Supplementary Material for "Unbiased Multi-Modality Guidance for Image Inpainting"

Yongsheng Yu[1,3], Dawei Du[2], Libo Zhang[1,3,4], and Tiejian Luo[3]

[1] Institute of Software Chinese Academy of Sciences, China
[2] Kitware, USA
[3] University of Chinese Academy of Sciences, China
[4] Nanjing Institute of Software Technology, China
yuyongsheng19@mails.ucas.ac.cn; cvdaviddo@gmail.com; libo@iscas.ac.cn; tjluo@ucas.ac.cn

1 Implementation Details

In this work, the layers of adaptive contexual bottlenecks in the encoder are set to $L = 8$, where the dilation rates of convolutions in each bottleneck are empirically set to $r \in R = \{1, 2, 3, 4\}$. In multi-scale spatial-aware attention, the patch size is set to $h = w = \{2, 4\}$. We use $\lambda_{\text{edge}} = 0.1, \lambda_{\text{seg}} = 0.5$ for the overall loss in Eq. 3. The weight of edge loss in Eq. 4 is empirically set to $w_1 = 5.0$.

We train our model with batch size of 20 using the Adam optimizer. We first use an initial learning rate of $2e-4$ and $1e-5$ and then $5e-5$ and $1e-6$ to train the main network and edge discriminator respectively. Following [6, 5], we scale the image size of all datasets to 256×256 as the input.

2 Visual Results

In this section, we compare our method with state-of-the arts on four large-scale datasets, i.e., CelebA-HQ dataset [3, 4], Outdoor dataset (OST) [8], and Cityscapes dataset [1].

2.1 Qualitative Facial Inpainted Results

Compared with the baselines (GC [9], CMGAN [10], ICT [7] and CTSDG [2]), the inpainted results of faces with various races are illustrated in Figure 1. In the first case (the black male with a square mask), GC [9] can only emerge blurred face outlines, and recent methods such as CMGAN [10], ICT [7] are still hard to generate complete face, while CTSDG [2] results in ripples that affect fidelity. In the second case (the white female with free-form masks), all baselines cannot repair the right eyebrow or lips, which our method restores symmetrically and reasonably. Similarly, only our method produces a harmonious eye shape in the third case (the Asian female with free-form masks). In addition, as shown in Figure 2, more facial cases demonstrate that our method can reconstruct images with semantically consistent and reasonable patches.

2.2 Qualitative Scene Inpainted Results

We provide some scene inpainted results in Figure 3 and 4 respectively. It indicates that our method can achieve state-of-the-art performance in various complex scenarios.

Fig. 1. Qualitative comparison results on the CelebA-HQ dataset.

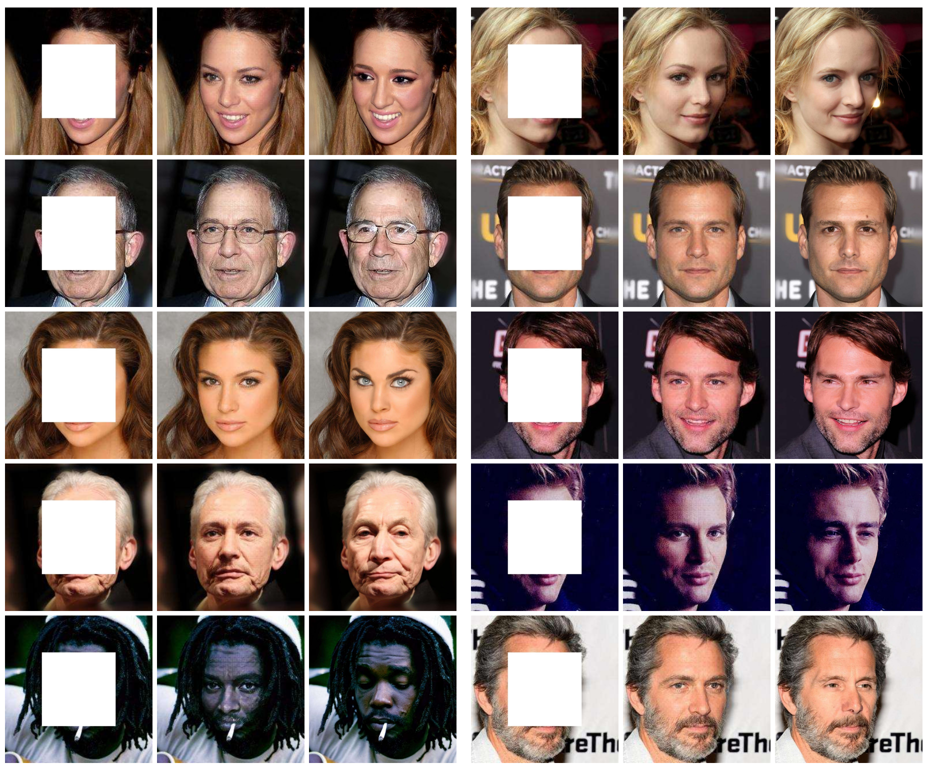

Fig. 2. Visual results on the CelebA-HQ dataset. From left to right are masked image, ours, and Ground Truth.

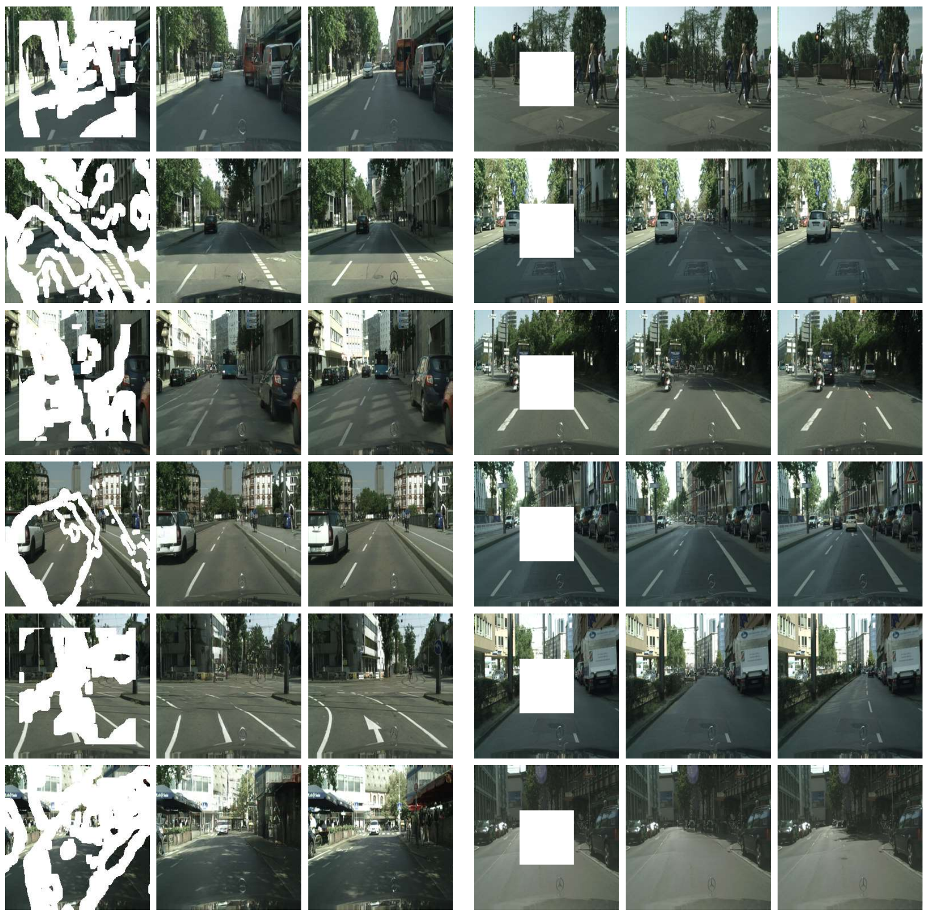

Fig. 3. Visual results on CityScape dataset. From left to right are masked image, ours, and Ground Truth.

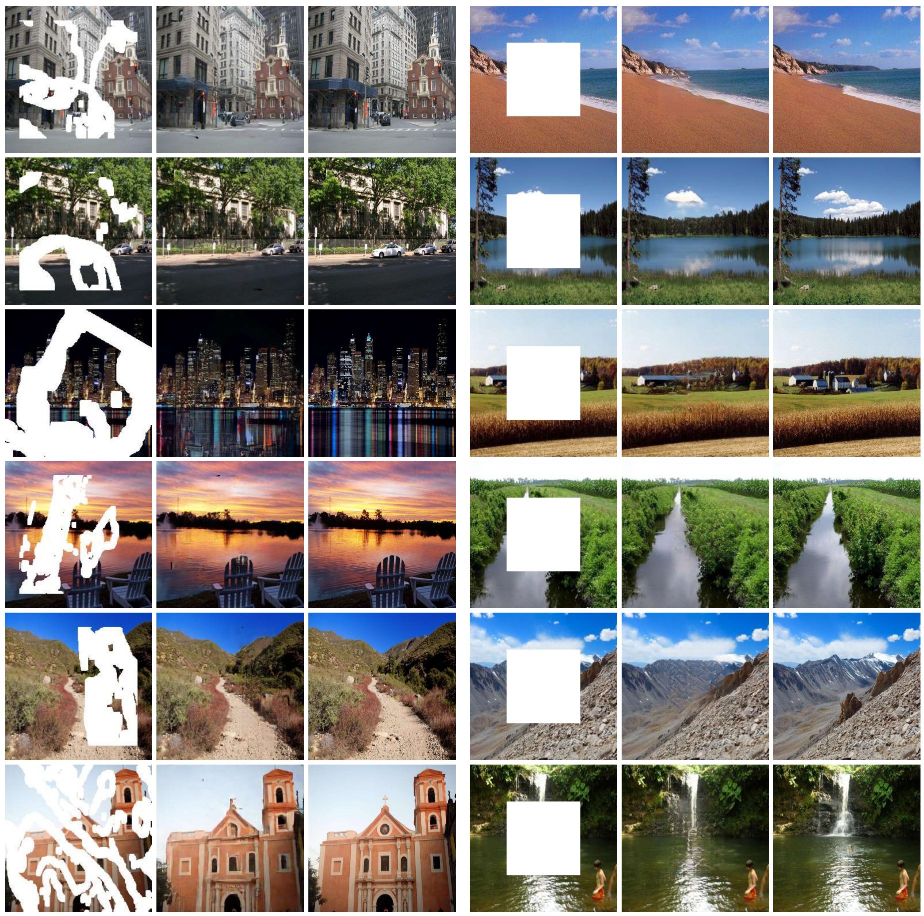

Fig. 4. Visual results on OST dataset. From left to right are masked image, ours, and Ground Truth.

References

1. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: CVPR. pp. 3213–3223 (2016)
2. Guo, X., Yang, H., Huang, D.: Image inpainting via conditional texture and structure dual generation. In: ICCV. pp. 14114–14123 (2021)
3. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. In: ICLR (2018)
4. Lee, C., Liu, Z., Wu, L., Luo, P.: Maskgan: Towards diverse and interactive facial image manipulation. In: CVPR. pp. 5548–5557 (2020)
5. Li, J., Wang, N., Zhang, L., Du, B., Tao, D.: Recurrent feature reasoning for image inpainting. In: CVPR. pp. 7757–7765 (2020)
6. Liao, L., Xiao, J., Wang, Z., Lin, C., Satoh, S.: Guidance and evaluation: Semantic-aware image inpainting for mixed scenes. In: ECCV. pp. 683–700 (2020)
7. Wan, Z., Zhang, J., Chen, D., Liao, J.: High-fidelity pluralistic image completion with transformers. In: ICCV. pp. 4672–4681 (2021)
8. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: CVPR. pp. 606–615 (2018)
9. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Free-form image inpainting with gated convolution. In: ICCV. pp. 4470–4479 (2019)
10. Zhao, S., Cui, J., Sheng, Y., Dong, Y., Liang, X., Chang, E.I., Xu, Y.: Large scale image completion via co-modulated generative adversarial networks. In: ICLR (2021)