Unpaired Portrait Drawing Generation via Asymmetric Cycle Mapping (Supplementary Material)

Ran Yi, Yong-Jin Liu*
CS Dept, BNRist
Tsinghua University, China
{yr16,liuyongjin}@tsinghua.edu.cn

Yu-Kun Lai, Paul L. Rosin
School of Computer Science and Informatics
Cardiff University, UK
{LaiY4,RosinPL}@cardiff.ac.uk

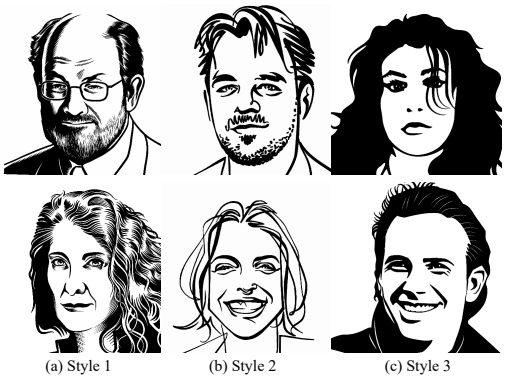
S1. Overview

This supplementary material includes:

- more style examples in the training set (Section S2);
- three more ablation studies (Section S3);
- a GAN metric evaluation to measure the similarity between the distributions of two drawing sets: one is the set of generated APDrawing and the other is the set of collected true drawings (Section S4);
- all evaluation material used in the user study in Section 4.3 of the main paper (Section S5.1);
- comparison with APDrawingGAN (Section S5.2);
- more test results on other face dataset (Section S5.3).

S2. More Style Examples in the Training Set

In the main paper, we introduce the selected three representative styles from the collected data and show three examples in Figure 2: the first style is from Yann Legendre and Charles Burns where parallel lines are used to draw shadows; the second style is from Kathryn Rathke where few dark regions are used and facial features are drawn using simple flowing lines; the third style is from vectorportal.com where continuous thick lines and large dark regions are utilized. Here we provide more examples in Figure S1.



(a) Style 1 (b) Style 2 (c) Style 3

Figure S1. More examples for the three styles in the training set.

S3. Three More Ablation Studies

In Section 4.4 of the main paper, we study three key factors in our model, i.e., relaxed cycle-consistency loss, local discriminators and HED edge extraction. Here, we present three more ablation studies: one focuses on the style feature and style loss, one focuses on the truncation loss, and the other focuses on how face region information is utilized in the discriminator.

In our proposed method, when inputting a face photo and a style feature, the system outputs an APDrawing with style specified by the style feature. If we remove the style feature input and style loss from our system, when inputting a face photo, an APDrawing can still be output. However, since the network is trained with mixed data, the output frequently exhibits different or mixed styles in different facial regions in an unpredictable way. Three examples are illustrated in Figure S2, in which all three photos contain a man face with beards. On the top of Figure S2(b), the generated APDrawing shows a parallel line style in the beard region. In the middle of Figure S2(b), thick line and dark region style appears in the eyes and hat regions, respectively. At the bottom of Figure S2(b), the generated APDrawing shows mixed styles. In comparison, as illustrated in Figures S2(c-e), after introducing style feature and style loss, our method can generate APDrawing results for each distinctive style, specified by the input style feature.

We further study the role of truncation loss and two examples are shown in Figure S3. The truncation loss is designed to prevent the generated drawings from hiding information in small values. Without the truncation loss, the results sometimes do not draw full outlines of facial features (e.g., nose). As shown in Figure S3(b), the nose in the first row lacks the left outline and the nose in the second

*Corresponding author

Figure S2. Ablation study on style feature input and style loss. From left to left to right: input photos, results of removing style feature input and style loss, our results (style1), our results (style2) and our results (style3).

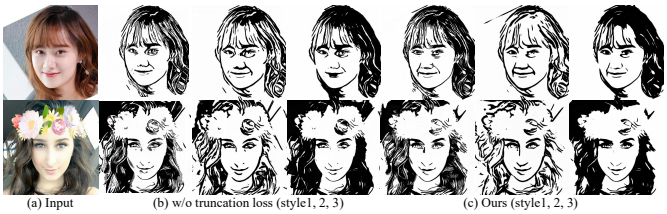


Figure S3. Ablation study on truncation loss. From left to right: input photos, results of removing truncation loss (style1, 2, 3), and our results (style1, 2, 3).

Figure S4. Comparison of results with our local discriminators (c) and replacing them with a new channel (b) for input photos (a).

row lacks the right outline. In comparison, by adding the truncation loss, our system can generate complete outlines of different facial features.

We further compare with the ablation of replacing local discriminators with a single discriminator with a new channel containing face region information. Our experiment shows that the results of this ablation are worse than those by our method, with partial facial features missing or messy (Figure S4). Also note that the face parsing masks are computed by an off-the-shelf face parsing network, and dilating the parsed eyes/nose/lips regions to make them cover the facial features. Some examples of the face parsing masks are shown in Figure S5. As can be seen, our system does not require accurate parsing masks.

S4. GAN Metric Evaluation

We adopt the Fréchet Inception Distance (FID) [3] to evaluate the similarity between the distributions of two

Figure S5. Examples of face parsing masks.

Table S1. Fréchet Inception Distance (FID) of our method and two multi-modal image translation methods. The FID values are computed between the set of generated APDrawing of each style and the collected true drawings of the corresponding style.

Methods	Style1	Style2	Style3
MUNIT [5]	206.2	281.2	248.8
ComboGAN [1]	151.6	163.3	142.0
Ours	**88.3**	**139.0**	**108.2**

drawing sets — one is the set of generated APDrawing for one style and the other is the set of collected true drawings for this style — where lower FID indicates better similarity. We translate all face photos in the test set into three styles of APDrawing (in our method this is achieved by changing the input style feature). The FID values between the set of generated APDrawing of each style and the collected drawings of the corresponding style are computed and summarized in Table S1. The results show that compared with the other multi-modal generation methods (MUNIT [5] and ComboGAN [1]), our method has lower FID on all three styles, indicating our method generates a closer distribution to the distribution of true drawings.

S5. More Results

S5.1. Material in the User Study

In Section 4.2 of the main paper, we compare our method with state-of-the-art methods in neural style transfer and image translation. In Section 4.3 of the main paper, we conduct a user study in which users sort the results of four methods (LinearStyleTransfer (LST) [7], ComboGAN [1], CycleGAN [9] and our method). In total 60 groups of images/drawings are evaluated in this user study and we show all of them in Figures S7, S8, S9, S10, S11, S12. Note that all these 60 groups are randomly chosen from the test set. Our method outperforms the other three methods in most groups in terms of style similarity, face structure preservation and image visual quality.

The results of the user study summarized in Section 4.3 of the main paper also demonstrate the advantage of our method, where 64.2% votes chose our method to be the best among the four methods. We present the user votes (rank1 percentage of each method) for each group in the last column of Figures S7 to S12.

S5.2. Comparison with APDrawingGAN

APDrawingGAN [8] is a deep neural network model specially designed for APDrawing generation by using a hi-

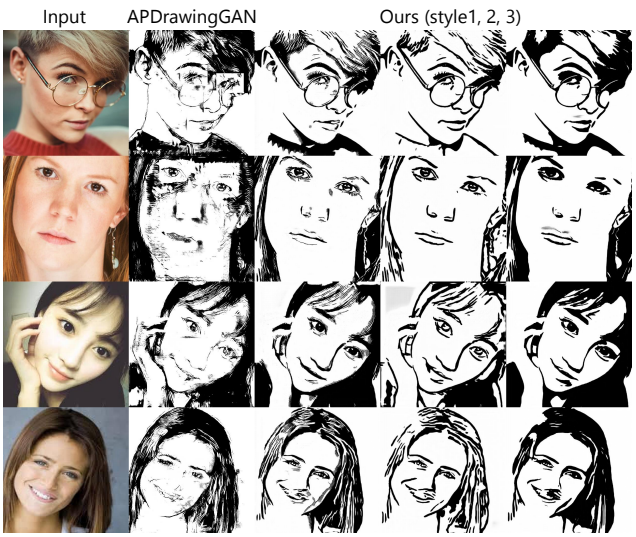


Figure S6. Comparisons of APDrawingGAN and our method on challenging photos with arbitrary head orientation. From left to right: input photos, APDrawingGAN results, and our results (style1, 2, 3).

erarchical structure and a distance transform loss. However, this method requires *paired* training data and cannot adapt well to face photos with unconstrained lighting in the wild due to the limited availability of paired training data. In comparison, our method uses *unpaired* training data, and then makes it possible to include more challenging photos into the training set. Therefore, our method can generate high quality APDrawings for challenging photos under various conditions. We compare the visual quality of APDrawingGAN and our method using some challenging examples as illustrated in Figure S13. These challenging examples include unconventional light conditions (1st-5th rows), unconventional expression or taking accessories like sunglasses (6th-8th rows), or blurry looking (9th-10th rows, zoom in to check). APDrawingGAN generates messy results on these challenging photos, while our method generates high-quality APDrawings with much better visual effect.

Moreover, APDrawingGAN uses a hierarchical network structure that feeds local rectangle regions around eyes, nose and mouth centers into local generators and discriminators. This setting cannot tolerate a large head tilt and requires that its input photos are in the upright orientation (i.e., the photo needs to be rotated so that the two eyes are on a horizontal line). Then the local regions of eyes, nose and mouth can be covered by rectangle regions. In comparison, although our model also has local discriminators, we use face masks (obtained from a face parsing network [2]), and the inputs to local discriminators are the masked eyes, nose, mouth regions. Therefore our method does not need the input images to be adjusted into the upright orientation. Comparisons of APDrawingGAN and ours on face photos with arbitrary head orientation are shown in Fig. S6. The results show that APDrawingGAN often generates messy results and some boundaries of rectangle local regions are clearly visible, whereas our results are clean and have good visual quality.

S5.3. More Tests on the CelebAMask-HQ Dataset

In the main paper, we test our model on photos collected from Internet. Here, we further test our method on photos from CelebAMask-HQ Dataset [6]. The results are summarized in Figure S14, and show our method generates high quality results with good image and line quality on the CelebAMask-HQ Dataset.

References

[1] Asha Anoosheh, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. ComboGAN: unrestrained scalability for image domain translation. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, pages 783–790, 2018. 2, 4, 5, 6, 7, 8, 9

[2] Shuyang Gu, Jianmin Bao, Hao Yang, Dong Chen, Fang Wen, and Lu Yuan. Mask-guided portrait editing with conditional GANs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3436–3445, 2019. 3

[3] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems, (NeurIPS)*, pages 6629–6640, 2017. 2

[4] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007. 10

[5] Xun Huang, Ming-Yu Liu, Serge J. Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *15th European Conference (ECCV)*, pages 179–196, 2018. 2

[6] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. MaskGAN: Towards diverse and interactive facial image manipulation. *CoRR*, abs/1907.11922, 2019. 3, 11

[7] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. Learning linear transformations for fast image and video style transfer. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3809–3817, 2019. 2, 4, 5, 6, 7, 8, 9

[8] Ran Yi, Yong-Jin Liu, Yu-Kun Lai, and Paul L. Rosin. APdrawingGAN: Generating artistic portrait drawings from face photos with hierarchical GANs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10743–10752, 2019. 2, 10

[9] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017. 2, 4, 5, 6, 7, 8, 9

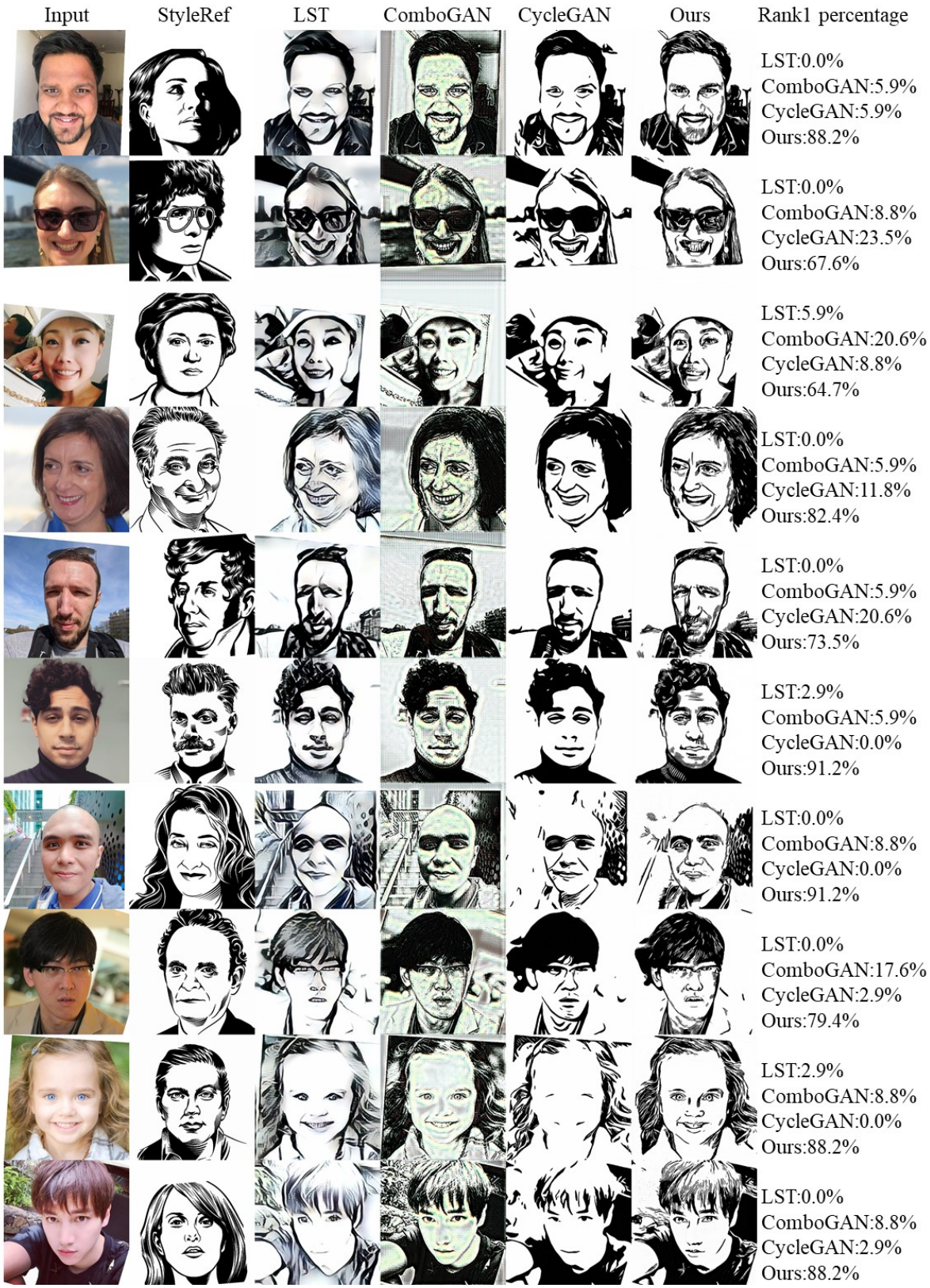


Figure S7. More qualitative comparisons of style1. From left to right: input face photos, the randomly chosen style images from style1 (used as the style input of LST), LinearStyleTransfer(LST) [7] results, ComboGAN [1] results, CycleGAN [9] results and our results. The last column shows user votes, i.e. the rank1 percentage of each method, for each group.

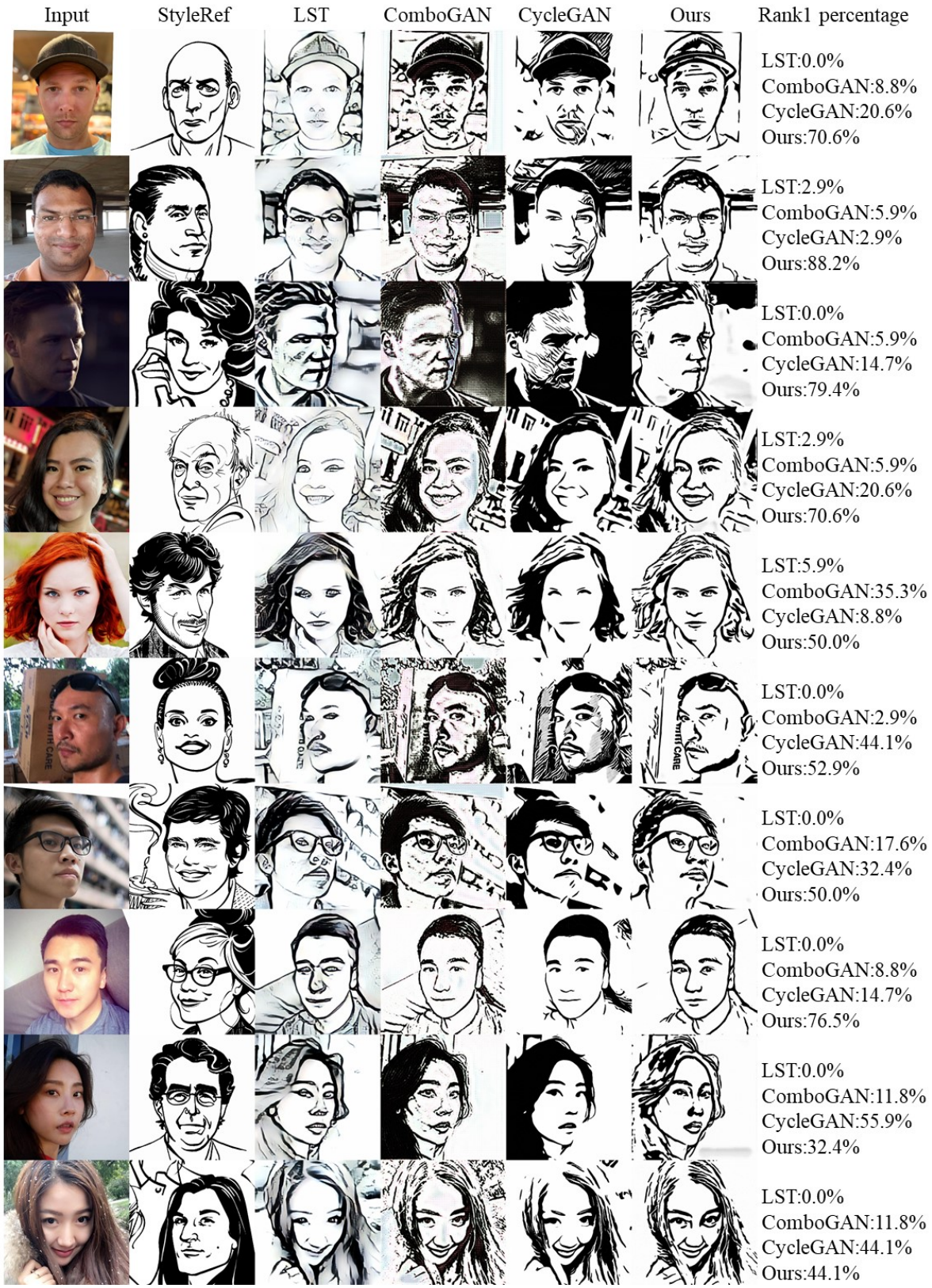


Figure S8. More qualitative comparisons of style2. From left to right: input face photos, the randomly chosen style images from style2 (used as the style input of LST), LinearStyleTransfer(LST) [7] results, ComboGAN [1] results, CycleGAN [9] results and our results. The last column shows user votes, i.e. the rank1 percentage of each method, for each group.

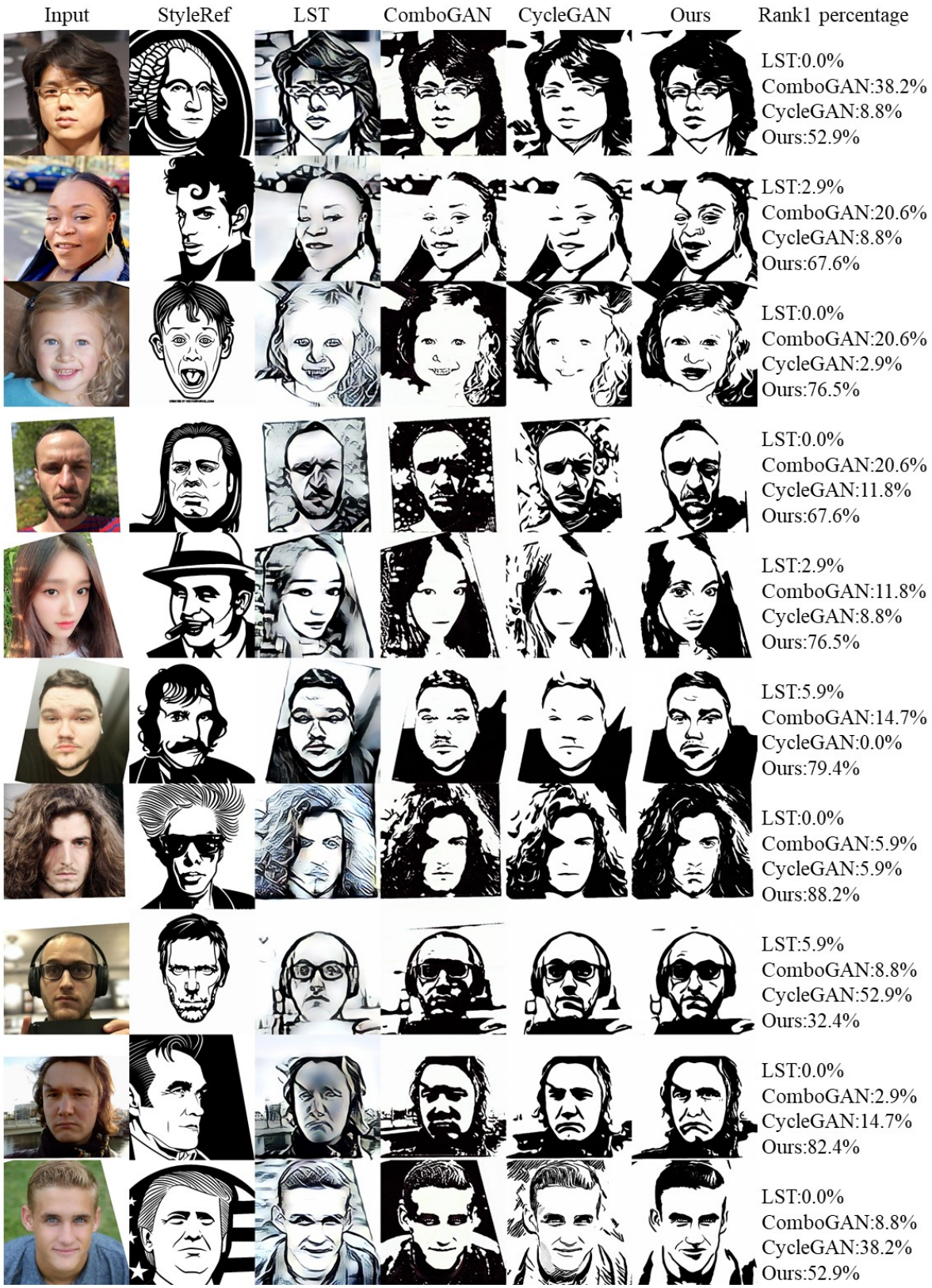


Figure S9. More qualitative comparisons of style3. From left to right: input face photos, the randomly chosen style images from style3 (used as the style input of LST), LinearStyleTransfer(LST) [7] results, ComboGAN [1] results, CycleGAN [9] results and our results. The last column shows user votes, i.e. the rank1 percentage of each method, for each group.

Figure S10. More qualitative comparisons of style1. From left to right: input face photos, the randomly chosen style images from style1 (used as the style input of LST), LinearStyleTransfer(LST) [7] results, ComboGAN [1] results, CycleGAN [9] results and our results. The last column shows user votes, i.e. the rank1 percentage of each method, for each group.

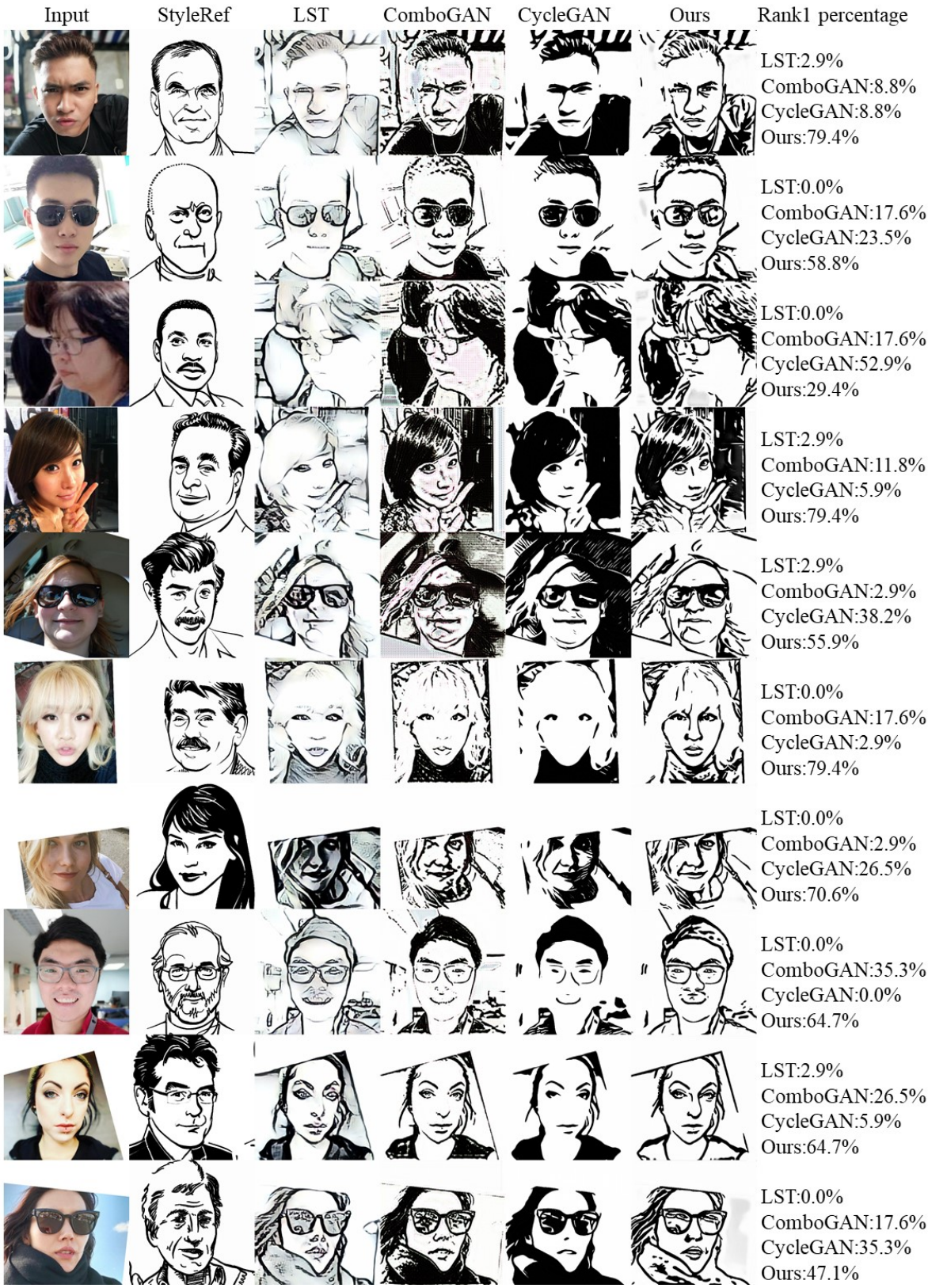


Figure S11. More qualitative comparisons of style2. From left to right: input face photos, the randomly chosen style images from style1 (used as the style input of LST), LinearStyleTransfer(LST) [7] results, ComboGAN [1] results, CycleGAN [9] results and our results. The last column shows user votes, i.e. the rank1 percentage of each method, for each group.

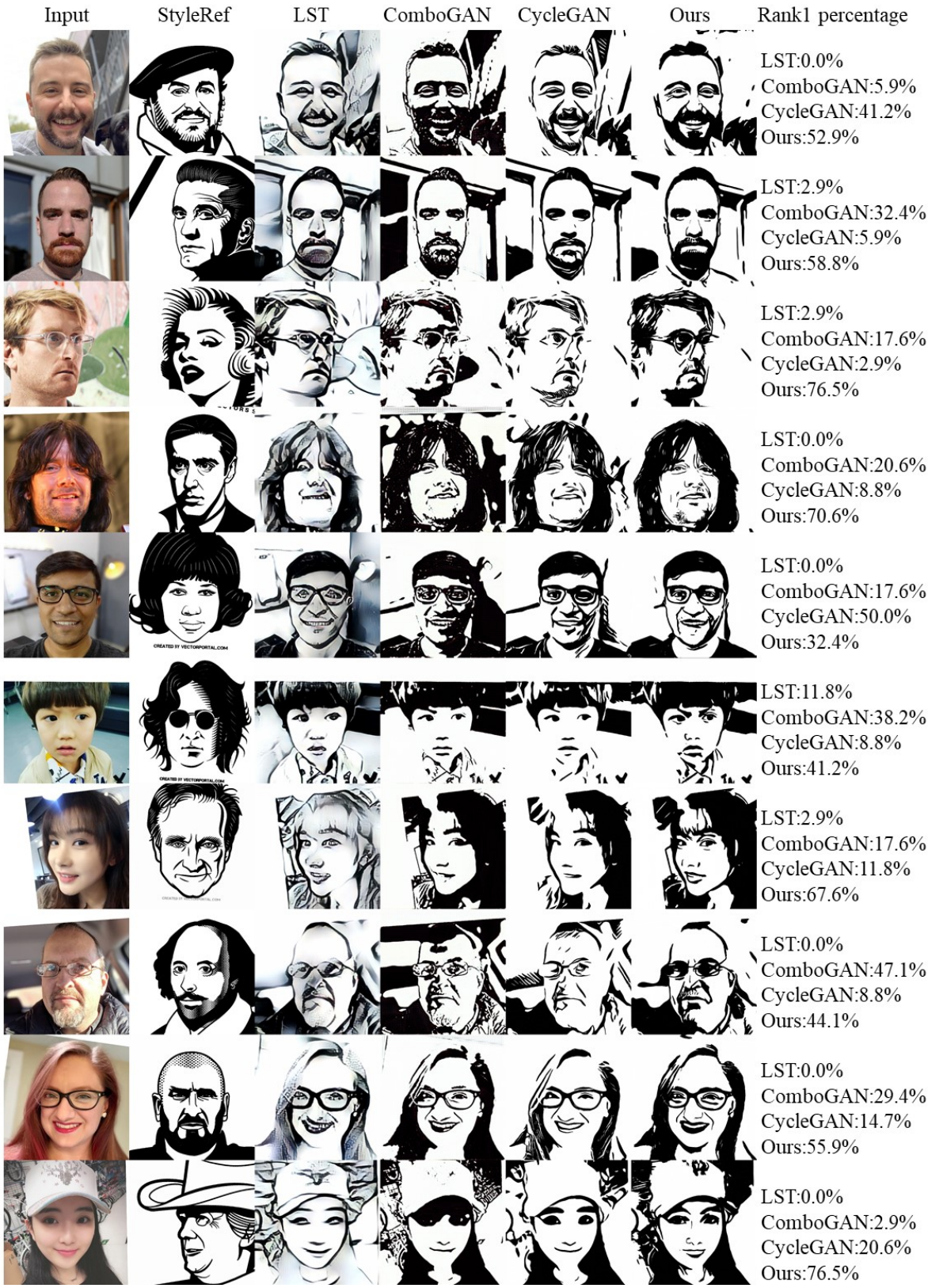


Figure S12. More qualitative comparisons of style3. From left to right: input face photos, the randomly chosen style images from style3 (used as the style input of LST), LinearStyleTransfer(LST) [7] results, ComboGAN [1] results, CycleGAN [9] results and our results. The last column shows user votes, i.e. the rank1 percentage of each method, for each group.

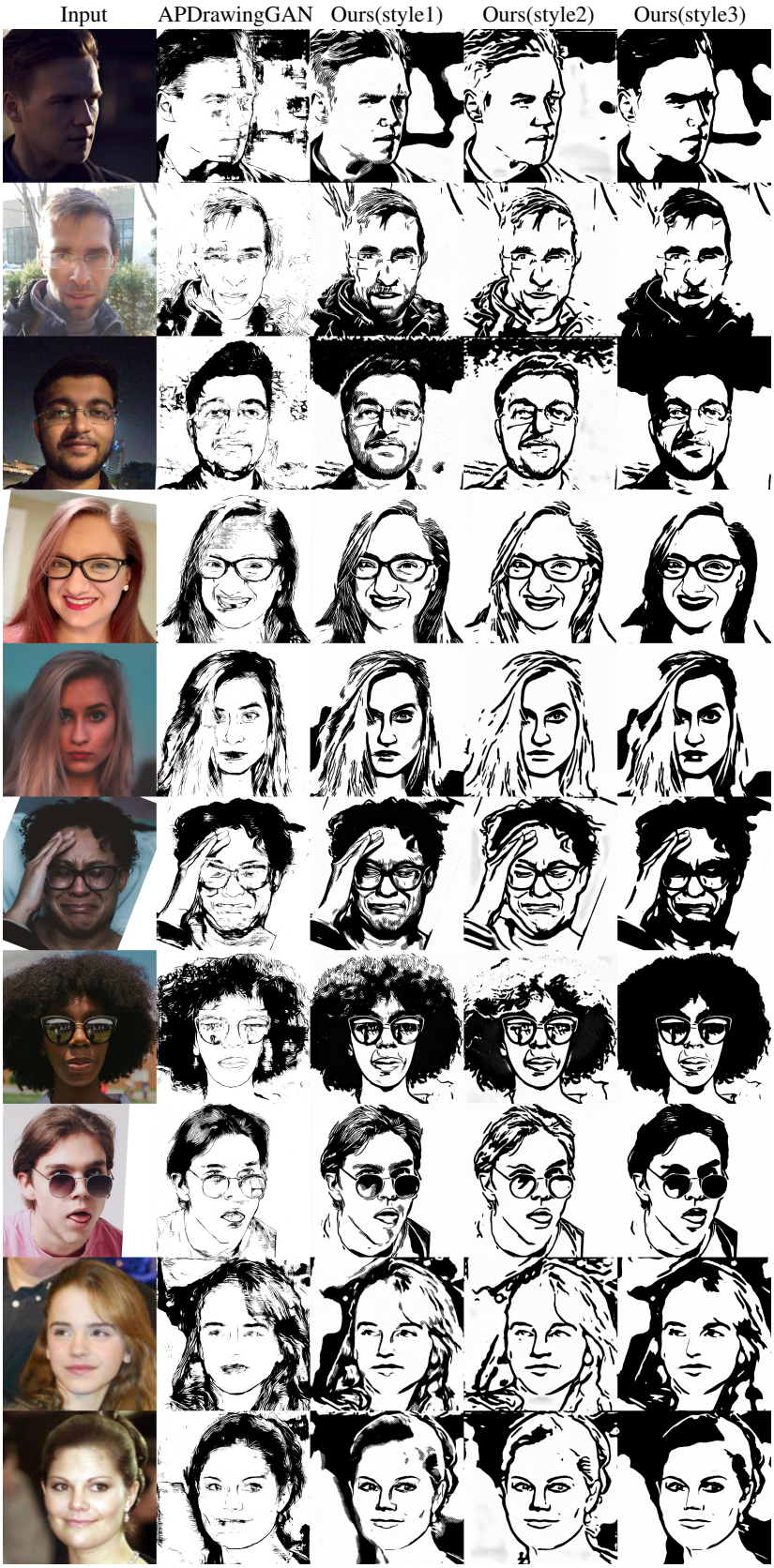


Figure S13. Comparison of APDrawingGAN [8] and our method on face photos under some challenging situations. From left to right: input face photos, APDrawingGAN [8] results, our results (style1), our results (style2), our results (style3). The last two input photos are from LFW Face Database [4].

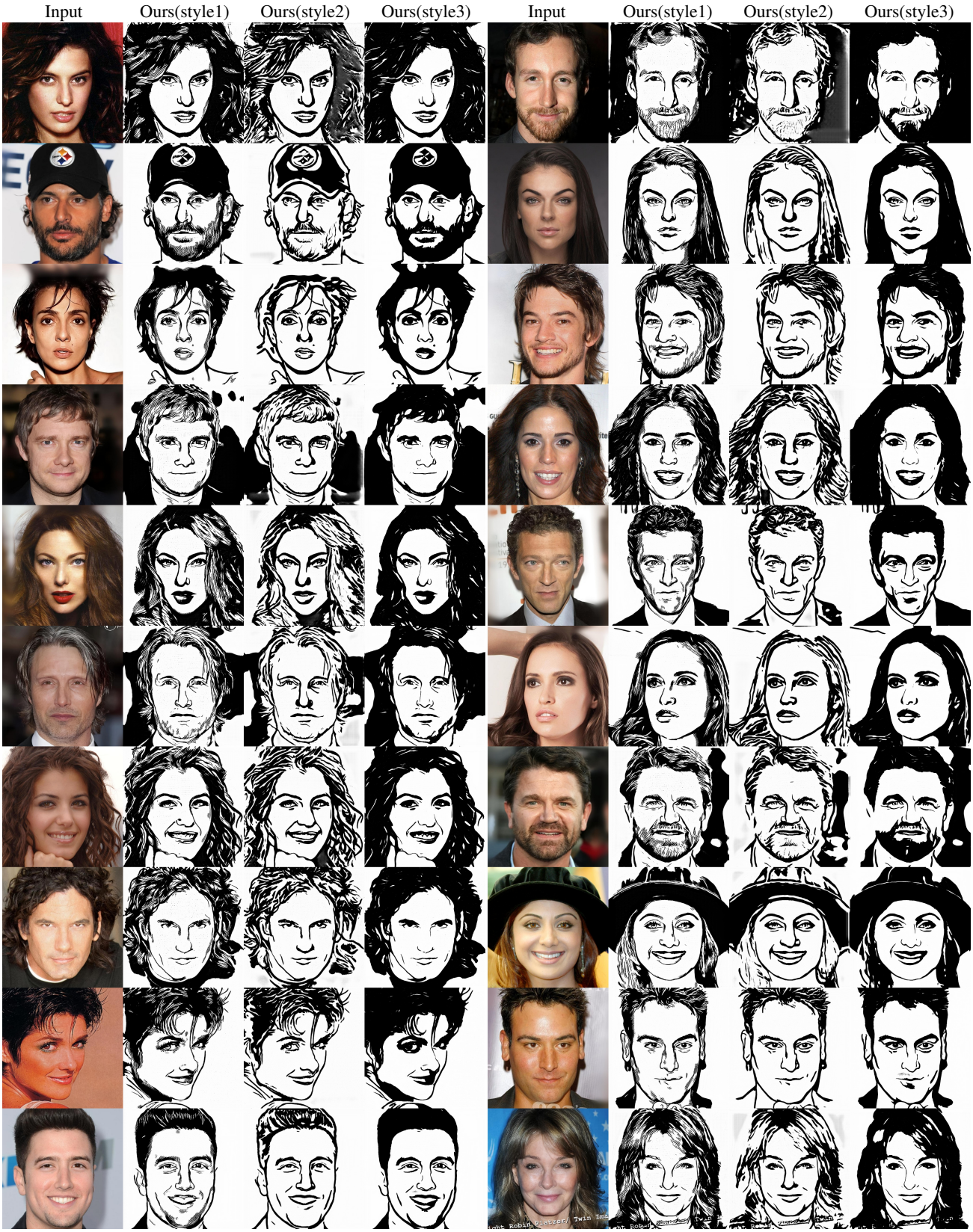


Figure S14. More test results on CelebAMask-HQ Dataset [6]. From left to right: input face photos, our results (style1), our results (style2), our results (style3), input face photos, our results (style1), our results (style2), our results (style3).