Disentangled Emotion-Controllable Color Palette Generation via β-CVAE with Conditional Prior

1st Mu Huan Sun
Department of Graphic Arts and Communications
National Taiwan Normal University
Taipei, Taiwan
61372003h@gapps.ntnu.edu.tw

2nd Tzren Ru Chou
Department of Graphic Arts and Communications
National Taiwan Normal University
Taipei, Taiwan
trchou@ntnu.edu.tw

***Abstract*—Color plays a crucial role in visual communication and multimedia content creation, yet generating color palettes that accurately convey specific emotions remains challenging. This paper presents an application-oriented generative framework that adapts established variational modeling techniques to the problem of emotion-conditioned color synthesis. Grounded in Kobayashi's Color Image Scale, the method maps continuous affective coordinates to three-color combinations. To address the scarcity of annotated data, an XGBoost-based pseudo-labeling strategy expands a small set of labeled samples into a dataset of 53,698 web-collected triplets. The study employs a β-Conditional Variational Autoencoder (β-CVAE) with a learned conditional prior, configured to reflect emotion-specific variations in the latent space. Direct alignment and decorrelation constraints are incorporated to disentangle emotion-controlling dimensions from residual stylistic factors, enabling independent manipulation of emotional intent and visual style. Experimental results show strong correlations between latent dimensions and affective axes (Spearman's $\rho = 0.926$ for warm–cool, 0.954 for soft–hard), demonstrating effective control and stylistic diversity. The proposed framework provides an interpretable and practical tool for emotion-aware multimedia design, visualization systems, and affective computing applications.**

***Index Terms*—Emotion-conditioned generation, Color Palette Generation, CVAE, Disentangled representation, Affective Computing**

I. Introduction

Color plays an essential role in multimedia and visual communication, influencing aesthetic quality, user attention, and overall experience. In applications such as interface design, branding, and data visualization [1]–[3], designers often rely on carefully crafted color palettes to convey specific impressions. However, creating palettes that express particular emotional qualities remains challenging and typically requires expert knowledge, motivating automated approaches for controlled palette generation.

Kobayashi's Color Image Scale (CIS) [4], [5] is a widely used reference for describing impressions of color combinations. CIS associates palettes with key image-words (e.g., "elegant," "pretty," "dynamic") and places them within a two-dimensional Warm–Cool and Soft–Hard space. While useful for design practice, its limited number of samples makes it insufficient for training generative models that require continuous and high-coverage affective supervision.

Traditional palette extraction techniques—such as histogram analysis, Principal Component Analysis (PCA), and K-means clustering [6], [7]—rely on low-level pixel statistics and cannot represent high-level affective semantics or generate diverse variations. Conversely, while deep generative models can synthesize harmonious palettes [8]–[11], most lack mechanisms for controlling outputs via continuous affective parameters and do not disentangle intended impressions from stylistic variation.

To address these gaps, we propose an emotion-controllable palette generation framework based on a disentangled β-Conditional Variational Autoencoder (β-CVAE). The model incorporates a learned conditional prior to capture impression-dependent structure and partitions the latent space into emotion-aligned and style-related components. Furthermore, we construct a large-scale pseudo-labeled dataset by extending CIS samples using XGBoost-based coordinate prediction, enabling continuous and robust generative modeling across the affective space.

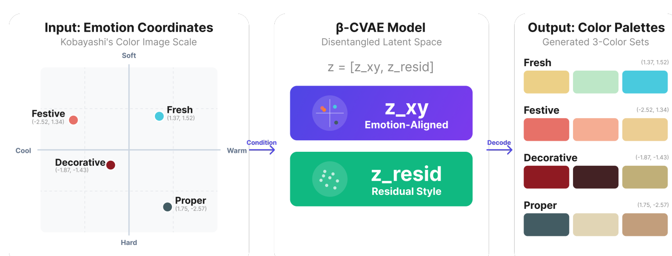

Fig. 1. Conceptual overview of the proposed emotion-controllable color palette generation framework. Given affective coordinates from the Color Image Scale, the β-CVAE encodes the emotion-aligned latent variables $\mathbf{z}_{xy}$ and residual style factors $\mathbf{z}_{\text{resid}}$, producing color palettes that reflect the intended emotional impression.

II. Related Work

A. Kobayashi's Color Image Scale

The Color Image Scale (CIS) has been widely adopted in applications such as image indexing, semantic annotation, painting-style emotion analysis, and color generative tasks [12]–[15]. These studies demonstrate that CIS offers a practical descriptive framework for associating color combinations with impression-related keywords, enabling its use across psychology, design, and computer vision research. Its structured Warm–Cool and Soft–Hard axes further support downstream tasks in affective design, color recommendation, and automated palette synthesis.

In the context of palette generation, several works have explored the use of CIS-labeled data. One approach introduces a Hue Spread Index (HSI) that quantifies subjective color harmony principles into computational metrics, enabling automated yet adjustable palette construction [11]. Another line of work employs autoencoder-based models to generate mixed-color palettes aligned with CIS impression categories [10]. While these studies illustrate the feasibility of using CIS for generative purposes, they remain constrained by the limited size of the original dataset and rely on relatively simple architectures. Autoencoder-based models, in particular, often struggle with controllability and disentanglement, making it difficult to separate affective intent from stylistic variation or to perform tasks such as style transfer.

B. Generative Palette Models

Deep generative models have been successfully applied to automated color tasks, including color harmony extraction [16], word–color association learning [17], and cross-modal emotion transfer [18]. While architectures such as VAEs and GANs can synthesize visually coherent palettes [8], [9], they typically prioritize visual fidelity over semantic controllability. Most approaches lack mechanisms for explicitly adjusting outputs through continuous affective parameters, limiting their use in tasks requiring precise emotional tuning.

A recent review on palette generation [19] highlights broader challenges in computational color design, emphasizing that color selection is inherently connected to emotional expression and cultural context. The review argues for systems that are not only harmonious but also semantically aligned with desired impressions. However, existing generative frameworks tend to conflate affective intent with stylistic variation in their latent representations. Consequently, modifying emotional characteristics often leads to unintended alterations in the palette's compositional structure, making independent control difficult.

III. METHODOLOGY

A. Problem Formulation

The goal of this work is to learn a conditional generative model that synthesizes three-color palettes consistent with specified emotional coordinates. Each palette is represented as $\mathbf{x} \in \mathbb{R}^9$, formed by concatenating three RGB triplets normalized to $[0, 1]$. The emotional condition $\mathbf{c} \in \mathbb{R}^2$ corresponds to the Warm–Cool and Soft–Hard dimensions of Kobayashi's Color Image Scale.

We model the conditional distribution $p(\mathbf{x}|\mathbf{c})$ using a Conditional Variational Autoencoder (CVAE), where an encoder approximates the posterior $q_\phi(\mathbf{z}|\mathbf{x}, \mathbf{c})$ and a decoder generates palettes from latent variables and conditions. The following section describes the latent structure and conditional prior used to enable emotion-controlled generation.

B. Model Architecture

The proposed framework consists of four components: an encoder, a decoder, a conditional prior network, and two auxiliary alignment heads.

Encoder. The encoder receives the concatenation of the palette $\mathbf{x} \in \mathbb{R}^9$ and condition $\mathbf{c} \in \mathbb{R}^2$, mapping them into the latent distribution parameters $(\boldsymbol{\mu}, \log \boldsymbol{\sigma}^2) \in \mathbb{R}^6$ of the approximate posterior $q_\phi(\mathbf{z}|\mathbf{x}, \mathbf{c})$. It is implemented as a lightweight multilayer perceptron with SiLU activations.

Decoder. The decoder reconstructs palettes from the concatenation $[\mathbf{z}, \mathbf{c}]$, producing an output in $[0, 1]^9$. A symmetric multilayer perceptron with a sigmoid output layer ensures valid RGB ranges.

Conditional Prior Network. To enable continuous control along affective dimensions, we introduce a conditional prior over the emotion-related subspace $\mathbf{z}_{xy}$. Instead of using a fixed isotropic Gaussian prior, a small neural network predicts the mean of the latent distribution conditioned on the emotional coordinates:

$$p_\psi(\mathbf{z}_{xy}|\mathbf{c}) = \mathcal{N}(\boldsymbol{\mu}_\psi(\mathbf{c}), \mathbf{I}), \tag{1}$$

where $\boldsymbol{\mu}_\psi(\mathbf{c}) \in \mathbb{R}^2$ is produced by a lightweight MLP. This design anchors the latent representation to emotion-dependent locations, allowing smooth traversal across the Warm–Cool and Soft–Hard axes. The remaining latent dimensions $\mathbf{z}_{\text{resid}}$ follow a standard Gaussian prior $p(\mathbf{z}_{\text{resid}}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$, providing stylistic variability that remains independent of emotional input. This hybrid prior formulation enables the model to jointly support emotion control and style diversity.

Auxiliary Alignment Heads. Two linear heads are incorporated during training: $f_{z\to c}$ predicts the emotional condition from $\mathbf{z}_{xy}$, reinforcing emotion-aligned encoding; $f_{c\to z}$ maps the condition back into the latent space for cycle-consistency regularization. These heads are used only for training losses and impose no overhead during inference.

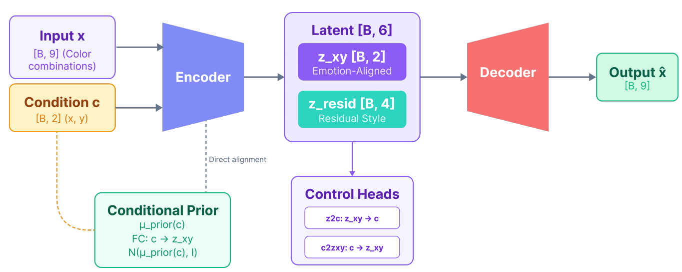

Fig. 2. Architecture of the proposed β-CVAE with a conditional prior. The encoder maps a color palette $\mathbf{x}$ and affective condition $\mathbf{c}$ to a disentangled latent representation $\mathbf{z} = [\mathbf{z}_{xy}, \mathbf{z}_{resid}]$, where $\mathbf{z}_{xy}$ encodes emotion-aligned factors and $\mathbf{z}_{resid}$ captures residual style. A conditional prior anchors $\mathbf{z}_{xy}$ to the affective coordinates during training. Auxiliary control heads and alignment constraints are used only for training to encourage disentanglement and are removed at inference time.

C. Training Objective

The model is trained by maximizing the evidence lower bound (ELBO) with additional regularization terms that encourage continuous emotion control and separation between emotion-related and stylistic factors. The full objective is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{KL}} + \lambda_1 \mathcal{L}_{\text{info}} + \lambda_2 \mathcal{L}_{\text{cycle}} + \lambda_3 L_{\text{direct}} + \lambda_4 \mathcal{L}_{\text{decorr}} \quad (2)$$

Reconstruction Loss. Palette reconstruction is optimized using an element-wise squared error:

$$\mathcal{L}_{\text{recon}} = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \quad (3)$$

KL Divergence with Capacity Control. The posterior is regularized toward the hybrid prior, consisting of a learned conditional prior for $\mathbf{z}_{xy}$ and a standard Gaussian prior for $\mathbf{z}_{\text{resid}}$:

$$\begin{aligned} D_{\text{KL}} = & D_{\text{KL}}\big(q_\phi(\mathbf{z}_{xy}|\mathbf{x},\mathbf{c}) \,\|\, p_\psi(\mathbf{z}_{xy}|\mathbf{c})\big) \\ & + D_{\text{KL}}\big(q_\phi(\mathbf{z}_{\text{resid}}|\mathbf{x},\mathbf{c}) \,\|\, \mathcal{N}(\mathbf{0},\mathbf{I})\big) \end{aligned} \quad (4)$$

To prevent posterior collapse while enabling the model to gradually encode more information, a capacity constraint inspired by β-VAE is applied:

$$\mathcal{L}_{\text{KL}} = \beta \cdot |D_{\text{KL}} - C(t)| \quad (5)$$

The target capacity $C(t)$ increases linearly from zero to a maximum $C_{\max}$ during training.

Information Preservation Loss. A linear head $f_{z\to c}$ predicts the emotional condition from $\mathbf{z}_{xy}$, encouraging emotion-relevant information to be retained in the emotion subspace:

$$\mathcal{L}_{\text{info}} = \|f_{z\to c}(\mathbf{z}_{xy}) - \mathbf{c}\|^2 \quad (6)$$

Cycle Consistency Loss. A reverse mapping $f_{c\to z}$ projects the condition back into the latent space, enforcing bidirectional consistency:

$$\mathcal{L}_{\text{cycle}} = \|f_{z\to c}(f_{c\to z}(\mathbf{c})) - \mathbf{c}\|^2 \quad (7)$$

Direct Alignment Loss. To stabilize emotion-conditioned latent structure, the encoder's mean for the emotion subspace is aligned with the prior mean:

$$L_{\text{direct}} = \|\boldsymbol{\mu}_{xy} - \boldsymbol{\mu}_\psi(\mathbf{c})\|^2 \quad (8)$$

Decorrelation Loss. To enforce independence between style and emotion, we minimize the squared Pearson correlation between the residual codes and each emotion dimension:

$$\mathcal{L}_{\text{decorr}} = \sum_{k=1}^{2} \rho(\mathbf{z}_{\text{resid}}, c_k)^2 \quad (9)$$

where c_k denotes the k-th dimension of the affective coordinates.

D. Implementation Details

The proposed framework was implemented in PyTorch. Both the encoder and decoder adopt symmetric MLP architectures with a hidden dimension of 512. The latent space has a total dimensionality of 6, consisting of an emotion-aligned subspace $\mathbf{z}_{xy} \in \mathbb{R}^2$ and a residual style subspace $\mathbf{z}_{\text{resid}} \in \mathbb{R}^4$. Optimization was carried out using Adam with a learning rate of 2×10^{-4}, weight decay of 1×10^{-5}, and a batch size of 64. Gradient clipping with a threshold of 5.0 was applied to stabilize training. All models were trained for 130 epochs.

For the β-VAE objective, the penalty coefficient was fixed at $\beta = 4.0$. To mitigate posterior collapse, a controlled capacity increase schedule was employed, where the target capacity C was linearly annealed from 0 to a maximum value of $C_{\max} = 4.0$ nats over the first 70% of the total training steps. In addition, a minimum information constraint (free-bits) of 0.05 was applied per latent dimension.

Regarding regularization, static loss weights were set to $\lambda_{\text{cycle}} = 0.1$ and $\lambda_{\text{decorr}} = 0.1$. For emotion-alignment objectives, a staged activation strategy was adopted to stabilize early training dynamics. The information alignment loss λ_{info} and direct alignment loss λ_{direct} were initialized to zero, activated after 20% and 25% of the total training steps respectively, and then linearly increased to their final values of $\lambda_{\text{info}} = 2.0$ and $\lambda_{\text{direct}} = 0.2$ over the subsequent 40% of training.

IV. Dataset

This work is based on Kobayashi's Color Image Scale (CIS), which provides over 1,000 three-color combinations positioned in a two-dimensional affective space defined by the Warm–Cool and Soft–Hard axes. Because the original charts provide only graphical placements, the affective coordinates were digitized by extracting the centroid position of each palette, resulting in 1,132 machine-readable samples.

To address the limited size of the CIS dataset, additional palettes were collected from online design platforms. These five-color schemes were decomposed into overlapping triplets (e.g., 1–2–3, 2–3–4, 3–4–5) to match the CIS format. An XGBoost regressor was trained on the digitized CIS data using an 80/20 train–validation split and achieved an R^2 score of 0.82 on the held-out validation set. The trained regressor was then used to predict affective coordinates for the web-collected triplets, yielding a pseudo-labeled dataset of 53,698 samples. For generative model training, this dataset was further randomly split into 80% for training and 20% for testing.

V. Experiments

A. Ablation Study on Latent Alignment

To evaluate the contribution of the auxiliary loss terms, we conducted an ablation study by training a baseline model in which all auxiliary losses were removed ($\lambda_{\text{info}} = \lambda_{\text{cycle}} = \lambda_{\text{direct}} = \lambda_{\text{decorr}} = 0$). This results in a standard β-CVAE with a learned conditional prior, trained only with reconstruction and KL objectives.

We measured the correspondence between the latent dimensions and the affective coordinates using Spearman's rank correlation coefficient (ρ). As shown in Table I, the baseline model exhibits noticeably weaker correlations. In comparison, the full model achieves high correlations of 0.926 on the Warm–Cool axis and 0.954 on the Soft–Hard axis, indicating that the auxiliary objectives play an important role in shaping the latent variables with respect to the affective dimensions.

TABLE I

ABLATION STUDY ON LATENT–EMOTION CORRELATION. THE BASELINE MODEL EXCLUDES ALL AUXILIARY LOSSES ($\lambda_{\text{AUX}} = 0$).

Method	Warm–Cool (z_1)	Soft–Hard (z_2)
Baseline ($\lambda_{\text{aux}} = 0$)	-0.016	0.200
Full Model	$\mathbf{0.926}$	$\mathbf{0.954}$

B. Emotion-Controllable Generation

To evaluate whether the model can generate palettes that consistently reflect specified emotional coordinates, we conducted an emotion-conditioned sampling experiment. For each target affective coordinate, the emotion-related latent variables $\mathbf{z}_{xy}$ were fixed according to the learned conditional prior, while the residual latent variables $\mathbf{z}_{\text{resid}}$ were randomly sampled to produce stylistic variations.

Figure 3 shows generated palettes for four representative affective targets: *Fresh*, *Pretty*, *Stylish*, and *Dynamic*. Within each condition, the generated palettes display coherent emotional characteristics consistent with their target coordinates. For example, *Fresh* palettes tend to exhibit higher-luminance cool tones, whereas *Dynamic* palettes often contain saturated warm hues. At the same time, variations across samples demonstrate that the residual latent space captures stylistic diversity independent of the affective dimensions.

These results indicate that the proposed model can control emotional intent while maintaining generative variability, enabling practical use in emotion-aware design workflows that require both consistency and diversity.

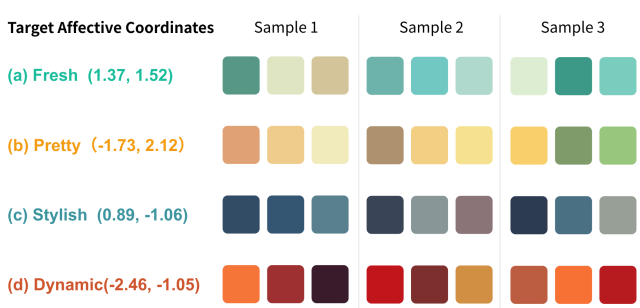

Fig. 3. Emotion-controllable color palette generation. For each affective target, the emotion-related code $\mathbf{z}_{xy}$ is fixed to the target coordinate, while the style code $\mathbf{z}_{\text{resid}}$ is randomly sampled. The model generates diverse palettes that remain consistent with the semantic characteristics of the target emotion.

C. Style-Consistent Emotion Transfer

To evaluate the separation between emotional and stylistic representations, we conducted a style transfer experiment in which the residual style vector $\mathbf{z}_{\text{resid}}$ from a source palette was combined with different target emotional coordinates.

Figure 4 presents the outputs. The source palette (*Modest*) features a distinctive light-to-dark luminance progression, and this structure is consistently preserved after transfer. For the four target emotions (*Vivid*, *Smart*, *Sharp*, *Free*), the model adjusts hue and saturation to match the intended affect while maintaining the original luminance ordering and contrast relationships.

These results suggest that $\mathbf{z}_{\text{resid}}$ encodes stable stylistic cues—such as luminance arrangement and basic compositional patterns—that remain intact regardless of the target emotion. This enables emotional attributes, controlled by $\mathbf{z}_{xy}$, to be modified without disrupting the palette's underlying structure.

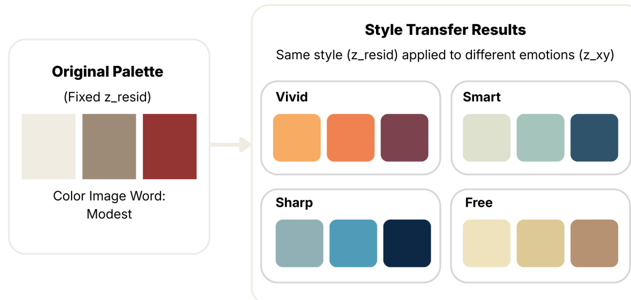

Fig. 4. Style-consistent emotion transfer. (Left) A source palette with a *Modest* impression, characterized by a light-to-dark luminance arrangement. (Right) Transfer results applying the same style code $\mathbf{z}_{\text{resid}}$ to four different target emotions. The model adapts the chromatic properties to the new emotions while strictly preserving the luminance hierarchy and compositional structure of the source.

D. Latent Space Traversal

To examine the continuity and organization of the learned emotion-related representations, we performed a latent interpolation experiment. The emotion-aligned subspace $\mathbf{z}_{xy}$ was traversed over a two-dimensional grid, while the residual style code $\mathbf{z}_{\text{resid}}$ was fixed at the prior mean.

Figure 5 visualizes the resulting palette manifold. The generated palettes exhibit smooth chromatic transitions across the affective plane. Moving horizontally and vertically across the grid yields gradual changes in hue, saturation, and luminance that align with the semantic structure of the Warm–Cool and Soft–Hard dimensions. For example, palettes in the upper-left region correspond to warm and high-contrast combinations, whereas palettes in the lower-right region transition toward cooler and softer tones.

The absence of abrupt discontinuities or degenerate outputs indicates that the model has learned a continuous and well-structured color–emotion manifold. This smooth traversal demonstrates that the proposed β-CVAE supports meaningful interpolation between affective states, enabling controllable navigation through the latent emotion space.

E. Application to Multimedia Interface Design

To illustrate the practical applicability of the proposed emotion-controllable palette generator in multimedia scenarios, we applied model-generated palettes to a mobile travel application interface. As shown in Figure 6, the layout, typography, and content are held constant across all variants; only

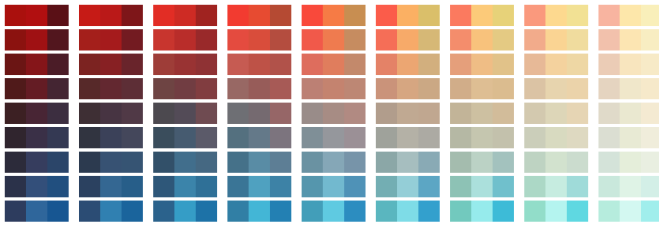

Fig. 5. Visualization of the learned color–emotion manifold. Each palette is generated by traversing the emotion-related latent subspace $\mathbf{z}_{xy}$ while keeping $\mathbf{z}_{\text{resid}}$ fixed. The smooth transitions from warm/intense (top-left) to cool/soft (bottom-right) indicate that the model captures a continuous and semantically organized affective space.

the color palette is replaced according to different affective targets.

The resulting interfaces exhibit clearly distinguishable affective impressions. For example, the *Fresh* palette produces a bright and welcoming tone, whereas the *Authoritative* palette conveys a more formal and structured atmosphere. Similarly, *Merry* yields a warm and cheerful presentation, while *Proper* results in a restrained and elegant aesthetic. Despite these perceptual differences, the underlying UI structure remains unchanged, demonstrating that the proposed model can be seamlessly integrated into interface design pipelines to support emotion-aware visual customization in multimedia applications.

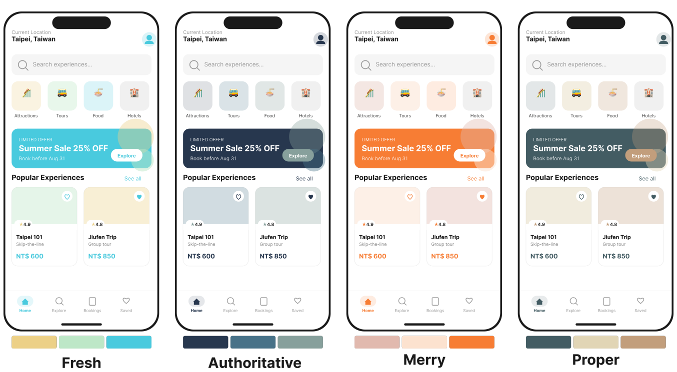

Fig. 6. Application to Interface Design. Four model-generated palettes applied to a fixed mobile travel interface. By varying only the latent affective coordinates, the model successfully alters the visual identity of the application across distinct emotional targets (*Fresh*, *Authoritative*, *Merry*, *Proper*) while maintaining strict structural consistency.

F. Application to Data Visualization

We also examined whether the generated palettes are suitable for analytical graphics, where clarity and category distinction are critical. Figure 7 applies four affective palettes to identical line and area chart layouts. Only the color assignments vary across conditions.

Across all examples, the charts remain readable, with consistent contrast between series and no ambiguity in class differentiation. This demonstrates that the proposed model can generate palettes that are not only affectively meaningful but also meet basic usability requirements for data visualization, enabling their use in dashboards and multimedia analytics tools.

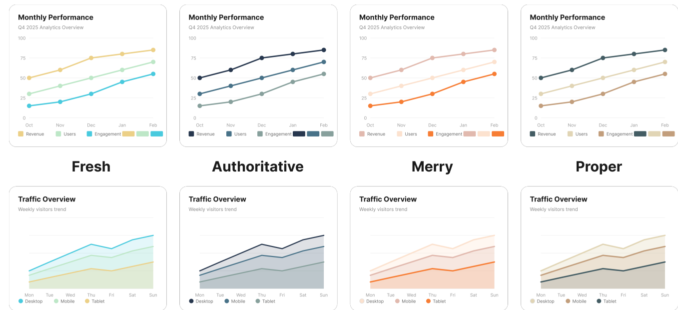

Fig. 7. Use of model-generated palettes in data visualization. Each chart uses identical data and layout, differing only in the applied affective palette. All palettes maintain sufficient contrast and category separability, indicating suitability for analytical graphics.

VI. CONCLUSION

In this work, we presented a β-CVAE with a learned conditional prior for emotion-controllable color palette generation. By integrating Kobayashi's Color Image Scale with a disentangled latent design, the model enables independent manipulation of emotional intent and stylistic characteristics. The pseudo-labeled dataset further expands the affective coverage of the training data, supporting robust learning in the two-dimensional emotion space.

Experimental results show that the model exhibits strong latent correlations with the target affective coordinates and generates palettes that remain semantically consistent while allowing substantial stylistic diversity. The framework supports style-consistent emotion transfer and smooth traversal across the learned affective manifold, and applications in UI design and data visualization demonstrate its practical utility in multimedia contexts.

Future work will explore extending the framework to support variable-length palettes, as well as incorporating multimodal conditioning—such as text prompts, images, or style exemplars—to enable more flexible and interactive affect-aware design tools. Additionally, conducting user studies and pursuing finer disentanglement of stylistic attributes, including luminance, saturation, and ordering patterns, represent promising directions for improving interpretability and expanding the model's applicability. Ultimately, this work contributes to the development of AI-driven multimedia systems that benefit from controllable color semantics and adaptive, emotion-aware visual design tools.

REFERENCES

[1] R. Liang, F. N. U. Feifan, Y. Liang, and Z. Ye, "Emotion-aware interface adaptation in mobile applications based on color psychology and multimodal user state recognition," *Frontiers in Artificial Intelligence Research*, vol. 2, no. 1, pp. 51–57, 2025.

[2] M. Shagyrov and P. Shamoi, "Color and sentiment: A study of emotion-based color palettes in marketing," in Proc. Joint 13th Int. Conf. Soft Comput. Intell. Syst. and 25th Int. Symp. Adv. Intell. Syst. (SCIS&ISIS), Nov. 2024, pp. 1–7.

[3] P. Weingerl, “Automated image-based user interface color theme generation,” Applied Sciences, vol. 14, no. 7, p. 2850, 2024.
[4] S. Kobayashi, “The aim and method of the color image scale,” Color Res. Appl., vol. 6, no. 2, pp. 93–107, 1981.
[5] S. Kobayashi, *Color Image Scale*, L. Matsunaga, Trans. Tokyo: Kodansha International, 1991.
[6] T. Lertrusdachakul, K. Ruxpaitoon, and K. Thiptarajan, ”Color palette extraction by using modified k-means clustering,” in Proc. 2019 7th Int. Electrical Engineering Congress (iEECON), Mar. 2019, pp. 1–4.
[7] S. C. Huang, “An efficient palette generation method for color image quantization,” Applied Sciences, vol. 11, no. 3, p. 1043, 2021.
[8] L. Xu, D. Liu, S. J. Park, and S. Lee, “Unified Color Harmony Model,” Color Research & Application, vol. 50, no. 4, pp. 346–371, 2025.
[9] A. Moussa and H. Watanabe, “Generation and extraction of color palettes with adversarial variational auto-encoders,” in Proc. 6th Int. Congr. Information and Communication Technology (ICICT), London, U.K., Sep. 2021, pp. 889–897.
[10] T. R. Chou and J. Y. Shao, “Color palette generation of mixed color images using autoencoder,” Sensors & Materials, vol. 36, 2024.
[11] M. X. Zou and T. R. Chou, “Color combination adaptation to hue spread index using an autoencoder,” in IET Conf. Proc. CP929, vol. 2025, no. 15, Stevenage, U.K., Apr. 2025, pp. 68–70.
[12] M. Solli and R. Lenz, “Color semantics for image indexing,” in Proc. CGIV 2010/MCS’10: 5th Eur. Conf. Colour in Graphics, Imaging, and Vision and 12th Int. Symp. Multispectral Colour Science, Joensuu, Finland, Jun. 14–17, 2010, pp. 353–358.
[13] D. Kang, H. Shim, and K. Yoon, “Mood from painting: Estimating the mood of painting by using color image scale,” in Proc. 21st Korea–Japan Joint Workshop Frontiers of Computer Vision (FCV), Jan. 2015, pp. 1–4.
[14] D. Kang, H. Shim, and K. Yoon, “A method for extracting emotion using colors comprise the painting image,” Multimedia Tools and Applications, vol. 77, no. 4, pp. 4985–5002, 2018.
[15] D. Kang, H. Shim, and K. Yoon, “A method for extracting emotion using colors comprise the painting image,” Multimedia Tools and Applications, vol. 77, no. 4, pp. 4985–5002, 2018.
[16] S. Liu, M. Tao, Y. Huang, C. Wang, and C. Li, “Image-driven harmonious color palette generation for diverse information visualization,” IEEE Trans. Visualization Comput. Graphics, vol. 30, no. 7, pp. 3089–3103, 2022.
[17] L. Xu, S. J. Park, and S. Lee, “Color2Vec: web-based modeling of word-color association with sociocultural contexts,” ACM Transactions on Computer-Human Interaction, vol. 30, no. 4, pp. 1–29, 2023.
[18] J. Hu, Y. He, T. Liang, C. Wang, and C. Li, “Music2Palette: Emotion-aligned color palette generation via cross-modal representation learning,” in Proc. 33rd ACM Int. Conf. Multimedia, 2025, pp. 6615–6624.
[19] Y. Gao, J. Liang, and J. Yang, “Color palette generation from digital images: A review,” Color Research & Application, vol. 50, no. 3, pp. 250–265, 2025.