Supplementary Material: Improved Diffusion-based Image Colorization via Piggybacked Models

Hanyuan Liu[1] Jinbo Xing[2] Minshan Xie[3] * Chengze Li[1] Tien-Tsin Wong[4]
[1] Saint Francis University [2] Alibaba Group
[3] Guangdong University of Technology [4] Monash University
{hliu, czli}@sfu.edu.hk jbxing@cse.cuhk.edu.hk msxie@gdut.edu.cn tt.wong@monash.edu

*Corresponding author.

A. Showcase for text-based image colorization

In contrast to existing text-based image colorization methods such as [**? ? ? ?**], our framework supports arbitrary text inputs. To demonstrate this capability, we present a set of colorization results with various text inputs in Figure 1.

Figure 1. Colorization results with various text inputs.

B. Additional comparative results

In the main paper, we only present a qualitative comparison with three of the most competitive colorization methods: Deoldify [**?**], ColTran [**?**], and Disco [**?**]. In this section, we provide a comprehensive visual comparison with recent colorization methods, including CIColor [**?**], UGColor [**?**], Deoldify [**?**], ChromaGAN [**?**], InstColor [**?**], ColTran [**?**], UniColor [**?**], and Disco [**?**]. The results are shown in Figure 2, Figure 3, Figure 4, and Figure 5.

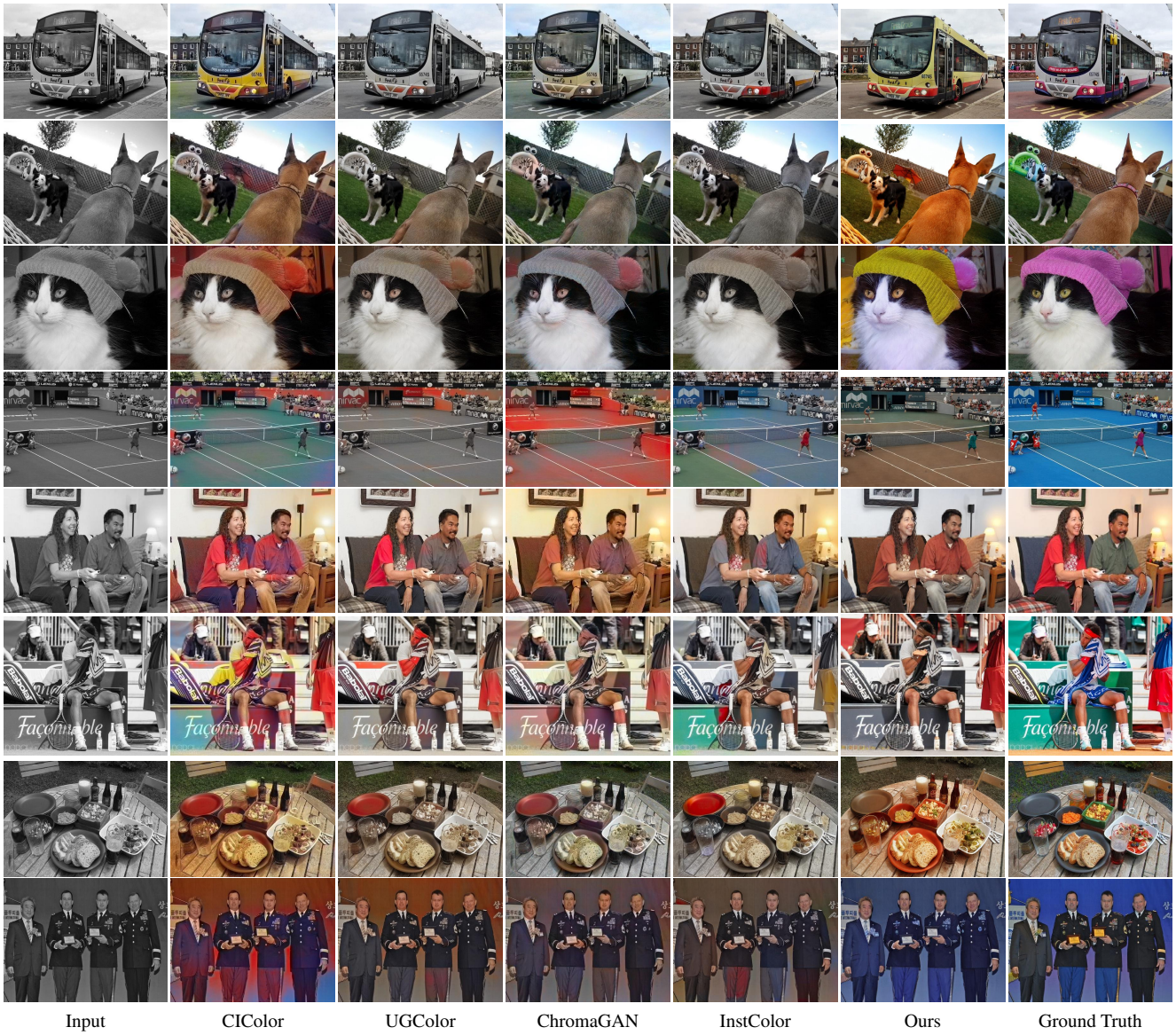

Figure 2. Comparative showcase of challenging cases for unconditional image colorization.

We also present a visual comparison with Palette [**?**], which is currently the state-of-the-art diffusion-based method for image colorization. However, it should be noted that the source code and pre-trained model of Palette are not publicly available. Therefore, we use the images provided in the original paper for the comparison. The results are shown in Figure 6.

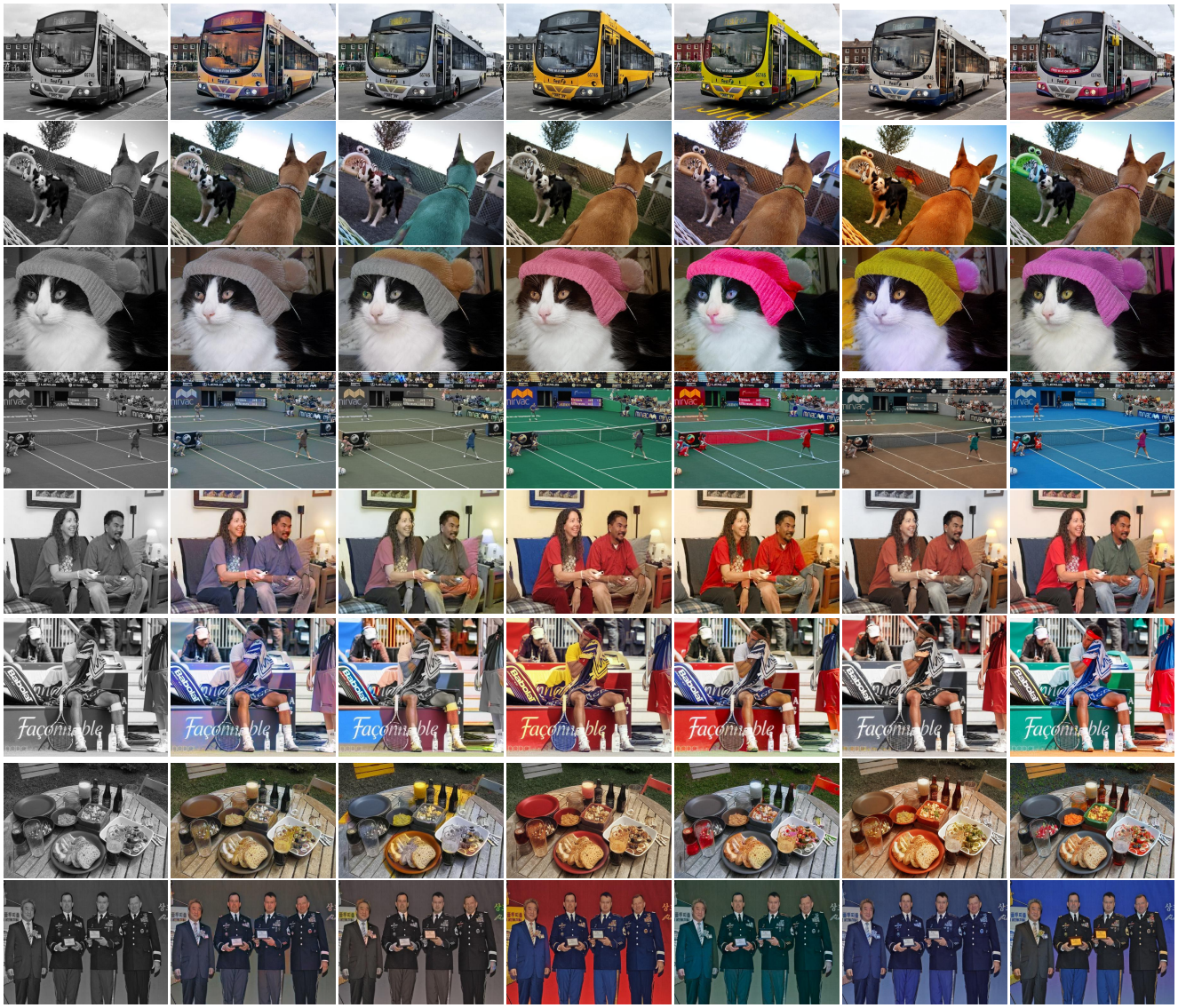

Figure 3. Comparative showcase of challenging cases for unconditional image colorization.

C. Network architecture

The network architecture of the lightness-aware VAE model is illustrated in Figure 7. The grayscale encoder is a replica of the original VAE encoder but with the exclusion of the middle blocks. The latent diffusion guider model is a replica of the stable diffusion model as proposed by [**?**], but with modifications made to the input layers to facilitate the acceptance of concatenated inputs. The grayscale input and color hint map is encoded with a encoder using the same architecture of original VAE encoder.

Input CIColor UGColor ChromaGAN InstColor Ours

Figure 4. Comparative showcase for old photo image colorization.

Input Deoldify Coltran Disco Unicolor Ours

Figure 5. Comparative showcase for old photo image colorization.

Figure 6. Visual comparison with Palette diffusion model.

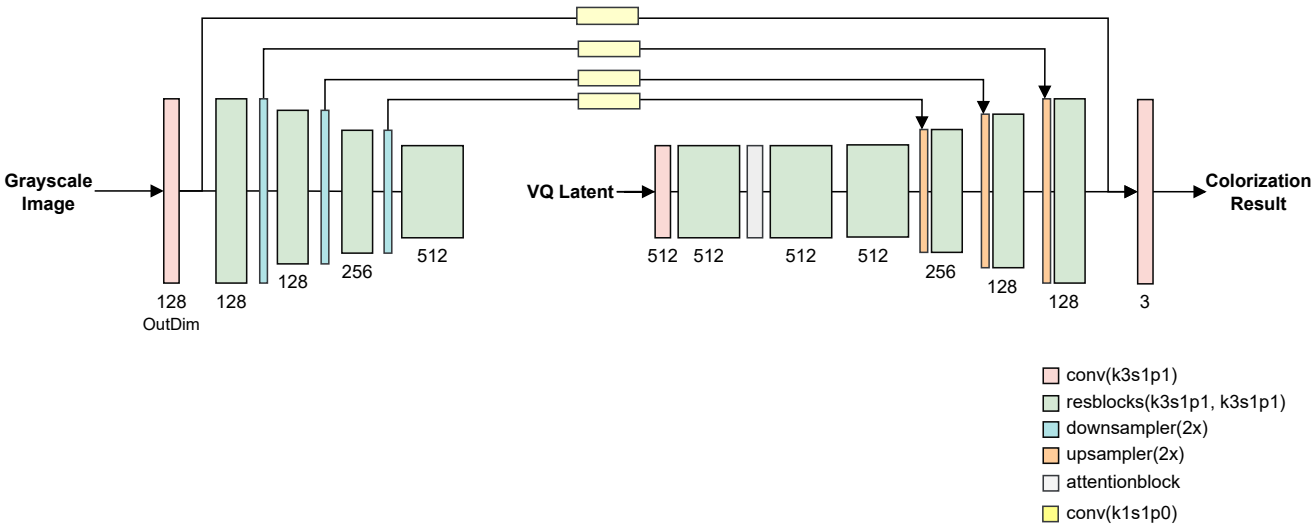

Figure 7. Network architecture of lightness-aware VAE model.