

# Line Drawings for Face Portraits from Photos using Global and Local Structure based GANs (Appendix)

Ran Yi, Mengfei Xia, Yong-Jin Liu, *Senior Member, IEEE*, Yu-Kun Lai, *Member, IEEE*,  
Paul L. Rosin, *Member, IEEE*

## A1 OVERVIEW

In this appendix, a theoretical explanation and more experimental results are provided, including:

- a theoretical explanation of the composite generator (Section A2);
- some illustrative examples from the APDrawing dataset (Section A3);
- more qualitative results that compare APDrawing-GAN++ with seven state-of-the-art style transfer methods: Gatys [1], CNNMRF [2], Deep Image Analogy [3], Pix2Pix [4], CycleGAN [5], Headshot Portrait [6] and APDrawingGAN [7] (Section A4);
- more details of the user study (Section A5);
- more details of the ablation study (Section A6);
- comparison of our training strategy with an alternative mixed training strategy (Section A7).

## A2 THEORETICAL EXPLANATION OF COMPOSITE GENERATOR

Our composite generator  $G$  defines a mapping  $g_\theta : \mathcal{Z} \rightarrow \mathcal{X}$ , where  $\mathcal{Z}$  is the latent space in which the input face photos are embedded with random perturbations<sup>1</sup>,  $\mathcal{X}$  is the data space containing all portrait drawings, and  $\theta$  denotes the set of parameters in the generator.

Our key observation is that artists usually draw different facial parts using different drawing techniques, e.g., eyes are drawn with fine details and hair is drawn with curves and flows. Accordingly, the APDrawing data of the same person lies in multiple disjoint manifolds of possibly different dimensions and we denote this manifold clustering as

- 
- R. Yi, M. Xia, Y.-J. Liu are with BNRIst, MOE-Key Laboratory of Pervasive Computing, the Department of Computer Science and Technology, Tsinghua University, Beijing, China. Y.-J. Liu is the corresponding author. E-mail: liuyongjin@tsinghua.edu.cn.
  - Y.-K. Lai and P.L. Rosin are with School of Computer Science and Informatics, Cardiff University, UK.

1. This is due to the randomness introduced in the dropout layer at test time.

$M = \{M_1, M_2, \dots\} \subset \mathcal{X}$ , in which any two data manifolds  $M_i$  and  $M_j$  are disconnected<sup>2</sup>.

Define the probability density function  $p_z : \mathcal{Z} \rightarrow \mathbb{R}$  for the distribution of face photos embedded in the latent space, and the probability density function  $p_x : \mathcal{X} \rightarrow \mathbb{R}$  for the distribution of APDrawing data that we want to learn. Note that  $p_x$  satisfies that the value outside the target manifold clusters  $M$  is 0, i.e. the support of  $p_x$  satisfies

$$\text{supp}(p_x) \triangleq \{x \in \mathcal{X} \mid p_x(x) \neq 0\} \subseteq M \quad (1)$$

The distribution function  $\mu$  on  $\mathcal{Z}$  can be derived from  $p_z$ , i.e., for an arbitrary subset  $U$  in  $\mathcal{Z}$ ,

$$\mu(U) = \int_U p_z(z) dz \quad (2)$$

Similarly, the distribution function  $\nu$  on  $\mathcal{X}$  can be derived from  $p_x$ .

The generator  $G$  is optimized via a minimax process and the goal is to find a proper  $\theta$  so that  $g_\theta \# \mu = \nu$ , where  $g_\theta \# \mu$  stands for the push forward of  $\mu$  by  $g_\theta$  [8]. I.e., for an arbitrary subset  $V$  in  $\mathcal{X}$ ,

$$g_\theta \# \mu(V) \triangleq \mu(g_\theta^{-1}(V)) = \nu(V) \quad (3)$$

where  $g_\theta^{-1}$  is the inverse mapping of  $g_\theta$ , and  $g_\theta^{-1}(V)$  is a subset in  $\mathcal{Z}$  (refer to Fig. A1).

Here we show that a single generator cannot learn such a proper  $\theta$  that satisfies  $g_\theta \# \mu = \nu$ . Therefore, designing a set of multiple generators like our  $G$  is necessary. In this study, we follow [9] to assume that all face photos of the same person lie in a low-dimensional manifold. Since these face photos lie in a single manifold, the probability density function  $p_z$  is non-zero on the whole latent space  $\mathcal{Z}$ ; in other words, the support of  $p_z$  is the whole  $\mathcal{Z}$ . Now we consider a special case, in which the manifold cluster  $M$  contains only two disconnected manifolds  $M_1$  and  $M_2$ . We have

**Theorem 1.** Let  $p_z$ ,  $p_x$  be the probability density functions on  $\mathcal{Z}$ ,  $\mathcal{X}$  respectively, and satisfy  $\text{supp}(p_z) = \mathcal{Z}$ ,  $\text{supp}(p_x) \subseteq$

2. Note that the condition  $M_i \cap M_j = \emptyset$  alone cannot guarantee that  $M_i$  and  $M_j$  are disconnected, e.g., in  $\mathbb{R}^2$ , the manifold  $\sin \frac{1}{x}$  and  $x = 0$  are connected, but their intersection is empty. On the other hand, if two manifolds are disconnected, their intersection is also empty.

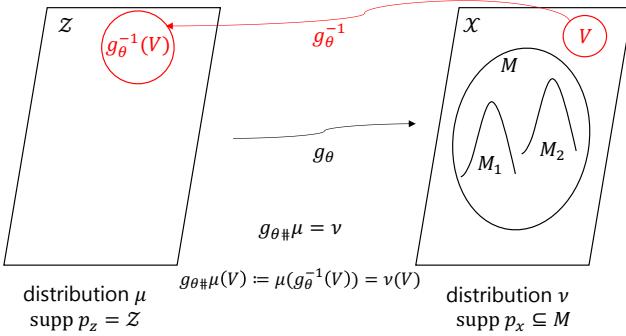


Fig. A1. The proof of Theorem 1.

$M = M_1 \cup M_2$ ,  $\text{supp}(p_x) \cap M_1 \neq \emptyset$ ,  $\text{supp}(p_x) \cap M_2 \neq \emptyset$ . Let  $\mu, \nu$  be the distribution functions corresponding to  $p_z, p_x$ , respectively. Then there does not exist a Lipschitz continuous function<sup>3</sup>  $g_\theta : \mathcal{Z} \rightarrow \mathcal{X}$  that makes  $g_\theta \# \mu = \nu$ .

*Proof.* We prove this theorem by contradiction. Assume that there exists such a Lipschitz continuous function  $g_\theta$  with which  $g_\theta \# \mu = \nu$ . For a subset  $V$  in  $\mathcal{X}$  that satisfies  $V \cap \text{supp}(p_x) = \emptyset$ , from the definition of  $\text{supp}(p_x)$ , we have  $\forall x \in V, p_x(x) = 0$ , thus  $g_\theta \# \mu(V) = \nu(V) = 0$ , i.e.,

$$\mu(g_\theta^{-1}(V)) = 0 \Rightarrow \int_{g_\theta^{-1}(V)} p_z(z) dz = 0 \quad (4)$$

Since  $\text{supp}(p_z) = \mathcal{Z}$ ,  $g_\theta^{-1}(V)$  is a zero set<sup>4</sup> in  $\mathcal{Z}$ . For the complement set of  $\text{supp}(p_x)$  in  $\mathcal{X}$ ,  $(\mathcal{X} \setminus \text{supp}(p_x)) \cap \text{supp}(p_x) = \emptyset$ , thus  $g_\theta^{-1}(\mathcal{X} \setminus \text{supp}(p_x))$  is a zero set in  $\mathcal{Z}$ .

Since  $g_\theta$  is Lipschitz continuous, we have  $g_\theta^{-1}(\mathcal{X} - \text{supp}(p_x)) = \emptyset$ . Then  $g_\theta$  is a function from  $\mathcal{Z}$  to  $M = M_1 \cup M_2$ , i.e.  $g_\theta : \mathcal{Z} \rightarrow M_1 \cup M_2$ . Since a continuous function maps a connected set to another connected set, but  $M_1 \cup M_2$  is not connected, this leads to a contradiction.  $\square$

### A3 ILLUSTRATIVE EXAMPLES IN APDRAWING DATASET

In Section 6 of the main paper, we present the construction of the APDrawing Dataset, which contains 140 pairs of face photos and corresponding portrait drawings, with all portrait drawings were drawn by a single professional artist. Figure A2 shows examples of four pairs. More examples are illustrated in Figures A5-A17, i.e., the pair is in the format of (input photo, ground truth).

### A4 MORE QUALITATIVE RESULTS

In Section 7.1.1 of the main paper, we compare APDrawingGAN++ with seven state-of-the-art style transfer methods: Gatys [1], CNNMRF [2], Deep Image Analogy [3], Pix2Pix [4], CycleGAN [5], Headshot Portrait [6] and APDrawingGAN [7].

To make this appendix self-explainable, we repeat here the comparison method in the main paper. For methods

3. We consider the Lipschitz continuous functions here due to the existence of ReLU, leaky ReLU and tanh functions in our network.

4. A zero set is a set with zero Lebesgue measure.



Fig. A2. Four examples of image pairs (each pair contains a face photo and an artist's portrait drawing) in our APDrawing dataset.

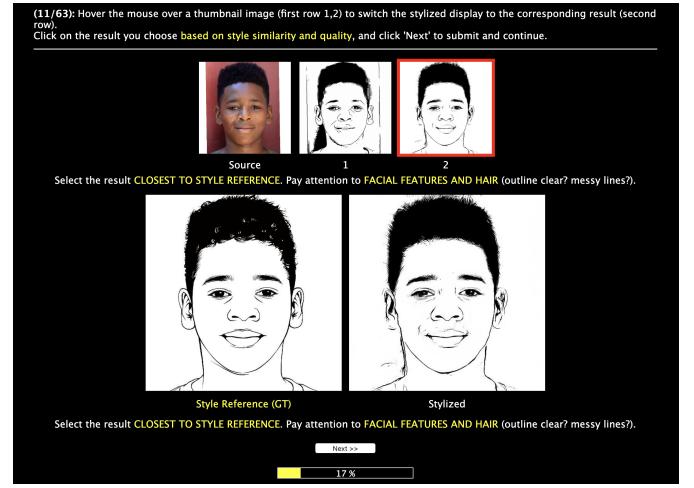


Fig. A3. A screenshot of the website for user study.

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 [1] 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 [5]. For CycleGAN, Pix2Pix and APDrawingGAN, we use the same training data as APDrawingGAN++ and default parameters to train the models.

The qualitative results on all test data are illustrated in Figures A13-A17. 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 A18. These results show that APDrawingGAN++ consistently generates high-quality and better APDrawings than other methods.

### A5 MORE DETAILS IN USER STUDY

In Section 7.1.2, we present a user study to compare CycleGAN [5], Pix2Pix [4], APDrawingGAN [7] and APDrawingGAN++. Here we present the detail of setup and experimental procedure in this user study.

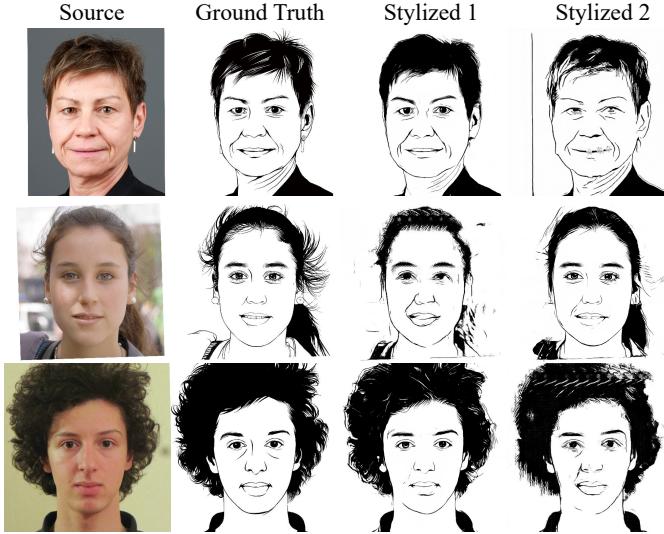


Fig. A4. Three checkpoints. (zoom in to see more details.) In checkpoints in the 1st and 3rd row, the stylized drawing on the left is obviously better than the right. In the checkpoint in the 2nd row, the stylized drawing on the right is obviously better than the left.

### A5.1 Experimental Procedure

All 70 test pairs in the APDrawing dataset were used in the user study. For each face photo, four stylized drawings were automatically generated by CycleGAN, Pix2Pix, AP-DrawingGAN and APDrawingGAN++. Then each image pair was expanded to a group of six images: one original face photo, one ground truth APDrawing, and four artificial stylized drawings. In a total of 70 groups of images, 10 groups were randomly assigned to each participant. Each time the original face photo and two stylized drawings were shown on the screen side by side. The participant can hover the mouse over each stylized drawing and the enlarged drawing will appear in the bottom for a detailed comparison with ground truth side by side. After checking the detail of each of two stylized drawings and comparing them with the original face photo and ground truth, the participant chose the one that was better as a masterful APDrawing based on style similarity and quality (clearer outline of facial features, less messy lines etc). See Figure A3 for a screenshot. For four stylized drawings in an image group, any two of them, i.e., (CycleGAN, Pix2Pix), (CycleGAN, APDrawingGAN), (CycleGAN, APDrawingGAN++), (Pix2Pix, APDrawingGAN), (Pix2Pix, APDrawingGAN++) and (APDrawingGAN, APDrawingGAN++), were shown once.

### A5.2 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 A4). 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.

### A5.3 Result Analysis

102 participants were recruited in this user study and 87 of them passed all the checkpoints. We performed statistical analysis on the valid inputs of these participants in three aspects.

First, we figure out the best ranked drawing from the four stylized drawings in each image group. For example, if A is ranked better than B, A is better than C and A is better than D by a participant, then the best result in (A, B, C, D) is A. If no single result is better than all three others, the votes of this participant for this image group are discarded. From all votes in 87 valid inputs, we compute the percentages that the four methods (CycleGAN, Pix2Pix, APDrawingGAN and APDrawingGAN++) are ranked best respectively. The ranking results are summarized in Table 1 of the main paper. We also compute the percentage of each method being preferred in pairwise comparison and summarize the results in the same table. The results in the main paper show that APDrawingGAN++ is much better than the other three methods.

Second, we conduct analysis of variance (ANOVA) for pairwise comparisons on ranked best percentage and preferred percentage. Pairwise ANOVA results are shown in Table 2 of the main paper. All of the  $p$ -values are  $\ll 0.01$ , justifying that the rejection of the null hypothesis and the differences between the means of our method and either method (Pix2Pix, CycleGAN or APDrawingGAN) are statistically significant.

Third, we investigate the improvement of APDrawingGAN++ over APDrawingGAN on dark faces. We concentrate on the pairwise comparison of (APDrawingGAN, APDrawingGAN++) and find that APDrawingGAN++ wins in this comparison in 77.67% of cases for dark faces and 74.69% of cases for light faces. This justifies that histogram matching augmentation in APDrawingGAN++ is useful for correcting dark faces that cannot be dealt with well by APDrawingGAN.

## A6 MORE INGREDIENTS IN THE ABLATION STUDY

In Section 7.2 of the main paper, we present an ablation study on:

- some key ingredients of APDrawingGAN++, including local networks, line-promoting DT loss  $L_{DT}$  and initialization using the model pre-trained on the NPR data; and
- major differences between APDrawingGAN++ and APDrawingGAN, including auto-encoders, lip and hair classifiers, DT nonlinear mapping, and histogram matching augmentation.

Here we present the study on more ingredients.

### A6.1 Loss Function

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 7.2 of the main paper), we

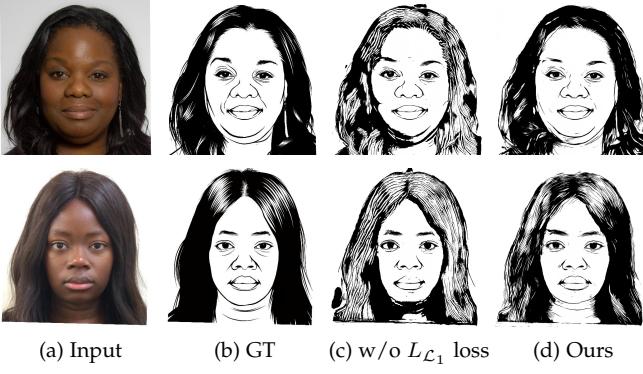


Fig. A5. Ablation study on the pixel-wise loss term  $L_{\mathcal{L}_1}$  in the loss function. From left to right: input face photos, ground truth, results of removing  $L_{\mathcal{L}_1}$  from the loss function, and our results.

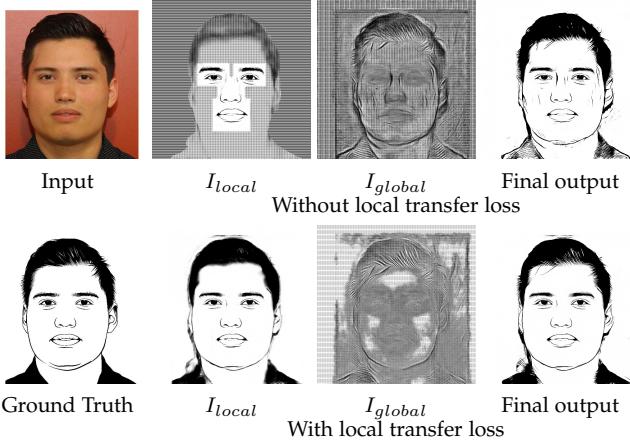


Fig. A6. Ablation study on the local transfer loss term  $L_{local}$  in the loss function. The first column shows the input face photo and ground truth APDrawing. The second, third and last columns show  $I_{local}$ ,  $I_{global}$  and the final output of generator  $G$ . Results of removing  $L_{local}$  from the loss function are shown in the top row, and results with  $L_{local}$  are shown in the bottom row.

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-truth drawings in a pixel-wise manner. As illustrated in Figure A5, 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 obvious artifacts in these regions (such as necks).

$L_{local}$  puts extra constraints on the intermediate output of six local generators in  $G_{ls}$ , and behaves as a regularization term in the loss function. As illustrated in Figure A6, 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. Overall, using the local transfer loss  $L_{local}$  decreases the average LPIPS distance value on the test set from 0.286 to

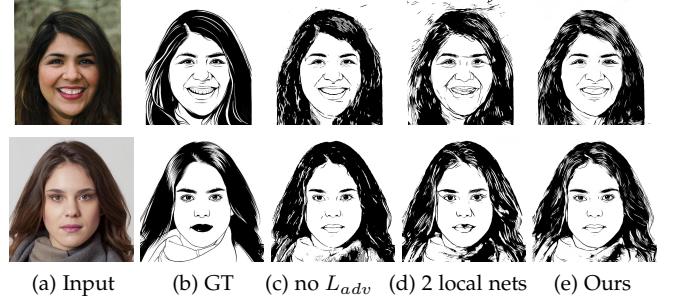


Fig. A7. Ablation study on GAN loss  $L_{adv}$  and using only 2 local nets (face and hair). From left to right: input face photos, ground truth, results of removing  $L_{adv}$  from the loss function, results of using only face and hair local nets, and our results.

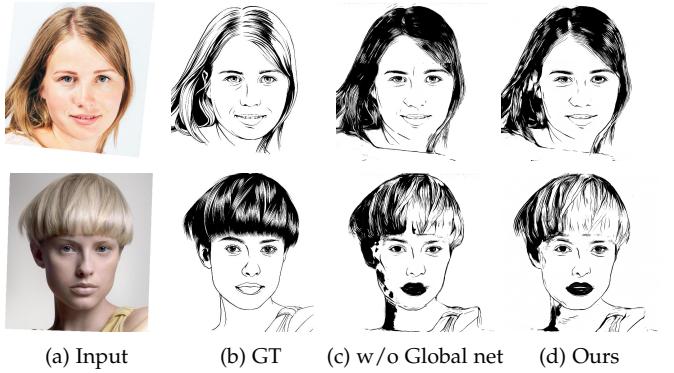


Fig. A8. Ablation study on the global net. From left to right: input face photos, ground truth, results of removing the global net, and our results.

0.258, and a user study<sup>5</sup> shows that our results are selected as better than the results of no  $L_{local}$  in 81.1% cases.

$L_{adv}$  is fundamental for the GAN architecture and guarantees better results than a CNN. As illustrated in Figure A7c, without GAN loss  $L_{adv}$ , the discriminator in APDrawingGAN++ is removed and the results tend to be blurry, i.e., delicate lines are absent especially in hair regions.

## A6.2 Local Network Structure

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 A7d, with only two local nets for face and hair, facial features are not well drawn, e.g. eyes in both results are much more messy than our results, and lips in the first result are drawn in a strange shape.

## A6.3 Global Network Structure

To explore the necessity of using the global network in generator, we conduct an ablation study on the global network. We remove the global network and the fusion network, and use the combination of local network results  $I_{local}$  as the output of the generator. As illustrated in Figure A8, without

<sup>5</sup> 48 participants were recruited in this user study and 40 of them passed checkpoints.

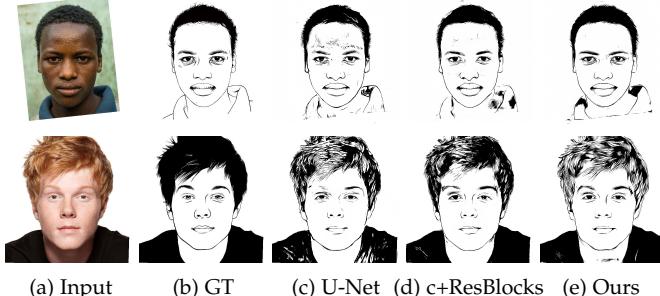


Fig. A9. Ablation study on residual blocks in the generator. From left to right: input face photos, ground truth, results of using U-Net in the generator, results of using U-Net and residual blocks in the generator, and our results (using CNN with residual blocks in the generator).

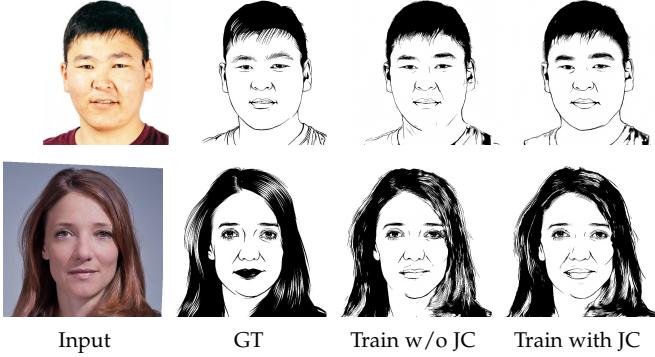


Fig. A10. Ablation study on adding jaw contours (JC) in coarse training data. From left to right: input face photos, ground truth, results of the APDrawingGAN++ model trained without adding JC, and results of the APDrawingGAN++ model trained with adding JC.

the global network, the hair is drawn less delicately and the boundaries of local regions (e.g. eyes) are more visible, degrading the quality of synthesized drawings. Overall, using the global network decreases the average LPIPS distance value on the test set from 0.274 to 0.258. A user study<sup>6</sup> shows that our results are selected as better than the results of no global net in 58.3% cases.

#### A6.4 Residual Blocks in the Generator

In Section 4 of the main paper, we introduce the architecture of our network. Compared with APDrawingGAN, we design the generator  $G$  using a CNN with residual blocks [10]. Residual blocks were designed to reduce the impact of vanishing gradients and speed up the training procedure. We compare the current generator structure with the U-Net structure used in APDrawingGAN and U-Net with residual blocks in Figure A9. It is shown that the results of using residual blocks (Figure A9(d)(e)) have less messy lines on the face and the results generated by our structure have finer details.

#### A6.5 Pre-training Strategy

In Section 6 of the main paper, we use a coarse-level pre-training to provide the training of APDrawingGAN++ with a good initialization. We collect 6,655 frontal face photos taken from ten face datasets [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]. For each photo, we generate a synthetic

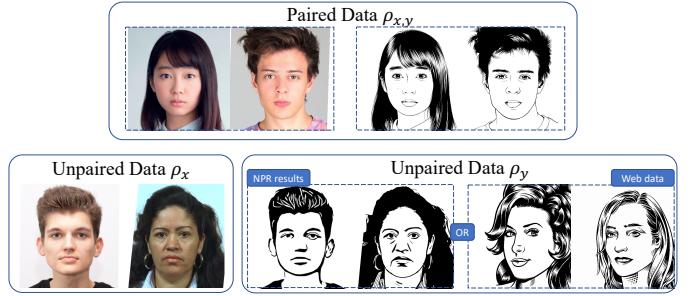


Fig. A11. Paired and Unpaired data used in the mixed training strategy.

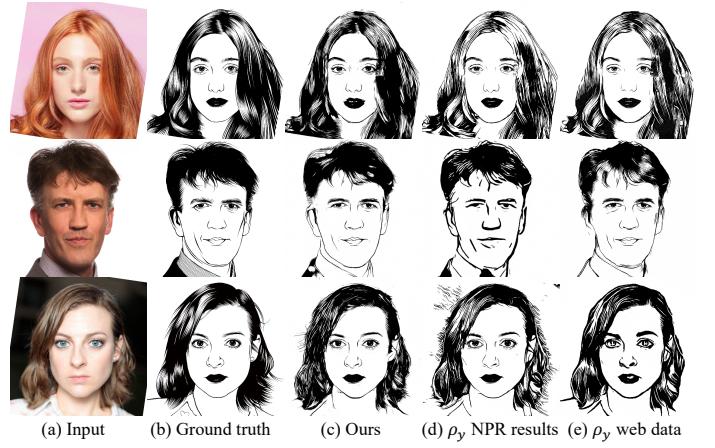


Fig. A12. Comparison results with mixed training strategy. From left to right: input photos, ground truth, our results, results of mixed training with NPR data, and results of mixed training with web data

drawing using the two-tone NPR algorithm in [21]. 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 [22] 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 A10, without this important preprocessing step, the trained APDrawingGAN++ (after formal training with the APDrawing dataset) cannot generate good jaw features in the synthesized APDrawings. This also demonstrates the benefits of pre-training, i.e., improved pre-training data can be efficiently obtained without manual effort.

#### A7 COMPARISON WITH MIXED TRAINING

We further compare our pre-training strategy with the mixed training strategy in [23]. The mixed training strategy jointly trains a generator  $G : x \rightarrow y$  and a discriminator  $D : y \rightarrow D(y) \in \mathbb{R}$  using both paired data  $\rho_{x,y}$  and unpaired data  $\rho_x$  and  $\rho_y$ . Since  $G$  only takes  $x$  as input and  $D$  only takes  $y$ , unrelated unpaired data can be used. Both GAN loss and supervised training loss are applied to paired data  $\rho_{x,y}$ , while only GAN loss is applied to unpaired data  $\rho_x$  and  $\rho_y$ . We compare with mixed training with two kinds of  $\rho_y$ : NPR results or web collected data, refer to Figure A11<sup>7</sup>.

As illustrated in Figure A12d, since NPR generated data is only a coarse approximation and is far from ideal

6. The same user study as the one in footnote 5.

7. The NPR results and photos are actually paired, but are used in an unpaired way here.

APDrawings, the difference in style leads to inconsistent results, i.e. sometimes results show NPR style (the second row), and sometimes results show APDrawing style (the first and third rows). As illustrated in Figure A12e, although the web collected data is more delicate and more similar to APDrawings than NPR generated data, the results still sometimes show inconsistent styles and contain thicker lines (the third row), while other times the results are better (the first row).

Overall, the LPIPS distance on the test set is 0.284 for mixed training with NPR generated data, 0.280 for mixed training with web data, 0.277 for no pre-training (i.e. using paired APDrawing data only) and 0.258 for our method. A user study<sup>8</sup> shows that our results are selected as better than the results of mixed training with NPR data in 77.5% cases, and better than the results of mixed training with web data in 77.4% cases; the results without pre-training are selected as better than the results of mixed training with web data in 60.4% cases. The results of mixed training are worse than ours, probably because the mixed datasets containing APDrawings and NPR/web data are of different characteristics (line width, abstraction level, etc), and this leads to inconsistent styles in results.

These results show that the mixed training strategy is sensitive to differences between training images, and the benefits of using unpaired data are limited and can produce worse results. In contrast, our pre-training benefits from easily obtained NPR data while avoiding the final results being affected by differences in styles.

It is still worth exploiting unpaired web data when paired data is limited, but further effort may be needed to deal with different styles.

## REFERENCES

- [1] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, 2016, pp. 2414–2423.
- [2] C. Li and M. Wand, "Combining markov random fields and convolutional neural networks for image synthesis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, 2016, pp. 2479–2486.
- [3] J. Liao, Y. Yao, L. Yuan, G. Hua, and S. B. Kang, "Visual attribute transfer through deep image analogy," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 120:1–120:15, 2017.
- [4] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '17)*, 2017, pp. 1125–1134.
- [5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networkss," in *IEEE International Conference on Computer Vision (ICCV '17)*, 2017, pp. 2223–2232.
- [6] Y. Shih, S. Paris, C. Barnes, W. T. Freeman, and F. Durand, "Style transfer for headshot portraits," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, pp. 148:1–148:14, 2014.
- [7] R. Yi, Y.-J. Liu, Y.-K. Lai, and P. L. Rosin, "ApdrawingGAN: Generating artistic portrait drawings from face photos with hierarchical gans," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '19)*, 2019, pp. 10743–10752.
- [8] C. Villani, *Optimal transport: old and new.* Springer Science & Business Media, 2008.
- [9] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268–2269, 2000.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, 2016, pp. 770–778.
- [11] M. Walker, S. Schönborn, R. Greifeneder, and T. Vetter, "The Basel face database: A validated set of photographs reflecting systematic differences in big two and big five personality dimensions," *PLoS ONE*, vol. 13, no. 3, p. e0193190, 2018.
- [12] D. S. Ma, J. Correll, and B. Wittenbrink, "The Chicago face database: A free stimulus set of faces and norming data," *Behavior Research Methods*, vol. 47, no. 4, pp. 1122–1135, 2015.
- [13] R. Courset, M. Rougier, R. Palluel-Germain, A. Smeling, J. M. Jonte, A. Chauvin, and D. Muller, "The caucasian and north african french faces (CaNAFF): A face database," *International Review of Social Psychology*, vol. 31, no. 1, pp. 22:1–22:10, 2018.
- [14] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.
- [15] P. Peer, Ž. Emeršič, J. Bule, J. Žganec-Gros, and V. Štruc, "Strategies for exploiting independent cloud implementations of biometric experts in multibiometric scenarios," *Mathematical Problems in Engineering*, vol. 2014, pp. 1–15, 2014.
- [16] N. C. Ebner, M. Riediger, and U. Lindenberger, "FACES-a database of facial expressions in young, middle-aged, and older women and men: Development and validation," *Behavior Research Methods*, vol. 42, no. 1, pp. 351–362, 2010.
- [17] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 6, pp. 902–913, 2010.
- [18] N. Strohminger, K. Gray, V. Chituc, J. Heffner, C. Schein, and T. B. Heagins, "The MR2: A multi-racial, mega-resolution database of facial stimuli," *Behavior Research Methods*, vol. 48, no. 3, pp. 1197–1204, 2016.
- [19] O. Chelnokova, B. Laeng, M. Eikemo, J. Riegels, G. Løseth, H. Maurud, F. Willoch, and S. Leknes, "Rewards of beauty: the opioid system mediates social motivation in humans," *Molecular Psychiatry*, vol. 19, pp. 746–751, 2014.
- [20] T. F. Vieira, A. Bottino, A. Laurentini, and M. De Simone, "Detecting siblings in image pairs," *The Visual Computer*, vol. 30, no. 12, pp. 1333–1345, 2014.
- [21] P. L. Rosin and Y. Lai, "Towards artistic minimal rendering," in *International Symposium on Non-Photorealistic Animation and Rendering (NPAR '10)*, 2010, pp. 119–127.
- [22] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.
- [23] E. Simo-Serra, S. Iizuka, and H. Ishikawa, "Mastering sketching: Adversarial augmentation for structured prediction," *ACM Trans. Graph.*, vol. 37, no. 1, pp. 11:1–11:13, 2018.

8. The same user study as the one in footnote 5.





Fig. A14. Qualitative results of our method and comparison with seven 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 [2] results, Deep Image Analogy [3] results, Headshot Portrait [6] results, Gatys [1] results, CycleGAN [5] results, Pix2Pix [4] results, APDrawingGAN [7] results, our APDrawingGAN++ results.



Fig. A15. Qualitative results of our method and comparison with seven 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 [2] results, Deep Image Analogy [3] results, Headshot Portrait [6] results, Gatys [1] results, CycleGAN [5] results, Pix2Pix [4] results, APDrawingGAN [7] results, our APDrawingGAN++ results.



Fig. A16. Qualitative results of our method and comparison with seven 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 [2] results, Deep Image Analogy [3] results, Headshot Portrait [6] results, Gatys [1] results, CycleGAN [5] results, Pix2Pix [4] results, APDrawingGAN [7] results, our APDrawingGAN++ results.



Fig. A17. Qualitative results of our method and comparison with seven 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 [2] results, Deep Image Analogy [3] results, Headshot Portrait [6] results, Gatys [1] results, CycleGAN [5] results, Pix2Pix [4] results, APDrawingGAN [7] results, our APDrawingGAN++ results.



Fig. A18. Qualitative results of our method and comparison with seven 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 [2] results, Deep Image Analogy [3] results, Headshot Portrait [6] results, Gatys [1] results, CycleGAN [5] results, Pix2Pix [4] results, APDrawingGAN [7] results, our APDrawingGAN++ results.