CAFE-GAN: Arbitrary Face Attribute Editing with Complementary Attention Feature -Supplementary material-

Jeong gi Kwak[1], David K. Han[2], and Hanseok Ko[1]

[1] Korea University, Seoul, Korea
[2] Army Research Laboratory, Adelphi, MD, USA

6 Appendix

In this section, we supplement additional quantitative and qualitative results that are not reported in the main paper. To demonstrate effectiveness of our methods in terms of visual quality and preventing undesired change, we first report two quantitative results i.e., FID score (subsection 6.1) and user study (subsection 6.2) with state-of-the-art methods [1, 2, 5] In addition, we present additional comparison results with [1, 2, 5, 7] in subsection 6.3 and then show extra results of CAFE-GAN in subsection 6.4.

6.1 FID Score

We first report FID (Fréchet Inception Distance) scores [3] which is commonly used as metrics of GAN models. We use the following methods and conditions to measure FID scores. First, we split the celebA test set in two subsets (TestA and TestB). TestA is used as source image of editing task and TestB is used to measure FID score. Therefore, we compare two sets of images (Test A′(set of fake images resulted from TestA) vs. TestB(real)). By doing this, we try to avoid the effect of similar input and output when measuring FID scores and the results are listed in Table. 1. Our model outperforms the other methods in all tasked attributes except for a few cases. Note that we set every image to undergo some changes (e.g., bang → no bang, no bang → bang if *Bangs* is given), so that an evaluation by FID score becomes meaningful.

Table 1: FID (Fréchet Inception Distance) scores. Lower is better.

	Bald	Bangs	Blond h.	Eyegl.	Gender	Musta.	Aged	Mouth	Avg.
AttGAN	18.51	14.38	18.11	34.89	22.35	9.52	13.02	9.11	17.49
StarGAN	25.15	17.62	25.16	26.71	20.59	16.85	18.83	13.67	20.57
STGAN	14.97	8.03	15.28	12.83	11.01	**4.99**	7.03	6.33	10.06
CAFE-GAN	**13.73**	**7.85**	**14.00**	**8.05**	**8.10**	5.47	**5.59**	**4.84**	**8.44**
Real img.									3.03

Table 2: Comparisons on the user preference. Numbers indicate the percentage of preference on each attribute.

	Bald	Bangs	Blond h.	Gender	Aged	Mouth	Avg.
AttGAN	1.34	**29.34**	13.42	13.2	14.37	15.29	14.49
StarGAN	2.39	21.73	11.19	5.26	13.64	4.64	9.81
STGAN	21.42	22.56	24.32	35.21	31.56	12.45	24.59
CAFE-GAN	**74.85**	26.37	**51.07**	**46.33**	**40.43**	**67.62**	**51.11**

6.2 User Study

Since there is no ground truth as to where to change or preserve in attribute editing, the criteria for this can vary from person to person. We thus report results of a user study with a survey platform. The number of attributes used in the evaluation is six. Each user evaluated 10 images per attribute, thus a total of 60 images were evaluated by each user, and they were directed to choose the best edited image with quality considering the specified attributes. The results are listed in Table 2. Our model achieves higher score that other models in five attributes and with a large gap in specific attributes such as *Bald*. Our model get lower score in *Bangs*.

6.3 Additional Comparison Results

We first present additional comparison of qualitative results with same models used in main paper, i.e., AttGAN [2], StarGAN [1], and STGAN [5] on CelebA [6] images (128×128) in Fig. 1. we also present comparison results on CelebA-HQ dataset [4] which contains 30,000 images in CelebA with high quality. We use 256×256 images for comparison with RelGAN [7] which use the difference between a target and a source vector (relative attributes) to constrain in addressing selected attributes. The results of RelGAN have some unintended changes in edited images. For example, when *Mustache* is given as an attribute to be inserted, RelGAN edits female look to like male as shown in Fig. 2, but CAFE-GAN edits only the regions related to mustache. Furthermore, CAFE-GAN expresses fine details better when *Open* or *Close Mouth* is given.

6.4 Extra Results of CAFE-GAN

Finally, we present extra qualitative results of CAFE-GAN in Fig. 3. Through the results, we demonstrate a superior visual quality and editing performance of our model.

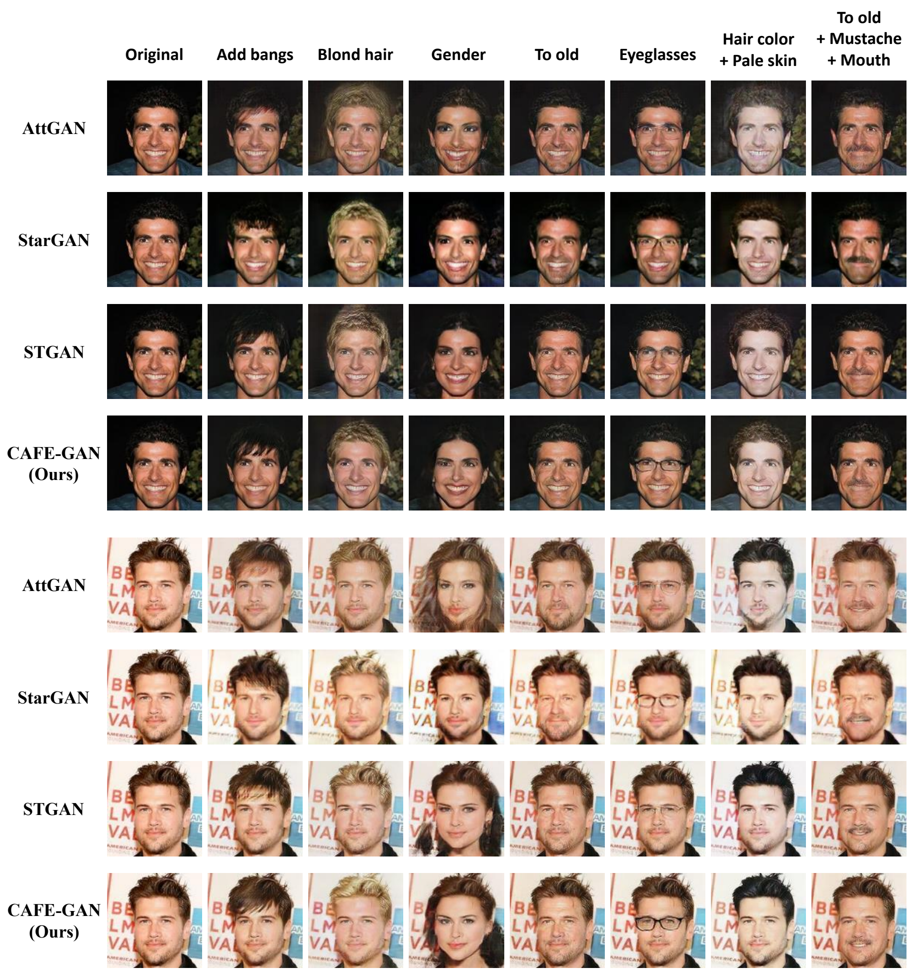

Fig. 1: Additional results of qualitative comparison on CelebA (128×128) images.

Fig. 2: Qualitative comparison with RelGAN on CelebA-HQ (256×256) images.

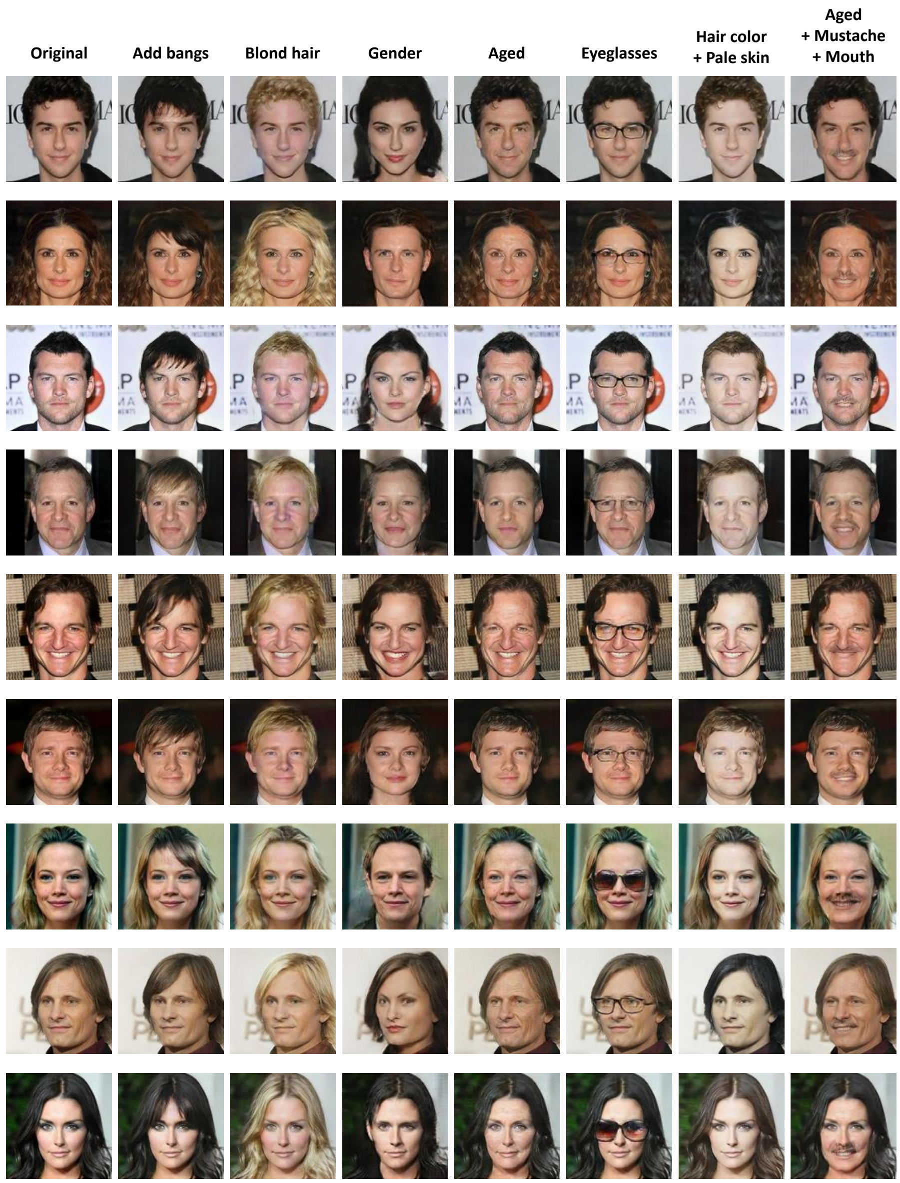

Fig. 3: Extra qualitative results of our model.

References

1. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
2. He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: AttGAN: Facial attribute editing by only changing what you want. IEEE Transactions on Image Processing (TIP) (2017)
3. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in Neural Information Processing Systems (NeurIPS) (2017)
4. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation (2017)
5. Liu, M., Ding, Y., Xia, M., Liu, X., Ding, E., Zuo, W., Wen, S.: STGAN: A unified selective transfer network for arbitrary image attribute editing. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
6. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
7. Wu, P.W., Lin, Y.J., Chang, C.H., Chang, E.Y., Liao, S.W.: Relgan: Multi-domain image-to-image translation via relative attributes. In: The IEEE International Conference on Computer Vision (ICCV) (2019)