SCIENCE CHINA
Information Sciences

• LETTER •

November 2020, Vol. 63 219103:1–219103:3
https://doi.org/10.1007/s11432-019-2890-8

Face-sketch learning with human sketch-drawing order enforcement

Liang CHANG[1], Lihua JIN[1], Lifen WENG[2], Wentao CHAO[3], Xuguang WANG[3], Xiaoming DENG[4*] & Qiulei DONG[5,6,7*]

[1]*School of Artificial Intelligence, Beijing Normal University, Beijing* 100875*, China;*
[2]*Department of Design Art, Xiamen University of Technology, Xiamen* 361024*, China;*
[3]*Department of Automation, North China Electric Power University, Baoding* 071003*, China;*
[4]*Beijing Key Laboratory of Human Computer Interactions, Institute of Software, Chinese Academy of Sciences, Beijing* 100190*, China;*
[5]*National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing* 100190*, China;*
[6]*School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing* 100049*, China;*
[7]*Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing* 100190*, China*

Received 27 November 2019/Revised 19 January 2020/Accepted 26 February 2020/Published online 12 October 2020

Citation Chang L, Jin L H, Weng L F, et al. Face-sketch learning with human sketch-drawing order enforcement. Sci China Inf Sci, 2020, 63(11): 219103, https://doi.org/10.1007/s11432-019-2890-8

Dear editor,

Although face-sketch synthesis generates a sketch from a given face photo automatically [1], it is an open research problem in computer vision [2–4]. Recently, several deep neural network (DNN) methods for face-sketch synthesis have been proposed with considerable results. However, the knowledge of the human sketch-drawing order is not yet exploited much in the existing DNN methods. Generally, in deep learning, if some intermediate knowledge is explicitly embedded in the DNN layers during learning, the problem of overfitting will be reduced significantly to achieve better performance. In neuroscience, the principle is commonly used in DNN-based sensory cortex modeling [5]. Moreover, by enforcing the DNN training with functional magnetic resonance imaging data, as well as electrophysiological data at different layers, the prediction performance of the trained model is increased significantly. Although our study is in line with these studies, we use the sequential sketching data to enforce the sketch predictions of different DNN layers in the training model. Thus, we propose a new DNN by constraining the face-sketch drawing orders using five key intermediate images in human sketching, which has the advantage of implicitly embedding the human cognitive knowledge in the DNN-based sketch learning.

Because the existing face-sketch datasets do not contain intermediate sketches for a given face photo and cannot support to train our face-sketch synthesis network, we present a new order-enforced face-sketch dataset named Ord-Sketch. It is based on the face photos from the CUHK and CUFSF datasets. In our proposed Ord-Sketch dataset, for each face photo, 25 intermediate sketches with incremental degrees of fineness are sequentially collected using standard drawing procedures. Then, we propose a multi-stage network for sketch synthesis by enforcing the sketch-drawing order, called SO-Net. Then, owing to the limitation of memory, only five sketches from the total of 25 intermediate sketches, are selected as the intermediate supervision for the proposed SO-Net. Hence, the SO-Net consists of five stages, each of which adopts a conditional GAN (cGAN) [6,7] architecture to learn the corresponding intermediate sketches.

Order-enforced photo-to-sketch dataset. Our Ord-Sketch dataset is constructed by face photos and the corresponding sketches with incre-

* Corresponding author (email: xiaoming@iscas.ac.cn, qldong@nlpr.ia.ac.cn)

mental degrees of fineness. The dataset contains 400 face photos of 400 identities, 188 photos for Chinese identities from the CUHK dataset, and 212 for non-Chinese identities from the CUFSF dataset. Figure 1(a) shows an example sequence of 25 sketches for a given face photo in the Ord-Sketch dataset. In over-all, we included 400 face photos and 10000 sketches in the Ord-Sketch dataset. Then, after data collection, we aligned all the sketches based on the key facial landmarks using 2D similarity transformations. Also, we cropped all the sketches and face photos and resized them to 256×256 pixels (see Appendix A for more details).

Face-sketch synthesis network with human sketch-drawing order enforcement. Figure 1(b) shows the architecture of the proposed SO-Net. Although there is a sequence of 25 sketches with incremental degrees of fineness for each identity in the Ord-Sketch dataset, only five key intermediate sketches (1st, 7th, 15th, 20th, and 25th) are utilized to train the SO-Net in this study. Further, it is a good recommendation for future work to use all the 25 intermediate sketches to train the sketch synthesis network. The network and training details can be found in Appendix B.

Hence, the proposed SO-Net is a five-stage network with face photos as input. It generates five intermediate sketches for each identity with a multi-stage network. At each stage of the network, it uses a cGAN [7] to learn the intermediate sketch, whereas a U-net [8] is adopted as the generator and a Patch-GAN [7] is also adopted as the discriminator.

Let P denote an input face photo in the Ord-Sketch dataset, Z denote the initial random noise, $\{Y_1, \ldots, Y_5\}$ denote the set of real sketches at each stage, and $Y'_1, \ldots, Y'_5$ denote the set of synthesized sketches at each stage. Also, we depict the random noise at stage i to be Z_i $(i = 2, 3, 4, 5)$. The generator G_i at stage i is trained to generate Y'_i as follows:

$$Y'_i = \begin{cases} G_i(Z|P), & i = 1, \\ G_i(Z_i|(P, Y'_{i-1})), & i = 2, 3, 4, 5. \end{cases}$$

Loss functions. In general, our total loss is the sum of three losses: reconstruction loss $\mathcal{L}^i_{\text{rec}}$, edge loss $\mathcal{L}^i_{\text{edge}}$, and adversarial loss $\mathcal{L}^i_{\text{adv}}$:

$$\mathcal{L} = \sum_{i=1}^{5} \lambda_1 \mathcal{L}^i_{\text{rec}} + \lambda_2 \mathcal{L}^i_{\text{edge}} + L^i_{\text{adv}}, \qquad (1)$$

where λ_1 and λ_2 are loss weights. However, the reconstruction loss is used to calculate the difference between the predicted sketch and the ground truth. Inspired by [7], we use L_1 norm for the reconstruction loss. The reconstruction loss for the generator at a stage i is expressed as follows:

$$\mathcal{L}^i_{\text{rec}} = \|Y_i - Y'_i\|_1.$$

Because image edges have an important influence on face sketches, we used edge loss to enforce the edge constraint of the synthesized sketches of each stage. The edge loss of stage i is expressed as follows:

$$\mathcal{L}^i_{\text{edge}} = \|K \cdot Y_i - K \cdot Y'_i\|_1,$$

where K is the Sobel kernel.

Then, the basic adversarial loss for a cGAN is used as the loss for the discriminator loss at each stage of our network, and it is expressed as follows:

$$\mathcal{L}^i_{\text{adv}} = \begin{cases} \log D_i(Y'_i|P) + \log(1 - D_i(G_i(Z|P))), i = 1, \\ \log D_i(Y'_i|P, Y'_{i-1}) \\ \quad + \log(1 - D_i(G_i(Z_i|P, Y'_{i-1}))), i = 2, 3, 4, 5, \end{cases}$$

where G_i and D_i represent the generator and the discriminator at the i-th stage, respectively.

The optimal generators and discriminators are obtained by solving a min-max optimization on G_i and D_i $(i = 1, \ldots, 5)$ with $\mathcal{L}$ in (1). We train our proposed SO-Net with an iterative strategy. At the beginning of the training iterations, we trained the network with gradient ascend and descend with all the network parameters. However, after some training epochs, e.g., 88×1000 steps, the training of our network becomes unstable, so we improve the training strategy by deploying the divide-and-conquer approach (see Algorithm B1 of Appendix B for training details).

Experimental results. We conducted experiments on face-sketch synthesis using our proposed SO-Net method. Then, comparing the results to the other four methods, our method achieves the best SSIM, MS-SSIM, and MR scores and competitive PSNR score. Also, the synthesized face sketches with our proposed SO-Net achieved the best NLDA scores in the sketch-based face recognition (see Appendixes C and D for details).

Conclusion. In this study, initially, we created an order-enforced face-sketch dataset, called Ord-Sketch. The dataset contains 400 face photos, and each is accompanied by a set of 25 sketches with incremental degrees of fineness. Based on the Ord-Sketch, we proposed a multi-stage sketch synthesis network, called SO-Net, where we used five intermediate sketches as sketch-drawing order constraints. Moreover, experimental results demonstrated that the proposed network outperforms several state-of-the-art methods. Although we achieve significant results with our dataset and

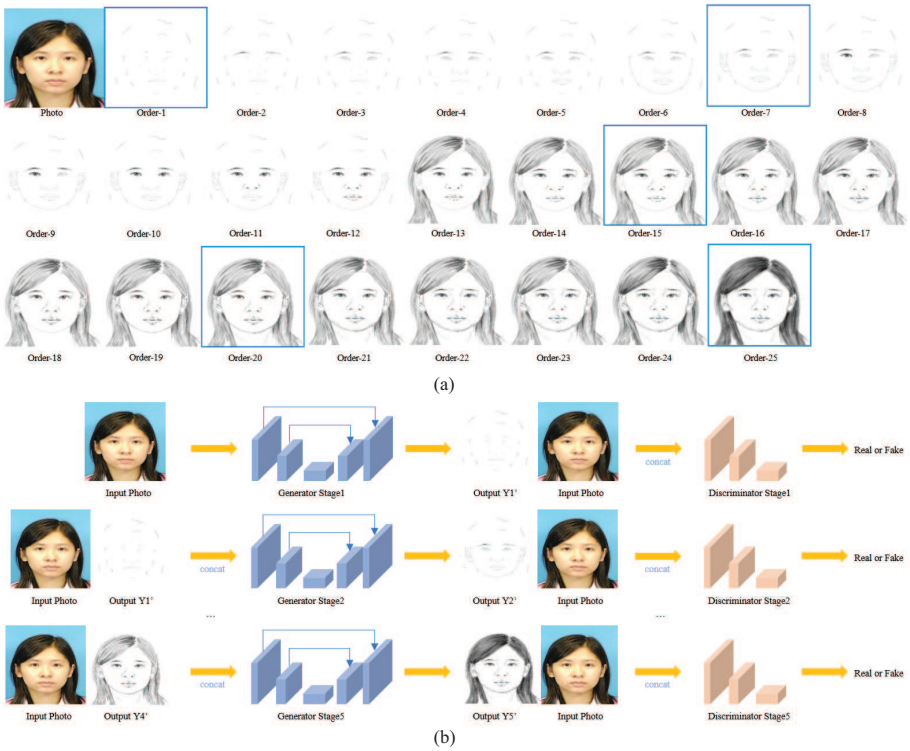

Figure 1 (Color online) (a) An example sequence of 25 sketches for a given photo in the Ord-Sketch dataset. The five key sketches selected for training are highlighted with blue rectangles. (b) The architecture of SO-Net, which is a five-stage cGAN model with photos as input.

model, there are still problems yet to be addressed, such as how to effectively utilize all the 25 sketches corresponding to a face photo for face synthesis. Besides, our Ord-Sketch dataset is expected to motivate related researches on face-sketch synthesis.

Acknowledgements This work was supported by National Key R&D Program of China (Grant Nos. 2017YFB1402105, 2019YFC1521100), National Natural Science Foundation of China (Grant Nos. U1805264, 61573359, 61672103, 61473276, 61402040), and Natural Science Foundation of Beijing (Grant No. L182052).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Wang X G, Tang X O. Face photo-sketch synthesis and recognition. IEEE Trans Pattern Anal Mach Intell, 2009, 31: 1955–1967
2 Zhang M, Wang N, Li Y, et al. Neural probabilistic graphical model for face sketch synthesis. IEEE Trans Neural Netw Learn Syst, 2019. doi: 10.1109/TNNLS.2019.2933590
3 Wang N, Gao X, Sun L, et al. Anchored neighborhood index for face sketch synthesis. IEEE Trans Circ Syst Video Technol, 2018, 28: 2154–2163
4 Zhang W, Wang X, Tang X. Coupled information-theoretic encoding for face photo-sketch recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011. 513–520
5 Khaligh-Razavi S M, Kriegeskorte N. Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput Biol, 2014, 10: e1003915
6 Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proceedings of Annual Conference on Neural Information Processing Systems (NIPS), 2014. 2672–2680
7 Isola P, Zhu J, Zhou T, et al. Image-to-image translation with conditional adversarial networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 5967–5976
8 Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention, 2015. 234–241