

Image Referenced Sketch Colorization Based on Animation Creation Workflow Supplementary Materials

Zhuoru Li³ *Dingkun Yan^{1, 4} *Xinrui Wang²
 Suguru Saito¹ Yusuke Iwasawa² Yutaka Matsuo² Jiaxian Guo²
¹Institute of Science Tokyo ²The University of Tokyo
³Project HAT ⁴Tokyo AI research institute, Shanda Group

1. Implementation details

We show the detailed structure of Diffusion UNet block in the proposed framework with split cross-attention in Figure 1. We also attach the source code of split cross attention with the supplementary in *lora.py* file.

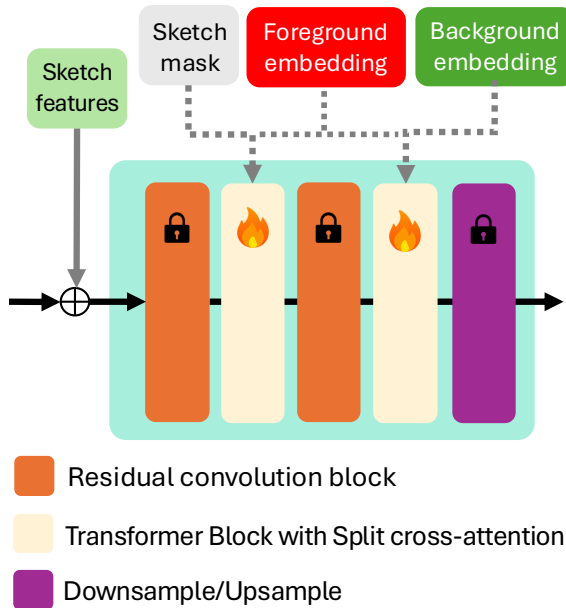


Figure 1. The implementation details of a block in the Diffusion UNet of proposed framework with split cross-attention.

2. Comparison on Color Reference Extraction

We show the comparison of local embeddings and CLS embeddings for color reference extraction in Figure 2. As is shown in Figure 2 (1), using CLS token as color reference causes blurry textures, as it only expresses semantic-level notions, such as objects, global style, image size and quality. High-dimensional local embeddings, on the other hand, enable the network to generate high-quality images by synthesizing sharper details and finer textures from the detailed spatial information it contains. 2 (2). Local embeddings achieve better quantitative performance evaluated by FID.

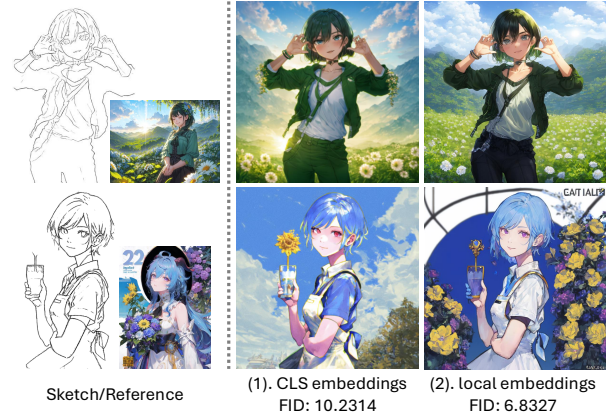


Figure 2. Comparison of local embeddings and CLS embedding for color reference extraction. Zoom in for details.



Figure 3. Example of failure cases.

However, adopting local embeddings leads to higher probability of spatial entanglement. We solve this with the proposed method in the main content.

3. Images Shown in the User Study

In the user study, we present 25 sets of sketch-reference pairs and results generated by all compared methods to users investigated. The images are shown in Figure 4 and 5. We also attach the high-resolution images in the zip file.

4. Failure Cases

Failure may happen when the animation segmentation tool applied in the framework fails to extract masks from sketches, where the Fig2Fig mode cannot effectively colorize the sketch image based on the reference. We illustrate the failure case in Figure 3.

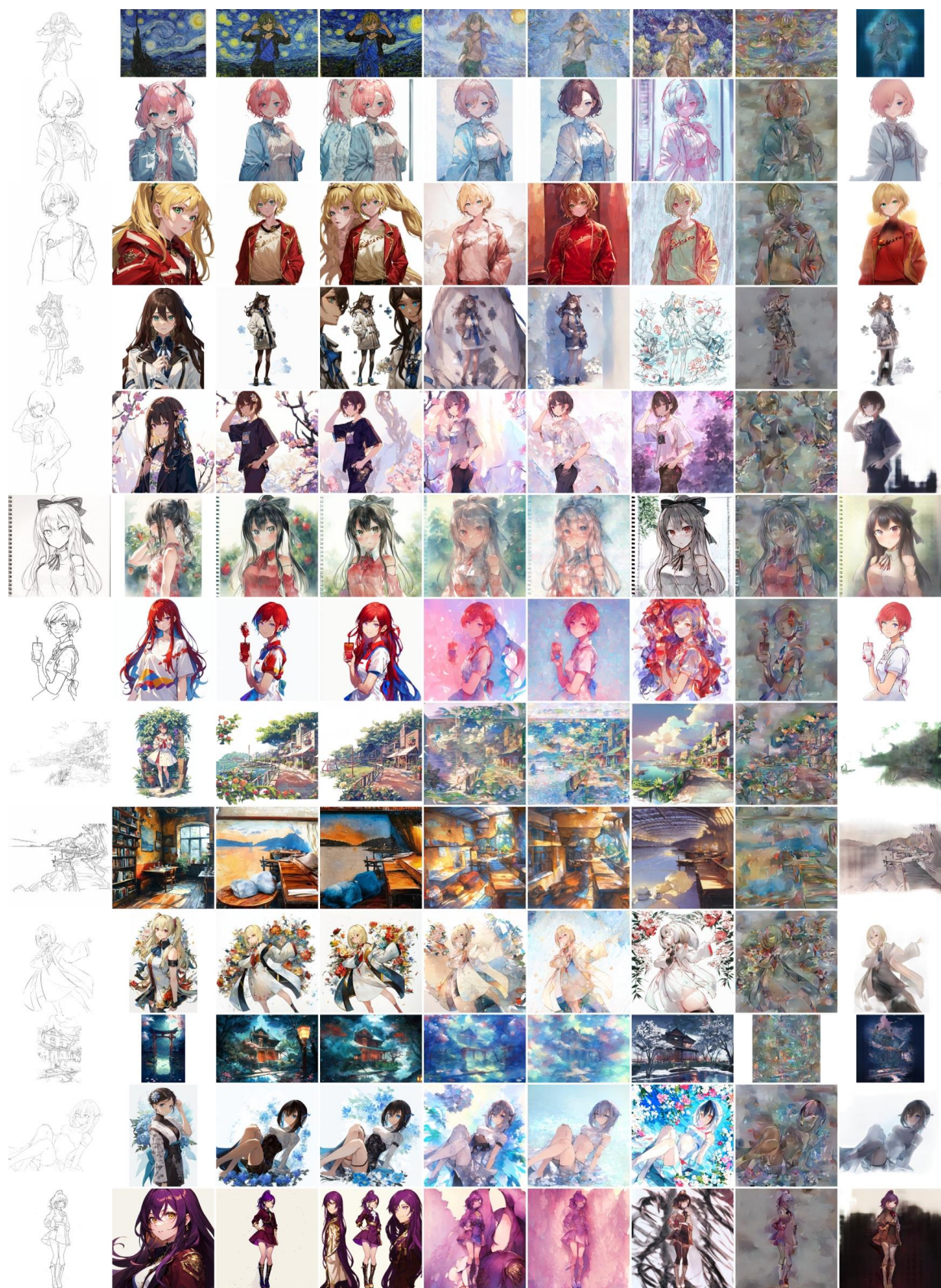


Figure 4. The images shown in the User study.

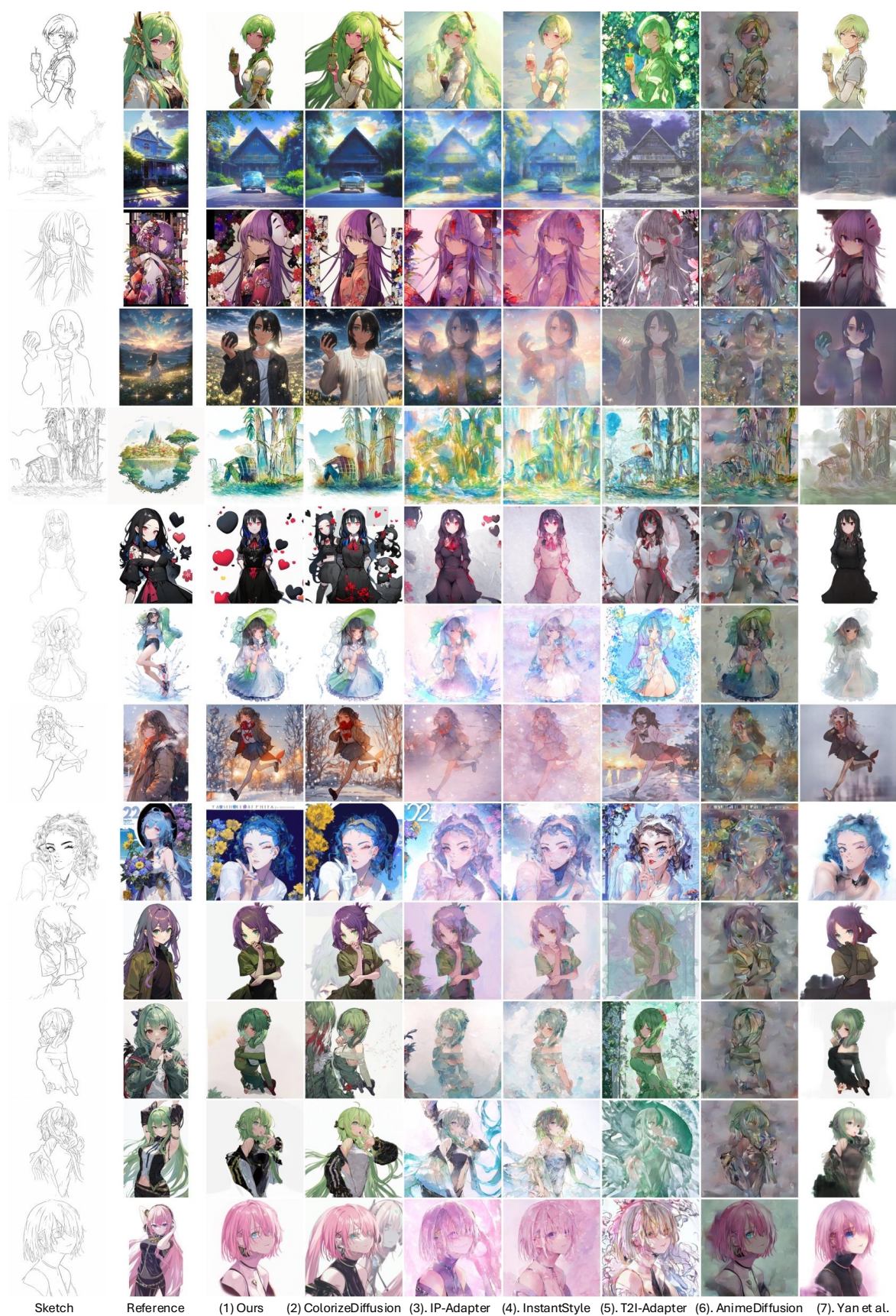


Figure 5. The images shown in the User study.