Supplementary Materials for Semantic-Sparse Colorization Network for Deep Exemplar-based Colorization

Yunpeng Bai[1], Chao Dong[2,3], Zenghao Chai[1], Andong Wang[1], Zhengzhuo Xu[1], and Chun Yuan[1,4](✉)

[1] Tsinghua Shenzhen International Graduate School, China
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[3] Shanghai AI Laboratory, China
[4] Peng Cheng National Laboratory, China
{byp20, wad20, xzz20}@mails.tsinghua.edu.cn, chao.dong@siat.ac.cn, zenghaochai@gmail.com, yuanc@sz.tsinghua.edu.cn

1 Implementation Details

In the LDT module, we use bi-linear interpolation to resize features. We trained our model on a single NVIDIA GeForce GTX 1080Ti GPU, and the training took around a week. We started from scratch to train all modules of the whole model, including VGG. During training, 20% of the images will be violently deformed to simulate semantically unrelated image pairs.

2 Experimental Results

2.1 More Visual Results

We show more visual results in Figure 1 and compare them with [1] method. We also extend our method to exemplar-based video colorization and show some results in Figure 2. As shown in Figure 4, the proposed SSCN also generates color bleeding artifacts sometimes. Color-bleeding is a common artifact at the edge of low contrast images in previous methods. Although SSCN also has difficulties in building accurate correspondence on these boundaries, it still shows fewer artifacts than other methods. We plan to introduce some edge detection methods to solve this problem in the future.

2.2 Runtime Discussion

As we have discussed the time complexity, we then present runtime comparison on inference process. By using the sparse attention, the time to process a 256×256 resolution image is reduced from 0.38s to 0.25s compared to the dense attention, and the time of 512×512 resolution is reduced from 0.87s to 0.42s.

✉ Corresponding author

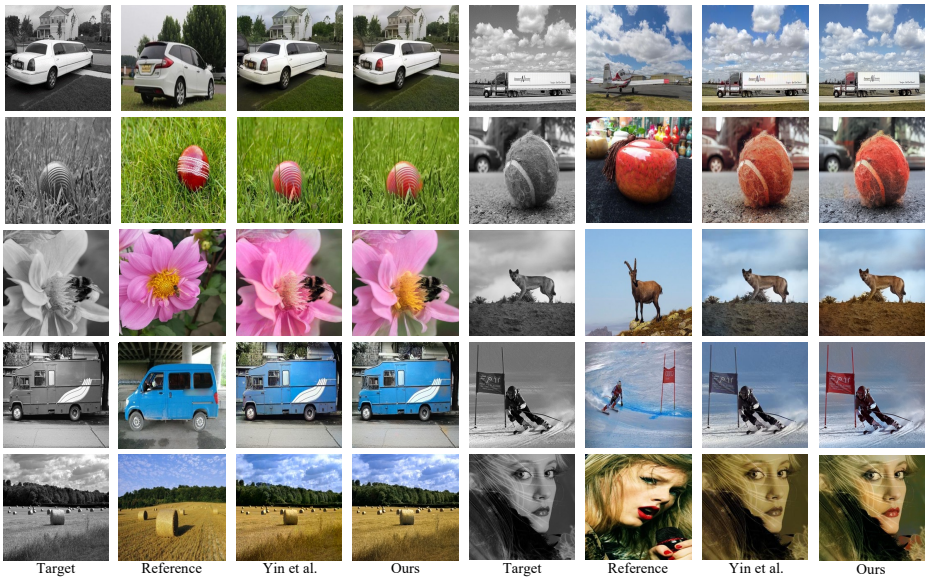

Fig. 1. The figure shows more results of our model. We compare them with the results of [1] method.

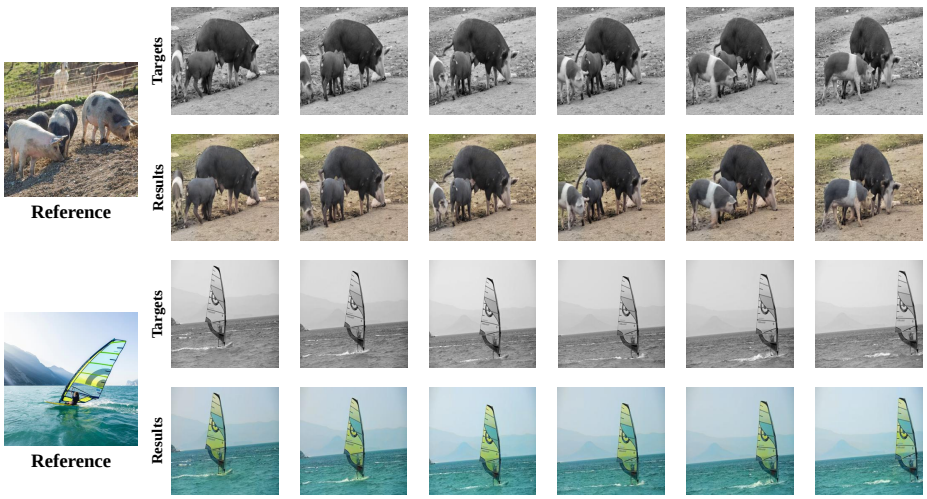

Fig. 2. Extending our method to video colorization. All black and white frames are independently colorized with the same reference to generate colorized results. Our method can also achieve satisfactory results in video colorization.

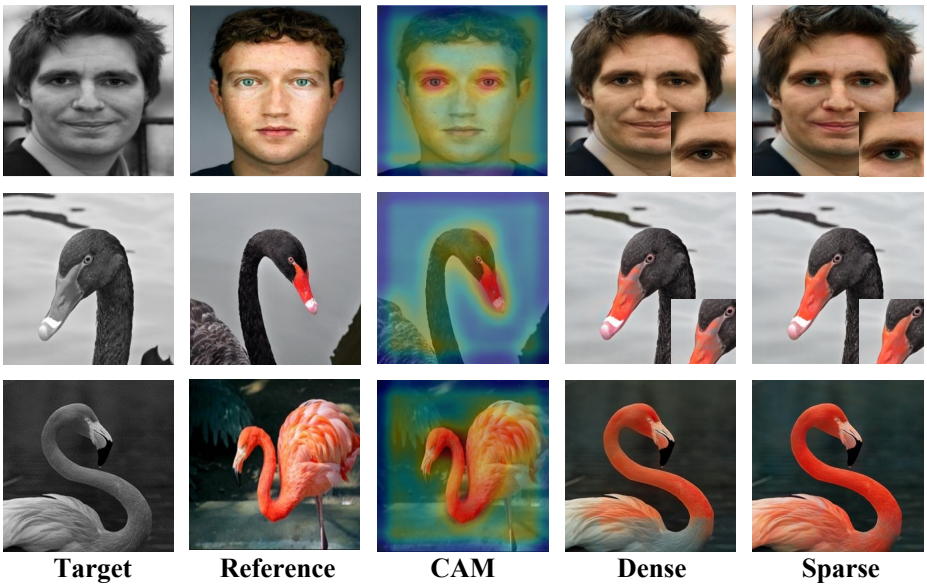

Fig. 3. More comparison results of dense and sparse correspondence strategies.

Fig. 4. Some cases the model prone to failure. Our method cannot color objects satisfactorily if their luminance gaps are too large, and there are also some color-breeding artifacts.

Fig. 5. Some examples of using self-augmented images as references to colorize target images.

References

1. Yin, W., Lu, P., Zhao, Z., Peng, X.: Yes, "attention is all you need", for exemplar based colorization. In: MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 - 24, 2021. pp. 2243–2251. ACM (2021) 1, 2