

– Supplemental Material –

Contextual Residual Aggregation for Ultra High-Resolution Image Inpainting

Zili Yi Qiang Tang Shekoofeh Azizi Daesik Jang Zhan Xu
Huawei Technologies Canada Co. Ltd.

Appendix

Network Architectures

In addition to Section 3.3 and Figure 2 in the main paper, we report more details of our network architectures. For simplicity, we denote them with K (kernel size), S (stride size), C (channel number) and D (dilation rate). D is neglected when $D=1$.

Coarse Network: $\text{downsample}(2\times)$ - K5S2C32 - K3S1C32 - K3S2C64 - K3S1C64 - K3S1C64 - K3S1C64 - K3S1C64 - K3S1C64 - K3S1C64 - K3S1C64 - K3S1C64D2 - K3S1C64D2 - K3S1C64D2 - K3S1C64D2 - K3S1C64D4 - K3S1C64D4 - K3S1C64D4 - K3S1C64D4 - K3S1C64D8 - K3S1C64D8 - K3S1C64 - K3S1C64 - K3S1C64 - $\text{upsample}(2\times)$ - K3S1C32 - $\text{upsample}(2\times)$ - K3S1C3 - clip - $\text{upsample}(2\times)$

Refine Network: K5S2C32 - K3S1C32 $[P^l=1]$ - K3S2C64 - K3S1C64 $[P^l=2]$ - K3S2C128 - K3S1C128 - K3S1C128 - K3S1C128D2 - K3S1C128D4 - K3S1C128D8 - K3S1C128D16 $[P^l=3]$ - concat - K3S1C128 - $\text{upsample}(2\times)$ - K3S1C64 - K3S1C64 - concat - $\text{upsample}(2\times)$ - K3S1C32 - K3S1C32 - concat - $\text{upsample}(2\times)$ - K3S1C3 - clip

Attention Computing Branch: $[P^l=3]$ - $\text{downsample}(2\times)$ - $[P]$ - ACM - ATM

Attention Transfer Branch ($P^l=3$): $[P^l=3]$ - ATM - K3S1C128 - concat

Attention Transfer Branch ($P^l=2$): $[P^l=2]$ - ATM - K3S1C64 - K3S1C64D2 - concat

Attention Transfer Branch ($P^l=1$): $[P^l=1]$ - ATM - K3S1C32 - K3S1C32D2 - concat

Table 1: Sources of some HD images used for test

Figure ID in the main paper	Image Source
Figure 3 top	http://www.sohu.com/a/117062677_189010
Figure 6 top	http://ow.ly/u8Wff
Figure 6 bottom	https://www.mafengwo.cn/yj/14103/s-0-0-0-1-0.html
Figure 1 topright	https://www.champaignoutdoors.com/kilimanjaro
Images in demo.pps	http://www.imecchina.com/news/1293274.html http://www.zdqx.com/wall/57962_6.html https://www.xuehua.us/2018/06/03/%E5%92%8C%E9%AB%98%E5%B0%94%E5%A4%AB%E5%98%89%E6%97%85%E4%B8%80%E9%81%93-%E6%8E%A2%E5%AF%BB%E4%BB%99%E6%B9%96%E8%BE%B9%E7%9A%84%E6%85%A2%E7%94%9F%E6%B4%BB/zh-tw/ https://you.autohome.com.cn/details/68005/727cc0cec7214dd62e92d8f009e7adf9 https://www.reyfoto.com/

Discriminator: K3S2C64 - K3S2C128 - K3S2C256 - K3S2C256 - K3S2C256 - K3S2C256 - fully connected to 1.

More Test Results on Places2

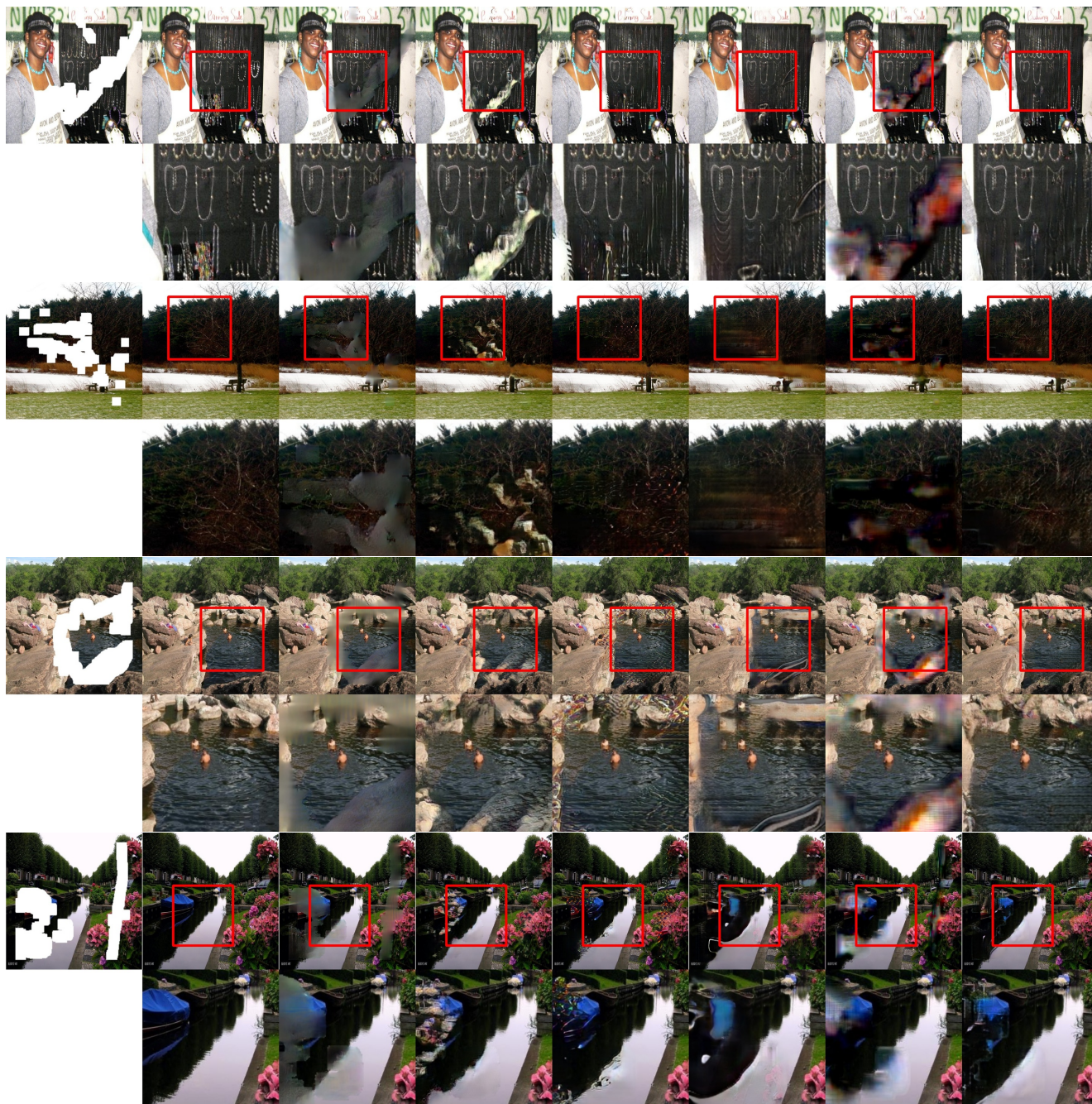
More test results on places2 are presented in Figures 1, 2, 3, with input size 512×512 , 1024×1024 , 2048×2048 respectively.

Sources of High-Resolution Images

Sources of the HD images in the main paper that are crawled from the internet are presented in Table 1.

Failure Examples & Limitation

Some failure examples of our model are presented in Figure 4. Our model is prone to fail when the majority parts of a background object are missing (Referring to the bicycle and dog face in Figure. 4).



(a) Input

(b) GT

(c) global-local

(d) DeepFillV1

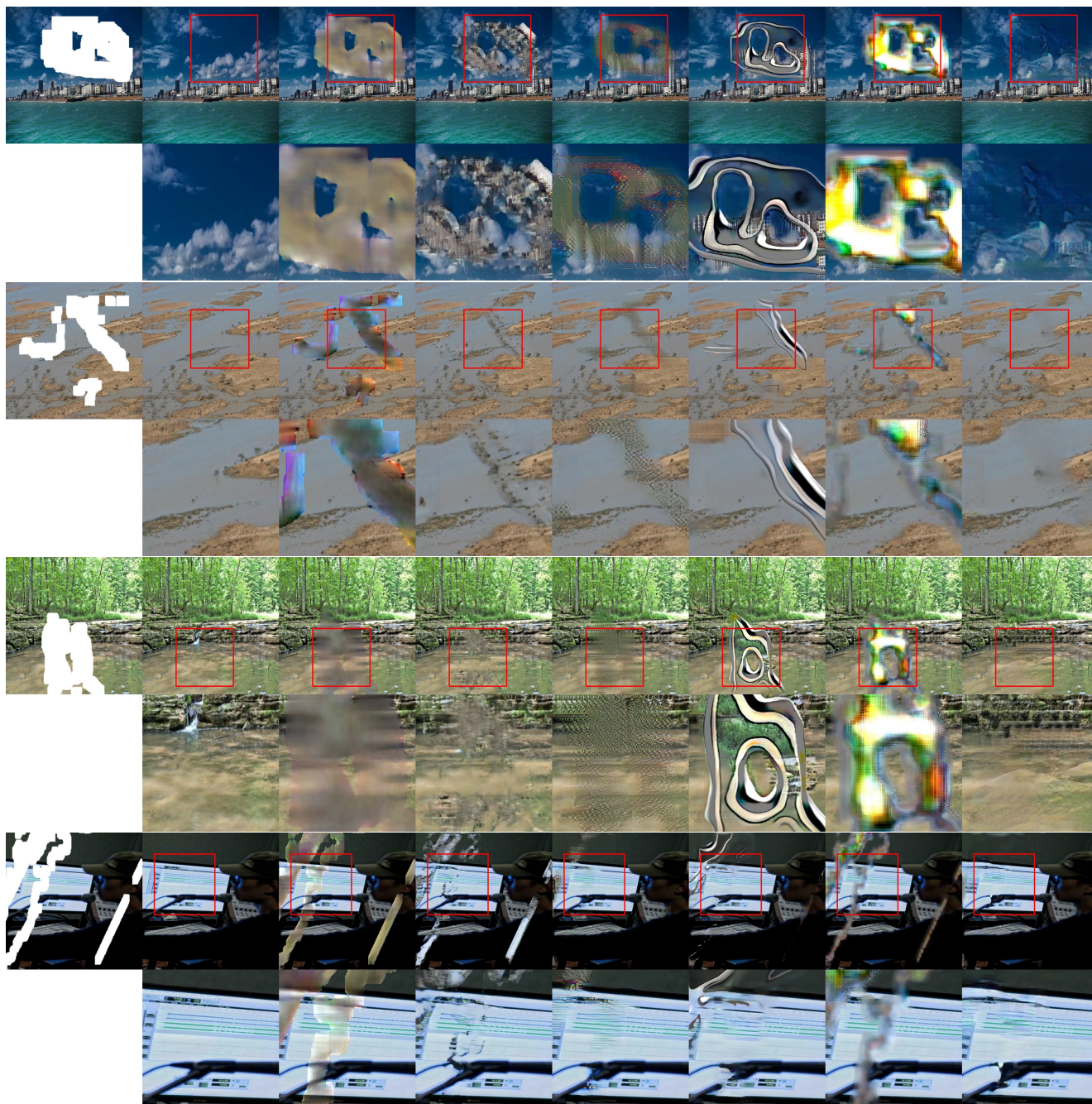
(e) DeepFillV2

(f) PEN-Net

(g) pconv

(h) Ours

Figure 1: Test results on places2 validation datasets with input size of 512×512 .



(a) Input

(b) GT

(c) global-local

(d) DeepFillV1

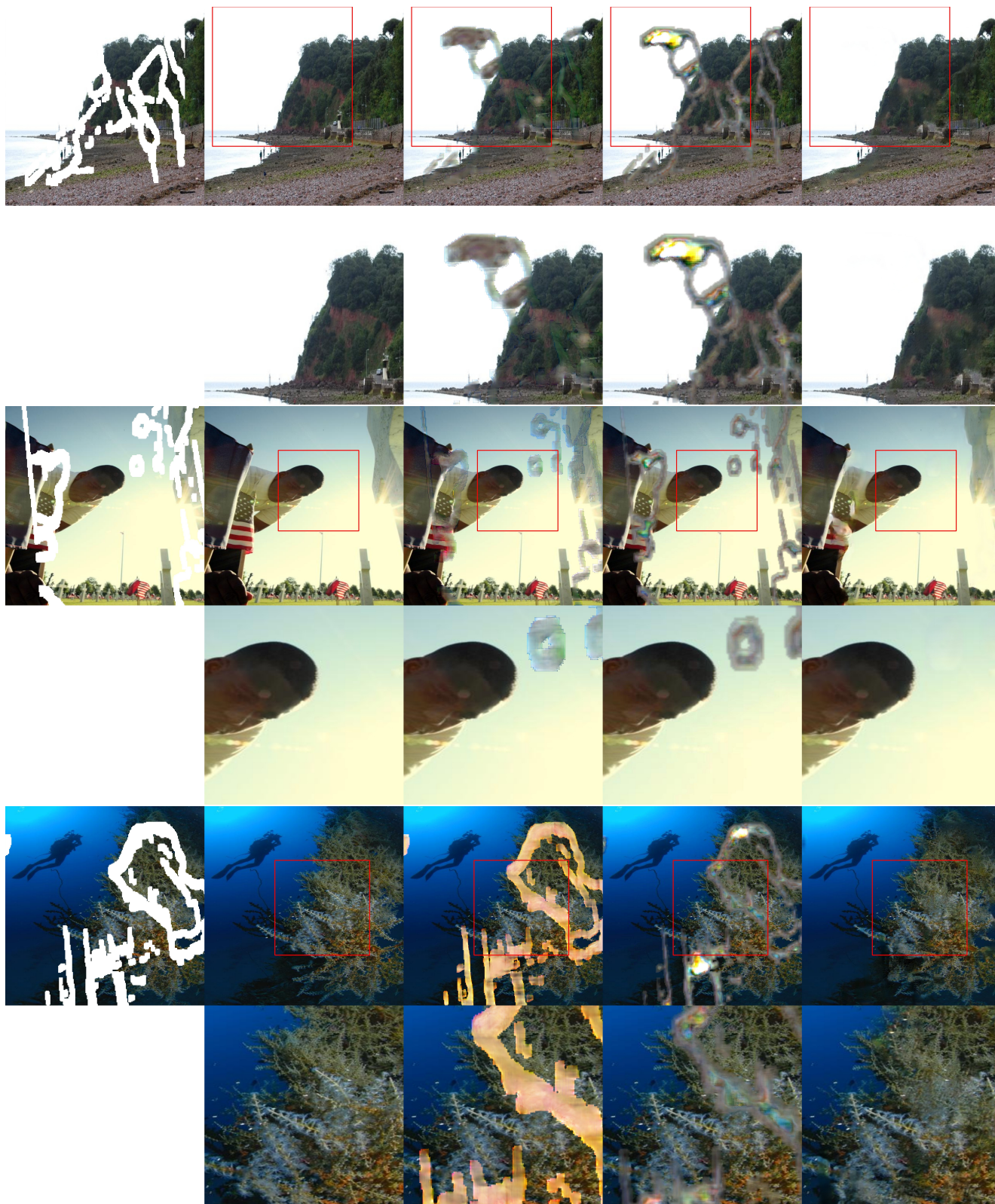
(e) DeepFillV2

(f) PEN-Net

(g) pconv

(h) Ours

Figure 2: Test results on places2 validation datasets with input size of 1024×1024 .



(a) Input

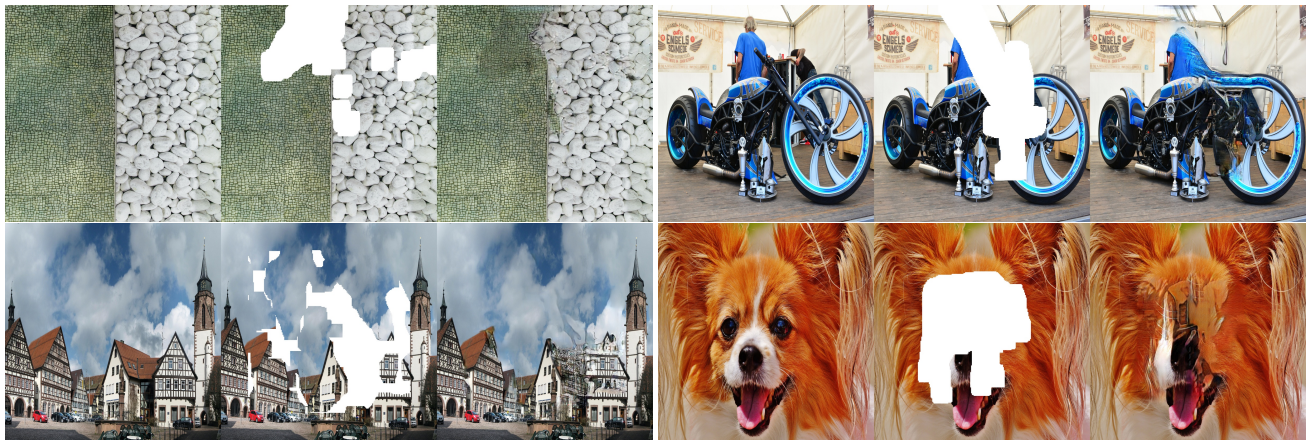
(b) GT

(c) global-local

(d) pconv

(e) Ours

Figure 3: Test results on places2 validation datasets with input size of 2048×2048 .



(a) Original

(b) Input

(c) Output

(d) Original

(e) Input

(f) Output

Figure 4: Failure examples of our model.