# Supplementary Material for Complete the Feature Space: Diffusion-Based Fictional ID Generation for Face Recognition

Myeong-Yeon Yi[1,2]
myeongyeon.yi@navercorp.com

DongJae Lee[1]
dongjae.lee@navercorp.com

Naeun Ko[1]
naeun.ko@navercorp.com

Yonghyun Jeong[1]
younghyun.jeong@navercorp.com

Sang-goo Lee[2]
sglee@europa.snu.ac.kr

Seunggyu Chang[1]
seunggyu.chang@navercorp.com

[1] Image Vision,
NAVER Cloud,
South Korea

[2] Department of Computer Science and
Engineering,
Seoul National University,
South Korea,

In this supplementary document, we report qualitative results that demonstrate the effects of variance-based interpolation and ID-NMS. Additionally, we provide additional quantitative comparisons between real and synthetic datasets on face recognition (FR) benchmarks. We also include a report on the training settings and configurations used for the FR models.

## A    Effect of Variance-Based Interpolation

In the main thesis, we demonstrate that variance-based interpolation improves FR performance by improving three key aspects of FR datasets: ID uniqueness, ID preservation, and intra-class diversity. This section presents a qualitative comparison of fictional IDs generated with and without variance-based interpolation. fig. 1 illustrates the qualitative distinction between samples of fictional IDs generated with and without variance-based interpolation.

When the fictional ID is generated without considering variances, the fictional ID may be too similar to a real ID with higher variance than the other real ID. On the other hand, variance-based interpolation enables the fictional ID to be distinct from the real IDs which are the interpolation endpoints. This observation aligns with the findings in table 3 of the main thesis, indicating that when fictional IDs are interpolated without variance, their uniqueness is degraded compared to when variance-based interpolation is used. Thus, employing variance-based interpolation contributes to minimizing redundancy in the generation of a fictional ID.
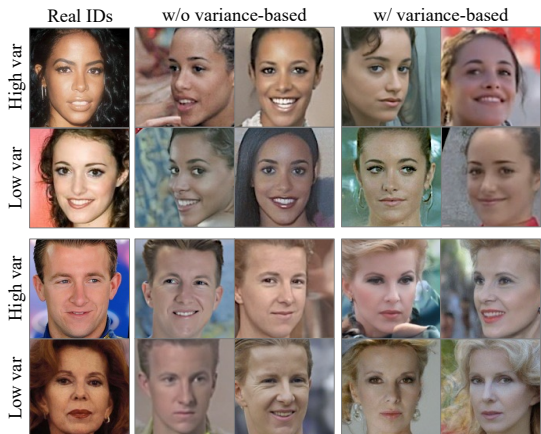


Figure 1: Comparison of samples without and with variance-based interpolation. The real IDs in the first column are the endpoints of interpolations. For each pair of real IDs, the upper ID has a greater average variance than the below one. Variance-based interpolation helps prevent the fictional IDs from being too similar to the real IDs with high variance.



Figure 2: Samples of **discarded** fictional IDs. Samples are generated based on the interpolations between features of the real IDs on their left and right sides. According to the ID-NMS, IDs are discarded when it is too similar to existing IDs or already selected IDs.

# B    Qualitative Results of ID-NMS

This section presents a qualitative comparison of fictional IDs that are discarded and selected according to ID-NMS. According to ID-NMS, it discards IDs when they are too similar to existing IDs or already selected IDs. Fig. 2 depicts samples of discarded fictional IDs. Samples are generated based on the interpolations between features of the real IDs on

Real IDs ................................ ► Selected fictional IDs ◄ ................................ Real IDs
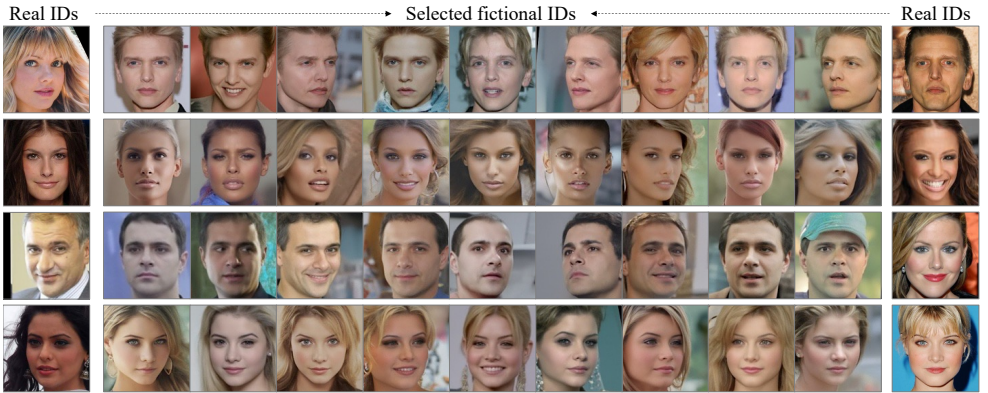


Figure 3: Samples of **selected** fictional IDs. Samples are generated based on the interpolations between features of the real IDs on their left and right sides. The mechanism of ID-NMS leads to an outcome where the selected IDs exhibit enhanced uniqueness, that are distinctly different from the interpolation endpoints.

their left and right sides. As can be seen in Fig. 2, discarded fictional IDs resemble the interpolation endpoints, as observed empirically. In contrast, the selected fictional IDs depicted in Fig. 3 exhibit a lesser degree of resemblance to the interpolation endpoints.

# C    Comparison of Real and Synthetic Datasets

The performance of FR models is observed to deteriorate when trained on synthetic datasets, in comparison to real datasets, due to the existing domain gap between them [2, 2]. In this section, we present an additional comparative analysis of real and synthetic datasets. The evaluation encompasses both qualitative and quantitative perspectives.

| Real/Synth | Authenticity | # of imgs | LFW | CFP-FP | AgeDB | CFP-FF | CALFW | CPLFW | Avg. | Gap to Real |
|---|---|---|---|---|---|---|---|---|---|---|
| Synth. | Real ID | 20 | 96.20 | 70.87 | 77.68 | 94.74 | 85.83 | 71.63 | 82.83 | 6.96 |
| | | 40 | **97.58** | **72.04** | **79.83** | **96.00** | **87.20** | 72.73 | **84.23** | **5.18** |
| | Fict. ID | 20 | 95.18 | 69.71 | 74.63 | 93.56 | 67.97 | 83.93 | 80.83 | 9.61 |
| | | 40 | 96.75 | 70.34 | 78.20 | 93.94 | 69.23 | **84.18** | 82.11 | 7.90 |
| Real | - | 20 | 97.88 | 80.29 | 86.25 | 97.90 | 90.78 | 78.52 | 88.60 | 0.00 |

Table 1: Verification accuracy (%) comparison on benchmark sets of real and synthetic datasets generated by DiffFR. "# of imgs" is the average number of images per ID. The number of IDs is fixed at 10,177 which is that of the original dataset, CelebA.

## C.1    FR Performance

We maintain the number of IDs at 10,177, which is that of the original dataset, CelebA while manipulating the number of images per ID and the authenticity of the identifiers by varying them between real and fictional. As can be seen in Table 1, a noticeable performance gap continues to exist between real and synthetic datasets.

Within synthetic datasets, those containing real IDs exhibit better performance in comparison to those with fictional IDs. The average performance on validation sets reveals

that the synthetic dataset, with 20 images per ID, exhibits an accuracy of 82.83%. In comparison, the real dataset performs significantly better, with an accuracy of 88.60%. However, when the number of images in the synthetic dataset is increased to 40, it achieves a comparable accuracy of 97.58% on LFW, while the real dataset achieves an accuracy of 97.88%. Additionally, in contrast to the other benchmark datasets, the use of fictional IDs yields better performance in CPLFW.

In line with the consistent observation, an increase in the number of images per ID positively impacts the overall performance. When the number of images per ID is 20, the synthetic dataset with real IDs demonstrates better average performance on validation sets, showing an accuracy of 82.83%, which exceeds that of the synthetic dataset employing fictional IDs, which achieves an accuracy of 80.83%.



Figure 4: Comparison of real and synthetic images. Images in each row have the same ID. In comparison to real images, synthetic images exhibit comparable photo-realism and competitive diversity.

## C.2 Qualitative Results

Fig. 4 provides a comparison of real and synthetic images. Within each row of the figure, the images share identical IDs. Synthetic images demonstrate comparable levels of photo-realism and competitive diversity in comparison to real images.

# D    Experimental Settings for FR

For training the FR, we employ a modified ResNet50 modified in [3] with CosFace [9] loss function. Optimizer SGD [5, 6, 8] is applied with a momentum of 0.9 and a weight decay of 5e-4. One NVIDIA Tesla A100 GPUs is used in the training with 1,024 batch size. The learning rate is initially set to 0.1 and decreased by 10 at 10, 16, 21, and 25 epochs and training terminates at 30 epochs. During training, we only use flip data augmentation. To ensure fairness, another pre-trained FR model from Insightface [4] with a modified ResNet50 backbone [3] trained on Glint360k [1], is employed for measuring the metrics.

# References

[1] Xiang An, Xuhan Zhu, Yang Xiao, Lan Wu, Ming Zhang, Yuan Gao, Bin Qin, Debing Zhang, and Fu Ying. Partial fc: Training 10 million identities on a single machine. In *Arxiv 2010.05222*, 2020.

[2] Gwangbin Bae, Martin de La Gorce, Tadas Baltrušaitis, Charlie Hewitt, Dong Chen, Julien Valentin, Roberto Cipolla, and Jingjing Shen. Digiface-1m: 1 million digital face images for face recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3526–3535, 2023.

[3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[4] Jia Guo, Jiankang Deng, Xiang An, Jack Yu, and Baris Gecer. insightface. https://github.com/deepinsight/insightface, 2021.

[5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[6] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.

[7] Haibo Qiu, Baosheng Yu, Dihong Gong, Zhifeng Li, Wei Liu, and Dacheng Tao. Synface: Face recognition with synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10880–10890, 2021.

[8] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR, 2013.

[9] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5265–5274, 2018.