AgileGAN3D: Few-Shot 3D Portrait Stylization by Augmented Transfer Learning (Supplementary)

Guoxian Song[1]   Hongyi Xu[1]   Jing Liu[1]   Tiancheng Zhi[1]   Yichun Shi[1]
Jianfeng Zhang[1,2]   Zihang Jiang[1,2]   Jiashi Feng[1]   Shen Sang[1]   Linjie Luo[1]
[1]ByteDance Inc   [2]National University of Singapore

In this supplementary material, we provide a user study to evaluate perceptual quality of our approach over baseline methods in Section 1. We further provide a qualitative ablation study on the influence of camera pose accuracy over the stylization quality (Section 2). In Section. 3, we even evaluate our approach on single-shot 3D stylization. Please also find our training exemplars in Fig. S3 and more visual results and comparison in the supplementary html page.

1. User Study

Table S1. User Preference Score(PS) for different 3D stylization methods

	Ours	AgileGAN-EG3D	Toonify-EG3D
PS↑	**75.5%**	0.6%	23.9 %

We conduct a user study with 20 participants to select the best stylized images in term of user similarity and perceptual quality, between our method and baseline approaches (AgileGAN-EG3D and Toonify-EG3D). Each participant was shown 16 questions randomly selected from a question pool containing 200 examples (using test images with indices 0-200 in the CelebA-HQ dataset). The preference score (PS) is evaluated as the preference ratio of all the selections. Table. S1 shows that results from our proposed method have the majority preference.

2. Camera Pose Ablation Study

In Fig. S1, we qualitatively evaluate the effect of camera pose accuracy over the stylization quality to further demonstrate the benefits of our style prior. The experiment is performed over Ukiyoe style, over which even the state-of-the-art pose estimators [2, 3] fail to detect human and obtain 3D camera pose. In contrast, our approach does not rely on pose estimation from stylized images but is always able to obtain accurate camera labels estimated from the paired real images, even for non-realistic or abstract styles. To conduct a fair comparison on 3D stylization, we use the same training data with three different settings for camera pose: 1) camera pose obtained from its paired real images with our style prior. 2) random camera poses towards frontal face. 3) a constant frontal camera pose. With accurately estimated camera poses for the volume rendering and dual discriminator, our method achieves the best 3D stylization quality.

3. Single-shot 3D Stylization

Our pipeline is a general framework for 3D stylization where the 2D style prior can be even created from a single style exemplar. Here we employ the recent one-shot 2D stylization pipeline [1] and train the style prior using a single style exemplar. Combined with our proposed augmented transfer learning strategy for 3D GAN, we successfully demonstrate 3D view-consistent stylization in Fig. S2. We note that compared to few-shot learning, one-shot stylization might reduce the generative diversity but it is more a limitation in 2D single-shot approaches.

References

[1] Min Jin Chong and David Forsyth. Jojogan: One shot face stylization. *arXiv preprint arXiv:2112.11641*, 2021. 1

[2] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *IEEE Computer Vision and Pattern Recognition Workshops*, 2019. 1

[3] Sicheng Xu, Jiaolong Yang, Dong Chen, Fang Wen, Yu Deng, Yunde Jia, and Xin Tong. Deep 3d portrait from a single image, 2020. 1

input   (a) ours   (b) random camera   (c) front camera
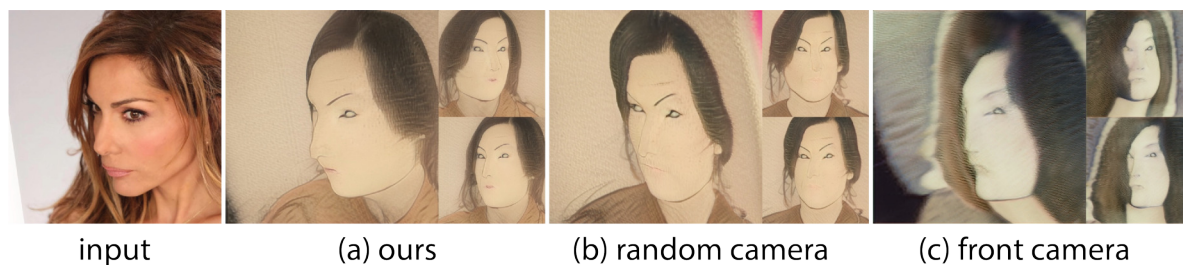
Figure S1. Our style prior enables accurate pose estimations from paired real images, with which we observe superior perceptual quality and user similarity in 3D stylization. Direct pose estimation from Ukiyoe style images fails.
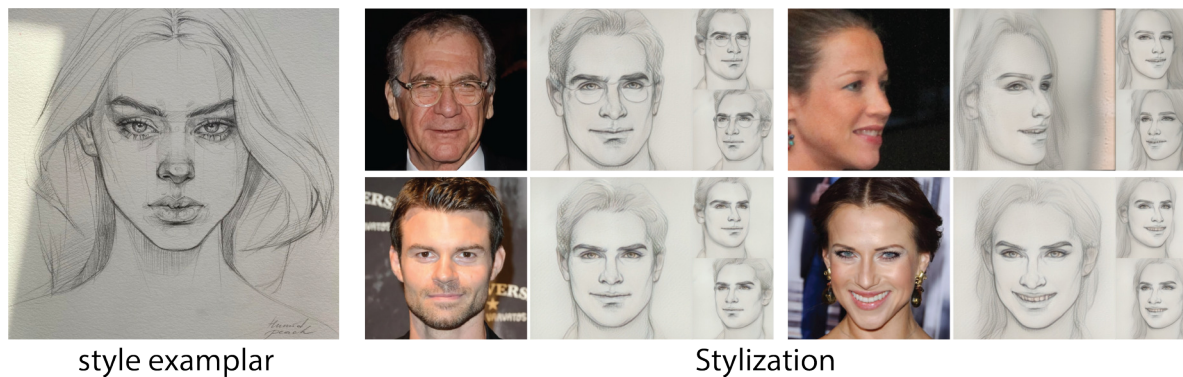

style examplar   Stylization

Figure S2. Our pipeline supports one-shot stylization and enables multi-view synthesising.
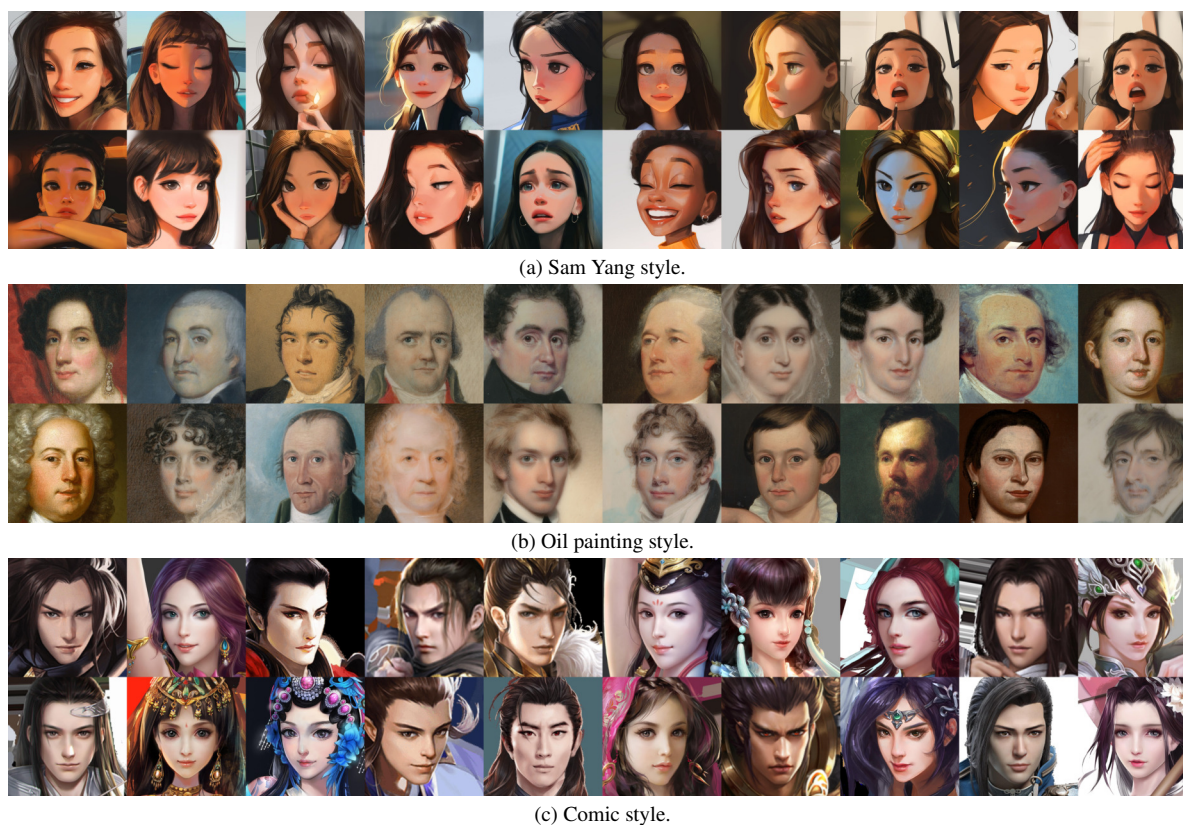

(a) Sam Yang style.

(b) Oil painting style.

(c) Comic style.

Figure S3. The few-shot training style exemplars.