion. Although such research is valuable for indicating overall trends, it still has significant limitations. First, traditional statistical methods generally assume simple relationships between variables, making it difficult to capture the complex nonlinear mapping between color perception and emotional response. Second, many studies focus on descriptive analysis, lack predictive model structures, and thus cannot be directly applied to real-world systems.

In recent years, machine learning and deep learning technologies have made significant progress in pattern recognition and computer vision, providing new technical means for color emotion perception research. Learning-based models can map color visual information to subjective emotional evaluations from image input, overcoming limitations of manual experience and rule-based design. However, supervised learning for emotion modeling typically requires large amounts of labeled data and often struggles to reveal latent structural characteristics of color samples at the emotion perception level.

Against this backdrop, unsupervised learning methods, especially clustering analysis, offer new insights into color emotion research. By clustering color samples in an emotion evaluation space or a color feature space, inherent grouping patterns can be discovered without relying on manual labeling. Different colors may naturally form groups centered on emotional characteristics such as "soothing," "active," or "neutral and stable." Applying clustering analysis to color emotion research is beneficial for revealing underlying structures of color perception from a data-driven perspective, thereby supporting subsequent modeling and interpretation. Combining unsupervised clustering with supervised learning and deep learning methods to establish a systematic framework for modeling color image emotion perception can help compensate for shortcomings of traditional approaches and provides new technical means to explore intrinsic laws of color perception.

From a theoretical perspective, this paper investigates the relationship between color images and subjective emotion perception by applying a combination of clustering analysis and deep learning to model color emotion perception and conducting multi-level exploration. Using K-means clustering to analyze the distribution of color samples across emotion evaluation dimensions, the study identifies potential inherent grouping structures at the emotion perception level, providing data support for understanding latent patterns in color perception. Furthermore, supervised learning models are established for emotion prediction, shifting the research from describing phenomena to constructing predictive models and enriching the theoretical view of color emotion research.

Methodologically, this paper combines unsupervised and supervised learning to form a relatively complete modeling pipeline for color emotion perception. First, clustering analysis is used to identify possible categorical structures of color samples in the emotion perception space. Then, traditional machine learning models and convolutional neural networks are employed to model pleasure regression and stimulation classification tasks from color images. Compared with studies that use a single modeling approach, this framework can both discover structure and make predictions, leading to more systematic and comprehensive analysis. In deep learning modeling, this paper uses color images directly as input, reducing reliance on handcrafted features. The model automatically learns visual features related to emotion perception. By visualizing

feature maps from intermediate layers of the convolutional neural network, the extracted features are further analyzed, improving interpretability to a certain extent and supporting reliable application of deep learning methods in emotion perception.

From an application perspective, the proposed color emotion perception modeling method can provide technical support for emotion-oriented color design and intelligent decision-making. The color emotion clusters obtained from clustering analysis can serve as reference templates in design, while deep learning-based prediction models can help assess emotional responses that new color schemes may evoke, assisting designers in making more scientific decisions. This research provides a methodological foundation for subsequent color generation and optimization under emotion constraints, and it has potential for further extension and application.

Overall, by combining unsupervised clustering analysis and deep learning, this paper systematically studies emotion perception in color images and provides new ideas and references for related research in emotion computing and intelligent design.

1.2. Literature Review

Artificial intelligence and machine learning technologies are developing rapidly and have been extensively studied and explored in art design and visual creativity. He Changming reported that artificial intelligence has profoundly influenced the creative methods and aesthetic concepts of art design. Machine learning has been widely applied in visual design, interaction design, and music composition, improving design efficiency and reshaping the generation of ideas and inspiration [1-3]. This line of research suggests that intelligent creative approaches based on human-machine collaboration are emerging, bringing new momentum and possibilities to innovation in the field of art design. It also outlines broad trends introduced by artificial intelligence in art design and provides a theoretical background for studies on specific techniques.

In more specific design research, many scholars have begun to consider the impact of artificial intelligence on environmental art and spatial design. Han Yong and Li Xinhui summarized and discussed the application of artificial intelligence in environmental art design from the perspective of sustainable development. They noted that artificial intelligence can play an important guiding role in promoting eco-friendly design, improving efficiency, and facilitating interdisciplinary collaboration [2,4-6]. However, this type of research is largely conducted at the macro or system level and does not sufficiently examine specific visual elements, such as color and graphics, or how these elements affect users' emotions and psychological experiences. In particular, quantitative modeling of emotional responses to specific visual elements remains limited.

Regarding intelligent modeling of color and other visual elements, some studies have begun to employ deep learning and generative models [7]. Zhang Jingtong used Regong art as the research object and applied generative adversarial networks (GANs) and variational autoencoders (VAEs) to model and generate color and pattern features of traditional art, further applying traditional cultural elements to modern clothing design [3-5]. These results suggest that intelligent generation techniques can improve the efficiency and creativity of color design and, to some extent, support the digital inheritance of cultural symbols [8]. However, such studies mainly focus on style generation and design applications, and rarely address emotional perception elicited by color itself [9,10].

Other scholars, such as Chen Zhengkai, applied machine learning methods to daily-use ceramic decoration design using Dunhuang murals as a case study, exploring intelligent extraction and application of traditional art elements in modern decorative design [4,11-15]. These studies have confirmed the feasibility and advantages of machine learning in extracting decorative patterns and color elements and have provided new technical means for integrating traditional art with modern design. Nevertheless, they

tend to emphasize formal characteristics and cultural expression rather than systematic modeling and analysis of user emotional feedback.

In summary, artificial intelligence and machine learning have produced substantial research outcomes in art design, environmental design, and color generation. However, most studies focus on improving design efficiency, generating styles, or inheriting cultural elements, while systematic and quantitative modeling of color-induced emotional perception remains insufficient. Moreover, existing research rarely integrates unsupervised learning with supervised learning to simultaneously mine structural patterns of color emotion from data and build predictive models. Based on this research gap, this paper takes color images as the research object and combines clustering analysis, traditional machine learning methods, and deep learning methods to model color emotion perception from a data-driven perspective. It analyzes emotional feedback in terms of pleasure and stimulation, aiming not only to reveal potential structural patterns in emotional perception but also to establish predictive models from color images to emotion evaluation. This work provides a quantifiable and interpretable analytical method for emotion-oriented color design and, to some extent, fills gaps in existing research on color emotion perception modeling while offering new directions for applying artificial intelligence in art design and emotion computing.

1.3. Research Ideas and Methods

This paper focuses on the core question of how color images evoke different emotional perceptions. Using a data-driven approach, it establishes a research framework for color emotion perception that integrates unsupervised clustering analysis with supervised learning. The overall approach unfolds in four stages: data acquisition and statistical analysis, emotion structure mining, emotion prediction modeling, and result analysis and interpretation. It examines emotion perception characteristics of color images from two perspectives: overall distribution patterns and individual predictive capability.

By analyzing questionnaire-based emotion data corresponding to color samples, a preliminary understanding of pleasure levels and calmness/stimulation levels across colors is obtained, providing a data foundation for subsequent modeling. Unsupervised clustering is then introduced to analyze the distribution of color samples in the emotion evaluation space and uncover latent structures in the emotion perception dimension. Based on clustering results, traditional machine learning models and convolutional neural network models are trained to map color images to emotion perception outcomes. Model performance is compared, and results are interpreted through both quantitative evaluation and intermediate feature visualization.

This paper uses color images as stimuli and quantifies color-induced emotion perception via questionnaire survey data. The questionnaire results include multiple evaluation dimensions, such as proportions of different levels of pleasure, calmness, and stimulation. For modeling and analysis, the original questionnaire statistics are organized into a structured dataset.

For emotion representation, this paper models the data from two perspectives: continuous emotion evaluation and discrete psychological state. First, a continuous pleasure score is constructed from the proportions of different pleasure levels and used as the regression target, reflecting the overall pleasure level evoked by each color. Second, samples are binarized using the stimulation proportion to construct a "stimulating" label, describing the psychological arousal response to color. In this way, subjective emotion evaluations are transformed into numerical labels suitable for machine learning and deep learning.

To explore potential structures in the emotion perception space, K-means clustering is applied to color samples in an unsupervised manner. The clustering features primarily include proportions of pleasure-related responses and the calmness/stimulation proportions, describing similarity between colors from the perspective of emotion

evaluation. Through clustering, natural groupings of colors at the emotion perception level can be discovered without predefined emotion categories.

In analyzing clustering results, semantic interpretations are provided based on the statistical characteristics of emotion indicators within each cluster. For example, clusters with high pleasure proportions and low stimulation proportions can be interpreted as "soothing" clusters, whereas clusters with high stimulation proportions can be interpreted as "active" clusters. This maps data-driven outcomes to interpretable emotional semantics and provides references for subsequent supervised modeling and application analysis.

In the supervised learning stage, several traditional machine learning methods are employed, including logistic regression, decision trees, and random forests. For the pleasure score prediction task, regression models are used to predict continuous emotion evaluation; for the stimulation judgment task, classification models are used to predict a binary emotional state. During model training, strategies such as bootstrap sampling and oversampling are used to address small sample sizes and class imbalance, according to task characteristics. By comparing model performance, the suitability of different machine learning methods for color emotion perception is evaluated, providing a baseline for subsequent deep learning experiments.

To further improve emotion perception from color images, this paper constructs end-to-end prediction models using convolutional neural networks (CNNs). Unlike traditional approaches that rely on handcrafted features, CNNs can automatically learn hierarchical visual features relevant to emotion perception from image inputs, making fuller use of color information. Separate CNN architectures are designed for regression and classification tasks to predict pleasure scores and stimulation labels, respectively. During training, input images are standardized in size and normalized to improve stability and generalization. Loss functions and evaluation metrics are recorded across epochs to analyze convergence behavior [16].

In the result analysis stage, the modeling performance is evaluated both quantitatively and qualitatively. Quantitatively, predictive performance is compared using metrics such as regression error, coefficient of determination, and classification accuracy. Qualitatively, feature maps from intermediate CNN layers are visualized to examine feature extraction patterns across different color samples, thereby exploring which regions and patterns the model attends to.

Through the above research ideas and methods, this paper completes a full pipeline from emotion structure discovery to emotion prediction modeling, providing a systematic method for emotion perception research based on color images [7-10].

2. Introduction to Data Analysis Technologies

2.1. Introduction to the Decision Tree Model

The decision tree algorithm is a powerful tool widely used in machine learning and data mining and can be applied to both classification and regression tasks.

2.1.1. Principle of Decision Trees

A decision tree is a tree-structured model that divides a dataset into different subsets, where each subset corresponds to a decision path. The model selects the best feature for partitioning the data by minimizing uncertainty, thereby achieving effective classification or prediction. Each node in a decision tree represents a feature, each edge represents a branch corresponding to a feature value, and each leaf node contains the final classification or regression output [11].

2.1.2. Construction Process of Decision Trees

The construction of a decision tree is recursive and typically includes the following steps:

- 1) Optimal feature selection: selecting the best splitting feature by maximizing information gain or reducing impurity.
- 2) Data splitting: dividing the data into multiple subsets based on the selected feature.
- 3) Recursively building subtrees: repeating the above process for each subset until stopping conditions (e.g., maximum depth or minimum number of samples) are met.

2.2. Introduction to the Random Forest Algorithm

Random forest is an ensemble learning method built on decision trees. It constructs multiple decision trees and combines their outputs to improve model performance and robustness. The core idea of random forest is "majority voting" (for classification) or "average prediction" (for regression). It performs well in a wide range of machine learning tasks, including classification, regression, and feature selection [12].

2.2.1. Ensemble of Decision Trees

Random forest reduces the risk of overfitting associated with individual decision trees and improves generalization by aggregating predictions from many trees.

2.2.2. Randomness

Random forest introduces two types of randomness:

- 1) Random sampling: each decision tree is trained on a bootstrap sample drawn from the original dataset (sampling with replacement), ensuring diversity across trees.
- 2) Random feature selection: at each split, a random subset of features is considered rather than evaluating all features, further increasing diversity and reducing correlation among trees.

3. Data Acquisition and Preprocessing

3.1. Data Sources and Preprocessing

3.1.1. Data source

The data used in this paper come from a questionnaire-based experiment on visual emotion perception of color. Different color images were used as stimuli, and participants were invited to evaluate their subjective emotional experiences evoked by the colors. Each color sample was presented as an image, with all images kept identical in size and presentation format to minimize interference from non-color factors in emotion perception.

In the questionnaire design, emotion evaluation was divided into two dimensions: (1) pleasure, describing the degree of pleasure evoked by the color, and (2) arousal, describing the degree of calmness or stimulation evoked by the color. After viewing each color image, participants selected emotion options based on their subjective feelings. After aggregating responses from multiple participants, the emotion proportions corresponding to each color were obtained.

During data processing, each color image file was mapped one-to-one with its questionnaire statistics, forming a structured data table. This table contains image path information and multiple statistical indicators reflecting color emotion perception, providing a data foundation for subsequent statistical analysis, clustering analysis, and machine learning and deep learning modeling.

3.1.2. Overall Structure Description of the Data

The final color dataset is stored in tabular form. Each row represents a color sample (i.e., one color image). Each column corresponds to statistical variables or auxiliary

information describing the sample across different emotional dimensions. The dataset includes numerical and categorical variables, and its structure is clear and suitable for subsequent analysis and modeling. A representative sample of the dataset is shown in Figure 1.

| Hue | Very Unpleasant | Unpleasant | Neutral | Pleasant | Very Pleasant | Calm | Arousal |
|-----|-----------------|------------|---------|----------|---------------|-------|---------|
| 0 | 3.17 | 5.67 | 37.19 | 32.20 | 21.77 | 39.23 | 60.77 |
| 0 | 1.13 | 5.67 | 42.40 | 33.79 | 17.01 | 75.06 | 24.94 |
| 0 | 5.44 | 17.69 | 39.68 | 21.54 | 15.65 | 49.43 | 50.57 |
| 0 | 3.40 | 12.93 | 31.75 | 28.12 | 23.81 | 70.07 | 29.93 |
| 0 | 2.27 | 7.03 | 38.55 | 30.39 | 21.77 | 85.49 | 14.51 |

Figure 1. Data sample.

In modeling, image data serve as the primary model input. The emotion statistics in the table are mainly used for label construction in supervised learning tasks, as well as for clustering and exploratory data analysis. In this way, a complete mapping from visual stimuli to subjective emotion evaluations is established.

The data fields used in this paper and their meanings are as follows.

(1) Color path

This field stores the path of the color image file in the local file system. During model training and testing, it is used to load the corresponding color image. It does not participate directly in numerical computations but is critical for data loading and sample indexing.

(2) Hue class

This field represents the hue category of a color and encodes hue differences using integer labels. In this paper, it is mainly used for auxiliary analysis and comparison and is not a direct input feature for deep learning models.

(3) Very unpleasant ratio

This field indicates the proportion of respondents who selected the "very unhappy" option in the questionnaire and reflects strong negative emotions elicited by the color.

(4) Unpleasant ratio

This field indicates the proportion of respondents who selected the "unhappy" option and reflects general negative emotions elicited by the color.

(5) Neutral (general) ratio

This field indicates the proportion of respondents who selected the "general" option and reflects emotional neutrality.

(6) Pleasant ratio

This field indicates the proportion of respondents who selected the "pleasant" option and reflects positive emotional experiences elicited by the color.

(7) Very pleasant ratio

This field indicates the proportion of respondents who selected the "very pleasant" option and reflects strong positive emotions elicited by the color.

3.2. Label value calculation

This paper aims to achieve quantitative modeling of emotion perception in color images. Therefore, the emotion proportion data obtained from the questionnaire need to be converted into numerical labels that can be used to train machine learning and deep learning models. In this paper, the pleasure dimension is converted into a continuous pleasure score using a weighted average, which is used as the regression target to reflect the overall pleasure evoked by a color.

In the questionnaire design, the pleasure dimension includes five levels: "very unhappy," "unhappy," "average," "pleasant," and "very pleasant." Each level corresponds to the proportion selected by participants. For numerical modeling, this paper assigns weights from 0 to 4 to these five levels in ascending order of pleasure intensity. The proportions are multiplied by their corresponding weights and summed to obtain a weighted value representing overall pleasure. Finally, the weighted value is divided by 100 to scale the pleasure score into the continuous range of 0 to 4.

This method preserves the contributions of different pleasure levels in the overall evaluation and avoids information loss that could occur when using only the maximum proportion. As a result, the pleasure score can more accurately reflect emotional differences among colors. This pleasure score serves as the prediction target for the regression task and is used to evaluate model performance in predicting color-induced pleasure.

In addition to pleasure, this paper also focuses on arousal-related perception, i.e., whether a color produces an obvious stimulating feeling. In the questionnaire, the arousal dimension is described using two options: "calm" and "stimulating," which are summarized as proportions.

To formulate a classification task, this paper generates a binary stimulation label using a thresholding strategy. If the proportion of "stimulating" responses for a sample exceeds a given threshold, the color is considered stimulating; otherwise, it is considered non-stimulating. Based on the data distribution, this paper uses a threshold of 50% and generates a binary variable `is_active`, where 1 represents a stimulating color and 0 represents a non-stimulating color.

This thresholding method is intuitive and interpretable, reflecting the majority judgment of participants regarding whether a color is stimulating. Moreover, modeling stimulation as a binary classification problem simplifies the model structure and makes predictions easier to interpret and apply in practice.

4. Exploratory Data Analysis

4.1. Overall Difference Analysis of Pleasure Scores under Different Hues

As shown in Figure 2, the bar chart illustrates the average pleasure scores for different hue categories. It can be observed that emotional perception varies across hues in the pleasure dimension. The average pleasure scores among hue categories are unbalanced, indicating that hue is an important factor affecting pleasure perception.

Some hue categories have significantly higher average pleasure scores, suggesting that these colors are more likely to evoke positive emotional experiences. In contrast, some hue categories have relatively lower average pleasure scores, implying that these colors tend to be more neutral or negative in emotional perception.

This observation verifies the differences in color emotion perception and provides an intuitive basis for subsequent emotion modeling based on hue-related characteristics.

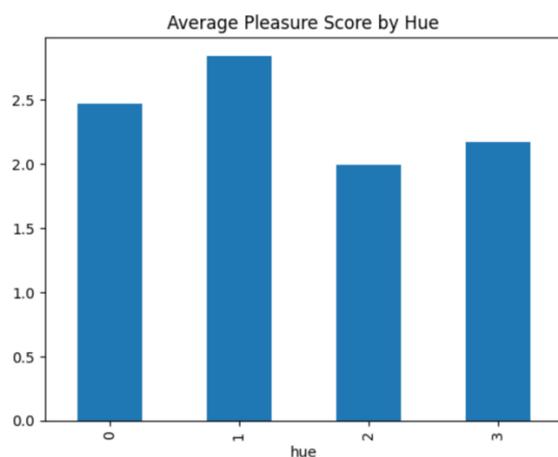


Figure 2. Overall differences in pleasure scores across hue categories.

4.2. Comparative Analysis of Calm and Stimulating Sensations under Different Hues

By comparing the average proportions of calmness and stimulation across different hue categories, differences in arousal responses can be identified. As shown in Figure 3, a higher calmness proportion in certain hues indicates that these colors are more likely to evoke low-arousal and stable emotions, whereas a higher stimulation proportion in other hues corresponds to higher arousal. This indicates that colors not only differ in pleasure but also show significant differences in emotional arousal.

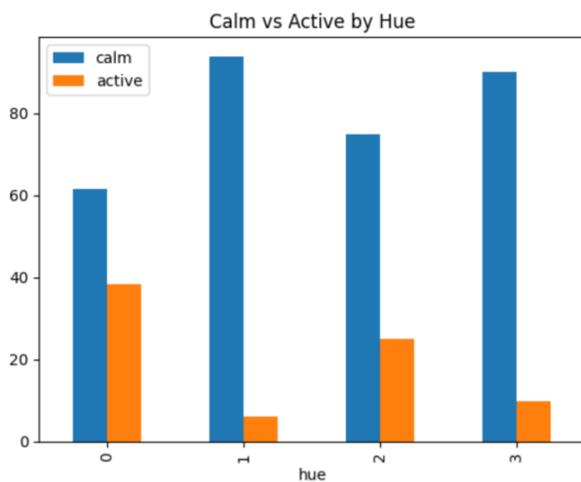


Figure 3. Comparison of calm and stimulating sensations across hue categories.

These findings provide an empirical basis for the subsequent stimulation classification model and also support semantic interpretation of emotion clusters in the clustering analysis.

4.3. Histogram Analysis of Pleasure Scores

From the histogram distribution of pleasure scores, as shown in Figure 4, it can be seen that the samples show a certain degree of concentration in the pleasure dimension. Pleasure scores are mainly distributed in the middle-to-high range, with relatively few samples at extremely low or extremely high values. This suggests that most color samples tend to evoke neutral or mildly pleasant emotions, whereas strong negative or extremely positive emotional experiences are less common. This distribution reflects, to some extent, the overall balance of color emotion perception and suggests that the impact of uneven sample distribution on model training should be considered in subsequent modeling.

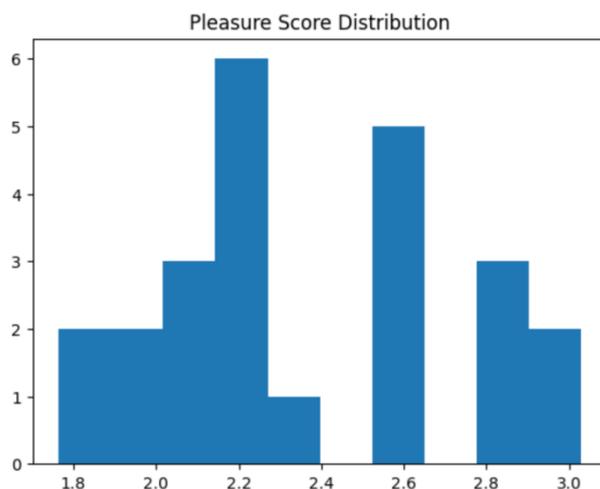


Figure 4. Histogram of pleasure scores.

4.4. Dispersion Analysis of Stimulation Ratio and Pleasure

By analyzing the scatter relationship between the stimulation ratio and the pleasure score, it can be observed that there is no simple linear correlation between these two variables. Some colors exhibit relatively high stimulation while still maintaining a high pleasure level, whereas some colors with low stimulation show varying levels of pleasure. This indicates that stimulation and pleasure are relatively independent yet interrelated aspects of color emotion perception; therefore, one cannot be inferred directly from the other. This result further supports the rationale for the multi-task modeling strategy adopted in this paper, where pleasure is modeled via regression and stimulation is modeled via classification (Figure 5).

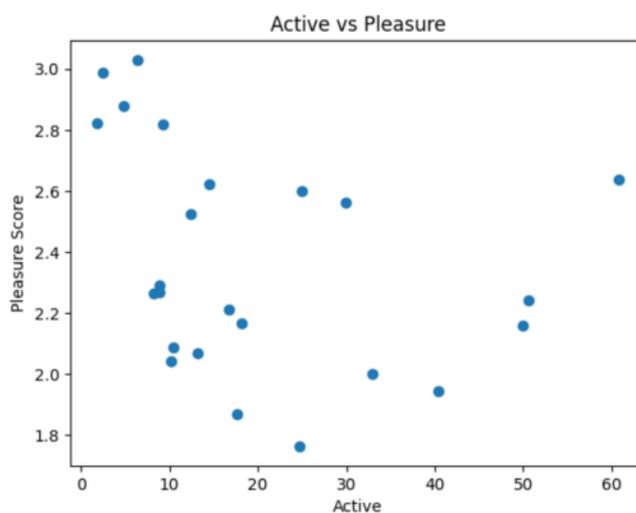


Figure 5. Scatter plot of stimulation ratio versus pleasure score.

4.5. Analysis of the Average Emotional Level

Using the stacked bar chart of average proportions across emotion levels, the composition of emotional responses to color samples can be visually examined. Overall, neutral and mildly pleasant emotions account for a relatively large proportion, while extreme emotion levels account for a relatively small proportion. This suggests that color emotion perception often changes gradually rather than exhibiting polarization. Meanwhile, the coexistence of different emotion-level proportions supports the use of a

weighted approach to construct a continuous pleasure score, which can more comprehensively reflect the emotional experience elicited by colors (Figure 6).

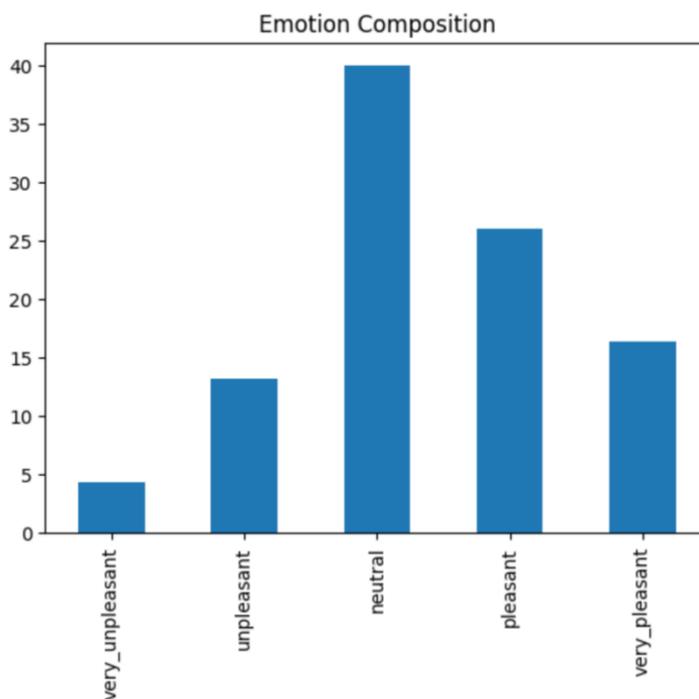


Figure 6. Distribution of average emotional levels.

5. Modeling Analysis

5.1. Regression Analysis

5.1.1. Feature Extraction

In machine learning-based emotion perception modeling, converting color images into numerical features is a key factor affecting model performance. Since this paper focuses on single-color images and emotion perception mainly arises from the color itself rather than complex spatial structures, a representation based on color statistical features is adopted to simplify the description of color information. This approach preserves key perceptual characteristics while reducing feature dimensionality, making it suitable for modeling with a small sample size.

For color feature representation, this paper does not use the RGB color space. Instead, images are converted to the HSV color space for feature extraction. From the perspective of human visual perception, HSV decomposes color into hue, saturation, and value (brightness), which aligns better with intuitive color perception. Hue refers to the category of color, saturation describes purity and vividness, and value describes perceived brightness. The HSV color space offers good interpretability and practicality for modeling color emotion perception (Figure 7).

| | H_mean | H_std | S_mean | S_std | V_mean | V_std |
|---|-----------|-----------|------------|-----------|------------|----------|
| 0 | 90.559644 | 86.404755 | 102.030601 | 97.375236 | 251.860714 | 2.996116 |
| 1 | 9.945427 | 9.489152 | 102.031148 | 97.375700 | 251.860631 | 2.996161 |
| 2 | 16.227224 | 15.482725 | 85.811198 | 81.895712 | 253.429877 | 1.498499 |
| 3 | 26.695680 | 25.470900 | 84.240779 | 80.397174 | 252.906557 | 1.997887 |
| 4 | 56.010794 | 53.441012 | 103.600519 | 98.873470 | 252.906629 | 1.997974 |
| 5 | 71.714816 | 68.424531 | 102.553789 | 97.874880 | 250.290276 | 4.494805 |
| 6 | 91.083491 | 86.904603 | 46.044298 | 43.943475 | 251.860714 | 2.996116 |
| 7 | 6.804261 | 6.492159 | 38.720149 | 36.953324 | 251.860631 | 2.996161 |
| 8 | 15.704153 | 14.983666 | 41.859035 | 39.949022 | 253.429980 | 1.498475 |

Figure 7. Example of feature extraction.

5.1.2. Comparison of Basic Models

To examine the modeling capability of different baseline models for predicting pleasure scores, this paper compares several commonly used regression models, including linear regression, decision trees, fully connected neural networks, random forests, and k-nearest neighbor (KNN). Mean squared error (MSE), root mean squared error (RMSE), and the coefficient of determination (R^2) are used as evaluation metrics. The experimental results for each model are shown in the table below.

Overall, model performance varies substantially across this task. Linear regression performs the worst: RMSE and MSE are much higher than those of other models, and the R^2 value is negative, indicating that linear regression fails to fit the relationship between color features and pleasure scores. This suggests that the mapping between color features and emotion perception is highly nonlinear.

The decision tree model shows some improvement, with lower error compared with linear regression. However, the R^2 value remains low or even negative. With a small sample size, a single decision tree is prone to overfitting or limited generalization. The fully connected neural network further reduces prediction error and yields a positive R^2 value, indicating that neural networks can better capture nonlinear relationships between color features and emotion, although there is still room for improvement.

Among ensemble methods, the random forest model achieves relatively good and stable results, with further reductions in RMSE and MSE and an R^2 of 0.406. This indicates that aggregating multiple decision trees can improve generalization and reduce instability caused by a single model. The KNN model achieves the best prediction results in this experiment: both RMSE and MSE are the smallest among all models, and R^2 is the highest. This suggests that, in the current feature space, similar color samples exhibit strong consistency in emotion perception, and distance-based local modeling can capture this relationship effectively.

In summary, for pleasure score prediction based on HSV statistical features, nonlinear models generally outperform linear models. Models based on sample similarity and ensemble learning show better performance, supporting the conclusion that the relationship between color features and emotion perception is complex and nonlinear. These results provide a baseline for subsequent end-to-end CNN experiments (Figure 8).

| | Model | RMSE | MSE | R2 |
|---|-------------------|----------|-----------|------------|
| 3 | Linear Regression | 3.319564 | 11.019507 | -18.259145 |
| 0 | Decision Tree | 0.798614 | 0.637784 | -0.114676 |
| 1 | FC Neural Network | 0.634787 | 0.402954 | 0.295745 |
| 2 | Random Forest | 0.582820 | 0.339679 | 0.406332 |
| 4 | KNN | 0.571030 | 0.326076 | 0.430107 |

Figure 8. Comparison of model performance metrics.

5.1.3. Explanation of Feature Importance

To further analyze the contribution of different color features to pleasure score prediction, this paper ranks the importance of input features after model training. The feature importance results are shown in the table, which includes the mean and standard deviation of each channel in the HSV color space. Feature importance analysis helps clarify how different color attributes contribute to emotion perception modeling.

Overall, features from the value (brightness) channel contribute the most. In particular, the standard deviation of value (V_std) has the highest importance (0.459), indicating that variation in brightness distribution is crucial for perceiving pleasant emotions. For single-color images, the standard deviation of brightness can reflect the uniformity and stability of brightness within the image, suggesting that brightness variation strongly influences emotional perception. Colors with relatively stable or moderate brightness variation may be more likely to evoke pleasant feelings.

The mean value (V_mean) is also highly important (0.310), further confirming the central role of brightness in color emotion perception. This implies that overall brightness is a key factor affecting the pleasure score. Higher or moderate brightness often corresponds to more positive emotional experiences, consistent with findings in psychological research that brighter colors tend to evoke more positive emotions.

Regarding saturation, the mean saturation (S_mean) shows moderate importance, suggesting that vividness also affects perceived pleasure. Higher saturation corresponds to more vivid and impactful colors, potentially increasing emotional arousal and enhancing pleasure perception. In contrast, the standard deviation of saturation (S_std) has relatively low importance, indicating that spatial variation in saturation is less influential than the overall saturation level.

Compared with brightness and saturation, hue-related features (H_mean and H_std) are generally less important, with H_std showing the lowest importance. This suggests that for single-color images, hue category plays a weaker role in pleasure prediction than brightness and saturation. In other words, how bright and vivid a color is may matter more for pleasure perception than which hue it belongs to.

In summary, feature importance analysis indicates that brightness and saturation play dominant roles in pleasant emotion perception modeling, while hue contributes relatively less in the single-color setting. These findings provide a reasonable explanation for the model results and also offer guidance for emotion-oriented color design: attention should be paid to the combination of brightness and saturation rather than focusing solely on hue selection (Figure 9).

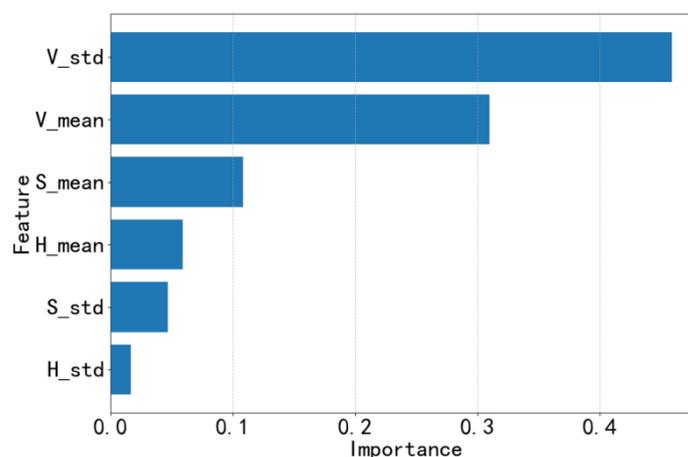


Figure 9. Feature importance map.

5.1.4. Deep Learning Regression Analysis

To further enhance the accuracy of color pleasure score prediction, this paper proposes a CNN-based regression model, SimpleCNN. The structure of this model is relatively simple. The model aims to extract features from single-color images and build a deep model. It consists of three convolutional blocks, each comprising a convolutional layer, a ReLU activation function, and a pooling layer. Features are extracted from spatial information and color distribution patterns across the three channels. The convolutional features are then passed through adaptive average pooling and a fully connected layer to produce an output representing the continuous pleasure score, enabling an end-to-end regression process. Compared with traditional regression methods that rely on handcrafted statistical features, CNN models directly learn color distributions from image pixels and can capture richer nonlinear relationships.

Model training uses mean squared error (MSE) as the loss function. In each epoch, the loss values and R^2 values on both the training set and the test set are recorded. The training process shows that R^2 is negative in the first few epochs, indicating that the network fails to fit the relationship between color and pleasure score at the beginning. However, as training progresses, model performance on both the training set and the test set improves significantly. For example, after the 10th epoch, the R^2 value reaches 0.7315 on the training set and 0.9095 on the test set, suggesting that the model begins to capture the nonlinear mapping between color features and pleasure perception.

After training for 30 epochs, the R^2 values on the training set and the test set reach 0.9532 and 0.9511, respectively, and the test-set MSE decreases to 0.0073, demonstrating strong predictive accuracy and good generalization. The overall training curves show stable convergence. Both training and test losses decrease substantially, indicating that SimpleCNN can effectively learn feature information from color images and achieve accurate prediction of pleasure scores.

Overall, the CNN-based regression model demonstrates advantages over traditional machine learning models in this task and can better learn complex nonlinear relationships between color features and emotion perception (Figure 10).

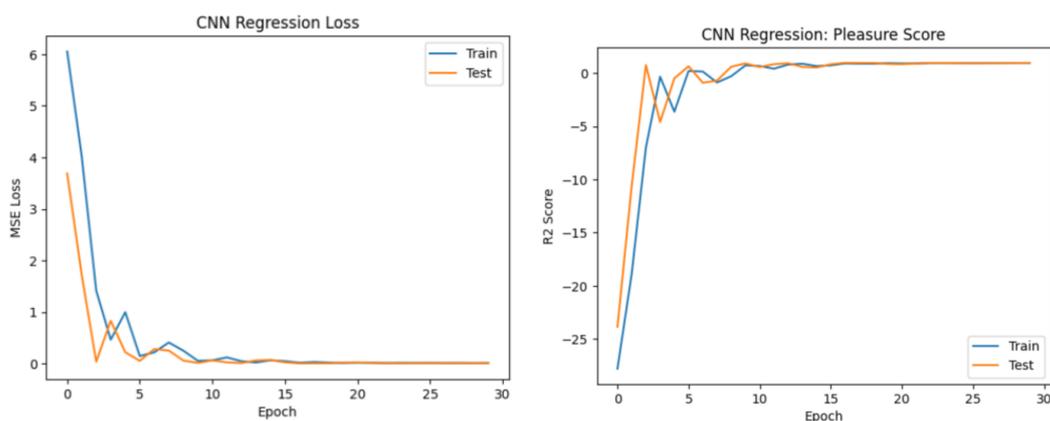


Figure 10. Training curves of the deep learning regression model.

5.2. Classification Analysis

For the binary classification task of whether a color is "stimulating," this paper selects three classic machine learning models—logistic regression, decision tree, and random forest—for baseline classification modeling. The classification target is binarized based on the sample's stimulation proportion, dividing samples into two categories: "low stimulation" (category 0) and "high stimulation" (category 1). This transforms subjective psychological feelings into numerical labels that can be used for machine learning modeling, quantifying, to some extent, the impact of color on the arousal dimension and providing clear and interpretable target variables for model training.

For model evaluation, this paper uses precision, recall, F1-score, and overall accuracy to assess performance. Experimental results show that all three baseline models achieve relatively good classification performance under a small-sample setting. For example, logistic regression achieves an overall accuracy of 0.89. The precision for category 0 (low stimulation) is 1.00, but recall is only 0.67, indicating missed detections when identifying low-stimulation samples. For category 1 (high stimulation), precision is 0.86 and recall reaches 1.00, indicating that the model can identify high-stimulation samples comprehensively. Decision trees and random forests show performance very close to logistic regression, with accuracies of 0.89, macro F1 of 0.86, and weighted F1 of 0.88, suggesting relatively balanced performance across categories.

From the category-level results, the recall of the three baseline models for the minority class (category 0) is relatively low, indicating that minority-class recognition remains challenging with small samples. Future work may consider strategies such as class balancing and weighted loss to improve minority-class recall. Overall, logistic regression, decision trees, and random forests can effectively distinguish high-stimulation and low-stimulation color samples, providing a robust baseline for stimulation perception classification and supporting subsequent deep learning classification experiments (Figure 11).

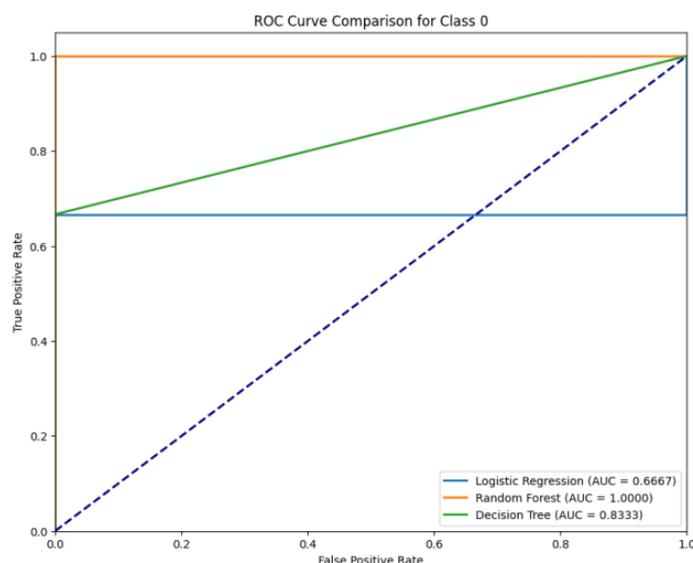


Figure 11. Comparison of machine learning classification results.

In the binary classification task of stimulation perception, this paper further builds a CNN-based deep learning model to automatically extract color features at the pixel level and perform end-to-end classification. The model uses multiple convolutional layers and activation functions to progressively identify local patterns and spatial distribution characteristics of colors and outputs a binary classification result through fully connected layers. Compared with baseline machine learning models, CNNs can learn complex nonlinear relationships directly from image data, potentially improving classification performance and generalization.

During training, both loss and accuracy change significantly across epochs. In the early stage, training and test accuracies are 57.1% and 55.6%, respectively, with relatively high loss values, indicating that the model has not yet learned the relationship between color and stimulation perception. As training continues, training accuracy gradually increases. By epoch 16, training accuracy reaches 100%, and test accuracy also reaches 100% at that time. Test loss decreases to 0.0757, indicating that the model fits the data well and achieves accurate predictions. In subsequent epochs, both training and test accuracies

remain stable at 1.0, and loss stays low (approximately 0.045-0.050), suggesting that the model converges and maintains good generalization on the test set.

Overall, the deep learning model achieves better performance than the baseline machine learning models in this stimulation classification task. The CNN can automatically identify key features of color images and explore nonlinear relationships between samples, enabling high-accuracy and stable classification. These results indicate that deep learning methods have clear advantages in modeling color emotion perception and are suitable for emotion recognition tasks involving multi-dimensional and complex feature patterns (Figure 12).

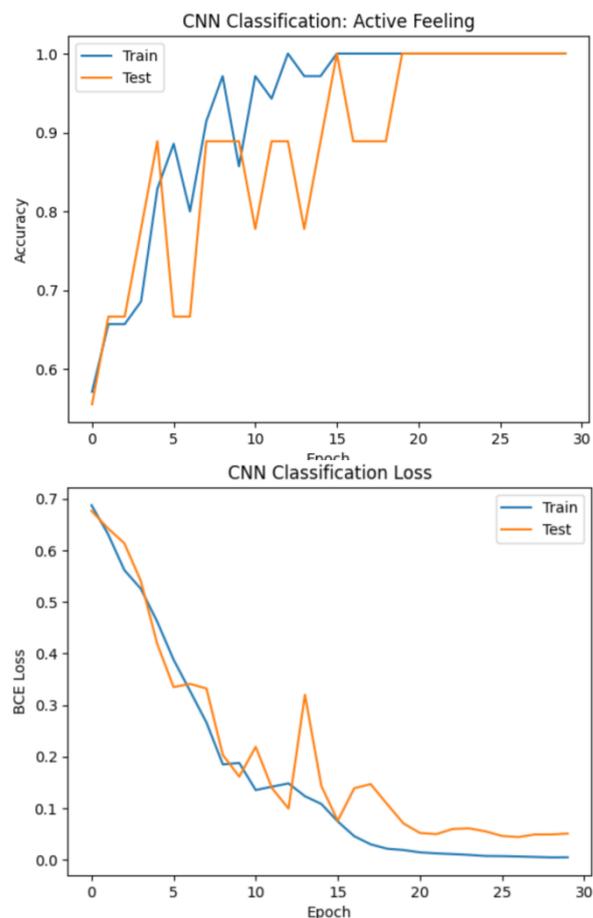


Figure 12. Training curves of the deep learning classification model.

5.3. Visualization Analysis of Deep Learning Feature Extraction

To enable readers to intuitively observe the feature extraction capability of the deep learning model for color-based emotion perception, this paper visualizes feature responses from intermediate convolutional layers. Several single-color image samples are randomly selected from the training set, and their feature maps after multi-layer convolution are examined. The feature maps differ substantially across samples.

For images with high brightness and saturation, the convolutional feature maps show strong responses in channels related to brightness and color intensity, and some channels highlight the brightness distribution and color contrast patterns in the image. In contrast, for dark or low-saturation images, feature responses mainly concentrate on subtle local brightness variations, with fewer highly activated regions in the feature maps. This indicates that the model can adapt its feature representations to differences in color attributes and capture subtle variations in brightness, saturation, and local distribution patterns.

Different feature-map channels reflect different decompositions of color information. Different convolutional kernels respond differently to the input image and can extract features at multiple levels, such as brightness cues, saturation changes, and spatial response patterns. Such multi-channel and multi-level feature representations are not available in traditional methods that rely solely on statistical features. Moreover, feature-map visualization suggests that the model can capture complex nonlinear relationships and provides distinctive representations for different color samples.

Overall, the deep learning model can automatically extract rich color-related features and generate differentiated representations across samples, providing strong feature support for pleasure score prediction and stimulation classification and demonstrating the advantages of deep learning for color emotion perception (Figure 13).

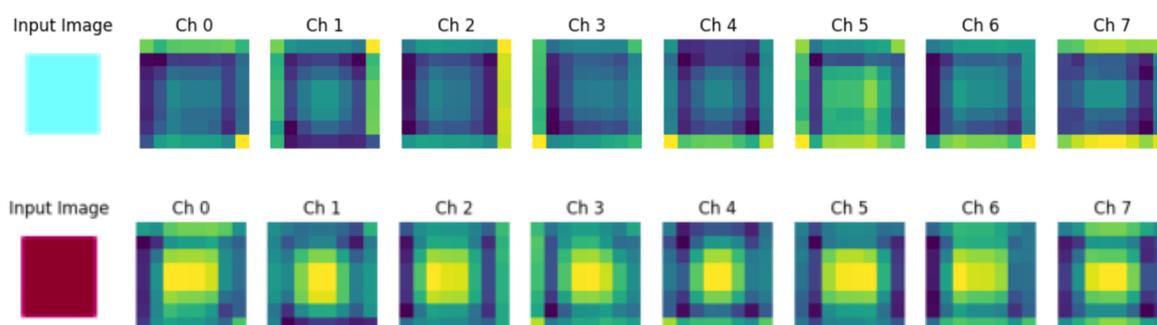


Figure 13. Visualization of deep learning feature extraction.

5.4. Cluster Analysis

5.4.1. Determining the Number of Clusters

To identify meaningful emotion-related clusters of color samples, this paper applies the K-means algorithm for unsupervised clustering. To determine an appropriate number of clusters (K), experiments are conducted with K ranging from 2 to 9, and silhouette scores are computed to evaluate intra-cluster compactness and inter-cluster separation.

The results show that the silhouette score generally increases as K increases, rising from 0.283 at K=2 to 0.506 at K=7, indicating improved clustering separation and compactness. When K=8, the silhouette score slightly decreases to 0.497, and it remains at 0.497 at K=9. Although K=7 yields the highest silhouette score, this paper selects K=8 as the final number of clusters to maintain consistency with the subsequent cluster interpretation (Clusters 0-7) and to better support semantic interpretability of the resulting groups. With K=8, clustering still achieves a clear separation among clusters while enabling more fine-grained and explainable emotional grouping of color samples.

These clustering results reflect different distributions of color samples in terms of pleasure and stimulation and provide a basis for mapping clusters to emotional semantics, offering data-driven support for subsequent color emotion analysis and application-oriented design (Figure 14).

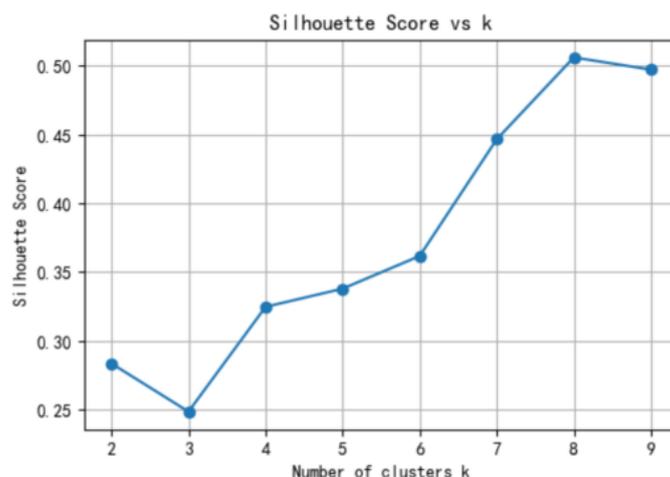


Figure 14. Silhouette score curve for different values of K.

5.4.2. Interpretability of Clustering Results

In the clustering analysis, this paper provides semantic interpretations of different clusters based on the emotion proportion characteristics within each cluster, enabling the data-driven clustering results to correspond to understandable emotional semantics. The interpretations are as follows:

Cluster 0 - Calm and Pleasant

The proportion of calm responses is very high (about 94%). The proportions of pleasant and very pleasant responses are also high (about 58%), while the proportion of stimulating responses is very low (about 6%). Overall, this cluster corresponds to soothing and gentle colors that feel visually comfortable and slightly pleasant.

Cluster 1 - Stimulating and Active

The stimulation proportion is relatively high (about 30%), the unpleasant proportion is relatively high (about 20%), calmness is moderate (about 70%), and pleasure is relatively low (about 28%). This cluster tends to evoke vitality and stimulation, with strong visual impact and a dynamic feeling.

Cluster 2 - Calm and Neutral

Calmness is high (about 90%), the general/neutral proportion is the highest (about 47%), pleasure is moderate (about 21%), and stimulation is low (about 10%). This cluster is dominated by soft and neutral colors, providing a visually comfortable and non-irritating experience, with an overall stable impression.

Cluster 3 - Calm and Pleasant

Calmness is high (about 90%), pleasure is relatively high (about 25%), and stimulation is low (about 10%). Overall, it is similar to Cluster 0 but with a slightly higher pleasure level. The colors appear soft and evoke a comfortable and pleasant feeling.

Cluster 4 - Vitality and Pleasure

Pleasure is high (about 48%), stimulation is relatively high (about 30%), and calmness is moderate (about 70%). Colors in this cluster are bright and lively, while still retaining a degree of calmness; vitality and calmness coexist.

Cluster 5 - High Pleasure and Calm

Calmness is extremely high (about 94%), pleasure is high (about 68%), and stimulation and unpleasant proportions are extremely low (about 6% and 4%, respectively). Colors in this cluster are very comfortable and pleasant, suitable for warm and soft design styles.

Cluster 6 - Strong Stimulation

Stimulation is high (about 55%), pleasure is moderate (about 45%), and calmness is relatively low (about 45%). This cluster is characterized by high saturation and strong visual impact, suitable for eye-catching or emphasis-oriented designs.

Cluster 7 - Calm Neutral with Mild Stimulation

Calmness is high (about 80%), stimulation is moderate (about 20%), the general/neutral proportion is high (about 47%), and pleasure is relatively low (about 16%). The overall impression is soft but includes a hint of stimulation, producing a color style that is stable yet energetic.

From these semantic interpretations, clear distribution patterns emerge in the emotion dimensions:

- 1) High calmness & low stimulation → calming clusters (Clusters 0, 2, 3, and 5). These clusters contain soft and soothing colors that are visually comfortable, with pleasure levels ranging from moderate to high.
- 2) High stimulation → active or strong-stimulation clusters (Clusters 1, 4, and 6). These clusters contain colors with strong visual impact and dynamism, suitable for energetic or emphasis-driven effects.
- 3) High general/neutral proportion → neutral-oriented clusters (Clusters 2 and 7). These clusters contain mild and stable colors that are visually comfortable, with moderate stimulation.
- 4) High pleasure proportion → pleasant clusters (Clusters 0, 3, 4, and 5). These clusters tend to evoke positive and relaxed emotional experiences.

Overall, clustering not only groups color samples in an emotion-feature space but also provides a quantitative basis for interpreting emotional semantics. It supports subsequent integration of clustering results with regression and classification models for color emotion perception modeling and application (Figure 15).

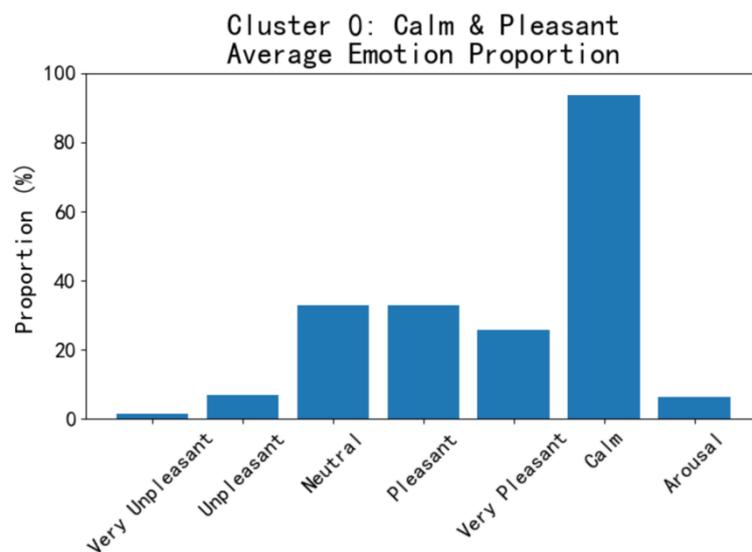


Figure 15. Visualization of clustering categories.

6. Summary and Prospect

This paper investigates a color emotion perception method integrating unsupervised clustering and deep learning and constructs a data-driven framework for color emotion analysis from four aspects: color feature extraction, regression and classification modeling, deep learning feature visualization, and clustering with semantic interpretation. In the regression analysis, statistical features in the HSV color space are used to represent single-color images. The results show that brightness and saturation features play important roles in predicting pleasure, while hue features have a relatively smaller influence. Comparative experiments with multiple baseline regression models show that nonlinear models-especially KNN and ensemble learning models based on sample similarity-perform well in predicting pleasure scores, providing a useful baseline for subsequent deep learning models.

For classification, baseline models trained with binary labels for stimulation perception, such as logistic regression, decision trees, and random forests, achieve high accuracy and F1-scores. This indicates that stimulation perception in the arousal dimension is also predictable. The deep learning model further demonstrates strong feature extraction capability. The convolutional network automatically extracts features related to brightness, saturation, and local response patterns, improving predictive performance during training. Feature-map visualization from intermediate layers shows clear differences in feature responses across different colors, indicating that deep learning models can learn discriminative representations for different color samples.

In clustering analysis, the combination of unsupervised K-means grouping and emotion proportion statistics enables semantic interpretation of color clusters. By analyzing proportions of pleasure, calmness, stimulation, and neutrality within each cluster, this paper assigns interpretable names such as "Calm and Pleasant," "Stimulating and Active," and "Vitality and Pleasure." Clear distribution patterns emerge: clusters with high calmness and low stimulation show soothing characteristics; clusters with high stimulation are more active or emphasis-oriented; and clusters with high pleasure tend to evoke positive and relaxed feelings. These clustering results provide data-driven support for emotion-oriented color design.

Overall, the proposed method is interpretable and practical for modeling color emotion perception and provides a foundation for combining clustering with supervised modeling. In future work, several improvements can be considered: (1) adopting more advanced end-to-end architectures, such as deeper CNNs or attention-based models, to capture more complex nonlinear relationships and local color patterns; (2) enriching the dataset with multi-color combinations and more diverse scenarios to explore interactions between colors and emotions; and (3) applying clustering semantics and prediction models to real design and human-computer interaction applications to support color scheme optimization and improve visual experience, thereby strengthening the connection between theory and practice.

References

1. M. Mazzone, and A. Elgammal, "Art, creativity, and the potential of artificial intelligence," In *Arts*, February, 2019, p. 26. doi: 10.3390/arts8010026
2. R. Zhang, "Development of environmental art design system under the background of artificial intelligence," In *Journal of Physics: Conference Series*, February, 2021, p. 042224. doi: 10.1088/1742-6596/1744/4/042224
3. W. Liu, R. Gao, and C. Fang, "Application and Optimization of Shanxi Fahua Art Elements Empowered by Artificial Intelligence in Clothing Design," In *International Conference on Human-Computer Interaction*, May, 2025, pp. 235-251. doi: 10.1007/978-3-031-93227-4_17
4. Y. Hu, S. Wu, Z. Ma, and S. Cheng, "Integrating deep learning and machine learning for ceramic artifact classification and market value prediction," *npj Heritage Science*, vol. 13, no. 1, p. 306, 2025. doi: 10.1038/s40494-025-01886-6
5. X. Li, "Innovative Research on Artificial Intelligence-Driven Art Design Education Courses," .
6. S. K. Gupta, A. Alemran, U. S. Basha, A. I. Zakari, S. Kim, R. S. K. Boddu, and S. K. Vohra, "Revolutionizing the way students learn photographic arts through experiential education using AI and AR systems," *Scientific Reports*, vol. 15, no. 1, p. 40705, 2025. doi: 10.1038/s41598-025-24415-8
7. G. Ganesh, S. Ambekar, A. K. Singh, A. K. Das, and M. Sanchihar, "Investigating the Transformative Role of Artificial Intelligence in Enhancing Creativity and Efficiency Across the Graphic Design Process," In *2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS)*, August, 2025, pp. 1-6. doi: 10.1109/worldsuas66815.2025.11199047
8. M. Muratbekova, and P. Shamoj, "Color-emotion associations in art: Fuzzy approach," *IEEE Access*, vol. 12, pp. 37937-37956, 2024. doi: 10.1109/access.2024.3375361
9. A. Dembe, "Advancing personalized learning through educational artificial intelligence: Challenges, opportunities, and future directions," *Res Invent J Eng Phys Sci*, vol. 3, no. 1, pp. 89-101, 2024.
10. J. Liang, "Enhanced art design education: Integrating artificial intelligence for resource optimization and innovative teaching," In *2024 4th International Conference on Internet Technology and Educational Informatization (ITEI 2024)*, November, 2024, pp. 239-247. doi: 10.2991/978-94-6463-560-7_29
11. L. Wei, "Exploring Innovative Pathways of Artificial Intelligence Empowering Art and Design Education," In *Proceedings of the 2024 10th International Conference on Frontiers of Educational Technologies*, June, 2024, pp. 32-37. doi: 10.1145/3678392.3678417

12. B. Zhou, "Application and Research of AIGC in Contemporary Art Design Curriculum," *Journal of Education, Humanities and Social Sciences*, vol. 41, pp. 67-70, 2024. doi: 10.54097/3nqmm282
13. J. Chen, "Digital Art: A Study on the Application Paths of Artificial Intelligence in Artistic Creation and Dissemination," *thinking*, vol. 7, no. 5, pp. 50-59, 2025.
14. X. Zheng, D. Bassir, Y. Yang, and Z. Zhou, "Intelligent art: The fusion growth of artificial intelligence in art and design," *International Journal for Simulation and Multidisciplinary Design Optimization*, vol. 13, p. 24, 2022. doi: 10.1051/smdo/2022015
15. Y. Sun, "Application of artificial intelligence in the cultivation of art design professionals," *International Journal of Emerging Technologies in Learning (ijET)*, vol. 16, no. 8, pp. 221-237, 2021.
16. Y. Luo, "Research on computer intelligent image recognition technology," In *Proceedings of the 7th International Conference on Cyber Security and Information Engineering*, September, 2022, pp. 735-737. doi: 10.1145/3558819.3565181

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of SOAP and/or the editor(s). SOAP and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.