One-shot Portrait Stylization via Geometric Alignment Supplementary Material

Xinrui Wang[1] Zilin Guo[2] Zhuoru Li[3] Jinze Yu[4]
Heng Zhang[5] Yusuke Iwasawa[1] Yutaka Matsuo[1] Jiaxian Guo[1]
[1]The University of Tokyo [2]Xiaomi Corporation [3]Project HAT
[4]Waseda University [5]Adaspace Technology Co.Ltd

1. Implementation Details.

In this paper, we use SDXL [3] as the backbone diffusion model, and LoRA weights for all styles are trained with AdamW optimizer [2] at learning rate 1×10^{-5} and batch size=1. We at first train the content LoRA on the content images for 250 steps, and then jointly train content LoRA and style LoRA for another 250 steps. During inference, we integrate ControlNet with canny edge into the framework for image-to-image translation.

We collect our style examples from the internet and use FFHQ [1] dataset as content images. Especially, images with number 0-999 are used for test data and images with number 1000-9999 are used for training. All images are firstly resized to 544*544 resolution, after alignment and TPS warping, they are center-cropped to 512*512 size to avoid the artifacts near the edge area caused by warping. During training, we use 28 facial landmarks for facial image alignment, and combine 8 edge points of the rectangle images with the 28 facial landmarks for TPS warping.

2. Extreme cases of style references.

Our landmark detector can precisely detect landmarks in non-photo realistic style examples in extreme poses and even animals. We illustrate some extreme cases with detected landmarks in Fig1. All style images and corresponding landmarks will be made available upon paper acceptance.

3. Images Shown in the User Study.

In the user study, we invited 40 participants and show each candidate stylized images with 4 different styles: Animation style, Watercolor style, Oilpaint style and Caricature style. We prepared 48 images in total with 12 images for each style, and show each candidate 8 images for each style and 32 images in total. Here we present the user interface of user study and all the 48 images shown in the user study in Figure 3 4, 5, 6

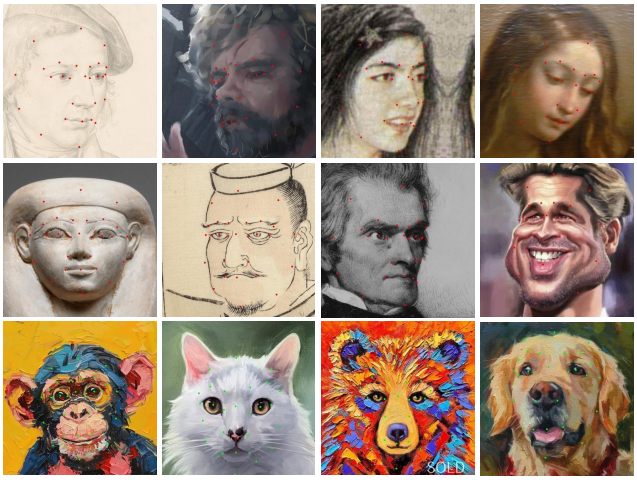

Figure 1. Extreme cases of style images with detected landmarks. Our method can even detect animal faces. Zoom in for details.

References

[1] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 1

[2] I Loshchilov. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 1

[3] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023. 1

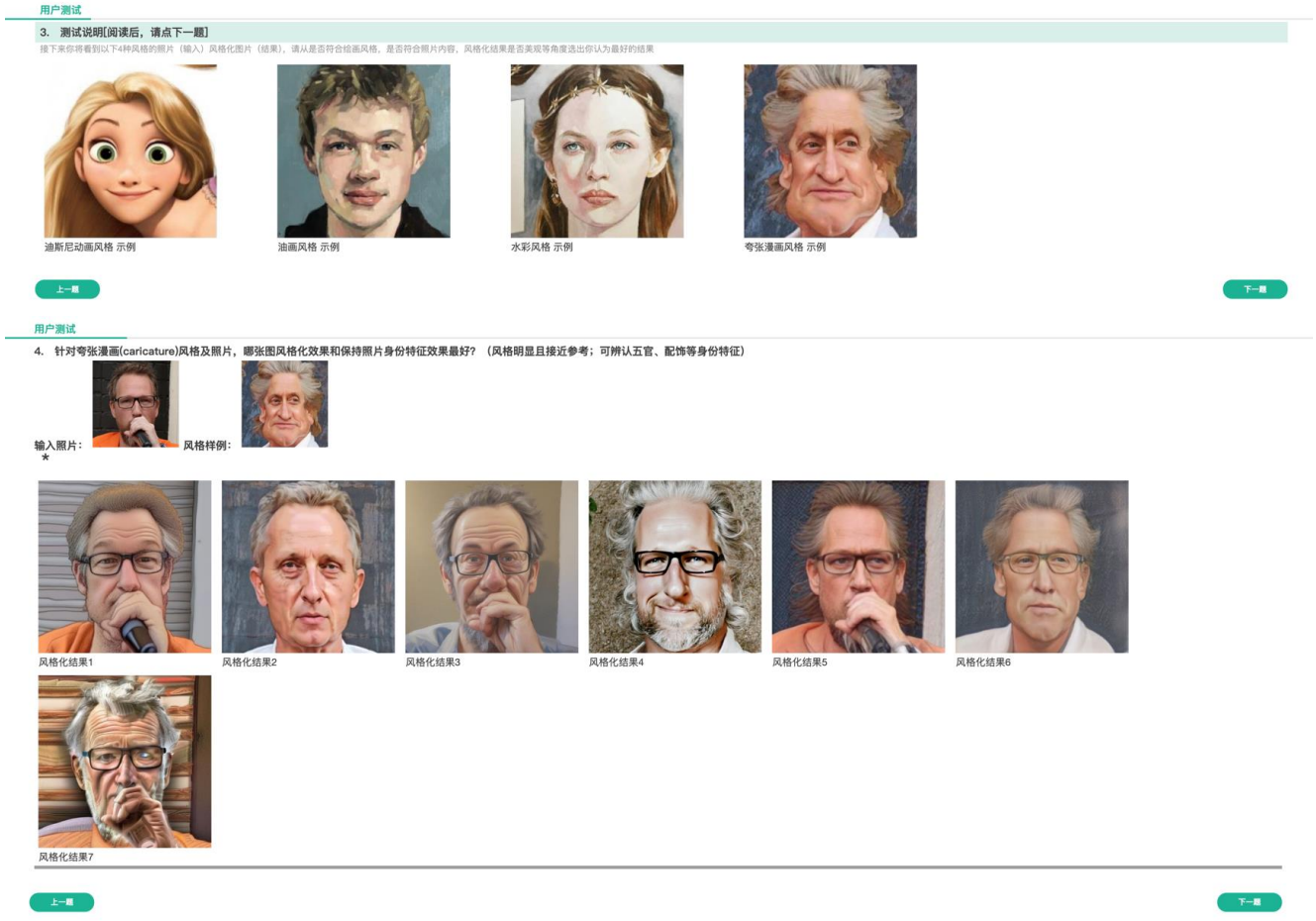

Figure 2. User interface used for the user study.

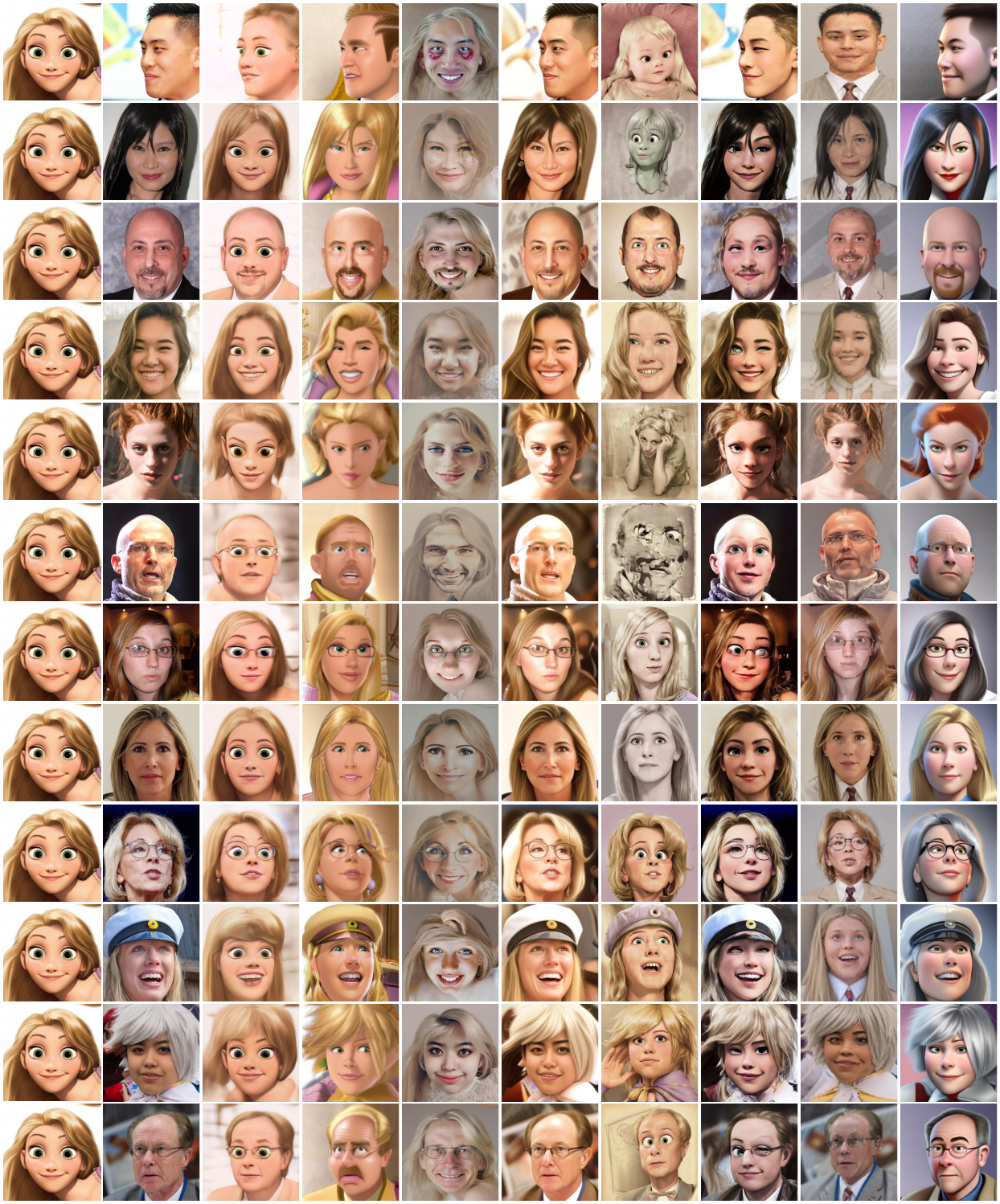

Figure 3. Stylization results of animation style shown in the user study.

Figure 4. Stylization results of watercolor style shown in the user study.

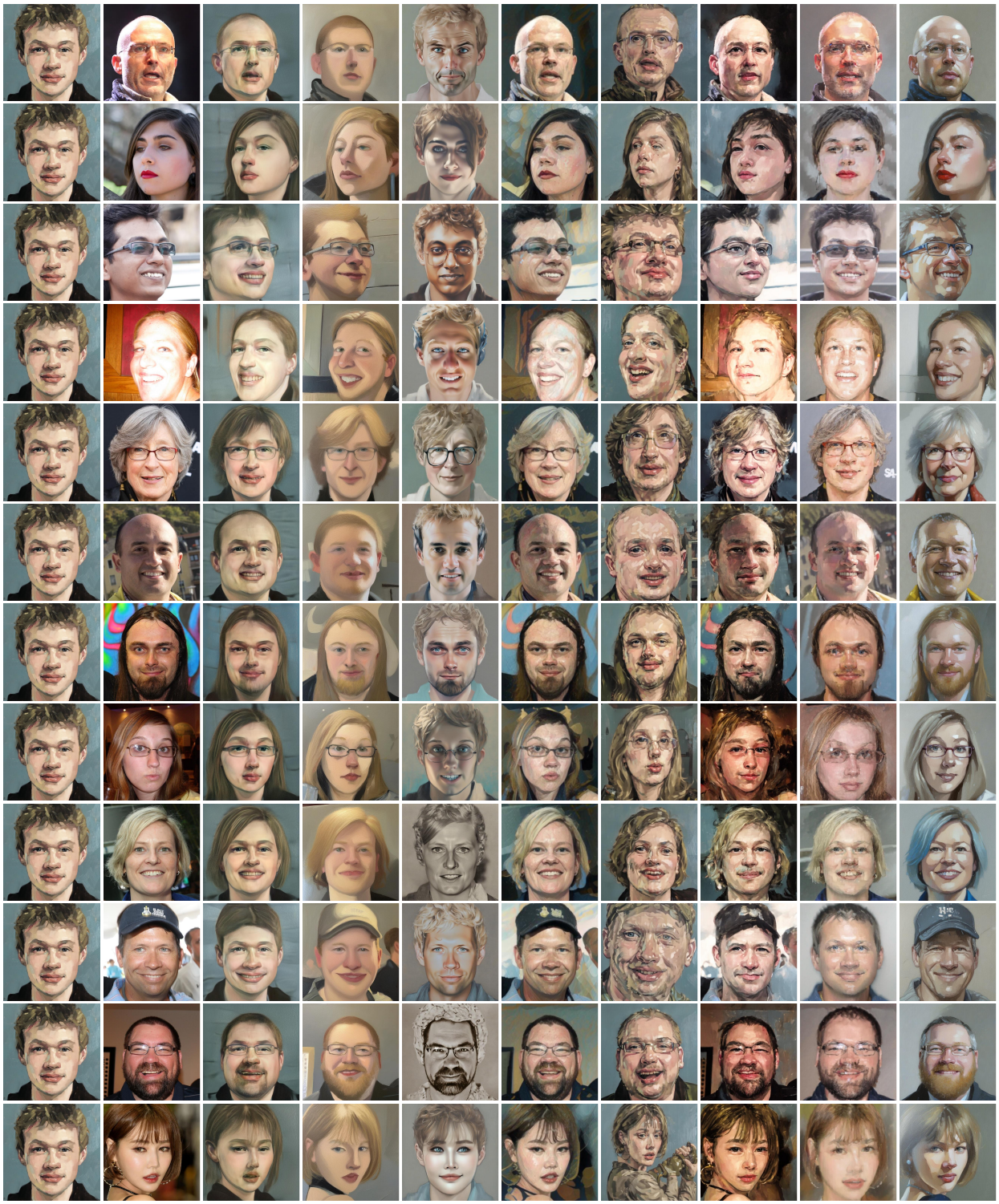

Figure 5. Stylization results of oilpaint style shown in the user study.

Figure 6. Stylization results of caricature style shown in the user study.