A Comparative Study on Synthetic Facial Data Generation Techniques for Face Recognition

Pedro Vidal*, Bernardo Biesseck*, Luiz Coelho ‡, Roger Granada‡, David Menotti*
*Depart. of Informatics, Federal University of Paraná, Curitiba, PR, Brazil {pbqv20, bjgbiesseck, menotti}@inf.ufpr.br
‡Unico - IDTech, Brazil {luiz.coelho, roger.granada}@unico.io

Abstract—**Face recognition has become a widely adopted method for user authentication and identification, with applications in various domains such as secure access, law enforcement, and locating missing persons. The success of this technology is largely attributed to deep learning, which leverages large datasets and effective loss functions to achieve highly discriminative features. Despite its advancements, face recognition still faces challenges in areas such as explainability, demographic bias, privacy and robustness against aging, pose variations, illumination changes, occlusions, and expressions. Additionally, the emergence of privacy regulations has led to the discontinuation of several well-established datasets, raising legal, ethical, and privacy concerns. To address these issues, synthetic facial data generation has been proposed as a solution. This technique not only mitigates privacy concerns but also allows for comprehensive experimentation with facial attributes that cause bias, helps alleviate demographic bias, and provides complementary data to enhance models trained with real data. Competitions, such as the FRCSyn and SDFR, have been organized to explore the limitations and potential of face recognition technology trained with synthetic data. This paper compares the effectiveness of established synthetic face datasets with different generation techniques in face recognition tasks. We benchmark the accuracy of seven mainstream datasets, providing a vivid comparison of approaches that are not explicitly contrasted in the literature. Our experiments highlight the diverse techniques used to address the synthetic facial data generation problem and present a comprehensive benchmark of the area. The results demonstrate the effectiveness of various methods in generating synthetic facial data with realistic variations, evidencing the diverse techniques used to deal with the problem.**

I. Introduction

Face recognition is a popular method for authenticating and identifying users, such as for access to secure facilities or devices, for law enforcement purposes, and to locate missing persons, among others. Facial data is also the main approach for biometric recognition and the success of this technology can be attributed to deep learning. By training with large amounts of data and effective loss functions based on margin loss, highly effective discriminative features are achieved.

However, it is still a challenging problem that encounters room for improvement in areas like explainability, demographic bias, privacy and robustness against aging, pose variations, illumination changes, occlusions, and expressions. Another problem is posed by the emergence of privacy regulations, which led to the discontinuation of many well-established datasets for training facial recognition models, such as MS-Celeb-1M [1]. This discontinuation is due to the data being collected from online sources without individual consent, raising legal, ethical, and privacy concerns.

To address this, solutions that generate synthetic facial data were proposed. This technique offers a valuable solution to the field, and it can be used to: mitigate demographic bias in facial recognition models [2]; provide a comprehensive way to experiment with facial attributes that influence recognition accuracy [3]; Supplement real data to enhance model performance [4].

Several competitions related to this topic were proposed. The 1st and 2nd editions of the FRCsyn competition [4], [5] have the objective of answering the following questions: What are the limitations of FR technology trained only with synthetic data? Can synthetic data help alleviate current limitations in FR technology? For the first edition, the organizers proposed subtasks that invited the participants to use synthetic data alone and in conjunction with real data to mitigate demographic bias and bring performance improvement. In the second edition, they extended to an unconstrained number of synthetic images, maintaining the same objectives. With a related objective, SDFR competition [6] proposed that participants submit original solutions to generate synthetic data for performance improvement and mitigate the synthetic-to-real gap.

As the competitors are encouraged to submit models with already established synthetic datasets or generate new ones, a lot of characteristics need to be considered. For example, the used dataset needs to have sufficient intra-class variations and changes in pose, aging, expressions, occlusions, and illumination. Adding to that, the datasets need to have sufficient interclass variations, for the proposed models to generalize to new unseen data. Given these challenges, generating sufficient intra-class and interclass is an active area of research.

With the synthetic facial data generation problem in mind, this work aims to compare different established datasets that use different techniques to deal with the problem. We use seven mainstream datasets to benchmark the accuracies in the face recognition task. We aim with these experiments to present a vivid benchmark of the area, contrasting approaches that are not explicitly compared in the literature and evidencing the diverse techniques used to approach this problem.

The remainder of this work is organized as follows: Section II presents related works, Section III describes the methodology, Section IV presents the results and discusses them, and Section V concludes by pointing out some future directions.

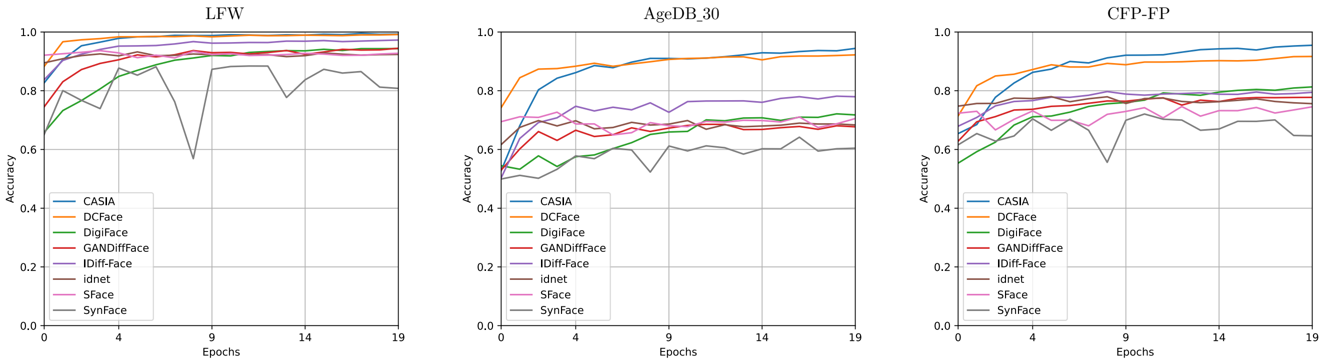

Fig. 1. Training accuracy results on (y-axis) in respect to the training epoch on (x-axis), for each benchmarked synthetic dataset, using the same training configurations.

TABLE I
BENCHMARK RESULTS OF VARIOUS SYNTHETIC FACIAL DATA GENERATION TECHNIQUES

Dataset	AGEBD	BUPT	CFP-FP	ROF	AVG ACC	GAP
CASIA	95.99	95.74	96.87	88.50	94.27	
DCFace	95.14	95.55	92.69	86.83	92.55	1.72
Idiffface	84.40	88.70	79.86	79.42	83.10	11.17
Digiface	79.10	82.35	82.17	75.47	79.77	14.5
Gandiff	76.26	82.46	77.19	75.44	77.84	16.43
Idnet	76.42	82.41	75.34	70.53	76.18	18.09
Sface	75.90	78.80	74.79	69.78	74.81	19.46
Synface	63.75	69.05	65.59	58.64	64.26	30.01

II. RELATED WORKS

Several approaches have been proposed in the last couple of years for facial image synthesis. Many of them use Generative models such as GAN's [7] and Diffusion Models [8]. Synface [9], Idnet [10], Idifface [11], SFace [12] use GAN's to generate the images and DCFace [13] uses diffusion models. Also, a combination of diffusion models and GAN's was applied by the GANdifface [14]. Another technique uses a computer graphics pipeline to synthesize the facial images and was proposed by Digiface. In the next paragraphs, it is described in more detail the mentioned state-of-the-art approaches. Synface [9] explores the performance gap between models trained on synthetic and real face images, identifying poor intra-class variations and domain gaps as key factors. To address these, the authors introduce identity mixup (IM) and domain mixup (DM) techniques, demonstrating significant improvements in face recognition with synthetic data. DCface [13] proposes training a diffusion models pipeline. This pipeline consists of two stages, (i) a sampling stage generating a new identity image, and (ii) a mixing stage combining the generated image with a style image from a style bank, creating a final image that blends both style and identity information. Using this approach, the authors significantly reduced the synthetic to real domain gap.

The SFace [12] dataset was generated using a class-conditional StyleGAN2-ADA. The dataset aims to address privacy and ethical concerns associated with using real-face datasets. The authors evaluate the identity relation between the synthetic and original datasets and propose three learning strategies for training face recognition models: multi-class classification, label-free knowledge transfer, and combined learning. Aiming to improve the previously mentioned work, the IDnet [10] dataset was generated using a class-conditioned StyleGAN2-ADA. They integrate the GAN min-max game with an identity separable loss, named ID3, and a domain adaptation loss. This approach enables the generator to learn encoded identity information, generating identity-separable synthetic samples, and minimizing the domain gap between synthetic and real data distributions.

Idifface [11] utilizes a diffusion model trained in the latent space of a pre-trained autoencoder. The diffusion model is also conditioned on identity context through feature representations obtained from a pre-trained face recognition model. They introduce a cross-attention mechanism to inject the identity condition into the intermediate representations of the diffusion model. The introduction of partial dropout in the components of the identity context during training to prevent overfitting and increase intra-class diversity is done.

Digiface [15] is a framework that can generate unique subjects based on 3D model parametric rendering. By also applying heavy data augmentation, they considerably bridge the synthetic to real domain gap. Gandiff [14] used a StyleGAN3 to generate synthetic images and further used transformation in the latent space to control the generated attributes that serve as input to a diffusion model to produce realistic intra-class variations.

In this work, we selected the seven mentioned approaches to compose our comparative study. Hence, Section III describes in more detail the methodology that delineates our benchmark.

III. Methodology

To compare the selected datasets, we choose the FRCsyn benchmark used to rank the submitted models to the competition. This benchmark is composed of four datasets, named AGEDB [16], BUPT [17], CFP-FP [18] and ROF [19]. The AgeDB dataset is focused on a comparison of images across age, and the most challenging scenario was used, named AgeDB-30, which has a gap of 30 years between the face images of the individuals and contains 3000 genuine pairs and 3000 impostor pairs. The BUPT dataset is designed to address performance disparities across different ethnic groups and has 4000 impostor and genuine pairs for each. On the other hand, the CFP-FP dataset contains images to evaluate differences in face pose (i.e., frontal and profile faces). The protocol includes 3500 genuine pairs and 3500 impostor pairs. The ROF dataset was also used, containing challenges related to occlusions and 1800 genuine and impostor pairs for each. The results of the benchmark are reported in terms of the best overall accuracy (ACC).

As we aim to fairly compare these works, we collect their publicly available datasets from Kim et al. [13][1], Boutros et al. [12][2] [11][3], Kolf et al. [10][4], Melzi et al. [14][5], Qiu et al. [9][6], and Bae et al. [15][7].

All the datasets' images were aligned and cropped using the RetinaFace [20] detector. For each selected dataset we trained a model using the Insightface library, and applied the Arcface loss [21]. As additional data augmentation, we apply random erasing and rand augment. We use the SGD optimizer and set the momentum to 0.9 and weight decay to 5.10^{-4}. The learning rate was set to 0.02 and decayed at each iteration. The models were trained for 20 epochs within a batch size of 128 on one NVIDIA TITAN Xp with 12GB memory.

IV. Results

In this section, we present and discuss the results of the datasets considered in the experiments, summarized in table I and figure 1. For comparison purposes, related to the synthetic to real domain gap, we selected a dataset containing real identities. In the first position, the DCface dataset proved to achieve the best performance in all four datasets, achieving an average accuracy of 92.55 and a synthetic to real accuracy gap of 1.72. It was also the main dataset choice for the FRCsyn 1st and 2nd editions. They used two diffusion models, one for identity generation and another for mixing two identities. The training employed a dataset containing real identities, named CASIA-Webface, which guided the model to learn how to mix identities. At the sampling stage, the generator and the mixer receive only synthetic images. However, the method uses labels of real identities and it was not allowed at the SDFR competition.

IDifface comes in second place, achieving an average accuracy of 83.10 and bridging the synthetic to real accuracy gap to 11.17. This result reveals the potential that diffusion models can have when generating images with variations in pose, age, expression, and illumination, containing unique information. As stated in DCFace, the diffusion models can generate a bigger number of unique identities than GAN's, which results in a dataset with more variety. Also, they use a pre-trained model to extract embeddings that are used by the conditional generator model based on diffusion. However, as identity labels were not used to train the face generator model, this dataset was allowed in the SDFR competition and was adopted by the majority of the competitors.

Digiface proposed a unique solution to the problem, by using a computer graphics pipeline to render the facial images. With this, they were able to achieve an average accuracy of 79.77 and a gap to the real of 14.5. They achieve a higher value on the CFP-FP dataset when compared to Iddiface, which indicates that profile images are more present in the training dataset of this solution. However, this approach is extremely computationally costly, and might not be available for research. With similar performance, comes the GANdiff face dataset that results in an average accuracy of 77.84 and a gap to the real of 16.43. They used StyleGAN3 to generate the images and further input those images into a diffusion model (ie. Dream Both), to generate the intra-class variations. The additional diffusion model enables the dataset to have a more realistic intra-class variation, in comparison to SFace, that employed StyleGAN3 solely and achieved an average accuracy of 74.84 and a domain gap of 19.46.

In sequence comes Idnet, which achieved an average accuracy of 76.18 and a domain gap of 18.09. According to the authors, "SFace suffers from relatively low identity separability which might lead to less optimal face verification accuracies when such synthetic data is used to train FR" [10]. To deal with this they integrate to the GAN min-max game and identity separable loss, named ID3, and a domain adaptation loss to make the generator learn to identify identity information encoded and generate more identity separable images. However, this comes with the cost of using identity labels in training the generative framework.

Lastly, comes Synface, which employed DiscoFace-GAN [22] and identity mixup and domain mixup techniques. The solution achieved an average accuracy of 64.26 and a synthetic to real domain gap of 30.01. This is caused by the low number of unique samples that the DiscoFaceGAN can generate, as stated in [22]. However, the work has the credit of being, to the best of our knowledge, one of the first to generate a synthetic dataset to train a facial recognition model.

V. Conclusion

In this study, we verify the synthetic to real domain gap of mainstream synthetic databases. We especially highlight the

[1] https://github.com/mk-minchul/dcface

[2] https://github.com/fdbtrs/SFace-Privacy-friendly-and-Accurate-Face-Recognition-using-Synthetic-Data

[3] https://github.com/fdbtrs/IDiff-Face.git

[4] https://github.com/fdbtrs/Three-Player-GAN-IDnet

[5] https://github.com/PietroMelzi/GANDiffFace

[6] https://github.com/haibo-qiu/SynFace

[7] https://github.com/microsoft/DigiFace1M

DCFace dataset, which achieved consistently better results on all four databases and was the main choice for the 1st and 2nd editions of the FRCsyn challenge.

We also present a benchmark of the area and contrast different approaches to deal with the problem, showing the techniques used to deal with the problem. As a natural extension of the comparative work done here, we aim to compare these datasets using a more diverse dataset consisting of surveillance scenarios and evaluate the accuracy achieved by demographic groups, another important aspect when generating new synthetic datasets.

Acknowledgment

This work was supported by a tripartite-contract, i.e., unico - idTech, UFPR (Federal University of Paraná), and FUNPAR (Fundação da Universidade Federal do Paraná). We also thank the National Council for Scientific and Technological Development (CNPq) (# 315409/2023-1) for supporting Prof. David Menotti.

References

[1] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 87–102.

[2] 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 *2023 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2023, pp. 1–9.

[3] G. Balakrishnan, Y. Xiong, W. Xia, and P. Perona, "Towards causal benchmarking of biasin face analysis algorithms," *Deep Learning-Based Face Analytics*, pp. 327–359, 2021.

[4] P. Melzi, R. Tolosana, R. Vera-Rodriguez, M. Kim, C. Rathgeb, X. Liu, I. DeAndres-Tame, A. Morales, J. Fierrez, J. Ortega-Garcia *et al.*, "Frcsyn-ongoing: Benchmarking and comprehensive evaluation of real and synthetic data to improve face recognition systems," *Information Fusion*, vol. 107, p. 102322, 2024.

[5] I. DeAndres-Tame, R. Tolosana, P. Melzi, R. Vera-Rodriguez, M. Kim, C. Rathgeb, X. Liu, A. Morales, J. Fierrez, J. Ortega-Garcia *et al.*, "Frcsyn challenge at cvpr 2024: Face recognition challenge in the era of synthetic data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 3173–3183.

[6] H. O. Shahreza, C. Ecabert, A. George, A. Unnervik, S. Marcel, N. Di Domenico, G. Borghi, D. Maltoni, F. Boutros, J. Vogel *et al.*, "Sdfr: Synthetic data for face recognition competition," *arXiv preprint arXiv:2404.04580*, 2024.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

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

[9] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, and D. Tao, "Synface: Face recognition with synthetic data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 880–10 890.

[10] J. N. Kolf, T. Rieber, J. Elliesen, F. Boutros, A. Kuijper, and N. Damer, "Identity-driven three-player generative adversarial network for synthetic-based face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 806–816.

[11] F. Boutros, J. H. Grebe, A. Kuijper, and N. Damer, "Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19 650–19 661.

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

[13] M. Kim, F. Liu, A. Jain, and X. Liu, "Dcface: Synthetic face generation with dual condition diffusion model," in *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, 2023, pp. 12 715–12 725.

[14] P. Melzi, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, D. Lawatsch, F. Domin, and M. Schaubert, "Gandifface: Controllable generation of synthetic datasets for face recognition with realistic variations," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3086–3095.

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

[16] 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 workshops*, 2017, pp. 51–59.

[17] 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*, 2020, pp. 9322–9331.

[18] 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 *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2016, pp. 1–9.

[19] M. E. Erakιn, U. Demir, and H. K. Ekenel, "On recognizing occluded faces in the wild," in *2021 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2021, pp. 1–5.

[20] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5203–5212.

[21] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.

[22] Y. Deng, J. Yang, D. C. F. Wen, and X. Tong, "Disentangled and controllable face image generation via 3d imitative-contrastive learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.