IJCB-AFMFR 2026: Competition on Adapting Foundation Models for Face Recognition Using Synthetic Training Data

Tahar Chettaoui[1,*], Guray Ozgur[1,2,*], Eduarda Caldeira[1,2,*], Arturas Nakvosas[3,+], Hatef Otroshi Shahreza[4,+] Sébastien Marcel[4,+], Rishabh Shukla[5,+], Aditya Takkar[5,+], Rushil Khullar[5,+], Lalak Yadav[5,+] Gourav Gupta[6,+], Anant Gupta[6,+], Shiqi Yu[7,*], Vitomir Struc[8,*], Naser Damer[1,2,*], Fadi Boutros[1,*]
[1]Fraunhofer Institute for Computer Graphics Research IGD, Germany
[2]Technical University of Darmstadt, Germany, [3]Vilnius University, Lithuania
[4]Idiap Research Institute, Switzerland, [5]Indian Institute of Technology Jammu, India
[6]ArogyaPandit Private Limited, India, [7]Southern University of Science and Technology, China
[8]University of Ljubljana, Slovenia

[*] Competition Organizer [+] Competition Participant
tahar.chettaoui@igd.fraunhofer.de

Abstract

This paper presents a summary of the Competition on Adapting Foundation Models for Face Recognition Using Synthetic Training Data (AFMFR), held at the 2026 International Joint Conference on Biometrics (IJCB 2026). The competition received a total of eight valid submissions from four distinct teams across two complementary tracks: a Full Data Track, in which participants adapt the CLIP ViT-L/14 foundation model using large-scale synthetic identity data, and a Limited Data Track, designed to reflect more resource-constrained adaptation regimes. All training data was generated exclusively using IDPERTURB. Submitted solutions are ranked based on verification and identification performance across a diverse suite of benchmarks, including LFW, CFP-FP, AgeDB-30, CALFW, CPLFW, IJB-B, IJB-C, and TinyFace, using the Borda count method. Fairness evaluation is additionally conducted on the RFW dataset across four demographic groups. The results demonstrate that adaptation of the CLIP foundation model with synthetic training data substantially improves over the off-the-shelf model and, in several cases, surpasses the baseline. Notably, full fine-tuning with Sub-Center ArcFace (DMSTI-Neurotechnology) leads the Full Data Track, while rank-stabilized LoRA adaptation (Idiap-BSP) proves most effective under limited-data conditions.

1. Introduction

Face recognition (FR) has achieved remarkable progress over the past decade, driven by advances in deep neural network architectures, innovations in training objectives, particularly large-margin softmax losses, and the availability of large-scale annotated face datasets [16, 22, 37]. These developments have enabled FR systems to achieve high performance on several unconstrained and challenging benchmarks and have facilitated their deployment across numerous real-world applications. Despite these advances, the continued development of FR systems remains fundamentally dependent on access to large, diverse, and accurately annotated face datasets. However, acquiring such datasets has become increasingly challenging due to legal, ethical, and privacy concerns surrounding the collection and use of biometric data [35, 17, 7]. In recent years, several widely used public datasets, including MS-Celeb-1M [18] and VGGFace2 [8], have been withdrawn or subjected to access restrictions, significantly limiting the availability of large-scale training data. These developments have accelerated research into synthetic face generation as a scalable and privacy-preserving alternative for training FR systems [26, 30, 13].

Recent advances in generative models, particularly diffusion-based methods, have substantially improved the realism, identity consistency, and diversity of synthetic face images. As a result, synthetic data has evolved from a simple data augmentation technique into a viable alternative to real training data for FR [2]. To systematically evaluate this emerging paradigm, the biometric community has organized several benchmarking initiatives dedicated to synthetic-data-driven FR. The FRCSyn challenge at WACV 2024 benchmark [27] established one of the

979-8-3195-1426-4/26/$31.00 ©2026 IEEE

first comprehensive evaluation protocols for assessing synthetic face generation methods and their effectiveness for training FR models. Later, the Synthetic Data for Face Recognition (SDFR) competition [30] compared state-of-the-art synthetic face generation pipelines under a unified evaluation framework. In these competitions, participants trained conventional FR models based on the ResNet-50 backbone from scratch using synthetic training data, enabling a fair comparison of synthetic data generation approaches. More recently, the second edition of the FRCSyn challenge at CVPR 2024 further demonstrated significant progress in synthetic face generation while revealing that a noticeable performance gap between models trained on synthetic and real data still remains [13]. Collectively, these initiatives established synthetic data as a promising direction for privacy-preserving biometrics while identifying data diversity, identity preservation, and intra-class variation as key factors governing downstream recognition performance. However, they did not investigate the emerging paradigm of adapting large pre-trained vision foundation models, which introduces fundamentally different challenges compared to training conventional FR networks (specifically the ResNet50 architecture) from scratch.

Concurrently, computer vision has undergone a fundamental transformation with the emergence of large-scale vision foundation models, such as CLIP [32] and DINOv2 [29]. Unlike conventional FR models that are trained from scratch on face datasets, these models learn rich and transferable visual representations from billions of image-text pairs or self-supervised objectives using internet-scale data. Consequently, they have become the backbone of numerous downstream vision applications through efficient fine-tuning or parameter-efficient adaptation. This paradigm has recently been extended to FR. In particular, FRoundation [10] demonstrated that adapting CLIP [32] and DINOv2 [29] using parameter-efficient fine-tuning on synthetic face datasets yields competitive recognition performance, outperforming vision transformers trained from scratch while requiring significantly fewer trainable parameters. These findings suggest that synthetic data can serve not only as a replacement for real training datasets but also as an effective means of adapting powerful pre-trained foundation models to biometric recognition tasks. Despite these promising developments, systematic and reproducible benchmarking of foundation model adaptation strategies for FR remains largely unexplored. Previous synthetic-data competitions primarily investigated the quality of synthetic images for training conventional FR models from scratch, whereas adapting foundation models introduces fundamentally different research questions concerning parameter-efficient fine-tuning, full-model adaptation, optimization strategies, and the interaction between synthetic identity diversity and pre-trained representations.

To address this gap, we introduce the IJCB-AFMFR 2026 Competition on Adapting Foundation Models for Face Recognition Using Synthetic Training Data. Unlike previous synthetic-data competitions [30, 27, 13] that focused on training conventional FR models, the proposed competition isolates the adaptation problem by fixing both the foundation backbone and the synthetic data generation pipeline, enabling a fair comparison of different adaptation strategies. Participants are required to adapt the OpenAI CLIP ViT-L/14 foundation model using synthetic face datasets generated exclusively with IDPERTURB [3], a geometry-driven framework that improves intra-class diversity through identity embedding perturbation. The competition comprises two complementary tracks. The Full Data Track provides approximately three million synthetic images covering 35,000 identities to evaluate large-scale adaptation, whereas the Limited Data Track restricts training to 250,000 images from 5,000 identities, reflecting realistic resource-constrained development scenarios.

Submitted methods are evaluated on a comprehensive set of nine FR benchmarks. Specifically, evaluation includes the common verification datasets LFW [20], CFP-FP [33], AgeDB-30 [28], CA-LFW [44], and CP-LFW [43], the challenging mixed-media benchmarks IJB-B [41] and IJB-C [25], and the low-resolution identification benchmark TinyFace [9]. In addition, demographic fairness is assessed using the RFW benchmark [39].

The competition attracted 28 registered teams from academia and industry across multiple countries, resulting in eight valid submissions from four participating teams across the two competition tracks. The submitted methods explore a diverse range of adaptation strategies, including parameter-efficient LoRA fine-tuning [19], rank-stabilized LoRA (rsLoRA) [21], and full backbone fine-tuning. Through a comprehensive analysis of recognition performance, demographic fairness, computational considerations, and adaptation strategies, this paper provides the first benchmark of synthetic-data-driven adaptation of foundation models for FR and establishes a reproducible reference for future research in this emerging area.

2. Backbone, Training Datasets, Evaluation Setup, and Participants

2.1. Backbone

CLIP Foundation Model: In this competition, participants are restricted to using the Contrastive Language–Image Pretraining (CLIP) foundation model [32], developed by OpenAI. CLIP is a multimodal model trained to learn joint representations of images and text by associating images with their corresponding textual descriptions through contrastive learning, enabling it to capture rich visual semantics grounded in natural language. The model

consists of two main components: an image encoder, which maps images into a shared embedding space, and a text encoder, which projects textual descriptions into the same space. This alignment allows the model to measure cross-modal similarity and has demonstrated strong generalization across a wide range of visual recognition tasks, making it a compelling backbone for adaptation. Depending on their approach, participants may leverage both the image and text encoders jointly, or restrict their adaptation to the image encoder alone.

Competition Backbone: Among the available CLIP variants, participants are required to use the pretrained ViT-L/14 model, which corresponds to the largest ViT architecture explored in the original CLIP paper [32]. When using the officially released pre-trained weights provided by the original authors and employing CLIP models purely as frozen feature extractors without additional fine-tuning, ViT-L/14 consistently achieved superior performance across multiple FR benchmarks compared to other CLIP variants [10]. The large-scale ViT-L/14 model contains approximately 0.3 billion parameters and also provides a variant, namely ViT-L/14@336, which was additionally pre-trained for one epoch at a higher input resolution of 336 pixels to further improve performance [36]. However, participants in the competition were restricted to using only the original ViT-L/14 variant operating at the standard 224×224 input resolution employed during the initial training phase [32]. The competition intentionally fixed the foundation model (CLIP) and backbone (ViT-L/14), as well as the synthetic dataset used for fine-tuning, to isolate the source of performance differences and ensure a fair, controlled comparison across adaptation strategies, rather than allowing variation in model capacity, pre-training data, or backbone choice to influence the results.

Baseline: As a baseline, we adopted FRoundation [10], which investigates the adaptation of CLIP [32] and DINOv2 [29] foundation models to FR via LoRA [19], a parameter-efficient fine-tuning mechanism that injects trainable components into the frozen model weights. Their study is directly relevant to our setting: they evaluate the ViT-L/14 CLIP architecture, demonstrate that fine-tuning on synthetic face data improves over both off-the-shelf foundation models and ViTs trained from scratch, and show that this advantage is most pronounced under limited data conditions. FRoundation therefore serves as a strong and principled baseline for our work, as it shares the same backbone, operates in the same low-data synthetic regime, and provides a well-validated fine-tuning protocol for FR. For the competition baseline, we follow the training protocol, hyperparameters, and fine-tuning methodology proposed in FRoundation. The CLIP ViT-L/14 backbone is fine-tuned on the provided synthetic training dataset using the same optimization settings and FR objective described in [10]. It follows the same adaptation data as the competition setup, along with all competition constraints and limitations. This baseline therefore serves as a reference implementation for assessing the effectiveness of alternative adaptation strategies and training approaches developed by participants.

2.2. Synthetic Training Dataset

The competition is organized into two complementary tracks designed to evaluate adaptation strategies under different data availability regimes. Track 1 focuses on large-scale adaptation, reflecting scenarios where abundant synthetic data can be leveraged to effectively adapt foundation models. Participants are provided with two datasets: a primary dataset containing 2.5 million images spanning 25,000 identities (100 images per identity), and a secondary dataset consisting of 500,000 images across 10,000 identities (50 images per identity). The secondary dataset provides different intra-class variation to enable more possibilities in the adaptation process. In contrast, Track 2 investigates adaptation under limited-data conditions, representing more resource-constrained settings where only a modest amount of synthetic data is available. This track provides a single dataset comprising 250,000 images distributed over 5,000 identities (50 images per identity). Across both tracks, all datasets consist exclusively of synthetic face imagery with identity annotations. Participants must perform model adaptation strictly using the datasets provided for their respective track, and the use of any external data sources is prohibited.

All training datasets provided across both tracks were generated using IDPERTURB [3], a geometry-driven synthetic face generation framework. The choice of this framework is motivated by a key limitation shared by many existing identity-conditioned diffusion models: while capable of producing photorealistic and identity-consistent face images, they often suffer from limited intra-class variation, a critical property for training robust and generalizable FR models. IDPERTURB directly addresses this shortcoming by perturbing identity embeddings within a constrained angular region on the unit hypersphere, producing a diverse set of embeddings without requiring any modification to the underlying generative model. Each perturbed embedding is then used as a conditioning vector for a pre-trained diffusion model, enabling the synthesis of visually varied yet identity-coherent face images. This makes IDPERTURB a principled and scalable solution for large-scale synthetic dataset construction, as it enhances intra-class diversity purely in the embedding space, without relying on auxiliary labels or network-level modifications. Empirically, FR models trained on datasets generated with IDPERTURB [3] have demonstrated improved performance across multiple benchmarks compared to existing synthetic data generation approaches, further motivating their adoption in

this competition. rev In addition to the SOTA performance achieved by the FR model trained on IDPERTURB [3], as stated in [3], we additionally validated our choice of IDPERTURB [3] by comparing the performances of foundation model, i.e., our CLIP baseline model, finetuned with several SOTA synthetic datasets, including IDiff-Face [5], DCFace [23], Arc2Face [31] and SFace2 [6] (Table 6).

2.3. Evaluation Benchmarks

FR evaluations: We evaluate the performance of the submitted models on several FR benchmarks. These include LFW [20], CFP-FP [33], AgeDB30 [28], CA-LFW [44], and CP-LFW [43]. We report verification accuracies (%) following the official evaluation protocols for each of these benchmarks. In addition, we evaluated on large-scale evaluation benchmarks, IJB-B [41], and IJB-C [25]. For IJB-C and IJB-B, we used the official 1:1 mixed verification protocol and reported the verification performance as true acceptance rates (TAR) at false acceptance rates (FAR) of $1e-3$, $1e-4$, and $1e-5$. These benchmarks were selected because they are commonly used to evaluate the latest advancements in FR and offer a diverse range of use cases [16, 37, 4, 12]. We also evaluate on the more challenging identification-based TinyFace [9] benchmark, which consists of unconstrained, low-resolution face images. Through this evaluation, we assess the model's robustness on a low-quality face dataset (average 20×16 pixels), highlighting its ability to generalize beyond controlled scenarios. This comprehensive setup enables us to examine the effectiveness of the submitted models across diverse conditions.

Bias evaluation: We evaluate the models on the RFW [39] dataset to assess model bias and performance across demographic groups, as it is widely used for fairness and bias evaluation [10, 40]. The RFW dataset contains four testing subsets corresponding to African, Asian, Caucasian, and Indian groups. We follow the reporting protocols and evaluation metrics associated with the evaluation datasets and recent works [10, 40, 39]. We report the results as verification accuracies in (%) on each subset and as average accuracies to evaluate general recognition performance on the benchmarks. To evaluate the bias, we report the STD between all subsets and the SER, which is given by $\frac{max_g Error_g}{min_g Error_g}$, where g represents the demographic group, as reported in [40, 38]. A higher STD value indicates more bias across demographic groups and vice versa. For SER, models with values closer to 1 are less biased.

2.4. Evaluation Criteria:

The competition is divided into two tracks, with teams allowed to submit up to two solutions per track, of which only the best-performing one is considered. Submissions are evaluated and ranked independently per track based on face verification and identification performance across multiple benchmarks. Specifically, verification accuracy is measured on CPLFW, CFP-FP, CALFW, AgeDB30, and LFW, true accept rate (TAR) at multiple false accept rate (FAR) thresholds is measured on IJB-B and IJB-C, and Rank-1 and Rank-5 identification accuracy is measured on TinyFace. For all metrics, higher values indicate better performance.

Rankings are determined using the Borda count method. For a given metric, all n submitted solutions are ranked by performance, and points are assigned such that the first-placed solution receives $n-1$ points, the second-placed $n-2$, and so on down to 0 points for the last-placed solution. The final score of each team is computed in three steps. Following common practice in FR evaluation [10, 16, 22, 37] and the protocol of previous competition on FR [30, 27, 13], we report single-run results and use an aggregated Borda count across multiple benchmarks to provide a consistent and fair ranking of all submissions. First, teams are ranked by their average accuracy across the five small verification benchmarks, and a single Borda count is assigned based on this ranking. Second, Borda counts are averaged across the different TAR thresholds of IJB-B and IJB-C separately, then averaged together. Third, Borda counts for Rank-1 and Rank-5 on TinyFace are averaged. The three resulting scores are finally averaged to produce the overall team score, on which the final ranking is based.

As a concrete example, consider $n = 4$ teams. A team finishing second on the small benchmarks receives a Borda count of 3. On IJB-C, if the same team finishes first, second, and third at TAR thresholds 10^{-3}, 10^{-4}, and 10^{-5} respectively, its IJB-C score is $(4+3+2)/3 = 3$. On TinyFace, finishing first on Rank-1 and third on Rank-5 gives $(4+2)/2 = 3$. The team's final score is then $(3+3+3)/3 = 3$.

Table 1. A summary of the valid submitted solutions, participating team members, affiliations, and type of institution. More details on the submitted algorithms are provided in Section 3.

Solution	Team Members	Affiliations	Type
ArogyaPandit	Gourav Gupta, Anant Gupta	Arogyapandit Private Limited, India	Industry
DMSTI-Neurotechnology	Arturas Nakvosas	Institute of Data Science and Digital Technologies, Vilnius University, Lithuania and Neurotechnology, Lithuania	Academia/Industry
Idiap-BSP	Hatef Otroshi Shahreza, Sébastien Marcel	Idiap Research Institute, Switzerland	Academia
STYK IITJ	Rishabh Shukla, Aditya Takkar, Rushil Khullar, Lalak Yadav	Indian Institute of Technology Jammu, India	Academia

2.5. Competition Participants

The competition was designed to attract a broad range of participants from both academia and industry, with diverse geographic and research backgrounds. To maximize outreach, the call for participation was disseminated through the International Joint Conference on Biometrics (IJCB) 2026 website, the competition website, social media platforms, and targeted mailing lists. The competition attracted 28 registered teams from both academia and industry of which 4 teams submitted 8 valid solu-

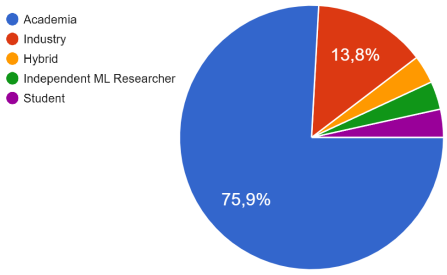

Figure 1. Distribution of registered teams by affiliation type. The competition attracted participants from academia, industry, academia–industry collaborations, independent machine learning researchers, and students, with academic teams accounting for more than 75% of all registrations.

tions. Participants represented a diverse range of backgrounds, including academic institutions, industry organizations, academia–industry collaborations, independent machine learning researchers, and students. As shown in Figure 1, more than 75% of the registered teams were affiliated with academia. The participants represented a diverse international community, spanning multiple countries across Asia, Africa, Europe, North and South America. Participating organizations included research universities, independent research institutes, and an industrial research startup, reflecting the broad interest in foundation models for FR. Each team was allowed to submit up to two solutions per track, resulting in a total of eight valid submissions across the two competition tracks from four different teams. A summary of the participating teams is provided in Table 1. Although the competition attracted 28 registered teams from academia and industry, only four teams ultimately submitted valid solutions. As in many research competitions, registration reflected initial interest, whereas completing a submission required substantial implementation and computational effort. Participants were required to adapt a large CLIP ViT-L/14 foundation model using large-scale synthetic datasets, comply with the prescribed competition constraints, and submit the final model in the required ONNX format. To facilitate participation, the organizers provided continuous support through the competition website [1], GitHub repository [2], and email communication. No organizational or technical issues preventing submission were reported to the organizers during the competition.

2.6. Submission and Evaluation Process

To participate in the competition, teams were first required to submit a registration request including their team name and institutional affiliation, upon which a link to the training data, hosted on a cloud service, was shared via email. Regarding model submission, participants were required to provide their final model in ONNX format, for which official export code was provided through the competition's GitHub repository. The submitted model must strictly adhere to the CLIP ViT-L/14 architecture, and the use of any external models beyond the provided CLIP image and text encoders is explicitly prohibited. This restriction notably excludes the use of pretrained FR models, for instance for knowledge distillation purposes. Additionally, no external setup, installation, or internet access is permitted at runtime. To ensure the integrity and fairness of the evaluation, participating teams were given no access to, or knowledge of, the evaluation benchmark at any point during the competition, including any information regarding the synthetic face generation framework used to produce the training datasets for both tracks.

3. Submitted Solutions

A total of 28 teams registered for the competition, reflecting broad interest in adapting foundation models for FR using synthetic training data. To encourage methodological exploration, each team was permitted to submit up to two adapted models per track. In the final evaluation phase, four distinct teams successfully submitted valid solutions for both competition tracks. Solution names, team members, affiliations, and type of the institution, i.e., academic, industry, or mixed, are summarized in Table 1. In the following, we provide a brief description of the valid submitted solutions:

- **ArogyaPandit:** The proposed approach adapts the OpenAI CLIP ViT-L/14 model [32] for FR through LoRA-based [19] fine-tuning of the visual encoder. Specifically, LoRA modules are applied to the Q, V attention projection layers and optimized using a temporary ArcFace identity classification head trained on the official synthetic identity datasets. For the limited-data track, an additional CLIP embedding-preservation loss is incorporated to improve feature stability under constrained training conditions. During inference preparation, the learned LoRA weights are merged into the original CLIP visual encoder, while the temporary ArcFace classifier is discarded, resulting in an ONNX model that preserves the original CLIP ViT-L/14 visual architecture.
- **DMSTI-Neurotechnology:** The proposed approach fine-tunes a CLIP ViT-L/14 backbone [32] for face verification. The original contrastive head is replaced with a Sub-Center ArcFace [14] classification head in Track 1, configured with $K = 5$, $s = 64$, and $m = 0.4$. For Track 2, a standard ArcFace [15] head with $K = 1$ is employed together with an exponential moving average (EMA) mechanism using a decay factor of 0.99.

[1] https://sites.google.com/view/ijcb-afmfr-2026/home

[2] https://github.com/TaharChettaoui/IJCB-AFMFR-2026

Table 2. Summary of the submitted approaches training details across the Full Data and Limited Data tracks. The table reports the adopted adaptation strategy (e.g., LoRA, rsLoRA, full fine-tuning), the set of adapted transformer components, rank and configuration details where applicable, and the employed loss functions. All methods are built upon the pre-trained CLIP ViT-L/14 backbone.

Team	Track	Adaptation Method	Adapted Layers	Rank / Config	Classification Head & Loss
Baseline [10]	Full Data Limited Data	rsLoRA [21]	Q, V	r=16, α=32	CosFace (m=0.3) [37]
ArogyaPandit	Full Data Limited Data	LoRA [19]	Q, V	r=16, α=32 r=4, α=8	ArcFace (m=0.42) [15] ArcFace (m=0.35) [15] + embedding-preservation loss
DMSTI-Neurotechnology	Full Data Limited Data	Full fine-tuning Full fine-tuning + EMA	All layers	– –	Sub-Center ArcFace (K=5, m=0.4) [14] ArcFace (m=0.4) [15]
Idiap-BSP	Full Data Limited Data	rsLoRA [21]	Self-attention blocks	r=256, α=16 r=1024, α=16	CosFace (m=0.3) [37]
STYK IITJ	Full Data Limited Data	LoRA [19] rsLoRA [21]	Q, V, output proj., LayerNorm Q, K, V, output proj., LayerNorm, FFN proj.	r=16, α=32	CosFace (m=0.3) [37] Sub-Center AdaFace (K=2, m=0.4) [22, 14]

Table 3. Data pre-processing and augmentation strategies across all submitted teams. Since all methods fine-tune a CLIP ViT-L/14 backbone, all teams uniformly resize input images to 224×224 pixels and apply standard CLIP normalization. The rightmost column reflect the applied data augmentation during fine-tuning.

Team	Track	Resize & CLIP Normalization	Training Augmentation
Baseline [10]	Full Data Limited Data	✓	Horizontal flipping, and RandAugment [11]
ArogyaPandit	Full Data Limited Data	✓	Random resized crop, horizontal flip, color jitter, random grayscale, Gaussian blur, random erasing
DMSTI-Neurotechnology	Full Data Limited Data	✓	Random resized crop, horizontal flip, color jitter, random grayscale, Gaussian blur, and random erasing
Idiap-BSP	Full Data Limited Data	✓	RandAugment [11], geometric transforms (shear, translate, rotate), photometric transforms (brightness, color, contrast, sharpness), perceptual transforms (AutoContrast, equalize, grayscale), horizontal flip
STYK IITJ	Full Data Limited Data	✓	Horizontal flip, RandAugment [11] Horizontal flip, RandAugment [11], random affine, random erasing

The entire backbone is fully unfrozen and optimized using AdamW with cosine learning rate scheduling and warmup, employing a backbone learning rate of 2×10^{-5}. Training is performed in bf16 precision on 8 NVIDIA H100 GPUs with extensive data augmentation, including RandAugment, RandomErasing, and GaussianBlur. During inference, the classification head is discarded, and cosine similarity is computed directly from the ℓ_2-normalized embeddings.

- **Idiap-BSP:** The proposed method fine-tunes the image encoder of a pre-trained CLIP ViT-L/14 model [32] for FR using the rsLoRA technique [21], a widely adopted strategy for efficient foundation model adaptation [34]. rsLoRA is applied within the self-attention blocks of the vision transformer using a rank of r=256 for Track 1 and r=1024 for Track 2, with a scaling factor of α=16 and a dropout rate of 0.25. The adapted image encoder is coupled with a CosFace classification head [37] configured with s=64 and m=0.3, and the entire framework is trained end-to-end using a cross-entropy objective over margin-modified cosine logits. Optimization is performed using AdamW with cosine annealing learning rate scheduling and gradient clipping at ℓ_2 norm 5. Training is conducted for 20 epochs on the full training set of each track, using a batch size of 368 on 8 NVIDIA RTX 3090 GPUs for Track 1, and a batch size of 184 on 4 NVIDIA RTX 3090 GPUs for Track 2.

- **STYK IITJ:** The proposed approach adapts a pre-trained CLIP ViT-L/14 [32] image encoder for masked face verification using LoRA [19] while keeping the majority of the backbone parameters frozen. In Track 1, LoRA modules are injected into the query (Q), value (V), and output projection layers using a rank of $r = 16$ with α/r scaling, and optimization is performed using CosFace-based classification losses. For Track 2, the adaptation strategy is extended by additionally applying adapters to the key (K) and feed-forward network projection layers, combined with rank-stabilized LoRA (rsLoRA) [21] using $\alpha/\sqrt{r}$ scaling. This track further incorporates Sub-Center AdaFace with $K = 2$ sub-centers and layer-wise learning rate decay with $\gamma = 0.85$. Across both tracks, LayerNorm parameters are unfrozen dur-

ing training, a CLIP text-anchor regularization term is employed to preserve semantic consistency, and an SVD-based intra-modal projection method (IsoCLIP) [24] is applied during inference.

Table 7 summarizes the architectural and training choices of all submitted methods, while Table 3 reports the corresponding preprocessing and data augmentation pipelines. These summaries provide a consolidated view of the design decisions explored by participants for adapting the CLIP ViT-L/14 backbone to FR using the provided synthetic training data.

4. Results

This section presents a comprehensive evaluation of all submitted approaches against the FRoundation baseline [10]. First, we assess standard FR performance on a diverse set of benchmarks, as described in Section 2.3. Second, we conduct a demographic fairness evaluation to measure the extent to which each approach mitigates bias across ethnic groups. Results are reported separately for the Full Data and Limited Data tracks, allowing a direct assessment of how data availability influences both recognition performance and fairness. CLIP [32] is included across all evaluations as a reference point to contextualize the gains brought by fine-tuning.

4.1. FR Evaluation

To evaluate FR performance across the submitted approaches, we report verification results on LFW [20], CFP-FP [33], AgeDB-30 [28], CALFW [44], and CPLFW [43]. We also report TAR at varying FAR thresholds on the large-scale benchmarks IJB-B [41] and IJB-C [25], as well as on the more challenging TinyFace benchmark [9]. The pre-trained CLIP ViT-L/14 [32] model is included for reference, highlighting that fine-tuning is necessary to achieve competitive performance. The FRoundation [10] baseline establishes the competitive benchmark for the Full and Limited Data tracks, respectively. Based on the reported results in Table 4, we make the following observations:

- In the Full Data Track, DMSTI-Neurotechnology achieves the strongest overall performance, obtaining the best average accuracy on the small-scale benchmarks (95.51%) and consistently leading on IJB-B and IJB-C across all FAR thresholds. Notably, it substantially surpasses the FRoundation baseline, particularly under strict operating conditions (e.g., 87.42% TAR@10^{-5} on IJB-C versus 76.71% for FRoundation), demonstrating superior robustness in low-FAR regimes. Idiap-BSP follows closely with strong overall performance on the small-scale and large-scale benchmarks and achieves the best results on TinyFace (65.50% Rank-1 and 71.14% Rank-5), indicating stronger robustness to low-resolution FR. STYK IITJ remains competitive on the small benchmarks (93.91%) but exhibits a noticeable performance drop at stricter FAR thresholds, suggesting weaker generalization under challenging verification settings.

- In the Limited Data Track, Idiap-BSP emerges as the most effective and consistent approach, achieving the best average accuracy on small-scale benchmarks (94.52%) while outperforming the FRoundation baseline across all reported benchmarks. In particular, it achieves substantial gains on IJB-B and IJB-C at low FAR thresholds, highlighting strong robustness despite the reduced training data regime. DMSTI-Neurotechnology remains competitive on the small benchmarks (94.08%) but experiences a significant degradation at stricter FAR thresholds, especially on IJB-C (60.49% TAR@10^{-5} compared to 76.50% for FRoundation), revealing reduced stability under limited-data constraints.

- Beyond overall ranking, the results reveal a clear trade-off between adaptation strategies. Full fine-tuning (DMSTI-Neurotechnology) achieves the best performance in the Full Data Track, including at strict FAR thresholds, but its performance collapses under the Limited Data Track, indicating that unconstrained updates to the full backbone require sufficient data to be effective and overfit when data is scarce. LoRA-based approaches, in particular Idiap-BSP's high-rank rsLoRA configuration, are more conservative and remain the most consistent method across both tracks, indicating that constrained adaptation offers greater robustness under limited-data conditions. This trade-off echoes prior findings that full fine-tuning tends to outperform LoRA on harder, more data-intensive tasks, while LoRA acts as a stronger regularizer against overfitting when data is limited [1].

- Increasing the LoRA rank from the baseline's rank-16 to Idiap-BSP's rsLoRA (rank-256 in the Full Data Track, rank-1024 in the Limited Data Track) yields consistent gains in both tracks: average accuracy on small benchmarks improves from 93.62% to 94.64% in the Full Data Track and from 94.15% to 94.52% in the Limited Data Track, with larger gains at strict FAR thresholds (IJB-C TAR@1e-5: 76.71% to 82.46% Full Data Track, 76.50% to 81.13% Limited Data Track). This indicates that rank capacity is a key driver of adaptation quality.

- In addition to the submitted participants' solutions, we introduce a minimal adaptation baseline, in which only

Table 4. The achieved verification performances by the baseline and submitted models. The results are reported in (%) on the small benchmarks, defined in Section 2.3, and as average accuracies. On IJB-B and IJB-C, the results are reported as TAR at FAR of 1e-3, 1e-4 and 1e-5. DMSTI-Neurotechnology achieves the best overall verification performance in the Full Data Track, while Idiap-BSP demonstrates superior robustness on TinyFace and emerges as the most consistent method under limited-data constraints.

Track	Approach	LFW	CFP-FP	AgeDB30	CALFW	CPLFW	Avg.	IJB-B			IJB-C			TinyFace	
								10^{-3}	10^{-4}	10^{-5}	10^{-3}	10^{-4}	10^{-5}	Rank-1	Rank-5
-	CLIP	95.90	90.66	79.82	83.10	82.73	86.44	62.72	40.90	20.52	64.74	44.69	25.23	34.71	43.96
Full Data Track	Baseline [10]	99.17	95.94	90.68	91.82	90.47	93.62	91.18	82.57	69.25	93.26	86.39	76.71	62.79	68.96
	Projection Layer	98.40	95.46	85.28	88.48	89.22	91.37	85.85	70.69	49.01	87.26	73.39	55.83	44.64	54.94
	ArogyaPandit	95.50	92.20	79.90	81.72	82.32	86.33	62.77	40.50	19.98	64.84	44.38	24.88	34.44	43.56
	DMSTI-Neurotechnology	99.47	97.39	94.43	94.40	91.88	**95.51**	**94.19**	**89.68**	**80.52**	**95.62**	**92.48**	**87.42**	65.18	70.25
	Idiap-BSP	99.35	96.86	92.97	93.07	90.98	94.64	92.96	86.58	74.18	94.59	89.83	82.46	**65.50**	**71.14**
	STYK IITJ	99.43	96.31	90.78	92.20	90.83	93.91	92.18	83.84	70.31	94.23	88.04	78.06	61.88	67.97
Limited Data Track	Baseline [10]	99.10	96.57	91.85	92.60	90.65	94.15	91.48	83.12	67.61	93.82	87.77	76.50	61.05	67.54
	Projection Layer	98.88	95.43	85.15	89.55	89.58	91.72	87.23	70.88	43.27	88.68	74.59	53.08	43.80	53.031
	ArogyaPandit	99.32	95.10	84.88	90.10	89.73	91.83	87.26	74.15	54.40	89.84	78.91	65.65	54.94	62.55
	DMSTI-Neurotechnology	99.35	96.46	91.02	92.62	90.95	94.08	91.68	82.40	54.65	93.67	86.08	60.49	58.99	65.91
	Idiap-BSP	99.47	96.51	92.53	93.53	90.53	**94.52**	**92.56**	**86.23**	**74.72**	**94.51**	**89.79**	**81.13**	**62.90**	**68.78**
	STYK IITJ	95.52	92.20	79.88	81.82	82.35	86.35	63.33	41.09	20.15	65.34	45.03	25.25	34.44	43.62

Table 5. Evaluation on RFW reported as average recognition performance in (%), standard deviation (STD) and skewed error ratio (SER) across four different demographic groups, as defined in Section 2.3. The higher STD indicates a more biased model and the higher average (avg) indicates, in general, better recognition performance. For SER, models with values closer to 1 are less biased. DMSTI-Neurotechnology leads the Full Data Track in both accuracy and fairness, while Idiap-BSP is the top performer in the Limited Data Track.

Track	**Approach**	African	Asian	Caucasian	Indian	Avg.	STD	SER
-	CLIP [32]	70.75	69.73	79.32	68.98	72.19	4.80	1.49
Full Data Track	Baseline [10]	85.23	85.42	91.52	86.12	87.07	2.99	1.74
	ArogyaPandit	74.03	72.15	82.60	73.15	75.48	4.81	**1.60**
	DMSTI-Neurotechnology	90.82	89.78	95.08	91.13	**91.70**	**2.33**	2.08
	Idiap-BSP	87.38	87.97	93.25	88.97	89.39	2.65	1.87
	STYK IITJ	84.42	84.90	92.33	87.03	87.17	3.62	2.03
Limited Data Track	Baseline [10]	84.48	85.77	91.72	86.52	87.12	3.18	1.87
	ArogyaPandit	79.78	81.25	87.70	80.82	82.39	3.59	1.64
	DMSTI-Neurotechnology	84.87	86.20	91.40	87.25	87.43	**2.82**	1.76
	Idiap-BSP	86.60	87.38	93.18	88.52	**88.92**	2.95	1.97
	STYK IITJ	74.27	72.15	82.60	73.17	75.55	4.78	**1.60**

the final projection layer is fine-tuned, while the rest of the backbone remains frozen. This already improves substantially over zero-shot CLIP, but the gap with the LoRA-based baseline widens at stricter FAR thresholds, showing that adapting only the final projection is not enough to reach the fine-grained discriminability needed at low false accept rates.

Based on the evaluation criteria, the top three teams for each track are as follows. In the Full Data Track, DMSTI-Neurotechnology ranks first, achieving the highest average accuracy on the small benchmarks and dominating across all IJB-B and IJB-C TAR thresholds. Idiap-BSP ranks second, leading on TinyFace Rank-1 and Rank-5 while remaining competitive across all other benchmarks. STYK IITJ ranks third, consistently outperforming ArogyaPandit across verification and identification metrics. In the Limited Data Track, Idiap-BSP ranks first, achieving the best performance across all benchmarks including the small verification sets, all IJB-B and IJB-C TAR thresholds, and both TinyFace metrics. DMSTI-Neurotechnology ranks second, performing comparably to Idiap-BSP on the small benchmarks while showing a notable drop at stricter FAR thresholds on IJB-B and IJB-C. ArogyaPandit ranks third, outperforming STYK IITJ across all metrics by a significant margin.

4.2. Bias Evaluation

To assess demographic fairness across submitted approaches, we evaluate each method on the RFW benchmark [39], which partitions the test set into four ethnic groups: African, Asian, Caucasian, and Indian. Results are reported in Table 5 in terms of average recognition accuracy, STD, and SER. CLIP (72.19%, STD 4.80) is included for reference, highlighting the substantial gain brought by fine-tuning. The FRoundation [10] baseline sets the competitive benchmark at 87.07% (STD 2.99) and 87.12% (STD 3.18)

for the Full and Limited Data tracks respectively. From the reported results, we make the following observations:

- In the Full Data Track, DMSTI-Neurotechnology surpasses the FRoundation baseline [10] and all competing approaches by a clear margin, achieving the best accuracy (91.70%) and lowest STD (2.33), improving both performance and fairness simultaneously. Idiap-BSP ranks second, achieving 89.39% (STD 2.65). STYK IITJ matches FRoundation in accuracy ($\sim$87%) but exhibits higher STD (3.62), indicating greater demographic disparity despite comparable average performance. ArogyaPandit falls below the baseline and other competing approaches in accuracy (75.48%) despite achieving the best SER (1.60), reflecting a trade-off between error balance and overall performance that limits its practical competitiveness.

- In the Limited Data Track, rankings shift notably. Idiap-BSP leads all participants (88.92%, STD 2.95), exceeding competing solutions and the FRoundation baseline [10] in both accuracy and fairness. DMSTI-Neurotechnology ranks second, achieving 87.43% (STD 2.82), and retains the lowest STD in this track, maintaining competitive fairness despite a drop relative to its Full Data performance.

- Across both tracks and all methods, the Caucasian group consistently yields the highest per-group accuracy, while the African and Asian groups systematically score lowest, reflecting demographic biases inherited from CLIP pre-training that fine-tuning on synthetic identity datasets only partially mitigates.

Overall, DMSTI-Neurotechnology represents the strongest submission in the Full Data Track, excelling in both recognition accuracy and demographic fairness, while Idiap-BSP emerges as the most competitive participant in the Limited Data Track, demonstrating generally robust and fair adaptation under constrained training conditions.

4.3. Evaluation of Synthetic Face Datasets

To justify the choice of the synthetic data generation method adopted for the competition, we fine-tune the baseline model [10] on several SOTA synthetic face datasets, including IDiff-Face [5], DCFace [23], Arc2Face [31], SFace2 [6], and the IDPerturb dataset [3], which was used to generate the synthetic training data for both the Full Data and Limited Data Tracks. All considered synthetic datasets contain the same number of samples, consisting of 0.5 million synthetic face images generated from 10K identities. Table 6 summarizes the verification performance on the considered FR evaluation benchmarks defined in Section 2.3.The baseline model fine-tuned with IDPerturb consistently outperforms models fine-tuned on the other synthetic datasets across all considered evaluation benchmarks. These results provide empirical justification for adopting IDPerturb as the synthetic data generation method for the competition.

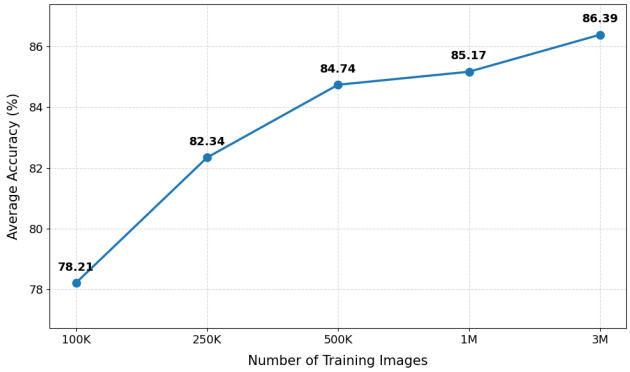

Figure 2. The effect of data volume on downstream FR performance. The y-axis reports IJB-C TAR at FAR of 1e-4, and the x-axis reports the size of the synthetic dataset subset used to fine-tune the baseline model (100K, 250K, 500K, 1M, and 3M images). Performance improves consistently with dataset size

To contextualize these results with respect to real-data training and prior work, we additionally report a real-data baseline trained on CASIA-WebFace [42], matched in scale to the synthetic datasets used in this competition (10K identities, 0.5M images). In the previous SDFR 2024 competition [30], models trained on synthetic data consistently underperformed relative to real-data training, leaving a substantial synthetic-to-real gap. In contrast, we observe that the baseline fine-tuned on the IDperturb dataset achieves better results than the model trained on CASIA-WebFace at comparable scale. This suggests that adapting foundation models for synthetic data generation is a promising direction for narrowing the synthetic-to-real gap identified in SDFR 2024.

4.4. Data Efficiency Analysis

To assess how performance changes with the amount of training data, we fine-tuned the baseline model on subsets of 100K, 250K, 500K, 1M, and 3M images and report IJB-C TAR at FAR of 1e-4 in Figure 2. Performance improves consistently with dataset size, from 78.21% at 100K to 86.39% at 3M images, with the largest gains observed between 100K and 500K images (78.21% to 84.74%), while further scaling to 1M and 3M images brings smaller additional improvements (85.17% and 86.39%, respectively).

4.5. Computational Cost

Table 7 reports the number of trainable parameters for each submitted approach. The full CLIP ViT-L/14 model, including both image and text encoders, comprises

Table 6. The achieved verification performances by the baseline approach, defined in Section 2, fine-tuned on different SOTA synthetic face datasets. We additionally report a real-data baseline trained on CASIA-WebFace, matched in scale to the synthetic datasets (10K identities, 0.5M images), to contextualize these results with respect to real-data training and prior work. The results are reported in (%) on the small benchmarks, defined in Section 2.3, and as average accuracies. On IJB-B and IJB-C, the results are reported as TAR at FAR of 1e-3, 1e-4 and 1e-5. IDPerturb consistently achieves the best performance across all benchmarks, demonstrating the effectiveness of the proposed synthetic data generation approach compared with prior state-of-the-art synthetic face datasets.

Training Dataset	LFW	CFP-FP	AgeDB30	CALFW	CPLFW	Avg.	IJB-B			IJB-C		
							10^{-3}	10^{-4}	10^{-5}	10^{-3}	10^{-4}	10^{-5}
CASIA-WebFace [42]	99.55	95.73	91.73	92.58	91.70	94.26	90.33	78.71	56.96	92.59	83.12	68.87
WebFace4M [45]	99.65	96.50	93.72	94.37	93.73	95.59	95.12	90.72	82.76	96.62	93.40	88.55
IDiff-Face (ICCV 2023) [5]	98.75	91.40	86.50	90.78	84.87	90.46	82.92	64.56	34.82	86.07	71.77	48.18
DCFace (CVPR 2023)[23]	98,62	94,07	90,45	91,33	87,10	92,31	86.36	73.27	51.65	89.38	77.63	56.13
Arc2Face (ECCV 2024) [31]	98,88	95,19	85,68	88,70	88,22	91,33	84.62	69.06	50.51	88.36	74.98	59.29
SFace2 (TBIOM 2024) [6]	97.88	89.87	79.03	86.50	83.88	87.43	75.31	54.86	25.14	78.64	60.81	37.73
IDperturb (CVPR 2026)[3]	99.37	96.76	92.05	92.92	90.917	**95.27**	**92.04**	**84.01**	**67.25**	**94.07**	**88.24**	**76.72**

427,616,513 parameters. Although participants were permitted to leverage the text encoder in their solutions, all submitted approaches relied solely on the image encoder, which accounts for 303,966,208 parameters. Within this image encoder, the number of trainable parameters varies substantially depending on the adaptation strategy, ranging from as few as 0.8M for ArogyaPandit's rank-8 LoRA configuration in the Limited Data Track, to 304M for DMSTI-Neurotechnology's full fine-tuning, which updates the entire image encoder. Idiap-BSP's rsLoRA configurations fall in between, using 25M trainable parameters in the Full Data Track (r=256) and 101M in the Limited Data Track (r=1024). Despite these differences in trainable parameters and training cost, all submitted models share the same CLIP ViT-L/14 image encoder architecture at inference time, resulting in identical inference cost 84.372 GFLOPs across submissions.

Table 7. Number of trainable parameters for submitted approach. While all methods share the same CLIP ViT-L/14 backbone at inference time, the number of trainable parameters vary depending on the adaptation strategy and the chosen hyperparameters.

Team	**Track**	**Adaptation Method**	**#Trainable Parameters(M)**
Baseline [10]	Full Data Limited Data	LoRA, r=16	1.6
ArogyaPandit	Full Data Limited Data	LoRA, r=16 LoRA, r=8	1.6 0.8
DMSTI-Neurotechnology	Full Data Limited Data	Full fine-tuning	304
Idiap-BSP	Full Data Limited Data	LoRA, r=256 LoRA, r=1024	25 101
STYK IITJ	Full Data Limited Data	LoRA, r=16	1.6

5. Conclusion

This paper presented the IJCB-AFMFR 2026 Competition on Adapting Foundation Models for FR Using Synthetic Training Data. The competition provided a controlled and reproducible benchmark for evaluating adaptation strategies for the CLIP ViT-L/14 foundation model, with all training data generated exclusively via the IDPERTURB synthetic face generation framework and without any use of real face images. Two tracks were organized to assess adaptation under large-scale and limited-data conditions, attracting 28 registered teams and four valid submitting teams from academic and industrial institutions across Asia, Africa, Europe, North and South America.

The evaluation results demonstrate that adapting CLIP with synthetic face data substantially improves recognition performance over the off-the-shelf model across all benchmarks. In the Full Data Track, DMSTI-Neurotechnology's full fine-tuning strategy with Sub-Center ArcFace achieves the strongest results in both recognition accuracy and demographic fairness, while in the Limited Data Track, Idiap-BSP's rsLoRA approach proves the most robust and consistent method, outperforming other submitted solution and the baseline FRoundation across all reported benchmarks. Fairness evaluation further reveals that, while fine-tuning significantly reduces demographic disparity relative to the pre-trained CLIP model for most submitted solutions, systematic performance gaps across ethnic groups persist, pointing to an important open challenge for the community. The competition protocol, evaluation benchmarks, and ranking methodology introduced in this work provide a reproducible framework for future comparisons in this emerging research direction. The findings motivate further investigation into the role of synthetic data diversity in shaping the generalization of adapted foundation models.

Acknowledgment

This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] D. Biderman, J. P. Portes, J. J. G. Ortiz, M. Paul, P. Greengard, C. Jennings, D. King, S. Havens, V. Chiley, J. Frankle, C. Blakeney, and J. P. Cunningham. Lora learns less and forgets less. *Trans. Mach. Learn. Res.*, 2024, 2024.

[2] P. Borsukiewicz, F. Boutros, I. E. Olatunji, C. Beumier, W. C. Ouedraogo, J. Klein, and T. F. Bissyandé. Beyond real faces: synthetic datasets can achieve reliable recognition performance without privacy compromise. *npj Artificial Intelligence*, 2026.

[3] F. Boutros, E. Caldeira, T. Chettaoui, and N. Damer. Idperturb: Enhancing variation in synthetic face generation via angular perturbation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

[4] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2022, New Orleans, LA, USA, June 19-20, 2022*, pages 1577–1586. IEEE, 2022.

[5] F. Boutros, J. H. Grebe, A. Kuijper, and N. Damer. Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion models. In *ICCV*, pages 19593–19604. IEEE, 2023.

[6] F. Boutros, M. Huber, A. T. Luu, P. Siebke, and N. Damer. Sface2: Synthetic-based face recognition with w-space identity-driven sampling. *IEEE Trans. Biom. Behav. Identity Sci.*, 6(3):290–303, 2024.

[7] F. Boutros, V. Struc, J. Fiérrez, and N. Damer. Synthetic data for face recognition: Current state and future prospects. *Image Vis. Comput.*, 135:104688, 2023.

[8] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *13th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2018, Xi'an, China, May 15-19, 2018*, pages 67–74. IEEE Computer Society, 2018.

[9] Z. Cheng, X. Zhu, and S. Gong. Low-resolution face recognition. In *ACCV (3)*, volume 11363 of *Lecture Notes in Computer Science*, pages 605–621. Springer, 2018.

[10] T. Chettaoui, N. Damer, and F. Boutros. Froundation: Are foundation models ready for face recognition? *Image Vis. Comput.*, 156:105453, 2025.

[11] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le. Randaugment: Practical automated data augmentation with a reduced search space. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, June 14-19, 2020*, pages 3008–3017. Computer Vision Foundation / IEEE, 2020.

[12] J. Dan, Y. Liu, H. Xie, J. Deng, H. Xie, X. Xie, and B. Sun. Transface: Calibrating transformer training for face recognition from a data-centric perspective. In *ICCV*, pages 20585–20596. IEEE, 2023.

[13] I. DeAndres-Tame, R. Tolosana, P. Melzi, R. Vera-Rodriguez, M. Kim, C. Rathgeb, X. Liu, L. F. Gomez, 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 frcsyn-ongoing: Winning solutions and post-challenge analysis to improve face recognition with synthetic data. *Information Fusion*, page 103099, 2025.

[14] J. Deng, J. Guo, T. Liu, M. Gong, and S. Zafeiriou. Sub-center arcface: Boosting face recognition by large-scale noisy web faces. In *ECCV (11)*, Lecture Notes in Computer Science, pages 741–757. Springer, 2020.

[15] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, pages 4690–4699. Computer Vision Foundation / IEEE, 2019.

[16] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):5962–5979, Oct. 2022.

[17] C. A. Fontanillo López and A. Elbi. On synthetic data: a brief introduction for data protection law dummies. `https://europeanlawblog.eu/2022/09/22/on-synthetic-data-a-brief-introduction-for-data-protection-law-dummies/`, 2022.

[18] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III*, volume 9907 of *Lecture Notes in Computer Science*, pages 87–102. Springer, 2016.

[19] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. Lora: Low-rank adaptation of large language models. In *ICLR*. OpenReview.net, 2022.

[20] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, Oct. 2008. Erik Learned-Miller and Andras Ferencz and Frédéric Jurie.

[21] D. Kalajdzievski. A rank stabilization scaling factor for fine-tuning with lora. 2023.

[22] M. Kim, A. K. Jain, and X. Liu. Adaface: Quality adaptive margin for face recognition. In *CVPR*, pages 18729–18738. IEEE, 2022.

[23] M. Kim, F. Liu, A. K. Jain, and X. Liu. Dcface: Synthetic face generation with dual condition diffusion model. In *CVPR*, pages 12715–12725. IEEE, 2023.

[24] S. Magistri, D. Goswami, M. Mistretta, B. Twardowski, J. van de Weijer, and A. D. Bagdanov. Isoclip: Decomposing clip projectors for efficient intra-modal alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 29315–29324, June 2026.

[25] B. Maze, J. C. Adams, J. A. Duncan, N. D. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, and P. Grother. IARPA janus benchmark - C: face dataset and protocol. In *ICB*, pages 158–165. IEEE, 2018.

[26] P. Melzi, R. Tolosana, R. Vera-Rodríguez, M. Kim, C. Rathgeb, X. Liu, I. DeAndres-Tame, A. Morales, J. Fiérrez, J. Ortega-Garcia, W. Zhao, X. Zhu, Z. Yan, X. Zhang, J. Wu, Z. Lei, S. Tripathi, M. Kothari, M. H. Zama, D. Deb, B. Biesseck, P. Vidal, R. Granada, G. P. Fickel, G. Führ, D. Menotti, A. Unnervik, A. George, C. Ecabert, H. Otroshi-Shahreza, P. Rahimi, S. Marcel, I. Sarridis, C. Koutlis, G. Baltsou, S. Papadopoulos, C. Diou, N. D. Domenico, G. Borghi, L. Pellegrini, E. Mas-Candela, Á. Sánchez-Pérez, A. Atzori, F. Boutros, N. Damer, G. Fenu, and M. Marras. Frcsyn-ongoing: Benchmarking and comprehensive evaluation of real and synthetic data to improve face recognition systems. *Inf. Fusion*, 107:102322, 2024.

[27] P. Melzi, R. Tolosana, R. Vera-Rodríguez, M. Kim, C. Rathgeb, X. Liu, I. DeAndres-Tame, A. Morales, J. Fiérrez, J. Ortega-Garcia, W. Zhao, X. Zhu, Z. Yan, X. Zhang, J. Wu, Z. Lei, S. Tripathi, M. Kothari, M. H. Zama, D. Deb, B. Biesseck, P. Vidal, R. Granada, G. P. Fickel, G. Führ, D. Menotti, A. Unnervik, A. George, C. Ecabert, H. Otroshi-Shahreza, P. Rahimi, S. Marcel, I. Sarridis, C. Koutlis, G. Baltsou, S. Papadopoulos, C. Diou, N. D. Domenico, G. Borghi, L. Pellegrini, E. Mas-Candela, Á. Sánchez-Pérez, A. Atzori, G. Fenu, F. Boutros, M. Marras, and N. Damer. Frcsyn challenge at WACV 2024: Face recognition challenge in the era of synthetic data. In *WACV (Workshops)*, pages 892–901. IEEE, 2024.

[28] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou. Agedb: the first manually collected, in-the-wild age database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop*, volume 2, page 5, 2017.

[29] M. Oquab, T. Darcet, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P. Huang, S. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jégou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski. Dinov2: Learning robust visual features without supervision. *Trans. Mach. Learn. Res.*, 2024, 2024.

[30] H. Otroshi-Shahreza, C. Ecabert, A. George, A. Unnervik, S. Marcel, N. D. Domenico, G. Borghi, D. Maltoni, F. Boutros, J. Vogel, N. Damer, Á. Sánchez-Pérez, E. Mas-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-Rodríguez, G. Perelli, G. Orrù, G. L. Marcialis, and J. Fiérrez. SDFR: synthetic data for face recognition competition. In *FG*, pages 1–9. IEEE, 2024.

[31] F. P. Papantoniou, A. Lattas, S. Moschoglou, J. Deng, B. Kainz, and S. Zafeiriou. Arc2face: A foundation model for id-consistent human faces. In *ECCV (37)*, volume 15095 of *Lecture Notes in Computer Science*, pages 241–261. Springer, 2024.

[32] 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 *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 2021.

[33] S. Sengupta, J. Chen, C. Castillo, V. Patel, R. Chellappa, and D. Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016. Institute of Electrical and Electronics Engineers Inc., May 2016. Publisher Copyright: © 2016 IEEE.; IEEE Winter Conference on Applications of Computer Vision, WACV 2016 ; Conference date: 07-03-2016 Through 10-03-2016.

[34] H. O. Shahreza and S. Marcel. Foundation models and biometrics: A survey and outlook. *Authorea Preprints*, 2025.

[35] The European Parliament and the Council of the European Union. General data protection regulation, 2016.

[36] H. Touvron, A. Vedaldi, M. Douze, and H. Jégou. Fixing the train-test resolution discrepancy. In *NeurIPS*, pages 8250–8260, 2019.

[37] 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 *CVPR*, pages 5265–5274. Computer Vision Foundation / IEEE Computer Society, 2018.

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

[39] M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *ICCV*, pages 692–702. IEEE, 2019.

[40] M. Wang, Y. Zhang, and W. Deng. Meta balanced network for fair face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(11):8433–8448, 2022.

[41] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. C. Adams, T. Miller, N. D. Kalka, A. K. Jain, J. A. Duncan, K. Allen, J. Cheney, and P. Grother. IARPA janus benchmark-b face dataset. In *CVPR Workshops*, pages 592–600. IEEE Computer Society, 2017.

[42] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *CoRR*, abs/1411.7923, 2014.

[43] T. Zheng and W. Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. Technical Report 18-01, Beijing University of Posts and Telecommunications, February 2018.

[44] T. Zheng, W. Deng, and J. Hu. Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments. *CoRR*, abs/1708.08197, 2017.

[45] 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 *CVPR*, pages 10492–10502. Computer Vision Foundation / IEEE, 2021.