APDrawingGAN: Generating Artistic Portrait Drawings from Face Photos with Hierarchical GANs (Supplemental 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

In this supplemental material, more experimental results are provided, including:

- more details about APDrawing dataset construction (Section S2);
- more factors in the ablation study (Section S3);
- a user study to subjectively evaluate CycleGAN, Pix2Pix and APDrawingGAN (Section S4);
- more qualitative results of comparison with six state-of-the-art style transfer methods: Gatys [5], CNN-MRF [8], Deep Image Analogy [9], Pix2Pix [7], CycleGAN [19] and Headshot Portrait [14] (Section S5).

S2. APDrawing Dataset Construction

To train and test the proposed APDrawingGAN, we build a dataset containing 140 pairs of face photos and corresponding portrait drawings. To make the training set distribution more consistent, all portrait drawings were drawn by a single professional artist. All images and drawings are aligned, downsampled and cropped to 512×512 size. Some examples are illustrated in Figure S1. We partition the dataset into two parts: 70 image pairs as the training set and the remaining 70 image pairs as the test set. All the evaluation results are based on the test set to ensure fairness.

S3. More Factors in Ablation Study

In Section 7.1 of the main paper, we study some key factors of APDrawingGAN in an ablation study, including local networks, line-promoting DT loss L_{DT} and initialization using the model pre-trained on the NPR data. Here we present the study on more factors.

*Corresponding author

Figure S1. Some examples of image pairs (each pair contains a face photo and an artist's portrait drawing) in our APDrawing dataset.

(a) Input (b) GT (d) W/O $L_{\mathcal{L}_1}$ loss (f) Ours

Figure S2. Ablation study on the pixel-wise loss term $L_{\mathcal{L}_1}$ in the loss function. From left to right: input face photos, ground truth, results of removing $L_{\mathcal{L}_1}$ from the loss function, and our results.

There are four terms in the loss function of APDrawingGAN (refer to Eq.(1) in the main paper). In addition to L_{DT} (studied in Section 7.1 of the main paper), we further study the other three terms: pixel-wise loss $L_{\mathcal{L}_1}$, local transfer loss L_{local} and adversarial loss L_{adv}.

$L_{\mathcal{L}_1}$ drives the synthesized drawings close to the ground-

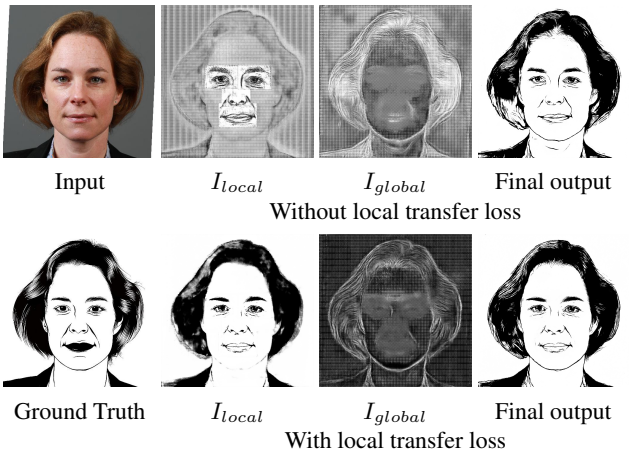

Figure S3. Ablation study on local transfer loss term L_{local} in the loss function. The first column shows the input face photo and ground truth APDrawing. The second, third and last columns show I_{local}, I_{global} and the final output of generator G. Results of removing L_{local} from the loss function are shown in the top row, and results with L_{local} are shown in the bottom row.

Figure S4. Ablation study on GAN loss L_{adv} and using only 2 local nets (face and hair). From left to right: input face photos, ground truth, results of removing L_{adv} from the loss function, results of using only face and hair local nets, and our results.

truth drawings in a pixel-wise manner. As illustrated in Figure S2, without this loss term, excessive white lines appear in the hair region, and meanwhile, regions without lines (such as the necks) become blurry. This is possibly because L_{DT} prefers to promote lines, and without the balance of $L_{\mathcal{L}_1}$, regions containing a few lines (such as hair) exhibit too many lines, while other regions without lines are still not controlled properly, leading to obviously blurry artifacts in these regions (such as necks).

L_{local} puts extra constraints on the intermediate output of six local generators in G_{l*}, and behaves as a regularization term in the loss function. As illustrated in Figure S3, without this loss term, both the intermediate results I_{local} (which is an aggregated drawing blending outputs of all local generators) and I_{global} (which is the output of G_{global}) are underconstrained, leading to unstable and poor generations.

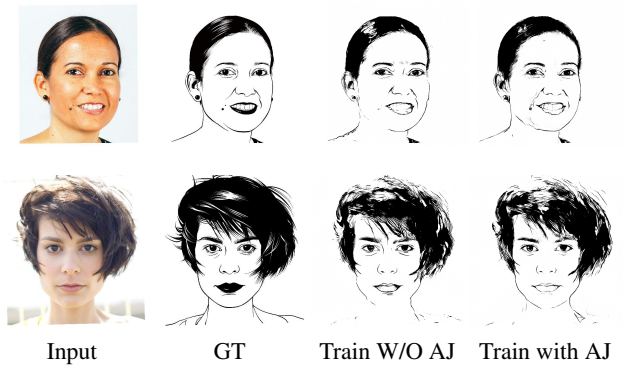

Figure S5. Ablation study on adding jaw contours in coarse training data (AJ). From left to right: input face photos, ground truth, results of the APDrawingGAN model trained without AJ, and results of the APDrawingGAN model trained with AJ.

L_{adv} is fundamental for the GAN architecture and guarantees better results than a CNN. As illustrated in Figure S4c, without GAN loss L_{adv}, the discriminator in our model is removed and the results tend to be blurry, i.e., delicate lines are absent especially in hair regions.

We use six local networks in APDrawingGAN, corresponding to the local facial regions of the left eye, right eye, nose, mouth, hair and the background. To explore the necessity of using six local networks, we conduct an ablation study on using only two local networks for face and hair. As illustrated in Figure S4d, with only two local nets for face and hair, facial features are not well drawn, e.g. noses in both results miss some details, and eyes in the second result are much more messy than our results.

In Section 6 of the main paper, we use a coarse-level pretraining to provide the training of APDrawingGAN with a good initialization. We collect 6,655 frontal face photos taken from ten face datasets [18, 10, 3, 12, 11, 4, 16, 15, 2, 17]. For each photo, we generate a synthetic drawing using the two-tone NPR algorithm in [13]. Since it often generates results without a clear jaw contour (due to low contrast in input images at these locations), we use the face model in OpenFace [1] to detect the landmarks on the jaws and subsequently add the jaw contour to the NPR results. We further study the effect of adding jaw contours in coarse training data. As illustrated in Figure S5, without this important preprocessing step, the trained APDrawingGAN model (after formal training with the APDrawing dataset) cannot generate good jaw features in the synthesized APDrawings. This also shows the benefits of pre-training as improved pretraining data can be efficiently obtained without manual effort.

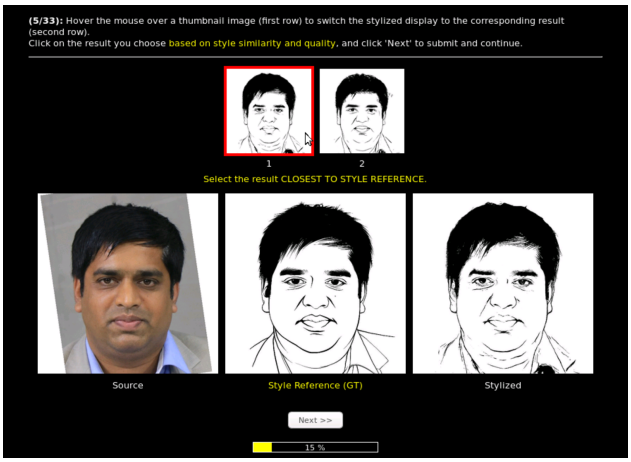

Figure S6. A screenshot of the website for user study.

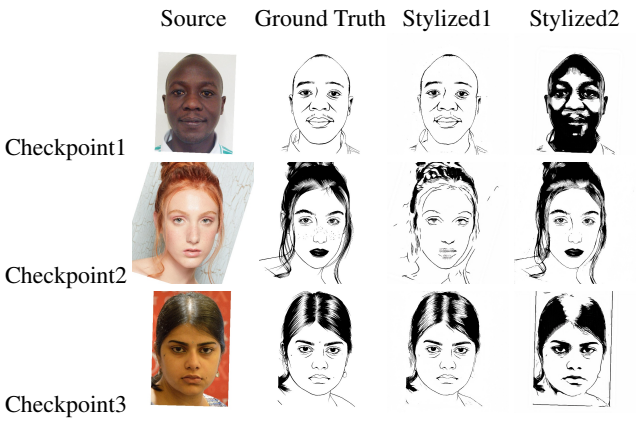

Figure S7. Three checkpoints. In checkpoints 1 and 3, the stylized drawing on the left is obviously better than the right. In checkpoint 2, the stylized drawing on the right is obviously better than the left.

S4. User Study

Due to the subjective nature of image styles, we also conduct a user study to compare our results to CycleGAN [19] and Pix2Pix [7].

Method. All 70 image pairs were used in the user study. For each face photo, three artificial stylized drawings were generated by CycleGAN, Pix2Pix and our APDrawingGAN. Then each image pair was expanded to a group of five images: one original face photo, one ground truth APDrawing, and three artificial stylized drawings. In a total of 70 groups of images, ten groups were randomly assigned to each participant. Each time two artificial stylized drawings were shown on the screen side by side. The participant can hover the mouse over each of them and the enlarged drawing will appear in the bottom for a detailed comparison with original face photo and ground truth side by side. After checking the details of each of two artificial stylized drawings and comparing them with the original face photo and ground truth, the participant chose the one which was better as an APDrawing based on style similarity and quality. See Figure S6 for a screenshot. For three artificial stylized drawings in an image group, any two of them, i.e., (CycleGAN, Pix2Pix), (CycleGAN, APDrawingGAN) and (Pix2Pix, APDrawingGAN), were shown once.

Table S1. Ranking statistics of the user study. For each of the three methods (CycleGAN, Pix2Pix and APDrawingGAN), the percentages of it being ranked best (1), middle (2) and worst (3) are summarized. In 71.39% of all cases, our APDrawingGAN is ranked best.

Methods	Rank 1	Rank 2	Rank 3
CycleGAN [19]	14.45%	30.90%	54.65%
Pix2Pix [7]	14.16%	44.92%	40.92%
APDrawingGAN	71.39%	24.18%	4.43%

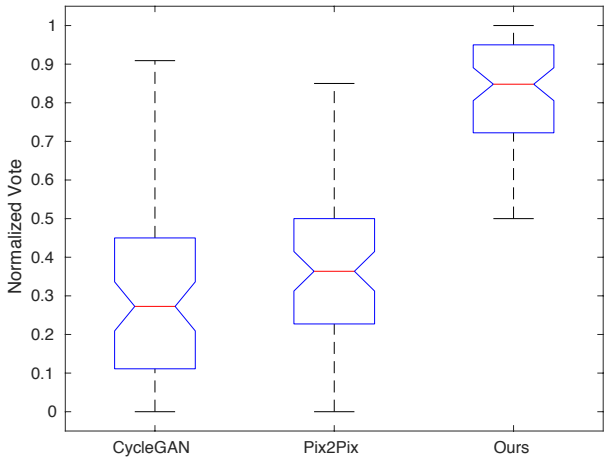

Figure S8. ANOVA test boxplot [6] of three methods.

Quality Control. To avoid unreliable input such as random selection, we add checkpoints in the process of the user study to control the quality of user input. We use three special pairs of stylized drawings with obvious preference as checkpoints (Figure S7). These three pairs randomly appeared in the process of user study. According to our preparatory experiments, participants with high concentration can easily choose the obviously better drawing, while those who just randomly select drawings are likely to fail in at least one checkpoint input. We discard the user input if one or more checkpoints failed.

Results. 81 paid participants attended this user study and 73 of them passed all the checkpoints. We performed statistical analysis on the valid inputs of these participants in two ways. First, we compute a global ranking for the three artificial stylized drawings in each image group. For example, if A is better than B, B is better than C and A is better than C, then the global ranking is A, B and C. If the local ranking is conflicted, e.g., A is better than B, B is better than C and C is better than A, the votes for this image group are discarded. From all votes in 73 valid inputs, we compute the percentages that the three methods (CycleGAN,

Pix2Pix and APDrawingGAN) are ranked best, middle or worst, respectively. The ranking results are summarized in Table S1, in which APDrawingGAN is ranked the best in 71.39% of all cases, significantly higher than CycleGAN and Pix2Pix where each of them is only ranked the best for about 14% of cases. Second, we conduct analysis of variance (ANOVA) on the normalized data, i.e., the votes normalized by maximum number of votes. The p-value is $2.04 \times 10^{-39} << 0.01$, justifying that the rejection of the null hypothesis and the differences between the means of the three methods are statistically significant. The boxplot is shown in Figure S8.

S5. More Qualitative Results of Comparison

In Section 7.2 of the main paper, we compare APDrawingGAN with six state-of-the-art style transfer methods: Gatys [5], CNNMRF [8], Deep Image Analogy [9], Pix2Pix [7], CycleGAN [19] and Headshot Portrait [14].

For methods that take one content image and one style image as input, i.e., CNNMRF, Deep Analogy and Headshot Portrait, we randomly select a style image in the training set. Gatys' method [5] by default takes one content image and one style image as input. But for fair comparison, we use all the style images in the training set and compute the average Gram matrix to model the target style as in [19]. For CycleGAN and Pix2Pix, we use the same training data as APDrawingGAN and default parameters to train the model. The qualitative results of randomly selected test data are illustrated in Figure S9. We also test our trained APDrawingGAN on arbitrary collected face photos which do not have ground truth artist's drawings, and the qualitative results are illustrated in Figure S10. These results show that our APDrawingGAN consistently generates high-quality and better APDrawings than existing methods.

References

[1] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. OpenFace: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016. 2

[2] Olga Chelnokova, Bruno Laeng, Marie Eikemo, Jeppe Riegels, Guro Løseth, Hedda Maurud, Frode Willoch, and Siri Leknes. Rewards of beauty: the opioid system mediates social motivation in humans. *Molecular Psychiatry*, 19:746–751, 2014. 2

[3] Rémi Courset, Marine Rougier, Richard Palluel-Germain, Annique Smeding, Juliette Manto Jonte, Alan Chauvin, and Dominique Muller. The caucasian and north african french faces (CaNAFF): A face database. *International Review of Social Psychology*, 31(1):22:1–22:10, 2018. 2

[4] Natalie C. Ebner, Michaela Riediger, and Ulman Lindenberger. FACES-a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42(1):351–362, 2010. 2

[5] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '16, pages 2414–2423, 2016. 1, 4, 5, 6

[6] Robert V. Hogg and Johannes Ledolter. *Engineering Statistics*. New York: MacMillan, 1987. 3

[7] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '17, pages 1125–1134, 2017. 1, 3, 4, 5, 6

[8] Chuan Li and Michael Wand. Combining markov random fields and convolutional neural networks for image synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '16, pages 2479–2486, 2016. 1, 4, 5, 6

[9] Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang. Visual attribute transfer through deep image analogy. *ACM Transactions on Graphics (TOG)*, 36(4):120:1–120:15, 2017. 1, 4, 5, 6

[10] Debbie S. Ma, Joshua Correll, and Bernd Wittenbrink. The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(4):1122–1135, 2015. 2

[11] Peter Peer, Žiga Emeršič, Jernej Bule, Jerneja Žganec-Gros, and Vitomir Štruc. Strategies for exploiting independent cloud implementations of biometric experts in multibiometric scenarios. *Mathematical Problems in Engineering*, 2014:1–15, 2014. 2

[12] P. Jonathon Phillips, Harry Wechsler, Jeffrey Huang, and Patrick J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998. 2

[13] Paul L. Rosin and Yu-Kun Lai. Towards artistic minimal rendering. In *International Symposium on Non-Photorealistic Animation and Rendering*, NPAR '10, pages 119–127, 2010. 2

[14] Yi-Chang Shih, Sylvain Paris, Connelly Barnes, William T. Freeman, and Frédo Durand. Style transfer for headshot portraits. *ACM Transactions on Graphics (TOG)*, 33(4):148:1–148:14, 2014. 1, 4, 5, 6

[15] Nina Strohminger, Kurt Gray, Vladimir Chituc, Joseph Heffner, Chelsea Schein, and Titus Brooks Heagins. The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods*, 48(3):1197–1204, 2016. 2

[16] Carlos Eduardo Thomaz and Gilson Antonio Giraldi. A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, 28(6):902–913, 2010. 2

[17] Tiago F. Vieira, Andrea Bottino, Aldo Laurentini, and Matteo De Simone. Detecting siblings in image pairs. *The Visual Computer*, 30(12):1333–1345, 2014. 2

[18] Mirella Walker, Sandro Schönborn, Rainer Greifeneder, and Thomas Vetter. The Basel face database: A validated set of photographs reflecting systematic differences in big

Figure S9. Qualitative results of our method and comparison with six state-of-the-art methods. From left to right: input face photos, ground truth APDrawings, the randomly-chosen style images for methods which take one content and one style image as input, CNNMRF [8] results, Deep Image Analogy [9] results, Headshot Portrait [14] results, Gatys [5] results, CycleGAN [19] results, Pix2Pix [7] results, our APDrawingGAN results.

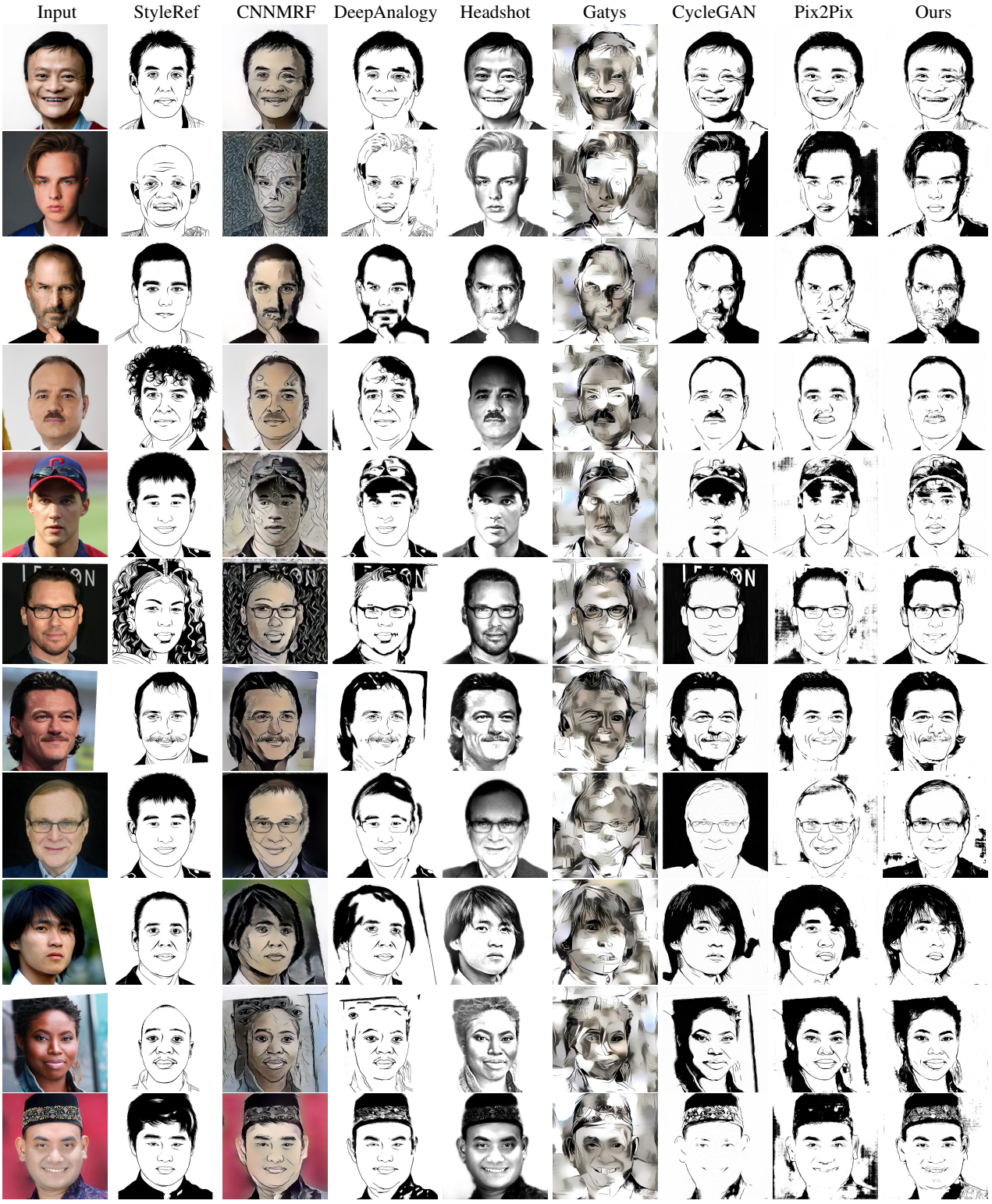

Figure S10. Qualitative results of our method and comparison with six state-of-the-art methods. From left to right: input face photos (collected from internet which do not have ground truth artist's drawings), the randomly-chosen style images for methods which take one content and one style image as input, CNNMRF [8] results, Deep Image Analogy [9] results, Headshot Portrait [14] results, Gatys [5] results, CycleGAN [19] results, Pix2Pix [7] results, our APDrawingGAN results.

two and big five personality dimensions. *PLoS ONE*, 13(3):e0193190, 2018. 2

[19] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *IEEE International Conference on Computer Vision*, ICCV '17, pages 2223–2232, 2017. 1, 3, 4, 5, 6