CooGAN: A Memory-Efficient Framework for High-Resolution Facial Attribute Editing–Supplementary Material

Xuanhong Chen[1,2⋆], Bingbing Ni[1,2†], Naiyuan Liu[1‡], Ziang Liu[1], Yiliu Jiang[1], Loc Truong[1], and Qi Tian[3]

[1] Shanghai Jiao Tong University, China
{chen19910528,nibingbing,acemenethil,jiangyiliu}@sjtu.edu.cn,
{liunaiyuan27,ttanloc}@gmail.com
[2] Huawei HiSilicon, China
{chenxuanhong,nibingbing}@hisilicon.com
[3] Huawei, China
tian.qi1@huawei.com

1 Content

In this supplementary material we elaborate on the detailed structures of the networks used, including LSTU-Net, CooGAN and HR modified versions of STGAN and AttGAN. We also give more generated images, including LR comparison results on facial attribute editing task, season transfer task and HR results of facial attribute editing. The contents are given in the following sequence:

1. The comparison experiments and network structure of LSTU-Net. This section contains two sub-sections: 1) detailed network structure; 2) qualitative comparisons of results on facial attribute editing tasks and season transfer tasks.
2. The comparison experiments and network structure of complete CooGAN. This part has two sub-sections: 1) details of network structure; 2) comparison experiments and HR results demonstration.

Images are best viewed in color and zoomed in. Our source code will be made available in `https://github.com/XHChen0528/CooGAN`.

2 Extra Results of LSTU-Net

2.1 Detailed Network Structure

The structure of our LSTU-Net (mentioned in our paper) and the detailed parameters are given in Table. 1.

⋆ Work done during an internship at Huawei HiSilicon.
† Corresponding author: Bingbing Ni.
‡ Contributed to the work while he was a research assistant at Shanghai Jiao Tong University.

l	G_{enc}	G_{dec}	D_{adv}	D_{att}
1	Conv(64, 4, 2),BN,LeakyReLU	DeConv(3, 4, 2),Tanh	Conv(64, 4, 2),IN,LeakyReLU	
2	Conv(128, 4, 2),BN,LeakyReLU	DeConv(128, 4, 2),BN,ReLU	Conv(128, 4, 2),IN,LeakyReLU	
3	Conv(256, 4, 2),BN,LeakyReLU	DeConv(256, 4, 2),BN,ReLU	Conv(256, 4, 2),IN,LeakyReLU	
4	Conv(512, 4, 2),BN,LeakyReLU	DeConv(512, 4, 2),BN,ReLU	Conv(512, 4, 2),IN,LeakyReLU	
5	Conv(1024, 4, 2),BN,LeakyReLU	DeConv(1024, 4, 2),BN,ReLU	Conv(1024, 4, 2),IN,LeakyReLU	
6			$FC(1024)$	$FC(1024)$
6			$FC(1)$	$FC(c)$

Table 1. The network parameter of LSTU-Net. $Conv(dim, k, s)$ and $DeConv(dim, k, s)$ denote the convolutional layer and the transposed convolutional layer with output channel dim, kernel size k and stride s. BN represents batch normalization [1] and c is the number of trained attributes, IN represents instance normalization [2].

2.2 Results Comparison

The facial attribute editing results comparison is demonstrated in Fig. 2. These results are generated using the dataset CelebA [3]. The visual effect of results from our LSTU-Net are compared with AttGAN [4], StarGAN [5] and STGAN [6]. The second block in the Fig. 2 is the results generated using the original image in the paper of STGAN. It can be observed that our LSTU performs more effectively.

We conduct season transfer experiments on the season dataset [7] and compare our model performance with other classic methods. The comparison results of AttGAN, STGAN and our LSTU-Net are given in Fig. 3. More results of our LSTU-Net are given in Fig. 4. We also conduct survey on a crowdsourcing platform to acquire people's general opinions of our LSTU-Net results on season transfer against AttGAN and STGAN. In the survey, people are randomly given images generated by different models and asked to determine whether it is successful in season transfer. The results are given in Table. 2.

Method	*Spring*	*Summer*	*Autumn*	*Winter*
AttGAN	66.8%	78.2%	76.8%	78.4%
STGAN	71.3%	91.7%	84.1%	88.8%
Ours	**73.9**%	**92.1**%	**83.9**%	**87.4**%

Table 2. User study results of our LSTU-Net, STGAN and AttGAN on season transfer tasks.

We also give the experiment results in Fig. 1, exhibiting the importance of LSTU in the in the translation framework. We evaluate the function and property of skip connection in attribute editing tasks by studying its influence on translation accuracy of specific attributes and the overall performance.

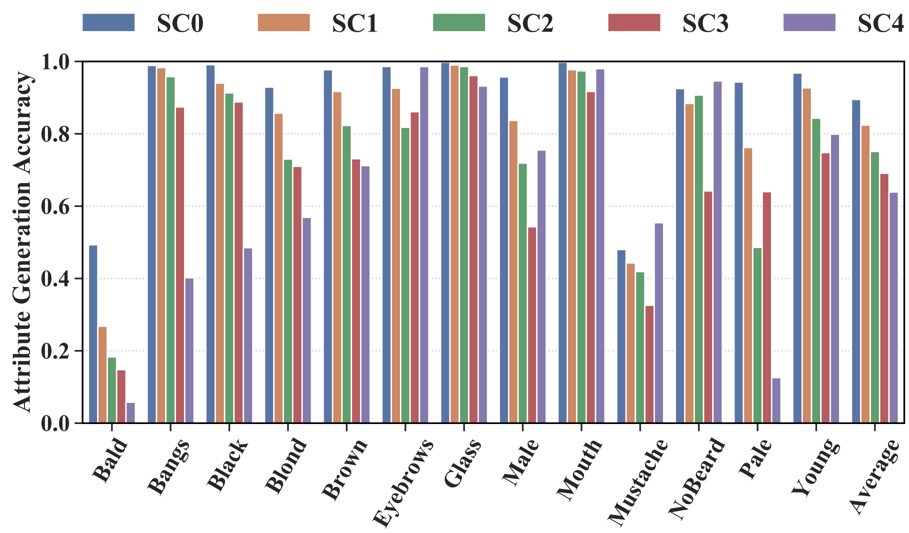

Fig. 1. The attribute translation accuracy of models with different numbers of skip connections.

3 Extra Results of CooGAN

3.1 Detailed Network Structure

The detailed parameters of the local module of CooGAN are given in Table. 3. Here we also give the network structures of AttGAN-HR and STGAN-HR in Table. 4, which are modified by us and used for comparison experiments.

3.2 Results Comparison

The HR results are generated using the CelebA-HQ [8] dataset. We compare the results of our CooGAN with AttGAN-HR and STGAN-HR. The comparison results are given in Fig. 5. More images generated by our CooGAN are given in Fig. 6 and Fig. 7.

l	G_{enc}	G_{dec}	D_{adv}	D_{att}
1	Conv(48,4,2),BN,LeakyReLU	DeConv(3,4,2),Tanh	Conv(48,4,2),IN,LeakyReLU	
2	Conv(96,4,2),BN,LeakyReLU	DeConv(96,4,2),BN,ReLU	Conv(96,4,2),IN,LeakyReLU	
3	Conv(192,4,2),BN,LeakyReLU	DeConv(192,4,2),BN,ReLU	Conv(192,4,2),IN,LeakyReLU	
4	Conv(384,4,2),BN,LeakyReLU	DeConv(384,4,2),BN,ReLU	Conv(384,4,2),IN,LeakyReLU	
5	Conv(768,4,2),BN,LeakyReLU	DeConv(768,4,2),BN,ReLU	Conv(768,4,2),IN,LeakyReLU	
6			$FC(1024)$	$FC(1024)$
7			$FC(1)$	$FC(c)$

l	G_{enc}	G_{dec}	D_{adv}	D_{att}
1	Conv(24,4,2),BN,LeakyReLU	DeConv(3,4,2),Tanh	Conv(24,4,2),IN,LeakyReLU	
2	Conv(48,4,2),BN,LeakyReLU	DeConv(48,4,2),BN,ReLU	Conv(48,4,2),IN,LeakyReLU	
3	Conv(96,4,2),BN,LeakyReLU	DeConv(96,4,2),BN,ReLU	Conv(96,4,2),IN,LeakyReLU	
4	Conv(192,4,2),BN,LeakyReLU	DeConv(192,4,2),BN,ReLU	Conv(192,4,2),IN,LeakyReLU	
5			$FC(256)$	$FC(256)$
6			$FC(1)$	$FC(c)$

Table 3. The network parameters of CooGAN. The first table is of the global module and the second table is of the local module. $Conv(dim, k, s)$ and $DeConv(dim, k, s)$ denote the convolutional layer and the transposed convolutional layer with output channel dim, kernel size k and stride s. BN represents batch normalization, IN represents instance normalization and c is the number of trained attributes.

l	$STGAN-HR$	
	encoder	decoder
1	Conv(32,4,2),BN,LeakyReLU	DeConv(3,4,2),Tanh
2	Conv(64,4,2),BN,LeakyReLU	DeConv(64,4,2),BN,ReLU
3	Conv(128,4,2),BN,LeakyReLU	DeConv(128,4,2),BN,ReLU
4	Conv(256,4,2),BN,LeakyReLU	DeConv(256,4,2),BN,ReLU
5	Conv(512,4,2),BN,LeakyReLU	DeConv(512,4,2),BN,ReLU
6	Conv(1024,4,2),BN,LeakyReLU	DeConv(1024,4,2),BN,ReLU
7	Conv(1024,4,2),BN,LeakyReLU	DeConv(1024,4,2),BN,ReLU

l	$AttGAN-HR$	
	encoder	decoder
1	Conv(64,4,2),BN,LeakyReLU	DeConv(3,4,2),Tanh
2	Conv(128,4,2),BN,LeakyReLU	DeConv(128,4,2),BN,ReLU
3	Conv(256,4,2),BN,LeakyReLU	DeConv(256,4,2),BN,ReLU
4	Conv(512,4,2),BN,LeakyReLU	DeConv(512,4,2),BN,ReLU
5	Conv(1024,4,2),BN,LeakyReLU	DeConv(1024,4,2),BN,ReLU
6	Conv(1024,4,2),BN,LeakyReLU	DeConv(1024,4,2),BN,ReLU
7	Conv(1024,4,2),BN,LeakyReLU	DeConv(1024,4,2),BN,ReLU

Table 4. The generator network parameters of STGAN-HR and AttGAN-HR.

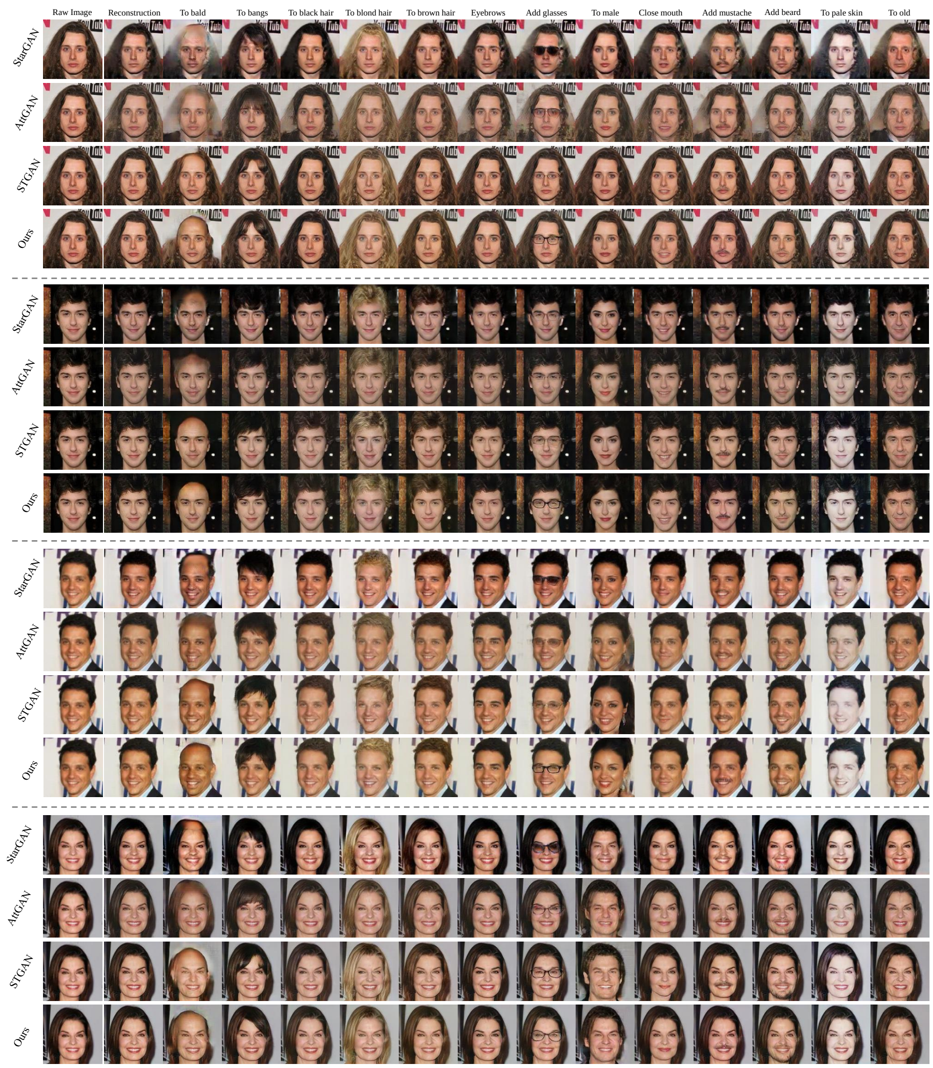

Fig. 2. LR (128 × 128) facial attribute editing results of StarGAN, AttGAN, STGAN and our LSTU-Net.

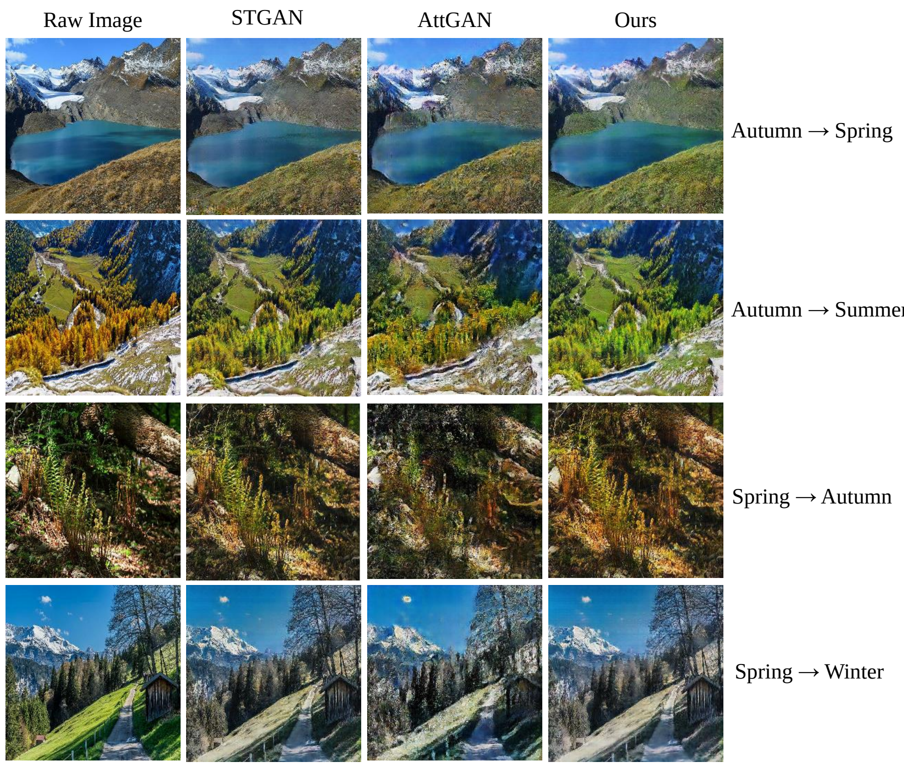

Fig. 3. Season transfer results comparison of AttGAN, STGAN and our LSTU-Net.

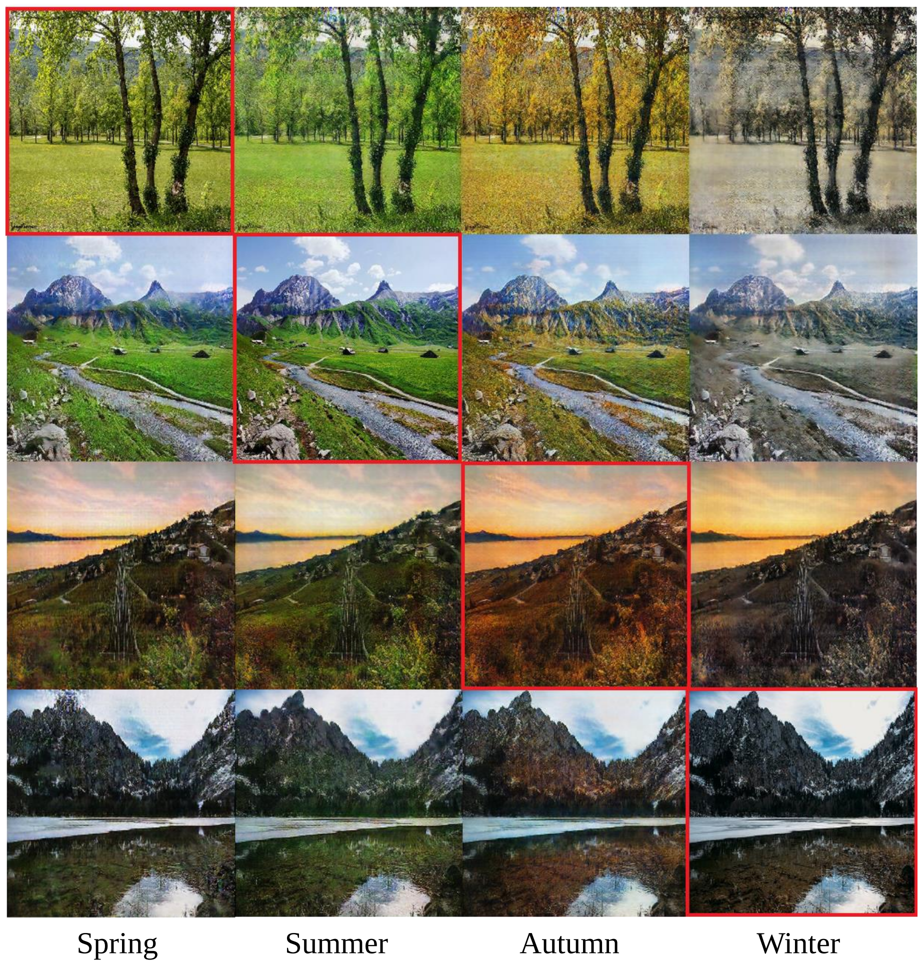

Fig. 4. Season transfer results of our LSTU-Net. The diagonal ones are original images.

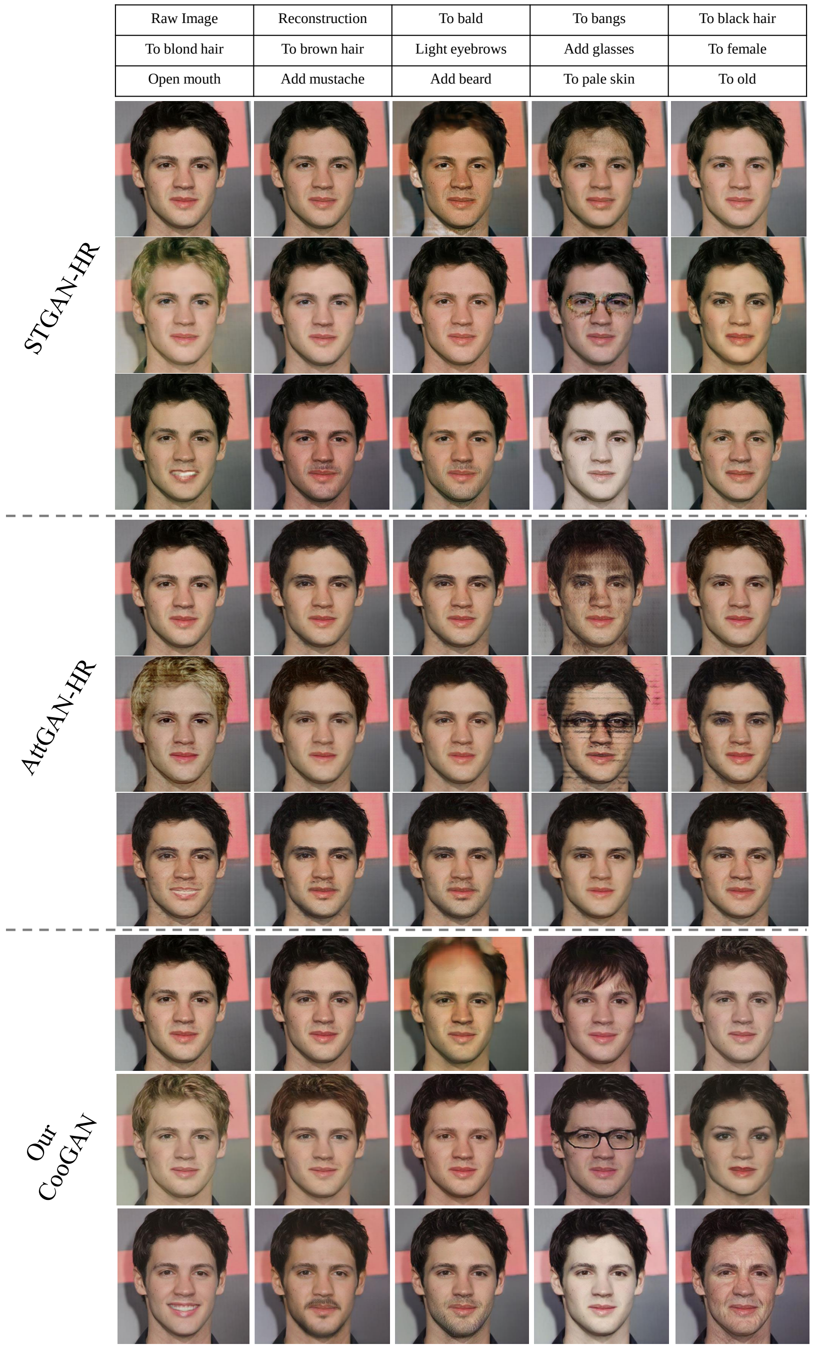

Fig. 5. HR (768× 768) results comparison of AttGAN-HR, STGAN-HR and our CooGAN.

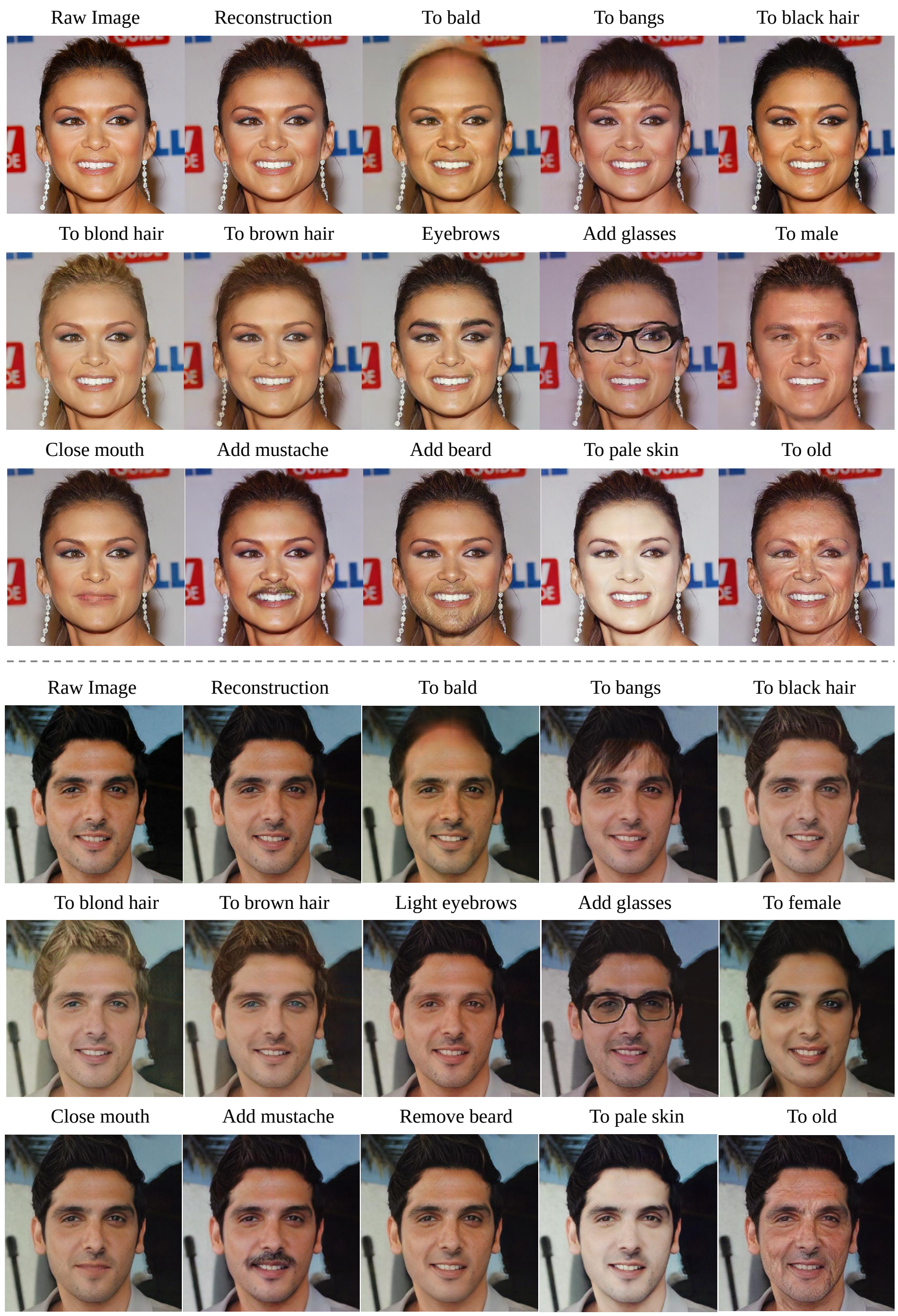

Fig. 6. HR (768× 768) results of our CooGAN.

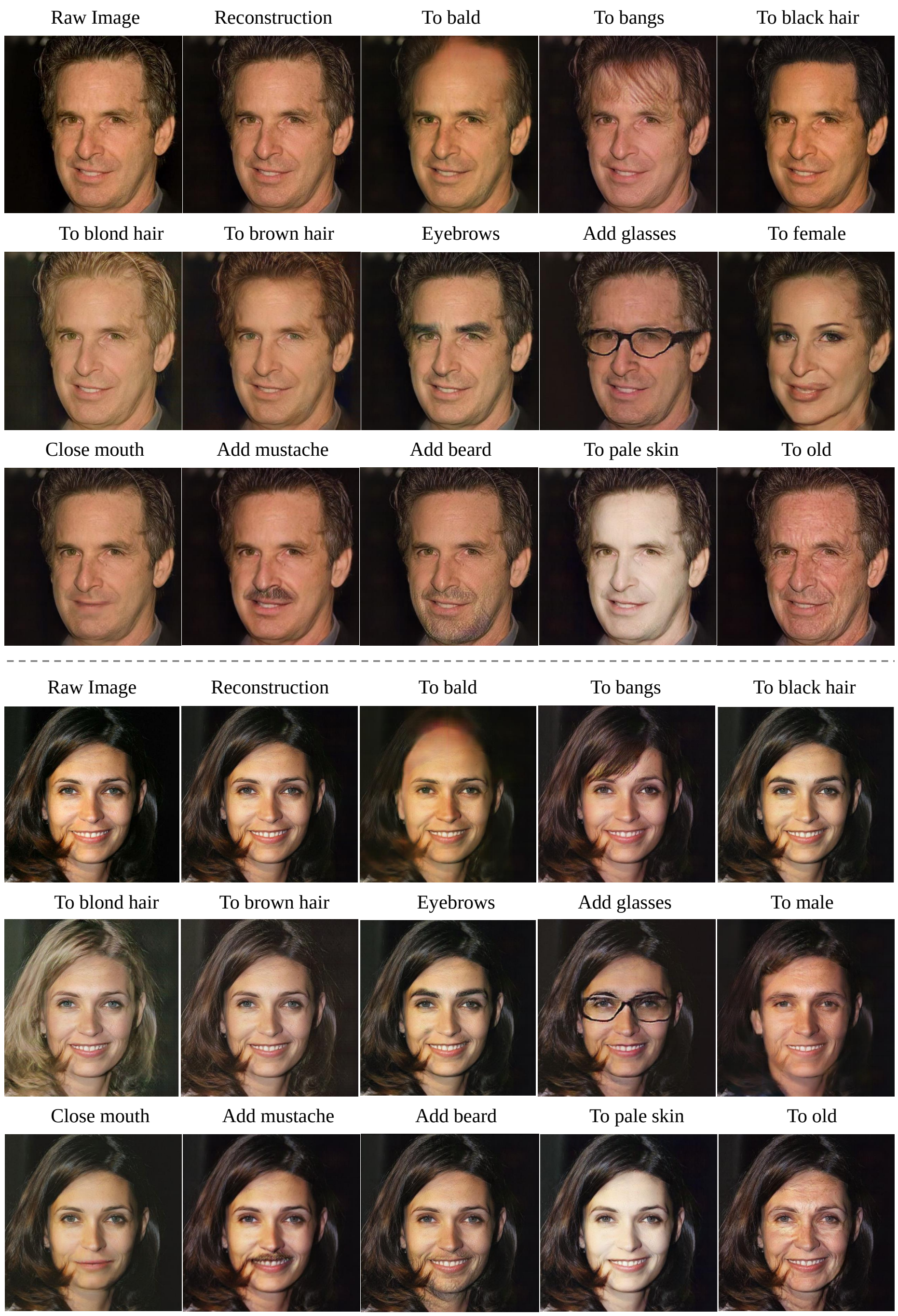

Fig. 7. HR (768× 768) results of our CooGAN.

References

1. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Bach, F.R., Blei, D.M., eds.: Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015. Volume 37 of JMLR Workshop and Conference Proceedings., JMLR.org (2015) 448–456
2. Ulyanov, D., Vedaldi, A., Lempitsky, V.S.: Instance normalization: The missing ingredient for fast stylization. CoRR **abs/1607.08022** (2016)
3. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV). (December 2015)
4. He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: Attgan: Facial attribute editing by only changing what you want. IEEE Trans. Image Processing **28**(11) (2019) 5464–5478
5. Choi, Y., Choi, M., Kim, M., Ha, J., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. (2018) 8789–8797
6. 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: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019. (2019) 3673–3682
7. Anoosheh, A., Agustsson, E., Timofte, R., Gool, L.V.: Combogan: Unrestrained scalability for image domain translation. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018. (2018) 783–790
8. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. In: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. (2018)