Rethinking Image Inpainting via a Mutual Encoder-Decoder with Feature Equalizations

Supplementary Materials

Paper ID 4179

In this supplementary material, we first introduce the training details of our model first. Meanwhile, we present the supplementary experiments to further validate the effectiveness of our method. Moreover, we process object removal on real-world images to verify the generalization ability of our model.

1 Training process of our model

For training, given a raw image I_{gt}, a binary image mask M (value 0 for known pixels and 1 denotes unknown ones) at a random position (irregular masks or centering mask). In this way, the input image I_{in} is obtained from the raw image as $I_{in} = I_{gt} \odot (1 - M)$. Our generator G takes $[I_{in}, M]$ as input, and produces prediction I_{out}. All input and output are linearly scaled to [-1,1]. Training procedure is shown in Algorithm 1.

Algorithm 1 Training of our proposed framework

Input: Training data CelebA or Places2 or Paris StreetView for image. Mask data Partial Conv. Structure data;

```
Sample batch images I_gt from training data;
Sample mask M for I_gt from Mask data ;
Sample structure label I_st from Structure data;
for T =1,2,3,...,max_epoch do
    for each mini_batch do
        Construct inputs I_in = I_gt ⊙ (1 − M);
        Get predictions I_out = G([I_in, M]);Get output of two branches I_ost and I_ote;
        Calculate the discriminators losses;
        Update local and global discriminators ;
        Calculate the Total Losses;
        Update Mutual Encoder-Decoder G;
    end for
    Update learning rate.
end for
```

2 More Comparisons on Paris StreetView, Places2 and CelebA

More comparisons with GC, SF, CSA and Ours on Paris StreetView and Places2 datasets with irregular masks. More comparisons with CE, CA, SH and Ours on the Celeba dataset with centering masks. Please refer to Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6 for more results on Paris StreetView and Places2. Please refer to Fig. 7, Fig. 8, Fig. 9 for more results on CelebA. It is obvious that our model outperforms state-of-the-art approaches in both structural consistency and detail richness. As a side contribution, we will release the pre-trained model and codes.

3 More Results on Paris StreetView, Places2 and CelebA

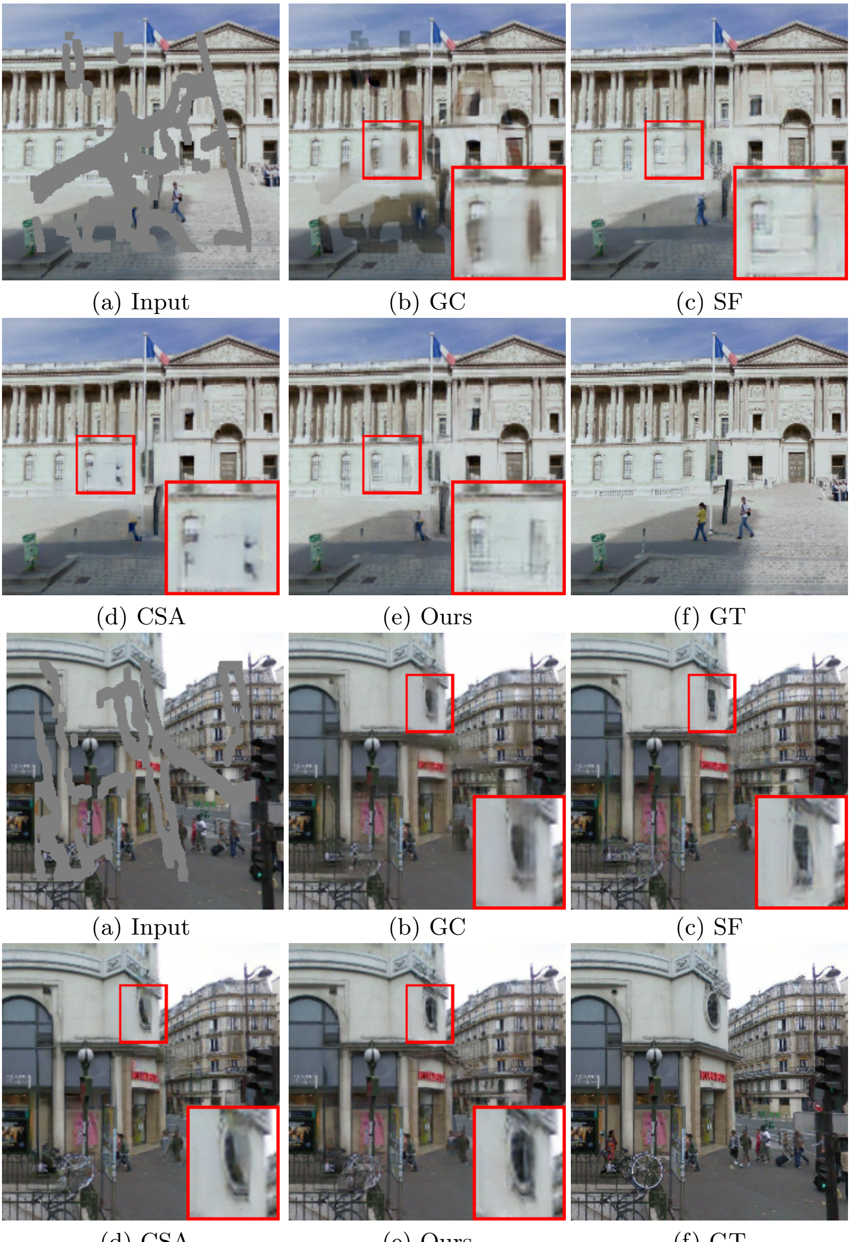
CelebA Fig. 10 and Fig. 11 show more results the obtained by our full model with centering masks, where the model is trained on CelebA dataset. We resize image to 256×256 for both training and evaluation.

Places2 Fig. 12 and Fig. 13 show more results obtained by our full model with irregular masks, where the model is trained on the Places2 dataset. We also resize the images to 256×256 for both training and evaluation.

Paris StreetView We also perform experiments on our full model trained on the Paris StreetView dataset with irregular masks, and the results are shown in Fig. 14 and Fig. 15. We resize image to 256×256 for both training and evaluation.

4 More object removal on real-world images

We apply our model trained on Paris StreetView or Places2 to process object removal on real-world images, as shown in Fig. 16, Fig. 17, Fig. 18, Fig. 19, Fig. 20 and Fig. 21 for results. These real-world images are complex for large area of distractors and complicated background. Even so, our model can handle them well, which indicates the effectiveness, applicability, and generality of our model.



(a) Input (b) GC (c) SF

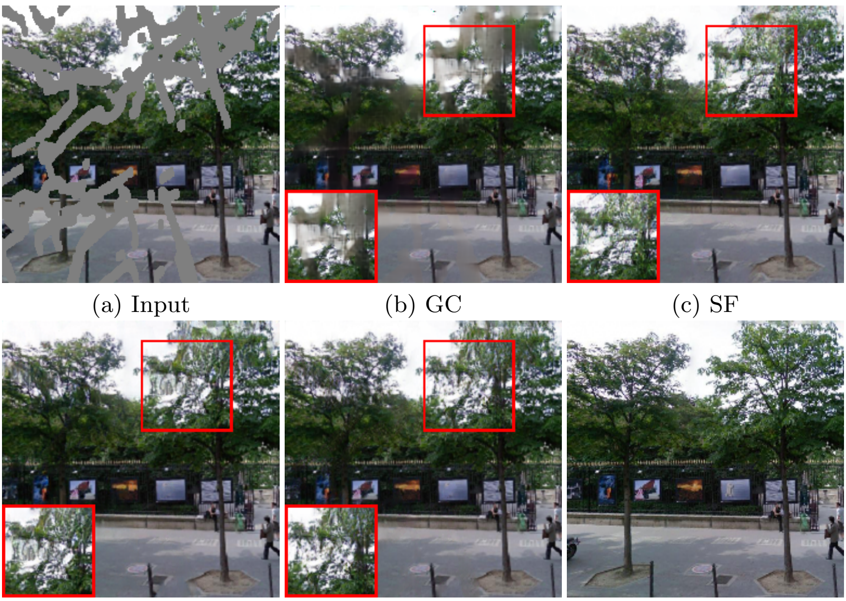
(d) CSA (e) Ours (f) GT

(a) Input (b) GC (c) SF

(d) CSA (e) Ours (f) GT

Fig. 1. Visual evaluations for filling irregular holes on the Paris StreetView dataset. Our method performs favorably against existing approaches to retain both structures and textures.

Fig. 2. Visual evaluations for filling irregular holes on the Paris StreetView dataset. Our method performs favorably against existing approaches to retain both structures and textures.



(a) Input (b) GC (c) SF

(d) CSA (e) Ours (f) GT

Fig. 3. Visual evaluations for filling irregular holes on the Paris StreetView dataset. Our method performs favorably against existing approaches to retain both structures and textures.

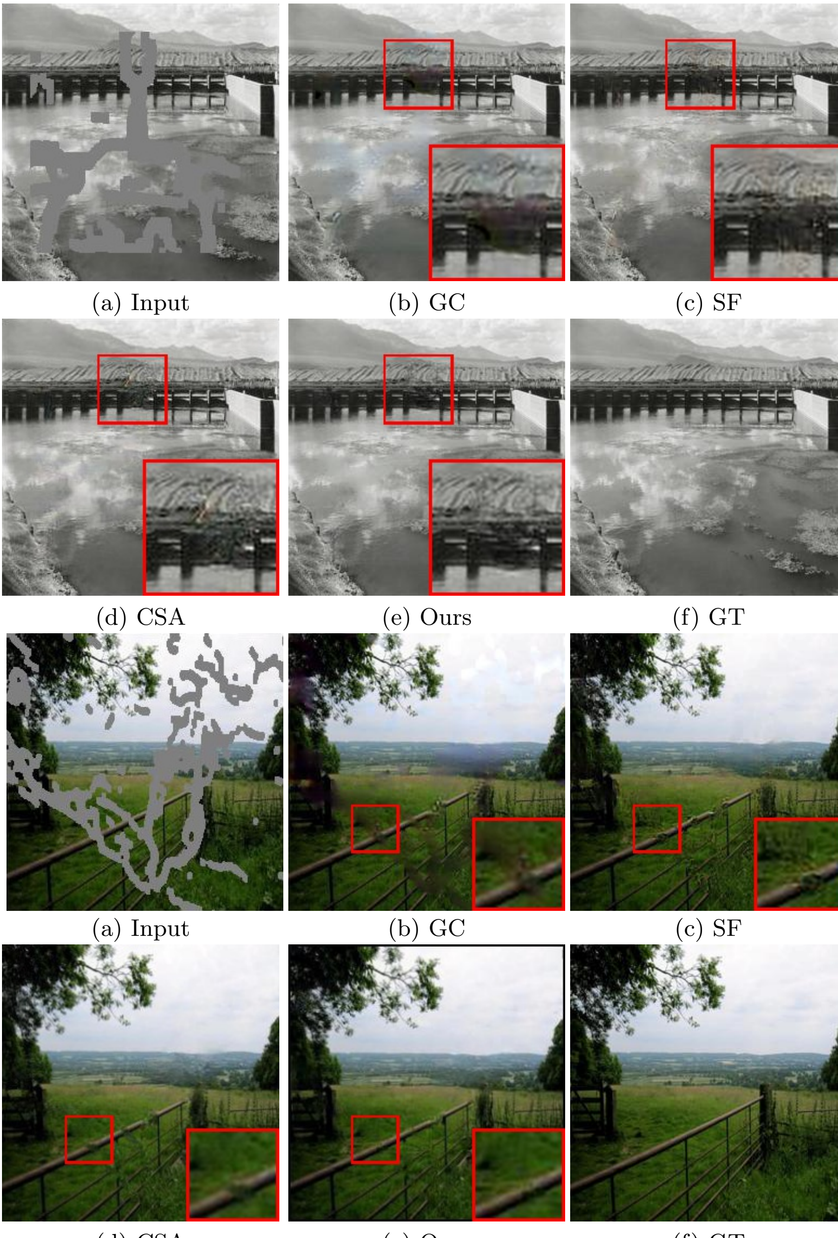
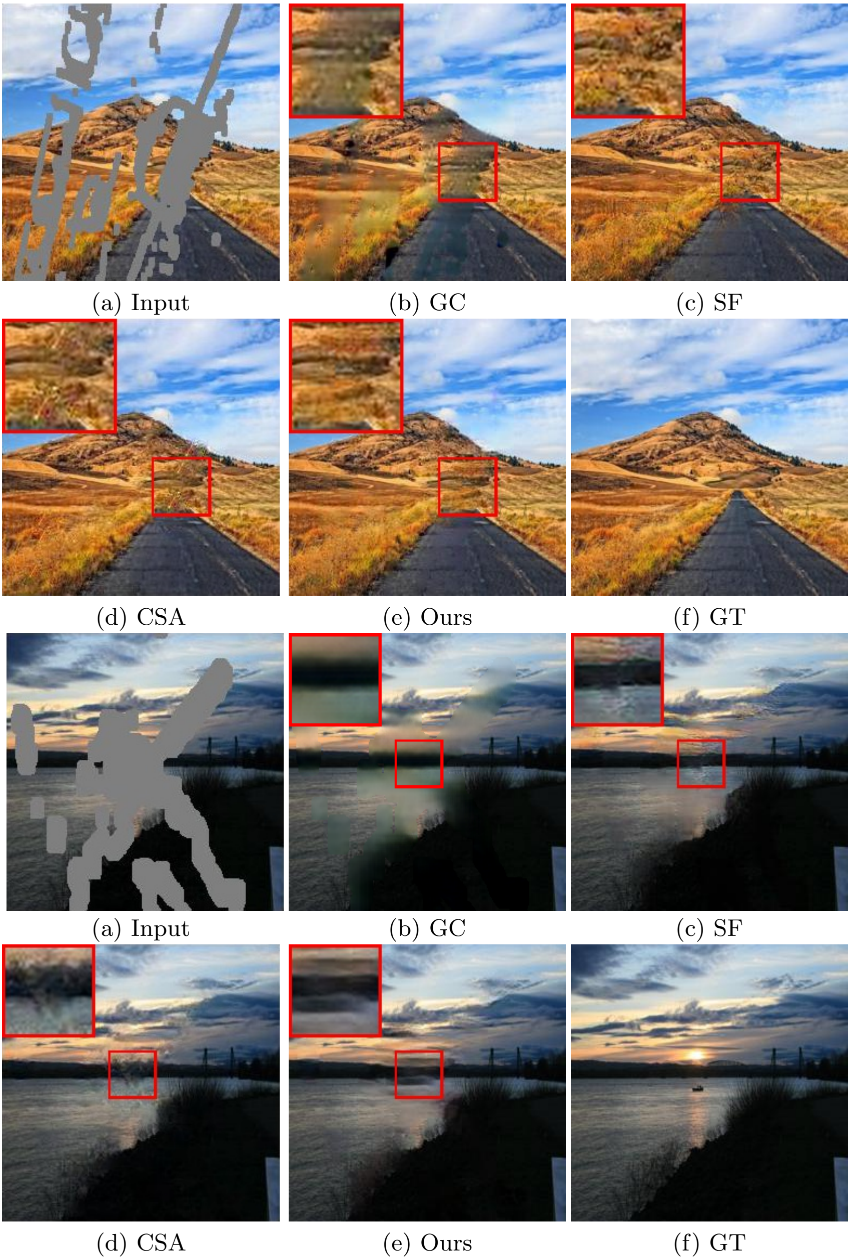


Fig. 4. Visual evaluations for filling irregular holes on the Places2 dataset. Our method performs favorably against existing approaches to retain both structures and textures.



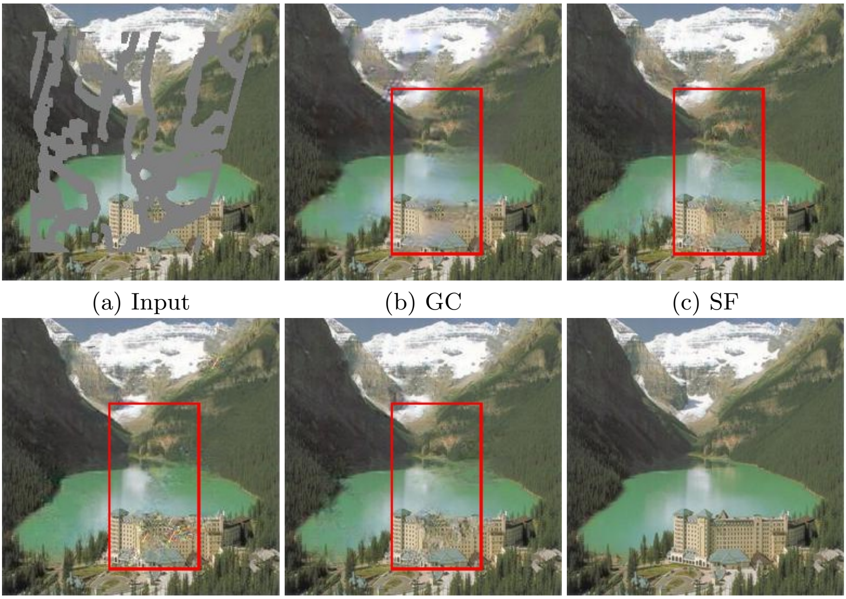
(a) Input (b) GC (c) SF

(d) CSA (e) Ours (f) GT

(a) Input (b) GC (c) SF

(d) CSA (e) Ours (f) GT

Fig. 5. Visual evaluations for filling irregular holes on the Places2 dataset. Our method performs favorably against existing approaches to retain both structures and textures.



(a) Input (b) GC (c) SF

(d) CSA (e) Ours (f) GT

Fig. 6. Visual evaluations for filling irregular holes on the Places2 dataset. Our method performs favorably against existing approaches to retain both structures and textures.

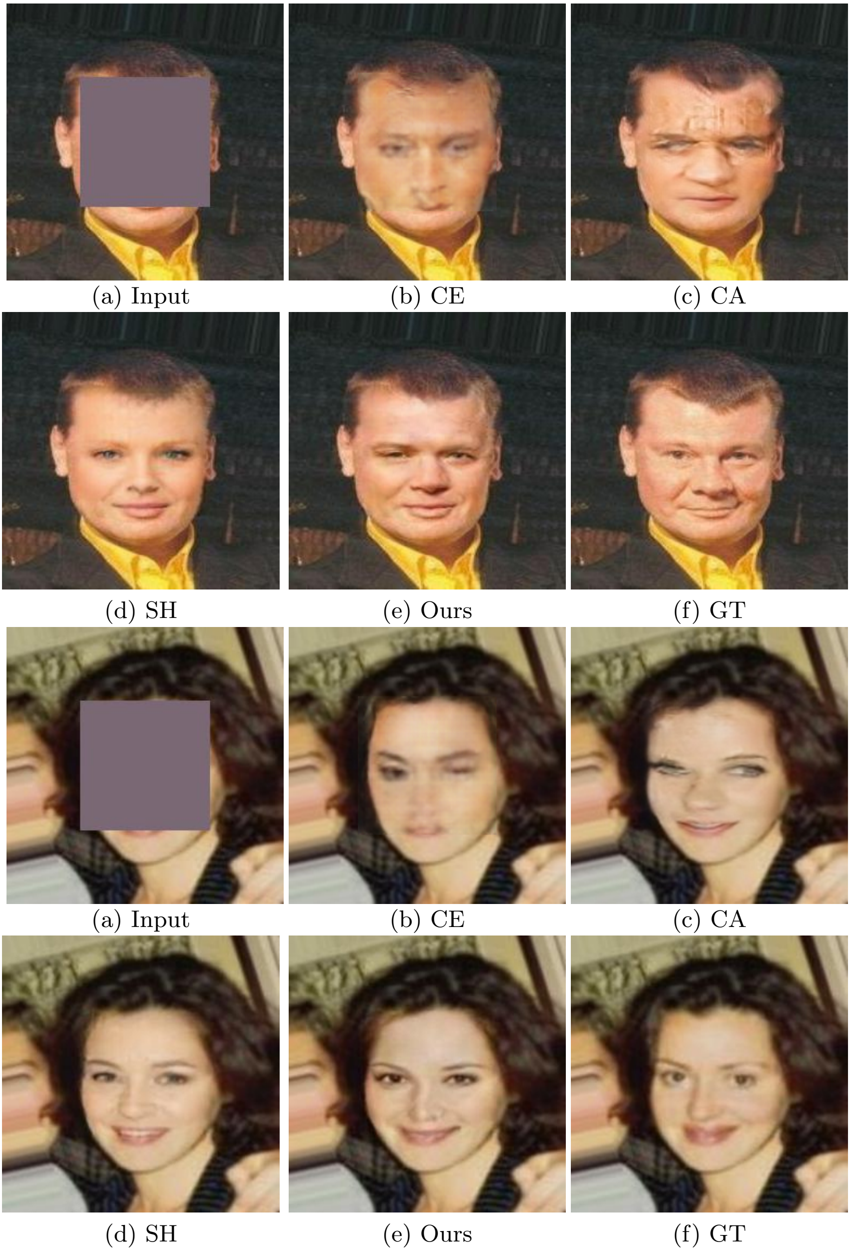
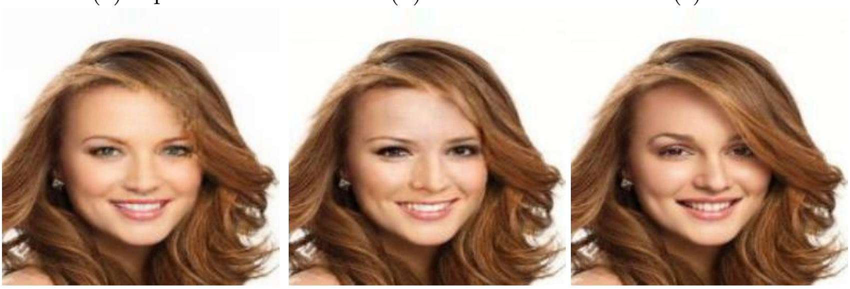


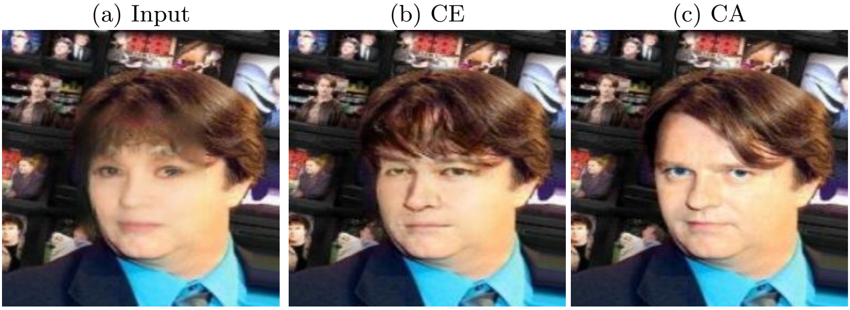
Fig. 7. Visual evaluations for filling centering holes on the CelebA dataset. Our method performs favorably against existing approaches to retain both structures and textures.

(a) Input (b) CE (c) CA



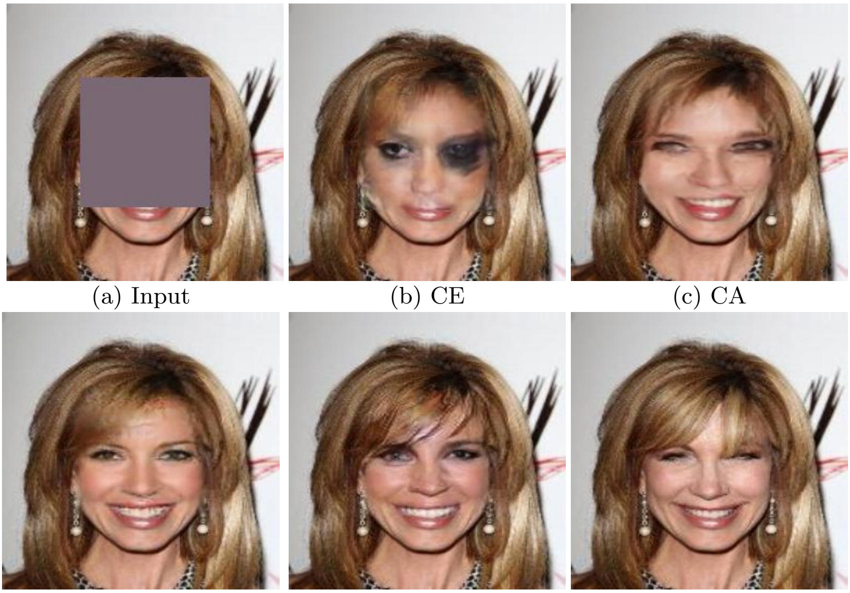
(d) SH (e) Ours (f) GT

(a) Input (b) CE (c) CA



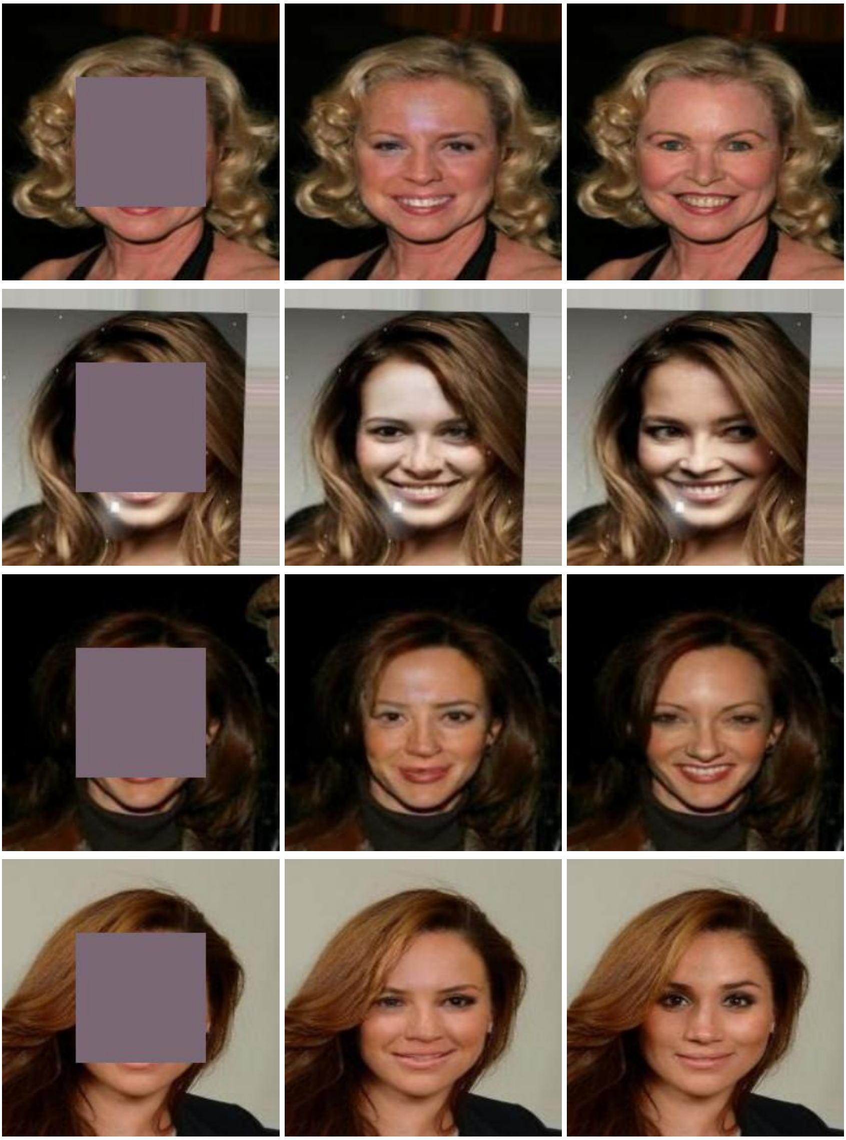
(d) SH (e) Ours (f) GT

Fig. 8. Visual evaluations for filling centering holes on the CelebA dataset. Our method performs favorably against existing approaches to retain both structures and textures.



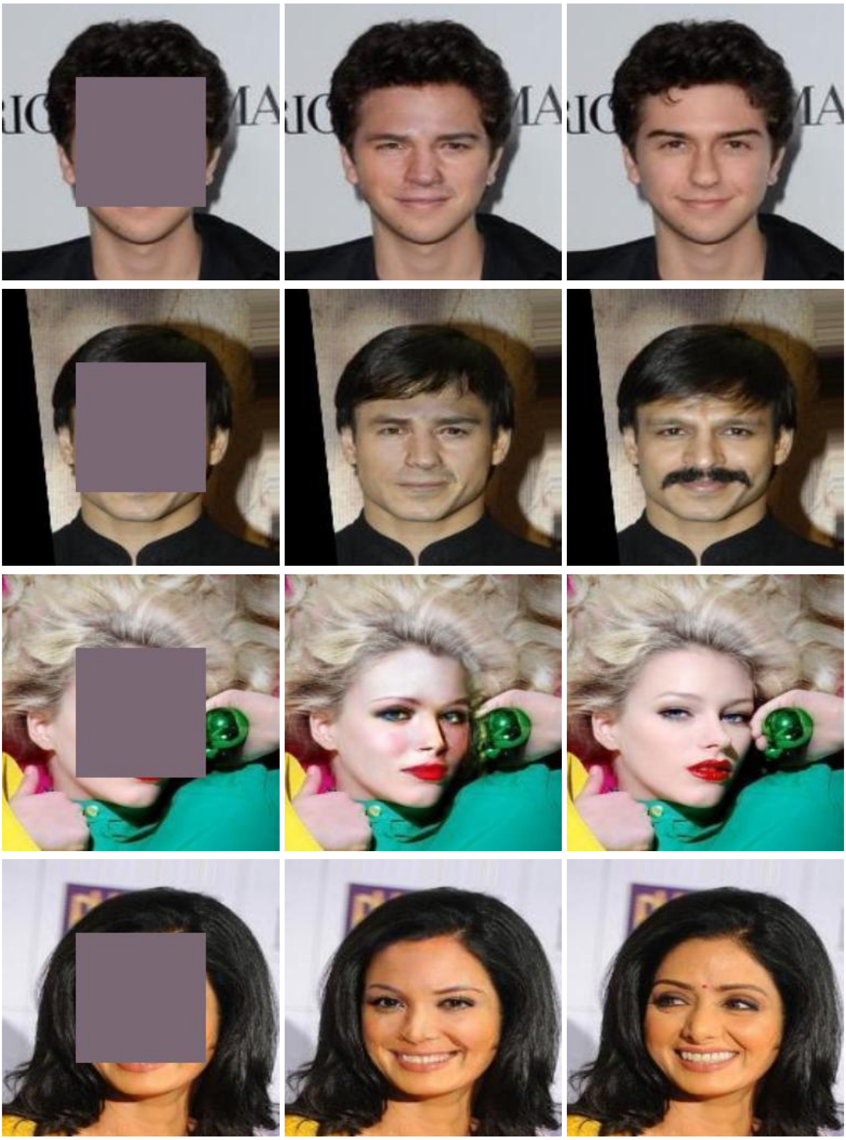
(a) Input (b) CE (c) CA (d) SH (e) Ours (f) GT

Fig. 9. Visual evaluations for filling centering holes on the CelebA dataset. Our method performs favorably against existing approaches to retain both structures and textures.



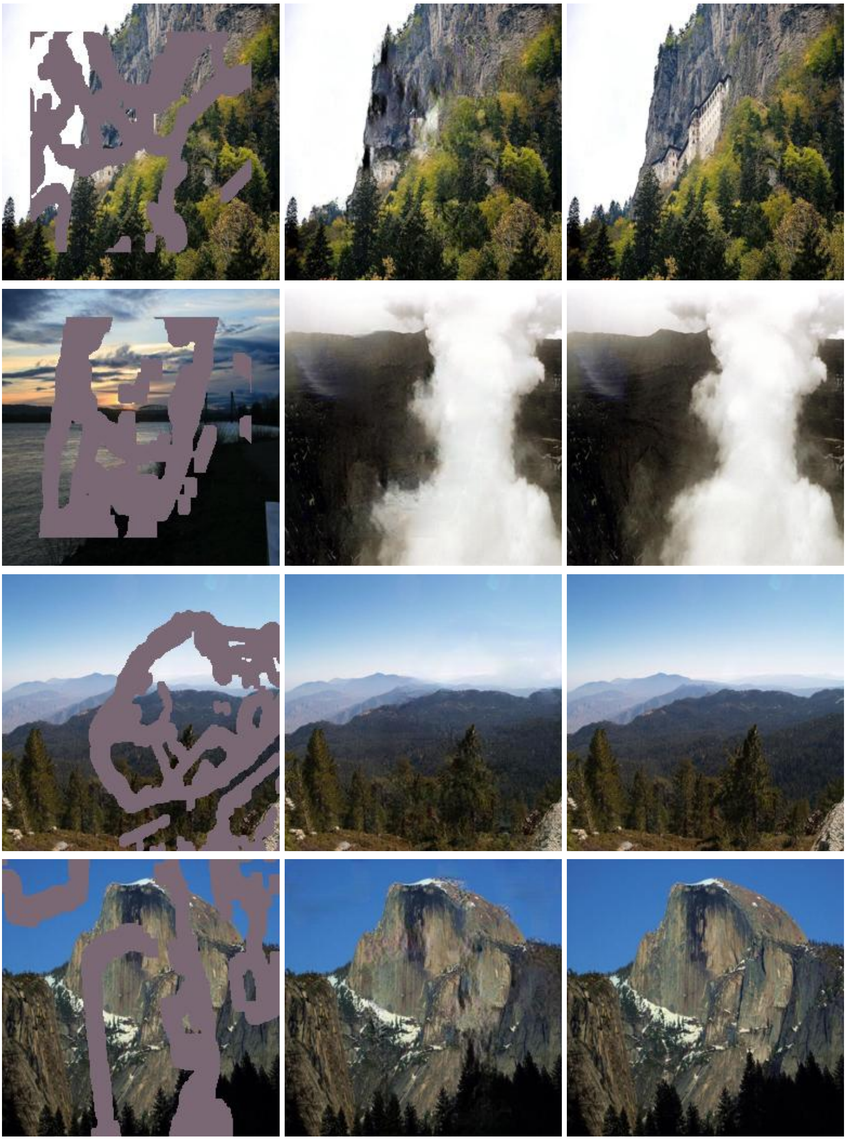
(a) Input (b) Ours (c) GT

Fig. 10. More results on the CelebA dataset.



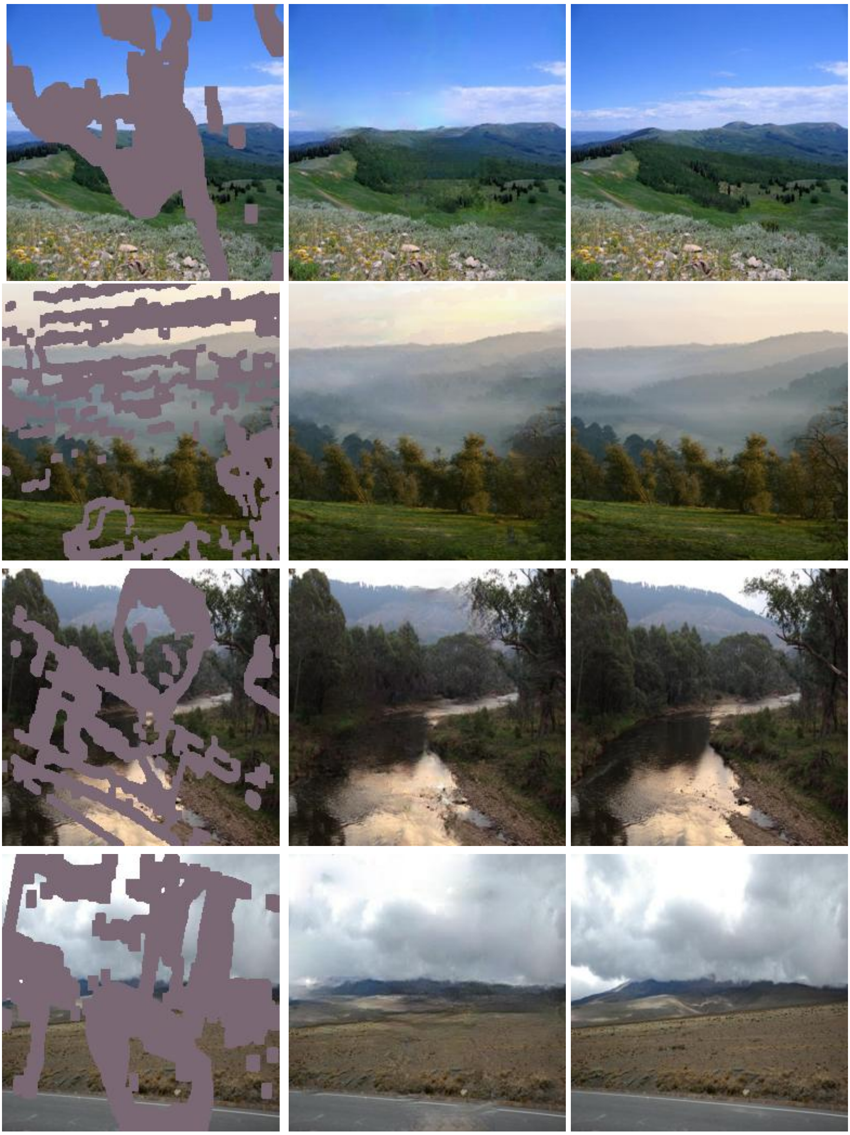
(a) Input (b) Ours (c) GT

Fig. 11. More results on the CelebA dataset.



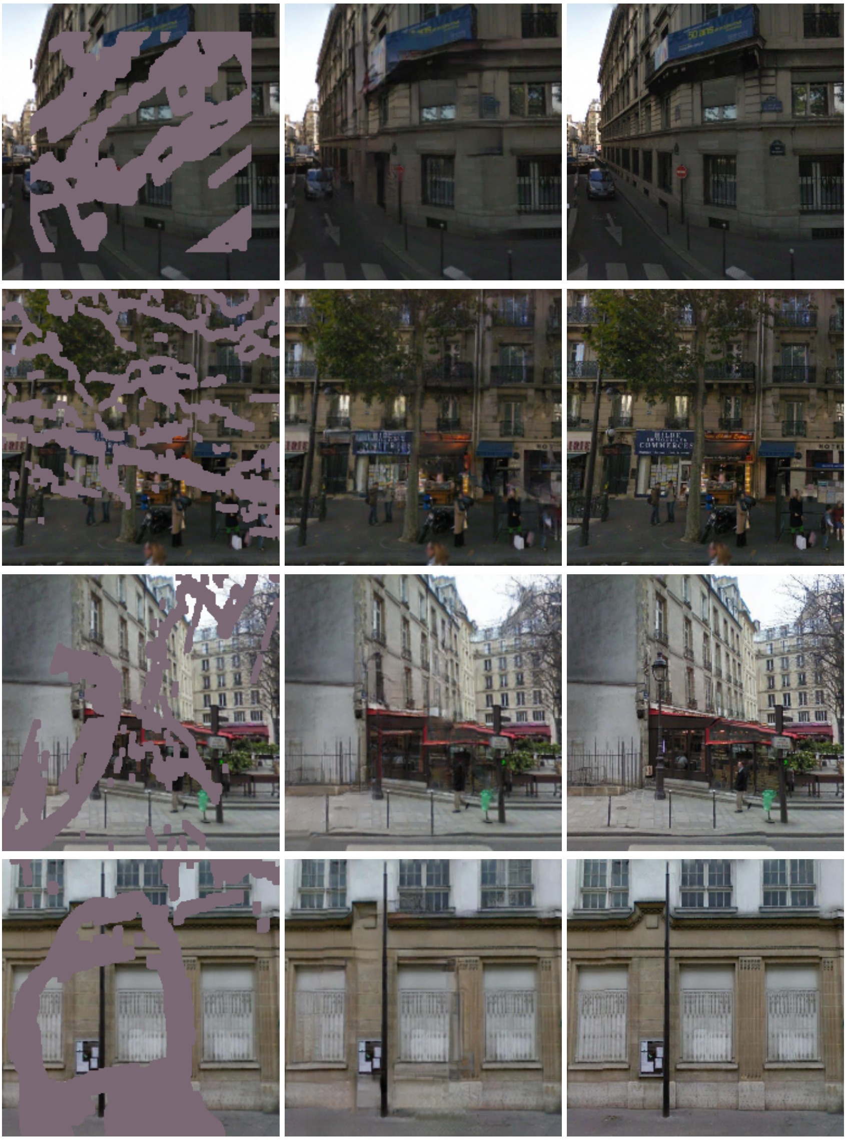
(a) Input (b) Ours (c) GT

Fig. 12. More results on the Places2 dataset.



(a) Input (b) Ours (c) GT

Fig. 13. More results on the Place2 dataset.

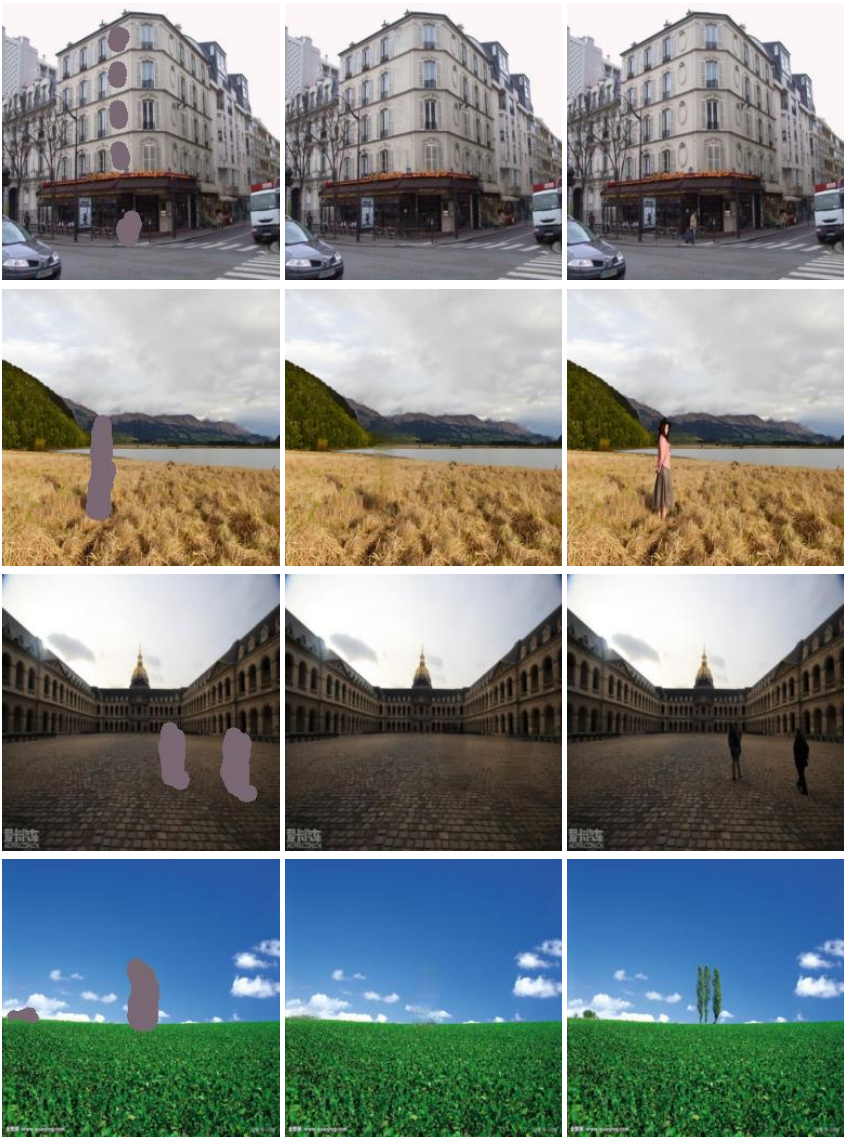


(a) Input (b) Ours (c) GT

Fig. 14. More results on the Paris StreetView dataset.

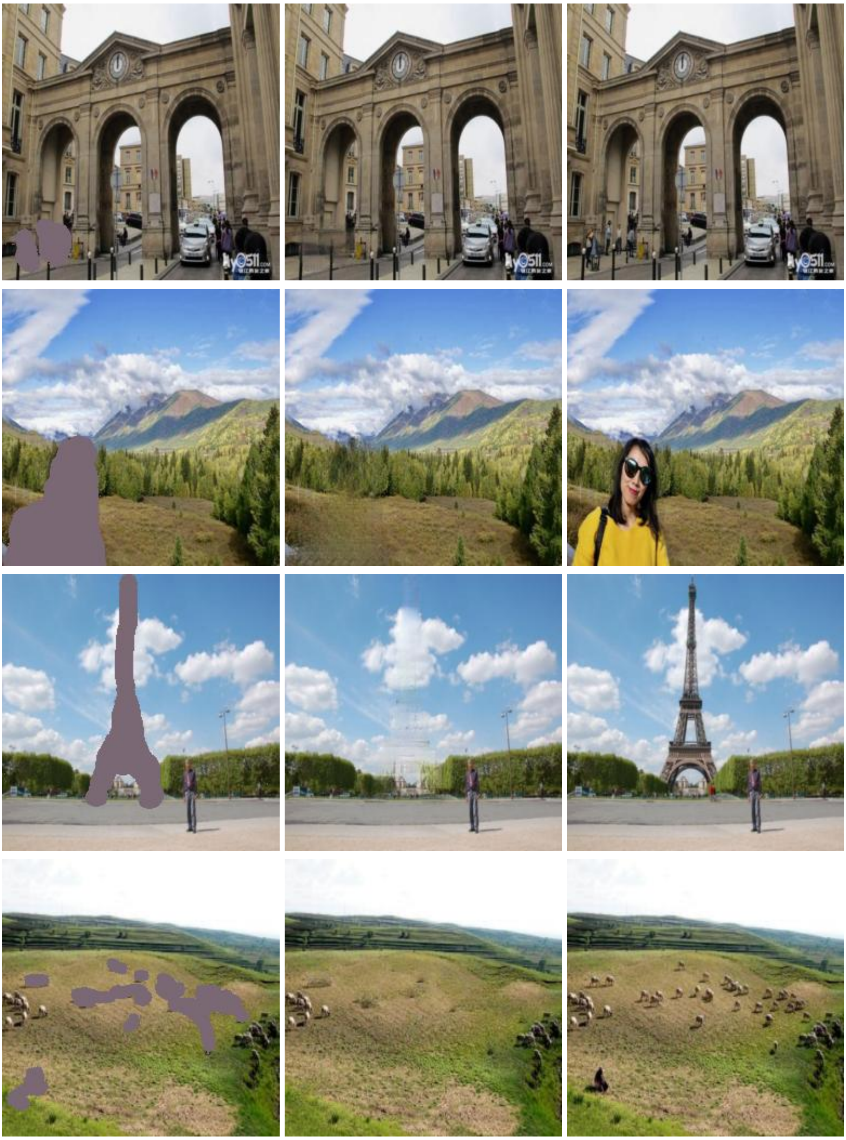
(a) Input (b) Ours (c) GT

Fig. 15. More results on the Paris StreetView dataset.



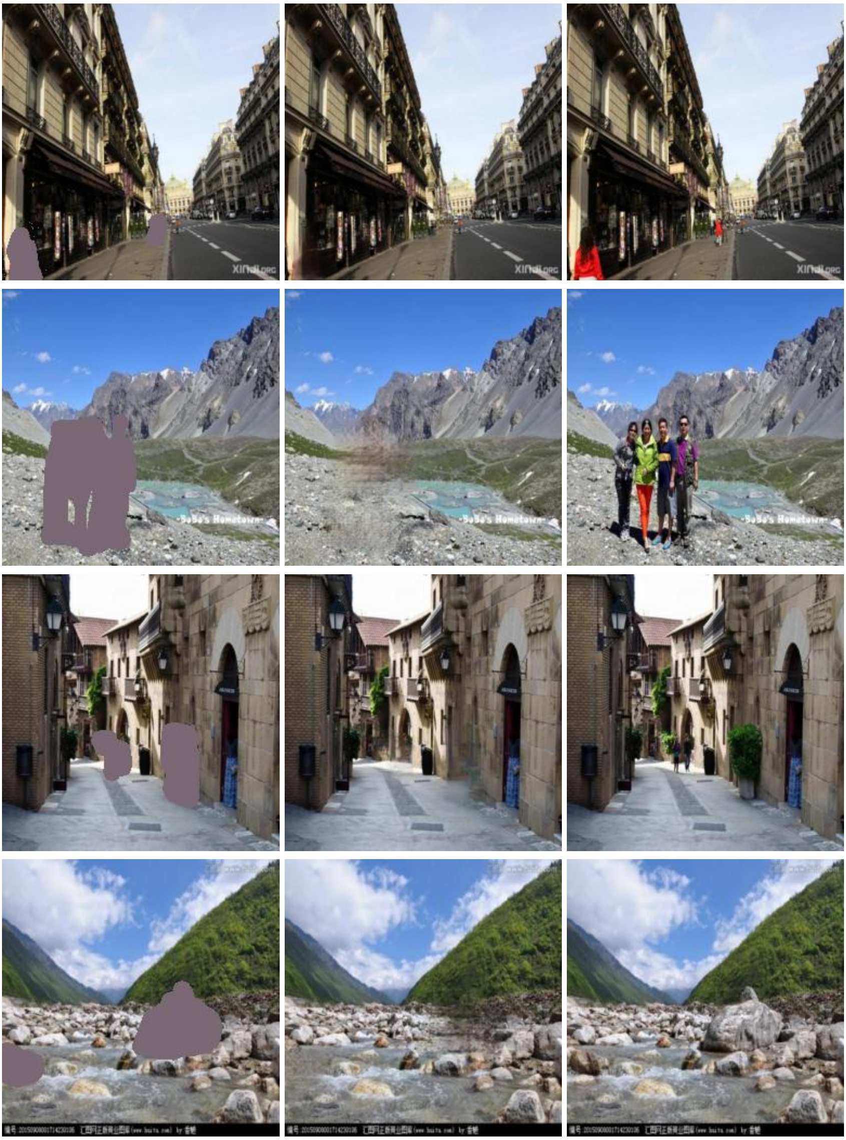
(a) Input (b) Ours (c) Real image

Fig. 16. Object removal on real images.



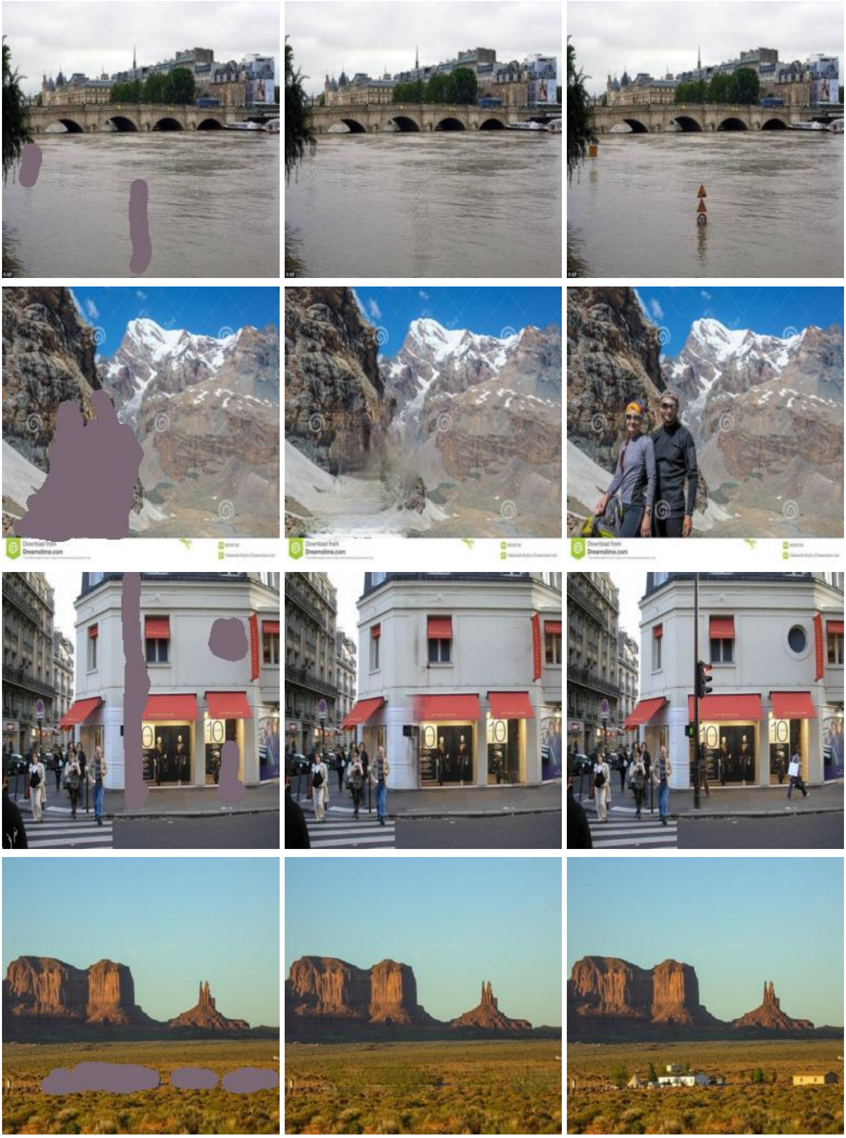
(a) Input (b) Ours (c) Real image

Fig. 17. Object removal on real images.



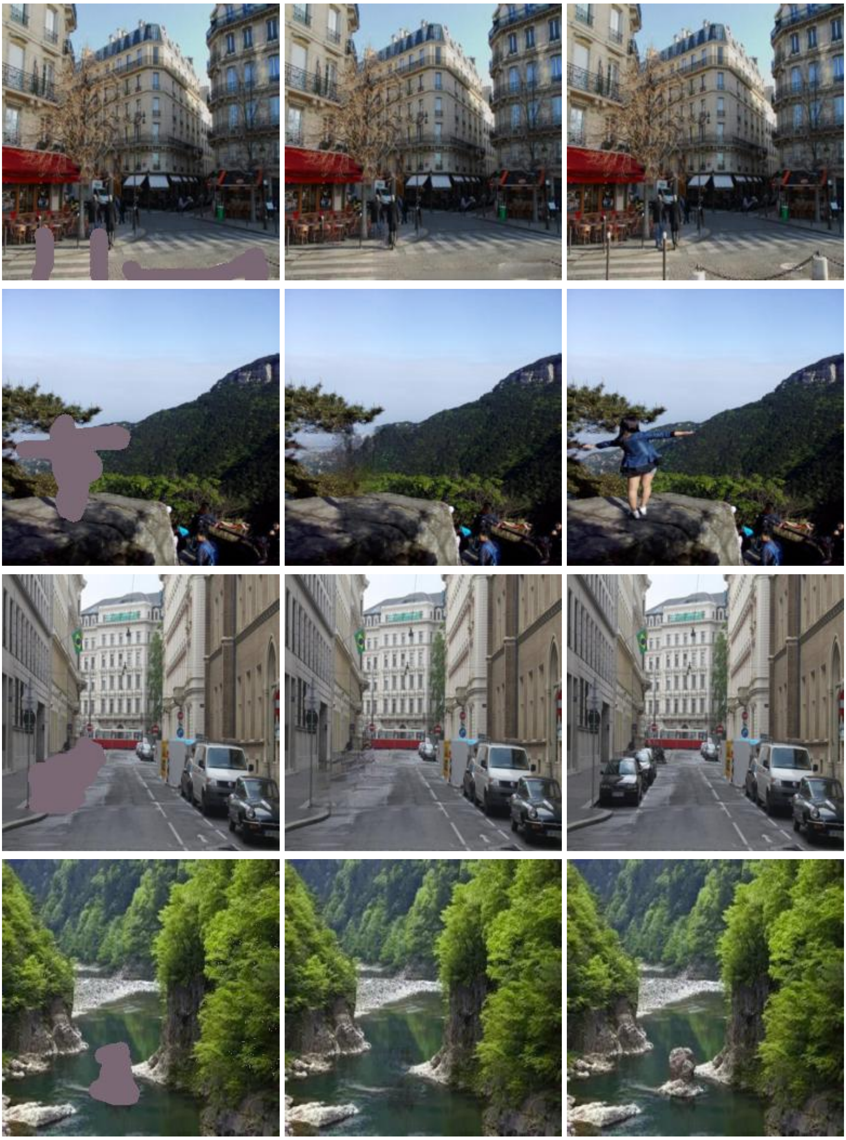
(a) Input (b) Ours (c) Real image

Fig. 18. Object removal on real images.



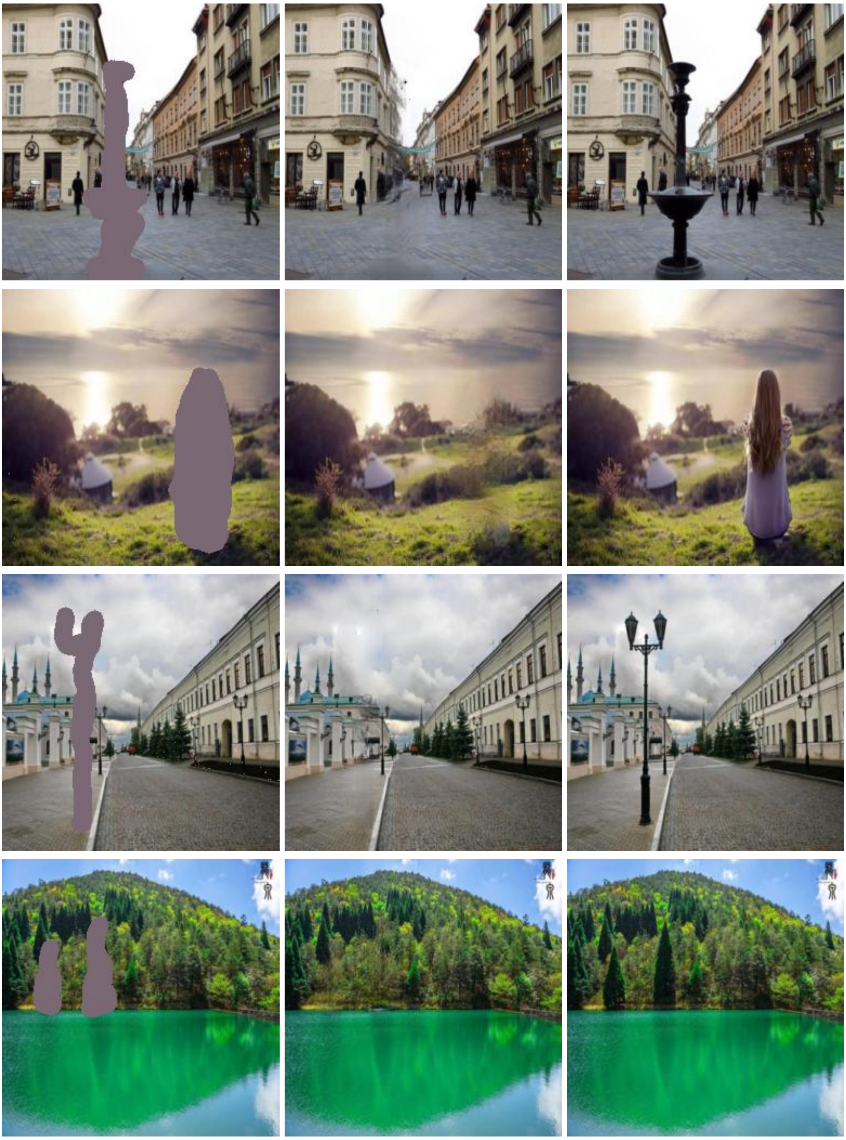
(a) Input (b) Ours (c) Real image

Fig. 19. Object removal on real images.



(a) Input (b) Ours (c) Real image

Fig. 20. Object removal on real images.



(a) Input (b) Ours (c) Real image

Fig. 21. Object removal on real images.