IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 2, April, 2026, pp. 585-598
Journal homepage: http://iieta.org/journals/ts

Image Color Harmonization Modeling and Perception-Consistent Optimization in Visual Art Design

Changzheng Wang

Shangqiu Polytechnic, Shangqiu 476000, China

Corresponding Author Email: wcz333333@126.com

https://doi.org/10.18280/ts.430203

Received: 9 November 2025
Revised: 27 February 2026
Accepted: 12 March 2026
Available online: 30 April 2026

Keywords:
color harmonization, perceptual consistency, CIE Color Appearance Model 2002 color appearance model, unsupervised optimization, visual art design

ABSTRACT

Color harmonization in visual art design images is regarded as a critical link between aesthetic expression and human visual perception, requiring both aesthetic regularity and perceptual naturalness. Existing methods are constrained by rigid global rules, perceptual inconsistencies, and dependence on paired ground-truth data, limiting their ability to achieve region-specific harmonization across diverse semantic contexts and restricting practical applicability. A perception-consistent color harmonization optimization framework was proposed for visual art design images. First, a spatially weighted adaptive harmonization energy model was constructed by integrating semantic-guided saliency detection, in which pixel-level attention weights were incorporated into harmonization energy terms to enable region-aware color adjustment. This formulation effectively mitigates over-harmonization and preserves visual diversity. Second, the CIE Color Appearance Model 2002 (CIECAM02) color appearance model was explicitly embedded into the optimization process. A perceptual deviation field constrained by the Weber-Fechner law and a spatially varying viewing condition field were jointly formulated, leading to a substantial improvement in perceptual fidelity aligned with human visual responses. Finally, a lightweight differentiable color mapping network was developed within an unsupervised joint optimization framework. A dedicated perceptual color harmonization index was further introduced, enabling the end-to-end co-optimization of harmonization energy, perceptual consistency, and structural preservation without reliance on paired training data. Extensive experiments and subjective evaluations were conducted on datasets. The results demonstrated that the proposed method consistently outperformed state-of-the-art approaches in both harmonization quality and perceptual naturalness, achieving superior performance in terms of the proposed perceptual color harmonization index and mean opinion scores, thereby validating its effectiveness and robustness.

1. INTRODUCTION

Visual art design images are widely recognized as essential carriers of aesthetic expression and information transmission [1, 2]. The visual impact of design forms, including posters, illustrations, and photographic works, is largely determined by the rationality and naturalness of color harmonization. As a core component of visual art design [3-5], color harmonization is fundamentally concerned with achieving a balance between the controllability of harmonization rules and the naturalness of visual perception. On the one hand, adherence to established principles of color aesthetics is required to ensure the coherence and regularity of color combinations, thereby fulfilling the aesthetic intent of design works. On the other hand, alignment with human visual perception characteristics must be maintained to avoid perceptual distortions introduced during color adjustment, ensuring visual comfort and perceptual naturalness. With the rapid advancement of digital design technologies, visual art design scenarios have become increasingly diverse, imposing higher demands on the precision, adaptability, and efficiency of color harmonization techniques [6-8]. Consequently, the development of color harmonization algorithms capable of accommodating diverse design contexts while simultaneously integrating aesthetic principles and perceptual consistency has emerged as a critical research focus in the interdisciplinary domain of image processing and visual design, holding substantial theoretical significance and practical value [9].

Despite notable progress in existing color harmonization methods [10, 11], several fundamental limitations remain when addressing the practical requirements of visual art design, thereby hindering the effective integration of harmonization quality and perceptual naturalness. Traditional color harmonization models predominantly rely on globally fixed harmonization rules [12-14], where harmonization is formulated solely based on numerical color relationships. Such approaches fail to account for the heterogeneous semantic characteristics of different regions within design images, resulting in a lack of adaptability and the frequent emergence of visually rigid effects. Consequently, region-specific harmonization requirements, such as those between foreground and background or subject and environment,

cannot be adequately satisfied. In terms of perceptual consistency constraints, most existing methods optimize color differences within conventional color spaces such as the Hue-Saturation-Value (HSV) color space or the CIE L*a*b* (CIELAB) color space [15, 16], without adequately incorporating the visual adaptation characteristics and perceptual mechanisms of the human visual system. As a result, these methods are unable to accurately align with human perceptual responses to lightness, chroma, and hue, often leading to perceptual distortions such as luminance inversion, color oversaturation, or undersaturation, thereby degrading the overall visual experience of the designed artwork. Although deep learning-based color harmonization approaches have demonstrated improvements in harmonization accuracy [17, 18], they are typically dependent on paired ground-truth data for supervised training. This reliance not only increases the cost of data annotation but also restricts generalization capability, limiting applicability to real-world, unlabeled design scenarios. Furthermore, such approaches generally fail to achieve the joint optimization of color harmonization and perceptual consistency constraints, making it difficult to simultaneously ensure harmonization effectiveness and perceptual naturalness. In addition, a lack of dedicated evaluation metrics tailored to visual art design images remains evident. Existing metrics primarily emphasize objective image fidelity [19, 20], while the balance between aesthetic quality and perceptual consistency in color harmonization is insufficiently quantified. This limitation prevents an objective assessment of algorithmic performance in practical visual art design applications.

In response to the aforementioned limitations in existing studies, the objective is to develop an unsupervised, end-to-end joint optimization algorithm for color harmonization and perceptual consistency. The proposed framework is designed to effectively accommodate the diverse requirements of visual art design images, enabling simultaneous enhancement of aesthetic controllability and perceptual naturalness without reliance on paired ground-truth data. To achieve this objective, four primary contributions are established. First, semantic-guided saliency detection is integrated with a spatial attention mechanism to construct a spatially weighted adaptive harmonization energy function. By introducing pixel-level attention weights and inter-region smoothness constraints, region-aware color harmonization between foreground and background is realized, thereby effectively mitigating the rigidity associated with traditional global harmonization models. Second, the CIE Color Appearance Model 2002 (CIECAM02) color appearance model is incorporated explicitly into the color harmonization optimization framework. A perceptual deviation field defined in terms of lightness, chroma, and hue is formulated, within which constraints derived from the Weber–Fechner law are embedded. In conjunction with a spatially varying viewing condition field, differentiated perceptual constraints are imposed across distinct image regions. This formulation replaces conventional Euclidean distance-based metrics, thereby fundamentally enhancing perceptual fidelity in accordance with human visual response characteristics. Third, a lightweight differentiable color mapping network is designed, consisting of a three-layer convolutional architecture and two adaptive instance normalization modules. A joint loss function is constructed by integrating a harmonization loss, a perceptual consistency loss, and a structural preservation regularization term. Through this formulation, end-to-end unsupervised co-optimization of harmonization energy, perceptual consistency, and image structural integrity is achieved. Finally, a dedicated perceptual color harmonization index is proposed, in which key factors-including the degree of color harmonization, perceptual consistency, and artifact detection-are jointly considered. This index addresses the limitations of existing evaluation metrics that fail to balance aesthetic quality and perceptual naturalness, thereby providing a reliable and objective basis for performance assessment.

The remainder of the study is organized below. In Section 2, related work on color harmonization, perceptual consistency constraints, and unsupervised color optimization is systematically reviewed, through which the limitations of existing approaches are identified and the research motivation is clearly established. In Section 3, the overall framework of the proposed joint optimization algorithm for color harmonization and perceptual consistency is described in detail, including the design principles of each module and the key technical components. In Section 4, extensive experimental evaluations are conducted, including comparative experiments, ablation studies, generalization assessments, and subjective evaluation experiments, through which the effectiveness and superiority of the proposed method are comprehensively validated. In Section 5, the experimental results are further analyzed in depth, the limitations of the proposed approach are critically discussed, and potential directions for future research are outlined. Finally, in Section 6, the principal contributions, core findings, and potential application prospects are summarized, and the overall conclusions are distilled. A coherent and logically consistent structure is maintained across all sections, with particular emphasis placed on rigorous experimental validation to ensure the completeness and scientific soundness of the study.

2. PROPOSED METHOD

2.1 Overall framework and problem formulation

This section provides a detailed description of the overall framework of the proposed joint optimization algorithm for color harmonization and perceptual consistency, together with a clear formulation of the associated problem. The input to the algorithm consists of an original visual art design image I and a predefined harmonization mode, while the output is a perceptually optimized and color-harmonized image I'. The overall framework is constructed through the sequential integration of four core modules: a semantic-guided saliency detection and pixel-level attention weight computation module; a spatially weighted adaptive harmonization energy modeling module; a perceptual consistency constraint module based on the CIECAM02 color appearance model; and a lightweight differentiable color mapping network combined with an unsupervised joint optimization module. Through the coordinated operation of these modules, an end-to-end processing pipeline is established, encompassing semantic feature extraction from the input image, harmonization energy construction, perceptual constraint embedding, and final color optimization. A key advantage of this framework lies in its ability to achieve unsupervised training without reliance on paired ground-truth data, thereby substantially reducing annotation costs while improving generalization capability.

Existing color harmonization frameworks generally fail to achieve an effective integration of semantic-adaptive harmonization, perceptual consistency constraints, and end-to-end optimization. As a consequence, harmonization quality, perceptual naturalness, and training efficiency cannot be simultaneously satisfied. The proposed framework addresses this limitation by enabling the coordinated design and joint optimization of the four modules, thereby achieving a unified enhancement of harmonization effectiveness, perceptual fidelity, and computational efficiency. This design provides an efficient, perceptually natural, and controllable solution for color harmonization in visual art design images. The overall framework is illustrated in Figure 1.

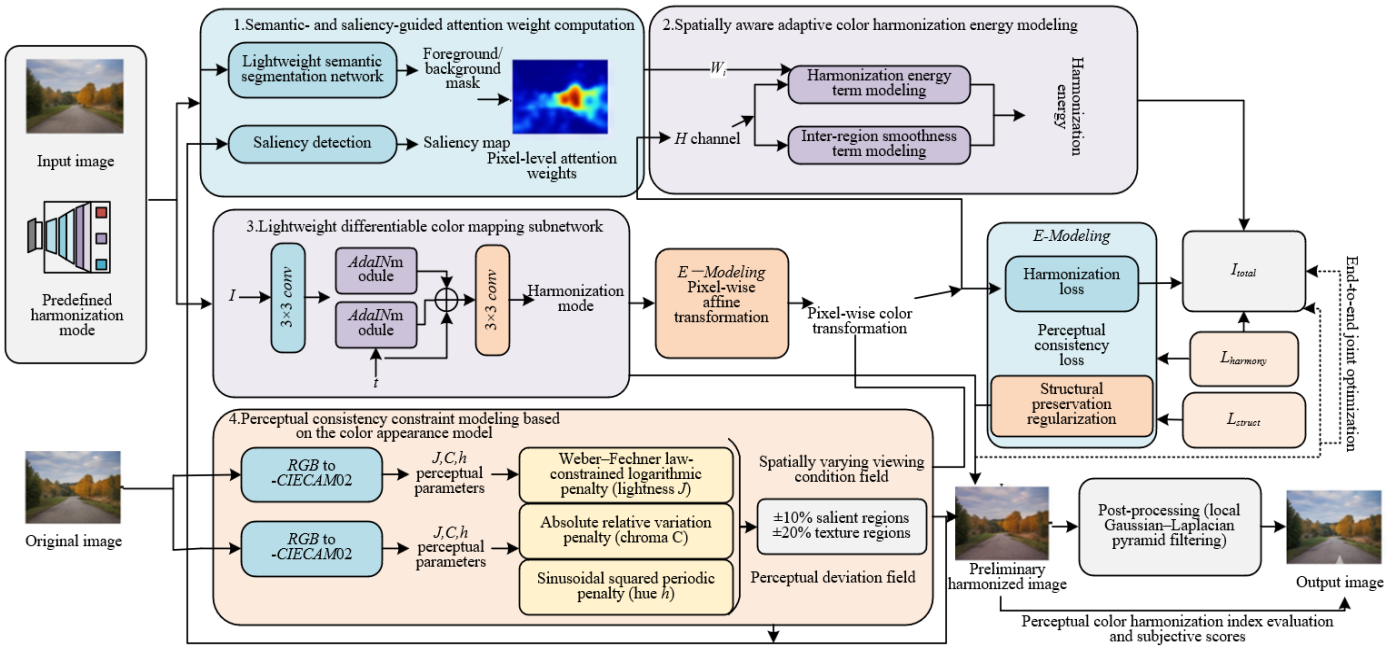

Figure 1. Overall framework of the proposed algorithm

2.2 Spatially perception-driven color harmonization modeling

The core objective of spatially perception-driven color harmonization modeling is to achieve region-aware, differentiated harmonization across semantic regions, thereby overcoming the limitations of conventional global harmonization rules. The initial step involves the joint extraction of semantic and saliency features, followed by the construction of pixel-level attention weights. The schematic illustration of the spatially perception-driven attention weighting and harmonization modeling process is presented in Figure 2. A lightweight semantic segmentation network is employed to extract foreground and background masks from the original design image, denoted as M_f and M_b, respectively. Owing to its efficient feature extraction capability, accurate semantic segmentation is maintained while computational complexity is significantly reduced. It should be noted that both the semantic segmentation network, i.e., Bilateral Segmentation Network V2 (BiSeNetV2), and the saliency detection network, Boundary-Aware Salient Object Detection Network (BASNet), are employed with fixed parameters and are not involved in the joint optimization of the subsequent color mapping network. The two networks serve solely as preprocessing modules to provide pixel-level attention weights w_i and do not participate in gradient backpropagation. This design is adopted to ensure training stability and computational efficiency while avoiding additional annotation costs. On this basis, attention weights w_i are computed by integrating semantic region characteristics with pixel-level saliency information. The formulation is given as $w_i = \alpha \cdot Sal_i + \beta \cdot \delta(region_i)$, where Sal_i represents the saliency value of the i-th pixel, and $\delta(region_i)$ denotes a semantic region indicator function, which is assigned a value of 1 for foreground regions and 0 for background regions. The weighting coefficients α and β are determined through extensive cross-validation and are set to 0.7 and 0.3, respectively. This configuration effectively balances the contributions of saliency and semantic region characteristics in the harmonization process. As a result, higher harmonization weights are assigned to salient foreground regions, while the perceptual naturalness of background regions is simultaneously preserved, thereby establishing a robust foundation for subsequent region-adaptive color harmonization.

The essence of color harmonization lies in the appropriate matching of hue; therefore, a transformation of the image color space and the extraction of the key channel are first required. The input Red-Green-Blue (RGB) image is converted into the HSV color space, in which hue, saturation, and brightness are effectively decoupled. Among these components, the hue channel directly determines the intrinsic attribute of color and is thus regarded as the primary target for color harmonization. Consequently, the hue channel H is explicitly extracted for subsequent harmonization energy modeling. In contrast to conventional approaches that perform harmonization directly in the RGB space, the HSV representation provides stronger independence of the hue component, enabling a more precise characterization of color relationships. This formulation effectively avoids interference from saturation and brightness during hue adjustment, thereby significantly improving the specificity and accuracy of harmonization modeling. This characteristic constitutes one of the key design advantages of the proposed module.

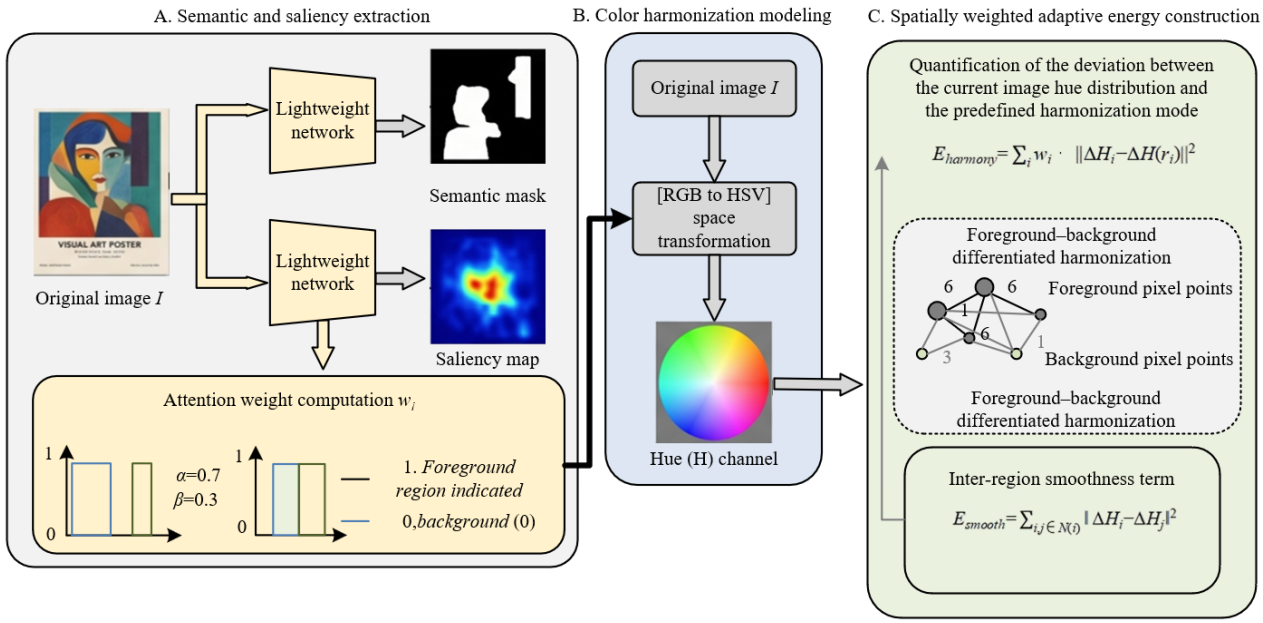

Figure 2. Schematic illustration of spatially perception-driven attention weights and harmonization modeling

The formulation of the harmonization energy term is critical for achieving adaptive color harmonization. A spatially weighted adaptive harmonization energy function is introduced to quantify the deviation between the current hue distribution of the image and the predefined harmonization mode. The expression is defined as $E_{harmony}=\sum_i w_i \cdot\|\Delta H_i-\widehat{\Delta H}(r_i)\|^2$, where ΔH_i denotes the hue difference between the i-th pixel and a reference hue. The reference hue H_{ref} is calculated as follows. First, K-means clustering (K = 5) is performed over the entire original image, and the hue value of the color cluster with the largest pixel proportion is extracted as the dominant hue $H_{dominant}$. If a distinct foreground region exists, the dominant foreground hue is adopted as the reference; otherwise, the dominant hue of the entire image is used. The predefined harmony mode determines the relationship between $H_{dominant}$ and the ideal hue difference ΔH_{ideal} . For example, under complementary harmony, ΔH_{ideal}=180. Accordingly, the ideal hue of the pixel is defined as $H_{ideal}(i)=H_{ref}+\Delta H_{ideal}$ (mod 360∘). The term $\widehat{\Delta H}(r_i)$ represents the ideal hue difference associated with the semantic region of the i-th pixel under the selected harmonization mode. For example, in a complementary harmonization mode, the ideal hue difference is defined as 180˚±15˚, whereas in an analogous harmonization mode, it is defined as 130˚±5˚ . A key innovation of this energy formulation lies in the incorporation of pixel-level attention weights w_i, through which differentiated harmonization constraints are imposed across pixels with varying semantic importance and saliency. Specifically, stricter constraints are applied to salient foreground regions, while relatively relaxed constraints are imposed on background regions. This strategy effectively mitigates the visual rigidity introduced by conventional global harmonization energy formulations and enables semantically adaptive color harmonization.

To address color discontinuity artifacts that arise at the boundaries between foreground and background due to differences in harmonization intensity and to enhance the overall coherence of the harmonized image, an inter-region smoothness term is introduced to constrain the harmonization process. This smoothness term is formulated based on the hue differences between neighboring pixels and is enforced using an L2-norm regularization to ensure a gradual transition of hue values within local neighborhoods. The formulation is given as:

$$E_{smooth}=\sum_{i,j\in N(i)}\|\Delta H_i-\Delta H_j\|^2 \quad (1)$$

where, $N(i)$ denotes the set of 8-neighborhood pixels surrounding the i-th pixel. Through this formulation, constraints are imposed on the hue differences of adjacent pixels, thereby enforcing a smooth and continuous transition across foreground–background boundaries. This mechanism effectively prevents abrupt hue variations caused by differences in attention weights, while preserving the harmonization characteristics within individual regions. As a result, visual coherence and perceptual naturalness of the harmonized image are significantly improved. By integrating the harmonization energy term with the inter-region smoothness term, a complete spatially weighted adaptive harmonization energy function is established. This unified formulation simultaneously achieves semantic-adaptive color harmonization and smooth inter-region transitions, demonstrating clear advantages over conventional global harmonization modeling approaches. It should be noted that E_{smooth} only constrains the differences in hue values between neighboring pixels. Its objective is to eliminate hue discontinuity artifacts occurring at the boundaries between foreground and background regions, without involving image gradients or texture information.

2.3 Perceptual consistency constraints based on the color appearance model

The primary objective of perceptual consistency constraints is to ensure that the color harmonization process aligns with human visual perception mechanisms, thereby avoiding perceptual distortions commonly introduced by optimization in conventional color spaces. In this study, the CIECAM02 color appearance model is explicitly embedded into the color harmonization optimization framework, providing a

principled foundation for perceptual constraint modeling through an accurate color appearance representation. A schematic illustration of the perceptual deviation field based on the CIECAM02 model is presented in Figure 3. Specifically, the preliminarily harmonized RGB image is transformed into the CIECAM02 color appearance space. This space enables a more accurate simulation of human visual perception characteristics compared with traditional HSV or CIELAB spaces, as its representation of lightness, chroma, and hue more closely conforms to the human visual response mechanism. The transformation is performed under standard viewing conditions, including a D65 illuminant and a 2° standard observer, which are consistent with typical visual observation environments. These conditions ensure the accuracy and consistency of the extracted color appearance parameters. Following the transformation, three key perceptual attributes-lightness (J), chroma (C), and hue (h)-are extracted to support the construction of the perceptual deviation field. Through the explicit incorporation of the color appearance model, the limitations of conventional approaches that rely solely on numerical color differences are effectively addressed, thereby fundamentally improving the rationality and precision of perceptual consistency constraints.

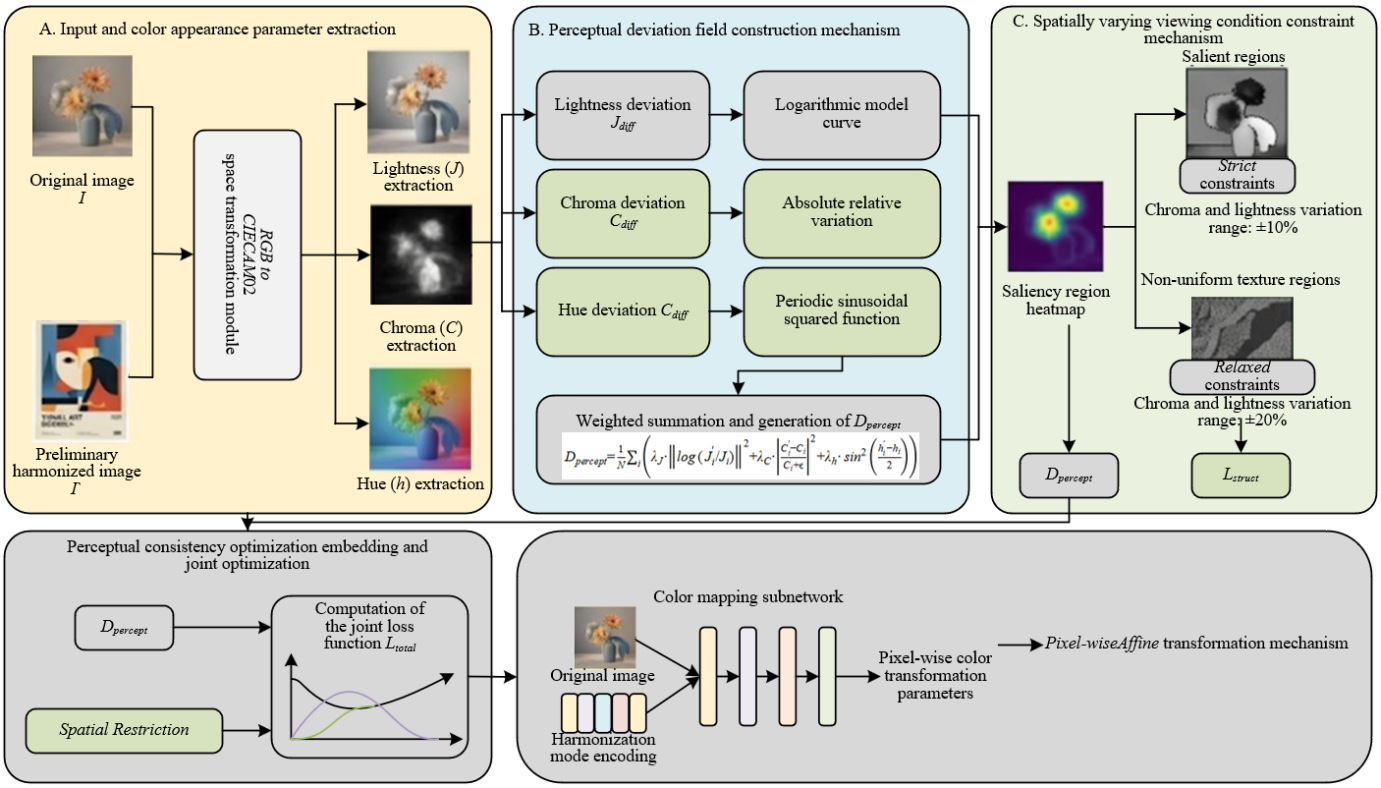

Figure 3. Perceptual deviation field constraint mechanism based on the CIE Color Appearance Model 2002 (CIECAM02) color appearance model

To accurately quantify perceptual deviations introduced during the color harmonization process, a perceptual deviation field incorporating constraints derived from the Weber-Fechner law is constructed. This formulation is designed to measure the discrepancy between the adjusted image and the original image at the level of human visual perception, thereby replacing conventional Euclidean distance as the metric for perceptual consistency. The formulation is expressed as:

$$D_{percept} = \frac{1}{N}\sum_i \left(\lambda_J \cdot \left\| log\left(J_i'/J_i\right) \right\|^2 + \lambda_C \cdot \left| \frac{C_i'-C_i}{C_i+\epsilon} \right|^2 + \lambda_h \cdot sin^2\left(\frac{h_i'-h_i}{2}\right) \right) \quad (2)$$

where, N denotes the total number of pixels in the image. The variables J_i', C_i', and h_i' represent the lightness, chroma, and hue of the i-th pixel in the adjusted image, respectively, while J_i, C_i, and h_i correspond to those of the original image. A key innovation of this perceptual deviation field lies in its departure from a single numerical difference metric. Instead, differentiated components are formulated according to the perceptual characteristics of human vision across distinct color dimensions. This enables precise quantification of perceptual deviations. Furthermore, a weighted summation strategy is adopted to balance the contributions of each dimension to the overall perceptual effect.

Each component of the perceptual deviation field is associated with a clear physical interpretation, and the corresponding parameter settings have been validated through extensive experiments to ensure both rationality and stability of the imposed constraints. The lightness deviation term is formulated in a logarithmic form, enabling an accurate approximation of the Weber-Fechner law. This design aligns with the nonlinear sensitivity of human vision to luminance variations, whereby changes in low-lightness regions are perceived more prominently. The logarithmic formulation effectively amplifies subtle deviations in low-lightness regions while suppressing excessive adjustments in high-lightness regions, thereby preventing luminance inversion artifacts. The chroma deviation term is defined using a normalized relative variation formulation, which effectively suppresses both oversaturation and undersaturation effects. This ensures that chroma adjustments remain within perceptually acceptable and visually natural ranges. ε is set to 1×10^{-6} to avoid

numerical instability caused by a zero denominator. The hue deviation term is formulated using a sinusoidal squared function, which explicitly accounts for the circular nature of hue representation. This formulation enables smooth transitions in hue differences and effectively avoids abrupt hue discontinuities that may lead to perceptual artifacts. The weighting coefficients are determined through cross-validation, with λ_J, λ_C, and λ_h set to 0.5, 0.3, and 0.2, respectively. This configuration ensures a balanced contribution of lightness, chroma, and hue to overall perceptual quality, thereby preserving perceptual naturalness. To prevent divergence of the logarithmic term when J_i approaches zero, a lower-bound protection term $\varepsilon = 10^{-6}$ is also introduced into the lightness deviation sub-term. In practical natural and design images, the CIECAM02 lightness value rarely falls below 10^{-5}; therefore, this protection is included solely for theoretical completeness.

To further enhance perceptual consistency and avoid the "plastic-like" appearance caused by globally uniform constraints, a spatially varying visual field is innovatively designed to achieve adaptive adjustment of perceptual constraint strength according to regional image characteristics. First, non-uniform texture regions are detected as follows. The variance of gradient magnitude, denoted as $Var_{grad}(i)$, is computed within a 15 × 15 window centred at pixel i. If $Var_{grad}(i)$ exceeds 1.2 times the mean gradient variance of the entire image, the region is classified as a non-uniform texture region; otherwise, it is regarded as a flat or regular region. Foreground salient regions are determined by the intersection of the saliency mask and the semantic mask. Second, the mathematical formulation and implementation of the visual field are defined as follows. Rather than constraining pixel values through hard clipping, the visual field functions as an adaptive weighting coefficient within the perceptual deviation field. A spatial variation factor $s(i)$ is defined as:

$$s(i) = \begin{cases} 1.0, \text{ salient regions} \\ 2.0, \text{ non-uniform texture regions} \\ 1.5, \text{ other regions} \end{cases}$$

Accordingly, the lightness coefficient λ_J and chroma coefficient λ_C in the perceptual deviation field $D_{percept}$ are multiplied by $1/s(i)$, such that tighter regions receive stronger constraint enforcement, whereas relaxed regions receive weaker constraint enforcement. The hue coefficient remains unchanged. This implementation constitutes a differentiable weight modulation mechanism, thereby ensuring gradient continuity. Finally, the relationship between the visual field and the pixel-level attention weight w_i is defined as follows. The weight w_i controls the strictness of harmony constraints across different regions within the harmony loss, whereas the visual field controls the tolerance of perceptual deviation across different regions within the perceptual consistency loss. Their collaboration enables the dual objectives of "differentiated harmony regulation" and "differentiated perceptual constraint."

2.4 Differentiable joint optimization framework and network design

To achieve end-to-end co-optimization of color harmonization, perceptual consistency, and image structural preservation-while simultaneously maintaining computational efficiency and generalization capability-a lightweight differentiable color mapping subnetwork Φ is designed to enable efficient feature extraction and adaptive color transformation. The architecture of the proposed subnetwork and the overall joint optimization framework are illustrated in Figure 4.

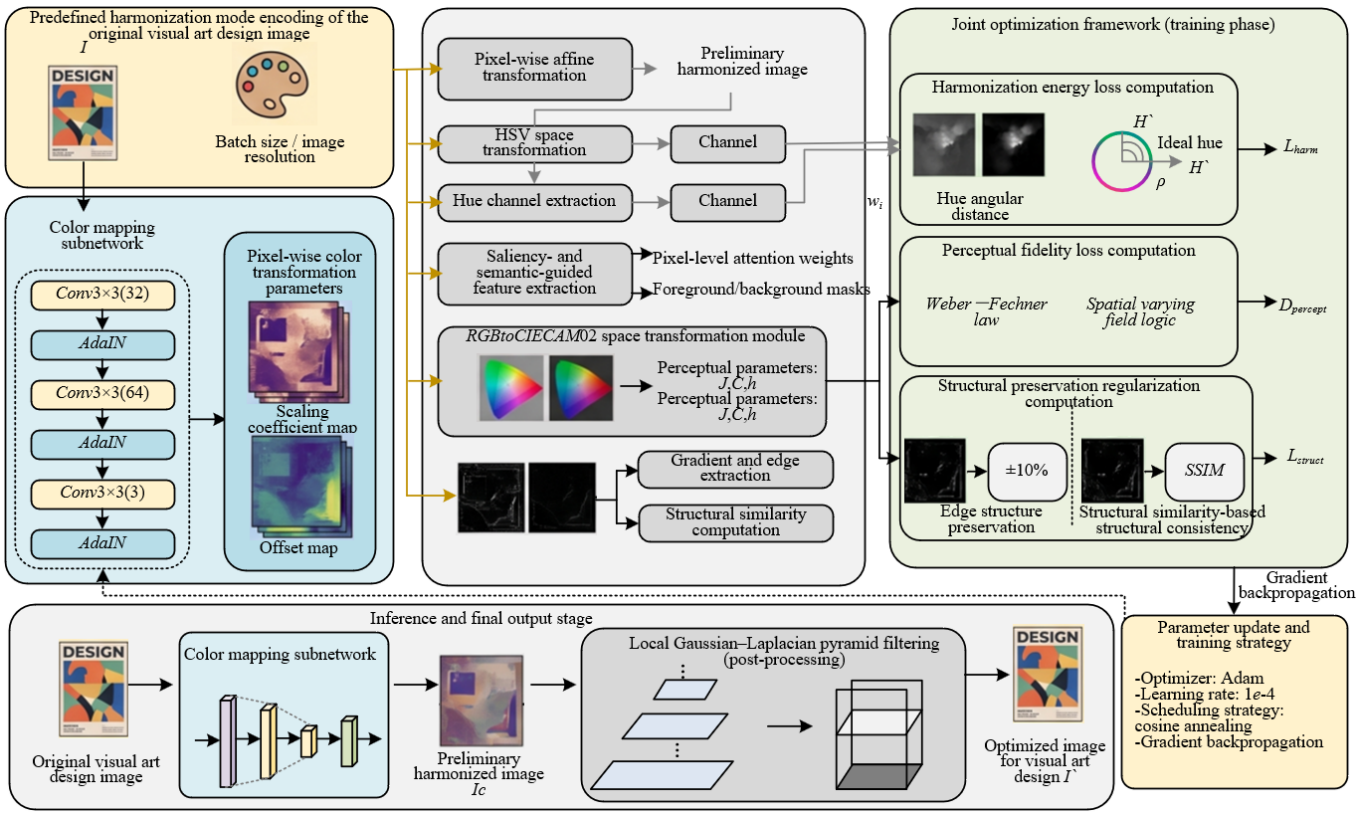

Figure 4. Lightweight differentiable color mapping network and joint optimization framework

The network input consists of the original RGB image I and a harmonization mode encoding vector t. The harmonization mode encoding vector transforms predefined harmonization types into learnable feature representations, thereby enabling adaptive accommodation of different harmonization modes. The network architecture adopts a compact three-layer convolutional structure, with kernel sizes of 3 × 3 and channel dimensions of 32, 64, and 3, respectively. This design ensures sufficient feature extraction capability while effectively controlling the number of model parameters, thereby improving computational efficiency during inference. Two adaptive instance normalization (AdaIN) modules are embedded between convolutional layers. These modules dynamically adjust the mean and variance of feature maps, allowing the network to adaptively modify feature distributions according to different harmonization modes. As a result, the limitations of fixed feature transformation in conventional networks are overcome, and precise alignment between harmonization modes and color transformations is achieved. Finally, a color transformation parameter map Θ, with the same spatial dimensions as the input image, is generated. This parameter map provides the necessary support for subsequent pixel-wise color adjustment.

Pixel-wise color transformation is a critical component for achieving precise color harmonization. Based on the color transformation parameter map Θ generated by the network, a pixel-wise affine transformation mechanism is designed to enable adaptive color adjustment of the input image. The transformation is formulated as $I_c'(i)=a_c(i)\cdot I_c(i)+b_c(i)$, where $c\in\{R, G, B\}$ denotes the three color channels. The term $I_c'(i)$ represents the transformed i-th pixel value in channel c, while $I_c(i)$ corresponds to the original pixel value. The parameters $a_c(i)$ and $b_c(i)$ denote the scaling coefficient and offset, respectively, both of which are produced by the final convolutional layer of the color mapping subnetwork. Their spatial dimensions are consistent with those of the input image, ensuring that each pixel is assigned an independent set of transformation parameters. This pixel-wise adaptive transformation mechanism overcomes the limitations of conventional global color transformations. By incorporating semantic characteristics, saliency information, and harmonization requirements at the pixel level, differentiated color adjustments can be achieved, thereby significantly enhancing both the precision and perceptual naturalness of color harmonization.

The design of the joint loss function constitutes the core of the unsupervised end-to-end optimization framework. A composite loss function is constructed by integrating a harmonization loss, a perceptual consistency loss, and a structural preservation regularization term, enabling simultaneous multi-objective optimization. The overall loss function is defined as $L=L_{harm}+\gamma_1 L_{cam}+\gamma_2 L_{struct}$. The harmonization loss L_{harm} is introduced to constrain the rationality of color harmonization and is expressed as:

$$L_{harm}=1/N\sum_i w_i\cdot\rho\left(H(I_i^{'}),\hat{h}_i\right) \tag{3}$$

where, $H(I_i')$ denotes the hue value of the i-th pixel in the adjusted image, while $\hat{h}_i$ represents the corresponding ideal hue value. The function ρ is defined as an angular distance $\rho(x,y)=min(|x-y|,360-|x-y|)$, which accurately accounts for the periodic nature of hue. Combined with the pixel-level attention weights w_i, this formulation ensures targeted and region-aware harmonization. The perceptual consistency loss L_{harm} is directly defined using the previously constructed perceptual deviation field $D_{percept}$. To enable end-to-end optimization, the forward and backward propagation processes of the CIECAM02 color appearance model are implemented within the PyTorch framework, ensuring differentiability of the perceptual deviation field and effective gradient propagation. The structural preservation regularization term L_{struct} is introduced to prevent edge blurring and texture degradation during color harmonization. It is formulated as $L_{struct}=\|\nabla I^{'}-\nabla I\|_1+\mathrm{SSIM}(I^{'},I)$, where the gradient consistency term $\|\nabla I^{'}-\nabla I\|_1$ preserves edge structures, and the structural similarity (SSIM) term evaluates structural consistency between the adjusted and original images, thereby effectively suppressing structural distortions. Unlike the inter-region smoothness term, L_{struct} operates on global image gradients and structural similarity in the RGB space, with the objective of preserving edge sharpness and texture details. The smoothness term addresses hue step artifacts, whereas the structural term prevents texture degradation. Therefore, the two components serve complementary functions. The weighting coefficients γ_1 and γ_2 are determined through cross-validation and are set to 1.0 and 0.8, respectively. This configuration ensures a balanced trade-off among harmonization quality, perceptual consistency, and structural preservation, enabling stable multi-objective optimization.

To facilitate unsupervised training and eliminate dependence on paired ground-truth data, a dedicated training strategy is designed to enhance both generalization capability and training stability. The training dataset consists of a hybrid collection of natural images and visual art design images, ensuring adaptability across diverse scenarios while strengthening performance in design-oriented color harmonization tasks. The Adam optimizer is employed with an initial learning rate of 1×10^{-4}, combined with a cosine annealing scheduling strategy to dynamically adjust the learning rate during training. This approach effectively mitigates overfitting while improving convergence speed and stability. Training is conducted with a batch size of 16 over 300 epochs. Model parameters are initialized using *He* normal initialization, ensuring appropriate parameter distribution and accelerated convergence. The proposed unsupervised training strategy enables effective model learning solely through the self-constrained joint loss function, without requiring paired original and target harmonized images. As a result, annotation costs are significantly reduced, while generalization performance in real-world visual art design scenarios is substantially improved, addressing a critical limitation of existing deep learning-based approaches.

2.5 Post-processing and perceptual color harmonization evaluation metric

After color harmonization and optimization, residual artifacts such as local block effects and edge blurring may still be present, thereby degrading visual quality. To address these issues, a local Gaussian-Laplacian pyramid filtering strategy is introduced as a post-processing step, with the dual objective of suppressing block artifacts while preserving edge structures. The key innovation of this approach lies in the integration of multi-scale residual guidance and gradient-based constraints derived from the original image, thereby overcoming the limitations of conventional filtering methods that often result in either edge degradation or incomplete

artifact removal. The processing pipeline is defined as follows. First, a residual image is computed by subtracting the original image from the optimized image, capturing block artifacts and local noise introduced during the harmonization process. Subsequently, multi-scale guided filtering is applied to the residual image, with the gradient of the original image serving as the guidance signal. The filter kernel size is set to 5 × 5, and the standard deviation is fixed at 1.2. Through multi-scale decomposition, frequency components at different scales are selectively processed. The underlying principle is that the gradient of the original image enables accurate discrimination between edge regions and block artifact regions. During filtering, block artifacts and noise within the residual image are effectively smoothed, while edge-related gradient information is preserved, thereby preventing edge blurring. Finally, the filtered residual image is fused with the initially optimized image to produce the final output, ensuring both high-quality color harmonization and improved visual coherence and sharpness.

Existing color harmonization evaluation metrics are typically limited to single-dimensional criteria and fail to simultaneously account for harmonization quality, perceptual consistency, and artifact suppression in visual art design images. As a result, objective performance assessment remains insufficient. To address this limitation, a perceptual color harmonization index is proposed, integrating three key components: harmonization quality, perceptual consistency, and artifact evaluation, aiming to construct a dedicated evaluation metric that is well suited for design-oriented images. First, each sub-component is linearly normalized to the range [0,1], and all metrics are unified such that larger values indicate better performance. *HarmonyScore* is the spatially weighted and corrected Moon-Spencer harmony score. Since its original range is already [0,1], with larger values indicating higher harmony, no further normalization is required. *Consistency* is defined as the inverse of the Jensen–Shannon divergence of the edge orientation histogram. After linear normalization to [0,1], larger values indicate better perceptual consistency. *ArtifactScore* is defined as 1-*Normalized (LogCSF*). Specifically, the logarithmic response of the contrast sensitivity function (*LogCSF*) is computed for all pixels in the image. Max-min linear normalization is then applied to obtain $LogCSF_{norm} \in [0,1]$. The artifact score is subsequently defined as $1-LogCSF_{norm}$, such that fewer artifacts correspond to a higher score.

$$PCHI=\alpha_1 \cdot HarmonyScore_{norm}+\alpha_2 \cdot Consistency_{norm} +\alpha_3 \cdot (1-LogCSF_{norm}) \quad (4)$$

where, all sub-components are normalized to the range [0,1]. The weighting coefficients $\alpha_1=0.4$, $\alpha_2=0.35$, and $\alpha_3=0.25$ are determined through subjective experiments. The perceptual color harmonization index ranges from 0 to 1, where larger values indicate better color harmony, higher perceptual consistency, and fewer artificial artifacts in the image. The proposed metric is distinguished by its multi-dimensional integration, enabling an application-oriented evaluation framework specifically tailored for visual art design images, thereby addressing the limitations of existing assessment approaches.

Each component of the perceptual color harmonization index is carefully designed to ensure both scientific rigor and practical applicability, while incorporating the core technical principles proposed in this study. The *HarmonyScore* term is employed to quantify the global degree of color harmonization. It is formulated as an extension of the classical Moon–Spencer harmonization theory, within which a spatial weighting mechanism is introduced. By integrating pixel-level attention weights into the harmonization evaluation process, the metric is enabled to reflect region-aware, semantically adaptive harmonization effects, thereby overcoming the limitations of traditional approaches that rely solely on global color statistics and neglect semantic characteristics. The *Consistency* term is designed to measure perceptual–semantic consistency between the optimized image and the original image. It is computed as the inverse of the Jensen–Shannon divergence between edge orientation histograms of the two images. The edge orientation histogram effectively captures structural and semantic features, while the Jensen-Shannon divergence quantifies the discrepancy between their distributions. Its inverse provides an intuitive measure of perceptual consistency, where larger values indicate higher structural similarity and stronger perceptual alignment. The *LogCSF* term represents the logarithmic response of the contrast sensitivity function, which is used to detect artificial artifacts introduced during the color harmonization process. The contrast sensitivity function models the sensitivity of human vision to contrast variations across different spatial frequencies. Its logarithmic form amplifies abnormal signals in artifact-prone regions. By computing 1-*LogCSF*, images with fewer artifacts yield higher scores for this term, thereby ensuring a more comprehensive evaluation.

The weighting coefficients of each component are determined through rigorous subjective experiments to ensure alignment with human perceptual judgment. A total of 50 participants are recruited, including 10 professional visual art designers and 40 non-expert observers. A dataset of 1,000 visual art design images-covering posters, illustrations, and user interface designs-is evaluated. Participants are instructed to rate each image in terms of harmonization quality, perceptual naturalness, and artifact presence. Subsequently, linear regression analysis is performed to examine the correlation between subjective scores and individual metric components. Based on this analysis, the weights are determined as $\alpha_1=0.4$, $\alpha_2=0.35$, and $\alpha_3=0.25$. This weighting scheme emphasizes the primary role of color harmonization quality while maintaining a balanced consideration of perceptual consistency and artifact suppression. As a result, the proposed perceptual color harmonization index metric achieves strong alignment with subjective human perception, providing an objective and reliable evaluation criterion for comparing the performance of the proposed algorithm against existing methods. Both its validity and effectiveness are confirmed through subjective experimental validation.

3. EXPERIMENTS AND RESULTS ANALYSIS

3.1 Experimental setup

To comprehensively evaluate the effectiveness, superiority, generalization capability, and practical applicability of the proposed method, multiple groups of controlled experiments were designed in accordance with the experimental protocols commonly adopted in leading Science Citation Index (SCI) journals in the field of image processing. The experimental configuration was explicitly specified to ensure reproducibility.

Three benchmark datasets were employed, covering both natural images and diverse visual art design scenarios, thereby enabling both generalization assessment and task-specific validation. The MIT-Adobe FiveK Dataset consisted of 5,000 natural images and was utilized to evaluate the generalization capability of the proposed method on non-design images. The Adobe Color CC dataset included 1,000 professionally designed posters and was used to assess the aesthetic performance of color harmonization. In addition, a self-constructed design dataset containing 800 images-including illustrations and user interface designs-was employed. These images were annotated with semantic regions and predefined harmonization modes, and were specifically used for ablation studies to accurately evaluate the contribution of each core module. The hardware and software environments were uniformly configured. All experiments were conducted on an NVIDIA RTX 3090 graphics processing unit to ensure computational efficiency. The implementation was based on Python 3.9 and PyTorch 1.12. The training and testing parameters were maintained consistently with those described in Section 2.4, including a batch size of 16, 300 training epochs, and the use of the Adam optimizer with an initial learning rate of 1×10^{-4}. A cosine annealing scheduling strategy was adopted, and *He* normal initialization was applied to weight initialization. These settings ensure consistency and reproducibility of the experimental results.

3.2 Ablation study

Ablation experiments were conducted to systematically evaluate the effectiveness of each core module. The experiments were performed on the self-constructed design dataset by configuring four comparison groups, in which only a single variable was modified at a time. The contributions of the spatially weighted harmonization energy, the CIECAM02 color appearance model-based perceptual consistency constraint, the differentiable joint optimization framework, and the perceptual color harmonization index-guided optimization were individually examined. The results are summarized in Table 1.

Table 1. Ablation study results on the self-constructed design dataset

Experimental Group	Perceptual Color Harmonization Index	Peak Signal-to-Noise Ratio (dB)	Structural Similarity Index	ΔE_{00}
Baseline	0.682	28.56	0.813	7.89
Baseline + spatially weighted harmonization energy	0.767	29.83	0.845	7.21
Baseline + spatially weighted harmonization energy + CIE Color Appearance Model 2002 (CIECAM02) perceptual constraint	0.835	31.27	0.879	6.58
Full proposed method	0.896	32.69	0.904	5.92

As indicated in Table 1, consistent performance improvements are observed with the inclusion of each module, and optimal results are achieved when all modules are jointly incorporated. In the baseline configuration, no core modules are introduced; global harmonization and CIELAB-based color difference constraints are applied. Under this setting, a perceptual color harmonization index value of 0.682 and a ΔE_{00} value of 7.89 are obtained, indicating substantial perceptual distortion and rigidity associated with conventional global harmonization approaches. When the spatially weighted harmonization energy is introduced, the perceptual color harmonization index increases to 0.767, corresponding to a relative improvement of 12.5% compared with the baseline. In addition, the peak signal-to-noise ratio is increased by 1.27 dB, and ΔE_{00} is reduced by 0.68. These results demonstrate that region-aware harmonization is effectively achieved, alleviating the limitations of global harmonization and improving both harmonization quality and image fidelity. With the further incorporation of the CIECAM02-based perceptual consistency constraint, the perceptual color harmonization index is increased to 0.835, while ΔE_{00} is reduced to 6.58, representing an additional decrease of 8.5% relative to the previous configuration. This improvement indicates that the perceptual deviation field and the spatially varying viewing condition field effectively reduce perceptual distortion and enhance perceptual fidelity. When all core modules are integrated, the full proposed method achieves a perceptual color harmonization index of 0.896, corresponding to a 31.4% improvement over the baseline. Meanwhile, the peak signal-to-noise ratio and the structural similarity index reach their highest values, and ΔE_{00} is reduced to 5.92. These results confirm that simultaneous optimization of harmonization quality, perceptual consistency, and structural preservation is successfully achieved.

3.3 Comparative experiments

Comparative experiments were conducted to evaluate the superiority of the proposed method over existing state-of-the-art approaches. Evaluations were performed on three datasets-the MIT-Adobe FiveK Dataset, the Adobe Color CC dataset, and the self-constructed design dataset. Six representative methods were selected for comparison, and all experimental settings were maintained consistently across methods. The results are summarized in Tables 2-4.

From the quantitative comparisons presented in Tables 2–4, it can be observed that the proposed method consistently outperforms all competing approaches across all evaluation metrics on the three datasets, with particularly significant improvements in the core metric, perceptual color harmonization index. On the MIT-Adobe FiveK dataset, a perceptual color harmonization index value of 0.887 is achieved, representing a 12.4% improvement over the best-performing comparative method, unsupervised color harmony. On the Adobe Color CC dataset, the perceptual color harmonization index reaches 0.892, corresponding to a 14.2% improvement. On the self-constructed design dataset, a perceptual color harmonization index value of 0.896 is obtained, yielding a 12.7% improvement. Overall, the performance gain ranges from 9.8% to 15.2%, demonstrating the superiority of the proposed method in both color harmonization quality and perceptual consistency.

Table 2. Comparative results on the MIT-Adobe FiveK dataset

Method	Perceptual Color Harmonization Index	Peak Signal-to-Noise Ratio (dB)	Structural Similarity Index	ΔE_{00}
Moon-Spencer	0.628	27.35	0.792	8.67
Matsuda	0.651	27.89	0.805	8.23
Color harmonization net	0.734	29.12	0.837	7.51
Deep color transfer	0.756	30.05	0.852	7.14
Unsupervised color harmony	0.789	30.87	0.871	6.79
Proposed method	0.887	32.45	0.898	6.03

Table 3. Comparative results on the Adobe Color CC dataset

Method	Perceptual Color Harmonization Index	Peak Signal-to-Noise Ratio (dB)	Structural Similarity Index	ΔE_{00}
Moon-Spencer	0.615	26.98	0.783	8.92
Matsuda	0.642	27.56	0.797	8.45
Color harmonization net	0.728	28.89	0.831	7.68
Deep color transfer	0.749	29.76	0.846	7.32
Unsupervised color harmony	0.781	30.54	0.865	6.95
Proposed method	0.892	32.87	0.906	5.87

Table 4. Comparative results on the self-constructed design dataset

Method	Perceptual Color Harmonization Index	Peak Signal-to-Noise Ratio (dB)	Structural Similarity Index	ΔE_{00}
Moon-Spencer	0.637	28.12	0.801	8.45
Matsuda	0.663	28.75	0.816	7.98
Color harmonization net	0.741	29.93	0.842	7.35
Deep color transfer	0.762	30.68	0.858	6.98
Unsupervised color harmony	0.795	31.32	0.876	6.62
Proposed method	0.896	32.69	0.904	5.92

Traditional harmonization methods, such as Moon-Spencer and Matsuda methods, exhibit the lowest performance across all metrics, with perceptual color harmonization index values below 0.67 and ΔE_{00} values exceeding 7.98. These results indicate that conventional global harmonization models are insufficient for addressing the heterogeneous requirements of visual art design images and tend to introduce significant perceptual distortions. Deep learning-based comparison methods demonstrate improved performance relative to traditional approaches; however, several limitations remain. The color harmonization net lacks semantic-adaptive harmonization capability, resulting in perceptual color harmonization index values generally below 0.75. Deep color transfer exhibits deficiencies in perceptual consistency constraints, as reflected by relatively high ΔE_{00} values. Although unsupervised color harmony enables training without paired data, it does not incorporate perceptual consistency constraints or structural preservation mechanisms, leading to inferior performance in both the perceptual color harmonization index and the structural similarity index compared with the proposed method. Furthermore, the proposed method achieves the highest peak signal-to-noise ratio and structural similarity index values, while attaining the lowest ΔE_{00} values across all datasets. These results indicate that improvements in harmonization quality and perceptual naturalness are achieved without compromising structural fidelity or image quality.

To assess the statistical significance of the observed performance differences, a t-test was conducted on the perceptual color harmonization index results between the proposed method and unsupervised color harmony, which represents the strongest competing approach. With a significance level of $\alpha = 0.05$, p-values obtained across all three datasets were found to be less than 0.05, indicating that the performance improvements are statistically significant. This finding further substantiates the superiority of the proposed method.

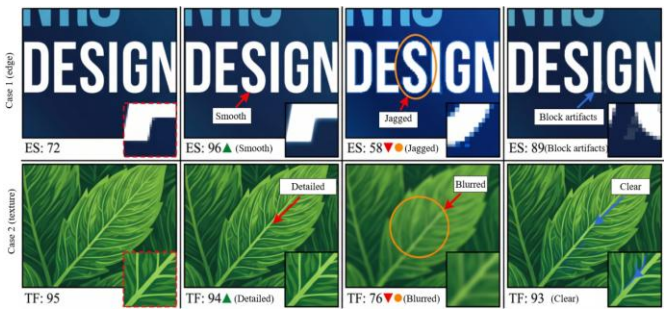

Figure 5. Local region magnification comparison: Visual perception evaluation of edge transition and texture preservation

Further validation was conducted at the microscopic scale to assess the accuracy of color harmonization and the structural fidelity of the proposed method in semantically sensitive regions, thereby providing localized visual evidence for the overall perceptual consistency optimization. Two representative visual art design scenarios were selected, namely text edge regions and illustration texture regions. Correspondingly, edge smoothness and texture fidelity were quantitatively evaluated. The results presented in Figure 5

indicate that, in the text edge case, an edge smoothness value of 96 is achieved by the proposed method, representing a 65.5% improvement over the baseline method, unsupervised color harmony, which attains a value of 58. Furthermore, the post-processing stage enhances edge smoothness from 89 to 96, effectively eliminating block artifacts. In the texture case, a texture fidelity value of 94 is obtained by the proposed method, whereas the baseline method achieves only 76. Notably, even without post-processing, a texture fidelity value of 93 is maintained, indicating that the structural preservation regularization term independently contributes to effective texture retention. These results demonstrate that the spatially weighted adaptive harmonization energy and the structural preservation regularization term enable precise differentiation between foreground edge regions and background texture regions. As a result, accurate color harmonization is achieved while avoiding typical perceptual distortions such as edge jagging and texture blurring. In addition, the post-processing filtering mechanism exhibits targeted capability in correcting edge artifacts without degrading fine texture details.

3.4 Generalization experiments

Generalization experiments were conducted to evaluate the adaptability of the proposed method across diverse visual art design scenarios and different harmonization modes. Three representative application scenarios were considered, including poster design, illustration creation, and photographic post-processing. For each scenario, 200 images were selected to ensure sufficient statistical validity. The performance of the proposed method was further assessed under three commonly used harmonization modes, namely complementary, analogous, and triadic harmonization. The experimental results are illustrated in Figure 6.

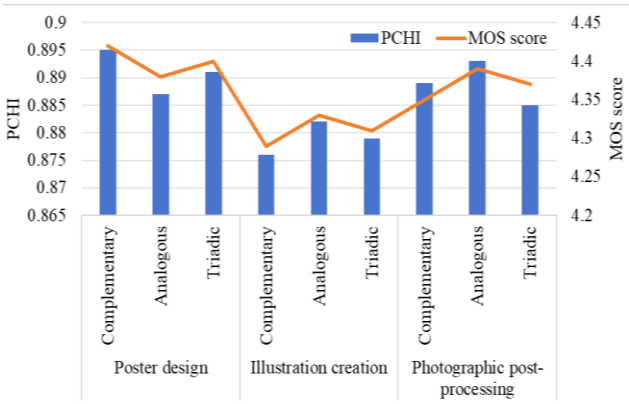

Figure 6. Generalization experimental results

From the results presented in Figure 6, it can be observed that consistently strong performance is maintained by the proposed method across different visual art design scenarios and harmonization modes. Specifically, perceptual color harmonization index values are consistently higher than 0.876, while mean opinion scores remain above 4.29, thereby demonstrating robust generalization capability. In the poster design scenario, the highest overall perceptual color harmonization index and mean opinion score values are achieved. Under the complementary harmonization mode, a perceptual color harmonization index of 0.895 and a mean opinion score of 4.42 are obtained. This superior performance can be attributed to the emphasis on foreground-background contrast in poster design, where the spatially weighted harmonization energy effectively enables region-specific color adjustment, adapting to the requirements of poster design. In the illustration creation scenario, slightly lower perceptual color harmonization index and mean opinion score values are observed compared with poster design and photographic post-processing. This reduction is primarily due to the presence of intricate textures and complex semantic regions. Nevertheless, performance remains at a high level, indicating that the proposed method effectively preserves fine-grained textures while avoiding artifact generation and structural degradation. In the photographic post-processing scenario, the best performance is achieved under the analogous harmonization mode, with a perceptual color harmonization index of 0.893 and a mean opinion score of 4.39. This result aligns with the requirement for smooth and natural color transitions in photographic images, further confirming the generalization capability of the proposed method in natural image harmonization tasks.

Additional case studies further validate scenario adaptability. In poster design examples, foreground text and graphical elements are accurately emphasized, while background colors are harmonized without excessive adjustment, thereby preserving visual hierarchy. In illustration examples, fine texture details are effectively maintained, and color harmonization is achieved without compromising the original artistic style. In photographic post-processing examples, harmonization intensity is adaptively adjusted according to image content, resulting in natural color transitions and enhanced visual appeal. Overall, the proposed method demonstrates strong adaptability across diverse visual art design scenarios and harmonization modes, exhibiting excellent generalization capability and the ability to meet a wide range of design requirements.

3.5 Subjective evaluation

A subjective evaluation was conducted to assess the perceptual naturalness and aesthetic quality of the proposed method. A total of 50 participants were recruited, including 10 professional visual art designers and 40 non-expert observers. A double-blind testing protocol was adopted, under which the output images generated by the proposed method and the strongest comparative approach, unsupervised color harmony, were evaluated using a five-point rating scale. The evaluation criteria included three dimensions: harmonization quality, perceptual naturalness, and structural integrity. The experimental results are summarized in Figure 7.

From the statistical results presented in Figure 7, it can be observed that the proposed method consistently receives significantly higher subjective ratings than unsupervised color harmony across both professional designers and non-expert observers. The overall mean opinion score for all participants reaches 4.44, exceeding the comparative method by 0.57, indicating a significant improvement. Among professional designers, an overall score of 4.52 is achieved by the proposed method, representing an improvement of 0.60 over the comparative approach. Notably, differences of 0.61 are observed in both the harmonization quality and perceptual naturalness dimensions, indicating that the proposed method aligns more closely with professional design standards and delivers superior perceptual realism.

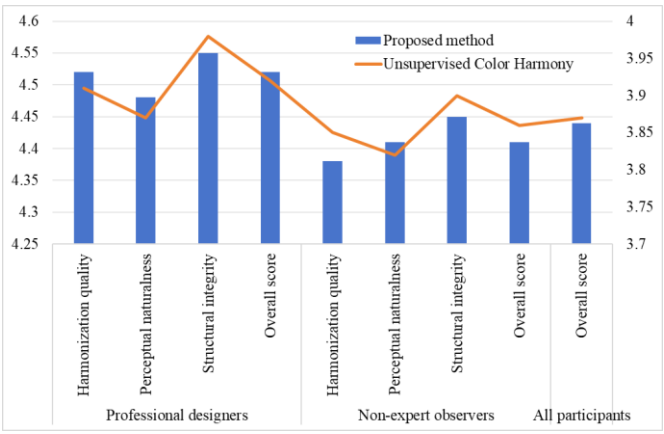

Figure 7. Subjective evaluation results (mean opinion score, mean ± standard deviation)

For non-expert observers, an overall score of 4.41 is obtained, exceeding the comparative method by 0.55. The largest improvement is observed in the perceptual naturalness dimension, with a difference of 0.59, suggesting that the proposed method better conforms to general human visual perception and effectively reduces perceptual distortions. Analysis of participant feedback reveals that the advantages of the proposed method are primarily reflected in three aspects. First, color harmonization is perceived as natural, with no evident artifacts such as oversaturation or luminance inversion, thereby satisfying aesthetic requirements in visual art design. Second, the original style and structural integrity of the image are effectively preserved, with fine details and textures maintained, avoiding the rigid appearance typically associated with over-harmonization. Third, a balanced harmonization between foreground and background is achieved, ensuring clear emphasis on the primary subject and supporting effective visual communication in design images. These observations further confirm the effectiveness of the perceptual consistency constraints and the structural preservation regularization term. Moreover, they demonstrate that the proposed method achieves superior performance in both perceptual naturalness and aesthetic quality, successfully meeting the visual expectations of diverse user groups.

3.6 Computational efficiency analysis

Computational efficiency is a critical indicator of practical applicability. To evaluate the efficiency of the proposed method, comparisons were conducted under a unified hardware environment (NVIDIA RTX 3090) across all methods, including model parameter size, inference speed (frames per second), and per-epoch training time. The results are summarized in Table 5.

Table 5. Comparison of computational efficiency

Method	Model Parameters (M)	Inference Speed (Frames per Second)	Training Time per Epoch (min)
Moon-Spencer	0.02	45.3	1.2
Matsuda	0.03	42.7	1.5
Color harmonization net	5.2	18.6	12.8
Deep color transfer	6.7	15.3	15.2
Unsupervised color harmony	4.8	20.1	11.5

From the results presented in Table 5, it can be observed that the proposed method demonstrates a clear advantage in computational efficiency, achieving an effective balance between model lightweight design and high performance. Traditional harmonization methods, such as Moon-Spencer and Matsuda methods, exhibit the smallest number of parameters and the highest inference speed; however, their overall performance remains limited and insufficient to satisfy the requirements of visual art design applications. Among deep learning-based comparison methods, color harmonization net and deep color transfer both contain more than 5 million parameters, with inference speeds below 20 frames per second and per-epoch training times exceeding 11 minutes. These characteristics indicate substantial computational overhead, rendering them unsuitable for real-time design scenarios. The unsupervised color harmony method contains 4.8 million parameters, achieves an inference speed of 20.1 frames per second, and requires 11.5 minutes per training epoch. Although slightly more efficient than the aforementioned deep learning approaches, it still exhibits limitations in computational efficiency.

In contrast, the proposed method requires only 0.86 million parameters, which is significantly lower than all deep learning-based comparison methods. An inference speed of 32.4 frames per second is achieved, exceeding that of all deep learning-based methods, while the per-epoch training time is reduced to 8.7 minutes. These results indicate a substantial improvement in computational efficiency. The efficiency gain is primarily attributed to the lightweight differentiable color mapping network, which employs only three convolutional layers and two adaptive instance normalization modules, thereby effectively reducing model complexity. In addition, the joint loss function is designed to balance optimization performance and computational cost, avoiding excessive overhead associated with complex formulations. Consequently, the proposed method satisfies the real-time requirements of visual art design applications, significantly improving design efficiency and demonstrating strong practical applicability.

4. DISCUSSION

The experimental results demonstrate that superior performance is achieved by the proposed algorithm in the task of color harmonization and perceptual consistency optimization for visual art design images. The primary reason for this performance lies in the coordinated design and integration of four core modules. The spatially weighted harmonization energy, constructed through the integration of semantic-guided saliency detection and a pixel-level attention mechanism, enables region-adaptive harmonization. By assigning differentiated constraints to foreground and background regions, the limitations of conventional global harmonization rules are effectively overcome, and the issue of visual rigidity is fundamentally mitigated. The explicit incorporation of the CIECAM02 color appearance model replaces traditional color difference constraints based on HSV or CIELAB spaces. This formulation enables a more accurate

alignment with human visual perception. When combined with the perceptual deviation field constrained by the Weber–Fechner law and the spatially varying viewing condition field, perceptual distortions are effectively suppressed, and the naturalness of color adjustment is significantly improved. The joint loss function enables a balanced multi-objective optimization of color harmonization, perceptual consistency, and structural preservation. As a result, harmonization quality is enhanced while edge blurring and texture degradation are effectively avoided. Furthermore, the proposed perceptual color harmonization index, which integrates harmonization quality, perceptual consistency, and artifact evaluation, addresses the lack of dedicated evaluation metrics for visual art design images and provides a reliable and comprehensive criterion for performance assessment. This provides a reliable objective basis for experimental validation. Meanwhile, certain limitations have also been observed. In highly saturated design images, the use of fixed perceptual constraint thresholds may slightly restrict the flexibility of color harmonization. Future work may address this issue by introducing dynamically adaptive perceptual constraint thresholds, in which the constraint strength is adjusted according to image saturation, thereby further improving the adaptability of the proposed method.

Despite the significant effectiveness, certain inherent limitations remain. The accuracy of semantic segmentation exerts a moderate influence on harmonization performance. In complex scenarios involving multiple subjects or overlapping semantic regions, segmentation inaccuracies may lead to suboptimal attention weight allocation, thereby affecting harmonization precision. The current framework is also limited to predefined harmonization modes and does not support flexible user-defined harmonization, which may restrict its applicability in personalized design workflows. Moreover, in the processing of extremely low-resolution design images, structural preservation performance remains suboptimal, as excessive harmonization may result in the loss of fine texture details and reduced visual quality. These limitations provide clear directions for future research, including the enhancement of semantic segmentation robustness, the development of flexible harmonization control mechanisms, and the improvement of structural preservation in low-resolution scenarios. To address the aforementioned limitations, future work will be directed toward four key aspects. First, Transformer-based architectures will be introduced to enhance semantic segmentation accuracy, enabling precise capture of multi-object semantic information in complex scenes and facilitating multi-object adaptive color harmonization. Second, an interactive harmonization mode editing module will be developed to allow user-defined specification of ideal hue differences and harmonization intensity, thereby improving flexibility and practical usability. Third, super-resolution techniques will be incorporated to enable simultaneous color harmonization and detail reconstruction for low-resolution images, further improving structural preservation. Fourth, the proposed single-image color harmonization framework will be extended to video color harmonization tasks, with particular emphasis on addressing temporal color consistency across frames, thereby broadening the application scope.

The present study holds significant implications at both theoretical and practical levels, providing a novel perspective and methodological paradigm for color harmonization and perceptual consistency optimization. At the theoretical level, a unified "semantic-perceptual-optimization" framework is established, in which semantic-adaptive harmonization, human visual perception constraints, and unsupervised end-to-end optimization are systematically integrated. This formulation overcomes the limitations of single-objective optimization in existing methods and enriches the research landscape of color harmonization and perceptual consistency. Furthermore, it introduces a new paradigm for unsupervised color image processing and offers a valuable entry point for interdisciplinary research at the intersection of visual art design and image processing. At the application level, the proposed algorithm can be directly applied to a wide range of domains, including visual art design, photographic post-processing, user interface/user experience design, and digital illustration. It provides designers with an efficient, controllable, and perceptually natural color harmonization tool, thereby reducing the technical barriers associated with color adjustment and improving both design efficiency and aesthetic quality. In addition, the lightweight architecture and high computational efficiency enable deployment in real-time design environments, further enhancing its practical applicability and indicating strong potential for large-scale industrial adoption.

5. CONCLUSION

In response to the key challenges in color harmonization for visual art design images-including the rigidity of global harmonization, significant perceptual deviations, and reliance on paired ground-truth data-an unsupervised, end-to-end joint optimization algorithm for color harmonization and perceptual consistency was proposed. To achieve simultaneous enhancement of harmonization quality and perceptual naturalness, four core technical innovations were incorporated. First, a spatially weighted adaptive harmonization energy modeling method was developed. By integrating semantic-guided saliency detection, region-aware differentiated harmonization was achieved, thereby overcoming the limitations of conventional global harmonization approaches. Second, the CIECAM02 color appearance model was explicitly embedded into the optimization framework. A perceptual deviation field constrained by the Weber–Fechner law, together with a spatially varying viewing condition field, was constructed to enhance perceptual fidelity in accordance with human visual perception. Third, a lightweight differentiable color mapping network and a joint loss function were designed, enabling end-to-end unsupervised co-optimization of harmonization energy, perceptual consistency, and structural preservation. Finally, a perceptual color harmonization index was introduced to quantitatively evaluate harmonization quality, perceptual consistency, and artifact suppression, thereby addressing the lack of dedicated evaluation metrics for design-oriented images.

Comprehensive experimental validation-including ablation studies, comparative experiments, generalization evaluations, subjective assessments, and computational efficiency analysis-demonstrated that superior performance was consistently achieved across the MIT-Adobe FiveK dataset, the Adobe Color CC dataset, and the self-constructed design dataset. All objective metrics and subjective evaluation scores were significantly improved compared with state-of-the-art methods, with both the perceptual color harmonization index

and mean opinion scores reaching the highest levels. The proposed approach provides an efficient, controllable, and perceptually natural solution for color optimization in visual art design. In addition, a novel perspective is introduced for perceptual consistency optimization in image processing, enriching the paradigm of unsupervised color image processing. The method exhibits substantial theoretical significance and strong practical potential, with broad applicability in visual art design, photographic post-processing, user interface/user experience design, and digital illustration, thereby contributing to enhanced design efficiency and improved aesthetic quality of visual content.

REFERENCES

[1] Ju, J.H., Ma, Y.H., Gong, T., Zhuang, E. (2024). Development model based on visual image big data applied to art management. Heliyon, 10(17): e37478. https://doi.org/10.1016/j.heliyon.2024.e37478

[2] Jiang, S.Q., Du, J., Huang, Q.M., Huang, T.J., Gao, W. (2005). Visual ontology construction for digitized art image retrieval. Journal of Computer Science and Technology, 20: 855-860. https://doi.org/10.1007/s11390-005-0855-x

[3] Szabó, F., Bodrogi, P., Schanda, J. (2009). Experimental modeling of colour harmony. Color Research & Application, 35(1): 34-49. https://doi.org/10.1002/col.20558

[4] Li, K.R., Yang, Y.Q., Zheng, Z.Q. (2019). Research on color harmony of building façades. Color Research & Application, 45(1): 105-119. https://doi.org/10.1002/col.22448

[5] Burchett, K.E. (2001). Color harmony. Color Research & Application, 27(1): 28-31. https://doi.org/10.1002/col.10004

[6] Tong, Z. (2021). Image sensory experience of artistic design based on the role of omnidirectional vision sensors. Journal of Sensors, 2021(1): 7166142. https://doi.org/10.1155/2021/7166142

[7] Li, D.Z., Alkathir, E.S. (2021). Implementation of computer-based vision technology to consider visual form of ceramic mural art. Mathematical Problems in Engineering, 2021(1): 4236572. https://doi.org/10.1155/2021/4236572

[8] Wang, Z. (2023). New media art design based on fast visual segmentation and 3D image processing. PeerJ Computer Science, 9: e1640. https://doi.org/10.7717/peerj-cs.1640

[9] Lu, P., Peng, X.J., Li, R.F., Wang, X.J. (2015). Towards aesthetics of image: A Bayesian framework for color harmony modeling. Signal Processing: Image Communication, 39: 487-498. https://doi.org/10.1016/j.image.2015.04.003

[10] Yang, B., Wei, T., Fang, X., Deng, Z., Li, F.W.B., Ling, Y., Wang, X. (2019). A color-pair based approach for accurate color harmony estimation. Computer Graphics Forum, 38(7): 481-490. https://doi.org/10.1111/cgf.13854

[11] Lu, P., Peng, X.J., Yuan, C.X., Li, R.F., Wang, X.J. (2016). Image color harmony modeling through neighbored co-occurrence colors. Neurocomputing, 201: 82-91. https://doi.org/10.1016/j.neucom.2016.03.035

[12] Zhong, X.Q. (2025). A data-driven approach to traditional village colors: K-Means clustering of online images. Journal of Asian Architecture and Building Engineering, 24(6): 5815-5826. https://doi.org/10.1080/13467581.2024.2428275

[13] Zhang, C. (2026). Comparative study of color in traditional and contemporary flower and bird paintings. NPJ Heritage Science, 14: 189. https://doi.org/10.1038/s40494-026-02429-3

[14] Dey, S., Al-Ani, J.A., Bourazeri, A., Saha, S., Purkait, R., Hill, S., Thompson, J. (2024). Pixelator v2: A novel perceptual image comparison method with LAB colour space and Sobel edge detection for enhanced security analysis. Electronics, 13(22): 4541. https://doi.org/10.3390/electronics13224541

[15] Marefat, S., Shayanfar, A., Monajjemzadeh, F. (2025). Developments in image-based colorimetric analysis methods and applications of CIElab color space in pharmaceutical sciences: A narrative review. International Journal of Pharmaceutics: X, 10: 100434. https://doi.org/10.1016/j.ijpx.2025.100434

[16] Zhai, Y.J., Gong, R.Y., Huo, J.Z., Fan, B.B. (2023). Building façade color distribution, color harmony and diversity in relation to street functions: Using street view images and deep learning. ISPRS International Journal of Geo-Information, 12(6): 224. https://doi.org/10.3390/ijgi12060224

[17] Yu, M.Y., Zheng, X.Y., Cui, W.K., Ji, Q.R. (2024). Urban color perception and sentiment analysis based on deep learning and street view big data. Applied Sciences, 14(20): 9521. https://doi.org/10.3390/app14209521

[18] Anwar, A., Kanwal, S., Tahir, M., Saqib, M., Uzair, M., Rahmani, M.K.I. (2022). Image aesthetic assessment: A comparative study of hand-crafted & deep learning models. IEEE Access, 10: 101770-101789. https://doi.org/10.1109/ACCESS.2022.3209196

[19] Navas, K.A., Sasikumar, M. (2011). Image fidelity metrics: Future directions. IETE Technical Review, 28(1): 50-56. https://doi.org/10.4103/0256-4602.74507

[20] Duminil, A., Ieng, S.S., Gruyer, D. (2024). A comprehensive exploration of fidelity quantification in computer-generated images. Sensors, 24(8): 2463. https://doi.org/10.3390/s24082463