Second Edition FRCSyn Challenge at CVPR 2024: Face Recognition Challenge in the Era of Synthetic Data

Ivan DeAndres-Tame[1] Ruben Tolosana[1] Pietro Melzi[1] Ruben Vera-Rodriguez[1]
Minchul Kim[2] Christian Rathgeb[3] Xiaoming Liu[2] Aythami Morales[1]
Julian Fierrez[1] Javier Ortega-Garcia[1] Zhizhou Zhong[4] Yuge Huang[5] Yuxi Mi[4]
Shouhong Ding[5] Shuigeng Zhou[4] Shuai He[6] Lingzhi Fu[6] Heng Cong[6]
Rongyu Zhang[6] Zhihong Xiao[6] Evgeny Smirnov[7] Anton Pimenov[7]
Aleksei Grigorev[7] Denis Timoshenko[7] Kaleb Mesfin Asfaw[8] Cheng Yaw Low[9]
Hao Liu[10] Chuyi Wang[10] Qing Zuo[10] Zhixiang He[10] Hatef Otroshi Shahreza[11,12]
Anjith George[11] Alexander Unnervik[11,12] Parsa Rahimi[11,12] Sébastien Marcel[11,13]
Pedro C. Neto[14,15] Marco Huber[16] Jan Niklas Kolf[16] Naser Damer[16]
Fadi Boutros[16] Jaime S. Cardoso[14,15] Ana F. Sequeira[14,15] Andrea Atzori[17]
Gianni Fenu[17] Mirko Marras[17] Vitomir Štruc[18] Jiang Yu[19] Zhangjie Li[19,20]
Jichun Li[19] Weisong Zhao[21] Zhen Lei[22] Xiangyu Zhu[22] Xiao-Yu Zhang[21]
Bernardo Biesseck[23,24] Pedro Vidal[23] Luiz Coelho[25] Roger Granada[25]
David Menotti[23]

[1]Universidad Autonoma de Madrid, Spain [2]Michigan State University, USA [3]Hochschule Darmstadt, Germany
[4]Fudan University, China [5]Tencent Youtu Lab, China [6]Interactive Entertainment Group of Netease Inc, China
[7]ID R&D Inc., USA [8]Korea Advanced Institute of Science & Technology, Korea [9]Institute for Basic Science, Korea
[10]China Telecom AI, China [11]Idiap Research Institute, Switzerland [12]EPFL, Switzerland [13]Université de Lausanne, Switzerland
[14]INESC TEC, Portugal [15]Universidade do Porto, Portugal [16]Fraunhofer IGD, Germany [17]University of Cagliari, Italy
[18]University of Ljubljana, Slovenia [19]Samsung Electronics (China) R&D Centre, China
[20]University of Science and Technology, China [21]IIE, CAS, China [22]MAIS, CASIA, China
[23]Federal University of Paraná, Brazil [24]Federal Institute of Mato Grosso, Brazil [25]unico - idTech, Brazil

Abstract

Synthetic data is gaining increasing relevance for training machine learning models. This is mainly motivated due to several factors such as the lack of real data and intra-class variability, time and errors produced in manual labeling, and in some cases privacy concerns, among others. This paper presents an overview of the 2nd edition of the Face Recognition Challenge in the Era of Synthetic Data (FRCSyn) organized at CVPR 2024. FRCSyn aims to investigate the use of synthetic data in face recognition to address current technological limitations, including data privacy concerns, demographic biases, generalization to novel scenarios, and performance constraints in challenging situations such as aging, pose variations, and occlusions. Unlike the 1st edition, in which synthetic data from DCFace and GANDiffFace methods was only allowed to train face recognition systems, in this 2nd edition we propose new subtasks that allow participants to explore novel face generative methods. The outcomes of the 2nd FRCSyn Challenge, along with the proposed experimental protocol and benchmarking contribute significantly to the application of synthetic data to face recognition.

1. Introduction

Face biometrics is a very popular area in the fields of Computer Vision and Pattern Recognition, finding applications in diverse domains such as person recognition [13, 47], healthcare [4, 17], or e-learning [10], among others. In the last years, with the rapid evolution of deep learning, we have witnessed a considerable performance improvement in areas such as face recognition (FR) [12, 23], outperforming the state-of-the-art on established benchmarks.

Figure 1. Examples of synthetic identities and variations for different demographic groups using GANDiffFace [27].

However, FR technology has still room for improvement in several research directions, such as explainability [9, 11], demographic bias [28, 42], privacy [29, 32], and robustness against challenging conditions [23], *e.g.,* aging, pose variations, illumination, occlusions, etc.

Synthetic data has recently appeared as a good solution to mitigate some of these drawbacks, allowing the generation of *i)* a huge number of facial images from different non-existent identities, and *ii)* variability in terms of demographic attributes and scenario conditions. Several approaches have been proposed in the last couple of years for the synthesis of face images, considering state-of-the-art deep learning methods such as Generative Adversarial Networks (GANs) [36, 51], Diffusion models [6, 20, 24], the combination of GAN and Diffusion models [27], or alternative methods [3, 49]. Examples of synthetic face images generated using GANDiffFace are shown in Figure 1.

However, beyond the generation of novel and realistic synthetic faces, a critical aspect lies in the possible application and benefits of synthetic data to better train FR technology. Recent preliminary studies in the literature have shown the existence of a performance gap between FR systems trained solely on synthetic data and those trained on real data [24, 36]. Nevertheless, the results achieved in the 1st edition of the Face Recognition Challenge in the Era of Synthetic Data (FRCSyn) demonstrate the importance of using synthetic data by itself or in combination with real data to mitigate challenges in FR such as demographic bias [30, 31]. It is important to highlight that in the 1st edition of the FRCSyn Challenge, only synthetic data from DCFace [24] and GANDiffFace [27] methods was allowed to train FR systems. In addition to novel generative methods, another possible improvement of the FR technology could be related to the specific design and training process, taking into account the domain gap between real and synthetic data in some scenarios. For example, we observed in the 1st edition of the FRCSyn Challenge that the majority of the teams considered the same deep learning architectures (*e.g.,* ResNet-100 [18]) and loss functions (*e.g.,* AdaFace [23]), popularly considered in FR systems trained with real data.

To promote the proposal of novel face generative methods and the creation of face synthetic databases, as well as specific approaches to better train FR systems with synthetic data, we have organized the 2nd edition of the FRCSyn Challenge as part of CVPR 2024[1]. In this 2nd edition, we introduce new sub-tasks enabling participants to train FR systems utilizing synthetic data obtained with the generative frameworks of their choice, offering more freedom compared to the 1st edition [30, 31]. In addition, we also consider new sub-tasks featured with different experimental settings, to investigate how FR systems can be trained in both constrained and unconstrained scenarios concerning the amount of synthetic training data. The FRCSyn Challenge aims to answer the following research questions:

1. What are the limits of FR technology trained only with synthetic data?
2. Can the use of synthetic data be beneficial to reduce the current limitations in FR technology?

These research questions have gained significant importance, particularly after the discontinuation of popular real FR databases due to privacy concerns[2] and the introduction of new regulatory laws[3].

The remainder of the paper is organized as follows. Section 2 focuses on the databases considered in this 2nd edition. Section 3 explains the experimental setup of the challenge, including the different tasks and sub-tasks, the experimental protocol, metrics, and restrictions. In Section 4, we describe the approaches proposed by the top-6 participating teams. Section 5 presents the best results achieved in the different tasks and sub-tasks of the 2nd edition, emphasizing the key results of the challenge. Finally, in Section 6, we provide some conclusions, highlighting potential future research directions in the field.

[1] https://frcsyn.github.io/CVPR2024.html
[2] https://exposing.ai/about/news/ *(March, 2024)*
[3] https://artificialintelligenceact.eu *(March, 2024)*

2. FRCSyn Challenge: Databases

2.1. Synthetic Databases

One of the main novelties of this 2nd edition of the FRCSyn Challenge is the absence of restrictions on the generative methods allowed to create synthetic data, unlike the 1st edition in which only synthetic data created using DCFace [24] and GANDiffFace [27] methods was allowed. As a reference, after the registration in the challenge, we provided all the participants with a list of possible state-of-the-art generative frameworks, including DCFace [24], GANDiffFace [27], DigiFace-1M [3], IDiff-Face [6], ID3PM [20], SFace [51], SYNFace [36], and ITI-GEN [49]. In addition, we also motivate participants to propose novel face generative methods. In the 2nd edition of the FRCSyn Challenge, synthetic data is exclusively utilized in the training stage of FR technology, replicating realistic operational scenarios.

2.2. Real Databases

For the training of the FR systems (depending on the sub-task, please see Section 3.1 for more details), participants are allowed to use only **CASIA-WebFace** [48]. This database contains $494,414$ face images of $10,575$ real identities collected from the web.

For the final evaluation of the proposed FR systems, we use the same four real databases of the 1st edition of the FRCSyn Challenge [30, 31]: *i)* **BUPT-BalancedFace** [46], designed to address performance disparities across different ethnic groups; *ii)* **AgeDB** [33], including facial images of the same subjects at different ages; *iii)* **CFP-FP** [37], presenting facial images from subjects with great changes in pose, including both frontal and profile images; and *iv)* **ROF** [15], consisting of occluded faces with both upper and lower face occlusions.

3. FRCSyn Challenge: Setup

3.1. Tasks

Similar to the 1st edition [30, 31], the challenge has been hosted on Codalab[4]. In this 2nd edition we also explore the application of synthetic data into the training of FR systems, with a specific focus on addressing two critical aspects in current FR technology: *i)* mitigating demographic bias, and *ii)* enhancing overall performance under challenging conditions that include variations in age and pose, the presence of occlusions, and diverse demographic groups. To investigate these two areas, we propose two distinct tasks, each comprising three sub-tasks considering different types (real/synthetic) and amounts of data for training the FR systems. Consequently, the 2nd edition of the FRCSyn Challenge comprises 6 different sub-tasks. In Table 1, we summarize the key aspects of the experimental protocol, metrics, and restrictions for each sub-task.

Task 1: synthetic data for **demographic bias mitigation** **Baseline**: training with only CASIA-WebFace [48]. **Metrics**: accuracy (for each demographic group). **Ranking**: average vs SD accuracy, see Section 3.3 for more details.
Sub-Task 1.1: [constrained] training exclusively with **synthetic** data **Train**: maximum $500K$ face images (*e.g.,* $10K$ identities and 50 images per identity). **Eval**: BUPT-BalancedFace [46].
Sub-Task 1.2: [unconstrained] training exclusively with **synthetic** data **Train**: no restrictions in terms of the number of face images. **Eval**: BUPT-BalancedFace.
Sub-Task 1.3: [constrained] training with **real and synthetic** data **Train**: CASIA-WebFace, and maximum $500K$ face synthetic images. **Eval**: BUPT-BalancedFace.
Task 2: synthetic data for **overall performance improvement** **Baseline**: training with only CASIA-WebFace. **Metrics**: accuracy (for each evaluation database). **Ranking**: average accuracy, see Section 3.3 for more details.
Sub-Task 2.1: [constrained] training with only **synthetic** data **Train**: maximum $500K$ face images. **Eval**: BUPT-BalancedFace, AgeDB [33], CFP-FP [37], and ROF [15].
Sub-Task 2.2: [unconstrained] training with only **synthetic** data **Train**: no restrictions in terms of the number of face images. **Eval**: BUPT-BalancedFace, AgeDB, CFP-FP, and ROF.
Sub-Task 2.3: [constrained] training with **real and synthetic** data **Train**: CASIA-WebFace, and maximum $500K$ face synthetic images. **Eval**: BUPT-BalancedFace, AgeDB, CFP-FP, and ROF.

Table 1. Tasks and sub-tasks for the 2nd FRCSyn Challenge and their respective metrics and databases. SD = Standard Deviation.

Task 1: The first proposed task focuses on the use of synthetic data to mitigate demographic biases in FR systems. To assess the effectiveness of the proposed systems, we generate lists of mated and non-mated comparisons using subjects from the BUPT-BalancedFace database [46]. We take into account eight demographic groups obtained from the combination of four ethnic groups (White, Black, Asian, and Indian) and two genders (Male and Female), and keep these groups balanced in the number of comparisons. In the case of non-mated comparisons, we only consider pairs of subjects within the same demographic group, as these hold greater relevance than non-mated comparisons involving subjects from different demographic groups.

Task 2: The second proposed task focuses on utilizing synthetic data to enhance the overall performance of FR systems under challenging conditions. To assess the effectiveness of the proposed systems, we utilize lists of mated and non-mated comparisons selected from subjects from the different evaluation databases, each one designed to address specific challenges in FR. Specifically, BUPT-BalancedFace is used to consider diverse demographic groups, whereas AgeDB, CFP-FP, and ROF to assess age, pose, and occlusion, respectively.

3.2. Experimental protocol

Training: The 6 sub-tasks introduced in the 2nd edition of the FRCSyn Challenge are mutually independent. This implies that participants have the flexibility to participate in any number of sub-tasks based on their preferences. For

[4] https://codalab.lisn.upsaclay.fr/competitions/16970

each selected sub-task, participants are required to develop and train the same FR system twice: *i)* using the authorized real database exclusively, i.e. CASIA-WebFace [48], and *ii)* following the specific requirements of the chosen sub-task, as summarized in Table 1. According to this protocol, participants must provide both the *baseline system* and the *proposed system* for the specific sub-task. The baseline system plays a critical role in evaluating the impact of synthetic data on training and serves as a reference point for comparing against the conventional practice of training solely with real databases. To maintain consistency, the baseline FR system, trained exclusively with real data, and the proposed FR system, trained according to the specifications of the selected sub-task, must have the same architecture.

Evaluation: In each sub-task, participants received the comparison files comprising both mated and non-mated comparisons, which are used to evaluate the performance of their proposed FR systems. Task 1 involves a single comparison file containing balanced comparisons of different demographic groups of the BUPT [46] database, while Task 2 comprises four comparison files, each corresponding to each of the specific real-world databases considered (i.e., BUPT, AgeDB [33], CFP-FP [37], and ROF [15]). During the evaluation of each sub-task, participants are required to submit via Codalab three files per database: *i)* the scores of the baseline system, *ii)* the scores of the proposed system, and *iii)* the decision threshold for each FR system *(i.e.,* baseline and proposed). The submitted scores must fall within the range of $[0, 1]$, with lower scores indicating non-mated comparisons, and vice versa.

3.3. Evaluation Metrics

We evaluate FR systems using a protocol based on lists of mated and non-mated comparisons for each sub-task and database. From the scores and thresholds provided by participants, we calculate the binary decision and the verification accuracy. Additionally, we calculate the gap to real (GAP) [24] as follows: $\text{GAP} = (\text{REAL} - \text{SYN})/\text{SYN}$, with REAL representing the verification accuracy of the baseline system and SYN the verification accuracy of the proposed system, trained with synthetic (or real + synthetic) data. Other metrics such as False Non-Match Rate (FNMR) at 1% False Match Rate (FMR), which are very popular for the analysis of FR systems in real-world applications, can also be computed from the scores provided by participants. Due to the lack of space, comprehensive evaluations of the proposed systems will be conducted in subsequent studies, including FNMRs and metrics for each demographic group and database used for evaluation. Next, we explain how participants are ranked in the different tasks.

Task 1: To rank participants and determine the winners of Sub-Tasks 1.1, 1.2, and 1.3, we closely examine the trade-off between the average (AVG) and standard deviation (SD) of the verification accuracy across the eight demographic groups defined in Section 3.1. We define the trade-off metric (TO) as follows: $\text{TO} = \text{AVG} - \text{SD}$. This metric corresponds to plotting the average accuracy on the x-axis and the standard deviation on the y-axis in 2D space. We draw multiple 45-degree parallel lines to find the winning team whose performance falls to the far right side of these lines. With this proposed metric, we reward FR systems that achieve good levels of performance and fairness simultaneously, unlike common benchmarks based only on recognition performance. The standard deviation of verification accuracy across demographic groups is a common metric for assessing bias and should be reported by any work addressing demographic bias mitigation.

Task 2: To rank participants and determine the winners of Sub-Tasks 2.1, 2.2, and 2.3, we consider the average verification accuracy across the four databases used for evaluation, described in Section 3.1. This approach allows us to evaluate four challenging aspects of FR simultaneously: *i)* diverse demographic groups, *ii)* pose variations, *iii)* aging, and *iv)* presence of occlusions providing a comprehensive evaluation of FR systems in real operational scenarios.

3.4. Restrictions

Regarding the FR system, participants have the freedom to choose any architecture for each sub-task, provided that the system's number of Floating Point Operations Per Second (FLOPs) does not exceed 50 GFLOPs. This threshold has been established to facilitate the exploration of innovative architectures and encourage the use of diverse models while preventing the dominance of excessively large models. Participants are also free to utilize their preferred training modality, with the requirement that only the specified databases are used for training. This means that no additional databases can be employed during the training phase, such as to adapt the verification thresholds. Participants are allowed to use non-face databases for pre-training purposes and employ traditional data augmentation techniques using the authorized training databases. Regarding the synthetic data used to train the FR system in each sub-task, we allow participants to use any existing/novel database and face generative framework, regardless of how the model is trained.

To maintain the integrity of the evaluation process, the organizers reserve the right to disqualify participants if anomalous results are detected or if participants fail to adhere to the challenge's rules.

4. FRCSyn Challenge: Systems Description

The 2$^{\text{nd}}$ edition of the FRCSyn Challenge received significant interest, with 78 international teams correctly registered, comprising research groups from both industry and academia. These teams work in various domains, including FR, generative AI, and other aspects of computer vi-

Team	Affiliations	Country	Sub-Tasks
ADMIS	4, 5	China	all
OPDAI	6	China	all
ID R&D	7	USA	all
K-IBS-DS	8, 9	South Korea	all
CTAI	10	China	all
Idiap-SynthDistill	11, 12, 13	Switzerland	1.2 - 2.2
INESC-IGD	14, 15, 16	Portugal and Germany	all
UNICA-IGD-LSI	16, 17, 18	Italy, Germany, Slovenia	all
SRCN_AIVL	19, 20, 21, 22	China	1.1
CBSR-Samsung	19, 21, 22	China	1.3 - 2.3
BOVIFOCR-UFPR	23, 24, 25	Brazil	1.2 - 2.1

Table 2. Description of the teams that ranked among the top-6 in at least one sub-task, ordered by the average rank in all the sub-tasks. The numbers reported in the column 'affiliations' refer to the ones provided in the title page.

sion, such as demographic fairness and domain adaptation. Finally, 23 teams submitted their scores, receiving all sub-tasks great attention. The submitting teams are geographically distributed, with fourteen teams from Asia, six teams from Europe, and three teams from America. Table 2 provides a general overview of the teams that ranked among the top-6 in at least one sub-task, including the sub-tasks in which they participated. Next, we describe briefly the approaches proposed for each team.

ADMIS (All sub-tasks): They used an IDiff-Face-based [6] Latent Diffusion Model (LDM) to synthesize face images. Specifically, they trained an identity-conditioned LDM using ID embeddings extracted from CASIA-WebFace [48] with a pre-trained ElasticFace [5] IResNet-101 [14] model. As the LDM takes the ID embeddings as context, they employed an unconditional Denoising Diffusion Probabilistic Model (DDPM) trained on the FFHQ database [21] as a context generator. This produced $400K$ images, from which they extracted approximately $30K$ unique ID embeddings with a 0.3 similarity threshold using the pre-trained ElasticFace model, creating a context database. Furthermore, they accelerated the sampling process of the LDM using a DDIM [41]. For the training of the FR model, they generated 49 images for each context. They adopted the ID oversampling strategy from DCFace [24] and performed it five times for each ID to enhance consistency. As a result, $10K$ contexts were utilized for Sub-Tasks 1.1 and 2.1, while $30K$ for Sub-Tasks 1.2 and 2.2. For Sub-Tasks 1.3 and 2.3, they expanded Sub-Tasks 1.1 and 2.1 with the CASIA-WebFace database. They applied the ArcFace [12] loss and random cropping augmentation during training. Both the baseline and proposed models used IResNet-101 architectures [14].
Code: https://github.com/zzzweakman/CVPR24_FRCSyn_ADMIS

OPDAI (All sub-tasks): They initially used the data provided by DCFace [24], generating then 10 more face images for each ID with large pose variations and occlusions using Photomaker [25]. They randomly replaced these images in the original DCFace data to ensure that the total number of samples meets the requirement of $500K$. During the Photomaker inference, they adopted a batch size of 1 and used random prompts including age, pose, and image quality to ensure the diversity of the generated samples. For Sub-Tasks 1.2 and 2.2, they combined this data with the $1.2M$ version of DCFace, while for Sub-Tasks 1.3 and 2.3, it was merged with CASIA-WebFace [48]. For Sub-Tasks 1.2, 1.3, 2.2, and 2.3 they did not merge nor denoise samples from different databases, following the Partial FC approach [1]. Also, they obtained the loss of different databases in independent AdaFace [23] heads, calculating the final loss as the average of the multiple heads. Both baseline and proposed models are based on IResNet-100 [14] architectures, with horizontal flipping.
Code: https://github.com/mightycatty/frcsyn_cvpr2024.git

ID R&D (All sub-tasks): To generate the synthetic data, they used two models trained on WebFace42M [53], one based on Hourglass Diffusion Transformers [8] and the other on StyleNAT [43], enhanced with a FR model [40]. They used classifier weights of the trained Prototype Memory [40] to get $50K$ identity vectors, of which $20K$ were randomly selected and $30K$ were uniformly sampled from the $1K$ clusters obtained using k-means, to get demographic diversity. For each identity, they generated 5 images using each of the two generative models. This data was used to train IResNet-200 [14] with UniFace [52] loss for 28 epochs. One network was trained with color, geometric augmentations, and FaceMix-B [16], and the other network used only random horizontal flipping. These two networks were combined in an "ensemble", where the first one received the original image, and the second one a mirrored copy. They used the same model for Sub-Tasks 1.1, 1.2, 2.1 and 2.2. For Sub-Tasks 1.3 and 2.3, they combined the synthetic data and CASIA-WebFace, training two models, one on the mixed data, and the other on the CASIA-WebFace.

K-IBS-DS (All sub-tasks): Inspired by Slacked-Face [26], they made two modifications to enhance the AdaFace [23] FR classifier. First, they made a more reliable weight initialization for uniformity across identity prototypes in the unit sphere and replaced the L2-norm with the face recognizability index from [26]. Regarding the synthetic data, they used DCFace [24] with $500K$ and $1.2M$ face images (depending on the sub-task). The training stage was in line with [24] and [3], including optimizer, learning rate, etc. For Sub-Tasks 1.3 and 2.3, the first $10K$ subjects of the CASIA-WebFace [48] were assigned for training, and the remaining ones for performance validation using random pairs with challenging conditions (identified

based on the poorest L2-norm values [27]). The final score is obtained by aggregating the comparison scores of ResNet with Squeeze-and-Excitation (SE) blocks [19] models of 50, 100, and 152 layers, along with the horizontally flipped instances through score fusion.
Code: https://github.com/kalebmes/cvpr_frcsyn

CTAI (All sub-tasks): By analyzing popular synthetic data, they found that intra-class and inter-class noise was widely present. Data cleaning can effectively remove the bad examples of synthetic data and retain important images from a large amount of synthetic data. In order to select the optimal synthetic data, they first trained an IResNet-100 [14] model with Squeeze-and-Excitation (SE) [19] blocks using CASIA-WebFace [48] to extract features of synthetic images from DCFace [24], GANDiffFace [27], and DigiFace [3]. Subsequently, they used DBSCAN clustering to segregate intra-class noise and removed IDs with a class center feature cosine similarity greater than 0.5. Finally, they used the cleaned synthetic data merged with CASIA-WebFace to finetune the IResNet-100 for a second data refinement. From the final refined synthetic dataset, they sampled $500K$ face images while retaining as many IDs as possible to build their synthetic training set. Regarding the FR model, in particular Sub-Task 2.3 in which they achieved their highest position among all sub-tasks, they trained IResNet-100 with AdaFace [23] loss (A1) and CosFace [44] loss (A2) with mask and occlusion augmentation on CASIA-WebFace and the refined synthetic data. They used an ensemble of A1, A2, and a model trained with only synthetic data. Furthermore, data augmentation was employed to enhance all features.
Code: https://github.com/liuhao-lh/FRCSyn-Challenge

Idiap-SynthDistill (Sub-Tasks 1.2 and 2.2): The proposed method was based on SynthDistill [38], which is an end-to-end approach, generating synthetic images and training the FR model in the same training loop. Instead of using the pre-trained model in a separate step, they directly used it in the training loop for supervision, while a new student FR model was trained fully using synthetic data generated from a StyleGAN model. For generating synthetic images, they trained StyleGAN2 [22] with the CASIA-WebFace database [48] and then dynamically generated synthetic images during training based on the training loss. For the dynamic image generation, they used the training loss to find the most difficult synthetic image in each batch, and then they generated a new batch of synthetic images by re-sampling the most difficult samples. Regarding the FR model, they used a model with the IResNet-101 [14] architecture and trained it with synthetic data using SynthDistill. They used the Adam optimizer with an initial learning rate of 0.001 and trained their student model with the same loss function as in [38]. For thresholding, a subset of DCFace [24] was used to determine the optimal threshold for maximizing verification accuracy, using a 10-fold cross-validation approach based on a random selection of identities and comparison pairs.
Code: https://gitlab.idiap.ch/bob/bob.paper.ijcb2023_synthdistill

INESC-IGD (All sub-tasks): In all sub-tasks they trained a ResNet-100 with ElasticCosFac-Plus loss [5] using the settings presented in [5]. For the training dataset, DCFace [24], IDiff-Face Uniform, and IDiff-Face Two-stage [6] datasets were merged and their images were labeled with ethnicity labels using a similar approach to [35]. For Sub-Tasks 1.1 and 2.1, they created a synthetic training dataset containing $500K$ face images by sampling $7K$ balanced identities, in terms of ethnicity labels. For Sub-Tasks 1.2 and 2.2, they created a synthetic training dataset containing $2.1M$ face images by sampling $50K$ identities from the training datasets. For Sub-Tasks 1.3 and 2.3, two instances of ResNet-100 were trained on CASIA-WebFace and a subset of synthetic datasets ($400K$ images of $9K$ identities), respectively. The synthetic datasets were sampled from DCFace and IDiff-Face. During the testing phase of Sub-Tasks 1.3 and 2.3, feature embeddings were obtained from trained models and the weighted sum of 0.5 score-level fusion was utilized. During the FR training of all sub-tasks, the training datasets were augmented using the RandAug utilized in IDiff-Face and occluded augmentation [34] with probabilities of 0.4.

UNICA-IGD-LSI (All sub-tasks): They used the DCFace [24] synthetic dataset as it led to remarkable performance gains under well-known evaluation benchmarks for face verification, while combined with real data [2]. They trained a ResNet-100 [18] network using CosFace loss [44] with a margin penalty of 0.35 and a scale term of 64. The similarity mean difference between real-only and synthetic-only samples was scaled and added to the loss value. They trained the model for 40 epochs with a batch size of 512 and an initial learning rate of 0.1, which was divided by 10 after 10, 22, 30, and 40 epochs. During the training phase, the synthetic samples were augmented using RandAugment with 4 operations and a magnitude of 16, following [2, 7]. For Sub-Tasks 1.3 and 2.3, the chosen synthetic dataset was combined with CASIA-Webface [48], obtaining a total of $1M$ images from $20,572$ identities.
Code: https://github.com/atzoriandrea/FRCSyn2

SRCN_AIVL (Sub-Task 1.1): They selected $400K$ samples from the DCFace [24] database and labeled the ethnicity of each subject, as they considered that the racial distribution gap may lead to bad performance in testing. Based

on this insight, they trained IDiff-Face [6] with CASIA-WebFace [48] database generating $100K$ synthetic face images of specific races. Regarding the FR system, they used two custom ResNet-101 [18] trained with AdaFace loss [23] function. The models were trained for 60 epochs with an initial learning rate of 0.1, which was adjusted at predefined milestones. Their training data underwent further preprocessing, including padding crop augmentation, low-resolution augmentation, photometric augmentation, random grayscale, and normalization. For the inference, data preprocessing involved an MTCNN [50] and resizing all data. After cropping and alignment, they fed the image and the flipped image into the two models. After obtaining the two feature embeddings, they combined them and performed the similarity calculation with these embeddings.
Code: https://github.com/Value-Jack/2nd-Edition-FRCSyn

CBSR-Samsung (Sub-Tasks 1.3 and 2.3): They first trained a FR model using CASIA-WebFace [48]. Then, they used it to de-overlap DCFace [24] from CASIA, as DCFace was trained using that real database. For the synthetic dataset, they compared the performance of models trained with three synthetic datasets, including GANDiff-Face [27], DCFace, and IDiffFace [6], and finally selected DCFace as the only synthetic training set. They created a validation dataset including three subsets for three different testing scenarios: *i)* random sample pairs from DCFace; *ii)* randomly positioned vertical bar masks to the images to simulate the self-occlusion due to pose; and *iii)* add a mask and sunglasses to images by detecting the landmarks [45]. All validation subsets consist of $6K$ positive pairs and $6K$ negative pairs. Finally, they concatenated these subsets as the validation set. Subsequently, they conducted an intra-class clustering for all datasets using DBSCAN (0.3 threshold) and removed the samples that were separated from the class center. They merged the refined datasets and trained IResNet-100 [14] with AdaFace loss [23]. In addition, they adopted two augmentation strategies, *i.e.,* photometric augmentation and rescaling. After that, they trained two FR models using occlusion augmentation with 10% and 30% probability, respectively. Finally, they submitted the average similarity score of the two models.

BOVIFOCR-UFPR (Sub-Tasks 1.2 and 2.1): They chose DCFace [24] as the synthetic dataset and ResNet-100 [18] as the backbone, trained with the ArcFace [12] loss function. The images used for training were augmented using a Random Flip with a probability of 0.5. They also applied random erasing and RandAugment as additional augmentations. The model was trained using the Insightface library for 20 epochs within a batch size of 128, running for approximately $78K$ iterations.
Code: https://github.com/PedroBVidal/insightface

Sub-Task 1.1 (Bias Mitigation): Synthetic Data (Constrained)

Pos.	Team	TO [%]	AVG [%]	SD [%]	GAP [%]
1	ID R&D	96.73	97.55	0.82	-5.31
2	ADMIS	94.30	95.10	0.80	1.47
3	SRCN_AIVL	94.06	95.12	1.07	-0.54
4	OPDAI	93.75	94.92	1.17	1.02
5	CTAI	93.21	94.74	1.53	-0.63
6	K-IBS-DS	92.91	94.11	1.20	1.58

Sub-Task 1.2 (Bias Mitigation): Synthetic Data (Unconstrained)

Pos.	Team	TO [%]	AVG [%]	SD [%]	GAP [%]
1	ID R&D	96.73	97.55	0.82	-5.31
2	ADMIS	95.72	96.50	0.78	-0.56
3	OPDAI	94.12	95.22	1.11	0.71
4	INESC-IGD	94.05	95.22	1.17	1.04
5	K-IBS-DS	93.72	94.88	1.16	0.77
6	CTAI	93.21	94.74	1.53	-0.63

Sub-Task 1.3 (Bias Mitigation): Synthetic + Real Data (Constrained)

Pos.	Team	TO [%]	AVG [%]	SD [%]	GAP [%]
1	ADMIS	96.50	97.25	0.75	-1.33
2	K-IBS-DS	96.17	96.92	0.75	-1.37
3	UNICA-IGD-LSI	96.00	96.70	0.70	-5.33
4	OPDAI	95.96	96.80	0.84	-0.03
5	INESC-IGD	95.65	96.33	0.67	-0.12
6	CBSR-Samsung	95.57	96.54	0.97	-24.43

Sub-Task 2.1 (Overall Improvement): Synthetic Data (Constrained)

Pos.	Team	AVG [%]	GAP [%]
1	OPDAI	91.93	3.09
2	ID R&D	91.86	2.99
3	ADMIS	91.19	2.78
4	K-IBS-DS	91.05	2.60
5	CTAI	90.59	-1.94
6	BOVIFOCR-UFPR	89.97	3.71

Sub-Task 2.2 (Overall Improvement): Synthetic Data (Unconstrained)

Pos.	Team	AVG [%]	GAP [%]
1	Idiap-SynthDistill	93.50	-0.05
2	ADMIS	92.92	0.21
3	OPDAI	92.04	3.00
4	ID R&D	91.86	2.99
5	K-IBS-DS	91.61	1.96
6	CTAI	90.59	-1.94

Sub-Task 2.3 (Overall Improvement): Synthetic + Real Data (Constrained)

Pos.	Team	AVG [%]	GAP [%]
1	K-IBS-DS	95.42	-2.15
2	OPDAI	95.23	-0.52
3	CTAI	94.56	-6.01
4	CBSR-Samsung	94.20	-4.40
5	ADMIS	94.15	-1.10
6	ID R&D	94.05	0.07

Table 3. Ranking for the six sub-tasks, according to the metrics described in Section 3.3. TO = Trade-Off, AVG = Average accuracy, SD = Standard Deviation of accuracy, GAP = Gap to Real.

5. FRCSyn Challenge: Results

Table 3 presents the rankings for the different sub-tasks considered in the 2$^{\text{nd}}$ edition of the FRCSyn Challenge. In general, the rankings for Sub-Tasks 1.1, 1.2, and 1.3 (bias mitigation), corresponding to the descending order of TO, closely align with the ascending order of SD (*i.e.,* from less to more biased FR systems). Notably, the winner of Sub-Tasks 1.1 and 1.2, ID R&D (96.73% TO), exhibits a considerable negative GAP value (-5.31%), indicating higher accuracy when training the FR system with synthetic data compared to real data (*i.e.,* CASIA-WebFace [48]). Furthermore, when the limitation in the number of synthetic images is removed (*i.e.,* Sub-Task 1.2), the TO value of most FR systems increases, obtaining a performance improvement and fairness simultaneously. For example, for the ADMIS team (top-2) the TO value increases to 95.72% (*i.e.,* 1.42% TO general improvement from Sub-Tasks 1.1 to 1.2), with a GAP value of -0.56%. These results highlight the advantages of synthetic data, including the potential for generating an infinite number of face images to reduce bias in current FR technology. Finally, for completeness, we analyze in Sub-Task 1.3 the case of adding real and synthetic data to the FR training process. In general, we can observe better TO values, in addition to negative GAP values for all the top-6 teams, *e.g.,* ADMIS (96.50% TO, -1.33 GAP), K-IBS-DS (96.17% TO, -1.37% GAP), and UNICA-IGD-LSI (96.00% TO, -5.33% GAP). These results prove that the combination of synthetic and real data achieves higher FR performance compared to training only with real data. In addition, it is also interesting to compare the best results achieved in Sub-Task 1.2, *i.e.,* unconstrained synthetic data, and Sub-Task 1.3, *i.e.,* constrained synthetic + real data. The ID R&D team achieves 96.73% TO in Sub-Task 1.2 whereas ADMIS achieves 96.50% TO in Sub-Task 1.3, proving that it is possible to obtain better results using only unlimited synthetic data than including real data.

For Task 2, the average accuracy across databases in the different sub-tasks is lower than the accuracy achieved for BUPT-BalancedFace [46] in Task 1, emphasizing the additional challenges introduced by AgeDB [33], CFP-FP [37], and ROF [15] real databases considered for evaluation. Also, although good results are achieved in Sub-Task 2.1 when training only with synthetic data (*e.g.,* 91.93% AVG for OPDAI), the positive GAP values provided by most of the top-6 teams are the greatest from all the sub-tasks, indicating that synthetic data alone currently struggles to completely replace real data for training FR systems. Nevertheless, in Sub-Task 2.2 in which there are no restrictions in the number of synthetic images to use, the Idiap-SynthDistill team (top-1) achieves much better results (93.50% AVG) with a GAP value of -0.05, proving that unlimited synthetic data by itself can even outperform limited real data. Finally, in Sub-Task 2.3, most of the teams report better AVG and higher negative GAP values (*e.g.,* 95.42% AVG and -2.15% GAP for the K-IBS-DS team, top-1), which suggests that synthetic data combined with real data can mitigate existing limitations within FR technology.

Finally, analyzing the contributions of all eleven top teams, a notable trend emerges, showing the prevalence of well-established methodologies. ResNet [18] or IResNet [14] backbones were chosen by all the teams for their wide adoption in state-of-the-art FR approaches. The AdaFace [23] and ArcFace [12] loss functions were widely used, featuring in the approaches of most of the teams, except for ID R&D which used the recent UniFace [52] or UNICA which used CosFace [44]. Notably, all the teams used DCFace [24] alone or combined with other databases like GANDiffFace [27], DigiFace [3], or IDiffFace [6], considering also interesting approaches based on synthetic data cleaning and selection for some teams such as CTAI and CBSR-Samsung. ID R&D and Idiap-SynthDistill were the only teams that used different approaches to generate the synthetic data. In particular, an Hourglass Diffusion Transformer [8] and StyleNAT [43] by the ID R&D team, and dynamic image generation using StyleGAN2 [38] by the Idiap-SynthDistill team.

6. Conclusion

The 2nd edition of the FRCSyn Challenge has presented a comprehensive exploration of the applications of synthetic data in FR, effectively addressing existing limitations in the field. In this 2nd edition, two additional sub-tasks have been introduced, showing that impressive results can be achieved using unlimited synthetic data, even outperforming in some cases the scenario of training with real data. With an increased number of participants in this edition, we have witnessed a considerable performance improvement in all subtasks in comparison to the 1st edition [30, 31]. This has been possible thanks to the proposal of novel methods to generate and select better synthetic data, as well as FR models and loss functions. These approaches can be compared across a variety of sub-tasks, with many being reproducible thanks to the materials made available by the participating teams. Future works will be oriented to a more detailed analysis of the results and comparison with recent challenges in the topic, such as SDFR [39]. We also plan to transform the CodaLab platform into an ongoing challenge, similar to what we did in the 1st edition [31].

Acknowledgements

This study has received funding from the European Union's Horizon 2020 TReSPAsS-ETN (No 860813) and is supported by INTER-ACTION (PID2021-126521OB-I00 MICINN/FEDER), Cátedra ENIA UAM-VERIDAS en IA Responsable (NextGenerationEU PRTR TSI-100927-2023-2) and R&D Agreement DGGC/ UAM/FUAM for Biometrics and Cybersecurity. It is also supported by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE. K-IBS-DS was supported by the Institute for Basic Science, Republic of Korea (IBS-R029-C2). UNICA-IGD-LSI was supported by the ARIS program P2-0250B.

References

[1] Xiang An, Xuhan Zhu, Yuan Gao, Yang Xiao, Yongle Zhao, Ziyong Feng, Lan Wu, Bin Qin, Ming Zhang, Debing Zhang, and Ying Fu. Partial FC: Training 10 Million Identities on a Single Machine. In *Proc. IEEE/CVF International Conference on Computer Vision Workshops*, 2021. 5

[2] Andrea Atzori, Fadi Boutros, Naser Damer, Gianni Fenu, and Mirko Marras. If It's Not Enough, Make It So: Reducing Authentic Data Demand in Face Recognition through Synthetic Faces. In *Proc. 18th International Conference on Automatic Face and Gesture Recognition*, 2024. 6

[3] Gwangbin Bae, Martin de La Gorce, Tadas Baltrušaitis, Charlie Hewitt, Dong Chen, Julien Valentin, Roberto Cipolla, and Jingjing Shen. DigiFace-1M: 1 Million Digital Face Images for Face Recognition. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023. 2, 3, 5, 6, 8

[4] Carmen Bisogni, Aniello Castiglione, Sanoar Hossain, Fabio Narducci, and Saiyed Umer. Impact of Deep Learning Approaches on Facial Expression Recognition in Healthcare Industries. *IEEE Transactions on Industrial Informatics*, 18(8): 5619–5627, 2022. 1

[5] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. ElasticFace: Elastic Margin Loss for Deep Face Recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022. 5, 6

[6] Fadi Boutros, Jonas Henry Grebe, Arjan Kuijper, and Naser Damer. IDiff-Face: Synthetic-based Face Recognition through Fizzy Identity-Conditioned Diffusion Model. In *Proc. IEEE/CVF International Conference on Computer Vision*, 2023. 2, 3, 5, 6, 7, 8

[7] Fadi Boutros, Marcel Klemt, Meiling Fang, Arjan Kuijper, and Naser Damer. Unsupervised Face Recognition using Unlabeled Synthetic Data. In *Proc. 17th International Conference on Automatic Face and Gesture Recognition*, 2023. 6

[8] Katherine Crowson, Stefan Andreas Baumann, Alex Birch, Tanishq Mathew Abraham, Daniel Z Kaplan, and Enrico Shippole. Scalable High-Resolution Pixel-Space Image Synthesis with Hourglass Diffusion Transformers. *arXiv preprint arXiv:2401.11605*, 2024. 5, 8

[9] Colton R. Crum, Patrick Tinsley, Aidan Boyd, Jacob Piland, Christopher Sweet, Timothy Kelley, Kevin Bowyer, and Adam Czajka. Explain To Me: Salience-Based Explainability for Synthetic Face Detection Models. *IEEE Transactions on Artificial Intelligence*, pages 1–12, 2023. 2

[10] Roberto Daza, Luis F Gomez, Aythami Morales, Julian Fierrez, Ruben Tolosana, Ruth Cobos, and Javier Ortega-Garcia. MATT: Multimodal Attention Level Estimation for e-learning Platforms. In *Proc. AAAI Workshop on Artificial Intelligence for Education*, 2023. 1

[11] Ivan Deandres-Tame, Ruben Tolosana, Ruben Vera-Rodriguez, Aythami Morales, Julian Fierrez, and Javier Ortega-Garcia. How Good Is ChatGPT at Face Biometrics? A First Look Into Recognition, Soft Biometrics, and Explainability. *IEEE Access*, 12:34390–34401, 2024. 2

[12] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 1, 5, 7, 8

[13] Hang Du, Hailin Shi, Dan Zeng, Xiao-Ping Zhang, and Tao Mei. The Elements of End-to-end Deep Face Recognition: A Survey of Recent Advances. *ACM Comput. Surv.*, 54(10s), 2022. 1

[14] Ionut Cosmin Duta, Li Liu, Fan Zhu, and Ling Shao. Improved Residual Networks for Image and Video Recognition. In *Proc. 25th International Conference on Pattern Recognition*, 2021. 5, 6, 7, 8

[15] Mustafa Ekrem Erakιn, Uğur Demir, and Hazιm Kemal Ekenel. On Recognizing Occluded Faces in the Wild. In *Proc. International Conference of the Biometrics Special Interest Group*, 2021. 3, 4, 8

[16] Nikita Garaev, Evgeny Smirnov, Vasiliy Galyuk, and Evgeny Lukyanets. FaceMix: Transferring Local Regions for Data Augmentation in Face Recognition. In *Proc. Neural Information Processing*, 2023. 5

[17] Luis F. Gomez, Aythami Morales, Juan R. Orozco-Arroyave, Roberto Daza, and Julian Fierrez. Improving Parkinson Detection Using Dynamic Features From Evoked Expressions in Video. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 1

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016. 2, 6, 7, 8

[19] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-Excitation Networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 6

[20] Manuel Kansy, Anton Raël, Graziana Mignone, Jacek Naruniec, Christopher Schroers, Markus Gross, and Romann M. Weber. Controllable Inversion of Black-Box Face Recognition Models via Diffusion. In *Proc. IEEE/CVF International Conference on Computer Vision Workshops*, 2023. 2, 3

[21] Tero Karras, Samuli Laine, and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. In *Proc. IEEE/CVF conference on Computer Vision and Pattern Recognition*, 2019. 5

[22] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and Improving the Image Quality of StyleGAN. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 6

[23] Minchul Kim, Anil K. Jain, and Xiaoming Liu. AdaFace: Quality Adaptive Margin for Face Recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 1, 2, 5, 6, 7, 8

[24] Minchul Kim, Feng Liu, Anil Jain, and Xiaoming Liu. DCFace: Synthetic Face Generation with Dual Condition Diffusion Model. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2, 3, 4, 5, 6, 7, 8

[25] Zhen Li, Mingdeng Cao, Xintao Wang, Zhongang Qi, Ming-Ming Cheng, and Ying Shan. PhotoMaker: Customizing Realistic Human Photos via Stacked ID Embedding. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 5

[26] Cheng Yaw Low, Jacky Chen Long Chai, Jaewoo Park, Kyeongjin Ann, and Meeyoung Cha. SlackedFace: Learning a Slacked Margin for Low-Resolution Face Recognition. In *Proc. 34th British Machine Vision Conference*, 2023. 5

[27] Pietro Melzi, Christian Rathgeb, Ruben Tolosana, Ruben Vera-Rodriguez, Dominik Lawatsch, Florian Domin, and Maxim Schaubert. GANDiffFace: Controllable Generation of Synthetic Datasets for Face Recognition with Realistic Variations. In *Proc. IEEE/CVF International Conference on Computer Vision Workshops*, 2023. 2, 3, 6, 7, 8

[28] Pietro Melzi, Christian Rathgeb, Ruben Tolosana, Ruben Vera-Rodriguez, Aythami Morales, Dominik Lawatsch, Florian Domin, and Maxim Schaubert. Synthetic Data for the Mitigation of Demographic Biases in Face Recognition. In *Proc. IEEE International Joint Conference on Biometrics*, 2023. 2

[29] Pietro Melzi, Hatef Otroshi Shahreza, Christian Rathgeb, Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Sébastien Marcel, and Christoph Busch. Multi-IVE: Privacy Enhancement of Multiple Soft-Biometrics in Face Embeddings. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, 2023. 2

[30] Pietro Melzi, Ruben Tolosana, Ruben Vera-Rodriguez, Minchul Kim, Christian Rathgeb, Xiaoming Liu, Ivan DeAndres-Tame, Aythami Morales, Julian Fierrez, Javier Ortega-Garcia, Weisong Zhao, Xiangyu Zhu, Zheyu Yan, Xiao-Yu Zhang, Jinlin Wu, Zhen Lei, Suvidha Tripathi, Mahak Kothari, Md Haider Zama, Debayan Deb, Bernardo Biesseck, Pedro Vidal, Roger Granada, Guilherme Fickel, Gustavo Führ, David Menotti, Alexander Unnervik, Anjith George, Christophe Ecabert, Hatef Otroshi Shahreza, Parsa Rahimi, Sébastien Marcel, Ioannis Sarridis, Christos Koutlis, Georgia Baltsou, Symeon Papadopoulos, Christos Diou, Nicolò Di Domenico, Guido Borghi, Lorenzo Pellegrini, Enrique Mas-Candela, Ángela Sánchez-Pérez, Andrea Atzori, Fadi Boutros, Naser Damer, Gianni Fenu, and Mirko Marras. FRCSyn Challenge at WACV 2024: Face Recognition Challenge in the Era of Synthetic Data. In *Proc. of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, 2024. 2, 3, 8

[31] Pietro Melzi, Ruben Tolosana, Ruben Vera-Rodriguez, Minchul Kim, Christian Rathgeb, Xiaoming Liu, Ivan DeAndres-Tame, Aythami Morales, Julian Fierrez, Javier Ortega-Garcia, Weisong Zhao, Xiangyu Zhu, Zheyu Yan, Xiao-Yu Zhang, Jinlin Wu, Zhen Lei, Suvidha Tripathi, Mahak Kothari, Md Haider Zama, Debayan Deb, Bernardo Biesseck, Pedro Vidal, Roger Granada, Guilherme Fickel, Gustavo Führ, David Menotti, Alexander Unnervik, Anjith George, Christophe Ecabert, Hatef Otroshi Shahreza, Parsa Rahimi, Sébastien Marcel, Ioannis Sarridis, Christos Koutlis, Georgia Baltsou, Symeon Papadopoulos, Christos Diou, Nicolò Di Domenico, Guido Borghi, Lorenzo Pellegrini, Enrique Mas-Candela, Ángela Sánchez-Pérez, Andrea Atzori, Fadi Boutros, Naser Damer, Gianni Fenu, and Mirko Marras. FRCSyn-onGoing: Benchmarking and Comprehensive Evaluation of Real and Synthetic Data to Improve Face Recognition Systems. *Information Fusion*, 107:102322, 2024. 2, 3, 8

[32] Aythami Morales, Julian Fierrez, Ruben Vera-Rodriguez, and Ruben Tolosana. SensitiveNets: Learning Agnostic Representations with Application to Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6): 2158–2164, 2021. 2

[33] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. AgeDB: The First Manually Collected, In-The-Wild Age Database. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2017. 3, 4, 8

[34] Pedro C. Neto, Fadi Boutros, Joao Ribeiro Pinto, Naser Damer, Ana F. Sequeira, Jaime S. Cardoso, Messaoud Bengherabi, Abderaouf Bousnat, Sana Boucheta, Nesrine Hebbadj, Mustafa Ekrem Erakın, Uğur Demir, Hazım Kemal Ekenel, Pedro Beber De Queiroz Vidal, and David Menotti. OCFR 2022: Competition on Occluded Face Recognition from Synthetically Generated Structure-Aware Occlusions. In *Proc. IEEE International Joint Conference on Biometrics*, 2022. 6

[35] Pedro C. Neto, Eduarda Caldeira, Jaime S. Cardoso, and Ana F. Sequeira. Compressed Models Decompress Race Biases: What Quantized Models Forget for Fair Face Recognition. In *Proc. International Conference of the Biometrics Special Interest Group*, 2023. 6

[36] Haibo Qiu, Baosheng Yu, Dihong Gong, Zhifeng Li, Wei Liu, and Dacheng Tao. SynFace: Face Recognition with Synthetic Data. In *Proc. IEEE/CVF International Conference on Computer Vision*, 2021. 2, 3

[37] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to Profile Face Verification in the Wild. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, 2016. 3, 4, 8

[38] Hatef Otroshi Shahreza, Anjith George, and Sébastien Marcel. SynthDistill: Face Recognition with Knowledge Distillation from Synthetic Data. In *Proc. IEEE International Joint Conference on Biometrics*, 2023. 6, 8

[39] Hatef Otroshi Shahreza, Christophe Ecabert, Anjith George, Alexander Unnervik, Sébastien Marcel, Nicolò Di Domenico, Guido Borghi, Davide Maltoni, Fadi Boutros, Julia Vogel, Naser Damer, Ángela Sánchez-Pérez, EnriqueMas-Candela, Jorge Calvo-Zaragoza, Bernardo Biesseck, Pedro Vidal, Roger Granada, David Menotti, Ivan DeAndres-Tame, Simone Maurizio La Cava, Sara Concas, Pietro Melzi, Ruben Tolosana, Ruben Vera-Rodriguez, Gianpaolo Perelli, Giulia Orrù, Gian Luca Marcialis, and Julian Fierrez. SDFR: Synthetic Data for Face Recognition Competition. In *Proc. 18th IEEE International Conference on Automatic Face and Gesture Recognition*, 2024. 8

[40] Evgeny Smirnov, Nikita Garaev, Vasiliy Galyuk, and Evgeny Lukyanets. Prototype Memory for Large-Scale Face Representation Learning. *IEEE Access*, 10:12031–12046, 2022. 5

[41] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. In *Proc. International Conference on Learning Representations*, 2021. 5

[42] Philipp Terhörst, Jan Niklas Kolf, Marco Huber, Florian Kirchbuchner, Naser Damer, Aythami Morales Moreno, Julian Fierrez, and Arjan Kuijper. A Comprehensive Study on Face Recognition Biases Beyond Demographics. *IEEE Transactions on Technology and Society*, 3(1):16–30, 2021. 2

[43] Steven Walton, Ali Hassani, Xingqian Xu, Zhangyang Wang, and Humphrey Shi. StyleNAT: Giving Each Head a New Perspective. *arXiv preprint arXiv:2211.05770*, 2022. 5, 8

[44] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 6, 8

[45] Jun Wang, Yinglu Liu, Yibo Hu, Hailin Shi, and Tao Mei. FaceX-Zoo: A PyTorch Toolbox for Face Recognition. In *Proc. 29th ACM International Conference on Multimedia*, 2021. 7

[46] Mei Wang and Weihong Deng. Mitigating Bias in Face Recognition Using Skewness-Aware Reinforcement Learning. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 3, 4, 8

[47] Mei Wang and Weihong Deng. Deep Face Recognition: A Survey. *Neurocomputing*, 429:215–244, 2021. 1

[48] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning Face Representation from Scratch. *arXiv preprint arXiv:1411.7923*, 2014. 3, 4, 5, 6, 7, 8

[49] Cheng Zhang, Xuanbai Chen, Siqi Chai, Chen Henry Wu, Dmitry Lagun, Thabo Beeler, and Fernando De la Torre. ITI-GEN: Inclusive Text-to-Image Generation. In *Proc. IEEE/CVF International Conference on Computer Vision*, 2023. 2, 3

[50] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016. 7

[51] Yaoyao Zhong, Weihong Deng, Jiani Hu, Dongyue Zhao, Xian Li, and Dongchao Wen. SFace: Sigmoid-Constrained Hypersphere Loss for Robust Face Recognition. *IEEE Transactions on Image Processing*, 30:2587–2598, 2021. 2, 3

[52] Jiancan Zhou, Xi Jia, Qiufu Li, Linlin Shen, and Jinming Duan. UniFace: Unified Cross-Entropy Loss for Deep Face Recognition. In *Proc. IEEE/CVF International Conference on Computer Vision*, 2023. 5, 8

[53] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, and Jie Zhou. WebFace260M: A Benchmark Unveiling the Power of Million-Scale Deep Face Recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 5