Supplementary Materials to “Pixel-Aware Stable Diffusion for Realistic Image Super-resolution and Personalized Stylization”

Tao Yang[1], Rongyuan Wu[2], Peiran Ren[3], Xuansong Xie[3], and Lei Zhang[2]

[1] Bytedance Inc.
[2] The Hong Kong Polytechnic University
[3] Alibaba Group

In this supplementary file, we provide the following materials:

- More visual comparisons of different Real-ISR models;
- Various stylization results generated by PASD;
- More visual comparisons of different old image restoration methods;
- Results of PASD on image colorization.

1 More Real-ISR Results

In Fig. 1, we show more visual comparisons between our method with state-of-the-art Real-ISR methods, including RealESRGAN [10], SwinIR [5], StableSR [9], DiffBIR [7] and SeeSR [12]. Similar conclusions to the main paper can be made. With the help of PACA and ANS modules, our PASD can provide adjustable pixel-level guidance on the image generation, reproducing more realistic fine details and less visual artifacts.

2 Various Stylization Results

As mentioned in the main paper, by simply switching the base diffusion model to a personalized one, our proposed PASD can do various stylization tasks without any additional training procedure. In the main paper, we have provided the results by using the ToonYou style. In Fig 2 of this supplementary file, we show more types of stylization results by using the personalized base models of Disney 3D, Oil painting and Shinkai. One can see that our PASD method can keep very well the pixel-wise image details while performing style transfer.

3 More Old Image Restoration Results

In Fig. 3, we show more visual comparisons between our method with state-of-the-art old image restoration methods, including Real-ESRGAN [10], FeMaSR [2], StableSR [9], and DiffBIR [7]. It can be seen that PASD can better recover semantic-aware and photo-realistic image details.

4 Image Colorization

Our PASD can serve as a generic solution for various pixel-wise image-to-image tasks. In addition to Real-ISR, personalized stylization and old photo restoration, we also apply it to image colorization and show the results in this supplementary file.

The proposed ANS in the main paper can effectively solve the non-zero SNR issue without any extra training in tasks such as Real-ISR. However, it cannot be directly applied to the task of image colorization, where the residual signal from RGB image cannot be compensated from the input grayscale image. To solve this problem, we follow the idea proposed in [6] to scale the noise schedule as follows:

$$\sqrt{\bar{\alpha'}_t} = \frac{\sqrt{\bar{\alpha}_t} - \sqrt{\bar{\alpha}_N}}{\sqrt{\bar{\alpha}_1} - \sqrt{\bar{\alpha}_N}}, \tag{1}$$

where $t \in \{1, 2...N\}$ and α' is the scaled α. One can see that $\sqrt{\bar{\alpha'}_{N-1}} \neq 0$ and $\sqrt{\bar{\alpha'}_N} = 0$, which means that α'_N has been successfully set to 0.

When SNR is zero, the ϵ prediction used in SD becomes a trivial task because the ϵ loss cannot guide the model to learn meaningful information from the data. We therefore follow [6,8] to finetune the PASD model with rescaled noise schedule and ν loss and prediction.

Figure 4 shows the qualitative comparisons between PASD and the state-of-the-art image colorization methods, including DeOldify [1], BigColor [4], CT2 [11] and DDColor [3]. One can see that our PASD generates more photo-realistic and vivid colorization results. In particular, it significantly alleviates the color bleeding effect, which often happens in the compared methods.

References

1. Anti, J.: jantic/deoldify: A deep learning based project for colorizing and restoring old image (and videos!) (2019), https://github.com/jantic/DeOldify 2
2. Chen, C., Shi, X., Qin, Y., Li, X., Han, X., Yang, T., Guo, S.: Real-world blind super-resolution via feature matching with implicit high-resolution priors. In: ACM MM (2022) 1
3. Kang, X., Yang, T., Ouyang, W., Ren, P., Li, L., Xie, X.: Ddcolor: Towards photo-realistic and semantic-aware image colorization via dual decoders. In: ICCV (2023) 2
4. Kim, G., Kang, K., Kim, S., Lee, H., Kim, S., Kim, J., Baek, S.H., Cho, S.: Bigcolor: Colorization using a generative color prior for natural images. In: ECCV (2022) 2
5. Liang, J., Cao, J., Sun, G., Zhang, K., Gool, L.V., Timofte, R.: Swinir: Image restoration using swin transformer. ArXiv (2021) 1
6. Lin, S., Liu, B., Li, J., Yang, X.: Common diffusion noise schedules and sample steps are flawed. In: ArXiv (2023) 2
7. Lin, X., He, J., Chen, Z., Lyu, Z., Fei, B., Dai, B., Ouyang, W., Qiao, Y., Dong, C.: Diffbir: Towards blind image restoration with generative diffusion prior. In: ArXiv (2023) 1
8. Salimans, T., Ho, J.: Progressive distillation for fast sampling of diffusion models. In: ICLR (2022) 2

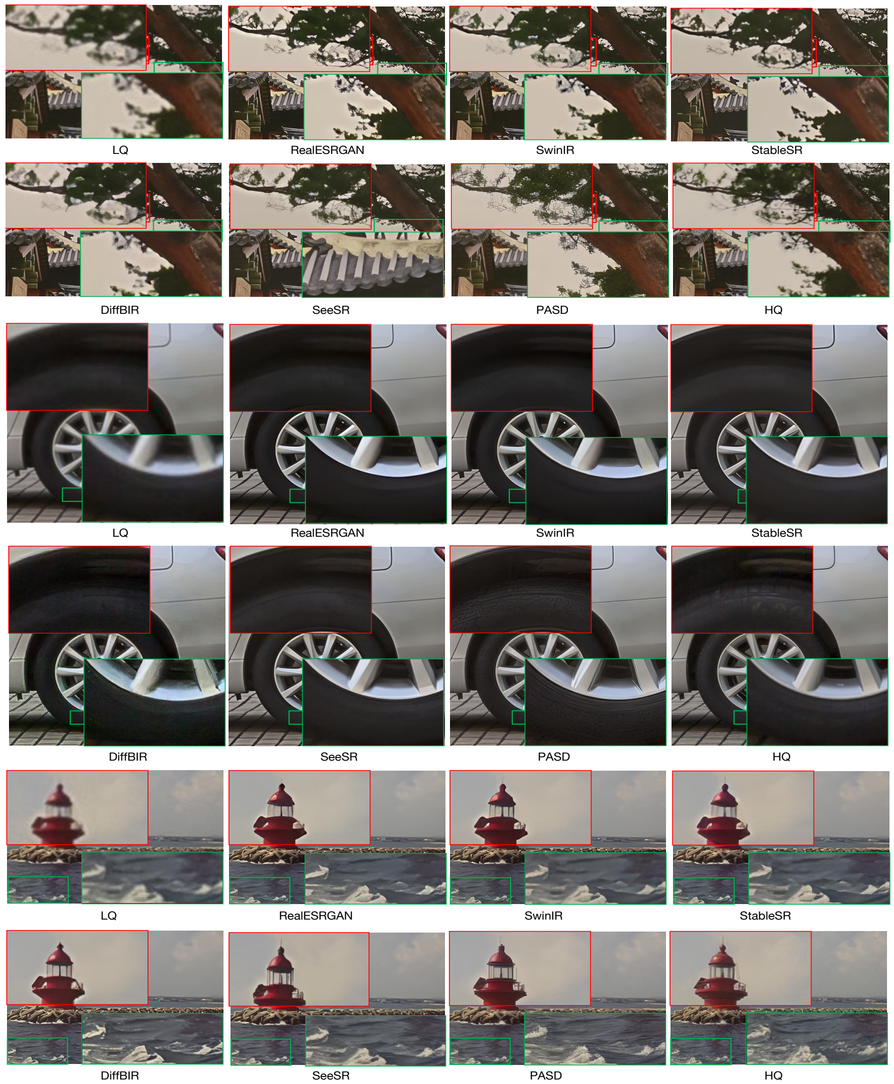

Fig. 1: Realistic image super-resolution results by different methods. Please zoom-in for better comparison.

9. Wang, J., Yue, Z., Zhou, S., Chan, K.C., Loy, C.C.: Exploiting diffusion prior for real-world image super-resolution. In: Arxiv (2023) 1
10. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: ICCVW (2021) 1
11. Weng, S., Sun, J., Li, Y., Li, S.: Ct2: Colorization transformer via color tokens. In: ECCV (2022) 2
12. Wu, R., Yang, T., Sun, L., Zhang, Z., Li, S., Zhang, L.: Seesr: Towards semantics-aware real-world image super-resolution. In: CVPR (2023) 1

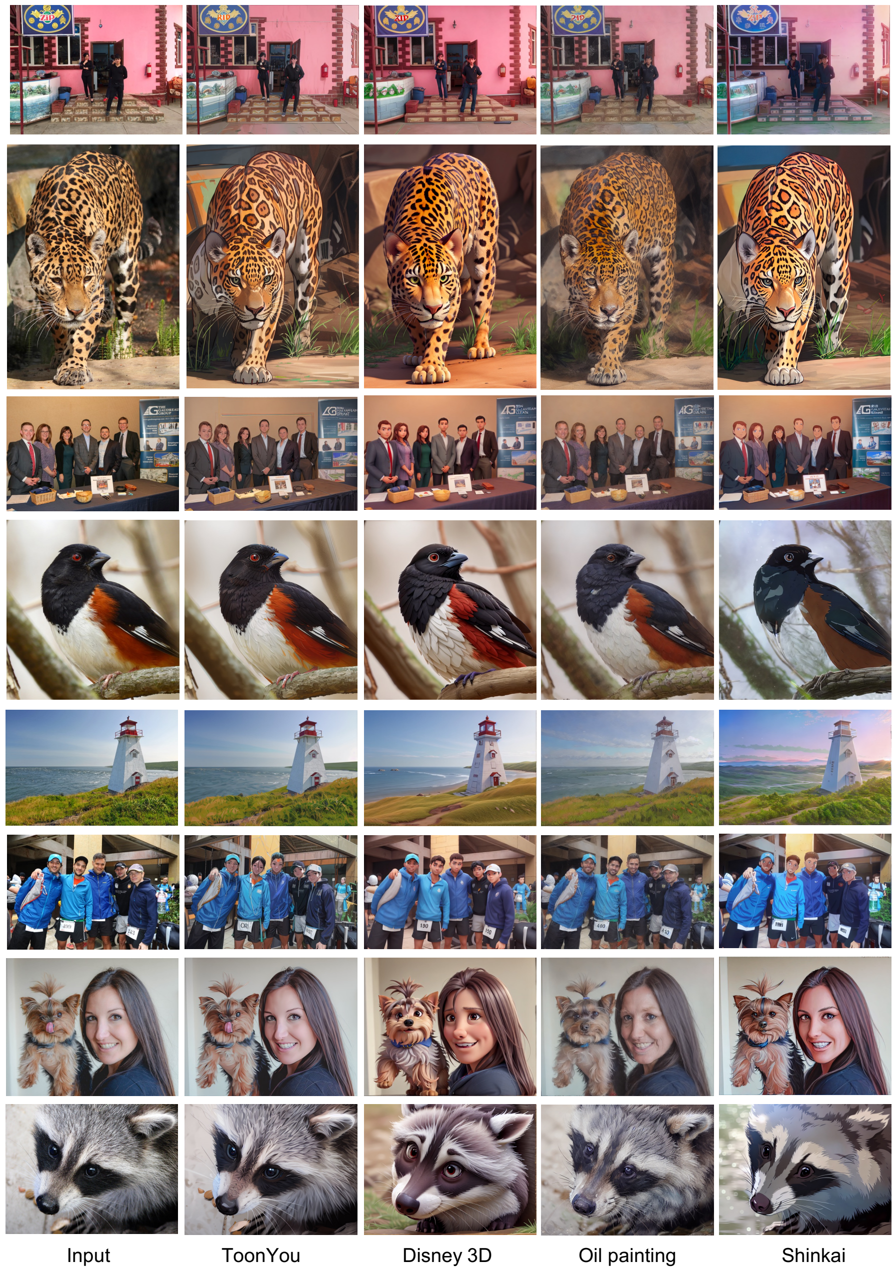

Fig. 2: Stylization results by PASD with different base models (ToonYou, Disney 3D, Oil painting, Shinkai) on real-world images.

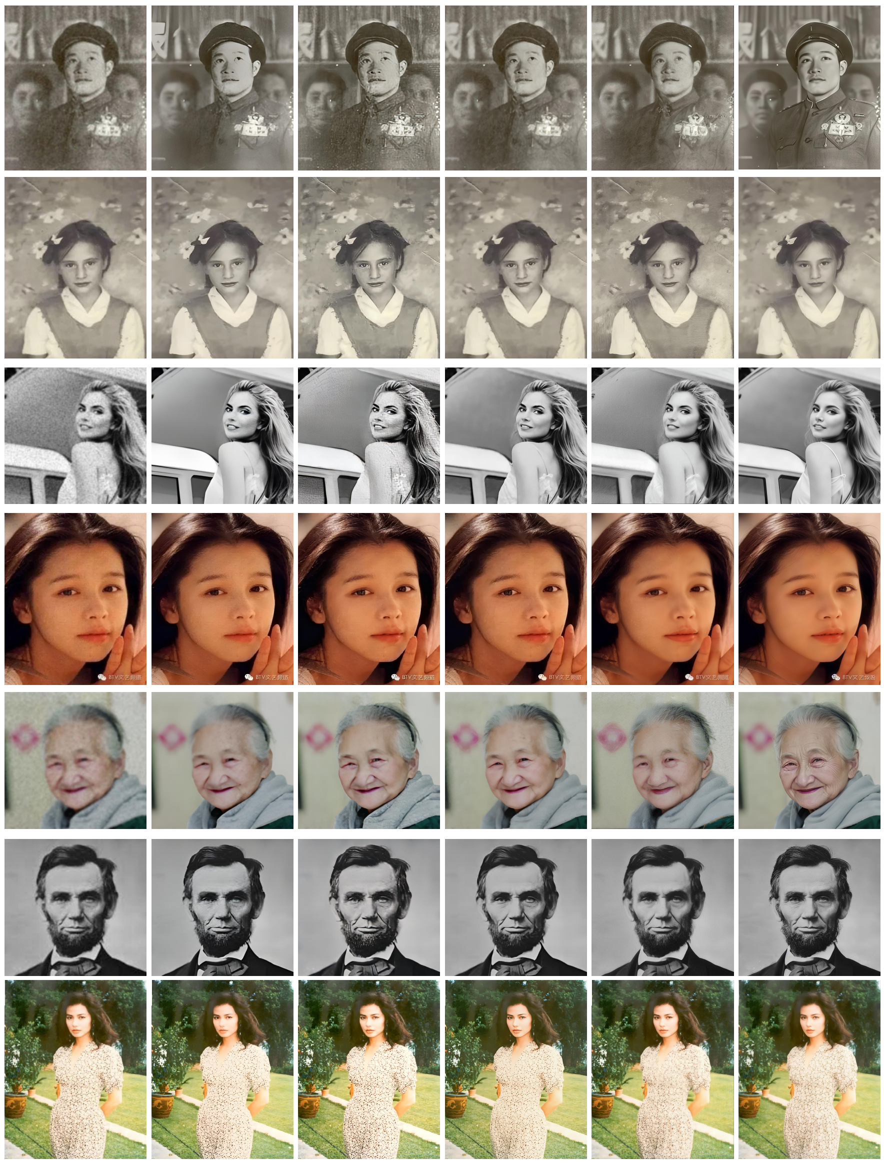

Fig. 3: Old image restoration results by different methods. Please zoom-in for better comparison.

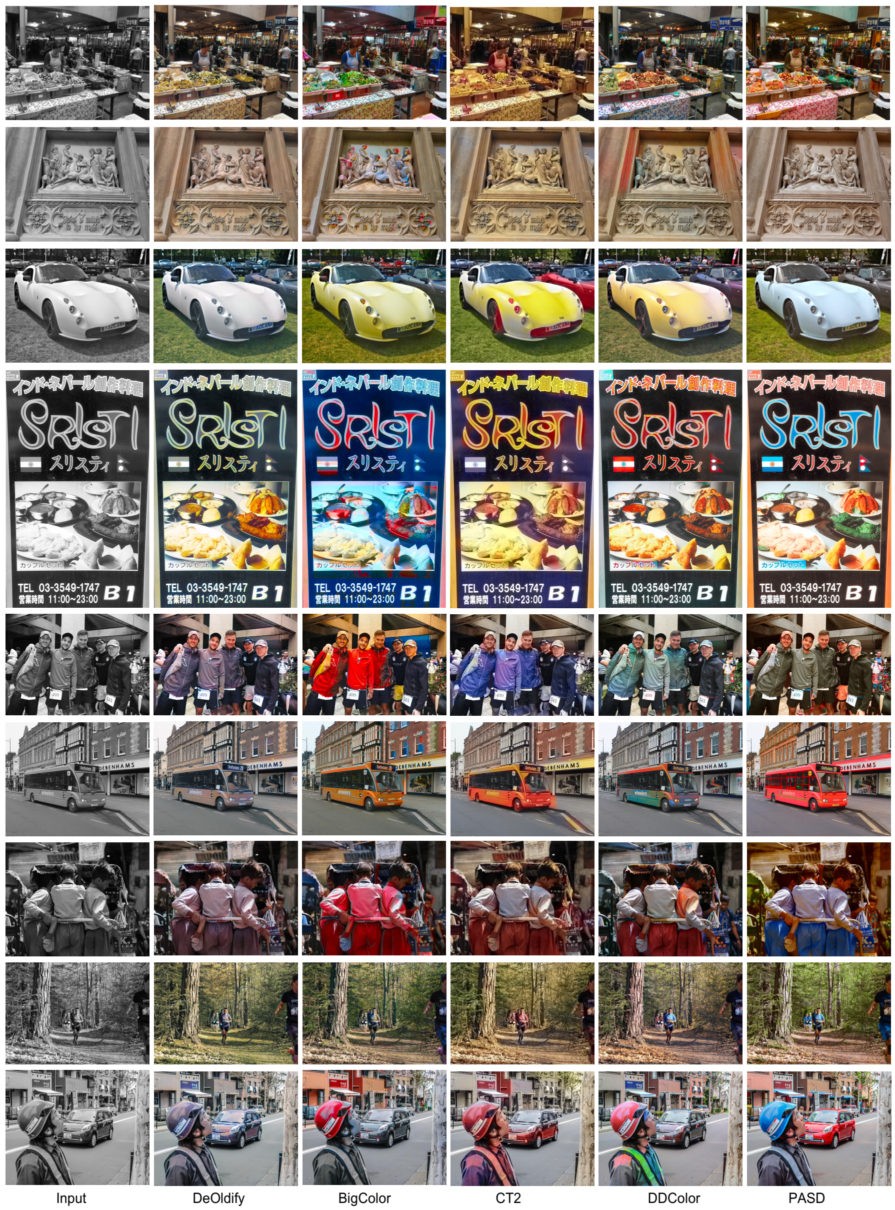

Fig. 4: Qualitative comparison of different colorization methods on real-world images.