Supplementary material: Component Divide-and-Conquer for Real-World Image Super-Resolution

Pengxu Wei[1], Ziwei Xie[1], Hannan Lu[2], Zongyuan Zhan[1], Qixiang Ye[3], Wangmeng Zuo[2], and Liang Lin[⋆1,4]

[1] Sun Yat-sen University, Guangzhou, China
[2] Harbin Institute of Technology, Harbin, China
[3] University of Chinese Academy of Sciences, Beijing, China
[4] DarkMatter AI

weipx3@mail.sysu.edu.cn xiezw5@mail2.sysu.edu.cn hannanlu@hit.edu.cn
zhanzy178@gmail.com qxye@ucas.ac.cn wmzuo@hit.edu.cn linliang@ieee.org

In this supplementary material, we provide more details about our released real-world dataset, the extraction of the component guidance mask, and additional visualization results.

1 DRealSR Dataset

Image Collection. In this paper, we follow the pinehole camera model and zoom DSLR cameras to capture a group of LR and HR images. Specifically, firstly we annotate each camera for different scaling factors to consistently obtain each group of HR-LR pairs. We fix cameras on a tripod, adjust camera positions to capture the corresponding regions in a calibration board shown in Fig.1, and record the focal length for $\times1\sim\times4$ scaling factors. When collecting images, we zoom cameras according to the annotated focal lengths and use Bluetooth remote controller to capture images. Exposure and white balance are set to the automatic mode.

Image Alignment. There are some factors that invite image misalignment between HR and LR. For example, the camera zoom makes it necessary to crop LR to algin the contents of HR; lens distortion aggravates this misalignment issue. Nevertheless, these are nearly unavoidable due to complex imaging conditions. Considering the great difficulty of image alignment between HR and LR, to build a well-prepared real-world SR dataset for evaluation, we mainly resort to the mature image registration method and repeated examination with algorithms and manual selection.

Specifically, we employ SIFT method for image registration to crop LR images to match the content of HR images globally. Then we crop the whole images into patches, select informative patches with their mean and variance, and manually select blurry patches which might result from the depth of field. We conduct image registration for these patches again for local image alignment.

⋆ Corresponding Author

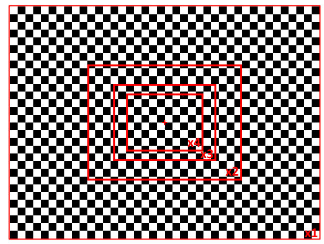

Fig. 1: Camera calibration board for four scaling factors.

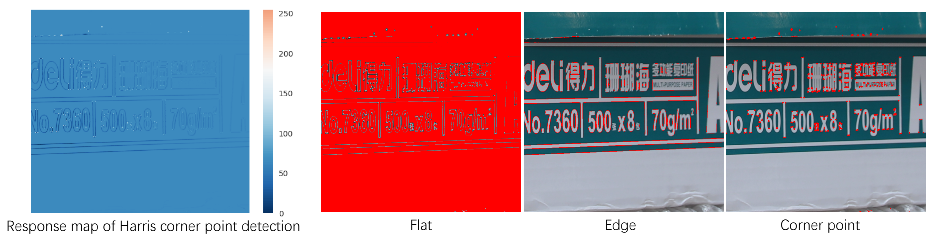

Fig. 2: An example of the response map extracted by Harris corner point detection method and the detected flat regions, edges and corner points denoted in red color.

These patches are converted into YCbCr color space to adjust the brightness of the Y channel using least square method. Their white balance is also adjusted using the Gray World method. Given the PSNR between HR patches and their upsampled LR, we select again patches with a threshold for training set. For our testing set, we further combine these aligned patches into a whole image and do not delete any patches; meanwhile, eight pixels along four sides in a whole image are removed to avoid the existence of black regions which result from the registration method.

2 Component guidance mask

In our proposed Component Divide-and-Conquer model (CDC), the component guidance mask $\boldsymbol{M}_e$ is extracted from an HR image in a CAB module that is specific to either *flat*, *edge* or *corner*. Specifically, Harris corner detection method [1] is employed on the HR image to derive a response map $\boldsymbol{R}$. The max value r_{max} of this response map is weighted by two scalars, *i.e.*, 0.01 and -0.001, respectively. We take these two weighted values as thresholds. On the response map, pixels whose values are smaller than the threshold $-0.001 * r_{max}$ are regarded as edges; those whose values are larger than the threshold $0.01 * r_{max}$ are regarded as corner points; those whose values are in the range $[-0.01 * r_{max}, 0.01 * r_{max}]$ are regarded as flat regions. In Fig. 2, we show an example of the response map and three images with the flat, edges and corner points denoted in red color.

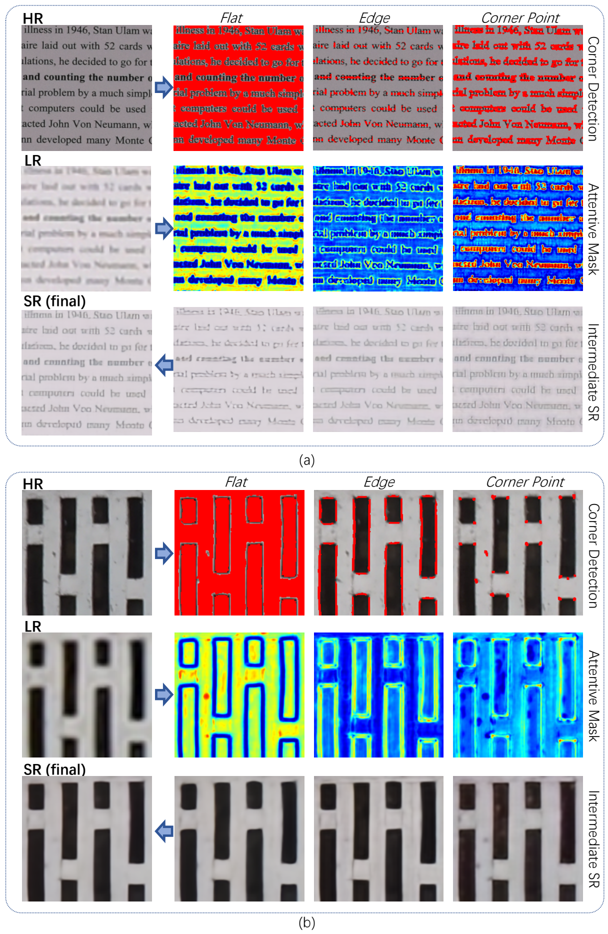

Fig. 3: Harris corner point detection, learned component-attentive masks and intermediate SR images from three CABs.

Accordingly, we generate three binary maps to indicate either flat regions, edges or corner points. These binary maps are our component guidance masks.

3 Visualization

The intermediate SR. To further analyze CDC, Fig. 3 provides more results of the generated intermediate SR. It is observed that three CABs focus on different image content reconstruction in different regions. For example, in Fig. 3 (a), the *flat* CAB focuses on the reconstruction of flat regions and its intermediate SR image has less details; the *edge* CAB generates sharp edges; the *corner point* CAB further reconstructs more details of the characters as well as sharp edges.

References

1. Harris, C.G., Stephens, M., et al.: A combined corner and edge detector. In: Alvey vision conference (1988)