Personalized Color Vision Deficiency Friendly Image Generation

A THESIS SUBMITTED TO
THE FACULTY OF ENGINEERING
OF THE UNIVERSITY OF SYDNEY
IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF PHILOSOPHY

THE UNIVERSITY OF
SYDNEY

SHUYI JIANG

Supervisor: Dr. Chang Xu

School of Computer Science

Faculty of Engineering

The University of Sydney

Australia

11 December 2023

Authorship Attribution Statement

The work presented in this thesis is published as 'Personalized Image Generation for Color Vision Deficiency Population' in the International Conference on Computer Vision (ICCV), 2023. The student designed the study, implemented the code, analyzed the data, and wrote the draft of the paper.

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Student: Shuyi Jiang

Date: 09 Oct 2023

As the supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Supervisor: Dr. Chang Xu

Date: 09 Oct 2023

Personalized Color Vision Deficiency Friendly Image Generation

Shuyi Jiang (Email: sjia6973@uni.sydney.edu.au)

Supervisor: Dr. Chang Xu

School of Computer Science

Faculty of Engineering

The University of Sydney

Copyright in Relation to This Thesis

Statement of Original Authorship

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

To Those Whom I love & Those Who Love Me.

Abstract

Approximately, 350 million people, a proportion of 8%, suffer from color vision deficiency (CVD). While image generation algorithms have been highly successful in synthesizing high-quality images, CVD populations are unintentionally excluded from target users and have difficulties understanding the generated images as normal viewers do. Although a straightforward baseline can be formed by combining generation models and recolor compensation methods as the post-processing, the CVD friendliness of the result images is still limited since the input image content of recolor methods is not CVD-oriented and will be fixed during the recolor compensation process. Besides, the CVD populations can not be fully served since the varying degrees of CVD are often neglected in recoloring methods.

To address these issues, we introduce a personalized CVD-friendly image generation algorithm distinguished by two key features: (i) the ability to produce CVD-oriented images that align with the needs of CVD populations, and (ii) the capacity to generate continuous personalized images for people with various CVD degrees through disentangling the color representation based on a triple-latent structure. Quantitative and qualitative experiments affirm the effectiveness of our proposed image generation model, demonstrating its practicality and superior performance compared to standard generation models and combination baselines across multiple datasets.

Keywords

Image Generation, Generative Adversarial Network, Disentanglement, Color Vision Deficiency, Healthcare AI.

Acknowledgements

I wish to extend my heartfelt appreciation to the many individuals whose unwavering support, guidance, and encouragement have played an indispensable role in my MPhil journey. Their assistance has been instrumental in bringing this thesis to fruition.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Chang Xu, for his unwavering guidance, invaluable insights, and unwavering support throughout my journey as an MPhil student. I would also like to extend my heartfelt appreciation to the postdoctoral researcher Dr.Daochang Liu. His collaboration, expertise, and willingness to share his knowledge have enriched my research experience. Their mentorship, dedication, and encouragement have been instrumental in shaping my research and academic growth. I am truly fortunate to have had such two exceptional mentors.

I'm also appreciative of my colleagues and fellow students for the camaraderie we've shared, the insightful discussions, and the vibrant academic atmosphere that defined our journey. The time we dedicated to pursuing our academic aspirations will forever remain etched in my memory.

Lastly, I would like to extend my heartfelt thanks to Linxin Sun and Linyan Dang for their unwavering encouragement, understanding, and patience throughout this challenging academic journey. The nights we shared in Rhodes hold a special place in my heart and will remain treasured memories for a lifetime. Their love and support have been my constant source of strength.

Table of Contents

CHAPTER 1

Introduction

In this section, we will first introduce the growing interest in image generation and the typical generative model structures. Additionally, the background of color vision deficiencies (CVD) and the gamut of individuals with CVD will be illustrated. Then, we will provide insights into the current GAPs and our motivation towards this work. Lastly, our main contribution will be summarized.

1.1 Introduction of Image Generation

In an era dominated by digital media and visual storytelling, the ability to create, manipulate, and generate images has become an indispensable facet of modern technology and communication. With the continuous advancement of deep learning technology, many outstanding image-generation algorithms have been introduced. Starting from the early Variational Autoencoders (VAEs) [1, 2], progressing to Generative Adversarial Networks (GANs) [3, 4, 5, 6, 7], and most recently, the Diffusion Models (DMs) [8, 9].

Specifically, the common architecture of a Variational Autoencoder (VAE) comprises two primary components: an encoder and a decoder. The encoder is responsible for encoding the input training image into a probability distribution in a lower-dimensional space, while the decoder's role is to acquire the ability to resemble the image from this distribution [2]. Though VAEs can generate new data samples that resemble the training data, they often tend to generate blurry and may struggle to capture highly complex data distributions [10]. Similarly, While Generative Adversarial Networks (GANs) and Diffusion Models (DMs) are capable of generating high-quality images, they each have their respective limitations.

In the context of Generative Adversarial Networks (GANs), there are two key components: a generator and a discriminator. The generator's primary task is to produce high-fidelity images with the objective of deceiving the discriminator. Conversely, the discriminator is tasked with distinguishing between genuine and synthesized images. Though high-quality images can be generated through GANs, the training process can encounter challenges such as instability and mode collapse [11]. On the counterpart, the typical training process of Diffusion Models (DMs) involves both a noising process and a denoising process. During the noising process, noise is introduced into the input training data following a Gaussian distribution. Subsequently, the model learns to predict and remove this noise during the denoising process. The following papers [11, 12, 13] introduce the classifier guidance and classifier-free generation training strategies to enhance the conditional generation and further incorporate the CLIP text encoder [14] to enhance the interactive generation process. Despite their capability to achieve exceptional performance, DMs are hindered by substantial computational costs, encompassing time, and resource requirements which can limit their efficiency. In pursuit of a balance between image generation quality and computational efficiency, we have opted to select the GAN as our primary baseline.

1.2 Introduction of Color Vision Deficiency

Human vision relies on three types of cone cells, known as L-cones (sensitive to long-wavelength light, including red and orange), M-cones (sensitive to medium-wavelength light, including green and cyan), and S-cones (sensitive to short-wavelength light, including blue and purple). Variations in the spectral sensitivity of these cones result in different forms of color vision deficiency (CVD), including protan for abnormal L-cones, deutan for abnormal M-cones, and tritan for abnormal S-cones. Colloquially, these conditions are often referred to as red-weak, green-weak, and blue-weak or even red-blind, green-blind, and blue-blind.

This variation in cone sensitivity has profound effects on how individuals perceive colors, leading to different color gamuts for people with CVD, as depicted in Fig. 1.1. The severity of CVD, denoted as δs, can be estimated as a percentage based on the shift $\Delta\lambda$ relative to

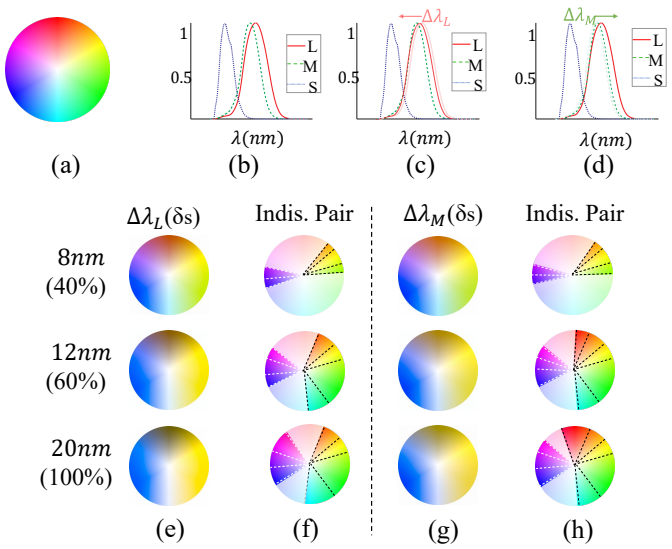

FIGURE 1.1. CVD color gamut and cone curves. Compared to the normal viewers' (a) and (b) [15], (c) and (d) are CVD cone curves with a shift of $\Delta\lambda_L$ and $\Delta\lambda_M$; (e) and (g) are the perceptual color gamut under varying severity δs; In (f) and (h), the gamut is indistinguishable between every two dotted lines with the same color. The white area is distinct to individuals with CVD.

the standard sensitivity curve, typically calibrated to a 20 nm shift. A 20 nm shift represents total cone dysfunction, akin to dichromacy (single-color-blindness). This shift is measured separately for L-cones and M-cones, denoted as $\Delta\lambda_L$ and $\Delta\lambda_M$, respectively.

Roughly 350 million people,constituting approximately 8% of the population, contend with color vision deficiency (CVD), and as of now, no efficient medical cure has been developed [16]. Nevertheless, this sizeable population is unintentionally excluded as the target audience of image generation, underscoring the need for the development of an image generation model that is friendly to a broader range of viewers, including those with color vision deficiency.

1.3 GAPs and Motivations in CVD-Friendly Generation

The term "CVD-friendly" refers to an image that can be comprehended similarly by both individuals with color vision deficiency (CVD) and those with normal color vision. A CVD-friendly image should preserve several essential characteristics, including the sharpness

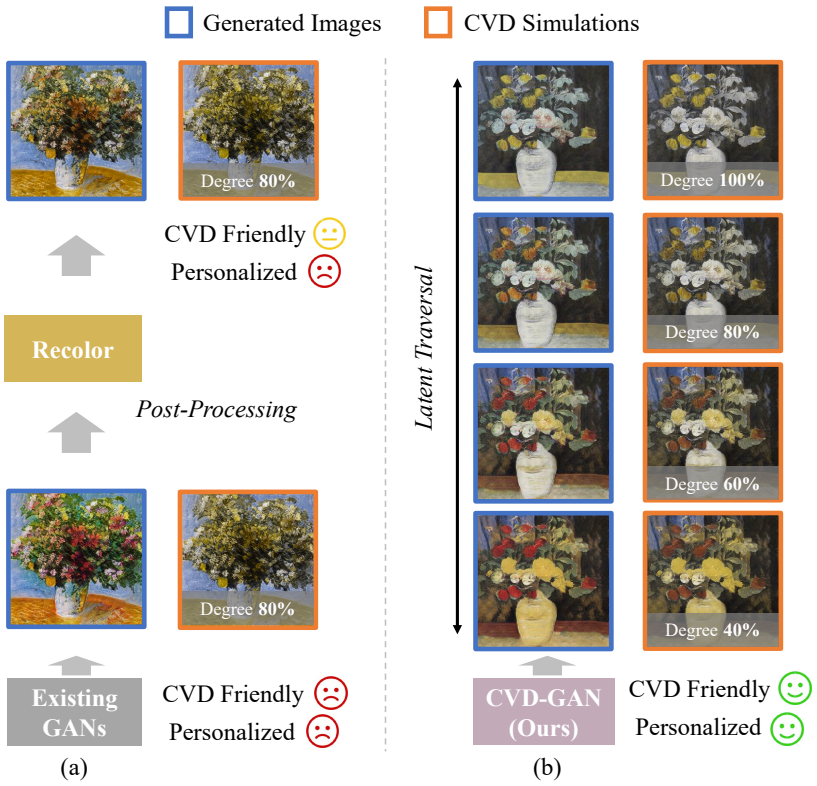

FIGURE 1.2. Compared to the combination baseline (a), the proposed CVD-GAN (b) can generate CVD-oriented images directly, enhancing the friendliness of the image for CVD populations. In addition, the model can generate personalized friendly images for CVD populations with varying degrees by disentangling the color representation based on the triple-latent structure.

of color transitions, consistency in color themes, and the retention of high-level semantic information, ensuring that individuals with CVD can perceive and interpret the image in a manner comparable to those with normal color vision.

So far, hardly any generation algorithm has offered to serve CVD populations. Some recoloring algorithms [17, 18, 19, 20, 21, 22, 23, 24, 25] can partly alleviate the problems by post-processing compensation based on the CVD simulation [26, 27] that provides the perspective of CVD populations of the given image. The process of recoloring can be summarized as providing CVD-unfriendly images as input, conducting color compensation or transformation, and outputting recolored images for CVD populations. As a result, a straightforward baseline for CVD-friendly generation can be formed by combining generation models

FIGURE 1.3. Potential limitations of the recoloring methods.

and recolor methods as the post-processing as Fig. 1.2 (a). However, this baseline still has many gaps in CVD-oriented and personalized generation.

The combination baseline is non-CVD-oriented, potentially restricting the user-friendliness of recolored images, where the generated content remains unchanged as recoloring methods solely concentrate on color transformation. This approach imposes a likely lower limit on the user-friendliness of the recolored images since the content may also matter in the CVD-friendly generation potentially and will somehow influence the performance of the recoloring methods. To illustrate, consider situations where indistinguishable colors (i.e. red, orange, and yellow) are present within a complex and intertwined color distribution or within small, densely populated areas (as depicted in the middle and right pairs in Fig. 1.3). In such scenarios, existing CVD-recolor methods often struggle to produce satisfactory results, as opposed to situations where these colors are showcased in larger patches with well-defined boundaries, as exemplified in the left pair of images in Fig. 1.3. (All figures are presented in pairs, with samples and deutan-simulated versions of the recolored samples.) Furthermore, despite the fact that CVD populations exhibit diverse requirements based on varying color impairment severity [28, 21], only a few recolor algorithms have addressed the issue of CVD diversity [20] so far.

Thus, our motivation is to propose a personalized CVD-friendly image generation algorithm, aiming to facilitate the visual experience of the diverse CVD population, allowing them to have a similar sensory experience to individuals with normal vision.

1.4 Contributions of Our Paper

To address the above gaps, we propose a CVD-oriented personalized image generation framework based on the adversarial network structure [4], as Fig. 1.2 (b). To generate CVD-aligned images, a framework that allows for unbiased perception among normal viewers and those with CVD is implemented. Further, in order to account for varying degrees of CVD, the color representation will be decoupled and controlled by a novel triple-latent structure, enabling the model to yield images with specified color distributions in accordance with the severity of the color impairment.

Particularly, a differential CVD simulator [17] posterior to the generated image, where CVD loss functions will be proposed and used to constrain the generated images and their corresponding simulation to achieve the CVD-oriented generation. Additionally, to reach the goal of personalized generation, triple-latent inputs will be established, where two latent codes serve as contrastive supervision and the other one controls the color pattern generation. Consequently, continuous CVD-friendly images towards various severity will be obtained through latent traversal.

Our proposed method evaluates the friendliness of generated images based on contrast decay, color information, and high-level perception across various types and degrees of CVD. Results indicate that our method outperforms existing image generation models and combination baselines on multiple datasets [29, 30, 24].

Our main contributions can be summarized as follows: (i) proposing an end-to-end CVD-oriented image generation framework, (ii) proposing a novel triple-latent structure to disentangle and control the color representation, enabling the model to generate continuous personalized CVD-friendly images aligned with all degrees of CVD populations. (iii) Extensive experiments on datasets [29, 30, 24] show that CVD-GAN can generate CVD-friendly images for CVD populations with varying types and severity.

CHAPTER 2

Literature review

Our contribution is related to prior works about generative adversarial networks, GAN Representation Disentanglement, and recoloring methods for CVD Compensation.

2.1 Generative Adversarial Network

In recent years, there has been a notable advancement in the domain of generative adversarial networks (GANs), encompassing substantial progress in both image quality enhancement and training stability, as evidenced by a substantial body of research [31, 6, 32, 7, 33, 34, 35, 36]. This commendable development is underscored by the transformation of generated images, which have transitioned from rudimentary representations such as handwritten digits to intricate and sophisticated compositions, including artistic paintings [37, 38] and exceptionally high-resolution visuals [33].

It is worth noting that the foundational concept of adversarial networks, initially conceived in the context of image generation, has transcended its origins and found extensive application across diverse domains [39, 40]. This broad-spectrum application underscores the far-reaching potential inherent in adversarial network frameworks.

However, amid this fervent pursuit of technological advancement and creative image synthesis, it is imperative to acknowledge an unintentional consequence of this progress. Specifically, individuals within the population afflicted by Color Vision Deficiency (CVD) may inadvertently find themselves excluded as target users of these generated images. Regrettably, the accessibility and comprehensibility of the content contained within these generated visuals

may elude those grappling with CVD, thereby limiting their ability to fully engage with and appreciate the evolving landscape of generative imagery. This inadvertent oversight highlights a pressing concern in the intersection of technology, accessibility, and inclusivity, warranting further exploration and potential remedies to ensure that these advancements benefit a wider and more diverse audience.

2.2 GAN Representation Disentanglement

The challenge of disentangling and controlling representations within the generative process, often regarded as a "black box," has been a subject of ongoing exploration and innovation in the field of generative adversarial networks (GANs). Several approaches have been proposed to address this intricate issue, each offering unique insights and solutions to the problem.

One notable approach is InfoGAN [41], which tackles the representation learning problem by maximizing the mutual information between latent variables and generated data. This framework seeks to uncover meaningful and interpretable representations within the latent space, thus enhancing control over the generative process. StyleGAN [6] introduces a distinctive architecture featuring intermediate latent variables. These intermediate variables enable the "mixing" of style information and can be progressively fed into different layers of the generator. This innovative structure provides finer-grained control over image style, allowing for the creation of diverse and customizable visuals. Building upon the foundations laid by StyleGAN, subsequent works have extended and refined this framework. Lee *et al.* [42] introduced techniques to fix noise patterns in StyleGAN, preserving a desired target style. Zhu *et al.* [43] proposed an automated mechanism for selecting style latent variables, facilitating semantic discovery and control.

However, it is important to note that the quest for unsupervised disentanglement is not without challenges. Locatello *et al.* [44] raised concerns about the reliability of some unsupervised disentanglement models. They emphasized the strong dependence of these models on random seeds and hyperparameters, highlighting the need for robustness and reproducibility in research. Additionally, Locatello's work advocated for making the inductive bias explicit

and underlining the practical benefits of disentanglement, aligning the pursuit of theoretical advances with real-world applicability. Furthermore, while advancements have been made in decoupling representations, the issue of controlling representations during latent traversal remains relatively underexplored [45]. This represents a compelling avenue for future research, as it could further empower users to interactively and intuitively shape generative outputs.

2.3 CVD Recoloring Compensation

In the domain of image recoloring, the pursuit of two primary objectives has been evident: the restoration of decaying contrast and the preservation of naturalness within the recolored images. These goals are essential not only for enhancing the overall visual quality but also for aiding individuals with Color Vision Deficiency (CVD) in discerning image content. Various research endeavors have been undertaken to address these objectives, leveraging a range of techniques and methodologies.

To enhance contrast and assist CVD users in distinguishing image content, several approaches have emerged. Some studies [17, 19, 18, 20] have focused on contrast compensation by optimizing objective functions that align given images with recolored image simulations. Alternatively, deep learning networks have been employed in works such as those by Li et al.[46] and Ma et al.[47] to perform color transformations, thereby improving contrast and color differentiation. Lau et al. [48] have implemented K-means algorithms to enhance contrast in adjacent areas, contributing to improved visual clarity.

While contrast enhancement is pivotal, preserving the naturalness of recolored images is equally crucial. Several approaches have proposed incorporating constraints between given images and their recolored counterparts as penalized regularization terms [21, 20, 22, 23, 24]. These constraints ensure that the recolored images maintain a sense of realism and fidelity to the original content. Additionally, Rigos et al. [25] introduced the concept of semantic segmentation to selectively transform the colors of objects while leaving other elements unchanged, further contributing to the preservation of naturalness.

However, despite these commendable advancements, the diverse demands of CVD populations, which can exhibit varying degrees of color vision impairment, have often been overlooked. For instance, Zhu et al. [20] required users to manually input configurations to obtain corresponding recolored images. This manual input approach may yield inappropriate results due to the sensitivity of the parameters and the potential for user error. Achieving personalized recoloring tailored to the unique needs of individuals with CVD remains a challenging endeavor.

CHAPTER 3

Methods

3.1 Overview

Our goal is to enable end-to-end CVD-aligned generation. Further, personalized generation will be achieved based on the novel triple-latent structure, adapting to varying degrees of CVD. Our method is established based on the generative adversarial network, training a generator $G(\cdot)$ that synthesizes images from noise z sampled from noise distribution p_{noise} to fool the discriminator and a discriminator $D(\cdot)$ to distinguish the fake images $G(z)$ based on the dataset distribution p_{data} adversarially at the same time. The loss function of GAN can be defined as:

$$\mathcal{L}_G = \mathbb{E}_{x\sim p_{\text{data}}}\Big[\log\big(1 - D(x)\big)\Big] + \mathbb{E}_{z\sim p_{\text{noise}}}\Big[\log\big(1 - D(G(z))\big)\Big]. \tag{3.1}$$

The GAN loss function only aims to generate images with the same distribution as the real images, where the demand of the CVD populations is disregarded. Hence, a CVD-oriented GAN is expected to assist the CVD populations.

As illustrated in Fig. 3.1, our model is comprised of two distinct functional components. The first component, referred to as "CVD-oriented generation" (depicted in Fig. 3.1 (b)), is designed with the specific objective of generating CVD-friendly images. This is achieved through the incorporation of the CVD simulation (introduced in Sec. 3.2) and a CVD-oriented loss function denoted as $\mathcal{L}_{\text{CVD}}$ (outlined in Sec. 3.3). Furthermore, recognizing that individuals with varying degrees of CVD possess different sensitivities to distinguishable colors, we have implemented a "color representation disentanglement" mechanism based on a triple-latent

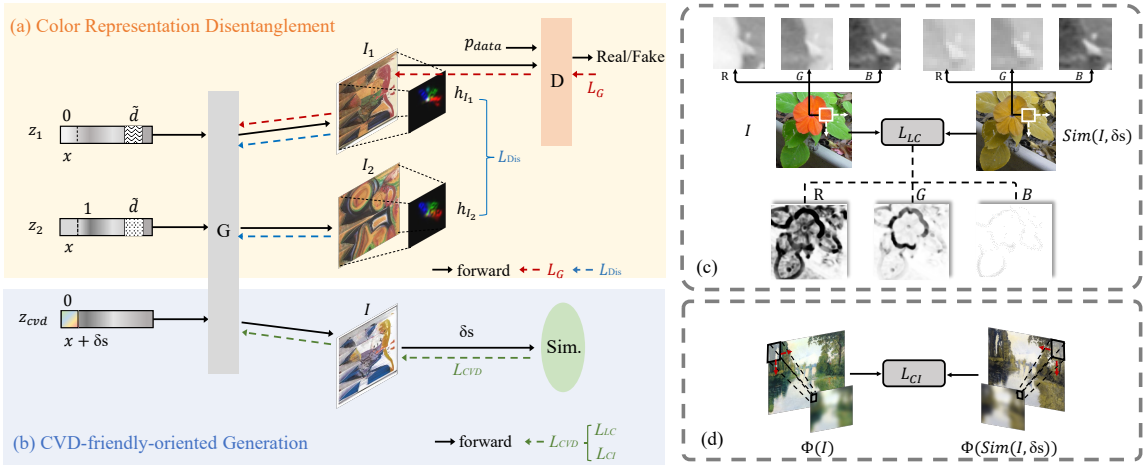

FIGURE 3.1. Structure of the CVD-GAN. In (a) and (b), z_1, z_2 and z_{cvd} are three latent codes with size of D. I_1, I_2 and I are images generated by the generator G. To enhance the dominance of the z^0, the dominance of other dimensions needs to be diminished. Hence, $\mathcal{L}_{\text{Dis}}$ is used to ensure the color histogram h_{I_1} and h_{I_1} have the same distribution. Meanwhile, an increment δs representing the CVD severity is added on the z^0_{cvd}, which is also passed into the CVD simulation Sim. to obtain the specified $\text{Sim}(\cdot)$ and constraints $\mathcal{L}_{\text{CVD}}$. Besides, discriminator $D(\cdot)$ discriminates whether I_1 is fake or not based on the real data distribution P_{data}. (c) and (d) present $\mathcal{L}_{\text{LC}}$ and $\mathcal{L}_{\text{CI}}$, which aim to retain the contrast and preserve the color information. In (c), $\mathcal{L}_{\text{LC}}$ retain the contrast by minimizing the decay of the local contrast of local maps in I as shown in the first row, which can be visualized in RGB channels and be summarized as the last row, where the darker regions indicate a more severe loss. In (d), $\mathcal{L}_{\text{CI}}$ calculated the loss of color information extracted by Gaussian Blur function Φ. $\mathcal{L}_{\text{CVD}}$ and $\mathcal{L}_{\text{Dis}}$ will be trained with the GAN loss $\mathcal{L}_G$.

structure (illustrated in Fig. 3.1 (a)). This adaptation serves to cater to the diverse requirements of individuals with differing degrees of CVD (explained in Sec. 3.4).

3.2 CVD Simulation

A two-stage model [27] is implemented to simulate CVD gamut, summarized as:

$$\text{Sim}(I, \delta s) = \Gamma^{-1}\Gamma_{\delta s} \cdot I, \tag{3.2}$$

where I is the input image, δs denotes the degree of the CVD, $\Gamma_{\delta s}$ is 3×3 matrix parameterized by δs. Γ is a constant matrix representing the perception of normal people, with the same size as $\Gamma_{\delta s}$. The detailed derived formulas will be presented in the appendix.

CVD simulation serves a crucial role in allowing individuals with normal color vision to gain insight into the visual perspective of those with color vision deficiencies (CVD). It enables the assessment of potential perception biases by employing pure matrix transformations, a differential process that will be seamlessly integrated into our framework.

3.3 CVD-Oriented Loss Functions

This section introduces the CVD-oriented loss $\mathcal{L}_{\text{CVD}}$, which aims to preserve image information after the corresponding CVD simulation to prevent perception bias. $\mathcal{L}_{\text{CVD}}$ includes two constraint losses $\mathcal{L}_{\text{LC}}(I, \delta s)$ and $\mathcal{L}_{\text{CI}}(I, \delta s)$ as:

$$\mathcal{L}_{\text{CVD}} = \mathcal{L}_{\text{LC}}(I, \delta s) + \mathcal{L}_{\text{CI}}(I, \delta s), \tag{3.3}$$

where I is image and δ_s represents the degree of CVD.

Local Contrast Loss. Due to color impairment, the patch boundaries of the image will be blurred if indistinguishable colors are distributed in adjacent pixels, discouraging the information acquisition for the CVD population. As shown in Fig. 3.1 (c), the boundaries of the petal and leaves become ambiguous due to color impairment. To retain the image distinct after simulation, the contrast within all of the local neighborhood maps of the image should be sustained after simulation. To evaluate the loss of contrast, the contrast term of the SSIM [49] is adopted as:

$$c\big(x, y\big) = \frac{2\sigma_x\sigma_y + \varepsilon}{\sigma_x^2 + \sigma_y^2 + \varepsilon}, \tag{3.4}$$

where σ_x and σ_y are the standard deviations of the input patch x and y as the first row of Fig. 3.1 (c), ε is a small constant to avoid instability. $c(\cdot)$ calculates the contrast similarity between corresponding local maps as Eq. (3.4). The loss $\mathcal{L}_{LC}$ is computed by aggregating the

local contrast losses in patches:

$$\mathcal{L}_{\mathrm{LC}}(I, \delta s) = 1 - \frac{1}{|\mathcal{N}|} \sum_{(x,y)\in\mathcal{N}} c\big(x, y\big), \tag{3.5}$$

where $\mathcal{N}$ is the set of corresponding local maps in the generated image I and its simulation $\mathrm{Sim}(I, \delta s)$; The $\mathcal{L}_{\mathrm{LC}}(I, \delta s)$ can be visualized in RGB channels as the last row of Fig. 3.1 (c), where the darker region presents a larger contrast loss.

Color Information Loss. Color itself carries a lot of information for images, including style, mood, temperature, etc., while the available color gamut for the CVD population is limited. Therefore, we expect the generated images can adapt to the CVD gamut and maintain the main colors after the simulation to avoid ambiguity. To extract the primary color of an image while avoiding excessive detail, a Gaussian kernel is applied to blur the image, as demonstrated in Fig. 3.1 (d). This optimization process can be summarized as:

$$\mathcal{L}_{\mathrm{CI}}(I, \delta s) = \Big\| \Phi\big(I\big) - \Phi\big(\mathrm{Sim}(I, \delta s)\big) \Big\|_1, \tag{3.6}$$

where I denotes the generated images; $\Phi(\cdot)$ means the Gaussian Blur process as pixel details are not needed; $\| \cdot \|_1$ is the L1 norm of a vector.

3.4 Triple-Latent Based Color Disentanglement

As people with distinct degrees of CVD have various sensitivities to discernable hues, color distribution generation is expected to be personalized to different users. To obtain images with varying color distribution for different requirements, two goals need to be achieved: 1) color representation should be disentangled; 2) color distribution can be controlled according to the specified requirement.

Therefore, a novel triple-latent structure is proposed to attain the goal. Specifically, the triple-latent can be divided into two groups, namely the contrastive group containing z_1 and z_2 that facilitates the first goal of color representation disentanglement and the control group z_{cvd} that accomplishes the second goal of the personalized generation.

Since color representation is entangled with the dimensions of the latent code in an ordinary GAN, changes in each dimension may cause changes in the color generation during the latent traversal. In other words, the dominance of the dimensions controlling color generation is diffused and irregular. Oppositely, a fixed dimension is expected to control the color. The contrastive group approach is designed based on the intuition that diminishing the influence of all other dimensions on color generation would result in the expected dimension dominating the color representation.

The contrastive group comprises two latent codes, z_1 and z_2, each with a dimensionality of D. Mathematically, $z_1 = \{z_1^d | d \in [0, D)\}$, $z_2 = \{z_2^d | d \in [0, D)\}$. Similarly, for the control latent code $z_{cvd} = \{z_{cvd}^d | d \in [0, D)\}$, where D is the dimension of latent codes, in which $z_1^0 = z_2^0, z_{cvd}^0 = z_1^0 + \delta_s$. δ_s is sampled from the uniform distribution of $[0.0, 1.0]$, indicating the severity of CVD. During the training, a randomly selected vector dimension $\tilde{d} \in [1, D)$ will be utilized to encourage the color representation to be fully decoupled. We ensure that 1) $z_1^{\tilde{d}} \neq z_2^{\tilde{d}}$; 2) $z_1^d = z_2^d, d \in [0, D), d \neq \tilde{d}$; 3) $z_1^d = z_{cvd}^d, d \in [1, D)$. As a result, the goal is to minimize the dominance of color representation of the $z^{\tilde{d}}$, persuading it to be dominated by the z^0.

To reduce the dominance of the $z^{\tilde{d}}$, z_1 and z_2 are sent into generator G as:

$$[I_1, I_2] = G([z_1, z_2]), \tag{3.7}$$

where $[I_1^{\tilde{d}}, I_2^{\tilde{d}}]$ is the image pair generated from the generator G. Further, to reduce the influence of $\tilde{d}$, a constraint will be utilized on the image pair $[I_1^{\tilde{d}}, I_2^{\tilde{d}}]$ to ensure the color distribution will keep unchanged no matter how the value of latent code $z^{\tilde{d}}$ on dimension $\tilde{d}$ changes as:

$$\mathcal{L}_{\text{Dis}} = \frac{1}{\sqrt{2}} || \sqrt{H(I_1)} - \sqrt{H(I_2)} ||_2^2, \tag{3.8}$$

where $H(\cdot)$ is a operation to obtain the 2D color histogram feature [50], $\| \cdot \|_2^2$ is the L2 norm. An example of color representation disentanglement is shown in Fig. 3.2. The impact of $z^{\tilde{d}}$ on the generation of color patterns is negligible because variations in the value of $z^{\tilde{d}}$ produce only slight modifications in the distribution of colors, then the color distribution generation can be predominantly influenced by z^0.

FIGURE 3.2. Color representation disentanglement. (a) The influence of the dimension $z^{\tilde{d}}$ on color pattern generation is minimal, as changes in the value of $z^{\tilde{d}}$ result in few alterations to the color distribution, (b) z^0 can dominate the color distribution generation.

This increment δs will be fed into the later objective function Eq. (3.5) and Eq. (3.6) as the CVD severity to obtain specified constraints as

$$\mathcal{L}_{\text{CVD}} = \mathcal{L}_{\text{LC}}\big(G(z_{cvd}), \delta s\big) + \mathcal{L}_{\text{CI}}\big(G(z_{cvd}), \delta s\big), \tag{3.9}$$

where $\mathcal{L}_{\text{LC}}(\cdot)$ and $\mathcal{L}_{\text{CI}}(\cdot)$ are local contrast and color information loss functions introduced in Sec. 3.3. As a result, $\mathcal{L}_{\text{CVD}}$ is able to provide different degrees of constraints for various severity of color impairment. Through training, CVD-GAN enables the generation of personalized images for different degrees of CVD by performing latent traversal on the dimension z^0, whereby increments of δs.

During training, the total losses $\mathcal{L}$ include constraints deployed for color representation disentanglement $\mathcal{L}_{Dis}$ and CVD-oriented loss functions $\mathcal{L}_{\text{CVD}}$, and GAN loss $\mathcal{L}_G$, which can be denoted as:

$$\mathcal{L} = \mathcal{L}_G + \alpha\mathcal{L}_{\text{Dis}} + \beta\mathcal{L}_{\text{CVD}}, \tag{3.10}$$

where α and β are loss weights.

CHAPTER 4

Results

4.1 Experiments Settings and Datasets

Datasets. To explore the CVD-oriented generation, the datasets [29, 30, 51] with flexible colors were selected. Flower [29] dataset contains 8,189 images with 103 classes. Abstract art [30] includes 15,022 artworks of the abstract genre from the Middle Ages to recent years. Still-Life and symbolic-painting are the subclasses of the wikiArt [51], which contain 4,799 images and 3,000 images depicting still objects and symbolic imagery, respectively.

Settings. StyleGAN-ada is served as the backbone, and the training setting mostly follows [32] with the Adam optimizer [52], the learning rate of 0.0025, batch size of 64, and 15000 steps. The weight α of the $\mathcal{L}_{\text{Dis}}$ is set to 15 while the weight β of the combination of $\mathcal{L}_{\text{LC}}(I, \delta s)$ and $\mathcal{L}_{\text{CI}}(I, \delta s)$ is set to 1. The trade-off between the weights and generated image quality will be discussed in Sec. 4.4. It is noted that, unlike StyleGAN, the latent codes with a length of 16 will be fed directly into the generation without a prior mapping transformation. The detailed network architecture is presented in the Table 4.1.

4.2 Qualitative Evaluation

In Figure 4.1, a comparative analysis is presented, evaluating the performance of three different approaches: StyleGAN [7], StyleGAN with recolor methods [20, 17], and the CVD-GAN proposed in this study. These evaluations are conducted using diverse datasets, including the still-life dataset [51], the flower dataset [29], and the symbolic painting dataset [51].

Generator
16 × 16 × 128 Learnable Constant
3 × 3 Deconv. ReLU
3 × 3 ModuConv. ReLU, Latents 4
3 × 3 Conv. ReLU
3 × 3 Conv. ReLU
3 × 3 Conv. + Noise ReLU
3 × 3 Conv. + Noise ReLU
3 × 3 Deconv. ReLU
3 × 3 ModuConv. ReLU, Latents 4
3 × 3 Conv. ReLU
3 × 3 Conv. ReLU
3 × 3 Conv. + Noise ReLU
3 × 3 Conv. + Noise ReLU
3 × 3 Deconv. ReLU
3 × 3 ModuConv. ReLU, Latents 4
3 × 3 Conv. ReLU
3 × 3 Conv. ReLU
3 × 3 Conv. + Noise ReLU
3 × 3 Conv. + Noise ReLU
3 × 3 Deconv. ReLU
3 × 3 ModuConv. ReLU, Latents 4
3 × 3 Conv. ReLU
3 × 3 Conv. ReLU
3 × 3 Conv. + Noise ReLU
3 × 3 Conv. + Noise ReLU
256 × 256 × 3

TABLE 4.1. Structure of CVD-GAN Generator.

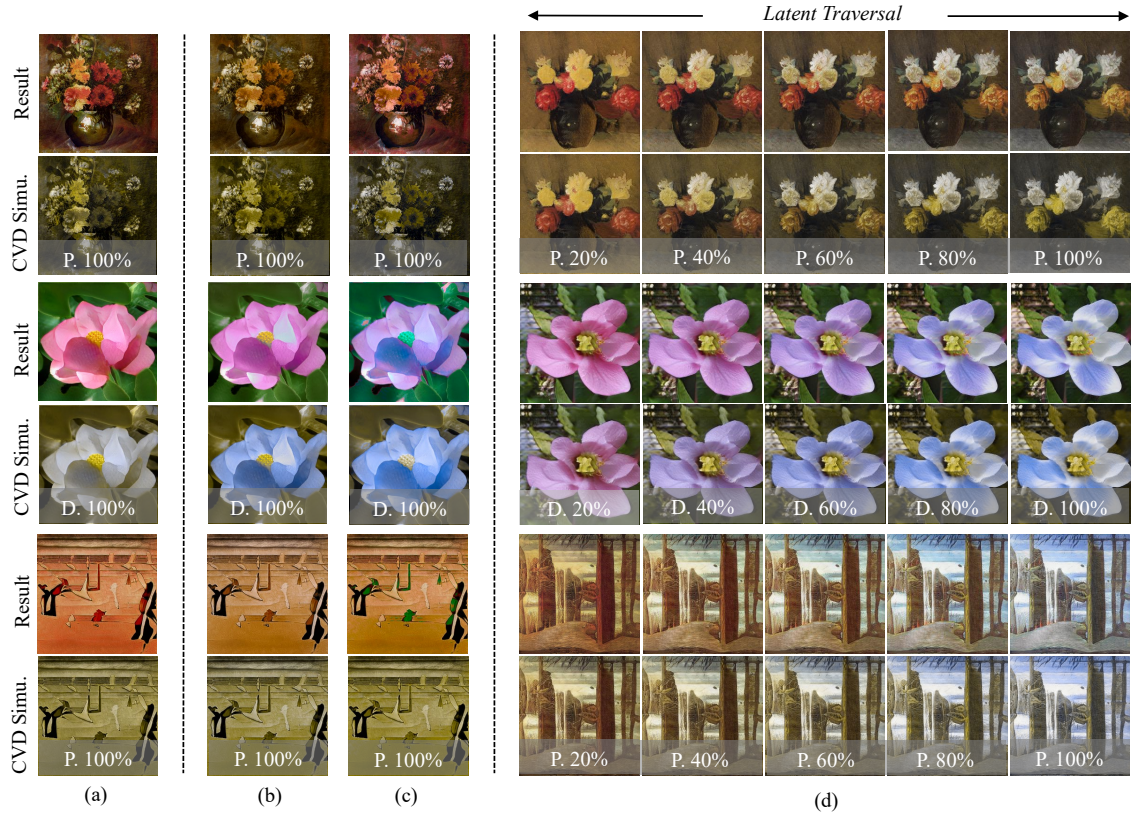

FIGURE 4.1. Qualitative comparison. (a) The results of StyleGAN [7], (b) and (c) present the results of StyleGAN with recolor methods [20, 17], (d) shows our results through latent traversal. For each, the first row shows the generation result (or after recolor compensation), and the second row shows the corresponding CVD simulation. “D." and “P." show the degree of deutan and protan, respectively.

In the context of the still-life dataset, images generated through StyleGAN are observed to exhibit a tendency to blur petals into the background, thereby creating ambiguity in the images. Even after the application of recolor compensation techniques, this ambiguity persists, potentially hindering the ability of individuals with Color Vision Deficiency (CVD) to distinguish the image content. In contrast, CVD-GAN demonstrates a distinctive capability to mitigate confusion by darkening the background as the degree of CVD increases and lightening the petals to a perceptible shade of yellow. This adjustment is particularly beneficial for protan populations, for whom the color red is often imperceptible.

When evaluating the flower dataset, StyleGAN's generated images are found to suffer from a severe decay of color and contrast, resulting in perceptual bias. While recolor compensation methods partially alleviate this bias, a noticeable gap between normal and CVD perspectives remains. In contrast, CVD-GAN exhibits a remarkable ability to generate CVD-oriented color distributions through latent traversal. Importantly, this transformation is achieved with minimal loss of information, as evidenced by the high-quality output images generated by the model.

The effectiveness of CVD-GAN is further exemplified in the context of the symbolic painting genre, where it consistently outperforms StyleGAN and recolor compensation methods in maintaining information fidelity and mitigating perceptual deviations for CVD populations.

To provide a more comprehensive assessment of CVD-GAN's performance, a user study has been conducted, the details of which are available in the appendix. This study reaffirms the effectiveness of CVD-GAN in improving accessibility and enhancing the visual experience for individuals with varying degrees of color vision impairment.

4.3 Quantitative Evaluation

In this section, several experiments will be conducted to compare the effectiveness among the generation baseline StyleGAN [7], StyleGAN with post-processing recolor methods [20, 17], and proposed CVD-GAN under various situations of degrees ($20\%, 40\%, 60\%, 80\%, 100\%$) with two different CVD types (protan and deutan) conditions.

As illustrated in Sec. 1.3, CVD-friendly pertains to an image that can be comprehended in a similar manner by both populations with Color Vision Deficiency (CVD) and those without. A CVD-friendly image should maintain the sharpness of color transitions, consistency of color theme, and high-level semantics for CVD-viewers. Our metrics adopted from [53, 50, 54] correspond to these aspects of CVD-friendliness.

Local Contrast Distance Decay. A CVD-friendly image should preserve the sharpness and clarity of color transitions. This entails ensuring that boundaries between different colors

Dataset	Type	Degree	StyleGAN [7]			StyleGAN with Zhu *et al.* [20]			StyleGAN with Huang *et al.* [17]			CVD-GAN (Ours)		
			LCD	H dis.	Perc. L.	LCD	H dis.	Perc. L.	LCD	H dis.	Perc. L.	LCD	H dis.	Perc. L.
Abstract Art [30]	Protan	20%	0.4663	0.0151	0.3629	0.4439	0.0150	0.3569	0.6712	0.0151	0.4334	**0.2155**	**0.0079**	**0.1094**
		40%	0.7639	0.0186	0.5950	0.7439	0.0181	0.5640	1.0699	0.0193	0.6929	**0.3355**	**0.0108**	**0.3230**
		60%	0.9573	0.0206	0.7715	0.7360	0.0199	0.6320	1.3085	0.0218	0.8898	**0.4002**	**0.0121**	**0.4165**
		80%	1.0762	0.0221	0.9149	0.6133	0.0209	0.6329	1.4391	0.0234	1.0482	**0.4301**	**0.0129**	**0.4856**
		100%	1.1218	0.0232	1.0350	0.5450	0.0218	0.6606	1.4848	0.0243	1.1756	**0.4378**	**0.0131**	**0.5333**
	Deutan	20%	0.5398	0.0159	0.4045	0.5209	0.0159	0.4509	0.7996	0.0178	0.4897	**0.2330**	**0.0086**	**0.2048**
		40%	0.8400	0.0193	0.6321	0.8388	0.0190	0.6165	1.2419	0.0217	0.7567	**0.3438**	**0.0113**	**0.3309**
		60%	1.0023	0.0212	0.7823	0.8293	0.0207	0.6827	1.4845	0.0238	0.9365	**0.3915**	**0.0122**	**0.4063**
		80%	1.0815	0.0225	0.8869	0.7350	0.0215	0.7053	1.6063	0.0251	1.0629	**0.4067**	**0.0129**	**0.4526**
		100%	1.1104	0.0232	0.9619	0.7007	0.0221	0.7415	1.6509	0.0257	1.1523	**0.4052**	**0.0129**	**0.4782**
Still-Life [51]	Protan	20%	0.4673	0.0101	0.3789	0.3982	0.0112	0.3368	0.7715	0.0125	0.4816	**0.2783**	**0.0075**	**0.2789**
		40%	0.7561	0.0148	0.6455	0.6293	0.0152	0.5369	1.2217	0.0183	0.7992	**0.4354**	**0.0114**	**0.4795**
		60%	0.9405	0.0182	0.8538	0.5777	0.0179	**0.6062**	1.4865	0.0219	1.0458	**0.5225**	**0.0138**	0.6272
		80%	1.0555	0.0207	1.2041	**0.4721**	0.0198	**0.6262**	1.6310	0.0243	1.2430	0.5660	**0.0155**	0.7366
		100%	1.1138	0.0224	1.1671	**0.4536**	0.0210	**0.6816**	1.6867	0.0257	1.4016	0.5800	**0.0167**	0.8181
	Deutan	20%	0.5261	0.0113	0.4207	0.4380	0.0123	0.3773	0.9581	0.0150	0.5432	**0.3044**	**0.0086**	**0.3091**
		40%	0.8041	0.0162	0.6820	0.6863	0.0162	0.5815	1.4718	0.0208	0.8700	**0.4408**	**0.0123**	**0.5061**
		60%	0.9476	0.0193	0.8620	0.6318	0.0189	0.6511	1.7424	0.0239	1.0979	**0.5095**	**0.0145**	**0.6309**
		80%	1.0145	0.0215	0.9891	0.5784	0.0208	0.7069	1.8698	0.0258	1.2585	**0.5285**	**0.0160**	**0.7068**
		100%	1.0379	0.0225	1.0817	0.5596	0.0217	0.7585	1.9093	0.0268	1.3709	**0.5265**	**0.0168**	**0.7585**
Symbolic-Painting [51]	Protan	20%	0.4190	0.0114	0.3363	0.3404	0.0128	0.2950	0.5508	0.0119	0.3725	**0.1980**	**0.0084**	**0.2252**
		40%	0.6840	0.0164	0.5715	0.4918	0.0172	0.4506	0.8788	0.0175	0.3259	**0.3055**	**0.0129**	**0.3780**
		60%	0.8564	0.0197	0.7528	0.4470	0.0198	0.5138	1.0754	0.0208	0.8214	**0.3626**	**0.0154**	**0.4845**
		80%	0.9661	0.0221	0.8996	0.3915	0.0214	0.5517	1.1888	0.0230	0.9779	**0.3882**	**0.0168**	**0.5489**
		100%	1.0245	0.0235	1.0236	**0.3770**	0.0224	**0.5957**	1.2403	0.024	1.1060	0.3947	**0.0176**	0.6125
	Deutan	20%	0.4532	0.0127	0.3741	0.3598	0.0141	0.3289	0.7079	0.0151	0.4402	**0.2107**	**0.0096**	**0.2468**
		40%	0.6880	0.0178	0.6031	0.4992	0.0185	0.4889	1.0853	0.0206	0.7095	**0.3034**	**0.0140**	**0.3937**
		60%	0.8038	0.0208	0.7559	0.4648	0.0210	0.5604	1.2805	0.0235	0.8932	**0.3387**	**0.0161**	**0.4810**
		80%	0.8530	0.0227	0.8620	0.4404	0.0224	0.6149	1.3692	0.0252	1.0212	**0.3448**	**0.0173**	**0.5323**
		100%	0.8654	0.0236	0.9401	0.4244	0.0231	0.6619	1.3937	0.0261	1.1121	**0.3388**	**0.0178**	**0.5627**
Flowers [29]	Protan	20%	0.5937	0.0179	0.5311	0.6829	0.0191	0.6047	0.9519	0.0164	0.6709	**0.2799**	**0.0118**	**0.3162**
		40%	0.9566	0.0233	0.8795	1.1452	0.0242	0.9067	1.5128	0.0222	1.0542	**0.4193**	**0.0168**	**0.5383**
		60%	1.1820	0.0263	1.1498	1.0872	0.0270	1.0920	1.8476	0.0256	1.3490	**0.4847**	**0.0196**	**0.7000**
		80%	1.3125	0.0282	1.3694	0.8756	0.0280	1.1231	2.0309	0.0278	1.5876	**0.5101**	**0.0211**	**0.8195**
		100%	1.3610	0.0294	1.5514	0.8437	0.0292	1.2508	2.0938	0.0289	1.7789	**0.5147**	**0.0218**	**0.9064**
	Deutan	20%	0.7323	0.0188	0.5777	0.8502	0.0199	0.6599	0.9952	0.0190	0.6889	**0.3423**	**0.0121**	**0.3334**
		40%	1.1509	0.0240	0.9187	1.3906	0.0246	1.0012	1.5518	0.0239	1.0431	**0.5071**	**0.0166**	**0.5460**
		60%	1.3896	0.0267	1.1614	1.3201	0.0268	1.0841	1.8641	0.0266	1.2846	**0.5829**	**0.0189**	**0.6860**
		80%	1.5178	0.0285	1.3386	1.1930	0.0270	1.0864	2.0269	0.0282	1.4560	**0.6123**	**0.0201**	**0.7778**
		100%	1.5756	0.0290	1.4645	1.1679	0.0274	1.1570	2.0917	0.0288	1.5749	**0.6179**	**0.0204**	**0.8346**

TABLE 4.2. Quantitative Results. Comparison with StyleGAN [7] and StyleGAN with recolor methods [20, 17]. For each method, three metrics, including Local Contrast Decay denoted as LCD, Hellinger distance of color histogram abbreviated as H.dis., and perceptual loss abbreviated as Perc.L., are implemented to evaluate. For all the metrics, the lower value means the higher friendliness of the image.

are distinct and easily distinguishable for individuals with CVD. Maintaining sharp color transitions is pivotal for preventing color confusion and enhancing the visual experience. In other words, the local contrast is expected to be preserved, otherwise, the image will be

FIGURE 4.2. The significance of the consistent color theme for CVD populations. Though hardly any indistinguishable patches occur under CVD's perspective, CVD individuals tend to misinterpret the original image (the left one) as an autumn scene.

ambiguous to distinct. Following the conventional thoughts [19, 55], decayed Euclidean distance between corresponding local maps of test images and their simulations will be employed to assess the sharpness loss. To be noted, test images will be transformed into CIE L*a*b* color space which better represents the human perception of colors [56] than RGB color spaces. The blue column in Table 4.2 shows the local contrast distance decay for each method.

Hellinger Distance of Color Histogram. In addition to contrasting aspects, our evaluation framework also incorporates a metric [50] for assessing color theme consistency. Given that colors inherently convey significant information, shifts in an image's tonal theme can profoundly affect its semantics. To illustrate, 4/5 of our CVD testers in the **user study** recognize Fig. 4.2 as an autumn scene, whereas it unequivocally represents spring from the perspective of individuals with normal vision.

To assess the preservation of the primary image color following simulation, we will employ the Hellinger distance to measure the dissimilarity between the color distributions [50] derived from the test image I and its simulation, denoted as $\mathrm{Sim}(I, \delta s)$. The less the distance is, the main color is more consistent after simulation, and the more friendly the image I is. The pink column in Table 4.2 shows the Hellinger distance between generated images and their simulations of the color histogram.

Perceptual Loss. Color perception impairment can lead to the loss of high-level information beyond content details. Consequently, we incorporate perceptual loss, a widely accepted

metric for evaluating the degradation of high-level features that might be neglected by local criteria [54], to further assess high-level semantics.

By leveraging a pre-trained VGG model, features and semantics from the images can be extracted, enabling us to evaluate the perceptual similarity between the test images I and their CVD-friendly simulations $\mathrm{Sim}(I, \delta s)$ more effectively. The pink column in Table 4.2 shows the perceptual loss between generated images and their simulations.

In summary, our CVD-GAN showcases state-of-the-art (SOTA) performance across the majority of datasets and exhibits robustness across varying levels of color vision deficiency (CVD) severity. This accomplishment underscores the effectiveness of our approach in generating visually coherent, accessible, and comprehensible images for individuals with color perception impairments. By advancing the field of CVD-friendly image generation, our work contributes to making visual content more inclusive and accommodating the diverse needs of individuals with different degrees of CVD.

4.4 Ablation Study

CVD Loss Functions. To further discuss the contribution of each of the CVD loss functions, $\mathcal{L}_{\mathrm{LC}}(I, \delta s)$ and $\mathcal{L}_{\mathrm{CI}}(I, \delta s)$ will be ablated to analyze. Note that the experiments are performed in the protan CVD type by default. As Table 4.3 shows, with the implementation of $\mathcal{L}_{\mathrm{LC}}(I, \delta s)$, the local contrast distance decay will decrease significantly, while the metric of Hellinger distance of color histogram will be better slightly. The opposite situation will happen when with the implementation of only $\mathcal{L}_{\mathrm{CI}}(I, \delta s)$. Also, It's surprisingly found that the high-level metric, perception loss, might be more relevant to local contrast preservation than general color preservation.

Color Representation Disentanglement. If color representation can be fully disentangled and controlled by the chosen dimension, the color histogram contributions will be consistent between images generated by latent codes that differ in other dimensions. Thus, to confirm the effect of the $\mathcal{L}_{\mathrm{Dis}}$, Hellinger distance is used again to calculate the similarity between

Method	Degree					
	40%			100%		
	LCD	H dis.	Perc. L.	LCD	H dis.	Perc. L.
StyleGAN	0.7639	0.0186	0.5950	1.1218	0.0232	1.0350
$+\mathcal{L}_{\text{LC}}$	**0.3784**	0.0158	**0.3726**	**0.5052**	0.0197	**0.6039**
$+\mathcal{L}_{\text{CI}}$	0.4659	**0.0114**	0.4112	0.6104	**0.0139**	0.6924
$+\mathcal{L}_{\text{LC}}+\mathcal{L}_{\text{CI}}$	**0.3355**	**0.0108**	**0.3230**	**0.4378**	**0.0131**	**0.5333**

TABLE 4.3. The ablation study of CVD loss under the degrees of 40% and 100% in protan type.

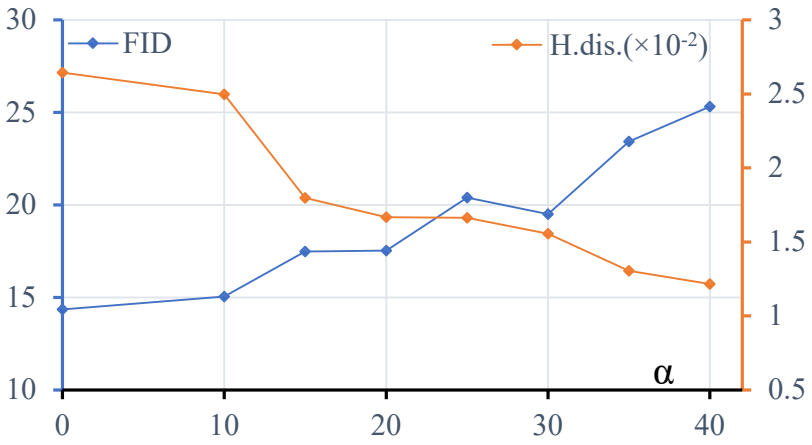

FIGURE 4.3. Effect of the color representation disentanglement and accordingly FID. α is the weight of the $\mathcal{L}_{\text{Dis}}$.

the color histogram feature extracted from the I_1 and I_2 denoted in the Fig. 3.1. Besides, to determine the value of the weight α of $\mathcal{L}_{\text{Dis}}$, the FID metric, used to evaluate the image quality, will be also considered. Fig. 4.3 presents the relationship between the α and FID.

It shows that with the increase of the weight α of $\mathcal{L}_{\text{Dis}}$, the image quality will decrease generally while the color representation disentanglement will be enhanced. When the weight equals 15, a balanced trade-off is reached to generate well-quality and disentangled images. As a result, the α is set to 15 in this paper.

Extra Baseline. To further substantiate the effectiveness of our approach, we conducted training with StyleGAN on the symbolic-painting dataset, which had undergone post-processing

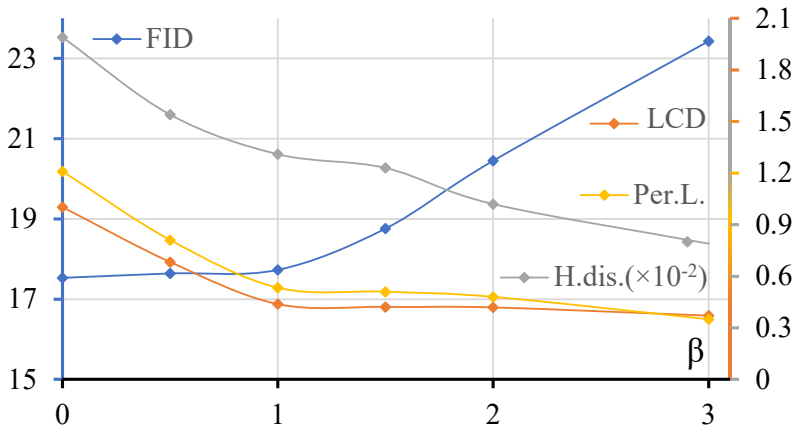

FIGURE 4.4. Trade-off of image quality. β is the weight of the $\mathcal{L}_{\text{CVD}}$. The blue, orange, gray, and yellow lines represent the FID, local contrast distance decay, Hellinger distance of color histogram, and perceptual Loss, respectively, based on the β.

as described by Zhu et al. [20]. The results in Table 4.4 clearly indicate the superiority of our method over this extra baseline across all evaluated aspects..

Method	P.40%			P.100%		
	LCD	H Dis.	Perc. L	LCD	H Dis.	Perc. L.
StyleGAN*	**0.295**	0.017	0.402	0.476	0.024	0.731
Ours	0.305	**0.013**	**0.378**	**0.395**	**0.018**	**0.613**

TABLE 4.4. Comparison with extra baseline under Protan's situation.

Trade-off of Generation Images Quality. The essence of all the CVD loss is to limit the color gamut of the generated images, which will cause a negative impact on the quality of generation. Fig. 4.4 presents the relationships between the β and FID metric with CVD metrics introduced in Sec. 4.3. The abscissa denotes the value of the weight of β, while the blue, orange, gray, and yellow lines represent the FID, local contrast distance decay, Hellinger distance of color histogram, and perceptual loss, respectively.

It is indicated that with the augment of the weight β of $\mathcal{L}_{\text{CVD}}$, the image is more suitable for CVD viewers at the cost of quality. After all, the β is set to 1 to reach a balanced trade-off between FID and CVD metrics.

In summary, we will compare the Fréchet Inception Distance (FID) of CVD-GAN across all datasets with the baseline, both with and without the post-processing recoloring method, as presented in Table 4.5. The results clearly illustrate that, in general, image quality diminishes as the severity of CVD increases, mainly because a higher degree of CVD leads to a more restricted gamut. However, when compared to various traditional post-processing recoloring methods, our approach consistently produces comparable results in terms of image quality across different datasets, even when dealing with varying degrees of CVD. It's worth noting that these experiments were conducted using a default CVD type of protan, and the dataset consisted of 4800 images.

Dataset	Degree	StyleGAN [7]	StyleGAN with		CVD-GAN (Ours)
			Zhu *et al.* [20]	Huang *et al.* [17]	
Abstract Art [30]	0%	14.35	-	-	17.73
	40%	-	16.68	-	18.27
	100%	-	23.44	16.86	19.58
Still-Life [51]	0%	18.96	-	-	22.10
	40%	-	23.42	-	24.09
	100%	-	26.36	21.91	25.36
Symbolic-Painting [51]	0%	28.20	-	-	31.66
	40%	-	29.26	-	28.37
	100%	-	30.55	28.76	28.01
Flowers [29]	0%	8.23	-	-	18.93
	40%	-	12.48	-	19.15
	100%	-	18.73	20.64	20.13

TABLE 4.5. FID of images generated by StyleGAN, post-processing recolor methods, and proposed CVD-GAN under various datasets, where the lower value indicates better image quality.

4.5 Limitations and Future work

CVD-GAN has demonstrated its ability to effectively generate personalized CVD-oriented images for protan and deutan types. However, it should be noted that it currently does not accommodate viewers with tritan (*e.g.* blue-perception cones impairment) or other complex color impairments due to constraints related to reference samples and volunteers. Also, the possibility of substituting the baseline with alternative generation models is encouraged to be considered, like diffusion models and VAEs. This could potentially yield improved results.

Furthermore, there is a need for an in-depth investigation to explore potential limitations of the recoloring algorithm, especially when dealing with content that may be considered "inherently unfriendly" to the recoloring process. Additionally, the advancement of large-scale vision models and multi-modal generation techniques has sparked a motivation to incorporate more user-friendly modalities into CVD-friendly image generation, such as text-to-image capabilities. This integration could potentially enhance the accessibility and user experience of generating images tailored for individuals with Color Vision Deficiency (CVD).

These aspects will be left for future exploration.

CHAPTER 5

Conclusions

The paper introduces an innovative approach to generate personalized images tailored for individuals with varying degrees of Color Vision Deficiency (CVD) using Generative Adversarial Networks (GANs). This method leverages deep learning techniques to address the needs of underrepresented populations. The key contributions of this model include:

1. End-to-End CVD-Oriented Image Generation: The model can seamlessly generate images specifically designed for individuals with CVD, providing a comprehensive solution.

2. Personalized Image Generation: It goes beyond generic CVD correction by producing personalized images for individuals with different CVD types and severity levels. This personalization is achieved by disentangling color representations through a triple-latent structure.

3. State-of-the-Art Performance: The proposed method demonstrates state-of-the-art performance on various datasets, encompassing both natural scenes and artistic paintings.

This research represents a significant advancement in the field of addressing visual impairments through deep learning techniques.

Bibliography

[1] A. Razavi, A. Van den Oord and O. Vinyals, 'Generating diverse high-fidelity images with VQ-VAE-2,' *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[2] J. Tomczak and M. Welling, 'VAE with a VampPrior,' in *International Conference on Artificial Intelligence and Statistics*, 2018, pp. 1214–1223.

[3] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, 'Generative adversarial networks: An overview,' *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[4] I. Goodfellow *et al.*, 'Generative adversarial networks,' *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[5] T. Miyato, T. Kataoka, M. Koyama and Y. Yoshida, 'Spectral normalization for generative adversarial networks,' in *International Conference on Learning Representations*, 2018.

[6] T. Karras, S. Laine and T. Aila, 'A style-based generator architecture for generative adversarial networks,' in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.

[7] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen and T. Aila, 'Training generative adversarial networks with limited data,' in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 12 104–12 114.

[8] A. Q. Nichol and P. Dhariwal, 'Improved denoising diffusion probabilistic models,' in *International Conference on Machine Learning*, 2021, pp. 8162–8171.

[9] J. Ho, A. Jain and P. Abbeel, 'Denoising diffusion probabilistic models,' *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[10] J. Bao, D. Chen, F. Wen, H. Li and G. Hua, 'Cvae-gan: Fine-grained image generation through asymmetric training,' in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2745–2754.

[11] P. Dhariwal and A. Nichol, 'Diffusion models beat gans on image synthesis,' *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[12] J. Ho and T. Salimans, 'Classifier-free diffusion guidance,' *arXiv preprint arXiv:2207.12598*, 2022.

[13] R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, 'High-resolution image synthesis with latent diffusion models,' in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[14] A. Radford *et al.*, 'Learning transferable visual models from natural language supervision,' in *International conference on machine learning*, PMLR, 2021, pp. 8748–8763.

[15] V. C. Smith and J. Pokorny, 'Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm,' *Vision Research*, vol. 15, no. 2, pp. 161–171, 1975.

[16] L. T. Sharpe, A. Stockman, H. Jägle and J. Nathans, 'Opsin genes, cone photopigments and color vision,' in *Color Vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe, Eds., Cambridge, UK: Cambridge University Press, 1999, ch. 1, pp. 3–51.

[17] J.-B. Huang, C.-S. Chen, T.-C. Jen and S.-J. Wang, 'Image recolorization for the colorblind,' in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 1161–1164. DOI: 10.1109/ICASSP.2009.4959795.

[18] G. M. Machado, 'A model for simulation of color vision deficiency and a color contrast enhancement technique for dichromats,' M.S. thesis, Universidade Federal do Rio Grande do Sul, 2010.

[19] G. M. Machado and M. M. Oliveira, 'Real-time temporal-coherent color contrast enhancement for dichromats,' *Computer Graphics Forum*, vol. 29, no. 3, pp. 933–942, 2010.

[20] Z. Zhu *et al.*, 'Personalized image recoloring for color vision deficiency compensation,' *IEEE Transactions on Multimedia*, vol. 24, pp. 1721–1734, 2022.

[21] Z. Zhu, M. Toyoura, K. Go, I. Fujishiro, K. Kashiwagi and X. Mao, 'Naturalness-and information-preserving image recoloring for red–green dichromats,' *Signal Processing: Image Communication*, vol. 76, pp. 68–80, 2019.

[22] M. F. Hassan and R. Paramesran, 'Naturalness preserving image recoloring method for people with red–green deficiency,' *Signal Processing: Image Communication*, vol. 57, pp. 126–133, 2017.

[23] H.-J. Kim, J.-Y. Cho and S.-J. Ko, 'Re-coloring methods using the HSV color space for people with the red-green color vision deficiency,' *Journal of the Institute of Electronics and Information Engineers*, vol. 50, no. 3, pp. 91–101, 2013.

[24] A. Rigos, S. Chatzistamatis and G. E. Tsekouras, 'A systematic methodology to modify color images for dichromatic human color vision and its application in art paintings,' *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, pp. 5015–5025, 2020.

[25] S. Chatzistamatis, A. Rigos and G. E. Tsekouras, 'Image recoloring of art paintings for the color blind guided by semantic segmentation,' in *Engineering Applications of Neural Networks*, 2020, pp. 261–273.

[26] H. Brettel, F. Viénot and J. D. Mollon, 'Computerized simulation of color appearance for dichromats,' *Journal of the Optical Society of America A*, vol. 14, no. 10, pp. 2647–2655, 1997.

[27] G. M. Machado, M. M. Oliveira and L. A. F. Fernandes, 'A physiologically-based model for simulation of color vision deficiency,' *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1291–1298, 2009. DOI: 10.1109/TVCG.2009.113.

[28] Z. Zhu, M. Toyoura, K. Go, I. Fujishiro, K. Kashiwagi and X. Mao, 'Processing images for red–green dichromats compensation via naturalness and information-preservation considered recoloring,' *The Visual Computer*, vol. 35, pp. 1053–1066, 2019.

[29] M.-E. Nilsback and A. Zisserman, 'Automated flower classification over a large number of classes,' in *Indian Conference on Computer Vision, Graphics & Image Processing*, 2008, pp. 722–729.

[30] G. Ogden, *Abstract art*, www.kaggle.com/datasets/goprogram/abstract-art, Accessed February 18, 2023 [Online]. [Online]. Available: https://www.kaggle.com/datasets/goprogram/abstract-art.

[31] T. Karras, T. Aila, S. Laine and J. Lehtinen, 'Progressive growing of GANs for improved quality, stability, and variation,' in *International Conference on Learning Representations*, 2018.

[32] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen and T. Aila, 'Analyzing and improving the image quality of StyleGAN,' in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8110–8119.

[33] A. Brock, J. Donahue and K. Simonyan, 'Large scale GAN training for high fidelity natural image synthesis,' in *International Conference on Learning Representations*, 2019.

[34] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. C. Courville, 'Improved training of Wasserstein GANs,' *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[35] M. Arjovsky, S. Chintala and L. Bottou, 'Wasserstein generative adversarial networks,' in *International Conference on Machine Learning*, 2017, pp. 214–223.

[36] N. Kodali, J. Abernethy, J. Hays and Z. Kira, 'On convergence and stability of GANs,' *arXiv preprint arXiv:1705.07215*, 2017.

[37] W. R. Tan, C. S. Chan, H. E. Aguirre and K. Tanaka, 'ArtGAN: Artwork synthesis with conditional categorical GANs,' in *IEEE International Conference on Image Processing*, 2017, pp. 3760–3764.

[38] W. R. Tan, C. S. Chan, H. E. Aguirre and K. Tanaka, 'Improved ArtGAN for conditional synthesis of natural image and artwork,' *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 394–409, 2019.

[39] S. Liu, K. Wang, X. Yang, J. Ye and X. Wang, 'Dataset distillation via factorization,' in *Advances in Neural Information Processing Systems*, 2022.

[40] Y. Wang, C. Xu, B. Du and H. Lee, 'Learning to weight imperfect demonstrations,' in *International Conference on Machine Learning*, PMLR, 2021, pp. 10 961–10 970.

[41] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever and P. Abbeel, 'InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets,' *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[42] D. Lee, J. Y. Lee, D. Kim, J. Choi, J. Yoo and J. Kim, 'Fix the noise: Disentangling source feature for controllable domain translation,' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 224–14 234.

[43] X. Zhu, C. Xu and D. Tao, 'ContraFeat: Contrasting deep features for semantic discovery,' *arXiv preprint arXiv:2212.07277*, 2022.

[44] F. Locatello *et al.*, 'Challenging common assumptions in the unsupervised learning of disentangled representations,' in *International Conference on Machine Learning*, 2019, pp. 4114–4124.

[45] A. Shoshan, N. Bhonker, I. Kviatkovsky and G. Medioni, 'Gan-control: Explicitly controllable gans,' in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 14 083–14 093.

[46] H. Li *et al.*, 'Color vision deficiency datasets & recoloring evaluation using GANs,' *Multimedia Tools and Applications*, vol. 79, pp. 27 583–27 614, 2020.

[47] Y. Ma, X. Gu and Y. Wang, 'Color discrimination enhancement for dichromats using self-organizing color transformation,' *Information Sciences*, vol. 179, no. 6, pp. 830–843, 2009.

[48] C. Lau, W. Heidrich and R. Mantiuk, 'Cluster-based color space optimizations,' in *International Conference on Computer Vision*, 2011, pp. 1172–1179.

[49] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, 'Image quality assessment: From error visibility to structural similarity,' *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[50] M. Afifi, M. A. Brubaker and M. S. Brown, 'HistoGAN: Controlling colors of GAN-generated and real images via color histograms,' in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7941–7950.

[51] B. Saleh and A. Elgammal, 'Large-scale classification of fine-art paintings: Learning the right metric on the right feature,' *arXiv preprint arXiv:1505.00855*, 2015.

[52] D. P. Kingma and J. Ba, 'Adam: A method for stochastic optimization,' in *International Conference on Learning Representations*, 2015.

[53] J.-B. Huang, Y.-C. Tseng, S.-I. Wu and S.-J. Wang, 'Information preserving color transformation for protanopia and deuteranopia,' *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 711–714, 2007.

[54] J. Johnson, A. Alahi and L. Fei-Fei, 'Perceptual losses for real-time style transfer and super-resolution,' in *European Conference on Computer Vision*, 2016, pp. 694–711.

[55] X. Wang, Z. Zhu, X. Chen, K. Go, M. Toyoura and X. Mao, 'Fast contrast and naturalness preserving image recolouring for dichromats,' *Computers & Graphics*, vol. 98, pp. 19–28, 2021.

[56] K. Leon, D. Mery, F. Pedreschi and J. Leon, 'Color measurement in L* a* b* units from RGB digital images,' *Food Research International*, vol. 39, no. 10, pp. 1084–1091, 2006.

[57] C. R. Ingling Jr and B. H.-P. Tsou, 'Orthogonal combination of the three visual channels,' *Vision Research*, vol. 17, no. 9, pp. 1075–1082, 1977.

5.1 Appendix A

CVD Simulation.

Based on the *two-stage theory*, this paper adopted a two-stage model to simulate the CVD gamut proposed by Machado [27]. Take the $\Delta\lambda$ as the offset distance, spectral curves of L-, M- and S-cone of CVD can be indicated as follows in the first stage:

$$L_a(\lambda) = L(\lambda + \Delta\lambda_L) \tag{5.1}$$

$$M_a(\lambda) = M(\lambda + \Delta\lambda_M) \tag{5.2}$$

$$S_a(\lambda) = S(\lambda + \Delta\lambda_S) \tag{5.3}$$

Then, in the second stage, the corresponding signals will be processed by the transformation matrix $T_{LMS_2O_{pp}}$ [57] into the opponent-color space as follows:

$$\begin{bmatrix} \text{WS}(\lambda) \\ \text{YB}(\lambda) \\ \text{RG}(\lambda) \end{bmatrix}_{pa} = T_{LMS_2O_{pp}} \begin{bmatrix} L_a(\lambda) \\ M(\lambda) \\ S(\lambda) \end{bmatrix} \tag{5.4}$$

$$\begin{bmatrix} \text{WS}(\lambda) \\ \text{YB}(\lambda) \\ \text{RG}(\lambda) \end{bmatrix}_{da} = T_{LMS_2O_{pp}} \begin{bmatrix} L(\lambda) \\ M_a(\lambda) \\ S(\lambda) \end{bmatrix} \tag{5.5}$$

$$\begin{bmatrix} \text{WS}(\lambda) \\ \text{YB}(\lambda) \\ \text{RG}(\lambda) \end{bmatrix}_{ta} = T_{LMS_2O_{pp}} \begin{bmatrix} L(\lambda) \\ M(\lambda) \\ S_a(\lambda) \end{bmatrix} \tag{5.6}$$

where pa, da, and ta represent the protan, deutan, and tritan deficiency; WS, YB and RG denote the channels of opponent-color space: white-black, yellow-blue, and red-green, respectively. By projecting the spectral power distribution $\varphi_R(\lambda)$, $\varphi_G(\lambda)$, and $\varphi_B(\lambda)$ of the RGB primaries, a transformation from RGB color space to the opponent-color space can be obtained as:

$$\begin{aligned}
\text{WS}_R &= \rho_{\text{WS}} \textstyle\int \varphi_R(\lambda) \text{WS}(\lambda) d\lambda, \\
\text{WS}_G &= \rho_{\text{WS}} \textstyle\int \varphi_G(\lambda) \text{WS}(\lambda) d\lambda, \\
\text{WS}_B &= \rho_{\text{WS}} \textstyle\int \varphi_B(\lambda) \text{WS}(\lambda) d\lambda, \\
\text{YB}_R &= \rho_{\text{YB}} \textstyle\int \varphi_R(\lambda) \text{YB}(\lambda) d\lambda, \\
\text{YB}_R &= \rho_{\text{YB}} \textstyle\int \varphi_G(\lambda) \text{YB}(\lambda) d\lambda, \\
\text{YB}_R &= \rho_{\text{YB}} \textstyle\int \varphi_B(\lambda) \text{YB}(\lambda) d\lambda, \\
\text{RG}_R &= \rho_{\text{RG}} \textstyle\int \varphi_R(\lambda) \text{RG}(\lambda) d\lambda, \\
\text{RG}_R &= \rho_{\text{RG}} \textstyle\int \varphi_G(\lambda) \text{RG}(\lambda) d\lambda, \\
\text{RG}_R &= \rho_{\text{RG}} \textstyle\int \varphi_B(\lambda) \text{RG}(\lambda) d\lambda,
\end{aligned} \tag{5.7}$$

where ρ_{WS}, ρ_{YB}, and ρ_{RG} are normalization factors, ensuring that

$$\begin{aligned} \text{WS}_R + \text{WS}_G + \text{WS}_B &= 1 \\ \text{YB}_R + \text{YB}_G + \text{YB}_B &= 1 \\ \text{RG}_R + \text{RG}_G + \text{RG}_B &= 1 \end{aligned} \tag{5.8}$$

Therefore, the transformation matrix can be concluded as a 3×3 matrix $\Gamma_{\delta s}$, where δs denotes the degree of CVD based on the $\Delta\lambda$:

$$\Gamma_{\delta s} = \begin{bmatrix} \text{WS}_R & \text{WS}_G & \text{WS}_B \\ \text{YB}_R & \text{YB}_G & \text{YB}_B \\ \text{RG}_R & \text{RG}_G & \text{RG}_B \end{bmatrix} \tag{5.9}$$

In summary, the general transformation from RGB color space to opponent-color space for CVD can be defined as a 3×3 matrix $\Gamma_{\delta s}$. Let Γ be the transformation matrix for normal viewers, then the CVD simulation of an RGB image can be defined as:

$$\begin{bmatrix} R_{sim} \\ G_{sim} \\ B_{sim} \end{bmatrix} = \Gamma^{-1}\Gamma_{\delta s} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{5.10}$$

Triple-Latent Based Color Disentanglement.

The triple-latent structure consists of a contrastive group to disentangle the color representation and a control group to ensure the personalized generation. Specifically, there are two latent codes in the contrastive group in order to eliminate the dominance of other dimensions $z^{\tilde{d}}$ $\left(\tilde{d} \in (1, D]\right)$ toward the color. To better evaluate the results of disentanglement, we assign $z^{\tilde{d}}$ and z^0 random values.

Fig. 5.1 shows the visualization results of random assignments. For each group divided by the dotted line, the first row presents the images generated from latent codes with random $z^{\tilde{d}}$, while the second row presents the ones generated from random z^0. It is shown that the color distribution in the image is maintained although changes in the $z^{\tilde{d}}$, and it will be

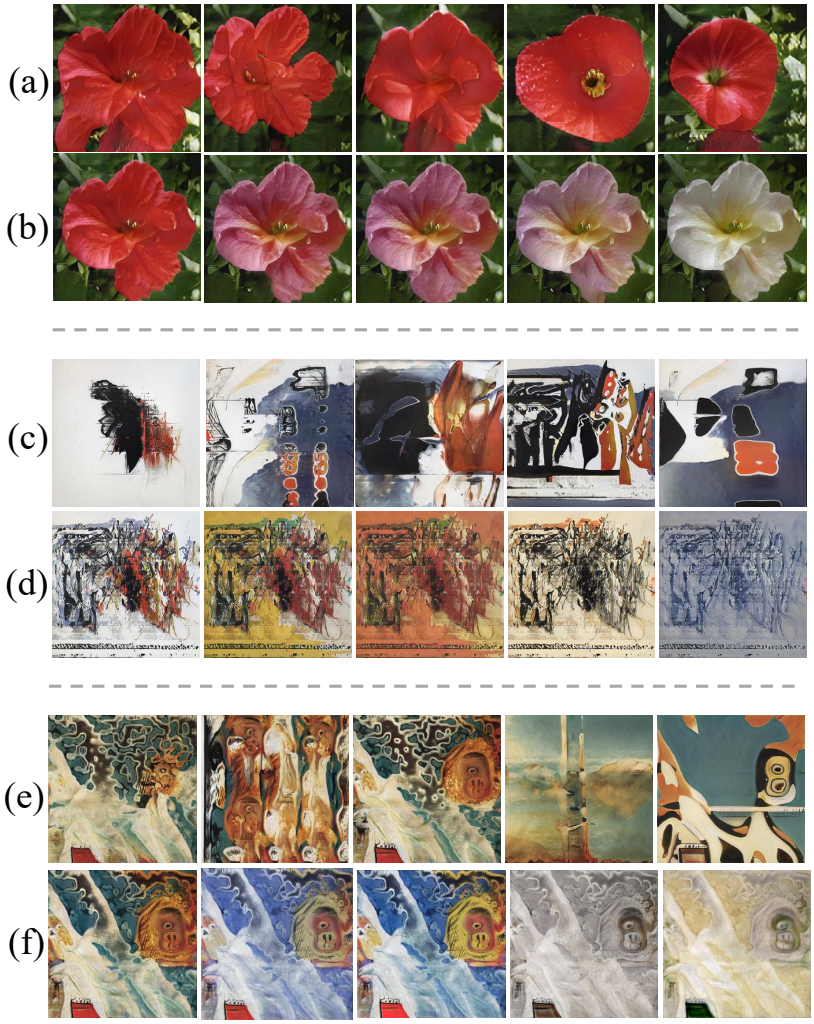

FIGURE 5.1. Examples of color representation disentanglement. For each group divided by the dotted line, the first row presents the images generated from latent codes with random $z^{\tilde{d}}$, while the second row presents the ones generated from random z^0.

modified significantly only when the changes in the z^0, which means the dominance of color representation has been decoupled with $z^{\tilde{d}}$. As a result, the z^0 can dominate the color pattern generation.

With the increment δs on the z^0 during the latent traversal, CVD-GAN can generate personalized images for CVD populations with varying degrees.

FIGURE 5.2. Examples of personalized generation of the symbolic-painting dataset, where "D" denotes deutan- and "P" denotes protan-simulation.

Fig. 5.2, Fig. 5.3, and Fig. 5.4 present the results of personalized generation with an increment of [0.05, 0.2, 0.4, 0.55, 0.7, 0.9, 1.0] on the z^0 and their corresponding simulations. The fewer the change in the image after simulation, the more friendly it is to CVD populations since fewer potential perception biases will occur. It is shown that for all degrees, CVD-GAN can generate friendly images with little perception bias.

User Study

As of now, our user study is still ongoing, and we have successfully recruited 17 CVD volunteers, covering a range of ages from 20 to 54 years old. These participants are categorized

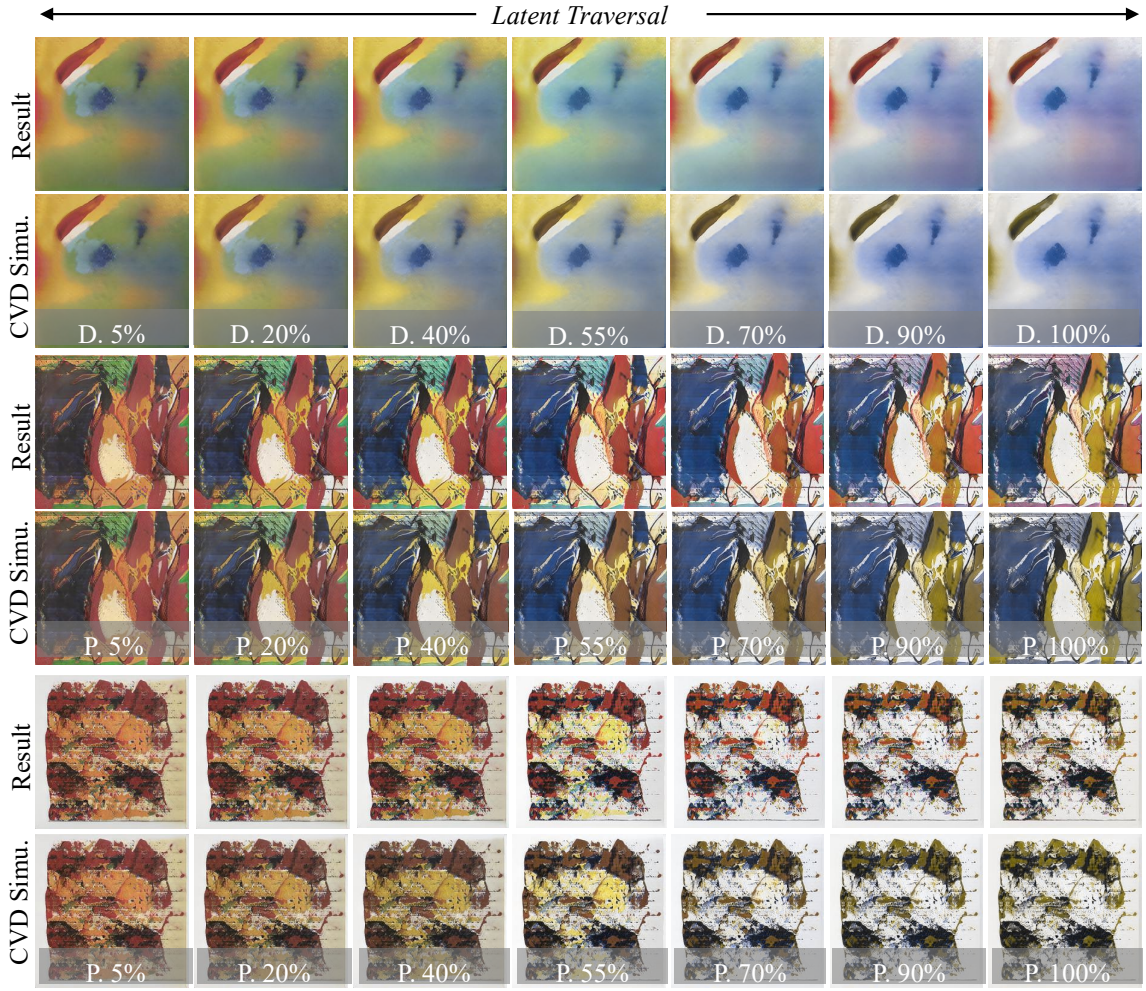

FIGURE 5.3. Examples of personalized generation of the abstract-art dataset.

into three levels: mild, medium, and severe, based on the Hue 100 test. Each volunteer rates 18 randomly selected images generated by three different methods: *StyleGAN (black box)*, *StyleGAN + Zhu (green box)*, and *our CVD-GAN (blue box)* using a Likert scale from 1 to 5. The ratings are based on the clearness and comfort level of the images, where a higher score indicates better results. The current outcomes are as Fig. 5.1.

According to the p-values of the t-test, ours achieves higher marks with statistical significance.

FIGURE 5.4. Examples of personalized generation of the still-life and flower datasets.

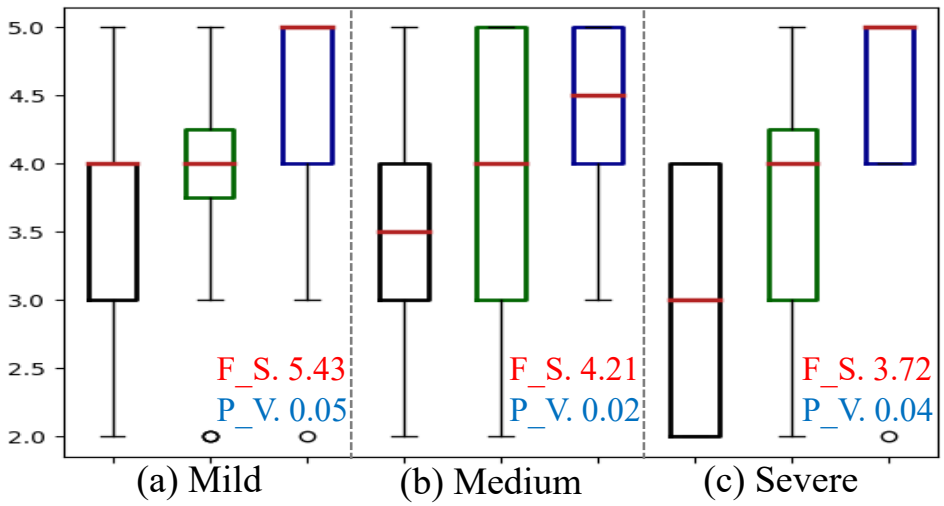

FIGURE 5.5. Result of the user study. (a), (b), and (c) showcase the ranking of populations with mild, medium, and severe CVD degrees, respectively. The notation F_S. indicates the F statistics and P_V represents the statistical significance of the collected preference results.