

# Culture-inspired Multi-modal Color Palette Generation and Colorization: A Chinese Youth Subculture Case

Yufan Li  
Tongji University  
Shanghai, China  
1831982@tongji.edu.cn

Jinggang Zhuo  
Tongji University  
Shanghai, China  
zhuojg1519@tongji.edu.cn

Ling Fan  
Tongji University; Tezign.com Ltd.  
Shanghai, China  
lfan@tongji.edu.cn

Harry Jiannan Wang  
University of Delaware  
Newark, DE, USA  
hjiwang@udel.edu

**Abstract**—Color is an essential component of graphic design, acting not only as a visual factor but also carrying cultural implications. However, existing research on algorithmic color palette generation and colorization largely ignores the cultural aspect. In this paper, we contribute to this line of research by first constructing a unique color dataset inspired by a specific culture, i.e., Chinese Youth Subculture (CYS), which is an vibrant and trending cultural group especially for the Gen Z population. We show that the colors used in CYS have special aesthetic and semantic characteristics that are different from generic color theory. We then develop an interactive multi-modal generative framework to create CYS-styled color palettes, which can be used to put a CYS twist on images using our automatic colorization model. Our framework is illustrated via a demo system designed with the human-in-the-loop principle that constantly provides feedback to our algorithms. User studies are also conducted to evaluate our generation results.

**Index Terms**—color palette generation, Chinese youth subculture, image colorization, conditional generative adversarial network

## I. INTRODUCTION

Color is an essential component of graphic design, acting as a visual factor while also conveying human inner emotions. These emotions and impacts are strongly influenced by the cultural implications behind colors. For example, the combination of red and green is unpleasant according to traditional Chinese color theory but this same color combination represents a cool and rebellious style for Chinese Youth Subculture (CYS) groups as illustrated by a number of posters found in popular CYS websites (Fig. 1). In addition, colors may have different semantic meanings for different cultures. For instance, the color directly related to money is green in Western culture, which is often gold and red in Eastern culture. Therefore, culture is an important factor that cannot be ignored when studying color semantics.

Color has also been an active research topic in the machine learning and AI areas [9] [10] [11] [12]. Many design tool companies have intelligent color design as one of their main product features, such as the Adobe Color tool by Adobe (color.adobe.com) and the Color Match tool by AI startup FacetAI (facet.ai). However, existing researches and tools on algorithmic color palette generation and colorization [8]

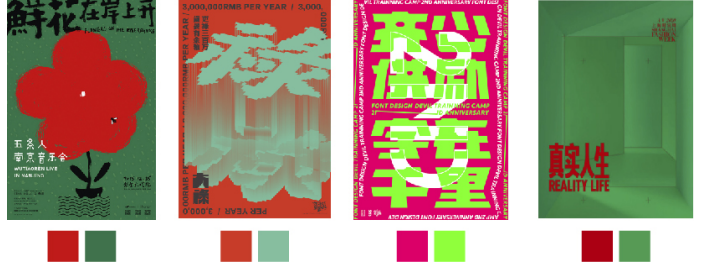


Fig. 1. Examples of ‘red with green’ designs in Chinese youth subculture

[13] largely ignore the cultural aspect. In this paper, we aim to extend existing work by incorporating the culture aspect. Created by and disseminated among Chinese young population, Chinese Youth Subculture has the unique balance of confliction and cooperation with the mainstream culture and its parent culture, symbolizing the identity of the young people in China [1]. We want to focus on the color research of the Chinese Youth Subculture for a number of reasons. First, based on our interviews of a number of graphic designers, who have done projects related to CYS, the colors used in CYS have special aesthetic and semantic characteristics that are different from existing color theories, which calls for more research. Second, previous research [2] has proved that the integration of subcultural style into the mainstream style is the driving force of design innovation, which keeps the mainstream culture alive and fresh. Finally, CYS is an vibrant and trending cultural group especially for the Gen Z population, which is one of the main driving forces of the 21st century economy and is of great interest to many businesses. Therefore, our study of CYS color system can have broader impact and practical design and business implications.

Besides color palette, there are other important factors that can affect the overall CYS style of the images, such as image elements, composition, text, and context. Take Fig. 1 as an example, with a similar palette of red with green, the first poster has a red flower on a green background while the second poster has a green Chinese character on a red background, indicating different roles of elements and composition. The

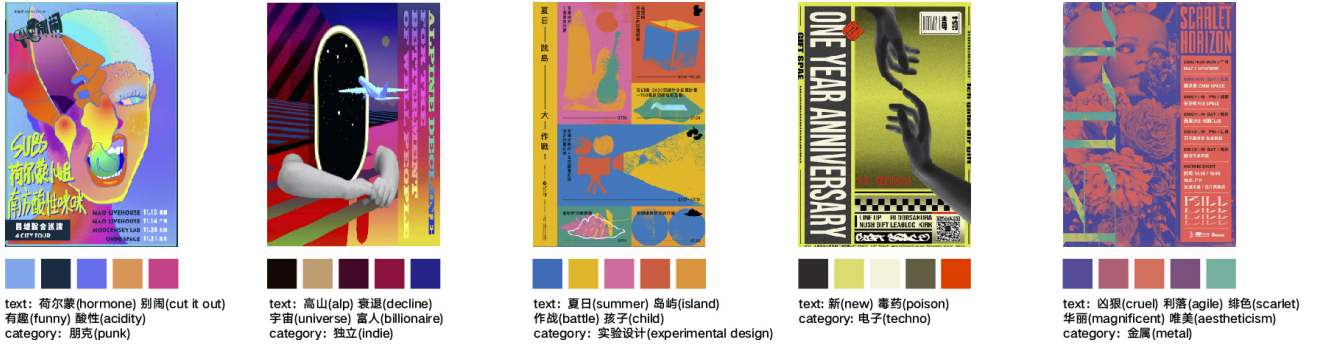


Fig. 2. CYS dataset examples

text in the image also has effect on the color, e.g., the text of the first poster is “flower blossoms on a river bank”, which implies the flower is a blossoming bright red flower on a fresh green bank. There is also additional text description of the poster that influences the CYS style design as we discuss in our dataset later. In addition, the context of the image also correlate with the CYS style. For instance, the first poster is for a live show of a rock band, which is casual and entertaining, while the fourth poster that is a commercial poster aimed at CYS-groups, which is more rigid and serious.

In this paper, we address the aforementioned CYS color research challenges by proposing an open-source culture-inspired multi-modal color palette generation and colorization framework. Our contributions are three-fold:

- We construct a unique open dataset of images, color palettes, text and categories for CYS to facilitate culture related color research.
- We introduce an innovative multi-modal generative architecture for culture-inspired color palette generation and colorization.
- We develop a demo system to illustrate our framework, which incorporates the human-in-the-loop principle to enable automatic system evolution.

The dataset, code, and demo have been open sourced.<sup>1</sup>

## II. RELATED WORK

Color and culture is an important branch in the study of art. The book “The Theory of Color and Culture” [3] comprehensively expounds on the relationship between color and culture. Sharkhuu et al. studied the use of colors in Buddhist cultures in different countries based on Buddhist pictures [4]. Kim et al. paid attention to the color of photos posted on the social platform Instagram and used colorfulness, color diversity, and color harmony to study the characteristics of different user groups [5]. Zhe et al. used the K-Means clustering method to extract the typical colors of traditional Chinese embroidery and studied the color characteristics of traditional Chinese cultural phenomena [6]. No study has made further algorithmic color generation attempt for a specific culture.

Color is not only an objective physical attribute but also strongly arouses people’s semantic associations. Koshiyaba [7] was the first to quantify the color-semantic relationship by using a color semantic space containing 180 adjectives expressing 1170 3-color palettes. Bahng et al. [8] proposed a GAN-based model to generate color palette based on a dataset with 4,312 words and 10183 text-palette pairs. In addition, many practical color semantic online tools have appeared, such as Adobe Color and Color Hex. Researchers have also used different interactive coloring methods for gray-scale images, such as reference images method [9] [10] and users’ stroke hints approach [11]. Text-based colorization has also been studied [8] [13]. However, no studies have considered the impact of culture on the text-palette relationship.

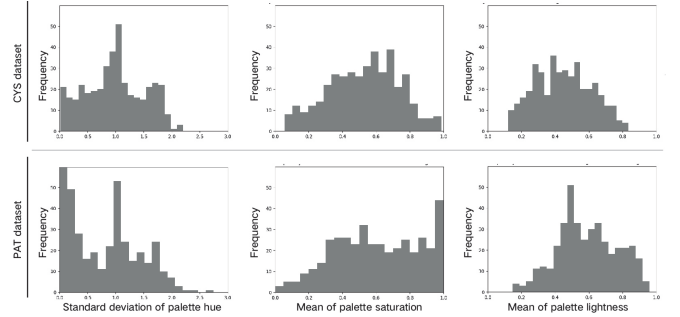


Fig. 3. Comparison between CYS dataset and PAT dataset

## III. CYS COLOR DATASET

We manually constructed the CYS color dataset, which contains 1263 images with corresponding 5-color palette, descriptive text, and category. Fig. 2 shows five examples from the dataset and the dataset development process is discussed next. Based on the interviews of designers who have done CYS-related projects, we obtain a list of websites targeting CYS population and manually collected 7051 images from those website, which are posters, album covers, and illustrations. Then, three designers manually selected 1263 images that have strong CYS style based on their judgement. From the web pages where the images are collected, we synthesized a list of keywords and phrases related to the CYS expression of the images by removing text irrelevant to color and CYS,

<sup>1</sup><https://github.com/teziglab/subculture-colorization>

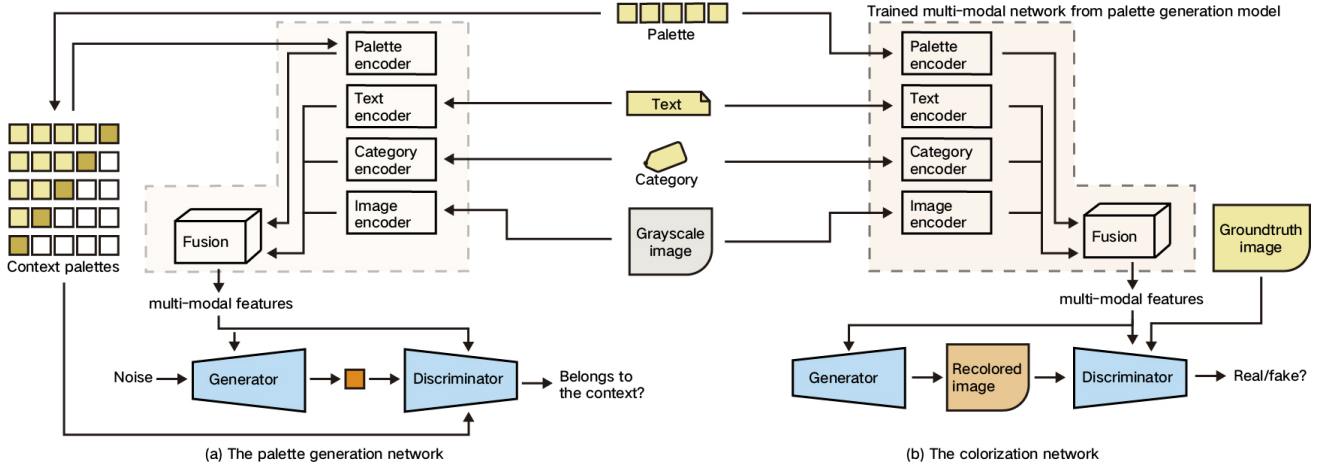


Fig. 4. The training structure of our framework

resulting in 2535 unique Chinese adjectives, nouns, and verbs. We also collected the categories data from the web pages, includes 14 categories, such as punk, hiphop, techno, etc. For each image, we first extract 10 colors using existing clustering based algorithm and then let designer select 5 colors to form the color palette according to the following rules: 1) removing similar colors that cannot be easily separated, 2) selecting the color combination that best represents the visual effect according to image elements and composition, and 3) selecting the colors that best reflects CYS and rank them accordingly. This step is critical because the clustering algorithm essentially ranks the colors based on their proportions in the images, which may miss many important colors for CYS style if we only choose the first 5 colors.

In order to show that CYS dataset is indeed different from generic color datasets, we conducted a comparison with the PAT dataset from Bahng et al. [8], which does not consider the cultural aspect. We randomly sampled 400 images from each dataset and calculated the overall distributions of the HSL (hue, saturation, lightness) values as shown in Fig. 3. The plot and T test results<sup>2</sup> clearly show the differences, e.g., CYS colors tend to have more diverse hue and lower saturation and lightness.

#### IV. FRAMEWORK AND DEMO

As shown in Fig. 4, our framework includes two separately trained networks: the color palette generation network and the colorization network. The first network is trained through a conditional GAN (cGAN) with a multi-modal input to generate CYS color palettes. Another cGAN in the second model is trained to color the input images according to the palette generated by the first network.

In order to augment the color dataset to provide more training samples, we leverage the relationships among the colors in the palette and ask the discriminator to determine whether the generated color is considered the next color in the

palette. Previous colors are fused together by a fusion function with the other three inputs to form the multi-modal context, which is fed to the cGAN as a condition to predict the next color. Finally, 5 generated colors are concatenated and output as the color palette.

We assign different weights to different elements in the multi-modality module using a parameter  $\lambda$  ( $\lambda_{text} = 0.5$ ,  $\lambda_{image} = 0.4$ ,  $\lambda_{category} = 0.1$  are default settings). The corresponding encoders are  $E_{image}$ ,  $E_{text}$  and  $E_{category}$  and the combined context of grayscale image, text and category is defined as  $c_1 = \sum_c \lambda_c E_c(x_c)$ ,  $c \in \{image, text, category\}$ . Given the encoded palette  $c_2 = E_{palette}(palette)$  and final multi-modal context is defined as  $y = f(c_1, c_2)$ . We follow the least-squares GAN (LSGAN) [14] to formulate the loss functions as:

$$L_D = \alpha(D(x, y) - 1)^2 + (1 - \alpha)(D(G(z, y), y))^2$$

$$L_G = \alpha(D(G(z, y), y) - 1)^2$$

where  $\alpha = 0.5$  is hyperparameter,  $x$  is the ground truth,  $y$  is the multi-modal context, and  $z$  is random noise.

The colorization network is structurally similar to the palette generation network, but use the generated 5-color palette as the input and different set of weights for the multi-modality context ( $\lambda_{text} = 0.3$ ,  $\lambda_{image} = 0.6$ , and  $\lambda_{category} = 0.1$ ).

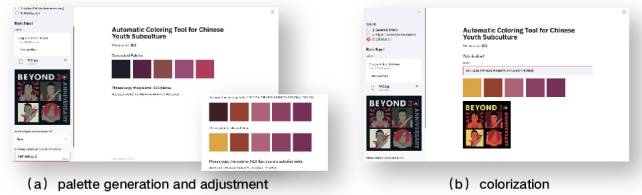


Fig. 5. The demo system

As shown in Fig. 5, we have developed a demo system to materialize our framework<sup>3</sup>, where users can obtain a

<sup>2</sup>p-values for hue standard deviation, lightness mean, and saturation mean are  $1.11 \times 10^{-7}$ ,  $1.93 \times 10^{-6}$ ,  $2.42 \times 10^{-29}$  respectively.

<sup>3</sup>video of the demo: <https://youtu.be/fLPwr-oX0ds>

image that is colored with the CYS style by following three steps: 1) After entering the text description, uploading the grayscale image and selecting the category, the user can get the generated CYS-styled palette, 2) the color palette can be adjusted according to the user’s preference, which is recorded by the system to improve the algorithm, 3) the image uploaded in the first step will be coloredized by our system to add CYS twist.

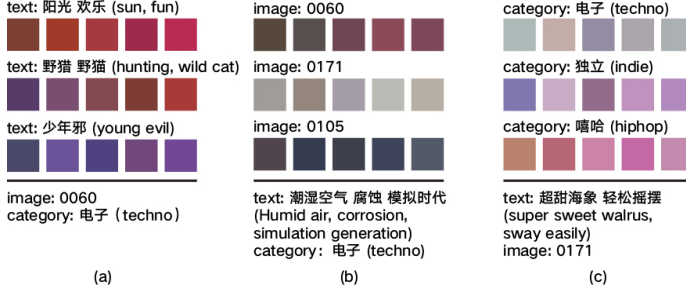


Fig. 6. Evaluation of multi-modality (a) different text with controlled image and category; (b) different image with controlled text and category; (c) different category with controlled text and image.

## V. EVALUATION

We want to evaluate our framework from two perspectives, i.e., whether the multi-modal context affects the generation results, and whether the colorized images have CYS styles.

As shown in Fig. 6, we did three sets of experiments for the first perspective: a) changing text while fixing image and category, b) changing image while fixing text and category, and c) changing category while fixing image and text. The results show that each of the multi-modal context components has direct impact on the generation results and our generation model has excellent controlled diversity. For example, in Fig. 6 (a), after changing the text from “sun, fun” to “evil”, the generated colors are apparently darker to reflect effect of the semantic textual changes.

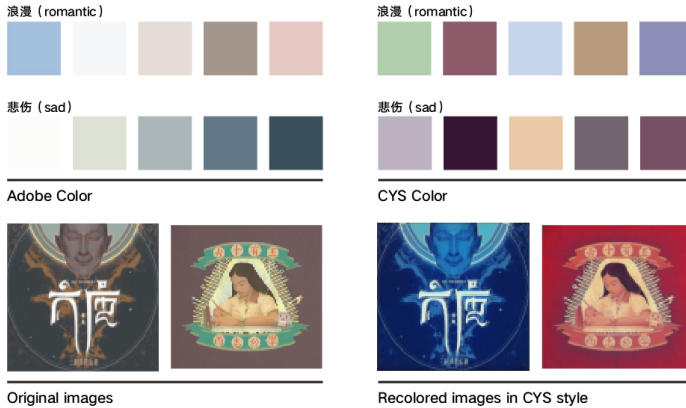


Fig. 7. User study examples

We design a user study to evaluate from the second perspective. We choose 20 keywords and search the Adobe Color website to get the cultural neutral color palettes. Then, we

generate our color palette using the keywords, 20 cultural neutral images that are from a graphic design website and category “indie” to form the color palette pairs. Then, we use the generated color palettes to colorize 20 images and form the image pairs. The palette pairs and image pairs are shown in Fig. 7. We then ask four designers to review the color palette and image pairs and determine which one is more CYS styled. The results show that 75% of our palettes and 76% of recolored images are considered to be more in line with the CYS-style, which verifies that our framework learned the color specificity of CYS.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a color generation and colorization framework based on a manually collected Chinese Youth Subculture dataset. A demo system of the framework had been developed and evaluated via a set of experiments and a user study. To the best of our knowledge, our framework is the first culture-inspired approach to automatic color generation and colorization. Note that our framework is not limited to CYS culture and can be used for any culture if trained using culture specific data. We are in the process of expanding the dataset and extending our framework by calibrating additional design features that may have important influence to culture-based color theory.

## REFERENCES

- [1] C. Yan, “The historical evolution of youth subculture in China.” Chinese Youth Social Sciences 38.03, 2019, pp.83–91.
- [2] X. Wang, “The power of the edge.” Central Academy of Fine Arts, PhD dissertation, 2014.
- [3] K. Zhang, “The theory of color and culture.” Zhejiang University Press, 2017.
- [4] M. Sharkhuu and I. K. Choi, “A comparative study on the color image of buddhism in Korea, Mongolia, China and Japan.” 2007.
- [5] J. H. Kim and Y. Kim, “Instagram user characteristics and the color of their photos: Colorfulness, color diversity, and color harmony,” Inf. Process. Manag., vol. 56, no. 4, pp. 1494–1505, Jul. 2019.
- [6] Z. Zhe, Q. Wang, and Y. Xing, “Research on big data analysis technology of chinese traditional culture yue embroidery color network,” J. Phys. Conf. Ser., vol. 1345, p. 022021, Nov. 2019.
- [7] S. Kobayashi, “The aim and method of the color image scale,” Color Res. Appl., vol. 6, no. 2, pp. 93–107, 1981.
- [8] H. Bahng et al., “Coloring with words: guiding image colorization through text-based palette generation,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 431–447.
- [9] Z. Xu, T. Wang, F. Fang, Y. Sheng, and G. Zhang, “Stylization-based architecture for fast deep exemplar colorization,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9363–9372.
- [10] P. Lu, J. Yu, X. Peng, Z. Zhao, and X. Wang, “Gray2ColorNet: transfer more colors from reference image,” in Proceedings of the 28th ACM International Conference on Multimedia, New York, NY, USA: Association for Computing Machinery, 2020, pp. 3210–3218.
- [11] H. Ren, J. Li, and N. Gao, “Two-stage sketch colorization with color parsing,” IEEE Access, vol. 8, pp. 44599–44610, 2020.
- [12] H. Chang, O. Fried, Y. Liu, S. DiVerdi, and A. Finkelstein, “Palette-based photo recoloring,” ACM Trans. Graph., vol. 34, no. 4, p. 139:1–139:11, Jul. 2015.
- [13] J. Li, H. Xiao, D. Tan, M. Zhang, and Y. Liu, “Image colorization based on texture by using of CNN,” in 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC), Jul. 2019, pp. 167–171.
- [14] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. Paul Smolley, “Least squares generative adversarial networks,” 2017, pp. 2794–2802.