Adaptive Nonlinear Latent Transformation for Conditional Face Editing —Supplmentary Material—

Zhizhong Huang[1] Siteng Ma[1] Junping Zhang[1] Hongming Shan[2,3]
[1] Shanghai Key Lab of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai 200433, China
[2] Institute of Science and Technology for Brain-inspired Intelligence and MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China
[3] Shanghai Center for Brain Science and Brain-inspired Technology, Shanghai 200031, China
{zzhuang19, stma21, jpzhang, hmshan}@fudan.edu.cn

1. Ablation study on the number of labeled samples.

A potential advantage of AdaTrans is the great generalization ability for the scenarios with few limited samples. This advantage benefits from the proposed disentangled learning strategy. To investigate the optimal and least labeled samples used for face editing, experiments are conducted to demonstrate the influence of the number of labeled samples.

Fig. A1 shows the quantitative results for this ablation. Interestingly, we found that training on 128 labeled samples achieves satisfactory performance versus 512 samples. AdaTrans failed with only 32 samples. Therefore, 128 samples are sufficient to train a good AdaTrans. Please refer to Fig. A4 for additional qualitative results.

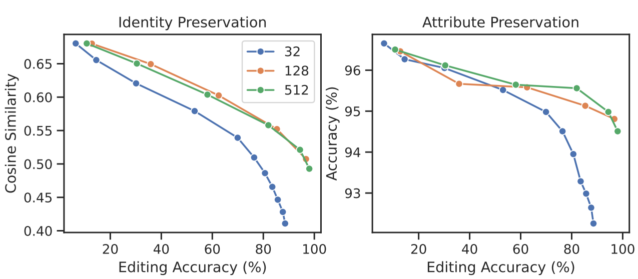


Figure A1. Quantitative results for the ablation of the number of labeled samples used for face editing.

2. Additional qualitative results.

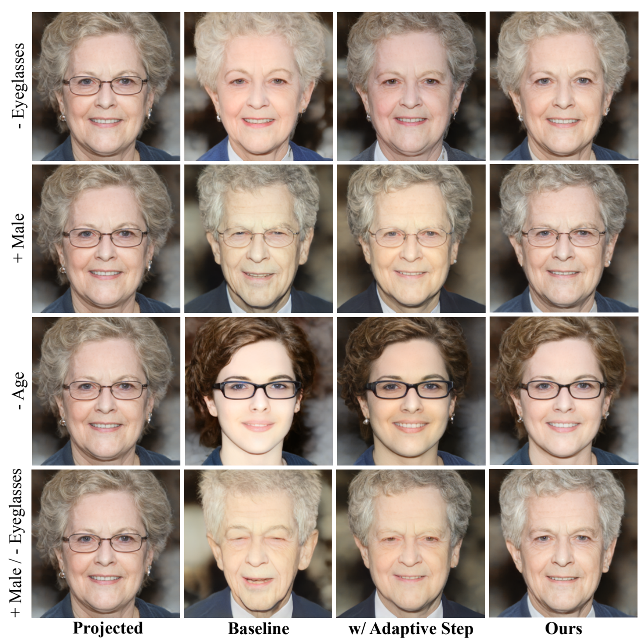


Figure A2. Qualitative results for the ablation of the proposed components in AdaTrans. The baseline has inevitably modified some unrelated facial attributes (*e.g.*, eyes close when editing male and eyeglasses), although the faces are successfully manipulated into target attributes. Thanks to the proposed latent regularization, the resultant faces are much more natural than previous two variants.

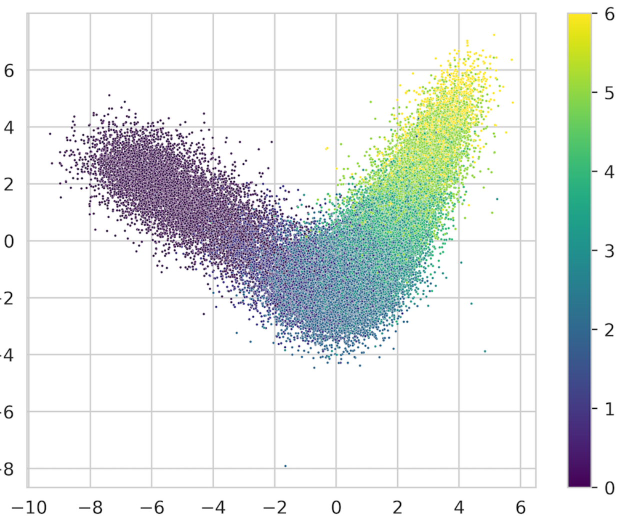


Figure A3. Visualization of 100k synthesized latent codes produced by StyleGAN2 mapping network, which is reduced to 2 dimensions by linear discriminant analysis. The points are colored by the age labels obtained using a pre-trained age classifier with 7 discrete age groups [1]. It validates that the transformation with finer ages cannot be accurately achieved by simple linear interpolation.

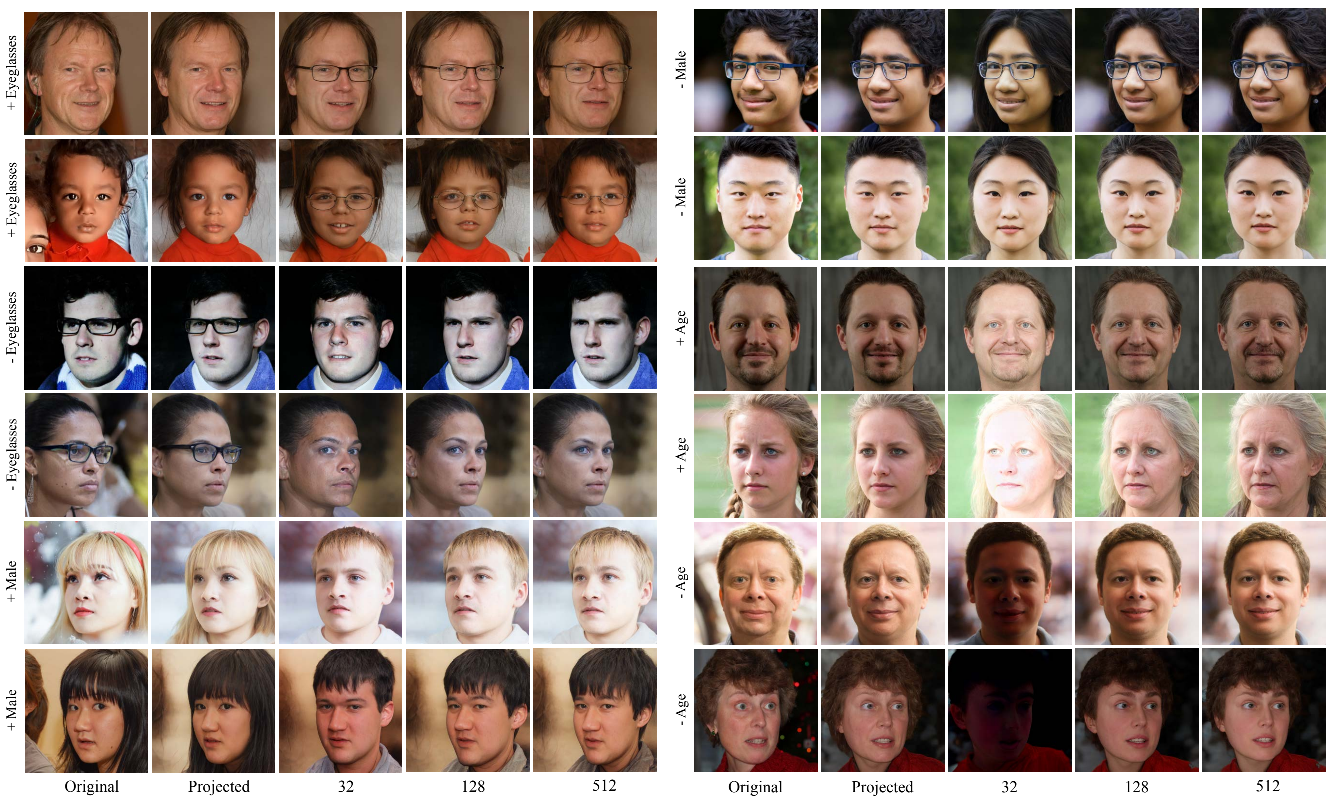


Figure A4. Qualitative results for the ablation of the number of labeled samples used for face editing. With only 128 labeled samples, AdaTrans can still achieve satisfactory face editing while 32 samples failed.

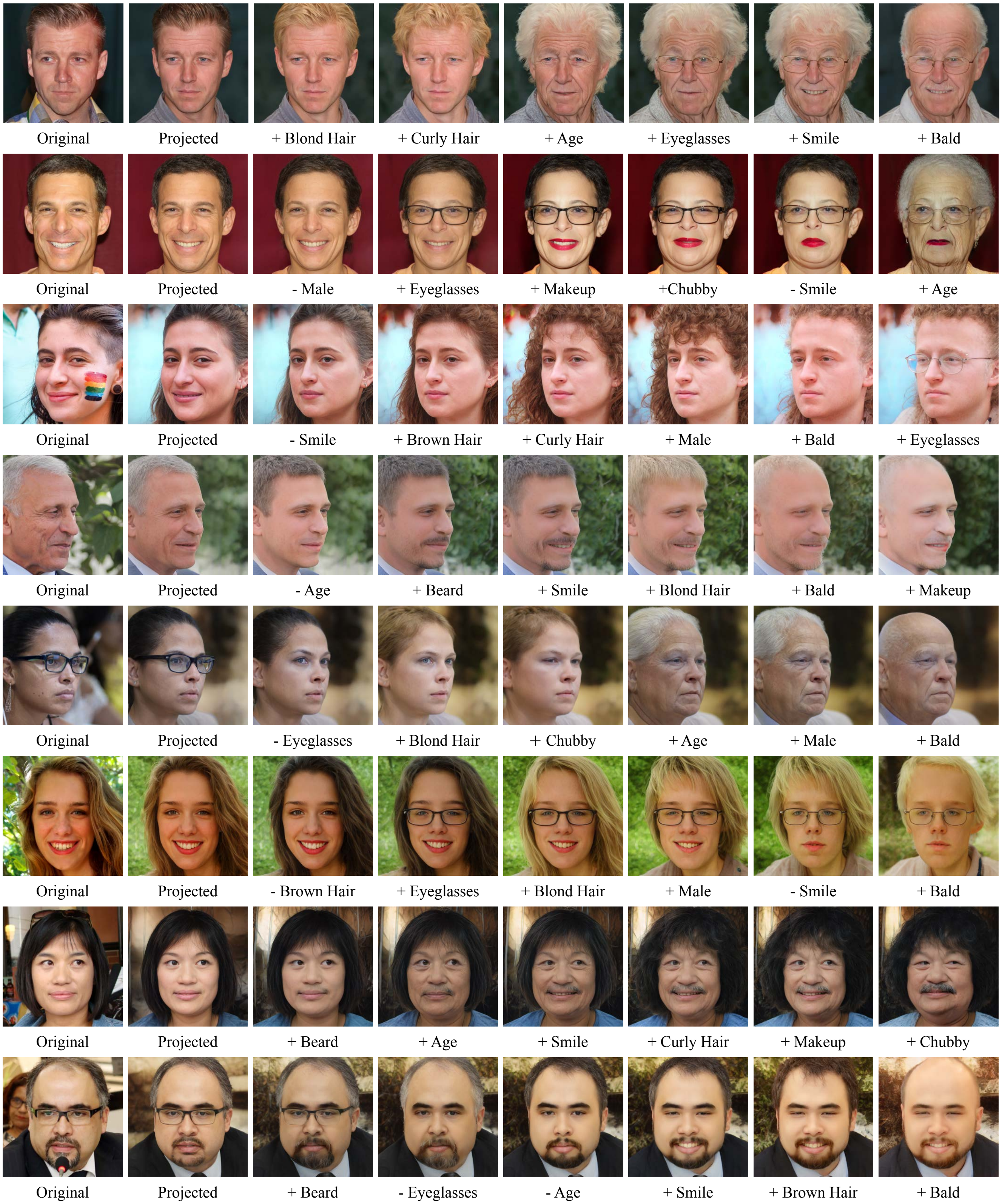


Figure A5. Additional results for sequential editing.

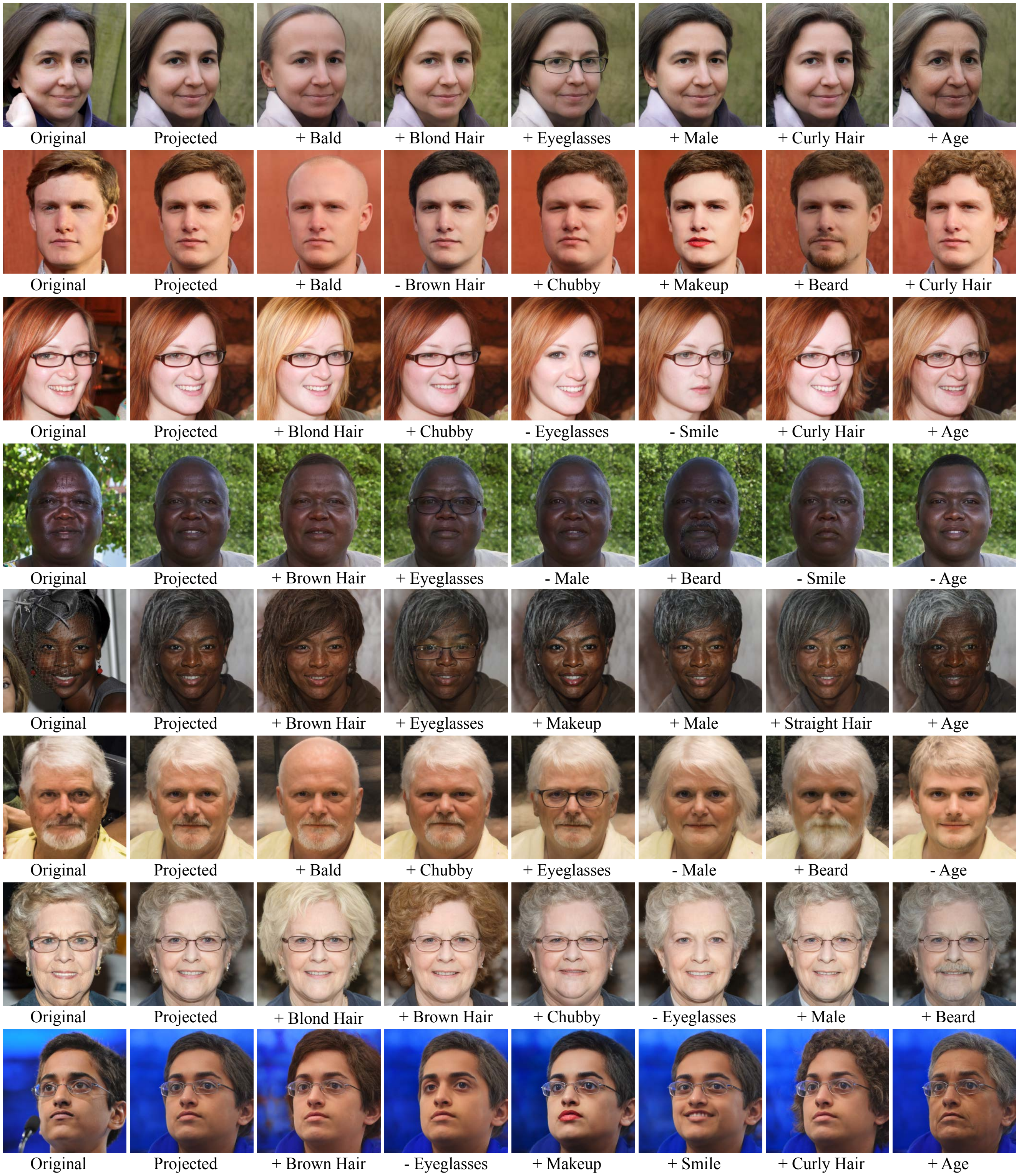
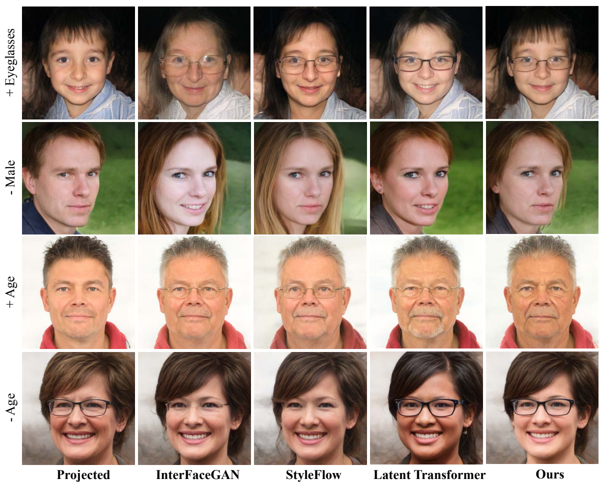
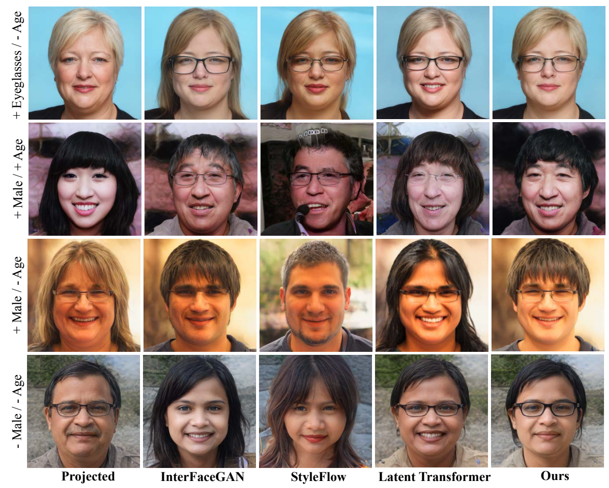


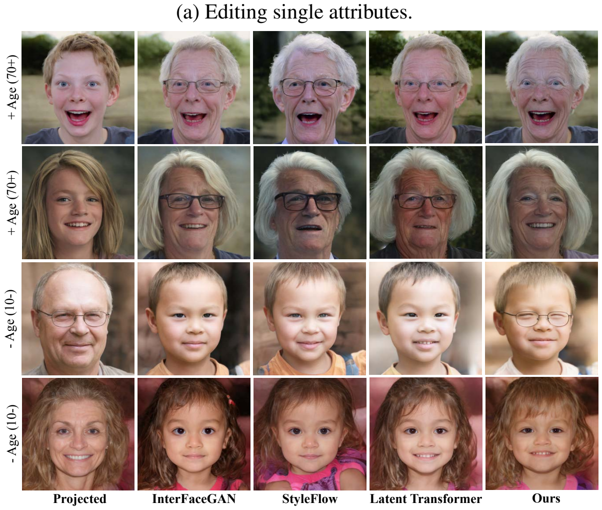
Figure A6. Additional results for editing a single attribute.



(a) Editing single attributes.



(b) Editing multiple attributes.



(c) Editing with large age gap.

(d) Editing with limited labeled data.

Figure A7. Additional results for the qualitative comparisons with recent methods for face editing under different experimental settings.

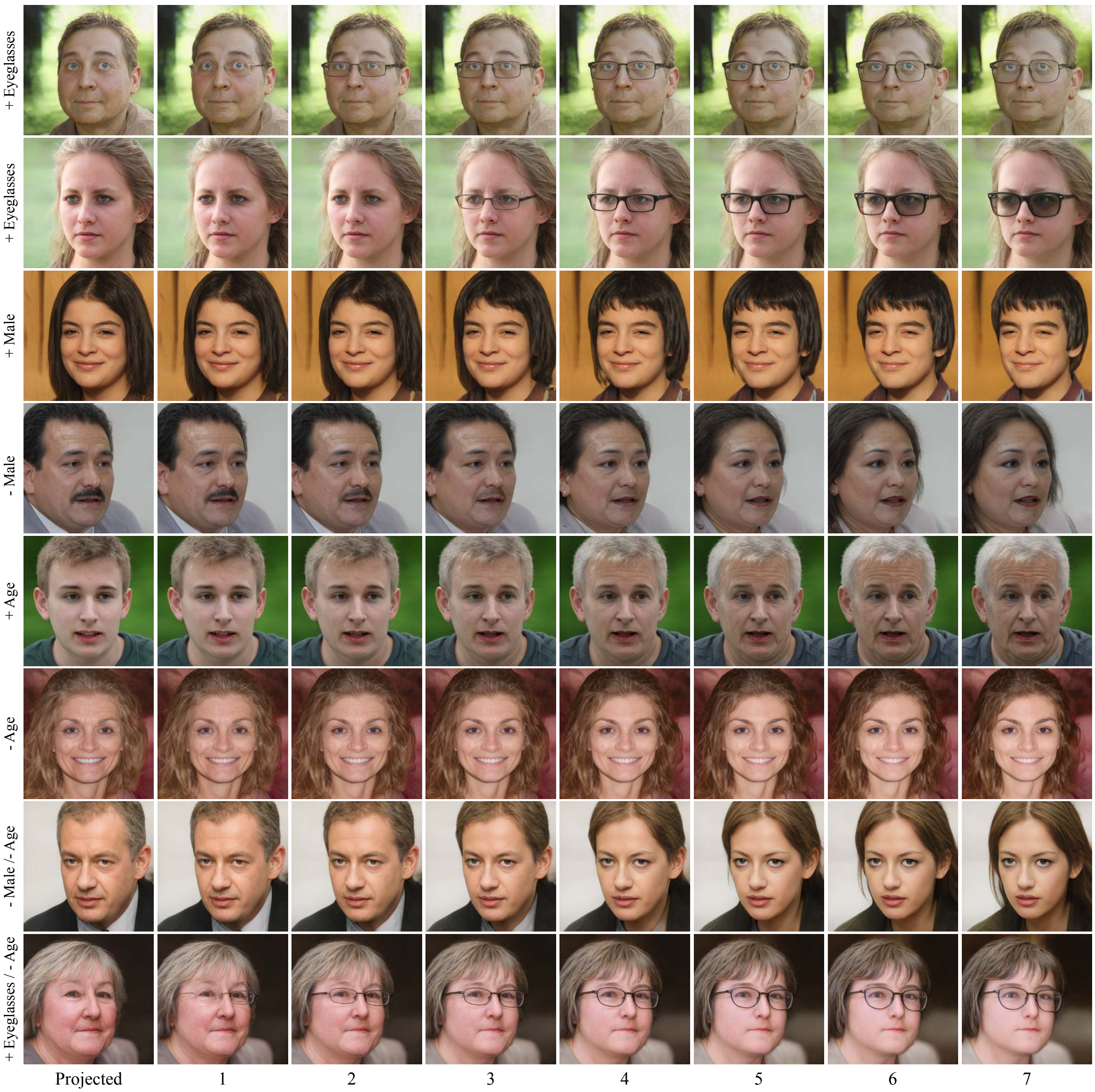


Figure A8. Visualization of the intermediate states for AdaTrans, with the step below. Although the maximum step is predefined as 5, AdaTrans can produce photorealistic results at different steps, and would not change unrelated attributes and identities when increasing the maximum step.

References

[1] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan age transformation synthesis. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 739–755. Springer, 2020.