Supplementary material for *Picture that Sketch*: Photorealistic Image Generation from Abstract Sketches

Subhadeep Koley^{1,2} Ayan Kumar Bhunia¹ Aneeshan Sain^{1,2} Pinaki Nath Chowdhury^{1,2} Tao Xiang^{1,2} Yi-Zhe Song^{1,2} ¹SketchX, CVSSP, University of Surrey, United Kingdom. ²iFlyTek-Surrey Joint Research Centre on Artificial Intelligence.

{s.koley, a.bhunia, a.sain, p.chowdhury, t.xiang, y.song}@surrey.ac.uk

A. Additional Results

Figs. 1-6 depict additional sketch-to-photo generation results with sketches from QMUL-ShoeV2 [1, 6], QMUL-ChairV2 [1, 6], and Handbag [7] datasets. A common observation seen in Figs. 1-6 is that, although photorealistic, the output quality is comparatively lower in ChairV2 and Handbag datasets than that in ShoeV2 due to a higher degree of sketch abstraction and a lower number of training pairs available in the former two datasets.



Figure 1. Results generated by the proposed method with sketches from the QMUL-ShoeV2 [1,6] dataset.



Figure 2. Results generated by the proposed method with sketches from the QMUL-ShoeV2 [1,6] [1,6] dataset.



Figure 3. Results generated by the proposed method with sketches from the QMUL-ChairV2 [1,6] dataset.



Figure 4. Results generated by the proposed method with sketches from the QMUL-ChairV2 [1,6] dataset.



Figure 5. Results generated by the proposed method with sketches from the Handbag [7] dataset.



Figure 6. Results generated by the proposed method with sketches from the Handbag [7] dataset.

B. Details on Human Study

Fig. 7 and Fig. 8 depict the login and scoring screen of the interface used to collect the MOS [3] values. Upon login, a user first selects a class (*i.e.*, Shoe, Chair, or Handbag) and draws a sketch of that class. Next, upon clicking on the "Generate" button, the system displays corresponding photo translations produced by our proposed method along with every other competing framework. The participant rates every generated photo and clicks on "Submit & Next" to continue. We anonymise the names of the competing methods (*e.g.*, pix2pix [4], MUNIT [2], etc.) to prevent the ratings from being influenced by the participant's past knowledge. For brevity and ease of the participants, following [3], we sub-divide the $1\rightarrow 5$ (bad→excellent) MOS levels into a nine-point discreet scale with the possible ratings $\{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$. For each method, we compute the final MOS [3] value by taking the mean (μ) and variance (σ) of all individual MOS responses.



Figure 7. Login screen of our human study interface



Figure 8. Scoring screen of our human study interface

C. Intermediate Photo Generation

Leveraging StyleGAN's [5] smooth latent space, our method allows us to generate realistic transitional photos given an initial and a final sketch. We achieve this via simple arithmetic operations (*e.g.*, interpolation) between the predicted latent codes of the two sketches. Fig. 9 and 10 shows how given *Sketch A* and *Sketch B*, our method can generate plausible intermediate photos.



Figure 10. Generating transitional handbag photos between Sketch A and B.

References

- Ayan Kumar Bhunia, Yongxin Yang, Timothy M Hospedales, Tao Xiang, and Yi-Zhe Song. Sketch Less for More: On-the-Fly Fine-Grained Sketch Based Image Retrieval. In CVPR, 2020. 1, 2, 3, 4
- [2] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal Unsupervised Image-to-Image Translation. In ECCV, 2018. 7
- [3] Quan Huynh-Thu, Marie-Neige Garcia, Filippo Speranza, Philip Corriveau, and Alexander Raake. Study of Rating Scales for Subjective Quality Assessment of High-Definition Video. *IEEE TBC*, 2010. 7
- [4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-Image Translation with Conditional Adversarial Networks. In CVPR, 2017. 7
- [5] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and Improving the Image Quality of StyleGAN. In CVPR, 2020. 8
- [6] Jifei Song, Kaiyue Pang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Learning to Sketch with Shortcut Cycle Consistency. In CVPR, 2018. 1, 2, 3, 4
- [7] Jifei Song, Qian Yu, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Deep Spatial-Semantic Attention for Fine-Grained Sketch-Based Image Retrieval. In *ICCV*, 2017. 1, 5, 6