Supplementary File for Efficient Mask Correction for Click-Based Interactive Image Segmentation

Fei Du, Jianlong Yuan, Zhibin Wang, Fan Wang
Alibaba Group
{dufei.df, gongyuan.yjl, zhibin.waz, fan.w}@alibaba-inc.com

1. Speed on GPUs

Table 1 shows the average inference time per click of different methods on GPUs. Our mask correction network is more efficient compared to RITM [3] and FocalClick [1]

	HRNet18s	HRNet32	SegB0	SegB3
RITM	30 / 22	59 / 40	-	-
FocalClick	35 / 26	61 / 45	21 / 16	44 / 34
Ours-FirstClick	38 / 30	70 / 47	23 / 17	54 / 36
Ours-MaskCorrection	9 / 7	10 / 7	9 / 7	10 / 7

Table 1. The average inference time (ms) per click of comparison methods with different backbones on NVIDIA P100/V100 GPUs.

2. Results on COCO

We further test our method on the COCO validation set [2]. As shown in Table 2, our method shows better or comparable performance to the FocalClick [1]. However, on this challenging dataset where more clicks are required to achieve a high IoU, the overall inference time of our method is significantly lower than FocalClick [1].

	Ours		FocalClick	
Backbone	NoC@85/90	Time@85/90	NoC@85/90	Time@85/90
hrnet18s	5.44 / 9.00	18 / 26min	5.96 / 9.43	31 / 46min
hrnet32	5.06 / 8.47	20 / 30min	5.47 / 8.92	41 / 63min
SegB0	5.62 / 9.05	17 / 24min	5.75 / 9.17	22 / 33min
SegB3	4.79 / 8.22	18 / 28min	4.85 / 8.11	30 / 45min

Table 2. Comparison with FocalClick on the COCO validation set. The total inference time is measured on 4 NVIDIA V100 GPU.

3. Work with preexisting masks

Our mask correction network can also take as input a preexisting mask generated by other tools. Table 3 shows the results with and without the preexisting mask on DAVIS-585 dataset provided by FocalClick [1]. Our method can work with preexisting masks to reduce the number of clicks even if our method is not optimized for this purpose.

	w/ preexisting mask		w/o preexisting mask	
Methods	NoC@85	NoC@90	NoC@85	NoC@90
Ours-hrnet18s	2.94	4.32	5.01	7.40
Ours-hrnet32	2.97	4.47	4.32	6.63
Ours-SegB0	3.43	5.05	5.03	7.56
Ours-SegB3	3.14	4.62	4.27	6.55

Table 3. The performance of our methods with and without the preexisting mask on DAVIS-585 dataset.

References

[1] 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

[2] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, volume 8693, pages 740–755, 2014. 1

[3] Konstantin Sofiiuk, Ilya A Petrov, and Anton Konushin. Reviving iterative training with mask guidance for interactive segmentation. In *ICIP*, pages 3141–3145, 2022. 1