

Focused and Collaborative Feedback Integration for Interactive Image Segmentation - Supplementary Material -

Qiaqiao Wei Hui Zhang Jun-Hai Yong
School of Software, BNRist, Tsinghua University, Beijing, China

wqq18@mails.tsinghua.edu.cn {huizhang, yongjh}@tsinghua.edu.cn

1. Network Architecture for the Collaborative Feedback Fusion Module

The precise layouts of the convolutional blocks Φ_1 , Φ_2 , and Φ_3 in the proposed Collaborative Feedback Fusion Module (CFFM) are shown in Fig. 1, Fig. 3, and Fig. 2, respectively. The stride is set to 1 for all convolutional layers, and the padding parameter is set to $(\#kernel_size - 1)/2$. Inspired by Xception [3] and FocalClick [2], we apply depthwise separable convolutions in the convolution block Φ_2 to update the feedback and the deep features.

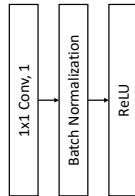
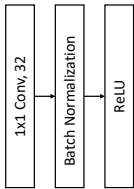


Figure 1. Specification for Φ_1 . Figure 2. Specification for Φ_3 .

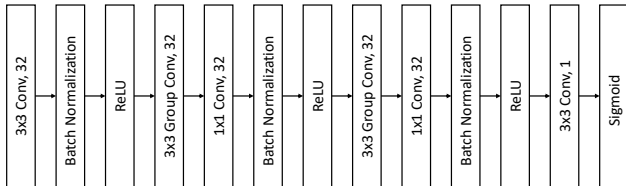


Figure 3. Specification for Φ_2 .

2. Additional Qualitative Results

In this section, we provide more qualitative results of our method. We visualize the interaction process and the resulting segmentation masks in Fig. 4. The examples are also illustrated in f-BRS [7]. Our method achieved 90% IoU with less than 3 clicks in those examples and has surpassed f-BRS in most cases. Even if our method failed in the first

interaction, it regained satisfactory segmentation results in the second interaction (see the fourth example in Fig. 4).

Fig. 5 illustrates some challenging examples of existing interactive image segmentation methods. The challenges include multimodality [4], occlusions, motion blur, thin structures, etc. In some cases, they can be addressed by providing more clicks to specify the position of the target object.

References

- [1] Xi Chen, Zhiyan Zhao, Feiwu Yu, Yilei Zhang, and Manni Duan. Conditional diffusion for interactive segmentation. In *ICCV*, pages 7345–7354, 2021. 3
- [2] Xi Chen, Zhiyan Zhao, Yilei Zhang, Manni Duan, Donglian Qi, and Hengshuang Zhao. Focalclick: Towards practical interactive image segmentation. In *CVPR*, pages 1300–1309, 2022. 1
- [3] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *CVPR*, pages 1251–1258, 2017. 1
- [4] Zhuwen Li, Qifeng Chen, and Vladlen Koltun. Interactive image segmentation with latent diversity. In *CVPR*, pages 577–585, 2018. 1
- [5] Zheng Lin, Zheng-Peng Duan, Zhao Zhang, Chun-Le Guo, and Ming-Ming Cheng. Focuscut: Diving into a focus view in interactive segmentation. In *CVPR*, pages 2637–2646, 2022. 3
- [6] Zheng Lin, Zhao Zhang, Lin-Zhuo Chen, Ming-Ming Cheng, and Shao-Ping Lu. Interactive image segmentation with first click attention. In *CVPR*, pages 13339–13348, 2020. 3
- [7] Konstantin Sofiiuk, Ilia Petrov, Olga Barinova, and Anton Konushin. f-brs: Rethinking backpropagating refinement for interactive segmentation. In *CVPR*, pages 8623–8632, 2020. 1, 3

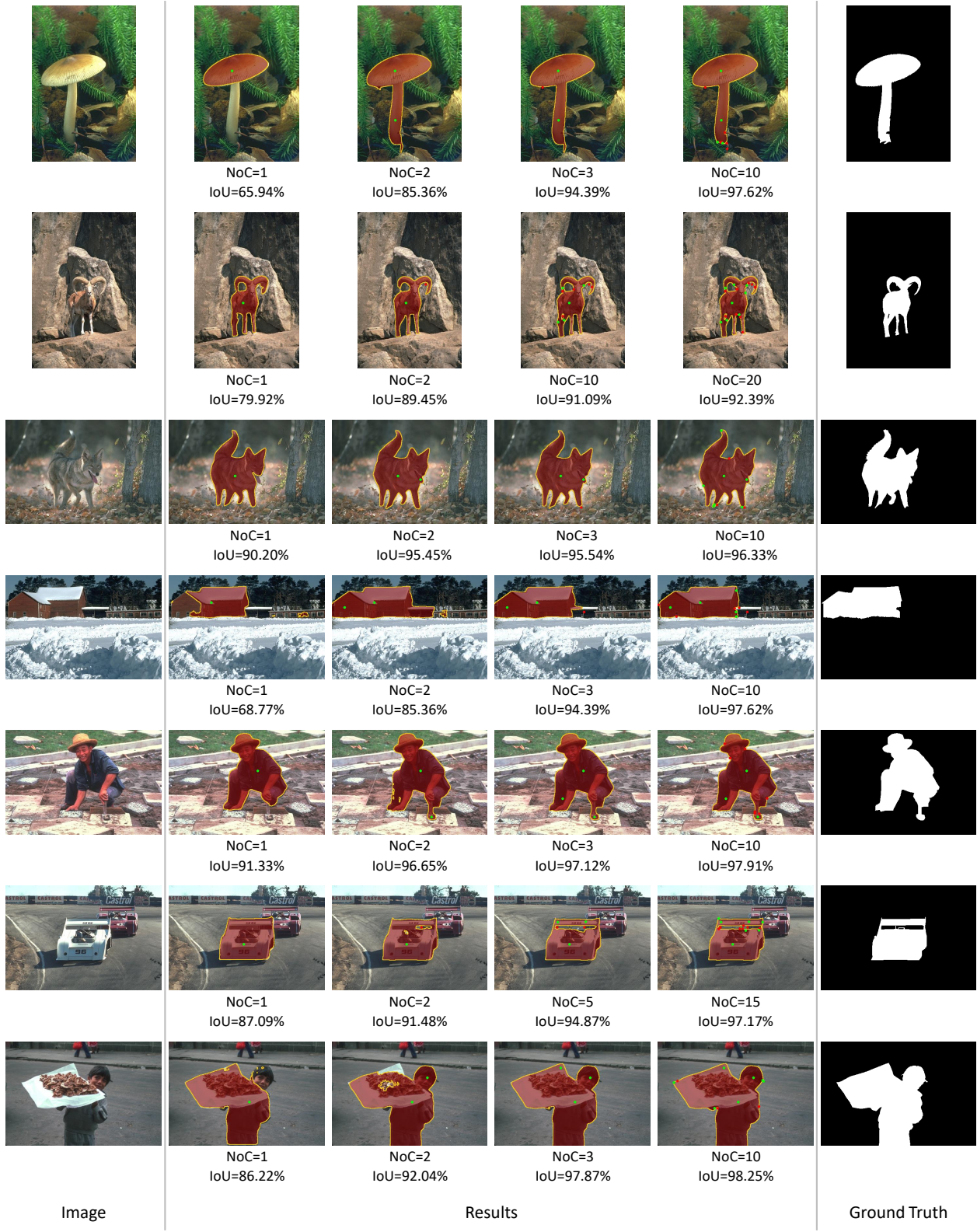


Figure 4. Visualization of the interaction process. Throughout this material, the green clicks denote foreground clicks, and the red ones denote background clicks.

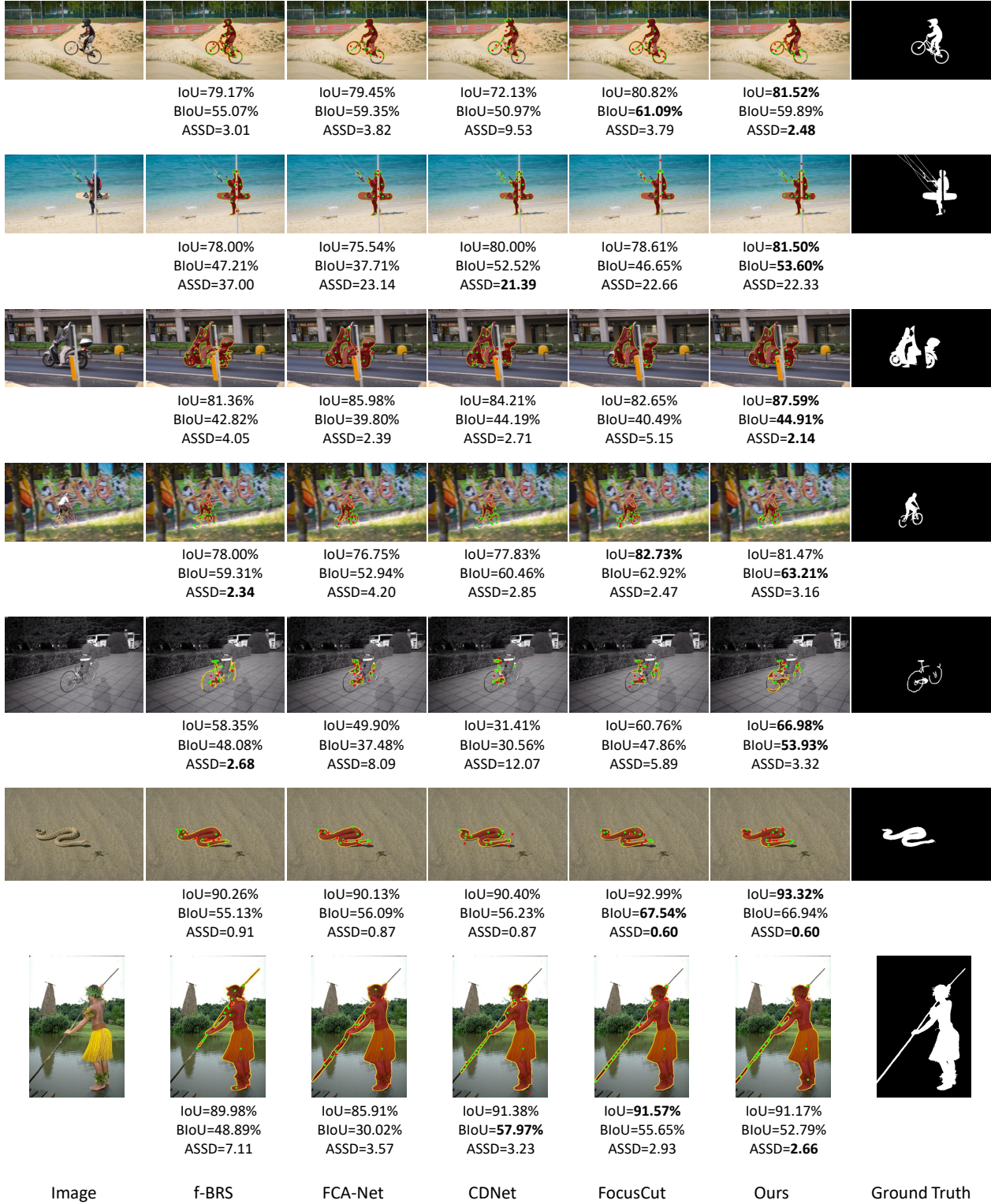


Figure 5. A comparison among f-BRS [7], FCA-Net [6], CDNet [1], FocusCut [5], and our method on some of the hard examples for existing interactive image segmentation methods. The number of clicks is 20 for all the visualized examples. The best results are in bold.