SDFD: Building a Versatile Synthetic Face Image Dataset with Diverse Attributes

Georgia Baltsou, Ioannis Sarridis, Christos Koutlis and Symeon Papadopoulos
Information Technologies Institute @ CERTH, Greece

***Abstract*— AI systems rely on extensive training on large datasets to address various tasks. However, image-based systems, particularly those used for demographic attribute prediction, face significant challenges. Many current face image datasets primarily focus on demographic factors such as age, gender, and skin tone, overlooking other crucial facial attributes like hairstyle and accessories. This narrow focus limits the diversity of the data and consequently the robustness of AI systems trained on them. This work aims to address this limitation by proposing a methodology for generating synthetic face image datasets that capture a broader spectrum of facial diversity. Specifically, our approach integrates a systematic prompt formulation strategy, encompassing not only demographics and biometrics but also non-permanent traits like make-up, hairstyle, and accessories. These prompts guide a state-of-the-art text-to-image model in generating a comprehensive dataset of high-quality realistic images and can be used as an evaluation set in face analysis systems. Compared to existing datasets, our proposed dataset proves equally or more challenging in image classification tasks while being much smaller in size.**

I. INTRODUCTION

AI systems are typically trained on large scale datasets. However, if such datasets are not balanced and diverse, there is a risk of ending up with unfair and inaccurate AI systems [42]. As a result, AI researchers and developers must be careful when selecting training and evaluation data. This is crucial for creating more reliable AI systems.

A particular application domain that calls for careful dataset composition is that of AI facial image analysis. More specifically, both the training and test images should be sufficiently broad and diverse to cover the numerous ways in which faces naturally differ. In the majority of existing face image datasets [16], [20], diversity is considered through the inclusion of faces from different demographic groups defined on the basis of age, gender and skin tone. However, these dimensions are not sufficient to capture the whole spectrum of facial variety. Previous studies showed that lack of diversity can lead to AI system failures [56]. Face verification systems, for example, may fail due to various types of occlusion [39], [55], [66]. This observation motivated us to rethink of important dimensions in facial appearance and include dimensions related to different hair colors and styles, head and facial accessories or attributes like glasses, hats, tattoos, head-scarfs, and even hijabs, turbans, etc. We focus on creating an evaluation set since we believe it is a vital part of the AI system assessment process.

In the present paper, we present a methodology for generating synthetic face image datasets with the goal of covering a wider range of the spectrum of facial diversity compared to existing ones. To this end, we move beyond demographics and biometrics to non-permanent characteristics like make-up, hair styles, accessories, etc. The proposed dataset creation methodology can be adjusted to the specific situation being examined. We present a particular instance of the proposed strategy, but AI researchers and practitioners may adapt it based on the particular context of use and task. Furthermore, we provide a new face image dataset, called *SDFD*, which can be used as an evaluation set in AI systems. *SDFD* comprises 1000 different face images depicting people of diverse races, genders, ages, wearing a variety of accessories, different types of make-up and expressing various emotions. Despite its relatively small size, the dataset captures a wide variety of different attributes and proves to be a challenging test set.

The contributions of our study are twofold:

- From a **modeling perspective**, we suggest a methodology for generating realistic synthetic face image datasets that can be useful to AI systems evaluation. The current study contributes to the differentiation of existing datasets by taking into account additional face traits beyond demographics and biometrics, which result in covering a wider spectrum of real-world face variety.
- From a **practical perspective**, we provide a first version of the synthetic face image dataset *SDFD*, created via the proposed methodology. This dataset has been designed to be as inclusive as possible in order to assist evaluating computer vision systems with respect to minority groups and outliers.

II. RELATED WORK

The widespread use of AI technologies in different aspects of society has generated concerns about fairness and equity [32], [52]. One area of concern is fair face analysis and recognition, which uses AI algorithms to identify, analyze, and recognize faces. Such processes should be designed to eliminate biases caused by factors such as poor representation in training data, algorithmic design decisions, or social prejudices. Most existing studies have either focused on recognizing bias in AI systems [3], [8], [46], or suggesting techniques to mitigate it [15], [59], [63]. Creating face image datasets that represent faces of diverse groups can help reduce bias in face analysis systems as they can easily represent various demographic groups [2], [20], [35], [36]. The major studies in this topic will be presented next, followed by methods for image generation.

A. Face Image Datasets

Existing face image datasets may consist of real-world or synthetic faces. Nowadays, there is a plethora of real-world face image datasets such as CelebA [28], FFHQ [21], and LFW [18]. However, real-world datasets suffer from some major problems. The first is about how most of these datasets were created. The images, in particular, were retrieved from the Internet, so no specific license was granted [12]. The second issue is that datasets are strongly biased toward specific attributes such as gender, race, or age. More precisely, since the majority of these images portray famous individuals, several of them depict e.g. white people with make-up and professional lighting [24], [35]. Consequently, such datasets do not fairly represent the real-world face distribution. The research community is aware of the former issues and significant efforts have been made to resolve them. DiF (Diversity in Faces) [36] constitutes a face image dataset that is constructed to ensure diversity in age, gender, skin tone, a set of craniofacial ratios, location, and a measure of lighting. Another work in the same direction is FairFace [20], a face image dataset that attempts to mitigate race, gender, and age bias.

Synthetic face image datasets are also explored in many works. In such datasets, the diversity issue can be mitigated by stating all attributes that, if depicted in it, would cover the real-world face variation prior to the image generation process. Furthermore, synthetic datasets have the advantage of reducing unintentional demographic bias caused by uneven representation of demographic groups in training data, since the quantity of images per category can be controlled [35], [53], [56]. An other advantage of AI-generated face images is that these days such images are so realistic that they cannot be separated from actual face photos, as a recent survey [38] reveals. Nevertheless, also works that focus on the generation of face images [4], [12], [13], [16], [23], [33], [45], [64] run into some problems. First, several of these datasets include low-quality or non-realistic images. Another problem, which is similar to that of real-world datasets, is that image generation tools have been mostly trained on images retrieved from the Internet. This might lead to bias in image generation. Besides, most of the synthetic datasets still focus on variety solely on the basis of demographics and biometric features. The objective of the present work is to fill this gap by suggesting a method to generate a fair and realistic face image dataset. Table I summarizes existing datasets along with some important characteristics.

B. Image Generation Methods

Text-to-image generation is an area that is rapidly progressing, with several studies appearing that lead to improved outcomes. As a result, many approaches have been proposed in the last years for generating images using text prompts. [22]. These fall into four main categories [67]: a) Generative Adversarial Networks (GAN) (e.g., StyleCLIP [43], GigaGAN [19]), b) autoregressive methods (e.g., Parti [65], Muse [10]), c) diffusion models (e.g., Imagen [50], DALL-E2 [47], Stable Diffusion [48], InstructPix2Pix [7]), and d) Neural Radiance Fields (NeRFs) (e.g., Magic3D [26], ProlificDreamer [62]). Besides, there are approaches that combine methods of the former categories like the GANDiffFace [34], which combines GAN and Diffusion models.

Despite their very good results, GANs may suffer from *mode collapse*, which causes the generator to produce restricted and repetitive outcomes, making them difficult to train. As stated in [57], there are several works that attempt to alleviate this issue. Furthermore, these approaches require a significant amount of training data to perform well.

Autoregressive techniques may produce high-fidelity images, but there is no guarantee that the whole data distribution is captured. In addition, such methods suffer from slow inference speed.

Diffusion models allow users to adjust the quality and variety of generated images by providing precise control over the generating process [17]. Although such methods can be computationally demanding, they produce remarkably realistic images. Creating realistic images is simply one aspect of an effective text-to-image model. Another key aspect to consider, is whether the output image matches the semantics of the input text description. Diffusion models are experimentally proven to have higher text-image alignment compared to other types of models discussed here [19], [25]. Also, like GAN methods, a substantial amount of training data is required for these approaches to perform successfully.

Finally, NeRF methods can result in 3D images, often of poor quality. In addition, NeRF approaches typically suffer from high computational costs.

III. METHODOLOGY FOR GENERATING SYNTHETIC FACE IMAGE DATASETS

Here, we present the general methodology that someone can follow in order to generate a diverse face image dataset. This involves a systematic process designed to produce diverse and realistic face images using the stable diffusion model, and is briefly illustrated in Figure 1.

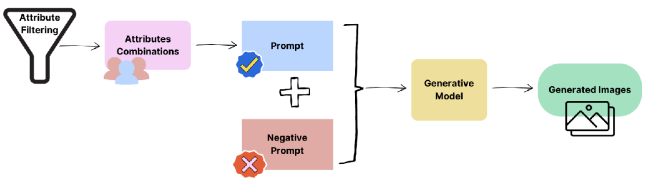


Fig. 1. Overview of proposed image generation process.

The following steps outline the proposed methodology:

1) **Attribute collection and filtering.** The first step is to compile a list of terms that will represent the attributes that one would like to be depicted in the target face images. These terms should be carefully picked and then filtered to eliminate specific words or phrases.
2) **Combinations of attributes.** After completing the basic attribute list, combinations of them should be created to generate the text prompt.

TABLE I

DATASETS AND THEIR MAIN CHARACTERISTICS.

Dataset	source	#Images	Unique Ids	#Labels/Image	Source of Labels
CelebA [28]	Real (Internet)	∼200k	∼10k	40	Professional Labeling Company
FFHQ [21]	Real (Flickr)	70k	N/A	-	-
LFW [18]	Real (Internet)	∼14k	∼6k	1 (identity)	Human annotators
DiF [36]	Real (Flickr)	∼970k	N/A	2	Human annotators
FairFace [20]	Real (Flickr, Twitter, Web)	∼101k	N/A	3	Human annotators
VGGFace2 [9]	Real (Web)	3.31m	∼9k	1 (identity)	Human annotators
BUPT [61]	Real (Web)	2m & 38k	1,3m & 28k	-	-
RFW [60]	Real ([1])	∼40k	∼12k	1 (identity)	Human annotators
Syn-Multi-PIE [64]	Synthetic (StyleGAN2)	100k	100k	70	Automatic Landmark Detection
Syn-LFW [45] (multiple)	Synthetic (DiscoFaceGAN)	multiple	multiple	5	MTCNN [68]
SymFace [16]	Synthetic (StyleGAN)	∼77k	∼26k	-	-
SFace [6]	Synthetic (StyleGAN2-ADA)	∼634k	∼11k	1 (identity)	Human annotators
[33]	Synthetic (StyleGAN)	∼10k	N/A	1 (race)	Downstream Models + Manually
IDiff-Face [5] (multiple)	Synthetic (Diffusion Model)	80k - 500k	5k - 10k	-	-
DigiFace-1M 1 & 2 [4]	Synthetic (Cycles Rendering)	720k & 500k	10k & 100k	-	-
DCFace 1 & 2 [23]	Synthetic (DDPM)	500k & 1m	10k & 20	-	-
Arc2Face [41]	Synthetic (SD v1.5)	21m	1m	-	-

3) **Prompt formulation.** Prompts are used as input text to the generative model, describing the image that should be constructed. The more exact the description, the closer the resulting image will be to what was envisioned. Besides, negative prompts are also important since they specify what should not be included in the generated image.
4) **Generative model.** Prompts, together with negative prompts, act as input to the generative model, which produces the final images. This might be a either a diffusion model, or any other state-of-the-art model.

The above steps might be adjusted depending on the context of use and task. For example, multiple attributes could be utilized, resulting in different prompts. The same is true for the generative model used.

It should be mentioned that the former image generation process may need to be performed several times before the user obtains the desired outcome. This, depends on the generative model used as well as the application.

In the next subsections, we discuss in further depth the steps that we followed to generate the proposed face image dataset and also the requirements we established.

A. Attribute Collection and Filtering

This work's aim is to create a photo-realistic face image dataset that spans the whole range of real-world facial variability. To do this, a list of terms or phrases was compiled that characterize individuals of diverse gender, race, and age, as well as a variety of other characteristics. This list, together with a corresponding list of negative terms, provides the input text according to which images are created.

Race options were selected based on the work of [20]. The race of *Pacific Islanders* was also added, as it is considered a group with distinct characteristics that is not included in the former. The other attributes' options were empirically selected, using as basis previous research works [31], [37], [51], [56]. More specifically, after many experiments with each word or phrase if the resulting image did not depict the input text sufficiently, the former was not used again. E.g., for the non-binary gender attribute 20 different words or phrases were tested before concluding on the four options presented in Table II. All terms used to produce facial images are listed in Table II.

B. Combinations of Attributes

The next step is to create meaningful attribute combinations that may be utilized as the prompt basis. Concerning race, in order to incorporate as many various faces from all over the world as possible, combinations of the eight distinct race groups were also created at random, such as Black-Asian, Indian-White, etc.

Besides race, all the other attributes were combined in an almost random way to create as diverse face images as possible. The combination is not fully random, since there are some combinations that are not allowed due to their incompatibility. For example, *hair style* attribute options except for *bald*, could be combined with a *hair color* attribute option. However, none option of *hat* attribute can be combined with *religious symbol* attribute options.

C. Prompt Formulation

Creating an appropriate prompt is a critical step in the acquisition of photo-realistic facial images [40]. The attribute combinations formed in the previous step are utilized as a prompt, describing the key characteristics of the image to be generated. Aside from this, one should also define traits that should not be displayed in the final image. These are the negative prompt components.

In the present work, some restrictions concerning the (negative) prompts were also applied to generate as realistic images as possible. The following restrictions apply to each generated image in terms of (negative) prompting: (a) must not have the appearance of a computer generated or animated image, (b) must contain only a person's face in a realistic background, and (c) stereotypical traits with reference to any attribute (e.g. Indian person wearing traditional clothes and jewellery) should be eliminated. It was generally attempted to use as objective terms as possible to describe the faces, but there is a possibility that some stereotypical language may have been used. However, the goal was to leverage stereotypical elements to create a more diverse dataset.

TABLE II

TERMS USED IN PROMPTS. OPTIONS ARE PRESENTED IN ALPHABETICAL ORDER.

Attribute	Options
gender	androgynous person, baby, boy, gender fluid person, girl, man, person having an androgynous appeal/male and female characteristics, woman
mood	angry, depressed, excited, happy, sad, scared, smiled, stressed, tired
race	Black, East Asian, Indian, Latino, Middle Eastern, Pacific Islander, Southeast Asian, White
hat	cap, hat, headscarf, helmet, pamela hat
vision	color contact lenses, glasses, sunglasses
hair colour	black, blonde, blue, brown, green, pink, purple, red, white
hair style	bald, braid, curly, dreadlocks, hair clips, headband, long, short, straight
facial hair	beard, moustache
face paints	heavy make-up, lipstick, tattoo
other facial	earrings, medical mask, wrinkles
religious item	al-amira, burqa, chado, ghoongha, hijab, khimar, kippah, niqab, nun hat, snood, tichel, turban, veil
resolution	4K, 8K, Ultra HD
camera	Canon Eos 5D, Fujifilm XT3, Nikon Z9, shot on iPhone
pose	front face, side profile

TABLE III

UNIVERSAL TERMS AND NEGATIVE TERMS USED IN PROMPTS OF ALL EXPERIMENTS.

Prompt	dry skin, realistic background, realistic dull skin noise, remarkable color, remarkable detailed pupils, skin fuzz, textured skin, tone mapping, ultra realistic, visible skin detail
Negative Prompt	animated, bad anatomy, extra fingers, low resolution, unreal engine, worst quality

Aside from these terms, some others were used in all of the experiments regardless of the attributes. The aim was to improve the realism of the generated images. These terms, hereafter called universal, are gathered in Table III.

The prompt formulation process is described in Algorithm 1. More precisely, each prompt is created by combining an attribute options list, the universal prompts and the negative prompts. The universal and negative prompts are consistent, as seen in Table III. The attribute options list is generated in a random way, by selecting one, more or even none option for each attribute as described in the Table II.

Algorithm 1 Prompt Formulation

$p = []$
$a_{optList} = []$
for $a \in attributes$ **do**
    $a_{opt} = random(a)$
    $a_{optList}.append(a_{opt})$
end for
$p.append(a_{optList}) + p.append(p_{uni}) + p.append(p_{neg})$

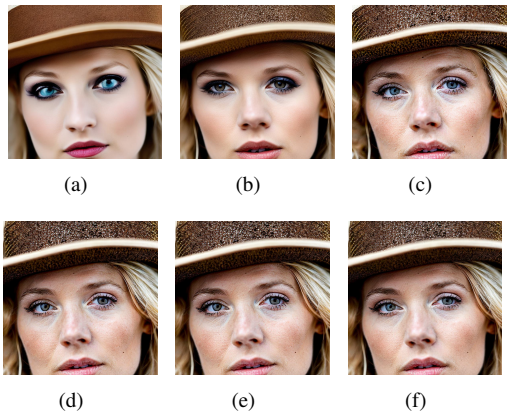
D. Diffusion Process

We selected a Denoising Diffusion Probabilistic Model (DDPM) for text-to-image generation. Diffusion models have been shown to outperform others like GANs on high-quality image generation [11], [14], [44]. These models have strong semantic knowledge on many concepts and have been trained on a wide variety of images, including faces, through language supervision. In particular, the pretrained Stable Diffusion (SD) version 2.1 [48] was used. SD versions 1.5 and xl were also tested, however the quality of generated face images were empirically found to be inferior to that of version 2.1, as shown in Figures 2. In general, the model works by breaking down the image generation process into various denoising steps. The process involves adding random noise to an image and gradually removing it in denoising steps to generate a final image. This diffusion process includes two phases: forward and backward. In the forward process, an image is transformed into noise through multiple steps, and in the backward process, the original image is restored by training a neural network to denoise a given noisy input. The model attempts to forecast the input of the forward process at each step.

(a) (b)

Fig. 2. Examples of generated images with other stable diffusion versions and their corresponding prompts without including universal prompt terms: (a) Stable Diffusion 1.5: White, tired, man, wearing glasses, front face, Ultra HD, Nikon Z9, (b) Stable Diffusion xl Turbo: Black, angry, red hair, androgynous person, wearing glasses, side profile, Fujifilm XT3.

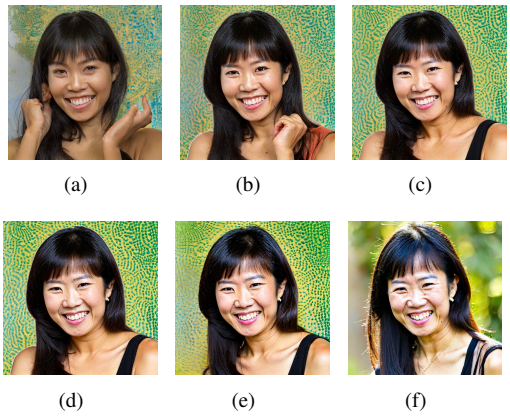
The scheduler is an important parameter when utilizing a diffusion model since it defines the whole denoising process: the number of diffusion steps, how much noise is added on each step, etc. There is a plethora of different schedulers in the literature [27], [30], [54], [69] each trying to make the diffusion process more efficient while preserving good image quality. For the current work, the DPM-Solver++ [30] was selected as it offers possibly the best speed-quality trade-off, with the number of inference (denoising) steps equal to 50. Typically, the more steps, the better the result, but the longer the creation takes. Additionally, a Classifier Free Guidance (CFG) with a weight of 7.5 was used. In general, the CFG weight affects the generator model's amount of flexibility when generating images, since the higher the weight, the greater the level of control by the provided prompt. For stable diffusion, values between 7 and 8.5 are typically appropriate selections [49], [58]. Figures 3, 4 show how different values for inference steps and CFG weight respectively, affect the resulting image.



(a) (b) (c)

(d) (e) (f)

Fig. 3. Example generated images with different values of inference steps for the prompt (apart from universal) "White, blonde woman, blue eyes, wearing hat": (a) 5 steps, (b) 10 steps, (c) 15 steps, (d) 30 steps, (e) 50 steps and (f) 70 steps.



(a) (b) (c)

(d) (e) (f)

Fig. 4. Example generated images with different values of CFG weight for the prompt (apart from universal) "Asian, woman, black hair, smiling": (a) weight = 2.5, (b) weight = 5, (c) weight = 7.5, (d) weight = 10, (e) weight = 12.5 and (f) weight = 15.

Example generated face images are shown in Figure 5. The recommended random method of prompt formulation yields intriguing images that may be difficult to obtain in another way. As a result, the dataset's face diversity is boosted. Figure 6 shows some examples of such images.

It should be emphasized that the "age" attribute is implicitly applied through the attributes of gender, hair colour and other facial attributes. The gender attributes baby, boy and girl directly indicate the age range of a person. In addition, the terms "white hair" and "wrinkles" allude to an older person. Figure 7 displays some generated face images that illustrate the age range of the dataset.

The former process of prompting and applying the stable diffusion model must be repeated as many times as necessary to obtain the final face image dataset. Regarding the number of trials with the same prompt, we determined that repeating an experiment more than 20 times does not increase image quality. Naturally, not every image that is generated in each phase meets the predetermined standards. As a result, the images underwent manual inspection and filtering to ensure their high quality. The criteria used for filtering stemmed from the constraints imposed during the prompt formulation

(a) (b) (c)

Fig. 5. Example generated images and their corresponding prompts without including universal prompt terms: (a) Pacific Islander, stressed, wearing headscarf, girl, black eyes, wearing lipstick, 8K, (b) East Asian, angry, red hair, man, wearing colour contact lenses, front face, Fujifilm XT3 and, (c) White, androgynous person, pink hair, blue eyes, Fujifilm XT3.

(a) (b) (c)

Fig. 6. Examples of some interesting generated images and their corresponding prompts except for universal ones: (a) Latino, stressed, green hair, person having male and female characteristics, front face, Fujifilm XT3, (b) Pacific Islander-Middle Eastern woman, having face tattoo, headband, having wrinkles, 8K, Nikon Z9 and, (c) Latino, baby girl, green hair, beard, black eyes, Fujifilm XT3.

process. Figure 8 shows some examples of images that were excluded from the final dataset.

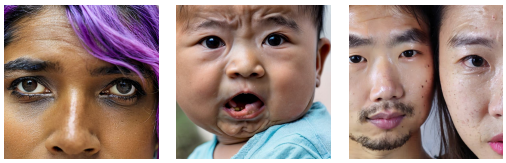
The final version of the *Stable Diffusion Face-image Dataset (SDFD)* consists of 1000 images accompanied with the relevant information. More specifically, it includes the name of each image, along with the corresponding prompt and the time needed to generate the image. We also keep the labels of all the attributes used in each prompt. *SDFD* is publicly available for research use online .

IV. DATASET EVALUATION AND COMPARISON ON ATTRIBUTE CLASSIFICATION

Since FairFace [20] is a face dataset that also aims to reduce bias, we decided to compare it with our generated face dataset (*SDFD*). We also chose the LFW dataset [18] to compare with *SDFD* since it constitutes a real face image dataset that spans the range of conditions typically encountered in everyday life, while it serves as a widely recognized benchmark for face analysis research. The comparison was

(a) (b) (c)

Fig. 7. Example generated images and their corresponding prompts except for universal ones, depicting different age groups: (a) Black, baby boy, short hair, 4K, Fujifilm XT3, (b) White, girl, headband, Ultra HD, Nikon Z9, (c) Southeast Asian-White, man, having piercing, dreadlocks, having wrinkles, Ultra HD, Canon Eos 5D.



(a) (b) (c)

Fig. 8. Generated images that were not selected for the final dataset: (a) the face is very zoomed making it difficult to recognize, (b) the mouth of the generated face is badly shaped and does not seem realistic, (c) the generated image contains two faces instead of one.

performed in terms of both image classification and image distribution in the space, with regard to specific attributes, to conclude if our suggested dataset covers a wider range of face characteristics.

Concerning the image classification task, we trained a ResNet-18 network using two publicly available face image datasets *CelebA dataset* [28] and *UTKface* [70], and then used *SDFD*, FairFace and LFW to infer gender or age as the target attribute. We also conducted experiments using these attributes as well as the race attribute as protected ones, by ensuring that each class of the protected attribute has sufficient representation in each dataset. We have to note that regarding gender classification we excluded from our generated images of *SDFD* the ones with gender label different from male/female. This was done to allow a fair comparison with FairFace and LFW that only include images of these two genders. For the same reason, we labeled *SDFD* images with young/non-young classes. Furthermore, we evaluated the three datasets in terms of certain metrics: accuracy (ACC), precision (PR), recall (R), F1 score (F1), standard deviation (StD), and skewed error ratio (SkER) computed as the ratio of (max group error)/(min group error). These were computed for each individual class separately. Table IV shows the corresponding results. In each row, the dataset that appears to have the worst performance, i.e. is the most challenging, is highlighted in bold. As can be seen, *SDFD* fared worse than FairFace and LFW in both gender and age classification in two of the four cases. We can attribute the lower performance of *SDFD* to its higher diversity, or its higher divergence from the datasets used for training, compared with FairFace and LFW.

According to Table IV, it appears that using the UTKface for training results in lower performance than CelebA, in both gender and age classification.

Table V presents the accuracy of classifying images from *SDFD*, FairFace and LFW using either gender as target attribute and age as the protected one, or vice versa. In each row, the dataset with the poorest performance is highlighted in bold. In this case, the findings demonstrate that *SDFD* and LFW perform worse in age classification but when the target attribute is gender FairFace performs worse.

Table V shows that UTKface continues to lead to lower performance compared with CelebA in terms of gender classification. Regarding the two genders examined, *SDFD* has lower accuracy in female prediction, while FairFace and LFW have the opposite effect. Again, CelebA dataset leads to better accuracy compared with UTKface when using age as the target attribute.

TABLE IV

METRICS EVALUATION ON IMAGE CLASSIFICATION WITH TARGET ATTRIBUTES (GENDER, AGE) AND SDFD, FAIRFACE AND LFW DATASETS.

Dataset	Trained on	Target	ACC	PR	R	F1	StD	SkER
SDFD	CelebA	Gender	0.70	0.65	0.87	0.74	0.52	3.70
FairFace	CelebA	Gender	0.75	0.72	0.86	0.78	0.49	2.52
LFW	CelebA	Gender	0.90	0.94	0.94	0.94	0.31	1.07
SDFD	UTKface	Gender	0.70	0.84	0.51	0.63	0.51	4.94
FairFace	UTKface	Gender	0.58	0.65	0.25	0.36	0.58	5.45
LFW	UTKface	Gender	0.73	0.40	0.40	0.40	0.52	1.02
SDFD	CelebA	Age	0.54	0.39	0.70	0.50	0.62	3.71
FairFace	CelebA	Age	0.68	0.82	0.71	0.76	0.56	1.92
LFW	CelebA	Age	0.58	0.24	0.79	0.37	0.54	12.16
SDFD	UTKface	Age	0.55	0.40	0.80	0.54	0.59	5.93
FairFace	UTKface	Age	0.60	0.74	0.69	0.71	0.63	1.25
LFW	UTKface	Age	0.43	0.36	0.86	0.51	0.59	10.86

TABLE V

ACCURACY EVALUATION ON IMAGE CLASSIFICATION WITH BOTH TARGET AND PROTECTED ATTRIBUTES (GENDER, AGE) AND SDFD, FAIRFACE AND LFW DATASETS.

Dataset	Trained on	Target	Protected	ACC
SDFD	CelebA	Gender	Age (young)	0.77
FairFace	CelebA	Gender	Age (young)	0.70
LFW	CelebA	Gender	Age (young)	0.89
SDFD	UTKface	Gender	Age (young)	0.77
FairFace	UTKface	Gender	Age (young)	0.54
LFW	UTKface	Gender	Age (young)	0.59
SDFD	CelebA	Age	Gender (female)	0.46
FairFace	CelebA	Age	Gender (female)	0.70
LFW	CelebA	Age	Gender (female)	0.77
SDFD	UTKface	Age	Gender (female)	0.38
FairFace	UTKface	Age	Gender (female)	0.58
LFW	UTKface	Age	Gender (female)	0.50
SDFD	CelebA	Age	Gender (male)	0.60
FairFace	CelebA	Age	Gender (male)	0.66
LFW	CelebA	Age	Gender (male)	0.58
SDFD	UTKface	Age	Gender (male)	0.67
FairFace	UTKface	Age	Gender (male)	0.55
LFW	UTKface	Age	Gender (male)	0.33

Table VI presents the accuracy of classifying images from *SDFD*, FairFace and LFW using either gender or age as target attribute and race as the protected one. The Pacific Islander is a race not included in the FairFace dataset, thus there are dashes in the corresponding table cells. The same happens for the LFW for races other than Black, Indian and White. In each row, the dataset that appears to have the worst performance is highlighted in bold. In this case, the findings demonstrate that *SDFD* performs the worst in most cases.

Table VI shows that in gender classification using race as a protected attribute, UTKface performs worse than CelebA. *SDFD* performs poorly in the *Latino* race with CelebA training and the *Pacific Islander* race with UTKface. FairFace underperforms in the *Black* race, regardless of the dataset used for training. In terms of age classification using race as a protected attribute, no training dataset outperforms the other. *SDFD* shows low performance in the *Indian* race with CelebA training and the *Southeast Asian* race with UTKface. FairFace performs poorly in the *Black* race again,

TABLE VI

ACCURACY ON IMAGE CLASSIFICATION WITH BOTH TARGET (GENDER, AGE) AND PROTECTED (RACE) ATTRIBUTES ON SDFD, FAIRFACE AND LFW DATASETS (PI: PACIFIC ISLANDER, ME: MIDDLE EASTERN, SEA: SOUTH-EAST ASIAN, EA: EAST ASIAN).

Dataset	**Trained on**	**Target**	**Protected**	**Accuracy**							
				Black	Indian	PI	ME	White	Latino	SEA	EA
SDFD	CelebA	Gender	Race	0.65	0.70	0.70	0.77	0.66	0.57	0.74	0.78
FairFace	CelebA	Gender	Race	0.67	0.72	-	0.81	0.77	0.78	0.77	0.74
LFW	CelebA	Gender	Race	0.84	0.92	-	-	0.91	-	-	-
SDFD	UTKface	Gender	Race	0.65	0.70	0.59	0.63	0.72	0.70	0.70	0.69
FairFace	UTKface	Gender	Race	0.52	0.57	-	0.69	0.59	0.51	0.58	0.55
LFW	UTKface	Gender	Race	0.72	0.71	-	-	0.73	-	-	-
SDFD	CelebA	Age	Race	0.57	0.50	0.55	0.67	0.52	0.62	0.53	0.52
FairFace	CelebA	Age	Race	0.62	0.66	-	0.66	0.69	0.68	0.69	0.74
LFW	CelebA	Age	Race	0.72	0.71	-	-	0.73	-	-	-
SDFD	UTKface	Age	Race	0.64	0.54	0.45	0.50	0.70	0.54	0.37	0.62
FairFace	UTKface	Age	Race	0.55	0.57	-	0.57	0.58	0.59	0.63	0.67
LFW	UTKface	Age	Race	0.72	0.71	-	-	0.73	-	-	-

independent of the dataset used during training. LFW has the lowest performance in the *Indian* race.

A tentative observation based on the previous results suggests that CelebA and UTKface may not be sufficient for training a model that generalizes well. This could explain the low accuracy values observed during cross-dataset validation. It is worth noting that these datasets contain images with limited diversity, which might contribute to these outcomes. Furthermore, a general observation on the experimental findings demonstrates that *SDFD* proved to be equally challenging for image classification compared with FairFace and LFW, but it is much smaller in size, making it easier and less costly to use.

Aside from comparing the three datasets using image classification, we also visualize their images in two dimensions to analyze their spatial distribution. Figure 9 illustrates the t-SNE visualization [29] of *SDFD*, FairFace and LFW datasets. More precisely, in Figure 9(a) it appears that our suggested dataset, *SDFD*, fills in some of the FairFace's gaps. Considering the sizes of the two datasets—*SDFD* has 1000 images, while FairFace has $10,953$—the former observation appears to be noteworthy. For efficiency considerations, we decided to compare only the FairFace dataset's validation set consisting of $10,953$ images rather than the entire collection ($108,501$ images).

Figure 9(b) displays the t-SNE visualization [29] of *SDFD* and LFW dataset. It appears also here that our suggested dataset, *SDFD*, fills in some of LFW's gaps. Considering the sizes of the two datasets—*SDFD* has 1000 images, while LFW has $14,089$—the former appears to be noteworthy. Additionally, we extracted a set of 40 facial attributes from the two datasets using the *Facer framework* [71]. These attributes are listed in Table VII. We next utilized t-SNE visualization to examine the dispersion of *Facer* characteristics across the datasets. Figure 9(c) illustrates the t-SNE visualization of face attributes of FairFace and *SDFD*, whereas Figure 9(d) illustrates LFW and *SDFD*. In both plots, we can see that, despite its smaller size compared to the FairFace and LFW datasets, *SDFD* covers a large attribute area. The former is more obvious in the plot of FairFace.

TABLE VII

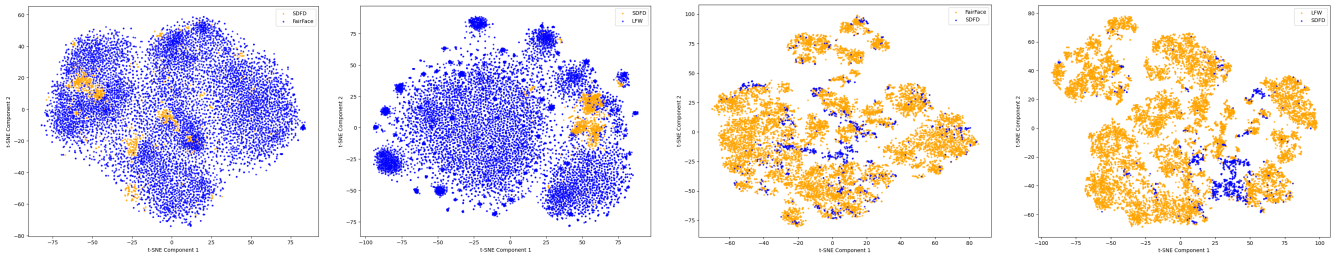
FACE ATTRIBUTES EXTRACTED USING *Facer* FRAMEWORK.

#	**Attribute Name**	**#**	**Attribute Name**
1	5 o' Clock Shadow	21	Male
2	Arched Eyebrows	22	Mouth Slightly Open
3	Attractive	23	Mustache
4	Bags Under Eyes	24	Narrow Eyes
5	Bald	25	No Beard
6	Bangs	26	Oval Face
7	Big Lips	27	Pale Skin
8	Big Nose	28	Pointy Nose
9	Black Hair	29	Receding Hairline
10	Blond Hair	30	Rosy Cheeks
11	Blurry	31	Sideburns
12	Brown Hair	32	Smiling
13	Bushy Eyebrows	33	Straight Hair
14	Chubby	34	Wavy Hair
15	Double Chin	35	Wearing Earrings
16	Eyeglasses	36	Wearing Hat
17	Goatee	37	Wearing Lipstick
18	Gray Hair	38	Wearing Necklace
19	Heavy Makeup	39	Wearing Necktie
20	High Cheekbones	40	Young

Diversity and inclusion are crucial features of the proposed dataset that may not be clearly valued but constitute important characteristics of *SDFD*.

V. DISCUSSION

The presented face image generation approach produced a realistic and high-quality dataset. However, there were several complications encountered during this process. More specifically, there were some terms describing attributes that were not possible or hard to be applied in the final images. Table VIII summarizes these problems along with their corresponding terms. The training datasets of the stable diffusion model could likely be the source of this issue. The former demonstrates the lack of attribute variety in available face image datasets. Although some of the words specified in Table VIII may be too particular to be shown in a generated image, e.g. the religious items, the majority of them should



(a) SDFD - FairFace (b) SDFD - LFW (c) SDFD - FairFace (Facer) (d) SDFD - LFW (Facer)

Fig. 9. t-SNE visualization of embeddings from visual features ((a), (b)) and Facer attributes ((c), (d)) of SDFD versus FairFace and LFW respectively.

TABLE VIII

ATTRIBUTE PROBLEMATIC CASES AND THEIR CORRESPONDING TERMS.

Attribute Problem	**Attribute Terms**
Terms unable to be applied on result image	angry, eyepatch, scar, pimples, freckles, teeth braces, teeth brackets, lips-botox, face lifting, eyeshadow, shayla, shpitzel, sheitel, zuchetto, mitre, chin in hand pose, hands up pose, hands-on-waist pose
Terms needing many iterations to be applied	tattoo
Terms that are not properly applied	excited, sunglasses, non-binary terms (depict two people), color mismatching, eyes' color/wrinkles result in zoomed face, long prompts
Combinations of terms suffering from above issues	heavy makeup+baby, beard+baby, Indian+helmet, bald+helmet

(a) (b) (c)

Fig. 10. Generated images that might reinforce race stereotypes: (a) the prompt given was a Black baby boy with blue eyes. However a White baby was generated, (b) a Pacific Islander girl with leafy background, (c) an Indian girl wearing traditional clothes.

be applicable. Figure 8 shows some examples of the above limitations.

Bias in the generated images was another problem that surfaced. Particularly when specific sentences were used in prompting, stereotypical patterns appeared in the resulting images. For instance, when the race of Pacific Islander was used, most of the times the background of the resulting image was one of foliage. Besides, concerning the Indian race, the majority of generated images depicted people with colorful, traditional attire and jewellery. In other cases, the combination of Black race and colorful eyes mostly resulted in White people with colorful eyes. Figure 10 shows some examples of images that were not selected for the final dataset due to perpetuating stereotypes. Despite our efforts in the opposite direction, the introduced dataset may contain various recognized biases or societal prejudices.

The dataset evaluation and comparison for attribute classification revealed that *SDFD* gives comparable results to the FairFace and LFW datasets for the metrics under consideration. In particular, none of the three datasets performed consistently well. However, in most cases, *SDFD* proved to be equally or even more challenging in image classification, with the obvious benefit of its substantially smaller size. The former makes *SDFD* easier to use as an evaluation dataset.

VI. CONCLUSIONS AND FUTURE WORK

In this work, we present a general methodology to generate synthetic face image datasets, which can be adapted according to the context of use. Following the steps and the specifications outlined, realistic and diversified face image datasets may be created. Furthermore, we introduce a novel face image dataset (*SDFD*), characterized by diversity and inclusiveness. SDFD may be utilized as an evaluation set for models that perform demographic attribute prediction. Including an evaluation set in such tasks facilitates the generalization of models to diverse populations, enhancing their applicability in real-world contexts. When we compared our suggested produced dataset to existing ones, we noticed that our approach made it equally or even more challenging to classify images based on gender and/or age, while being significantly smaller in size, and therefore easier and less costly to use. Although the size of *SDFD* is relatively small, it sufficiently covers the diversity implied by the utilized attributes. Besides, a two-dimensional visualization of the datasets revealed that *SDFD* has a very good spatial dispersion. The former might imply that *SDFD* is a more inclusive dataset with higher face attribute variety. In the future, we are intend to experiment with more generative models and expand the scale of our dataset, while also including even more facial attributes such as scars, braces, and other features that we were unable to integrate in the current dataset version. Finally, we aim to experiment with different traits like disabilities and disfigurements in order to include more underrepresented groups.

VII. ACKNOWLEDGMENTS

This research work was funded by the European Union under the Horizon Europe MAMMOth project, Grant Agreement ID: 101070285.

REFERENCES

[1] Face feature test/trillion pairs. http://http://trillionpairs.deepglint.com/overview. Accessed: 2024-03-05.

[2] M. Alvi, A. Zisserman, and C. Nellåker. Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.

[3] A. Amini, A. P. Soleimany, W. Schwarting, S. N. Bhatia, and D. Rus. Uncovering and mitigating algorithmic bias through learned latent structure. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 289–295, 2019.

[4] G. Bae, M. de La Gorce, T. Baltrušaitis, C. Hewitt, D. Chen, J. Valentin, R. Cipolla, and J. Shen. Digiface-1m: 1 million digital face images for face recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3526–3535, 2023.

[5] F. Boutros, J. H. Grebe, A. Kuijper, and N. Damer. Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19650–19661, 2023.

[6] F. Boutros, M. Huber, P. Siebke, T. Rieber, and N. Damer. Sface: Privacy-friendly and accurate face recognition using synthetic data. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–11. IEEE, 2022.

[7] T. Brooks, A. Holynski, and A. A. Efros. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18392–18402, 2023.

[8] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018.

[9] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 67–74. IEEE, 2018.

[10] H. Chang, H. Zhang, J. Barber, A. Maschinot, J. Lezama, L. Jiang, M.-H. Yang, K. Murphy, W. T. Freeman, M. Rubinstein, et al. Muse: Text-to-image generation via masked generative transformers. *arXiv preprint arXiv:2301.00704*, 2023.

[11] X. Chen and S. Lathuilière. Face aging via diffusion-based editing. *arXiv preprint arXiv:2309.11321*, 2023.

[12] L. Colbois, T. de Freitas Pereira, and S. Marcel. On the use of automatically generated synthetic image datasets for benchmarking face recognition. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2021.

[13] Y. Deng, J. Yang, D. Chen, F. Wen, and X. Tong. Disentangled and controllable face image generation via 3d imitative-contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5154–5163, 2020.

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

[15] S. Dooley, R. Sukthanker, J. Dickerson, C. White, F. Hutter, and M. Goldblum. Rethinking bias mitigation: Fairer architectures make for fairer face recognition. *Advances in Neural Information Processing Systems*, 36, 2024.

[16] M. Grimmer, H. Zhang, R. Ramachandra, K. Raja, and C. Busch. Generation of non-deterministic synthetic face datasets guided by identity priors. *arXiv preprint arXiv:2112.03632*, 2021.

[17] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[18] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[19] M. Kang, J.-Y. Zhu, R. Zhang, J. Park, E. Shechtman, S. Paris, and T. Park. Scaling up gans for text-to-image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10124–10134, 2023.

[20] K. Karkkainen and J. Joo. Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1548–1558, 2021.

[21] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[22] M. Z. Khan, S. Jabeen, M. U. G. Khan, T. Saba, A. Rehmat, A. Rehman, and U. Tariq. A realistic image generation of face from text description using the fully trained generative adversarial networks. *IEEE Access*, 9:1250–1260, 2020.

[23] M. Kim, F. Liu, A. Jain, and X. Liu. Dcface: Synthetic face generation with dual condition diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12715–12725, 2023.

[24] A. Kortylewski, B. Egger, A. Schneider, T. Gerig, A. Morel-Forster, and T. Vetter. Analyzing and reducing the damage of dataset bias to face recognition with synthetic data. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2261–2268. IEEE, 2019.

[25] T. Lee, M. Yasunaga, C. Meng, Y. Mai, J. S. Park, A. Gupta, Y. Zhang, D. Narayanan, H. B. Teufel, M. Bellagente, et al. Holistic evaluation of text-to-image models. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.

[26] C.-H. Lin, J. Gao, L. Tang, T. Takikawa, X. Zeng, X. Huang, K. Kreis, S. Fidler, M.-Y. Liu, and T.-Y. Lin. Magic3d: High-resolution text-to-3d content creation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 300–309, 2023.

[27] L. Liu, Y. Ren, Z. Lin, and Z. Zhao. Pseudo numerical methods for diffusion models on manifolds. In *International Conference on Learning Representations*, 2021.

[28] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[29] M. LJPvd and G. Hinton. Visualizing high-dimensional data using t-sne. *J Mach Learn Res*, 9(2579-2605):9, 2008.

[30] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022.

[31] S. Luccioni, C. Akiki, M. Mitchell, and Y. Jernite. Stable bias: Evaluating societal representations in diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024.

[32] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6):1–35, 2021.

[33] K. A. Mekonnen. Balanced face dataset: Guiding stylegan to generate labeled synthetic face image dataset for underrepresented group. *arXiv preprint arXiv:2308.03495*, 2023.

[34] P. Melzi, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, D. Lawatsch, F. Domin, and M. Schaubert. Gandiffface: Controllable generation of synthetic datasets for face recognition with realistic variations. *arXiv preprint arXiv:2305.19962*, 2023.

[35] P. Melzi, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, A. Morales, D. Lawatsch, F. Domin, and M. Schaubert. Synthetic data for the mitigation of demographic biases in face recognition. *arXiv preprint arXiv:2402.01472*, 2024.

[36] M. Merler, N. Ratha, R. S. Feris, and J. R. Smith. Diversity in faces. *arXiv e-prints*, pages arXiv–1901, 2019.

[37] Microsoft. Azure ai vision documentation. https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/concept-face-detection.

[38] E. J. Miller, B. A. Steward, Z. Witkower, C. A. Sutherland, E. G. Krumhuber, and A. Dawel. Ai hyperrealism: Why ai faces are perceived as more real than human ones. *Psychological Science*, page 09567976231207095, 2023.

[39] M. O. Oloyede, G. P. Hancke, and H. C. Myburgh. A review on face recognition systems: recent approaches and challenges. *Multimedia Tools and Applications*, 79:27891–27922, 2020.

[40] L. Papa, L. Faiella, L. Corvitto, L. Maiano, and I. Amerini. On the use of stable diffusion for creating realistic faces: from generation to detection. In *2023 11th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2023.

[41] F. P. Papantoniou, A. Lattas, S. Moschoglou, J. Deng, B. Kainz, and S. Zafeiriou. Arc2face: A foundation model of human faces. *arXiv preprint arXiv:2403.11641*, 2024.

[42] O. Parraga, M. D. More, C. M. Oliveira, N. S. Gavenski, L. S. Kupssinskü, A. Medronha, L. V. Moura, G. S. Simões, and R. C. Barros. Fairness in deep learning: A survey on vision and language research. *ACM Computing Surveys*, 2023.

[43] O. Patashnik, Z. Wu, E. Shechtman, D. Cohen-Or, and D. Lischinski. Styleclip: Text-driven manipulation of stylegan imagery. In *Proceed-*

ings of the IEEE/CVF International Conference on Computer Vision, pages 2085–2094, 2021.

[44] M. V. Perera and V. M. Patel. Analyzing bias in diffusion-based face generation models. *arXiv preprint arXiv:2305.06402*, 2023.

[45] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, and D. Tao. Synface: Face recognition with synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10880–10890, 2021.

[46] I. D. Raji and J. Buolamwini. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 429–435, 2019.

[47] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. Hierarchical text-conditional image generation with clip latents. *arXiv e-prints*, pages arXiv–2204, 2022.

[48] 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 (CVPR)*, pages 10684–10695, June 2022.

[49] M. Rost and S. Andreasson. Stable walk: An interactive environment for exploring stable diffusion outputs. 2023.

[50] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. Gontijo Lopes, B. Karagol Ayan, T. Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35:36479–36494, 2022.

[51] P. Samangouei, V. M. Patel, and R. Chellappa. Facial attributes for active authentication on mobile devices. *Image and Vision Computing*, 58:181–192, 2017.

[52] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 59–68, 2019.

[53] I. Serna, A. Pena, A. Morales, and J. Fierrez. Insidebias: Measuring bias in deep networks and application to face gender biometrics. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3720–3727. IEEE, 2021.

[54] J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[55] M. Taskiran, N. Kahraman, and C. E. Erdem. Face recognition: Past, present and future (a review). *Digital Signal Processing*, 106:102809, 2020.

[56] P. Terhörst, J. N. Kolf, M. Huber, F. Kirchbuchner, N. Damer, A. M. Moreno, J. Fierrez, and A. Kuijper. A comprehensive study on face recognition biases beyond demographics. *IEEE Transactions on Technology and Society*, 3(1):16–30, 2021.

[57] S. Tomar and A. Gupta. A review on mode collapse reducing gans with gan's algorithm and theory. *GANs for Data Augmentation in Healthcare*, pages 21–40, 2023.

[58] D. Valevski, D. Lumen, Y. Matias, and Y. Leviathan. Face0: Instantaneously conditioning a text-to-image model on a face. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–10, 2023.

[59] M. Wang and W. Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9322–9331, 2020.

[60] M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *Proceedings of the ieee/cvf international conference on computer vision*, pages 692–702, 2019.

[61] M. Wang, Y. Zhang, and W. Deng. Meta balanced network for fair face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 44(11):8433–8448, 2021.

[62] Z. Wang, C. Lu, Y. Wang, F. Bao, C. Li, H. Su, and J. Zhu. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. *arXiv preprint arXiv:2305.16213*, 2023.

[63] Z. Wang, K. Qinami, I. C. Karakozis, K. Genova, P. Nair, K. Hata, and O. Russakovsky. Towards fairness in visual recognition: Effective strategies for bias mitigation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8919–8928, 2020.

[64] E. Wood, T. Baltrušaitis, C. Hewitt, S. Dziadzio, T. J. Cashman, and J. Shotton. Fake it till you make it: face analysis in the wild using synthetic data alone. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3681–3691, 2021.

[65] J. Yu, Y. Xu, J. Y. Koh, T. Luong, G. Baid, Z. Wang, V. Vasudevan, A. Ku, Y. Yang, B. K. Ayan, et al. Scaling autoregressive models for content-rich text-to-image generation. *Transactions on Machine Learning Research*, 2022.

[66] D. Zeng, R. Veldhuis, and L. Spreeuwers. A survey of face recognition techniques under occlusion. *IET biometrics*, 10(6):581–606, 2021.

[67] F. Zhan, Y. Yu, R. Wu, J. Zhang, S. Lu, L. Liu, A. Kortylewski, C. Theobalt, and E. Xing. Multimodal image synthesis and editing: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[68] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016.

[69] Q. Zhang and Y. Chen. Fast sampling of diffusion models with exponential integrator. *arXiv preprint arXiv:2204.13902*, 2022.

[70] Z. Zhang, Y. Song, and H. Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5810–5818, 2017.

[71] Y. Zheng, H. Yang, T. Zhang, J. Bao, D. Chen, Y. Huang, L. Yuan, D. Chen, M. Zeng, and F. Wen. General facial representation learning in a visual-linguistic manner. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18697–18709, 2022.