Personalized Image Generation for Color Vision Deficiency Population

Shuyi Jiang[1], Daochang Liu[1], Dingquan Li[2], Chang Xu[1]
[1]The University of Sydney, [2]Peng Cheng Laboratory
sjia6973@uni.sydney.edu.au, {daochang.liu, c.xu}@sydney.edu.au, lidq01@pcl.ac.cn

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. Instead, we propose a personalized CVD-friendly image generation algorithm with two key characteristics: (i) generating CVD-oriented images aligned with the needs of CVD populations; (ii) generating continuous personalized images for people with various CVD degrees through disentangling the color representation based on a triple-latent structure. Quantitative and qualitative experiments indicate our proposed image generation model can generate practical and compelling results compared to the normal generation model and combination baselines on several datasets. The code is available at: https://github.com/Jiangshuyi0V0/CVD-GAN.git

1. Introduction

In the image generation area, many outstanding generative models, such as variational auto-encoder (VAE) [37, 44], generative adversarial network (GAN) [7, 8, 33, 17, 16], and diffusion models [34, 11], are proposed for high-quality images generation. However, all the generative algorithms are only centric to normal viewers, aiming to facilitate the distribution of generated images close to the dataset established under normal viewers' perspective. The needs of underrepresented populations, like color-impairment populations, are often neglected in image generation tasks, causing perception deviation of the generated images. Currently, 350 million people, a proportion of 8% [22], suffer from color vision deficiency (CVD) resulting from the abnormality of cone cells distributed on the retina of the eyes. Nevertheless, this sizeable population is unintentionally excluded as the target audience of image generation, necessitating the development of an image generation model that is inclusive of all viewers.

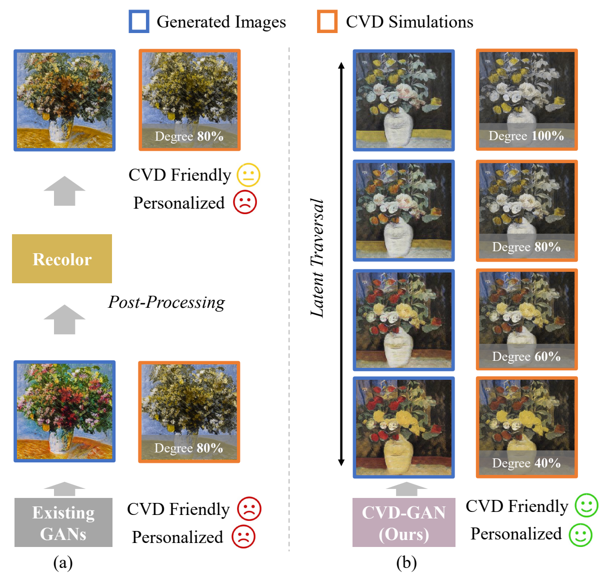

Figure 1. Compared to the combination baseline (a), the proposed CVD-GAN (b) can generate CVD-oriented images directly, enhancing the friendliness of the image for CVD populations. In addition, the model can generate personalized friendly images for CVD populations with varying degrees by disentangling the color representation based on the triple-latent structure.

So far, hardly any generation algorithm has offered to serve CVD populations. Some recoloring algorithms [12, 30, 31, 51, 49, 10, 19, 38, 5] can partly alleviate the problems by post-processing compensation based on the CVD simulation [3, 32] that provides the perspective of CVD populations of the given image. There are two main goals in recoloring methods: restoring the decayed contrast [26, 29, 12, 30, 31, 51] and maintaining the naturalness [49, 51, 10, 19, 38, 5] of a 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 and recolor methods as the post-processing as Fig. 1 (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. Consequently, this approach imposes a likely upper limit on the user-friendliness of the recolored images. Furthermore, despite the fact that CVD populations exhibit diverse requirements based on varying color impairment severity [50, 49], only a few recolor algorithms have addressed the issue of CVD diversity [51] thus far.

To address the above gaps, we propose a CVD-oriented personalized image generation framework based on the adversarial network structure [8], as Fig. 1 (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 [12] 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 [35, 36, 38].

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 [35, 36, 38] show that CVD-GAN can generate CVD-friendly images for CVD populations with varying types and severity.

2. Related Work

Generative Adversarial Network. Recently, the generative adversarial network has been improving in both aspects of image quality [15, 17, 18, 16, 4] and training stability [9, 2, 21]. The generated images have evolved from handwritten digits to complicated images like art painting [42, 43] and high-resolution images [4]. The concept of the adversarial network is also widely applied in various fields [27, 46], indicating the immense potential of it. Despite the success in synthesizing the images, CVD populations are unintentionally excluded as target users and may fail to access the content within those generated images.

GAN Representation Disentanglement. Since the generation process is a "black box", how to disentangle and control the representations is challenging. InfoGAN [6] learned the representations by maximizing the mutual information, StyleGAN [17] proposed the special structure with intermediate latent variables, which can "mix the style" and be progressively fed into the different layers of the generator to control the image style. Besides, many other works were proposed based on the StyleGAN structure, Lee *et al.* [24] fixed the noise of StyleGAN to maintain the target style, and Zhu *et al.* [48] automatically selected the style latent variables for semantic discovery. However, Locatello *et al.* [28] argued that some unsupervised disentangle models might not be reliable enough due to strong dependence on random seeds and hyperparameters through extensive experiments. The paper [28] also suggested that the role of inductive bias should be explicit and practical benefits of disentanglement should be emphasized. Besides, though the representation can be decoupled, how to control the representation [40] during the latent traversal is still underexploited.

Recoloring for CVD Compensation. There are two main goals in recoloring methods: restoring the decayed contrast and maintaining the image's naturalness. To enhance the contrast as well as help CVD users to distinguish the image content, works [12, 31, 30, 51] compensated the contrast by optimizing the objective functions between the given image and recolored image simulation, while works [26, 29] used deep learning networks to perform the color transformation. Lau *et al.* [23] implemented K-means algorithms to enhance the contrast in adjacent areas. To maintain naturalness, works [49, 51, 10, 19, 38] proposed the constraints between the given image and recolored image as a penalized regularization, while Rigos *et al.* [5] deployed the semantic segmentation to transform the colors of objects and keep the other unchanged. Despite all the improvements, the demands of CVD populations with varying degrees are neglected. Zhu *et al.* [51] requested the user to manually input configurations to obtain the corresponding recolored image, which may output inappropriate images due to the

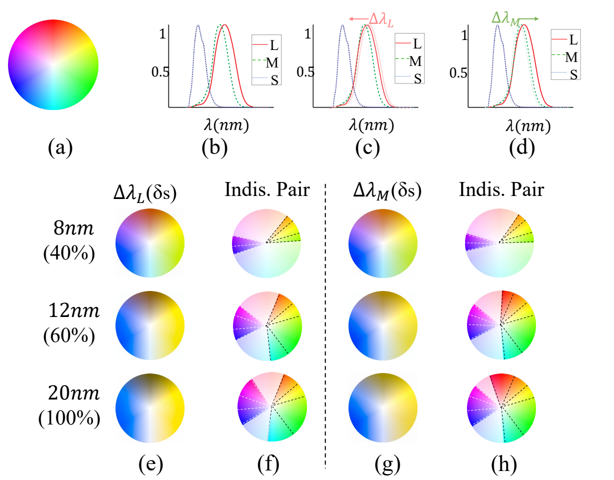

Figure 2. CVD color gamut and cone curves. Compared to the normal viewers' (a) and (b) [41], (c) and (d) are CVD cone curves with a shift of $\Delta\lambda_L$ and $\Delta\lambda_M$; (e) and (g) are the perceptual color gamut under varying severity δs; In (f) and (h), the gamut is indistinguishable between every two dotted lines with the same color. The white area is distinct to individuals with CVD.

sensitivity of the parameters. Personalized recoloring for CVD is still challenging. We aim to achieve personalized generation by disentangling and controlling the color representation in the latent space.

3. Background

Color Gamut of CVD. There are three kinds of cone cells [32] sensitive to long- (e.g. red and orange), medium- (e.g. green and cyan), and short-wavelength (e.g. blue and purple) light called L-, M-, and S- cones respectively. As a result, people with abnormal L-, M- and S- cones' photopigment spectral sensitivity will be referred to as protan, deutan, and tritan respectively, or red-, green-, and blue-weak/blind colloquially. To illustrate, different gamuts will be observed as shown by CVD in Fig. 2. The severity δs of CVD can be estimated as a percentage of the spectral sensitivity curve shift $\Delta\lambda$ relative to 20 nm, as a shift of 20 nm means totally dysfunctional for a cone and equivalent to dichromacy (single-color-blind), where $\Delta\lambda_L$ and $\Delta\lambda_M$ denote the shift on the L- and M- cone accordingly.
CVD Simulation. A two-stage model [32] is implemented to simulate CVD gamut, summarized as:

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

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 supplementary material.

CVD simulation helps normal viewers to perceive the perspective of CVD populations and evaluate the potential perception bias through pure matrix transformations, which are also differential and will be included in our framework.

4. Method

4.1. 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:

$$\begin{aligned}\mathcal{L}_G = \mathbb{E}_{x\sim p_{\text{data}}}\Big[\log\big(1-D(x)\big)\Big]+\\ \mathbb{E}_{z\sim p_{\text{noise}}}\Big[\log\big(1-D(G(z))\big)\Big].\end{aligned} \tag{2}$$

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 shown in Fig. 3, our model consists of two parts based on functional roles. The first part is *CVD-oriented generation* (shown in Fig. 3 (b)), which aims to generate CVD-friendly images with the help of CVD-oriented loss function $\mathcal{L}_{\text{CVD}}$ (in Sec. 4.2). Further, since people with various degrees of CVD have different sensitivities toward perceivable colors, we then implemented *color representation disentanglement* based on the triple-latent structure (shown in Fig. 3 (a)) to meet various needs (in Sec. 4.3).

4.2. 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:

$$\mathcal{L}_{\text{CVD}} = \mathcal{L}_{\text{LC}}(I, \delta s) + \mathcal{L}_{\text{CI}}(I, \delta s), \tag{3}$$

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 (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

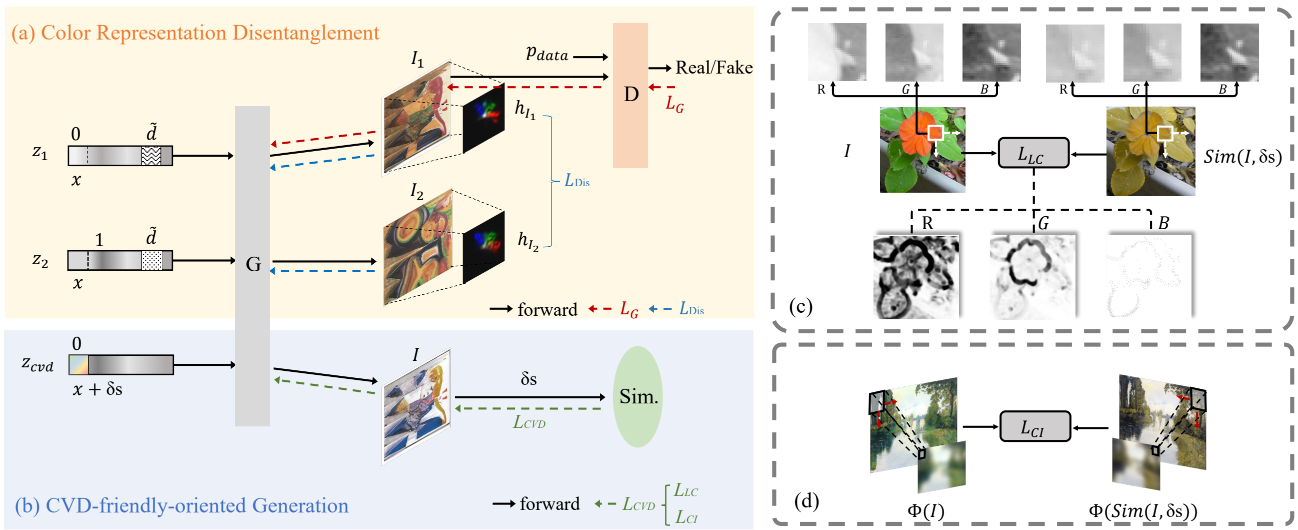

Figure 3. Structure of the CVD-GAN. In (a) and (b), z_1, z_2 and z_{cvd} are three latent codes with size of D. I_1, I_2 and I are images generated by the generator G. To enhance the dominance of the z^0, the dominance of other dimensions needs to be diminished. Hence, $\mathcal{L}_{\text{Dis}}$ is used 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 $\mathrm{Sim.}$ to obtain the specified $\mathrm{Sim}(\cdot)$ 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 $\mathcal{L}_G$.

term of the SSIM [47] is adopted as:

$$c(x,y) = \frac{2\sigma_x\sigma_y + \varepsilon}{\sigma_x^2 + \sigma_y^2 + \varepsilon}, \tag{4}$$

where σ_x and σ_y are the standard deviations of the input patch x and y as the first row of Fig. 3 (c), ε is a small constant to avoid instability. $c(\cdot)$ calculates the contrast similarity between corresponding local maps as Eq. (4). The loss $\mathcal{L}_{LC}$ is computed by aggregating the local contrast losses in patches:

$$\mathcal{L}_{\text{LC}}(I, \delta s) = 1 - \frac{1}{|\mathcal{N}|} \sum_{(x,y)\in\mathcal{N}} c(x,y), \tag{5}$$

where $\mathcal{N}$ is the set of corresponding local maps in the generated image I and its simulation $\mathrm{Sim}(I, \delta s)$; The $\mathcal{L}_{\text{LC}}(I, \delta s)$ can be visualized in RGB channels as the last row of Fig. 3 (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 (d). This optimization process can be summarized as:

$$\mathcal{L}_{\text{CI}}(I, \delta s) = \left\| \Phi(I) - \Phi\big(\mathrm{Sim}(I, \delta s)\big) \right\|_1, \tag{6}$$

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.

4.3. 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

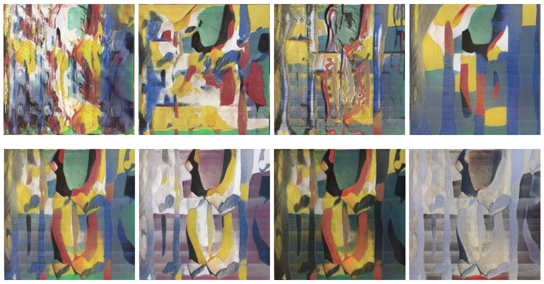

Figure 4. Color representation disentanglement. (a) The influence of the dimension $z^{\tilde{d}}$ on color pattern generation is minimal, as changes in the value of $z^{\tilde{d}}$ result in few alterations to the color distribution, (b) z^0 can dominate the color distribution generation.

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.

For three control latent codes $z_1 = \{z_1^d | d \in [0, D)\}$, $z_2 = \{z_2^d | d \in [0, D)\}$, and $z_{cvd} = \{z_{cvd}^d | d \in [0, D)\}$, 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, for a randomly selected vector dimension $\tilde{d} \in [1, D)$, 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:

$$[I_1, I_2] = G([z_1, z_2]), \tag{7}$$

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:

$$\mathcal{L}_{\text{Dis}} = \frac{1}{\sqrt{2}} || \sqrt{H(I_1)} - \sqrt{H(I_2)} ||_2^2, \tag{8}$$

where $H(\cdot)$ is a operation to obtain the 2D color histogram feature [1], $\|\cdot\|_2^2$ is the L2 norm. An example of color representation disentanglement is shown in Fig. 4. 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.

This increment δs will be fed into the later objective function Eq. (5) and Eq. (6) as the CVD severity to obtain specified constraints as

$$\mathcal{L}_{\text{CVD}} = \mathcal{L}_{\text{LC}}\big(G(z_{cvd}), \delta s\big) + \mathcal{L}_{\text{CI}}\big(G(z_{cvd}), \delta s\big), \tag{9}$$

where $\mathcal{L}_{\text{LC}}(\cdot)$ and $\mathcal{L}_{\text{CI}}(\cdot)$ are local contrast and color information loss functions introduced in Sec. 4.2. 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:

$$\mathcal{L} = \mathcal{L}_G + \alpha \mathcal{L}_{\text{Dis}} + \beta \mathcal{L}_{\text{CVD}}, \tag{10}$$

where α and β are loss weights.

5. Experiment

5.1. Experiments Settings and Datasets

Datasets. To explore the CVD-oriented generation, the datasets [35, 36, 39] with flexible colors were selected. Flower [35] dataset contains 8,189 images with 103 classes. Abstract art [36] 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 [39], 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 [18] with the Adam optimizer [20], 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. 5.4. 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 will be presented in the supplementary material.

5.2. Qualitative Evaluation

The Fig. 5 compares StyleGAN [16], StyleGAN with recolor methods [51, 12], and the proposed CVD-GAN using diverse datasets [35, 36, 39]. Based on the still-life [39] dataset, StyleGAN blurs petals into the background, which remains ambiguous after recolor compensation, hindering CVD populations from distinguishing the content. CVD-GAN avoids confusion by darkening the background as the degree increases and lightening the petals to yellow, as red is imperceptible to protan populations. For the flower dataset [35], StyleGAN generates images with severe decay

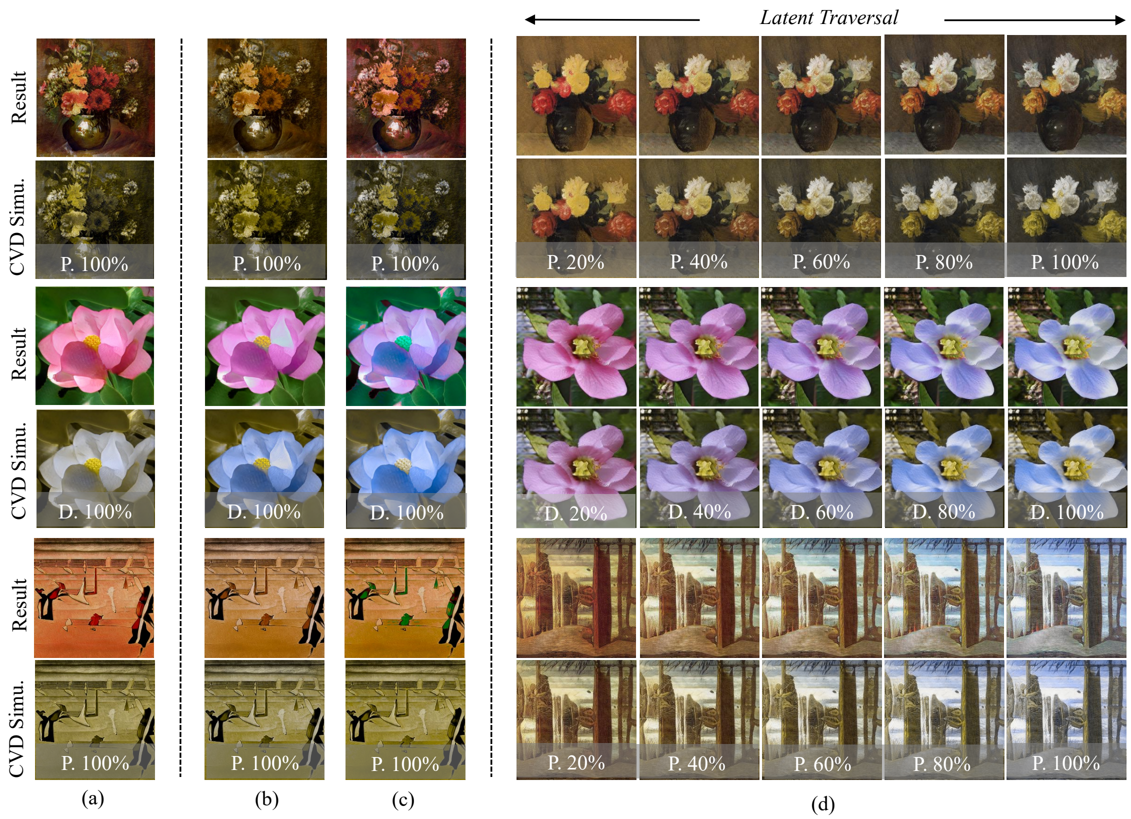

Figure 5. Qualitative comparison. (a) The results of StyleGAN [16], (b) and (c) present the results of StyleGAN with recolor methods [51, 12], (d) shows our results through latent traversal. For each, the first row shows the generation result (or after recolor compensation), and the second row shows the corresponding CVD simulation. "D." and "P." show the degree of deutan and protan, respectively.

of color and contrast, causing perceptual bias. Recolor compensation relieves the gap between normal and CVD perspectives but it still remains. In contrast, CVD-GAN generates CVD-oriented color distribution through latent traversal, with little loss of information after simulation. Similar qualitative results can be obtained through the symbolic painting [39] genre. The effectiveness of CVD-GAN is further validated through a user study, details of which are available in the supplementary material.

5.3. Quantitative Evaluation

Based on the three CVD-friendliness metrics adopted from [13, 1, 14], several experiments will be conducted to compare the results among the generation baseline StyleGAN [16], StyleGAN with post-processing recolor methods [51, 12] and proposed CVD-GAN under various situations of degrees $(20\%, 40\%, 60\%, 80\%, 100\%)$ with two different CVD types (protan and deutan) conditions.

Local Contrast Distance Decay. For a CVD-friendly image, the local contrast is expected to be preserved, otherwise, the image will be ambiguous to distinct. To be specific, decayed Euclidean distance between corresponding local maps of test images and their simulations will be employed [31, 45]. To be noted, test images will be transformed into CIE L*a*b* color space which better represents the human perception of colors [25] than RGB color spaces. The blue column in Table 1 shows the local contrast distance decay for each method.

Hellinger Distance of Color Histogram. To evaluate the main color of the image is whether maintained after simulation, Hellinger distance will be adopted to calculate the distance between color distributions [1] extracted from the test image I and its simulation $\mathrm{Sim}(I, \delta s)$. 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 1 shows the Hellinger distance between generated images and their simulations of the color histogram.

Perceptual Loss. Due to color perception impairment, high-level information except for content details may be lost. As a result, VGG pre-trained model will be adopted to extract the abstract features from the test image I and its simulation $\mathrm{Sim}(I, \delta s)$, then Perceptual loss [14] will be adopted to evaluate CVD-friendliness at the highest level. The pink column in Table 1 shows the perceptual loss be-

Dataset	Type	Degree	StyleGAN [16]			StyleGAN with						CVD-GAN (Ours)		
						Zhu *et al.* [51]			Huang *et al.* [12]					
			LCD	H dis.	Perc. L.	LCD	H dis.	Perc. L.	LCD	H dis.	Perc. L.	LCD	H dis.	Perc. L.
Abstract Art [36]	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 [39]	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 [39]	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 [35]	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 1. Quantitative Results. Comparison with StyleGAN [16] and StyleGAN with recolor methods [51, 12]. 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.

tween generated images and their simulations.

5.4. Ablation Study

CVD Loss Functions. To further discuss the contribution of each of the CVD loss functions, $\mathcal{L}_{\text{LC}}(I, \delta s)$ and $\mathcal{L}_{\text{CI}}(I, \delta s)$ will be ablated to analyze. Note that the experiments are performed in the protan CVD type by default. As Table 2 shows, with the implementation of $\mathcal{L}_{\text{LC}}(I, \delta s)$, 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 $\mathcal{L}_{\text{CI}}(I, \delta s)$. 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 $\mathcal{L}_{\text{Dis}}$, Hellinger distance is used again to calculate the similarity between the color histogram feature extracted from

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

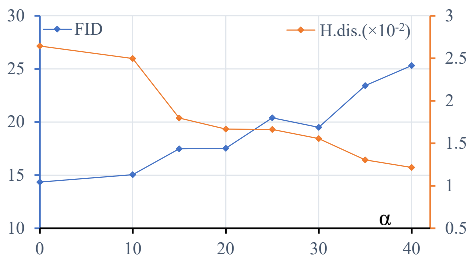

Figure 6. Effect of the color representation disentanglement and accordingly FID. α is the weight of the $\mathcal{L}_{\text{Dis}}$.

the I_1 and I_2 denoted in the Fig. 3. 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. 6 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.

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. 7 presents the relationships between the β and FID metric with CVD metrics introduced in Sec. 5.3. 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, the FID of CVD-GAN on all datasets will be compared to the baseline as the Table 3. More comparisons between CVD-GAN and baseline with post-processing recolor methods under different CVD types and degrees will be presented in the supplementary materials. The proposed contributions are demonstrated to have a minimal impact on image quality in primary datasets with flexible color distributions. However, for natural scenes with fixed color distribution, the change of color may result in a

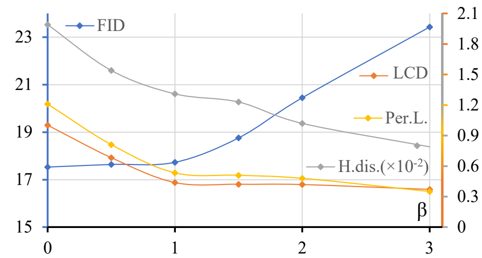

Figure 7. Trade-off of image quality. β is the weight of the $\mathcal{L}_{\text{CVD}}$. The blue, orange, gray, and yellow lines represent the FID, local contrast distance decay, Hellinger distance of color histogram, and perceptual Loss, respectively, based on the β.

Method	Dataset			
	Abstract [36]	Still [39]	Symbolic [39]	Flowers [35]
StyleGAN [16]	14.35	18.96	28.20	8.23
CVD-GAN (Ours)	17.73	22.10	31.66	18.93

Table 3. FID of images generated by StyleGAN and proposed CVD-GAN under various datasets, where the lower value indicates better image quality.

negative effect on image quality.

5.5. Limitations and Future work

CVD-GAN successfully generates personalized CVD-oriented images for protan and deutan types; however, it does not account for viewers with tritan or other complex color impairments due to limited reference samples. Replacing the baseline with alternative generation models may lead to enhanced outcomes. Additionally, further investigation is warranted to explore potential limitations of the recoloring algorithm, particularly concerning the presence of “inherently unfriendly” content. These aspects will be left for future exploration.

6. Conclusion

The paper proposed a personalized CVD-oriented image generation method based on the generative adversarial network, which can generate CVD-oriented and personalized images for varying degrees of CVD populations, adopting deep learning algorithms in the area of underrepresented populations. The model can 1) generate CVD-oriented images end-to-end; 2) generate personalized images for people with various CVD types and degrees by disentangling the color representation based on a triple-latent structure. Our method achieves state-of-the-art performances on several datasets including natural scenes and art paintings.

7. Acknowledgement

This work was supported in part by the Australian Research Council under Project DP210101859 and the University of Sydney Research Accelerator (SOAR) Prize.

References

[1] Mahmoud Afifi, Marcus A Brubaker, and Michael S Brown. HistoGAN: Controlling colors of GAN-generated and real images via color histograms. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7941–7950, 2021.

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223, 2017.

[3] Hans Brettel, Françoise Viénot, and John D Mollon. Computerized simulation of color appearance for dichromats. *Journal of the Optical Society of America A*, 14(10):2647–2655, 1997.

[4] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019.

[5] Stamatis Chatzistamatis, Anastasios Rigos, and George E Tsekouras. Image recoloring of art paintings for the color blind guided by semantic segmentation. In *Engineering Applications of Neural Networks*, pages 261–273, 2020.

[6] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in Neural Information Processing Systems*, 29, 2016.

[7] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[9] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of Wasserstein GANs. *Advances in Neural Information Processing Systems*, 30, 2017.

[10] Mohd Fikree Hassan and Raveendran Paramesran. Naturalness preserving image recoloring method for people with red–green deficiency. *Signal Processing: Image Communication*, 57:126–133, 2017.

[11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[12] Jia-Bin Huang, Chu-Song Chen, Tzu-Cheng Jen, and Sheng-Jyh Wang. Image recolorization for the colorblind. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1161–1164, 2009.

[13] Jia-Bin Huang, Yu-Cheng Tseng, Se-In Wu, and Sheng-Jyh Wang. Information preserving color transformation for protanopia and deuteranopia. *IEEE Signal Processing Letters*, 14(10):711–714, 2007.

[14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711, 2016.

[15] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018.

[16] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In *Advances in Neural Information Processing Systems*, volume 33, pages 12104–12114, 2020.

[17] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.

[18] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020.

[19] Hyun-Ji Kim, Jae-Young Cho, and Sung-Jea 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*, 50(3):91–101, 2013.

[20] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

[21] Naveen Kodali, Jacob Abernethy, James Hays, and Zsolt Kira. On convergence and stability of GANs. *arXiv preprint arXiv:1705.07215*, 2017.

[22] Amir Kosari. Colorblind people population! statistics, Nov 2022. www.colorblindguide.com/post/colorblind-people-population-live-counter, Accessed February 18, 2023 [Online].

[23] Cheryl Lau, Wolfgang Heidrich, and Rafal Mantiuk. Cluster-based color space optimizations. In *International Conference on Computer Vision*, pages 1172–1179, 2011.

[24] Dongyeun Lee, Jae Young Lee, Doyeon Kim, Jaehyun Choi, Jaejun Yoo, and Junmo Kim. Fix the noise: Disentangling source feature for controllable domain translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14224–14234, 2023.

[25] Katherine Leon, Domingo Mery, Franco Pedreschi, and Jorge Leon. Color measurement in L* a* b* units from RGB digital images. *Food Research International*, 39(10):1084–1091, 2006.

[26] Hongsheng Li, Liang Zhang, Xiangdong Zhang, Meili Zhang, Guangming Zhu, Peiyi Shen, Ping Li, Mohammed Bennamoun, and Syed Afaq Ali Shah. Color vision deficiency datasets & recoloring evaluation using GANs. *Multimedia Tools and Applications*, 79:27583–27614, 2020.

[27] Songhua Liu, Kai Wang, Xingyi Yang, Jingwen Ye, and Xinchao Wang. Dataset distillation via factorization. In *Advances in Neural Information Processing Systems*, 2022.

[28] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier

Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *International Conference on Machine Learning*, pages 4114–4124, 2019.

[29] Yu Ma, Xiaodong Gu, and Yuanyuan Wang. Color discrimination enhancement for dichromats using self-organizing color transformation. *Information Sciences*, 179(6):830–843, 2009.

[30] Gustavo Mello Machado. A model for simulation of color vision deficiency and a color contrast enhancement technique for dichromats. Master's thesis, Universidade Federal do Rio Grande do Sul, 2010.

[31] Gustavo M Machado and Manuel M Oliveira. Real-time temporal-coherent color contrast enhancement for dichromats. *Computer Graphics Forum*, 29(3):933–942, 2010.

[32] Gustavo M. Machado, Manuel M. Oliveira, and Leandro A. F. Fernandes. A physiologically-based model for simulation of color vision deficiency. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1291–1298, 2009.

[33] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018.

[34] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171, 2021.

[35] Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *Indian Conference on Computer Vision, Graphics & Image Processing*, pages 722–729, 2008.

[36] George Ogden. Abstract art. www.kaggle.com/datasets/goprogram/abstract-art, Accessed February 18, 2023 [Online].

[37] Ali Razavi, Aaron Van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with VQ-VAE-2. *Advances in Neural Information Processing Systems*, 32, 2019.

[38] Anastasios Rigos, Stamatis Chatzistamatis, and George 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*, 9:5015–5025, 2020.

[39] Babak Saleh and Ahmed Elgammal. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv preprint arXiv:1505.00855*, 2015.

[40] Alon Shoshan, Nadav Bhonker, Igor Kviatkovsky, and Gerard Medioni. Gan-control: Explicitly controllable gans. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14083–14093, 2021.

[41] Vivianne C Smith and Joel Pokorny. Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research*, 15(2):161–171, 1975.

[42] Wei Ren Tan, Chee Seng Chan, Hernán E Aguirre, and Kiyoshi Tanaka. ArtGAN: Artwork synthesis with conditional categorical GANs. In *IEEE International Conference on Image Processing*, pages 3760–3764, 2017.

[43] Wei Ren Tan, Chee Seng Chan, Hernan E Aguirre, and Kiyoshi Tanaka. Improved ArtGAN for conditional synthesis of natural image and artwork. *IEEE Transactions on Image Processing*, 28(1):394–409, 2019.

[44] Jakub Tomczak and Max Welling. VAE with a VampPrior. In *International Conference on Artificial Intelligence and Statistics*, pages 1214–1223, 2018.

[45] Xinyi Wang, Zhenyang Zhu, Xiaodiao Chen, Kentaro Go, Masahiro Toyoura, and Xiaoyang Mao. Fast contrast and naturalness preserving image recolouring for dichromats. *Computers & Graphics*, 98:19–28, 2021.

[46] Yunke Wang, Chang Xu, Bo Du, and Honglak Lee. Learning to weight imperfect demonstrations. In *International Conference on Machine Learning*, pages 10961–10970. PMLR, 2021.

[47] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[48] Xinqi Zhu, Chang Xu, and Dacheng Tao. ContraFeat: Contrasting deep features for semantic discovery. *arXiv preprint arXiv:2212.07277*, 2022.

[49] Zhenyang Zhu, Masahiro Toyoura, Kentaro Go, Issei Fujishiro, Kenji Kashiwagi, and Xiaoyang Mao. Naturalness- and information-preserving image recoloring for red–green dichromats. *Signal Processing: Image Communication*, 76:68–80, 2019.

[50] Zhenyang Zhu, Masahiro Toyoura, Kentaro Go, Issei Fujishiro, Kenji Kashiwagi, and Xiaoyang Mao. Processing images for red–green dichromats compensation via naturalness and information-preservation considered recoloring. *The Visual Computer*, 35:1053–1066, 2019.

[51] Zhenyang Zhu, Masahiro Toyoura, Kentaro Go, Kenji Kashiwagi, Issei Fujishiro, Tien-Tsin Wong, and Xiaoyang Mao. Personalized image recoloring for color vision deficiency compensation. *IEEE Transactions on Multimedia*, 24:1721–1734, 2022.