SynFace: Face Recognition with Synthetic Data
(Supplementary Materials)

Haibo Qiu[1]*, Baosheng Yu[2]*, Dihong Gong[3], Zhifeng Li[3], Wei Liu[3], Dacheng Tao[1,2]
[1] JD Explore Academy, China [2] The University of Sydney, Australia
[3] Tencent Data Platform, China
qiuhaibo1@jd.com, baosheng.yu@sydney.edu.au, gongdihong@gmail.com,
michaelzfli@tencent.com, wl2223@columbia.edu, dacheng.tao@gmail.com

In this appendix, we illustrate plenty of face images (from both Syn_10K_50 and CASIA-WebFace) to further demonstrate our observations: 1) the synthetic dataset usually lacks of intra-class variations which significantly degrades the performance (Appendix. A), and 2) the generated face images have limited diversity on facial expressions which are mainly "smiling" with slight differences (Appendix. B).

A. Intra-class Variations

Recalling that there is a clear performance gap (88.98% *vs*. 99.18%) on LFW [4] between SynFace and RealFace. We notice that the fundamental purpose of face synthesis model (*e.g*., DiscoFaceGAN [3]) is to generate high-quality and clean face images, while the face recognition model is usually required to recognize those face images in the wild (*e.g*., LFW [4]) with complex conditions. Therefore, this kind of domain gap leads to the model trained on synthetic data intrinsically lacking well generalization ability.

Then we explore the potential factors which are responsible for the simplicity of Syn_10K_50. Figure 1 demonstrates multiple face images of different people from both CASIA-WebFace (Figure 1a) and Syn_10K_50 (Figure 1b), in which face images of one row belong to the same person. As we can observe, the variations of real face images are clearly larger than the synthetic images. For example, comparing to the synthetic face images, the real face images in the wild usually have the large motion blur and illumination variations. If we augment the synthetic face images with the ColorJitter transformation in PyTorch and MotionBlur from Albumentations [2] for training, the face recognition performance is boosted from 88.98% to 91.23%. Hence, we conclude that the lack of intra-class variations by synthetic dataset leads to its simplicity which significantly degrades the face recognition performance.

B. Expression Diversity

We randomly select three classes from the "Expression" dataset (which means only varying the facial expression of face images while fixing the other attributes) and visualize all the samples (50 images per identity) in Figure 2. Apparently, the differences of images inside the same class are marginal and only reflected by the mouth variations, which reveal the limited expression diversity of "Expression" that is responsible for the worst performance. We conjecture that the 3D priors from 3DMM [1] and the training images from web lack of the expression variations, which result in the limited expression diversity of synthetic face images generated by DiscoFaceFAN [3].

References

[1] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual Conference on Computer graphics and interactive techniques*, pages 187–194, 1999. 1

[2] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020. 1

[3] Yu Deng, Jiaolong Yang, Dong Chen, Fang Wen, and Xin Tong. Disentangled and controllable face image generation via 3d imitative-contrastive learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5154–5163, 2020. 1

[4] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. 2008. 1

*Equal contribution

(a) CASIA-WebFace

(b) Syn_10K_50

Figure 1. Comparison of real and synthetic face images. Each row indicates the same person with different face images. Obviously, comparing the real-world dataset, the synthetic dataset significantly lacks of the intra-class variations.

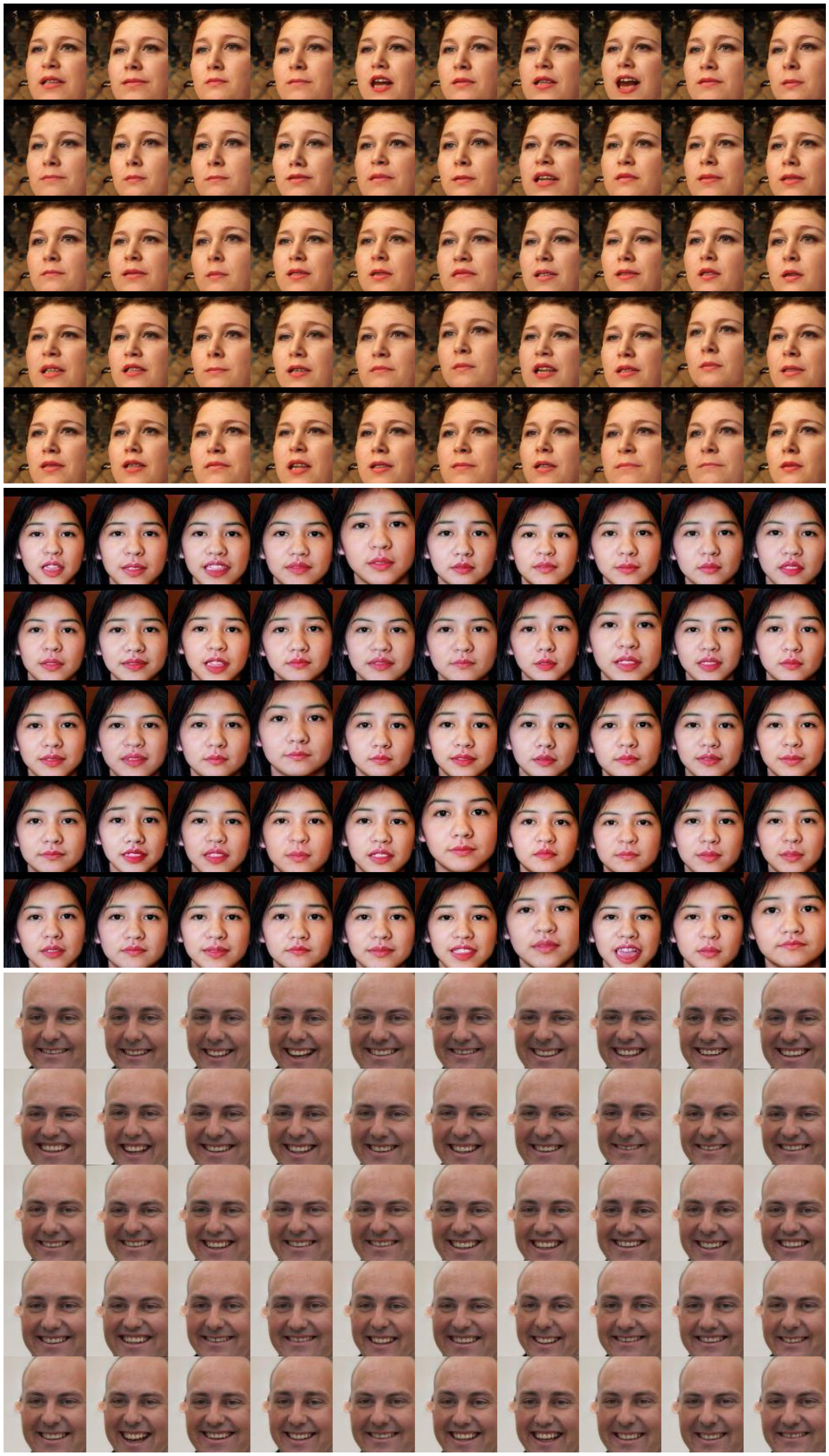

Figure 2. Visualizations of all the samples from three different classes. The generated expressions of face images are mainly "smiling" despite of slight differences, which reveals the limited expression diversity of "Expression".