

# Second FRCSyn-onGoing: Winning Solutions and Post-Challenge Analysis to Improve Face Recognition with Synthetic Data

Ivan DeAndres-Tame, Ruben Tolosana, Pietro Melzi, Ruben Vera-Rodriguez, Minchul Kim, Christian Rathgeb, Xiaoming Liu, Luis F. Gomez, Aythami Morales, Julian Fierrez, Javier Ortega-Garcia, Zhizhou Zhong, Yuge Huang, Yuxi Mi, Shouhong Ding, Shuigeng Zhou, Shuai He, Lingzhi Fu, Heng Cong, Rongyu Zhang, Zhihong Xiao, Evgeny Smirnov, Anton Pimenov, Aleksei Grigorev, Denis Timoshenko, Kaleb Mesfin Asfaw, Cheng Yaw Low, Hao Liu, Chuyi Wang, Qing Zuo, Zhixiang He, Hatef Otroshti Shahreza, Anjith George, Alexander Unnervik, Parsa Rahimi, Sébastien Marcel, Pedro C. Neto, Marco Huber, Jan Niklas Kolf, Naser Damer, Fadi Boutros, Jaime S. Cardoso, Ana F. Sequeira, Andrea Atzori, Gianni Fenu, Mirko Marras, Vitomir Štruc, Jiang Yu, Zhangjie Li, Jichun Li, Weisong Zhao, Zhen Lei, Xiangyu Zhu, Xiao-Yu Zhang, Bernardo Biesseck, Pedro Vidal, Luiz Coelho, Roger Granada, David Menotti

*Abstract*—Synthetic data is gaining increasing popularity for face recognition technologies, mainly due to the privacy concerns and challenges associated with obtaining real data, including diverse scenarios, quality, and demographic groups, among others. It also offers some advantages over real data, such as the large amount of data that can be generated or the ability to customize it to adapt to specific problem-solving needs. To effectively use such data, face recognition models should also be specifically designed to exploit synthetic data to its fullest potential. In order to promote the proposal of novel Generative AI methods and synthetic data, and investigate the application of synthetic data to better train face recognition systems, we introduce the 2<sup>nd</sup> FRCSyn-onGoing challenge, based on the 2<sup>nd</sup> Face Recognition Challenge in the Era of Synthetic Data (FRCSyn), originally launched at CVPR 2024. This is an ongoing challenge that provides researchers with an accessible platform to benchmark *i)* the proposal of novel Gen-

erative AI methods and synthetic data, and *ii)* novel face recognition systems that are specifically proposed to take advantage of synthetic data. We focus on exploring the use of synthetic data both individually and in combination with real data to solve current challenges in face recognition such as demographic bias, domain adaptation, and performance constraints in demanding situations, such as age disparities between training and testing, changes in the pose, or occlusions. Very interesting findings are obtained in this second edition, including a direct comparison with the first one, in which synthetic databases were restricted to DCFace and GANDiffFace.

*Index Terms*—FRCSyn, Face Recognition, Synthetic Data, Generative AI, Demographic Bias, Benchmark, Privacy

## I. INTRODUCTION

Face biometrics is a very popular area within Computer Vision and Pattern Recognition, finding applications across various domains such as person recognition [1], [2], healthcare [3], [4], and e-learning [5], among others. In recent years, with the fast development of deep learning, significant

2<sup>nd</sup> FRCSyn Challenge Organizers: Ruben Tolosana, Ivan DeAndres-Tame, Pietro Melzi, Ruben Vera-Rodriguez, Minchul Kim, Christian Rathgeb, Xiaoming Liu, Aythami Morales, Julian Fierrez and Javier Ortega-Garcia. Information related to each author is included at the end of the article. Link to 2<sup>nd</sup> FRCSyn Challenge: <https://frcsyn.github.io/CVPR2024.html>

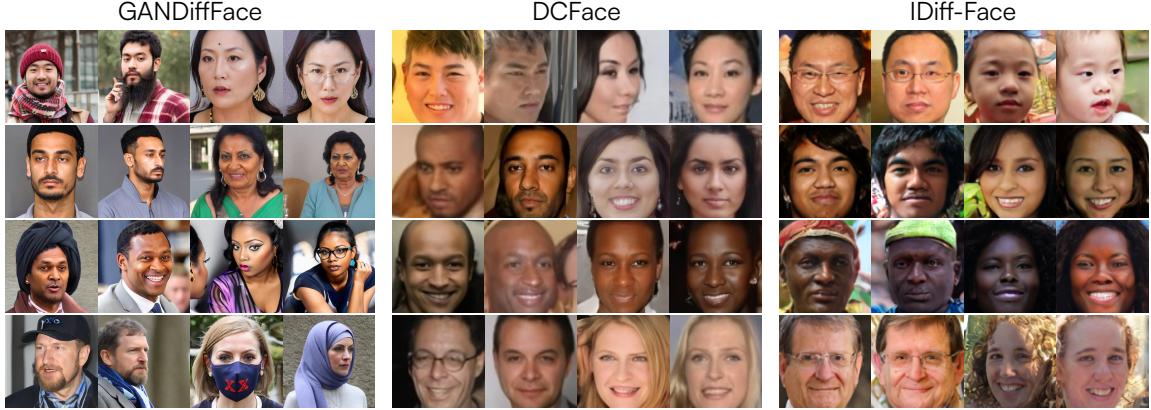


Fig. 1. Examples of synthetic identities and variations for different demographic groups using GANDiffFace [21] (left), DCFace [22] (middle) and IDiff-Face [23] (right).

advances have been made in areas like face recognition (FR) [6], [7], surpassing previous benchmarks. However, FR technology still faces challenges in several research directions, including explainability [8]–[10], demographic bias [11], [12], privacy [13]–[15], and robustness against adverse conditions [7], such as aging [16], pose variations [17], [18], illumination changes [19], and occlusions [20].

Synthetic data has gained popularity as a good solution to mitigate some of these drawbacks [24], [25], allowing the generation of *i*) a large number of facial images from different non-existent identities, and *ii*) variability in terms of demographic attributes and scenario conditions. In this context, we refer to demographics as whole societies or smaller groups defined by criteria such as ethnicity, sex, or age [11], [26]. Several Generative AI approaches have been presented in the last couple of years for the synthesis of face images, considering state-of-the-art deep learning methods. The most popular approaches to generate synthetic facial images are Generative Adversarial Networks (GANs) [27], Diffusion Models [28] or the combination of both [21]. However, there are other less common approaches that do not rely on these models and still achieve a high level of realism in image generation [29], [30]. Some examples of synthetic face images generated using some of these methods are shown in Figure 1.

Beyond the generation of synthetic faces, another

critical aspect lies in understanding the potential applications and benefits of synthetic data in enhancing FR technology. Recent studies have highlighted a performance gap between FR systems trained only with synthetic data and those trained on real data [22], [31]. Nevertheless, the results achieved in the 1<sup>st</sup> edition of the Face Recognition Challenge in the Era of Synthetic Data (FRCSyn) [32], [33], emphasize the relevance of synthetic data, either alone or merged with real data, in mitigating challenges in FR, such as demographic bias [32], [33]. Notably, in the 1<sup>st</sup> FRCSyn-onGoing, only synthetic data from DCFace [22] and GANDiffFace [21] methods were allowed for training FR systems. Additionally, together with novel generative methods, improving FR technology involves refining the design and training processes to address domain gaps between real and synthetic data in certain scenarios. For instance, observations from the 1<sup>st</sup> FRCSyn-onGoing revealed that most teams considered similar deep learning architectures (e.g., ResNet-100 [34]) and loss functions (e.g., AdaFace [7]), commonly used in FR systems trained with real data. Moreover, the use of synthetic facial data is not limited only to FR. With the emerging popularity of foundational models [35], [36], synthetic facial data can also be leveraged to provide these large models with a general understanding of what a human face looks like, serving as pretraining for many other

tasks [37]–[39]. Additionally, the fact that these data can be labeled as they are generated allows the creation of novel datasets for various tasks such as attribute detection, facial expression recognition, and more. Although in the field of synthetic faces there are still not many public datasets in these lines of research, several studies have already highlighted the benefits of synthetic data for biometric tasks such as facial expression recognition [40], signature verification [41], action recognition [42], or pose estimation [43].

In order to promote the development of novel face generative methods and the creation of synthetic face databases, as well as investigate the application of synthetic data to better train FR systems, we have organized the 2<sup>nd</sup> FRCSyn-onGoing Challenge, which is based on the 2<sup>nd</sup> FRCSyn Challenge as part of CVPR 2024<sup>1</sup> [44]. In this 2<sup>nd</sup> edition, we introduce new sub-tasks allowing participants to train FR systems using synthetic data generated with their preferred generative frameworks, offering more flexibility compared to the 1<sup>st</sup> edition [32], [33]. Additionally, new sub-tasks with varied experimental settings are included to explore how FR systems can be trained under both constrained and unconstrained scenarios regarding the amount of synthetic training data. The FRCSyn Challenge aims to address the following research questions:

- 1) What are the limitations of FR technology trained only with synthetic data?
- 2) Can synthetic data help alleviate current limitations in FR technology?

These questions have become increasingly relevant after the discontinuation of popular real FR databases due to privacy concerns<sup>2</sup> and the introduction of new regulatory laws<sup>3</sup>.

The foundation of the present article was established in an earlier publication [44], with the current version notably extending it through: *i*) a more extensive description and analysis of the top synthetic face generation methods and FR systems presented so far in this 2<sup>nd</sup> FRCSyn-onGoing, including key

graphical representations of the proposed systems to improve the understanding of the reader, *ii*) incorporating additional metrics in the evaluation of the proposed FR systems in order to analyze different operational scenarios, *iii*) presenting an in-depth analysis of the performance achieved for various demographic groups and databases used for evaluation, together with novel figures and tables, and *iv*) a direct comparison between the results obtained in this 2<sup>nd</sup> edition and the ones obtained in 1<sup>st</sup> edition [32], [33], highlighting very interesting findings.

The remainder of the article is organized as follows. Section II describes the databases considered at the 2<sup>nd</sup> FRCSyn-onGoing. Section III explains the experimental setup of the challenge, including the different tasks and sub-tasks, the experimental protocol, metrics, and restrictions. In Section IV, we describe the approaches proposed by the top-6 participating teams. Section V presents the best results achieved so far in the different tasks and sub-tasks of 2<sup>nd</sup> FRCSyn-onGoing, emphasizing the key results of the challenge. Finally, in Section VI, we provide some conclusions, highlighting potential future research directions in the field.

## II. SECOND FRCSYN-ONGOING: DATABASES

### A. Synthetic Databases

One of the main novelties of the 2<sup>nd</sup> FRCSyn-onGoing is that there are no restrictions in terms of the generative methods used to create synthetic data. Unlike the 1<sup>st</sup> FRCSyn-onGoing, where only synthetic data created using DCFace [22] and GAN-DiffFace [21] was available, in this 2<sup>nd</sup> edition we allow participants to use any generative framework of their choice to create synthetic data, limiting in some sub-tasks the number of synthetic face images used to train the FR systems (more details in Section III-A). As a reference, after the registration in the challenge, we provide all the participants with a list of possible state-of-the-art generative frameworks. For completeness, we summarize next and in Table I the most popular approaches available at the beginning of the challenge:

<sup>1</sup><https://frcsyn.github.io/CVPR2024.html>

<sup>2</sup><https://exposing.ai/about/news/> (March, 2024)

<sup>3</sup><https://artificialintelligenceact.eu> (March, 2024)

TABLE I

DESCRIPTION OF SOME POSSIBLE GENERATIVE METHODS AND SYNTHETIC DATABASES THAT CAN BE USED BY PARTICIPANTS IN THE 2<sup>ND</sup> FRCSYN-ONGOING. ID = IDENTITIES, IMG = IMAGES

| Database                  | Framework           | # Id | # Img/Id | # Img  |
|---------------------------|---------------------|------|----------|--------|
| DCFace [22]               | Diffusion Model     | 20K  | 50       | 1,200K |
|                           |                     | 40K  | 5        |        |
| GANDiffFace [21]          | GAN Diffusion Model | 10K  | 50       | 500K   |
| IDiff-face Uniform [23]   | Diffusion Model     | 10K  | 50       | 500K   |
| IDiff-face Two-Stage [23] | Diffusion Model     | 10K  | 50       | 500K   |
| DigiFace-1M [29]          | 3D-model            | 10K  | 72       | 1.2M   |
|                           |                     | 100K | 5        |        |
| ID3PM [45]                | Diffusion Model     | -    | -        | -      |
| SFace [46]                | GAN                 | 10K  | 60       | 0.6M   |
| SYNFace [31]              | GAN                 | 10K  | 100      | 1M     |
| ITI-GEN [30]              | CLIP                | -    | -        | -      |

- **DCFace**<sup>4</sup> [22]: This framework is entirely based on Diffusion models, composed of a sampling stage for the generation of synthetic identities  $X_{ID}$ , and a mixing stage for the generation of images  $X_{ID,sty}$  with the same identities  $X_{ID}$  from the sampling stage and the style selected from a “style bank” of images  $X_{sty}$ .
- **GANDiffFace**<sup>5</sup> [21]: This framework combines StyleGAN [47] and a Diffusion Model, *i.e.*, DreamBooth [48], to generate fully synthetic FR databases with desired properties such as human face realism, controllable demographic distributions, and realistic intra-class variations (*e.g.*, changes in pose, expression, and occlusions). Graphical examples are shown in Figure 1.
- **IDiff-Face**<sup>6</sup> [23]: This framework uses a Diffusion Model conditioned on identity context, which allows the model to either generate variations of existing authentic images by us-
- ing authentic embeddings or to generate novel synthetic identities by using synthetic face embeddings. The authors presented two distinct datasets: one by generating identity context in a two-stage process, and the other through a synthetic uniform representation.
- **DigiFace-1M**<sup>7</sup> [29]: This framework can generate large-scale synthetic face images with many unique subjects based on 3D parametric model rendering. It considers the method introduced by Wood *et al.* [49], tackling the ethical and labeling problems associated with the generation of synthetic data.
- **ID3PM** [45]: This framework considers a Diffusion Model to perform an inversion of a FR model generating new images from Gaussian noise with various backgrounds, lighting, poses, and expressions while preserving the identity.
- **SFace**<sup>8</sup> [46]: This framework uses a conditional GAN to synthetically generate face im-

<sup>4</sup><https://github.com/mk-minchul/dcface>

<sup>5</sup><https://github.com/PietroMelzi/GANDiffFace>

<sup>6</sup><https://github.com/fdbtrs/IDiff-Face>

<sup>7</sup><https://github.com/microsoft/DigiFace1M>

<sup>8</sup><https://github.com/fdbtrs/SFace-Privacy-friendly-and-Accurate-Face-Recognition-using-Synthetic-Data>

TABLE II  
DESCRIPTION OF THE REAL DATABASES CONSIDERED FOR EVALUATION IN THE 2<sup>ND</sup> FRCSYN-ONGOING.  
ID = IDENTITIES, IMG = IMAGES

| Database               | Framework | # Id  | # Img/Id | # Img |
|------------------------|-----------|-------|----------|-------|
| CASIA-WebFace [51]     | Real      | 10.5K | 47       | 500K  |
| BUPT-BalancedFace [52] | Real      | 24K   | 45       | 1M    |
| AgeDB [53]             | Real      | 570   | 29       | 17K   |
| CFP-FP [54]            | Real      | 500   | 14       | 7K    |
| ROF [55]               | Real      | 180   | 31       | 6K    |

ages with an adaptive discriminator augmentation to increase the diversity of the training database.

- **SYNFace**<sup>9</sup> [31]: This framework uses DiscoFaceGAN [50] to generate face images with different identities from a Mixup Face Generator.
- **ITI-GEN**<sup>10</sup> [30]: This framework uses CLIP [35] to generate embeddings to translate the visual attribute differences into natural language differences and perform a Text-to-Image generation that is inclusive.

These are just some possible generative frameworks, with the corresponding synthetic databases available, that can be used by participants. But, as indicated before, the purpose of the 2<sup>nd</sup> FRCSyn-onGoing is to promote the proposal of novel generative methods and the creation of better synthetic databases to improve the performance of FR systems. It is important to mention that in the 2<sup>nd</sup> FRCSyn-onGoing, synthetic data is exclusively used in the training stage of FR technology, replicating realistic operational scenarios.

### B. Real Databases

For the training of the FR systems participants are allowed to use only the **CASIA-WebFace** [51] as real data (depending on the sub-task, please see Section III-A). 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 consider the same four real databases

<sup>9</sup><https://github.com/haibo-qiu/SynFace>

<sup>10</sup><https://github.com/humansensinglab/ITI-GEN>

used at the 1<sup>st</sup> FRCSyn Challenge [32], [33], as they consider key challenges in FR such as demographic bias, pose variations, aging, and occlusions. We summarize next and in Table II each of them:

- **BUPT-BalancedFace** [52] is designed to address performance disparities across different ethnic groups. We relabel it according to the FairFace classifier [56], which provides labels for ethnicity (White, Black, Asian, Indian) and gender (Male, Female). We then consider the eight demographic groups obtained from all possible combinations of four ethnic groups and genders. We are aware that these groups do not comprehensively represent the entire spectrum of real world ethnic diversity. Nevertheless, the selection of these categories, while imperfect, is primarily driven by the need to align with the demographic categorizations used in BUPT-BalancedFace to facilitate easier and more consistent evaluation.
- **AgeDB** [53] contains facial images featuring the same subjects at different ages in different environmental contexts.
- **CFP-FP** [54] presents facial images from subjects with great changes in pose, with both frontal and profile images, and different environmental contexts.
- **ROF** [55] consists of occluded faces with both upper face occlusion, due to sunglasses, and lower face occlusion, due to masks.

Finally, it is important to highlight that, as different databases are considered for training and evaluation, we also intend to analyze the generalization ability of the proposed FR systems.

TABLE III

TASKS AND SUB-TASKS FOR THE 2<sup>ND</sup> FRCSYN-ONGOING AND THEIR RESPECTIVE METRICS AND DATABASES.  
TO = TRADE-OFF. GAP = GAP TO REAL. AVG = AVERAGE VERIFICATION ACCURACY. SD = STANDARD DEVIATION. FLOPS = FLOATING POINT OPERATIONS PER SECOND. SYN = ACCURACY PROPOSED MODEL. REAL = ACCURACY BASELINE MODEL. ID = IDENTITY.

|   |
|---|
| <b>Task 1:</b> synthetic data for <b>demographic bias mitigation</b>  |
| <b>Baseline:</b> training with only CASIA-WebFace [51].   |
| <b>Ranking:</b> Trade-Off, see Section III-C for more details.  |
| <b>Metrics:</b> $TO = AVG - SD$<br>$GAP = (REAL - SYN)/SYN$   |
| <b>Sub-Task 1.1: [constrained]</b> training exclusively with <b>synthetic</b> data  |
| <b>Train:</b> maximum 500K face images (e.g., 10K IDs and 50 images per ID).  |
| <b>Eval:</b> BUPT-BalancedFace [52].  |
| <b>Sub-Task 1.2: [unconstrained]</b> training exclusively with <b>synthetic</b> data  |
| <b>Train:</b> no restrictions in terms of the number of face images.  |
| <b>Eval:</b> BUPT-BalancedFace.   |
| <b>Sub-Task 1.3: [constrained]</b> training with <b>real and synthetic</b> data   |
| <b>Train:</b> CASIA-WebFace, and maximum 500K face synthetic images.  |
| <b>Eval:</b> BUPT-BalancedFace.   |
| <b>Task 2:</b> synthetic data for <b>overall performance improvement</b>  |
| <b>Baseline:</b> training with only CASIA-WebFace.  |
| <b>Ranking:</b> average accuracy, see Section III-C for more details.   |
| <b>Metrics:</b> $AVG$<br>$GAP = (REAL - SYN)/SYN$   |
| <b>Sub-Task 2.1: [constrained]</b> training with only <b>synthetic</b> data   |
| <b>Train:</b> maximum 500K face images.   |
| <b>Eval:</b> BUPT-BalancedFace, AgeDB [53], CFP-FP [54], and ROF [55].  |
| <b>Sub-Task 2.2: [unconstrained]</b> training with only <b>synthetic</b> data   |
| <b>Train:</b> no restrictions in terms of the number of face images.  |
| <b>Eval:</b> BUPT-BalancedFace, AgeDB, CFP-FP, and ROF.   |
| <b>Sub-Task 2.3: [constrained]</b> training with <b>real and synthetic</b> data   |
| <b>Train:</b> CASIA-WebFace, and maximum 500K face synthetic images.  |
| <b>Eval:</b> BUPT-BalancedFace, AgeDB, CFP-FP, and ROF.   |
| <b>Restrictions:</b> FLOPS $\leq$ 50 GFLOPS<br>Only the specified databases can be used for training.<br>Generative models cannot be used to generate supplementary data. |

### III. SECOND FRCSYN-ONGOING: SETUP

Due to the success of the 1<sup>st</sup> FRCSyn-onGoing [32], [33], we also decided to run the 2<sup>nd</sup> edition in Codalab<sup>11</sup>, an open-source framework designed for conducting scientific competitions and benchmarks. On this platform, participants can find the competition's requirements and limitations and can submit their scores to automatically obtain *i*) the evaluation metrics of their system, and *ii*) the

<sup>11</sup><https://codalab.lisn.upsaclay.fr/competitions/16970>

position on the challenge leaderboard. Table III provides an overview of the key aspects of the experimental protocol, metrics and restrictions for each sub-task. More detailed explanations can be found in their respective subsections.

#### A. Tasks

Similar to the 1<sup>st</sup> FRCSyn-onGoing [32], [33], in this 2<sup>nd</sup> edition we also explore the application of synthetic data for training 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 consider two different tasks, each comprising three sub-tasks. Each sub-task considers different types (real/synthetic) and amounts of data for training the FR systems. Consequently, the 2<sup>nd</sup> edition comprises 6 different sub-tasks.

**Task 1:** The first task focuses on using synthetic data to mitigate demographic biases within FR systems. To evaluate the performance of these systems, we create sets of mated and non-mated comparisons using subjects from the BUPT-BalancedFace database [52]. We consider the eight demographic groups defined in Section II-B, which result from the combination of four ethnicities (White, Black, Asian, and Indian) and two genders (Male and Female), ensuring a balanced representation across these groups in the comparison lists. For non-mated comparisons, we exclusively pair subjects within the same demographic group, as these hold greater relevance compared to non-mated comparisons involving subjects from different demographic groups.

**Task 2:** The second proposed task focuses on using synthetic data to enhance the overall performance of FR systems under challenging conditions. To assess the effectiveness of the proposed systems, we use 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 challenges respectively.

### B. Experimental protocol

**Training:** The six sub-tasks introduced in the 2<sup>nd</sup> FRCSyn-onGoing are mutually independent. This implies that participants have the flexibility to participate in any number of sub-tasks based on their preferences. For each selected sub-task, participants are required to develop a FR system

and train it twice: *i*) using the authorized real databases exclusively, *i.e.*, CASIA-WebFace [51], and *ii*) following the specific requirements of the chosen sub-task, as summarized in Table III. According to this protocol, participants must provide both the *baseline system* and the *proposed system* for each 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 the proposed model against the conventional practice of training only 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 and training protocol.

**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 [52] database, while Task 2 comprises four comparison files, each corresponding to every specific real-world databases considered (*i.e.*, BUPT, AgeDB [53], CFP-FP [54], and ROF [55]). During the evaluation of each sub-task, participants are required to submit two files per database through the Codalab platform: *i*) the scores of the baseline system, and *ii*) the scores of the proposed system. Finally, for each sub-task, participants must submit a file including 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.

### C. Evaluation Metrics

We evaluate the 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) [22] as follows:  $GAP = (\text{REAL} - \text{SYN}) / \text{SYN}$ , 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 a fixed operational point, or the Area Under the ROC Curve which are very popular for the analysis of FR systems in real-world applications, are also computed from the scores provided by participants. 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 II-B. We define the trade-off metric (TO) as follows:  $TO = AVG - SD$ . This metric involves plotting the average accuracy on the x-axis and the standard deviation on the y-axis in a 2D space. Multiple 45-degree parallel lines are drawn to identify the winning team, whose performance is located on the far right 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 establish the winners in Sub-Tasks 2.1, 2.2, and 2.3, we examine the average verification accuracy from the four different databases designated for evaluation, as described in Section II. This approach enables us to assess four main challenges of FR technologies: *i)* the representation of diverse demographic groups, *ii)* the impact of aging on recognition, *iii)* the variations in facial pose, and *iv)* the challenges made by occlusions. This evaluation provides a comprehensive overview of FR systems in real operational scenarios.

#### D. Restrictions

Participants have the freedom to choose the FR system for each task as long as the number of Floating Point Operations Per Second (FLOPs) of the

system 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 use their preferred training modality, with the requirement that only the specified databases are used for training. Generative models cannot be used to generate supplementary data. Participants are allowed to use non-face databases for pre-training purposes and use traditional data augmentation techniques using the authorized training databases. 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.

## IV. SECOND FRCSYN-ONGOING: SYSTEMS DESCRIPTION

In this 2<sup>nd</sup> FRCSyn-onGoing, we encourage participants to propose novel Generative AI methods for the creation of synthetic data. Besides, we also give participants the freedom to choose the FR architecture and training methods. Table IV summarizes for each team the key information in terms of the proposed synthetic data and FR system. This table serves as a quick reference, while more detailed explanations of each team's approach and methodology can be found in the corresponding subsections. Teams are arranged by their average ranking in the 6 sub-tasks from the 2<sup>nd</sup> FRCSyn-onGoing. In general, we can see that most teams have decided to use synthetic data from DC-Face [22] and IDiff-Face [23] databases, improving also the original data through cleaning and selection approaches, among other more sophisticated techniques. Also, regarding the FR technologies, most of them are based on ResNet [34] and IResNet [57] architectures, with AdaFace [7] and ArcFace [6] as the main used losses. However, some teams proposed their own methods to generate synthetic facial images, as well as to train their FR models. Next, we describe the specific details of the top-11 proposed systems in the 2<sup>nd</sup> FRCSyn-onGoing.

TABLE IV  
DESCRIPTION OF THE KEY INFORMATION IN TERMS OF THE PROPOSED SYNTHETIC DATA AND FR SYSTEM. LDM = LATENT DIFFUSION MODEL, DDPM = DIFFUSION PROBABILISTIC MODEL, DDM = DENOISING DIFFUSION IMPLICIT MODELS, HDT = HOURGLASS DIFFUSION TRANSFORMER

| Team               | Country                      | Sub Task   | Fig. | Synthetic Database          | Synthetic Data - Team Improvements   | FR Model - Training method  |
|--------------------|------------------------------|------------|------|-----------------------------|--|---|
| ADMIS              | China                        | All        | 2    | Novel                       | They used an LDM [58] to create the faces. The LDM uses ID embeddings as context. These contexts were generated with a DDPM [28]. They used cosine similarity to ensure the quality and a DDM [59] to accelerate the sampling. They finally oversampled the ID to enhance consistency. | They applied a 25% dropout on the dimensions from the feature embeddings.<br><b>Loss:</b> ArcFace [6]<br><b>Backbone:</b> ResNet-101 [57]   |
| OPDAI              | China                        | All        | 3    | DCFace [22]                 | They generated 10 new faces/ID with Photomaker [60] and replaced them randomly in DCFace.  | They used different heads for different databases and calculated the final loss<br><b>Loss:</b> AdaFace [7]<br><b>Backbone:</b> IResNet-100 [57]  |
| ID R&D             | USA                          | All        | 4    | Novel                       | They used an HDT [61] model to generate the synthetic images. It is trained with identity embeddings and style embeddings processed through a VQVAE [62]. They also used a StyleNAT [63] to generate variability from the images generated with the HDT.                               | They trained 2 models, one with color, geometric augmentations, and FaceMix-B [64] and the second with horizontal flipping augmentation. The final score was obtained by combining the outputs.<br><b>Loss:</b> UniFace [65]<br><b>Backbone:</b> IResNet-200 [57]   |
| K-IBS-DS           | S. Korea                     | All        | 5    | DCFace [22]                 | -  | They modified AdaFace following SlackedFace [66] by changing the initialization and replacing the L2-norm with the p-norm. They used 3 different backbones and the final score is the combination of them.<br><b>Loss:</b> SlackedFace [66]<br><b>Backbone:</b> IResNet-50/100/150 [57] w/ Squeeze-and-Excitation blocks [67] |
| CTAI               | China                        | All        | -    | DCFace [22]<br>GANDiff [21] | They cleaned the data using an IResNet with Squeeze-and-Excitation blocks. They also used DBSCAN to segregate intra-class noise and removed the IDs that were far from the ID class center.  | They used two models trained with different losses and occlusion augmentation. The final score was obtained by combining the outputs.<br><b>Loss:</b> AdaFace & CosFace [68]<br><b>Backbone:</b> ResNet-100 [57]  |
| Idiap-SynthDistill | Switzerland                  | 1.2<br>2.2 | 6    | Novel                       | They used an end-to-end method that generates synthetic images while training the FR model. The base data was generated with a StyleGAN2 [69] model and during training, they dynamically generated more images based on the training loss.  | They used a pretrained model to train by model distillation a new model using only the synthetic data they generated.<br><b>Loss:</b> SynthDistill [70]<br><b>Backbone:</b> ResNet-101 [57]   |
| INESC-IGD          | Portugal<br>Germany          | All        | 7    | DCFace [22]<br>IDiff [23]   | They labeled the images from the public databases with the ethnicity and balanced their final data.  | They trained two models using the same loss and applying RandAug and occlusion augmentation. The final score was obtained by combining the outputs from both models.<br><b>Loss:</b> ElasticCosFac-Plus [71]<br><b>Backbone:</b> ResNet-100 [34]  |
| UNICA-IGD-LSI      | Italy<br>Germany<br>Slovenia | All        | 8    | DCFace [22]                 | They used the public data from DCFace and generated data with the IDiff-Face Framework and ExFaceGAN [72].   | They took the similarity mean between real and synthetic samples and added it to the loss value.<br><b>Loss:</b> CosFace [68]<br><b>Backbone:</b> ResNet-101 [34]   |
| SRCN_AIVL          | China                        | 1.1        | 9    | DCFace [22]<br>IDiff [23]   | They balanced DCFace in ethnicity and trained IDiff-Face to be balanced as well.   | They trained two models and for the inference, they ingested one with the original image and the other with the flipped image. Then they merged the two resulting vectors and used the resulting vector to get the distance scores.<br><b>Loss:</b> AdaFace [7]<br><b>Backbone:</b> ResNet-101 [34]                           |
| CBSR-Samsung       | China                        | 1.3<br>2.3 | 10   | DCFace [22]                 | They de-overlaped DCFace from CASIA.   | They created a validation dataset using a subset of the DCFace, adding masks and sunglasses and positioning vertical bars in the images.<br><b>Loss:</b> ArcFace [6]<br><b>Backbone:</b> ResNet-100 [34]  |
| BOVFOCR-UFPR       | Brazil                       | 1.2<br>2.1 | 11   | DCFace [22]                 | -  | They used RandAugment, random erasing and random flip as augmentation techniques.<br><b>Loss:</b> ArcFace [6]<br><b>Backbone:</b> ResNet-100 [34]   |

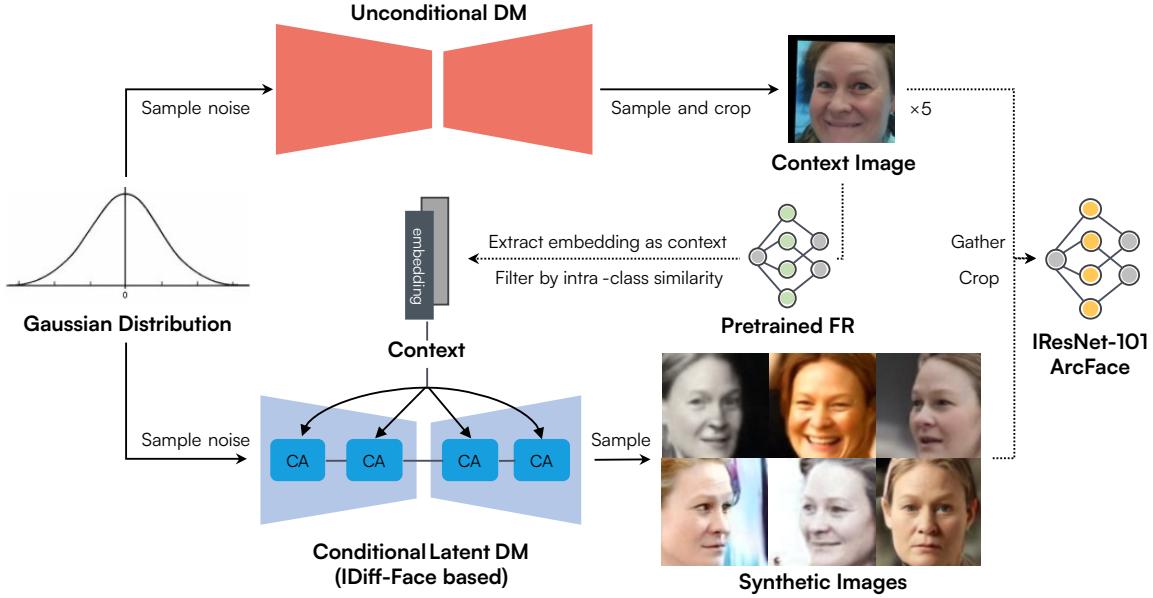


Fig. 2. Framework proposed by the ADMIS team.

**ADMIS (All sub-tasks):** This team comprises members from Fudan University and Tencent YouTu Lab, China. They used a Latent Diffusion Model (LDM) [58] based on IDiff-Face [23] to synthesize faces. While IDiff-Face uses a noise embedding sampled from a Gaussian distribution to serve as the LDM's, the ADMIS team uses identity embeddings as contexts, extracted from faces by a pretrained ElasticFace [71] model with an IResNet-101 [57] backbone. They trained the LDM with the CASIA-WebFace [51] database. As the LDM takes the ID embeddings as context, they considered an unconditional Denoising Diffusion Probabilistic Model (DDPM) trained on the FFHQ database [69] as a context generator. Specifically, they used the DDPM to generate 400K faces with arbitrary identities, known as context faces. They exploited the same ElasticFace model to extract the embeddings from the context faces. To encourage the quality and the distinctiveness of identity from later LDM-generated faces, they filtered the embeddings by setting a minimum cosine similarity threshold of 0.3 between arbitrary pairs of embeddings. This yields  $\approx 30K$  embeddings with discriminative iden-

tities. Furthermore, they accelerated the sampling process of the LDM by Denoising Diffusion Implicit Models (DDIM) [59]. For the training of the FR model, they generated 49 images for each context. They adopted the ID oversampling strategy from DCFace [22] and performed it five times for each ID to enhance consistency. As a result, 10K contexts were used 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. Both the baseline and proposed FR models used IResNet-101 architectures. They applied the ArcFace [6] loss with a batch size of 64 and an initial learning rate of 0.1 for 40 epochs. The learning rate was divided by 10 at epochs 22, 28, and 32. They also used random cropping augmentation during training. Their proposed architecture is described in Figure 2.

Code: [https://github.com/zzweakman/CVPR24\\_FRCSyn\\_ADMINIS](https://github.com/zzweakman/CVPR24_FRCSyn_ADMINIS)

**OPDAI (All sub-tasks):** This team comprises members from the Interactive Entertainment Group of Netease Inc., China. They initially used the DCFace [22] database, generating then 10 more

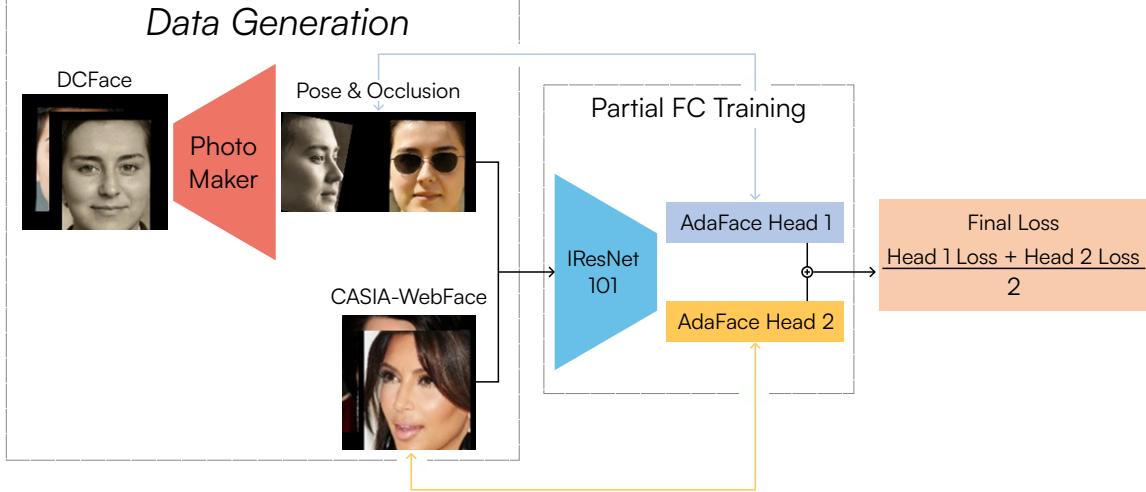


Fig. 3. Framework proposed by the OPDAI team.

face images for each ID with large pose variations and occlusions using Photomaker [60]. Given a few input ID images, Photomaker can generate diverse personalized ID photos based on a text prompt while preserving the identity information from the input image. 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 [51]. 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 [73], which consists in a sparse variant of model parallel architecture for training FR models. Regarding the FR model, they obtained the loss of different databases in independent AdaFace [7] heads, calculating the final loss as the average of the multiple heads. Both baseline and proposed models are based on IResNet-100 [57] architectures, with horizontal flipping. Their proposed architecture is described in Figure 3.

Code: [https://github.com/mightycatty/frcsyn\\_cvpr2024.git](https://github.com/mightycatty/frcsyn_cvpr2024.git)

**ID R&D (All sub-tasks):** This team comprises members from ID R&D Inc, USA. To generate the synthetic data, they used two models trained on WebFace42M [74]. The first model was based on Hourglass Diffusion Transformers (HDT) [61], which combines the scalability of Transformer architectures with the efficiency of convolutional U-Nets [75]. It was trained focusing on conditional flow matching [76] and following the classifier-free guidance approach [77]. Identity embeddings were used directly, whereas style embeddings were processed through a Vector Quantised-Variational AutoEncoder (VQVAE) [62]. Specifically, for head position, 32 embeddings were allocated for VQ-VAE processing, while age and facial expression attributes were represented with 8 embeddings each. During inference, combinations of these embeddings were randomly selected. The second model used to generate synthetic data was based on Style-NAT [63], enhanced with a FR model. This network was trained using auxiliary sources of supervision: a pre-trained FR network with Prototype Memory [78] (for identity supervision) and a pre-trained face attribute classification network (for style supervision). To create their synthetic data, they used classifier weights of the trained Prototype Memory to get 50K identity embeddings, of which 20K

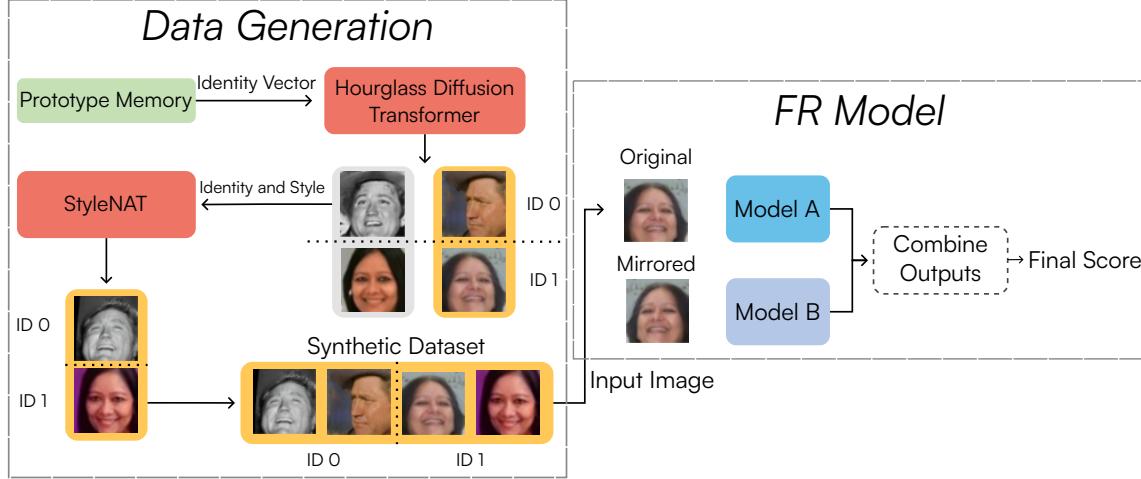


Fig. 4. Framework proposed by the ID R&D team.

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 face images using each of the two generative models. In the first stage, identity embeddings were used by the HDT model to get 10 images for each identity. Then 5 of these images were included in the training dataset. Identity and style embeddings were taken from the remaining 5 images and used as a condition to generate 5 different images with the StyleNAT model. These images were also put into the training dataset. Regarding the FR model, they trained an IResNet-200 [57] with UniFace [65] loss for 28 epochs. One network was trained with color, geometric augmentations, and FaceMix-B [64], and the other one using 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 with the CASIA-WebFace [51] data, training two models, one on the mixed data, and the other on the CASIA-WebFace. Their proposed architecture is described in Figure 4.

**K-IBS-DS (All sub-tasks):** This team comprises members from the Korea Advanced Institute of Science & Technology and the Institute for Basic Science, South Korea. They used DCFace [22] with 500K and 1.2M face images (depending on the sub-task). Regarding the FR model, they used several IResNet [57] models of 50, 100, and 152 layers with Squeeze-and-Excitation blocks [67], which are architectural components designed to enhance the representational power of convolutional neural networks by dynamically adjusting channel-wise features. Inspired by SlackedFace [66], they made two modifications to enhance the AdaFace [7] FR classifier: *i*) they used renormalized uniform initialization as a more reliable weight initialization for uniformity across identity prototypes in the unit sphere and *ii*) replaced the L2-norm with the powered-norm (p-norm), or face recognizability index from [66], which integrates the L2-norm with the learned embedding proximity. The training stage was in line with [22] and [29], including optimizer, learning rate, etc. For Sub-Tasks 1.3 and 2.3, the first 10K subjects of the CASIA-WebFace [51] 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 [21]). The

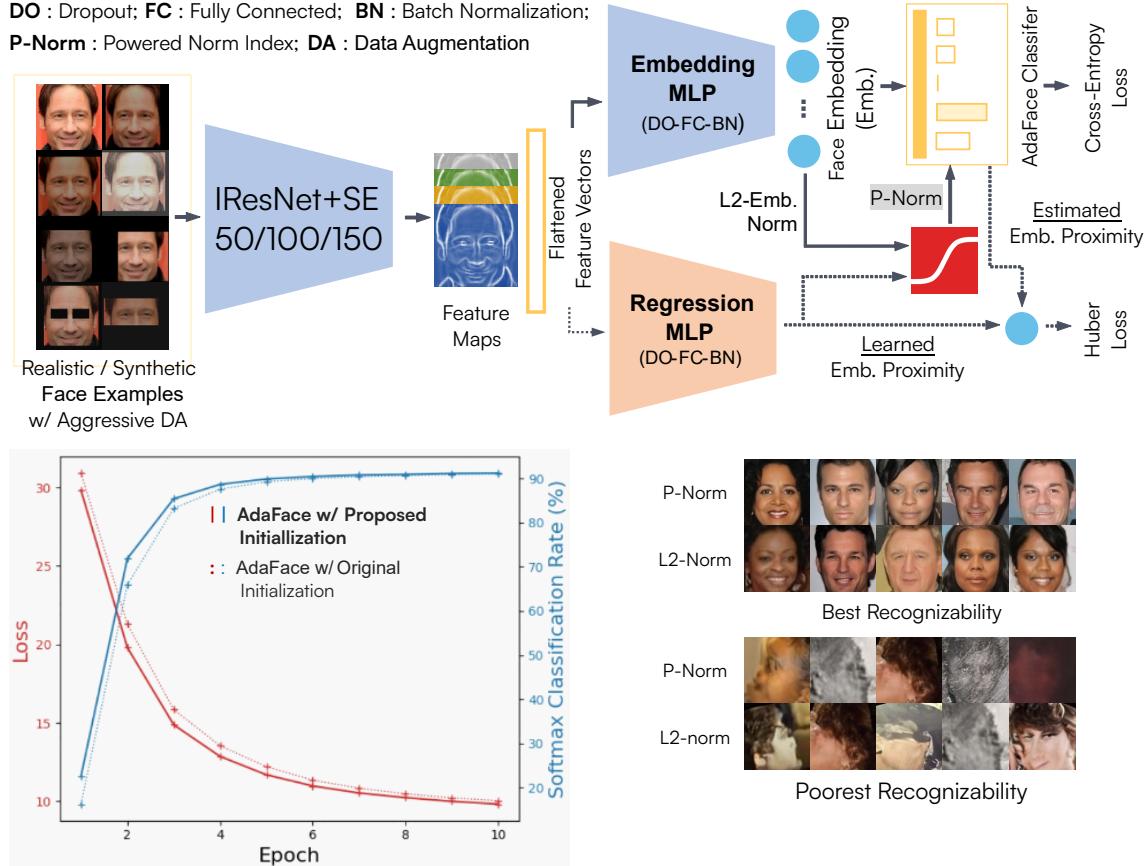


Fig. 5. Framework proposed by the K-IBS-DS team.

final score was obtained by aggregating the comparison scores of the different IResNet models, along with the horizontally flipped instances through score fusion. All training and test sets were realigned using RetinaFace [79], followed by a similarity transformation. For all sub-tasks, aggressive data augmentations were applied, including random horizontal flipping, photometric operations, cropping, resizing, and the addition of sunglasses and masks. Their proposed architecture is described in Figure 5. Code: [https://github.com/kalebmcs/cvpr\\_frcsyn](https://github.com/kalebmcs/cvpr_frcsyn)

**CTAI (All sub-tasks):** This team comprises members from China Telecom AI, China. 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 [57] model with Squeeze-and-Excitation blocks [67] using CASIA-WebFace [51] to extract features of synthetic images from DCFace [22], GANDiffFace [21], and DigiFace-1M [29]. Subsequently, they used the DBSCAN clustering method 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-

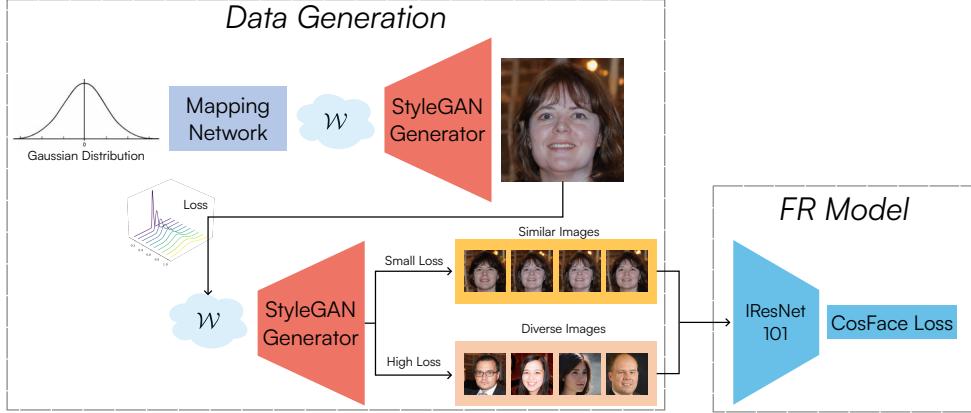


Fig. 6. Framework proposed by the Idiap-SynthDistill team.

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 loss [7] (A1) and CosFace loss [68] (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 considered to enhance all features.

Code: <https://github.com/liuhao-lh/FRCSyn-Challenge>

**Idiap-SynthDistill (Sub-Tasks 1.2 and 2.2):** This team comprises members from the Idiap Research Institute, EPFL, and Université de Lausanne, Switzerland. The proposed method was based on SynthDistill [70], 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 [69]. For generating synthetic images, they trained StyleGAN2 [80] with the CASIA-WebFace database [51] and then dynamically generated 20M synthetic images during training based on the training loss.

For the dynamic image generation, they used the training loss from every iteration as feedback to find the most difficult synthetic image in each batch and then re-sampled a new batch of synthetic images in the intermediate latent space  $\mathcal{W}$  of StyleGAN near the latent vector of the most difficult sample. If the loss value was high, they re-sampled with a relatively small standard deviation around the difficult sample and generated similar images, but if the loss value was small they re-sampled with a higher standard deviation, generating images with more variations. Throughout the process, the generated images were resized to  $112 \times 112$  before being fed to the FR models. They used a pre-trained FR model with IResNet-101 [57] architecture trained with CosFace [68] loss on a subset of the WebFace260M database [74] and trained a new model as a student network with the same architecture using synthetic data with their dynamic synthetic image generation approach. 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 [70]. For thresholding, a subset of DCFace [22] 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. Their proposed architecture is described in Figure 6.

Code: [https://gitlab.idiap.ch/bob/bob.paper.ijcb2023\\_synthdistill](https://gitlab.idiap.ch/bob/bob.paper.ijcb2023_synthdistill)

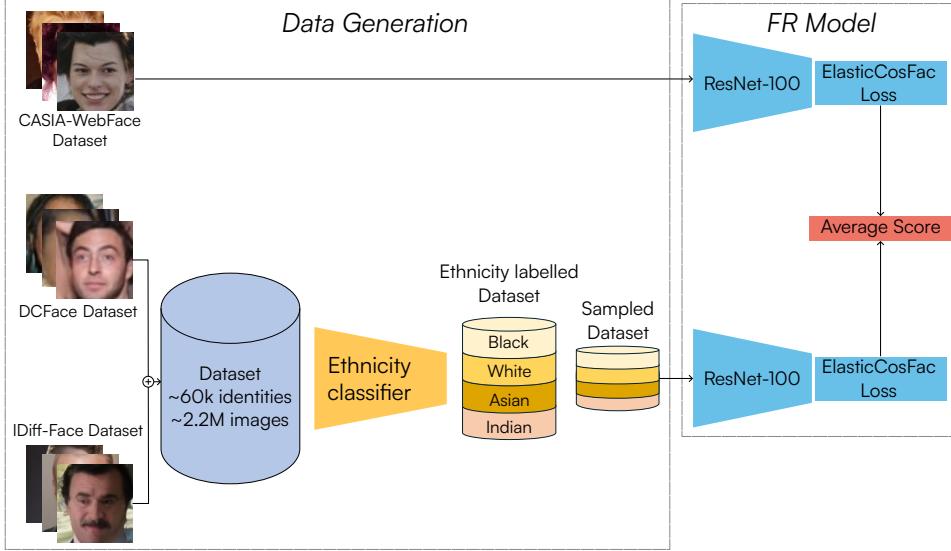


Fig. 7. Architecture proposed by the INESC-IGD team.

**INESC-IGD (All sub-tasks):** This team comprises members from INESC TEC and Universidade do Porto, Portugal, and Fraunhofer IGD, Germany. For the training dataset, they merged DCFace [22], IDiff-Face Uniform, and IDiff-Face Two-stage [23] databases and then labeled the data with ethnicity labels using a similar approach to [81]. 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 [34] were trained, one on CASIA-WebFace [51] and the other on a subset of synthetic datasets (*e.g.*, 400K images of 9K identities). For all sub-tasks, they trained a ResNet-100 with ElasticCosFac-Plus loss [71] using the settings presented in [71]. 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 used. During the FR training of all sub-tasks, the training datasets were augmented using RandAug and occluded augmentation [82] with probabilities

of 0.4. The occluded augmentation followed protocol 4 proposed in [82], leading to occlusions on the eyes, lower face, upper face, or a combination of the eyes occlusion with the others. Occluded augmentations boosted the performance, as synthetic data has a lower frequency of natural occlusions such as beard and makeup [83]. Their proposed architecture is described in Figure 7.

Code: <https://github.com/NetoPedro/Equilibrium-Face-Recognition>

**UNICA-IGD-LSI (All sub-tasks):** This team comprises members from Fraunhofer IGD, Germany, University of Cagliari, Italy, and University of Ljubljana, Slovenia. They used the DCFace [22] synthetic database as it led to remarkable performance gains under well-known evaluation benchmarks for face verification, while combined with real data [84]. Also, they considered synthetic data generated with IDiff-Face [23] and ExFaceGAN [72] in Sub-Tasks 1.1, 1.2, 2.1, and 2.2. The ExFaceGAN data was generated using an identity disentanglement approach on pretrained GAN-Control [85]. Regarding the FR model, they trained a ResNet-100 [34] network using CosFace loss [68] with a margin penalty of 0.35 and a scale term of

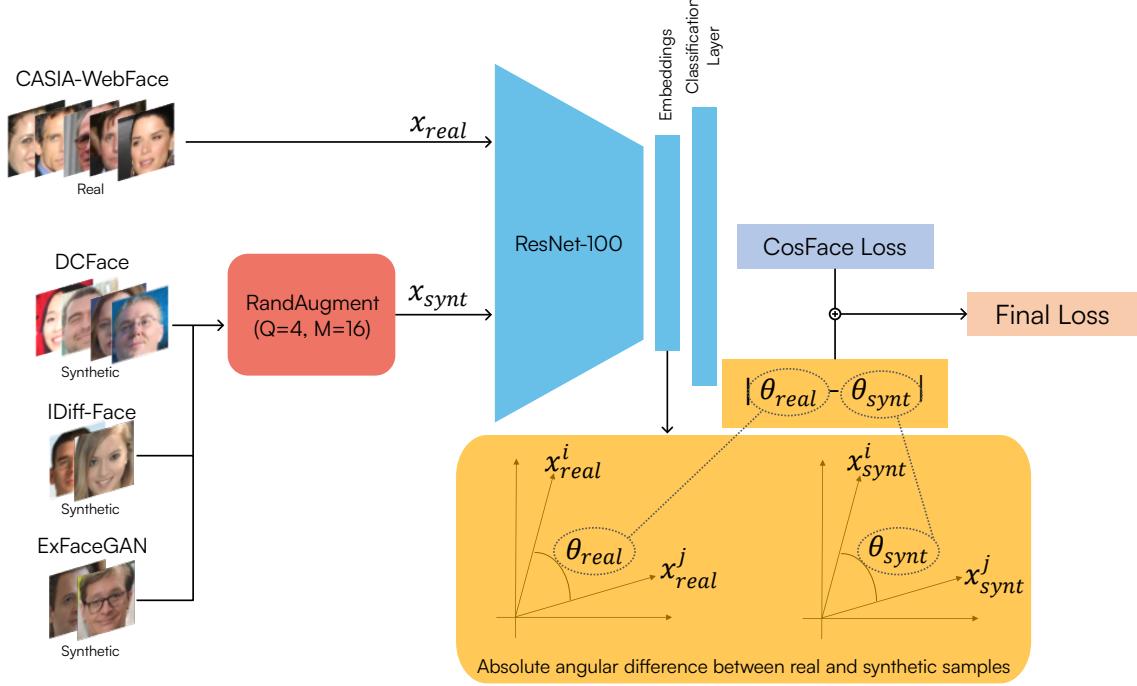


Fig. 8. Framework proposed by the UNICA-IGD-LSI team.

64. The similarity mean difference between real-only and synthetic-only samples ( $|\theta_{real} - \theta_{synt}|$ ) was scaled and added to the loss value. They trained the FR 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 [84], [86]. For Sub-Tasks 1.3 and 2.3, the selected synthetic dataset was combined with CASIA-Webface [51], obtaining a total of 1M images from 20,572 identities. Their proposed architecture is described in Figure 8.

Code: <https://github.com/atzoriandrea/FRCSyn2>

**SRCN\_AIVL (Sub-Task 1.1):** This team comprises members from Samsung Electronics (China) R&D Centre, University of Science and Technology, IIE, CAS, and MAIS, CASIA, China. They selected 400K samples from the DCFace [22] database and labeled the ethnicity of each subject, as the racial distribution gap may lead to bad

performance in testing. Based on this approach, they trained IDiff-Face [23] with CASIA-WebFace [51] database generating 100K synthetic face images of specific races. Regarding the FR system, they used two custom ResNet-101 [34] trained with AdaFace loss [7] function. The models were trained for 60 epochs with an initial learning rate of 0.1 and a batch size of 512, 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. The threshold was determined by the 10-fold optimal threshold in the validation set. For the inference, data preprocessing involved an MTCNN [87] for the alignment and resizing all data to  $112 \times 112$ . After cropping and alignment, they fed the image and the flipped image into the two models. Finally, after obtaining the two feature embeddings, they combined them and performed the similarity calculation with these embeddings.

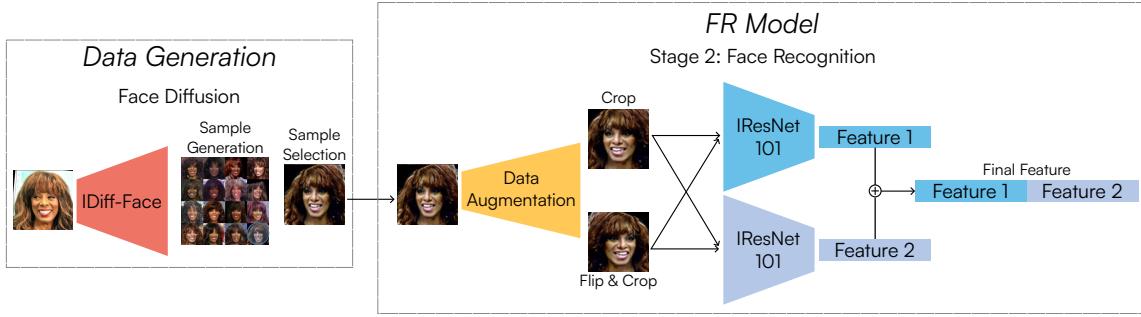


Fig. 9. Framework proposed by the SRCN\_AIVL team.

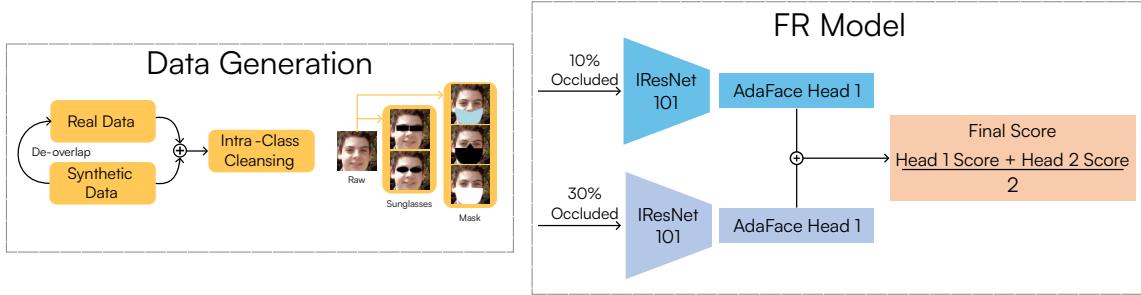


Fig. 10. Framework proposed by the CBSR-Samsung team.

Their proposed architecture is described in Figure 9. Code: <https://github.com/Value-Jack/2nd-Edition-FRCSyn>

#### **CBSR-Samsung (Sub-Tasks 1.3 and 2.3):**

This team comprises members from Samsung Electronics (China) R&D Centre, IIE, CAS, and MAIS, CASIA, China. They first trained a FR model using CASIA-WebFace [51]. Then, they used it to de-overlap DCFace [22] from CASIA, as DCFace was trained using that real database. For the synthetic dataset, they compared the performance of models trained with three synthetic databases, including GANDiffFace [21], DCFace, and IDiff-Face [23], 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, simulating age variability and demographic groups as in AgeDB [53] and BUPT-Balanced [52] databases, respectively; *ii*) randomly positioned vertical bar masks to the images to simulate the self-occlusion due to as considered in CFP-FP database [54];

and *iii*) add mask and sunglasses to the images by detecting the landmarks via FaceX-Zoo [88], simulating the ROF database [55]. This is done following [89]. 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. Regarding the FR model, they merged the refined datasets and trained IResNet-100 [57] with AdaFace loss [7]. 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. Their proposed architecture is shown in Figure 10.

#### **BOVIFOCR-UFPR (Sub-Tasks 1.2 and 2.1):**

This team comprises members from Federal Uni-

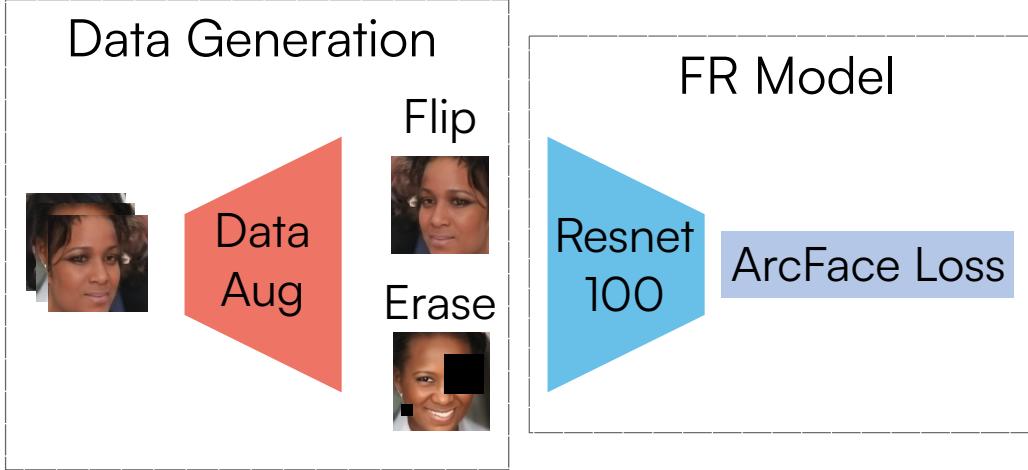


Fig. 11. Framework proposed by the BOVIFOOCR-UFPR team.

versity of Paraná, Federal Institute of Mato Grosso, and unico - idTech, Brazil. They chose DCFace [22] as the synthetic database and randomly removed 910 identities with 55 images per ID to reduce the number to follow the rules. For the FR model, they used a ResNet-100 [34] as the backbone, trained with the ArcFace [6] 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. To validate their model they subsampled images from DCFace, generating genuine and impostor pairs, and used these pairs to select the best threshold to classify the proposed model output scores. Their proposed architecture is described in Figure 11.

Code: <https://github.com/PedroB Vidal/insightface>

## V. SECOND FRC SYN-ONGOING: RESULTS

Next, we describe in Sections V-A and V-B the main results achieved in Tasks 1 and 2, respectively. These results are further analyzed in Section V-C focusing on specific demographic groups and individual databases. Finally, we discuss in Section V-D common trends among the different teams and compare the results with those obtained in the 1<sup>st</sup> edition. We present in Table V the ranking and key results of the 2<sup>nd</sup> FRCSyn-onGoing.

### A. Task 1: Bias Mitigation

Table VI shows the results achieved by participants in Task 1, focused on demographic bias mitigation. Teams are ranked by descending order of TO, which tends to correlate to 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), demonstrates a significant negative GAP value (-5.31%), showing a higher performance when training the FR system with synthetic data compared to real data (*i.e.*, CASIA-WebFace [51]). Furthermore, in Sub-Task 1.1, we can observe that teams with negative GAP values considered Diffusion Models for the generation of synthetic data (*i.e.*, ID R&D uses HDT, SRCN\_AIVL combines DCFace and IDiff-Face, and CTAI combines DCFace and GANDiff-Face), showing that this generation method may work better than real data in scenarios with limited data. Next, after removing the limitation on the number of synthetic images (*i.e.*, Sub-Task 1.2), the TO value of most FR systems increases, which leads to performance and fairness improvement simultaneously. For instance, for the ADMIS team (ranked top-2 in both Sub-Task 1.1 and 1.2), the TO value increases to 95.72% in Sub-Task 1.2 (*i.e.*, 1.42% TO improvement compared to Sub-Task 1.1).

TABLE V

RESULTS 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. FOR EACH TEAM, WE REPORT THE RANKING METRIC AND THE POSITION ACROSS ALL THE SUB-TASKS. THE BEST RESULT OF EACH SUB-TASK IS HIGHLIGHTED IN BOLD. WE MARK WITH A ‘-’ IF THE TEAM DID NOT PARTICIPATE IN A SUB-TASK. TO = TRADE-OFF, AVG = AVERAGE ACCURACY

| Team               | Task 1: Bias Mitigation |                    |                    | Task 2: Overall Improvement |                     |                     |
|--------------------|-------------------------|--------------------|--------------------|-----------------------------|---------------------|---------------------|
|                    | Task 1.1<br>TO [%]      | Task 1.2<br>TO [%] | Task 1.3<br>TO [%] | Task 2.1<br>AVG [%]         | Task 2.2<br>AVG [%] | Task 2.3<br>AVG [%] |
| ADMIS              | 94.30 (2)               | 95.72 (2)          | <b>96.50 (1)</b>   | 91.19 (3)                   | 92.92 (2)           | 94.15 (5)           |
| OPDAI              | 93.75 (4)               | 94.12 (3)          | 95.96 (4)          | <b>91.93 (1)</b>            | 92.04 (3)           | 95.23 (2)           |
| ID R&D             | <b>96.73 (1)</b>        | <b>96.73 (1)</b>   | 86.73 (8)          | 91.86 (2)                   | 91.86 (4)           | 94.05 (6)           |
| K-IBS-DS           | 92.91 (6)               | 93.72 (5)          | 96.17 (2)          | 91.05 (4)                   | 91.61 (5)           | <b>95.42 (1)</b>    |
| CTAI               | 93.21 (5)               | 93.21 (6)          | 95.41 (7)          | 90.59 (5)                   | 90.59 (6)           | 94.56 (3)           |
| Idiap-SynthDistill | -                       | 89.70 (9)          | -                  | -                           | <b>93.50 (1)</b>    | -                   |
| INESC-IGD          | 92.28 (7)               | 94.05 (4)          | 95.65 (5)          | 83.16 (8)                   | 85.40 (7)           | 89.43 (8)           |
| UNICA-IGD-LSI      | 91.89 (8)               | 91.89 (7)          | 96.00 (3)          | 87.80 (7)                   | 87.80 (8)           | 92.79 (7)           |
| SRCN_AIVL          | 94.06 (3)               | -                  | -                  | -                           | -                   | -                   |
| CBSR-Samsung       | -                       | -                  | 95.57 (6)          | -                           | -                   | 94.20 (4)           |
| BOVIFOCR-UFPR      | -                       | 90.48 (8)          | -                  | 89.97 (6)                   | -                   | -                   |

Also, the GAP value decreases from 1.47% to -0.56%, obtaining better results when increasing the amount of synthetic data in comparison to limited real data (*i.e.*, CASIA-WebFace). Another example is the OPDAI team, which raised from top-4 to top-3 positions between Sub-Task 1.1 and 1.2. Its TO value increases to 94.12% (*i.e.*, 0.37% TO improvement from Sub-Tasks 1.1 to 1.2), and the GAP value is reduced from 1.02% to 0.71%. These findings emphasize the potential of generating large number of synthetic face images from different demographic groups to mitigate bias in existing FR technology. Finally, we analyze in Sub-Task 1.3 the case of using both, real and synthetic data, in the FR training process. In general, we can observe considerable improvements in terms of TO values, along with higher 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). Moreover, Table VII shows the performance of each demographic group for the baseline models, *i.e.*, those trained only with real data, and the models proposed by the top-ranked teams for each of the sub-tasks of Task 1. The ID R&D team achieves higher performance across all demographic

groups except for white males and females (98.40% vs. 97.90% AVG, and 97.60% vs. 96.50% AVG, respectively). We believe this is produced due to the class imbalance in the real data for this demographic group, which also leads to achieving the best performance for the case of training only with real data, due to overfitting. Similarly, the ADMIS team achieves higher performance across all demographic groups except for black males (98.00% vs. 97.10% AVG). Moreover, when we analyze the scenario of training only with synthetic data, not only does the overall performance increase, but the bias between demographic groups is also reduced (0.82% vs 4.40% STD for Sub-Tasks 1.1 and 1.2, and 0.75% vs 1.81% STD for Sub-Task 1.3), which implies a better model in terms of fairness across all demographic groups. This is mainly produced as with synthetic data we can control how balanced the data is with respect to these groups. Finally, it is noteworthy 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, showing that unlimited synthetic data for training can even outperform FR systems trained

TABLE VI

RANKING FOR THE THREE SUB-TASKS CONSIDERED IN TASK 1. FOR EACH SUB-TASK, WE HIGHLIGHT IN BOLD THE BEST TEAM ACCORDING TO THE TRADE-OFF. TO = TRADE-OFF, AVG = AVERAGE ACCURACY, SD = STANDARD DEVIATION OF ACCURACY, FNMR = FALSE NON-MATCH RATE, FMR = FALSE MATCH RATE, AUC = AREA UNDER CURVE, GAP = GAP TO REAL.

| <b>Sub-Task 1.1 (Bias Mitigation): Synthetic Data (Constrained)</b> |                   |              |              |             |                 |                  |              |
|---|-------------------|--------------|--------------|-------------|-----------------|------------------|--------------|
| Pos.  | Team              | TO [%]       | AVG [%]      | SD [%]      | FNMR@FMR=1%     | AUC [%]          | GAP [%]      |
| <b>1</b>  | <b>ID R&amp;D</b> | <b>96.73</b> | <b>97.55</b> | <b>0.82</b> | <b>3.17 (1)</b> | <b>99.57 (1)</b> | <b>-5.31</b> |
| 2   | ADMIS             | 94.30        | 95.10        | 0.80        | 11.38 (3)       | 98.96 (3)        | 1.47         |
| 3   | SRCN_AIVL         | 94.06        | 95.12        | 1.07        | 10.72 (2)       | 98.83 (4)        | -0.54        |
| 4   | OPDAI             | 93.75        | 94.92        | 1.17        | 11.85 (4)       | 99.51 (2)        | 1.02         |
| 5   | CTAI              | 93.21        | 94.74        | 1.53        | 14.38 (5)       | 98.33 (6)        | -0.63        |
| 6   | K-IBS-DS          | 92.91        | 94.11        | 1.2         | 15.03 (6)       | 98.47 (5)        | 1.58         |

| <b>Sub-Task 1.2 (Bias Mitigation): Synthetic Data (Unconstrained)</b> |                   |              |              |             |                 |                  |              |
|---|-------------------|--------------|--------------|-------------|-----------------|------------------|--------------|
| Pos.  | Team              | TO [%]       | AVG [%]      | SD [%]      | FNMR@FMR=1%     | AUC [%]          | GAP [%]      |
| <b>1</b>  | <b>ID R&amp;D</b> | <b>96.73</b> | <b>97.55</b> | <b>0.82</b> | <b>3.17 (1)</b> | <b>99.57 (2)</b> | <b>-5.31</b> |
| 2   | ADMIS             | 95.72        | 96.50        | 0.78        | 6.33 (2)        | 99.51 (3)        | -0.56        |
| 3   | OPDAI             | 94.12        | 95.22        | 1.11        | 10.78 (3)       | 98.92 (4)        | 0.71         |
| 4   | INESC-IGD         | 94.05        | 95.22        | 1.17        | 11.03 (4)       | 98.70 (5)        | 1.04         |
| 5   | K-IBS-DS          | 93.72        | 94.88        | 1.16        | 12.75 (5)       | 99.66 (1)        | 0.77         |
| 6   | CTAI              | 93.21        | 94.74        | 1.53        | 14.38 (6)       | 98.33 (6)        | -0.63        |

| <b>Sub-Task 1.3 (Bias Mitigation): Synthetic + Real Data (Constrained)</b> |               |              |              |             |                 |                  |              |
|--|---------------|--------------|--------------|-------------|-----------------|------------------|--------------|
| Pos.   | Team          | TO [%]       | AVG [%]      | SD [%]      | FNMR@FMR=1%     | AUC [%]          | GAP [%]      |
| <b>1</b>   | <b>ADMIS</b>  | <b>96.50</b> | <b>97.25</b> | <b>0.75</b> | <b>3.90 (1)</b> | <b>99.72 (1)</b> | <b>-1.33</b> |
| 2  | K-IBS-DS      | 96.17        | 96.92        | 0.75        | 5.88 (4)        | 99.54 (2)        | -1.37        |
| 3  | UNICA-IGD-LSI | 96.00        | 96.70        | 0.70        | 5.90 (5)        | 99.49 (3)        | -5.33        |
| 4  | OPDAI         | 95.96        | 96.80        | 0.84        | 4.90 (2)        | 99.54 (2)        | -0.03        |
| 5  | INESC-IGD     | 95.65        | 96.33        | 0.67        | 6.15 (6)        | 99.18 (5)        | -0.12        |
| 6  | CBSR-Samsung  | 95.57        | 96.54        | 0.97        | 5.00 (3)        | 99.41 (4)        | -24.43       |

with limited synthetic + real data. These results motivate the use of synthetic data for demographic bias mitigation, improving at the same time privacy as no real identities are seen by the network.

#### B. Task 2: Overall Improvement

Table VIII provides the results achieved by participants in Task 2, focusing not only on bias mitigation but also other challenges in FR such as age, pose, and occlusions. Teams are ranked in descending order based on the average verification accuracy across the four databases. Notably, in all

sub-tasks, the AVG is lower than the achieved in Task 1 for the BUPT-BalancedFace [52] database, showing the additional challenges introduced by the AgeDB [53], CFP-FP [54], and ROF [55] databases. This trend can also be observed in GAP results, which tend to be worse for Sub-Tasks 2.1 and 2.2 compared to Sub-Tasks 1.1 and 1.2, suggesting that it is far more difficult to emulate the conditions of these real databases with synthetic data by itself. For example, in Sub-Task 2.1, although the top-3 teams achieve high AVG results, they exhibit a considerable positive GAP value (*i.e.*, OPDAI

TABLE VII

COMPARISON BETWEEN THE BASELINE MODEL (TRAINED EXCLUSIVELY WITH REAL DATA) AND THE PROPOSED MODEL (TRAINED WITH SYNTHETIC DATA) FOR EACH FINALIST OF TASK 1. THE PERFORMANCE OF EACH DEMOGRAPHIC GROUP IS REPRESENTED BY ITS AVG [%].

| Sub Task | Team   | Model    | Asian Female | Asian Male   | Black Female | Black Male   | Indian Female | Indian Male  | White Female | White Male   | AVG [%]      | STD [%]     | TO [%]       | GAP [%] |
|----------|--------|----------|--------------|--------------|--------------|--------------|---------------|--------------|--------------|--------------|--------------|-------------|--------------|---------|
| 1.1      | ID R&D | Proposed | <b>96.80</b> | <b>96.90</b> | <b>97.20</b> | <b>97.80</b> | <b>98.10</b>  | <b>99.20</b> | 96.50        | 97.90        | <b>97.55</b> | <b>0.82</b> | <b>96.73</b> | -5.31   |
|          |        | Baseline | 85.70        | 90.00        | 94.10        | 95.20        | 87.10         | 90.80        | <b>97.60</b> | <b>98.40</b> | 92.36        | 4.40        | 87.96        |         |
| 1.3      | ADMIS  | Proposed | <b>96.20</b> | <b>97.30</b> | <b>97.30</b> | 97.10        | <b>96.20</b>  | <b>97.40</b> | <b>98.60</b> | <b>97.90</b> | <b>97.25</b> | <b>0.75</b> | <b>96.50</b> | -1.33   |
|          |        | Baseline | 92.90        | 94.60        | 97.00        | <b>98.00</b> | 93.70         | 96.90        | 97.10        | 97.50        | 95.96        | 1.81        | 94.15        |         |

91.93% AVG, 3.09% GAP; ID R&D 91.86% AVG, 2.99% GAP; ADMIS 91.19% AVG, 2.78% GAP), showing that a FR model trained only with real data (*i.e.*, CASIA-WebFace) adapts better to adverse image conditions such as aging, pose, or occlusions. Focusing on Sub-Task 2.2, the team ranked top-1, *i.e.*, Idiap-SynthDistill, achieves much better results compared to the best result of Sub-Task 2.1 (*i.e.*, 93.50% vs. 91.93% AVG), proving that unlimited synthetic data can further improve the performance of the system. Finally, in Sub-Task 2.3, most teams report better AVG and higher negative GAP values (*e.g.*, K-IBS-DS achieves 95.42% AVG, -2.15% GAP), proving again that synthetic data combined with real data can alleviate existing limitations within FR technology.

Furthermore, Table IX compares the performance across the different evaluation databases, highlighting the variations between the baseline models, which were trained only on real data, and the models proposed by the top-ranked teams for the sub-tasks of Task 2. In Sub-Task 2.1, where the model proposed by OPDAI is trained with limited synthetic data due to competition rules, it underperforms compared to the baseline across all databases (3.09% GAP). This suggests that 500,000 synthetic samples are insufficient to fully replace real data. Moreover, in Sub-Task 2.2, where training synthetic data is unlimited, the performance of the model proposed by Idiap-SynthDistill is almost identical to the baseline (-0.05% GAP). This suggests that a large synthetic dataset can closely replicate the distribution of real data. We can also observe that in AgeDB and BUPT databases, the baseline still outperforms the proposed model, which indicates that

synthetic data does not fully capture the differences between aging or demographic groups. Finally, in Sub-Task 2.3, which combines real and unlimited synthetic data, the model proposed by K-IBS-DS outperforms the baseline in all databases (-2.16% GAP). This supports the idea of all our previous experiments, that synthetic data complements real data, improving generalization.

### C. Demographic Groups and Evaluation Databases

This section provides an in-depth analysis of the results in terms of the different demographic groups and individual databases considered in the 2<sup>nd</sup> FRCSyn-onGoing. Figure 12 (left), shows the Detection Error Tradeoff (DET) curves of Sub-Tasks 1.1, 1.2, and 1.3, including the results achieved for the top-1 team in each demographic group. For completeness, the information and graphical representations for all the teams can be found on the challenge Codalab platform<sup>12</sup>.

For Sub-Tasks 1.1 and 1.2, the team that achieves the first place, ID R&D, demonstrates high performance across the different demographic groups considered (above 96.50% Accuracy for all demographic groups). However, a slight gender bias can be observed (improvements of  $\approx 1\%$  between ‘Male’ and ‘Female’ labels for some ethnicities). Regarding the ethnicity, the proposed FR model showed better results for subjects from the ‘Indian’ ethnicity (99.20% Accuracy for Indian Male; 98.10% Accuracy for Indian Female). Finally, for Sub-Task 1.3, the winning team, ADMIS, also performs well across all demographic groups (all above

<sup>12</sup><https://codalab.lisn.upsaclay.fr/competitions/16970#results>

TABLE VIII

RANKING FOR THE THREE SUB-TASKS CONSIDERED IN TASK 2. FOR EACH SUB-TASK, WE HIGHLIGHT IN BOLD THE BEST TEAM ACCORDING TO THE AVERAGE ACCURACY. AVG = AVERAGE ACCURACY, FNMR = FALSE NON-MATCH RATE, FMR = FALSE MATCH RATE, AUC = AREA UNDER CURVE, GAP = GAP TO REAL.

| <b>Sub-Task 2.1 (Overall Improvement): Synthetic Data (Constrained)</b>        |                           |              |                  |                  |              |
|--|---------------------------|--------------|------------------|------------------|--------------|
| Pos.   | Team                      | AVG [%]      | FNMR@FMR=1%      | AUC [%]          | GAP [%]      |
| <b>1</b>   | <b>OPDAI</b>              | <b>91.93</b> | <b>17.63 (2)</b> | <b>97.30 (2)</b> | <b>3.09</b>  |
| 2  | ID R&D                    | 91.86        | 10.36 (1)        | 97.48 (1)        | 2.99         |
| 3  | ADMIS                     | 91.19        | 20.41 (3)        | 97.04 (3)        | 2.78         |
| 4  | K-IBS-DS                  | 91.05        | 24.87 (6)        | 96.09 (6)        | 2.60         |
| 5  | CTAI                      | 90.59        | 21.88 (4)        | 96.40 (5)        | -1.94        |
| 6  | BOVIFOCR-UFPR             | 89.97        | 24.04 (5)        | 96.70 (4)        | 3.71         |
| <b>Sub-Task 2.2 (Overall Improvement): Synthetic Data (Unconstrained)</b>      |                           |              |                  |                  |              |
| Pos.   | Team                      | AVG [%]      | FNMR@FMR=1%      | AUC [%]          | GAP [%]      |
| <b>1</b>   | <b>Idiap-SynthDistill</b> | <b>93.50</b> | <b>9.17 (1)</b>  | <b>97.17 (4)</b> | <b>-0.05</b> |
| 2  | ADMIS                     | 92.92        | 14.45 (3)        | 97.76 (1)        | 0.21         |
| 3  | OPDAI                     | 92.04        | 16.48 (4)        | 97.41 (3)        | 3.00         |
| 4  | ID R&D                    | 91.86        | 10.36 (2)        | 97.48 (2)        | 2.99         |
| 5  | K-IBS-DS                  | 91.61        | 22.48 (6)        | 96.54 (5)        | 1.96         |
| 6  | CTAI                      | 90.59        | 21.88 (5)        | 96.40 (6)        | -1.94        |
| <b>Sub-Task 2.3 (Overall Improvement): Synthetic + Real Data (Constrained)</b> |                           |              |                  |                  |              |
| Pos.   | Team                      | AVG [%]      | FNMR@FMR=1%      | AUC [%]          | GAP [%]      |
| <b>1</b>   | <b>K-IBS-DS</b>           | <b>95.42</b> | <b>9.49 (5)</b>  | <b>98.14 (6)</b> | <b>-2.15</b> |
| 2  | OPDAI                     | 95.23        | 7.54 (1)         | 98.70 (1)        | -0.52        |
| 3  | CTAI                      | 94.56        | 8.85 (4)         | 98.41 (3)        | -6.01        |
| 4  | CBSR-Samsung              | 94.20        | 8.62 (3)         | 98.17 (4)        | -4.40        |
| 5  | ADMIS                     | 94.15        | 10.99 (6)        | 98.46 (2)        | -1.10        |
| 6  | ID R&D                    | 94.05        | 8.00 (2)         | 98.16 (5)        | 0.07         |

96% Accuracy). However, there exists variability in performance between different demographic groups. For example, ‘Asian Females’ and ‘Indian Females’ have the lowest Accuracy (96.20%) while ‘White Females’ have the highest Accuracy (98.60%).

Figure 12 (right) shows the DET curves of

Sub-Tasks 2.1, 2.2, and 2.3, including the results achieved for the top-1 team in each database. Analyzing the FR model proposed by the OPDAI team for Sub-Task 2.1, the spread of the curves indicates variability in the system performance across different databases, with the results from AgeDB (94.54% Accuracy) outperforming others.

TABLE IX

COMPARISON BETWEEN THE BASELINE MODEL (TRAINED EXCLUSIVELY WITH REAL DATA), AND THE PROPOSED MODEL (USING SYNTHETIC DATA) FOR EACH FINALIST OF TASK 2. THE PERFORMANCE OF EACH DATABASE IS REPRESENTED BY ITS AVG [%].

| Sub-Task | Team                  | Model    | AgeDB        | BUPT         | CFP-FP       | ROF          | AVG [%]      | STD [%]     | GAP [%] |
|----------|-----------------------|----------|--------------|--------------|--------------|--------------|--------------|-------------|---------|
| 2.1      | OPDAI                 | Proposed | 94.54        | 93.42        | 92.51        | 87.22        | 91.93        | 2.81        | 3.09    |
|          |                       | Baseline | <b>96.78</b> | <b>93.47</b> | <b>97.29</b> | <b>91.39</b> | <b>94.73</b> | <b>2.42</b> |         |
| 2.2      | Idiap<br>SynthDistill | Proposed | 96.72        | 93.85        | <b>96.14</b> | <b>87.31</b> | 93.5         | <b>3.74</b> | -0.05   |
|          |                       | Baseline | <b>97.6</b>  | <b>97.22</b> | 95.1         | 84.14        | <b>93.52</b> | 5.5         |         |
| 2.3      | K-IBS-DS              | Proposed | <b>96.89</b> | <b>96.88</b> | <b>97.51</b> | <b>90.39</b> | <b>95.42</b> | <b>2.91</b> | -2.16   |
|          |                       | Baseline | 96.2         | 93.76        | 95.97        | 87.58        | 93.38        | 3.48        |         |

Moreover, in Sub-Task 2.2 the Idiap-SynthDistill FR model significantly improves the performance of Sub-Task 2.1 for AgeDB and CFP-FP databases (*i.e.*, 96.72% and 96.14% Accuracy, respectively). Finally, for Sub-Task 2.3, the curves from the K-IBS-DS FR model are closely aligned for the AgeDB (96.89% Accuracy), BUPT (96.88% Accuracy), and CFP-FP (97.51% Accuracy), showing consistent and reliable performance across these databases. However, the curve from the ROF database remains the worst in each sub-task (*i.e.*, 87.22%, 87.31%, and 90.39% Accuracy for Sub-Tasks 2.1, 2.2, and 2.3, respectively), reflecting that it is the most difficult database to emulate with synthetic data.

#### D. Post-Challenge Analysis and Comparison with 1<sup>st</sup> Edition

Analyzing the contributions of all eleven top teams, we can observe the prevalence of well-established methodologies. Notably, most teams used DCFace [22] either independently or in conjunction with other synthetic databases such as GANDiffFace [21], DigiFace-1M [29], or IDiffFace [23]. DCFace is a Dual Condition Face Generator based on a diffusion model, designed to create facial images of the same subject in various styles while maintaining identity consistency. Its key component is the *Patch-wise Style Extractor*, which extracts style features from an image while minimizing identity information. This forces the model to rely on a separate input for identity data. Unlike previous approaches like StyleGAN [69], DCFace

retains essential spatial details, such as pose, ensuring greater variability between subjects of the same identity. This results in images with a similar style that enhance the performance of facial recognition models in identifying subjects. Furthermore, several teams, including CBSR-Samsung, INESC-IGD, and CTAI adopted interesting approaches involving synthetic data cleaning and selection. These approaches include: *i*) de-overlapping the data from DCFace, as it is trained with CASIA-WebFace and some data could be very similar, deteriorating the training, *ii*) balancing the data with respect to the demographic information, and *iii*) removing the images that are far from the class center using clustering techniques such as DBSCAN. It is interesting to highlight the ID R&D and Idiap-SynthDistill teams as they considered novel methods to generate synthetic data. Specifically, the ID R&D team used an HDT [61] to generate synthetic facial images along with identity and style embeddings, which were used by a StyleNAT [63] model to generate more variability in the synthetic data. Another example is the Idiap-SynthDistill team, which proposed an end-to-end method that dynamically generated facial images through StyleGAN2 [70] and trained a FR model through model distillation. Regarding the backbone architecture, all teams opted for either ResNet [34] or IResNet [57], mainly for their widespread adoption in state-of-the-art FR methodologies. Both architectures use residual connections to improve the training, but while in ResNet, the skip connections bypass one or more layers to address the vanishing gradient problem, IResNet optimizes the informa-

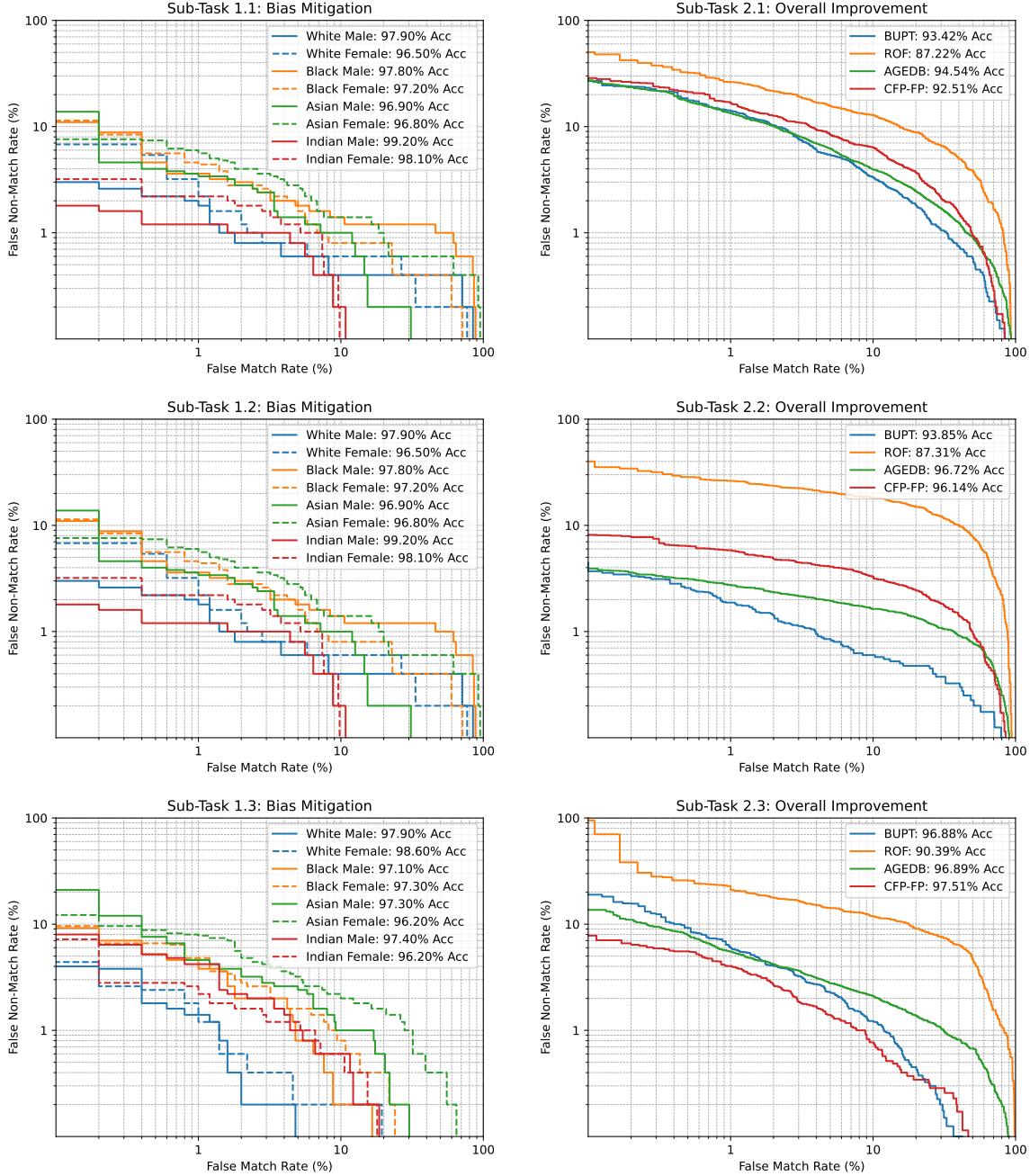


Fig. 12. DET curves of Task 1 (left) and Task 2 (right). Sub-Tasks 1.1 (left-top), 1.2 (left-middle), and 1.3 (left-bottom), including the results achieved for the top-1 team (i.e., ID R&D, ID R&D, ADMIS, respectively) in each demographic group. Sub-Tasks 2.1 (right-top), 2.2 (right-middle), and 2.3 (right-bottom), including the results achieved for the top-1 team (i.e., OPDAI, Idiap-SynthDistill, K-IBS-DS, respectively) in each demographic group.

TABLE X  
DESCRIPTION OF THE BEST RESULTS ACHIEVED IN THE 1<sup>ST</sup> AND 2<sup>ND</sup> FRCSYN-ONGOING. SUB-TASKS 1.2 AND 2.2 OF THE 2<sup>ND</sup> FRCSYN-ONGOING ARE NOT INCLUDED IN THE TABLE AS THEY ARE NOVEL SUB-TASKS ONLY AVAILABLE IN THE 2<sup>ND</sup> EDITION.

| Sub-Task 1.1: Bias Mitigation  |        |              |                                |              |              | Sub-Task 2.1: Overall Improvement |         |              |                                |              |         |
|--------------------------------|--------|--------------|--------------------------------|--------------|--------------|-----------------------------------|---------|--------------|--------------------------------|--------------|---------|
| 1 <sup>st</sup> FRCSyn-onGoing |        |              | 2 <sup>nd</sup> FRCSyn-onGoing |              |              | 1 <sup>st</sup> FRCSyn-onGoing    |         |              | 2 <sup>nd</sup> FRCSyn-onGoing |              |         |
| Team                           | TO [%] | GAP [%]      | Team                           | TO [%]       | GAP [%]      | Team                              | AVG [%] | GAP [%]      | Team                           | AVG [%]      | GAP [%] |
| LENS                           | 92.25  | -0.74        | ID R&D                         | <b>96.73</b> | <b>-5.31</b> | BOVIFOCR                          | 90.50   | <b>2.66</b>  | OPDAI                          | <b>91.93</b> | 3.09    |
| Sub-Task 1.3: Bias Mitigation  |        |              |                                |              |              |                                   |         |              |                                |              |         |
| 1 <sup>st</sup> FRCSyn-onGoing |        |              | 2 <sup>nd</sup> FRCSyn-onGoing |              |              | Sub-Task 2.3: Overall Improvement |         |              | 2 <sup>nd</sup> FRCSyn-onGoing |              |         |
| Team                           | TO [%] | GAP [%]      | Team                           | TO [%]       | GAP [%]      | Team                              | AVG [%] | GAP [%]      | Team                           | AVG [%]      | GAP [%] |
| CBSR                           | 95.25  | <b>-2.10</b> | ADMIS                          | <b>96.50</b> | -1.33        | BOVIFOCR                          | 94.95   | <b>-3.69</b> | K-IBS-DS                       | <b>95.42</b> | -2.15   |

tion flow through the network, allowing for the training of extremely deep architectures without increasing model complexity. Finally, the selection of the loss functions was also similar among the teams, with AdaFace [7] and ArcFace [6] the prevalent choices. Both losses use angular margin-based loss functions to improve facial feature discrimination, but while in ArcFace the margin is fixed for all samples, AdaFace adapts the margin dynamically based on the quality of each image. Nevertheless, there were exceptions such as for the ID R&D team that used the recent UniFace [65], or the UNICA team that considered CosFace [68].

Aditinally, we compare the results achieved in the 2<sup>nd</sup> FRCSyn-onGoing with the results of the 1<sup>st</sup> edition [32]. Table X shows the best results achieved in the 1<sup>st</sup> and 2<sup>nd</sup> editions of the challenge, including also the GAP values. It is important to remark that Sub-Tasks 1.2 and 2.2 of the 2<sup>nd</sup> FRCSyn-onGoing are not included in the analysis as they are novel sub-tasks only available in the 2<sup>nd</sup> edition. Notably, two observations can be made: *i*) the main metric for ranking teams (*i.e.*, TO and AVG) shows improvements across all cases in this 2<sup>nd</sup> edition for both Task 1 (*e.g.*, 96.73% vs. 92.25% TO in Sub-Task 1.1 and 96.50% vs. 95.25% TO in Sub-Task 1.3) and Task 2 (*e.g.*, 91.93% vs. 90.50% AVG in Sub-Task 2.1 and 95.42% vs. 94.95% AVG in Sub-Task 2.3), and *ii*) in terms of the GAP

value, the FR models of this 2<sup>nd</sup> edition follow a similar trend compared to the 1st edition, achieving in most sub-tasks negative GAP values, remarking the benefits for training using synthetic data. In particular, for Sub-Task 1.1, a much higher negative GAP value is observed in this 2<sup>nd</sup> edition (*i.e.*, -5.31% vs. -0.74%). This result, together with the higher TO value, seems to be due to the generation of better synthetic data by the ID R&D team, with the proposal of novel generative methods, as indicated before. In addition, several conclusions can be drawn. First, the improvement in the main metric can be associated with the freedom to select the methodology to generate the synthetic data to train the FR models, as well as the application of data cleaning and selection techniques. We observe that the increasing GAP value can be associated also with the enhancement of the FR models, due to the proposal of different architectures and loss functions, as indicated before.

Finally, we observe that when it comes to the quality of the generated synthetic data, higher quality does not imply better performance in recognition tasks. Most teams use DCFace as the main dataset for their training, even though it generates images with lower resolution and less detail from a qualitative perspective. Nevertheless, analyzing the results achieved using this database, we conclude that FR models might not require highly detailed

images to learn to match identities, at least for the FR databases considered in this challenge. This suggests that instead of focusing on realism, synthetic data for FR should generate diverse images that help models better learn class variability while reducing noise.

## VI. CONCLUSION

The 2<sup>nd</sup> FRCSyn-onGoing has presented a comprehensive exploration of the applications of synthetic data in FR, effectively addressing existing limitations in the field. In this 2<sup>nd</sup> 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 only real data. With an increased number of participants in this last edition, we have witnessed a considerable performance improvement in all sub-tasks in comparison to the 1<sup>st</sup> edition [32], [33]. 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 studies will include recent AI techniques [90], to make sure that only the databases available by the challenge are used by participants. We will also perform a more detailed analysis of the results and comparison with recent challenges in the topic, such as SDFR [91], or evaluate over a more diverse set of databases that include other FR challenges, like quality, surveillance, or large distance images [92]. Finally, we plan to focus on the explainability of these FR models and the frameworks that generate synthetic images [8]. Debiasing face recognition models using synthetic datasets is an important task and in this challenge, we found that the use of synthetic data can further increase the performance. However, the concept of “bias” itself is complex and often subjective. What constitutes a “fair” representation can vary significantly depending on cultural context, individual experiences, and even personal beliefs. Therefore, debiasing efforts should be approached with an

accurate understanding of the multifaceted nature of bias. Simply generating synthetic data to reflect a particular demographic distribution might not fully address the complexities of real-world inequalities. Rather than seeking to eliminate bias, perhaps a more productive approach is to pursue research in the direction of transparency, interpretability, and controllability. This translates to research questions that allow researchers to easily define what bias should be and allow them to fine-tune their models accordingly. Ultimately, the goal should be to develop face recognition systems that are not only accurate but also fair and ethical.

## ACKNOWLEDGEMENTS

This study is supported by INTER-ACTION (PID2021-126521OB-I00 MICINN/FEDER), Cátedra ENIA UAM-VERIDAS en IA Responsable (NextGenerationEU PRTR TSI-100927-2023-2), R&D Agreement DGGC/UAM/FUAM for Biometrics and Cybersecurity, and PowerAI+ (SI4/PJI/2024-00062, funded by Comunidad de Madrid through the grant agreement for the promotion of research and technology transfer at UAM). 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] M. Wang and W. Deng, “Deep Face Recognition: A Survey,” *Neurocomputing*, vol. 429, pp. 215–244, 2021. [1](#)
- [2] H. Du, H. Shi, D. Zeng, X.-P. Zhang, and T. Mei, “The Elements of End-to-end Deep Face Recognition: A Survey of Recent Advances,” *ACM Comput. Surv.*, vol. 54, no. 10s, pp. 1–42, 2022. [1](#)
- [3] C. Bisogni, A. Castiglione, S. Hossain, F. Narducci, and S. Umer, “Impact of Deep Learning Approaches on Facial Expression Recognition in Healthcare Industries,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5619–5627, 2022. [1](#)
- [4] L. F. Gomez, A. Morales, J. R. Orozco-Arroyave, R. Daza, and J. 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](#)

- [5] R. Daza, L. F. Gomez, A. Morales, J. Fierrez, R. Tolosana, R. Cobos, and J. Ortega-Garcia, "MATT: Multimodal Attention Level Estimation for e-learning Platforms," in *Proc. AAAI Workshop on Artificial Intelligence for Education*, 2023. 1
- [6] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2, 8, 9, 10, 18, 25
- [7] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality Adaptive Margin for Face Recognition," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2, 8, 9, 11, 12, 14, 16, 17, 25
- [8] I. Deandres-Tame, R. Tolosana, R. Vera-Rodriguez, A. Morales, J. Fierrez, and J. Ortega-Garcia, "How Good Is ChatGPT at Face Biometrics? A First Look Into Recognition, Soft Biometrics, and Explainability," *IEEE Access*, vol. 12, pp. 34 390–34 401, 2024. 2, 26
- [9] C. R. Crum, P. Tinsley, A. Boyd, J. Piland, C. Sweet, T. Kelley, K. Bowyer, and A. Czajka, "Explain To Me: Salience-Based Explainability for Synthetic Face Detection Models," *IEEE Transactions on Artificial Intelligence*, pp. 1–12, 2023. 2
- [10] Y. Shen, C. Yang, X. Tang, and B. Zhou, "InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 2004–2018, 2022. 2
- [11] 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*, vol. 3, no. 1, pp. 16–30, 2021. 2
- [12] 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," in *Proc. IEEE International Joint Conference on Biometrics*, 2023. 2
- [13] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, "SensitiveNets: Learning Agnostic Representations with Application to Face Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 2158–2164, 2021. 2
- [14] P. Melzi, H. O. Shahreza, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, J. Fierrez, S. Marcel, and C. 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
- [15] P. Melzi, C. Rathgeb, R. Tolosana, R. Vera, and C. Busch, "An Overview of Privacy-Enhancing Technologies in Biometric Recognition," *ACM Computing Surveys*, 2024. 2
- [16] J. Zhao, S. Yan, and J. Feng, "Towards Age-Invariant Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 474–487, 2022. 2
- [17] R. Valle, J. M. Buenaposada, and L. Baumela, "Multi-Task Head Pose Estimation in-the-Wild," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2874–2881, 2021. 2
- [18] L. Tran, X. Yin, and X. Liu, "Representation Learning by Rotating Your Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 3007–3021, 2019. 2
- [19] S. P. Mudunuri and S. Biswas, "Low Resolution Face Recognition Across Variations in Pose and Illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 5, pp. 1034–1040, 2016. 2
- [20] H. Qiu, D. Gong, Z. Li, W. Liu, and D. Tao, "End2End Occluded Face Recognition by Masking Corrupted Features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6939–6952, 2022. 2
- [21] 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," in *Proc. IEEE/CVF International Conference on Computer Vision Workshops*, 2023. 2, 3, 4, 9, 12, 13, 17, 23
- [22] M. Kim, F. Liu, A. Jain, and X. 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, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 23
- [23] F. Boutros, J. H. Grebe, A. Kuijper, and N. Damer, "IDiff-Face: Synthetic-based Face Recognition through Fuzzy Identity-Conditioned Diffusion Model," in *Proc. IEEE/CVF International Conference on Computer Vision*, 2023. 2, 4, 8, 9, 10, 15, 16, 17, 23
- [24] F. Boutros, V. Struc, J. Fierrez, and N. Damer, "Synthetic Data for Face Recognition: Current State and Future Prospects," *Image and Vision Computing*, vol. 135, p. 104688, 2023. 2
- [25] I. Joshi, M. Grimmer, C. Rathgeb, C. Busch, F. Bremond, and A. Dantcheva, "Synthetic Data in Human Analysis: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 7, pp. 4957–4976, 2024. 2
- [26] Y. Sun, M. Zhang, Z. Sun, and T. Tan, "Demographic Analysis from Biometric Data: Achievements, Challenges, and New Frontiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 332–351, 2018. 2
- [27] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Proc. Advances in Neural Information Processing Systems*, 2014. 2
- [28] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," in *Proc. Advances in Neural Information Processing Systems*, 2020. 2, 9
- [29] 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 *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023. 2, 4, 12, 13, 23
- [30] C. Zhang, X. Chen, S. Chai, C. H. Wu, D. Lagun, T. Beeler, and F. De la Torre, "ITI-GEN: Inclusive Text-to-Image Generation," in *Proc. IEEE/CVF International Conference on Computer Vision*, 2023. 2, 4, 5
- [31] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, and D. Tao, "SynFace: Face Recognition with Synthetic Data," in *Proc. IEEE/CVF International Conference on Computer Vision*, 2021. 2, 4, 5

- [32] P. Melzi, R. Tolosana, R. Vera-Rodriguez *et al.*, “FRCSSyn-onGoing: Benchmarking and Comprehensive Evaluation of Real and Synthetic Data to Improve Face Recognition Systems,” *Information Fusion*, vol. 107, p. 102322, 2024. [2](#), [3](#), [5](#), [6](#), [25](#), [26](#)
- [33] ———, “FRCSSyn 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](#), [5](#), [6](#), [26](#)
- [34] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016. [2](#), [8](#), [9](#), [15](#), [16](#), [18](#), [23](#)
- [35] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, “Learning Transferable Visual Models From Natural Language Supervision,” in *Proc. 38th International Conference on Machine Learning*, 2021. [2](#), [5](#)
- [36] M. Oquab, T. Dariseti, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby *et al.*, “DinoV2: Learning Robust Visual Features Without Supervision,” *arXiv preprint arXiv:2304.07193*, 2023. [2](#)
- [37] Y. Wang, Y. Wu, S. Tang, W. He, X. Guo, F. Zhu, L. Bai, R. Zhao, J. Wu, T. He *et al.*, “Hulk: A Universal Knowledge Translator For Human-Centric Tasks,” *arXiv preprint arXiv:2312.01697*, 2023. [3](#)
- [38] S. Tang, C. Chen, Q. Xie, M. Chen, Y. Wang, Y. Ci, L. Bai, F. Zhu, H. Yang, L. Yi, R. Zhao, and W. Ouyang, “HumanBench: Towards General Human-Centric Perception With Projector Assisted Pretraining,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. [3](#)
- [39] R. Khirodkar, T. Bagautdinov, J. Martinez, S. Zhaoen, A. James, P. Selednik, S. Anderson, and S. Saito, “Sapiens: Foundation for Human Vision Models,” in *Proc. European Conference on Computer Vision*, 2025. [3](#)
- [40] B. Bozorgtabar, M. S. Rad, H. K. Ekenel, and J.-P. Thiran, “Using Photorealistic Face Synthesis and Domain Adaptation to Improve Facial Expression Analysis,” in *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, 2019. [3](#)
- [41] R. Tolosana, P. Delgado-Santos, A. Perez-Uribe, R. Vera-Rodriguez, J. Fierrez, and A. Morales, “DeepWriteSYN: On-Line Handwriting Synthesis via Deep Short-Term Representations,” *Proc. AAAI Conference on Artificial Intelligence*, vol. 35, pp. 600–608, 2021. [3](#)
- [42] H. Hwang, C. Jang, G. Park, J. Cho, and I.-J. Kim, “ElderSim: A Synthetic Data Generation Platform for Human Action Recognition in Eldercare Applications,” *IEEE Access*, vol. 11, pp. 9279–9294, 2023. [3](#)
- [43] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid, “Learning From Synthetic Humans,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017. [3](#)
- [44] I. DeAndres-Tame, R. Tolosana, P. Melzi, R. Vera-Rodriguez, M. Kim, C. Rathgeb, X. Liu, A. Morales, J. Fierrez, J. Ortega-Garcia, Z. Zhong, Y. Huang, Y. Mi, S. Ding, S. Zhou, S. He, L. Fu, H. Cong, R. Zhang, Z. Xiao, E. Smirnov, A. Pimenov, A. Grigorev, D. Timoshenko, K. M. Asfaw, C. Y. Low, H. Liu, C. Wang, Q. Zuo, Z. He, H. O. Shahreza, A. George, A. Unnervik, P. Rahimi, S. Marcel, P. C. Neto, M. Huber, J. N. Kolf, N. Damer, F. Boutros, J. S. Cardoso, A. F. Sequeira, A. Atzori, G. Fenu, M. Marras, V. Štruc, J. Yu, Z. Li, J. Li, W. Zhao, Z. Lei, X. Zhu, X.-Y. Zhang, B. Biesseck, P. Vidal, L. Coelho, R. Granada, and D. Menotti, “Second Edition FRCSSyn Challenge at CVPR 2024: Face Recognition Challenge in the Era of Synthetic Data,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. [3](#)
- [45] M. Kansy, A. Raël, G. Mignone, J. Naruniec, C. Schroers, M. Gross, and R. M. Weber, “Controllable Inversion of Black-Box Face Recognition Models via Diffusion,” in *Proc. IEEE/CVF International Conference on Computer Vision Workshops*, 2023. [4](#)
- [46] F. Boutros, M. Huber, P. Siebke, T. Rieber, and N. Damer, “SFace: Privacy-friendly and Accurate Face Recognition using Synthetic Data,” in *Proc. IEEE International Joint Conference on Biometrics*, 2022. [4](#)
- [47] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, “Alias-Free Generative Adversarial Networks,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 852–863, 2021. [4](#)
- [48] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. [4](#)
- [49] 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 *Proc. IEEE/CVF International Conference on Computer Vision*, 2021. [4](#)
- [50] Y. Deng, J. Yang, D. Chen, F. Wen, and X. Tong, “Disentangled and Controllable Face Image Generation via 3D Imitative-Contrastive Learning,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. [5](#)
- [51] D. Yi, Z. Lei, S. Liao, and S. Z. Li, “Learning Face Representation from Scratch,” *arXiv preprint arXiv:1411.7923*, 2014. [5](#), [6](#), [7](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#)
- [52] M. Wang and W. Deng, “Mitigating Bias in Face Recognition Using Skewness-Aware Reinforcement Learning,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. [5](#), [6](#), [7](#), [17](#), [20](#)
- [53] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, “AgeDB: The First Manually Collected, In-The-Wild Age Database,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2017. [5](#), [6](#), [7](#), [17](#), [20](#)
- [54] S. Sengupta, J.-C. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, “Frontal to Profile Face Verification in the Wild,” in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, 2016. [5](#), [6](#), [7](#), [17](#), [20](#)
- [55] M. E. Erakan, U. Demir, and H. K. Ekenel, “On Recognizing Occluded Faces in the Wild,” in *Proc. International Conference of the Biometrics Special Interest Group*, 2021. [5](#), [6](#), [7](#), [17](#), [20](#)
- [56] K. Karkkainen and J. Joo, “FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias

- Measurement and Mitigation,” in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021. 5
- [57] I. C. Duta, L. Liu, F. Zhu, and L. Shao, “Improved Residual Networks for Image and Video Recognition,” in *Proc. 25th International Conference on Pattern Recognition*, 2021. 8, 9, 10, 11, 12, 13, 14, 17, 23
- [58] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-Resolution Image Synthesis With Latent Diffusion Models,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 9, 10
- [59] J. Song, C. Meng, and S. Ermon, “Denoising Diffusion Implicit Models,” in *Proc. International Conference on Learning Representations*, 2021. 9, 10
- [60] Z. Li, M. Cao, X. Wang, Z. Qi, M.-M. Cheng, and Y. Shan, “PhotoMaker: Customizing Realistic Human Photos via Stacked ID Embedding,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 9, 11
- [61] K. Crowson, S. A. Baumann, A. Birch, T. M. Abraham, D. Z. Kaplan, and E. Shippole, “Scalable High-Resolution Pixel-Space Image Synthesis with Hourglass Diffusion Transformers,” *arXiv preprint arXiv:2401.11605*, 2024. 9, 11, 23
- [62] A. van den Oord, O. Vinyals, and k. kavukcuoglu, “Neural Discrete Representation Learning,” in *Proc. Advances in Neural Information Processing Systems*, 2017. 9, 11
- [63] S. Walton, A. Hassani, X. Xu, Z. Wang, and H. Shi, “StyleNAT: Giving Each Head a New Perspective,” *arXiv preprint arXiv:2211.05770*, 2022. 9, 11, 23
- [64] N. Garaev, E. Smirnov, V. Galyuk, and E. Lukyanets, “FaceMix: Transferring Local Regions for Data Augmentation in Face Recognition,” in *Proc. Neural Information Processing*, 2023. 9, 12
- [65] J. Zhou, X. Jia, Q. Li, L. Shen, and J. Duan, “UniFace: Unified Cross-Entropy Loss for Deep Face Recognition,” in *Proc. IEEE/CVF International Conference on Computer Vision*, 2023. 9, 12, 25
- [66] C. Y. Low, J. C. L. Chai, J. Park, K. Ann, and M. Cha, “SlackedFace: Learning a Slacked Margin for Low-Resolution Face Recognition,” in *Proc. 34th British Machine Vision Conference*, 2023. 9, 12
- [67] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 9, 12, 13
- [68] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, “CosFace: Large Margin Cosine Loss for Deep Face Recognition,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 9, 14, 15, 25
- [69] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” in *Proc. IEEE/CVF conference on Computer Vision and Pattern Recognition*, 2019. 9, 10, 14, 23
- [70] H. Otroshi Shahreza, A. George, and S. Marcel, “Knowledge Distillation for Face Recognition Using Synthetic Data With Dynamic Latent Sampling,” *IEEE Access*, vol. 12, pp. 187 800–187 812, 2024. 9, 14, 23
- [71] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, “ElasticFace: Elastic Margin Loss for Deep Face Recognition,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022. 9, 10, 15
- [72] F. Boutros, M. Klemt, M. Fang, A. Kuijper, and N. Damer, “ExFaceGAN: Exploring Identity Directions in GAN’s Learned Latent Space for Synthetic Identity Generation,” in *IEEE International Joint Conference on Biometrics*, 2023. 9, 15
- [73] X. An, X. Zhu, Y. Gao, Y. Xiao, Y. Zhao, Z. Feng, L. Wu, B. Qin, M. Zhang, D. Zhang, and Y. Fu, “Partial FC: Training 10 Million Identities on a Single Machine,” in *Proc. IEEE/CVF International Conference on Computer Vision Workshops*, 2021. 11
- [74] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du, and J. 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. 11, 14
- [75] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Proc. Medical image computing and computer-assisted intervention*, 2015. 11
- [76] A. Tong, K. FATRAS, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, G. Wolf, and Y. Bengio, “Improving and Generalizing Flow-Based Generative Models with Minibatch Optimal Transport,” *Transactions on Machine Learning Research*, 2024. 11
- [77] J. Ho and T. Salimans, “Classifier-Free Diffusion Guidance,” *arXiv preprint arXiv:2207.12598*, 2022. 11
- [78] E. Smirnov, N. Garaev, V. Galyuk, and E. Lukyanets, “Prototype Memory for Large-Scale Face Representation Learning,” *IEEE Access*, vol. 10, pp. 12 031–12 046, 2022. 11
- [79] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, “RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 13
- [80] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and Improving the Image Quality of StyleGAN,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 14
- [81] P. C. Neto, E. Caldeira, J. S. Cardoso, and A. 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. 15
- [82] P. C. Neto, F. Boutros, J. R. Pinto, N. Damer, A. F. Sequeira, J. S. Cardoso, M. Bengherabi, A. Bousnat, S. Boucheta, N. Hebbadj, M. E. Erakin, U. Demir, H. K. Ekenel, P. B. De Queiroz Vidal, and D. Menotti, “OCFR 2022: Competition on Occluded Face Recognition from Synthetically Generated Structure-Aware Occlusions,” in *Proc. IEEE International Joint Conference on Biometrics*, 2022. 15
- [83] P. C. Neto, R. M. Mamede, C. Albuquerque, T. Gonçalves, and A. F. Sequeira, “Massively Annotated Datasets for Assessment of Synthetic and Real Data in Face Recognition,” *arXiv preprint arXiv:2404.15234*, 2024. 15
- [84] A. Atzori, F. Boutros, N. Damer, G. Fenu, and M. 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. 15, 16

- [85] A. Shoshan, N. Bhonker, I. Kviatkovsky, and G. Medioni, “GAN-Control: Explicitly Controllable GANs,” in *Proc. IEEE/CVF International Conference on Computer Vision*, 2021. [15](#)
- [86] F. Boutros, M. Klemm, M. Fang, A. Kuijper, and N. Damer, “Unsupervised Face Recognition using Unlabeled Synthetic Data,” in *Proc. 17th International Conference on Automatic Face and Gesture Recognition*, 2023. [16](#)
- [87] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016. [16](#)
- [88] J. Wang, Y. Liu, Y. Hu, H. Shi, and T. Mei, “FaceX-Zoo: A PyTorch Toolbox for Face Recognition,” in *Proc. 29th ACM International Conference on Multimedia*, 2021. [17](#)
- [89] M. L. Ngan, P. J. Grother, and K. K. Hanaoka, “OnGoing Face Recognition Vendor Test (FRVT) Part 6b: Face Recognition Accuracy with Face Masks Using Post-Covid-19 Algorithms,” 2020. [17](#)
- [90] D. DeAlcala, A. Morales, G. Mancera, J. Fierrez, R. Tolosana, and J. Ortega-Garcia, “Is my Data in your AI Model? Membership Inference Test with Application to Face Images,” *arXiv preprint arXiv:2402.09225*, 2024. [26](#)
- [91] H. O. Shahreza, C. Ecabert, A. George, A. Unnervik, S. Marcel, N. D. Domenico, G. Borghi, D. Maltoni, F. Boutros, J. Vogel, N. Damer, Ángela Sánchez-Pérez, EnriqueMas-Candela, J. Calvo-Zaragoza, B. Biesseck, P. Vidal, R. Granada, D. Menotti, I. DeAndres-Tame, S. M. L. Cava, S. Concas, P. Melzi, R. Tolosana, R. Vera-Rodriguez, G. Perelli, G. Orrù, G. L. Marcialis, and J. Fierrez, “SDFR: Synthetic Data for Face Recognition Competition,” in *Proc. 18th IEEE International Conference on Automatic Face and Gesture Recognition*, 2024. [26](#)
- [92] F. Liu, R. Ashbaugh, N. Chmitt, N. Hassan, A. Hassani, A. Jaiswal, M. Kim, Z. Mao, C. Perry, Z. Ren, Y. Su, P. Varghaei, K. Wang, X. Zhang, S. Chan, A. Ross, H. Shi, Z. Wang, A. Jain, and X. Liu, “FarSight: A Physics-Driven Whole-Body Biometric System at Large Distance and Altitude,” in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024. [26](#)

## VII. BIOGRAPHY SECTION

**Ivan DeAndres-Tame** is a Ph.D. at Universidad Autónoma de Madrid at Madrid, Spain. Contact him at ivan.deandres@uam.es.

**Ruben Tolosana** is an associate professor at Universidad Autónoma de Madrid at Madrid, Spain. Contact him at ruben.tolosana@uam.es.

**Pietro Melzi** got his Ph.D. at Universidad Autónoma de Madrid at Madrid, Spain. Contact him at pietro.melzi@uam.es.

**Ruben Vera-Rodriguez** is an associate professor at Universidad Autónoma de Madrid at Madrid, Spain. Contact him at ruben.vera@uam.es.

**Minchul Kim** is a Ph.D. at Michigan State University, at Michigan, USA. Contact him at kimminc2@msu.edu.

**Christian Rathgeb** is a full professor at Hochschule Darmstadt at Darmstadt, Germany. Contact him at christian.rathgeb@h-da.de.

**Xiaoming Liu** is a full professor at Michigan State University, at Michigan, USA. Contact him at liuxm@cse.msu.edu.

**Luis F.Gomez** got his Ph.D. at Universidad Autónoma de Madrid at Madrid, Spain. Contact him at luisf.gomez@uam.es.

**Aythami Morales** is an associate professor at Universidad Autónoma de Madrid at Madrid, Spain. Contact him at aythami.morales@uam.es.

**Julian Fierrez** is a full professor at Universidad Autónoma de Madrid at Madrid, Spain. Contact him at rjulian.fierrez@uam.es.

**Javier Ortega-Garcia** is a full professor at Universidad Autónoma de Madrid at Madrid, Spain. Contact him at javier.ortega@uam.es.

**Zhizhou Zhong** is an M.S. student at Fudan University at Shanghai, China. Contact him at zzzhong22@m.fudan.edu.cn.

**Yuge Huang** is a senior researcher at Tencent YouTu Lab at Shanghai, China. Contact him at yugehuang@tencent.com.

**Yuxi Mi** is a Ph.D. at Fudan University at Shanghai, China. Contact him at yxmi20@fudan.edu.cn.

**Shouhong Ding** is a principal researcher at Tencent YouTu Lab at Shanghai, China. Contact him at ericshding@tencent.com.

**Shuigeng Zhou** is a full professor at Fudan University at Shanghai, China. Contact him at sgzhou@fudan.edu.cn.

**Shuai He** is a Software Engineer at the Interactive Entertainment Group of Netease Inc at Guangzhou, China. Contact him at heshuai03@corp.netease.com.

**Lingzhi Fu** is a Software Engineer at the Interactive Entertainment Group of Netease Inc at Guangzhou, China. Contact him at fulingzhi@corp.netease.com.

**Heng Cong** is a Software Engineer at the Interactive Entertainment Group of Netease Inc at Guangzhou, China. Contact him at congheng@corp.netease.com.

**Rongyu Zhang** is a Software Engineer at the Interactive Entertainment Group of Netease Inc at Guangzhou, China. Contact her at zhangrongyu@corp.netease.com.

**Zhihong Xiao** is a Software Engineer at the Interactive Entertainment Group of Netease Inc at Guangzhou, China. Contact him at xiaozhihong@corp.netease.com.

**Evgeny Smirnov** is a ML engineer at ID R&D Inc. in Barcelona, Spain. Contact him at evgeny.smirnov@idrnd.net.

**Anton Pimenov** is a ML engineer at ID R&D Inc. in Barcelona, Spain. Contact him at pimenov@idrnd.net.

**Aleksei Grigorev** is a ML Engineer at ID R&D Inc. in Barcelona, Spain. Contact him at grigoriev@idrnd.net.

**Denis Timoshenko** is a ML engineer at ID R&D Inc. in New York, USA. Contact him at timoshenko@idrnd.net.

**Kaleb Mesfin Asfaw** is an undergraduate student researcher with the Korea Advanced Institute of Science and Technology at Daejeon, South Korea. Contact him at kalebmes06@kaist.ac.kr.

**Cheng Yaw Low** is a research associate with the Institute for Basic Science at Daejeon, South Korea. Contact him at chengyawlow@ibs.re.kr.

**Hao Liu** is a M.S at China Telecom AI, China. Contact him at liuh9@chinatelecom.cn

**Chuyi Wang** is a M.S at China Telecom AI, China. Contact him at wangcy30@chinatelecom.cn.

**Qing Zuo** is a Ph.D. at China Telecom AI, China. Contact him at zuoq2@chinatelecom.cn

**Zhixiang He** is a Ph.D. at China Telecom AI, China. Contact him at hezx3@chinatelecom.cn.

**Hatef Otroshi Shahreza** is a Ph.D. student at EPFL at Lausanne, Switzerland and a Research Assistant at Idiap Research Institute at Martigny, Switzerland. Contact him at hatef.otroshi@idiap.ch.

**Anjith George** is a Research Associate at Idiap Research Institute at Martigny, Switzerland. Contact him at anjith.george@idiap.ch.

**Alexander Unnervik** is a Ph.D. student at EPFL at Lausanne, Switzerland and a Research Assistant at Idiap Research Institute at Martigny, Switzerland. Contact him at alex.unnervik@idiap.ch.

**Parsa Rahimi** is a Ph.D. student at EPFL at Lausanne, Switzerland and a Research Assistant at Idiap Research Institute at Martigny, Switzerland. Contact him at parsa.rahami@idiap.ch.

**Sébastien Marcel** is a Senior Researcher and the Head of Biometrics Security and Privacy group at Idiap Research Institute at Martigny, Switzerland and a Professor at UNIL at Lausanne, Switzerland. Contact him at sebastien.marcel@idiap.ch.

**Pedro C. Neto** is a research assistant at INESC TEC at Porto, Portugal. Contact him at Pedro.d.carneiro@inesctec.pt

**Marco Huber** is a research associate at Fraunhofer IGD at Darmstadt, Germany. Contact him at marco.huber@igd.fraunhofer.de

**Jan Niklas Kolf** is a research associate at Fraunhofer IGD at Darmstadt, Germany. Contact him at jan.niklas.kolf@igd.fraunhofer.de

**Naser Damer** is a research associate at Fraunhofer IGD at Darmstadt, Germany. Contact him at naser.damer@igd.fraunhofer.de

**Fadi Boutros** is a research associate at Fraunhofer IGD at Darmstadt, Germany. Contact him at fadi.boutros@igd.fraunhofer.de

**Jaime S. Cardoso** is a full professor at FEUP at Porto, Portugal. Contact him at jaime.cardoso@fe.up.pt

**Ana F. Sequeira** is an assistant researcher at INESC TEC at Porto, Portugal. Contact her at ana.f.sequeira@inesctec.pt

**Andrea Atzori** is a Ph.D. student at University of Cagliari at Cagliari, Italy. Contact him at andrea.atzori@unica.it.

**Gianni Fenu** is a Ph.D. student at University of Cagliari at Cagliari, Italy. Contact him at gianni.fenu@unica.it.

**Mirko Marras** is a Ph.D. student at University of Cagliari at Cagliari, Italy. Contact him at mirko.marras@acm.org.

**Vitomir Štruc** is a Full Professors at the University of Ljubljana at Ljubljana, Slovenia. Contact him at vitomir.struc@fe.uni-lj.si

**Jiang Yu** is a engineer at Samsung Electronics (China) R&D Centre, China. Contact him at jiang0922.yu@samsung.com.

**Zhangjie Li** is a student at University of Science and Technology of China, China. Contact him at lizhangjie@mail.ustc.edu.

**Jichun Li** is a engineer at Samsung Electronics (China) R&D Centre, China. Contact him at jichun.li@samsung.com.

**Weisong Zhao** is a student at Institute of Information Engineering, Chinese Academy of Sciences at Beijing, China. Contact him at zhaoweiwang@iie.ac.cn.

**Zhen Lei** is a professor at MAIS, Institute of Automation, Chinese Academy of Sciences at Beijing, China. Contact him at zlei@nlpr.ia.ac.cn.

**Xiangyu Zhu** is an associated professor at Institute of Automation, Chinese Academy of Sciences at Beijing, China. Contact him at xiangyu.zhu@ia.ac.cn

**Xiao-Yu Zhang** is a professor at Institute of Information Engineering, Chinese Academy of Sciences at Beijing, China. Contact him at zhangxiaoyu@iie.ac.cn

**Bernardo Biesseck** is a Ph.D. student at Federal University of Paraná, at Curitiba, Brazil. Contact him at bjg-biesseck@inf.ufpr.br.

**Pedro Vidal** is an Undergraduate student at Federal University of Paraná, at Curitiba, Brazil. Contact him at pbqv20@inf.ufpr.br.

**Luiz Coelho** is an ML Engineer at Unico IDTech at Belo Horizonte, Brazil. Contact him at luiz.coelho@unicio.io.

**Roger Granada** is a Ph.D./ML Engineer at Unico IDTech at Porto Alegre, Brazil. Contact him at roger.granada@unico.io.

**David Menotti** is an Associate Professor at Federal University of Paraná, at Curitiba, Brazil. Contact him at menotti@inf.ufpr.br.