f-BRS: Rethinking Backpropagating Refinement for Interactive Segmentation Supplementary materials

Konstantin Sofiiuk Ilia Petrov Olga Barinova Anton Konushin
{k.sofiiuk, ilia.petrov, o.barinova, a.konushin}@samsung.com
Samsung AI Center – Moscow

1. Analysis of the average IoU according to the number of clicks

We computed the mean IoU score according to the number of clicks for GrabCut, Berkeley, SBD and DAVIS datasets (see Figure 1). We also evaluated the BRS [1] model from authors' public repository for a fair comparison.

On the plots you can see that f-BRS-B has drops on DAVIS and SBD datasets at the number of clicks 9. This is due to the fact that f-BRS can sometimes fall into a bad local minimum. This issue can be solved by setting a higher regularization coefficient λ in the BRS loss function. However, with the increase of the λ, the convergence of the method at a large number of clicks becomes worse.

2. Measuring the limitation of f-BRS

We decided to find out the limit of accuracy that can be obtained using only f-BRS, adjusting scales and biases for an intermediate layer in the DeepLabV3+ head. For this, we first evaluated the model for 20 clicks using the standard protocol. Then we continued with L-BFGS-B optimization for scales and biases using ground truth mask as loss target instead of interactive clicks. It equals to using all pixels of the image as input clicks (positive click for each foreground pixel and negative for each background pixel). We estimated the mean IoU score for each dataset which is shown in Figure 1 (f-BRS-B Oracle).

The figure illustrates that the accuracy limit the algorithm can reach is highly dependent on the dataset. DAVIS and SBD datasets are much harder than GrabCut and Berkeley. DAVIS has many complex masks labeled with pixel perfect precision, which is closer to the task of image matting. On the contrary, SBD has many masks with rough or inaccurate annotation.

3. Full evaluation results for all our methods

We report the NoC@85 and NoC@90 metrics for GrabCut, Berkeley, SBD and DAVIS datasets for all BRS variations with different backbones (ResNet-34, ResNet-50 and ResNet-101). The use of BRS leads to consistent improvement in accuracy. All these results are presented in Table 1.

Overall, the choice of a backbone only slightly affects the methods' accuracy on GrabCut and Berkeley datasets. However, we noticed a significant difference between ResNet-34 and ResNet-101 while testing on SBD validation dataset, which has the closest distribution to the training one. In most cases, DistMap-BRS shows slightly worse NoC compared to RGB-BRS.

4. Additional interactive segmentation results

We also provide more results of our interactive segmentation algorithm (f-BRS-B with ResNet-50) on different images. Figure 2 and 3 represent good cases, while Figure 4 represents bad cases when testing on Berkeley dataset.

Figure 5 shows some of the worst results of testing on DAVIS dataset. The algorithm does not even match 85% IoU in 20 clicks.

References

[1] Won-Dong Jang and Chang-Su Kim. Interactive image segmentation via backpropagating refinement scheme. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5297–5306, 2019. 1

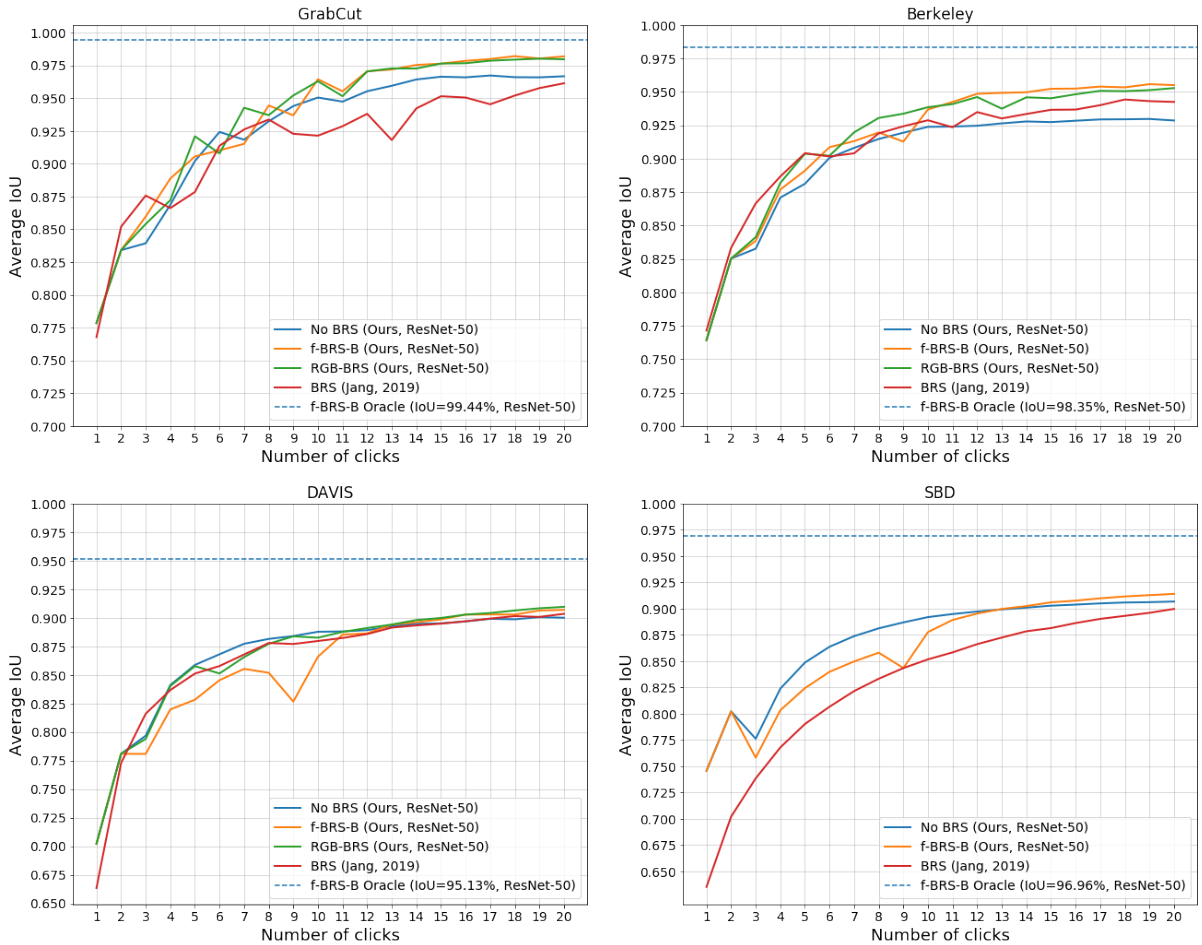

Figure 1. Comparison of the average IoU scores according to the number of clicks on GrabCut, Berkeley, DAVIS and SBD datasets. The dashed horizontal line shows the average IoU limit that can theoretically be reached by f-BRS-B method (for more details see Section 2).

Method		GrabCut		Berkeley		SBD		DAVIS	
		NoC@85	NoC@90	NoC@85	NoC@90	NoC@85	NoC@90	NoC@85	NoC@90
Ours w/o BRS	ResNet-34	2.52	3.20	3.09	5.31	5.51	8.58	5.47	8.51
	ResNet-50	2.64	3.32	3.29	5.18	5.10	8.01	5.39	8.18
	ResNet-101	2.50	3.18	3.45	6.25	5.28	8.13	5.12	8.01
Ours RGB-BRS	ResNet-34	2.00	2.52	2.51	4.28	4.72	7.45	5.30	7.86
	ResNet-50	2.38	2.94	2.65	4.08	4.45	7.12	5.28	7.58
	ResNet-101	2.00	2.48	2.26	4.21	4.17	6.69	4.95	7.09
Ours DistMap-BRS	ResNet-34	1.98	2.54	2.45	4.41	4.85	7.66	5.34	8.11
	ResNet-50	2.36	2.90	2.67	4.17	4.63	7.37	5.35	7.93
	ResNet-101	2.00	2.46	2.21	4.41	4.42	7.10	5.03	7.63
Ours f-BRS-A	ResNet-34	1.94	2.54	2.66	4.36	5.11	8.17	5.39	8.09
	ResNet-50	2.54	3.06	2.74	4.44	4.94	7.97	5.37	7.54
	ResNet-101	2.08	2.62	2.39	4.79	4.68	7.58	5.01	7.21
Ours f-BRS-B	ResNet-34	2.00	2.46	2.60	4.65	5.25	8.30	5.39	8.21
	ResNet-50	2.50	2.98	2.77	4.34	5.06	8.08	5.39	7.81
	ResNet-101	2.30	2.72	2.52	4.57	4.81	7.73	5.04	7.41
Ours f-BRS-C	ResNet-34	2.10	2.54	2.72	4.48	5.23	8.11	5.47	8.35
	ResNet-50	2.60	3.10	2.89	4.90	5.05	7.97	5.50	7.90
	ResNet-101	2.18	2.68	2.64	4.64	4.85	7.64	5.11	7.37

Table 1. Evaluation results on GrabCut, Berkeley, SBD and DAVIS datasets.

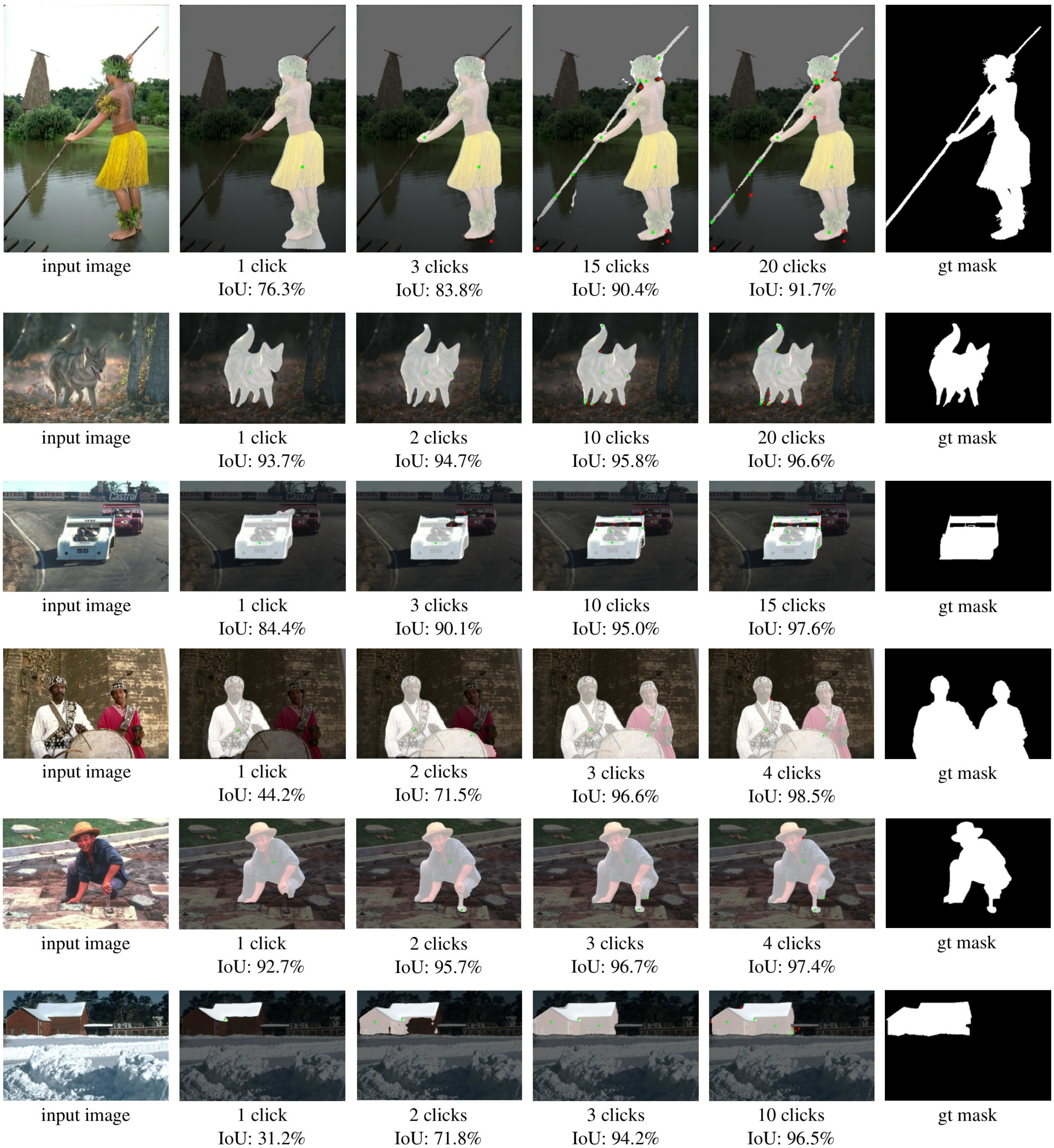

Figure 2. Examples of good convergence of the proposed f-BRS-B method with ResNet-50 backbone on Berkeley dataset.

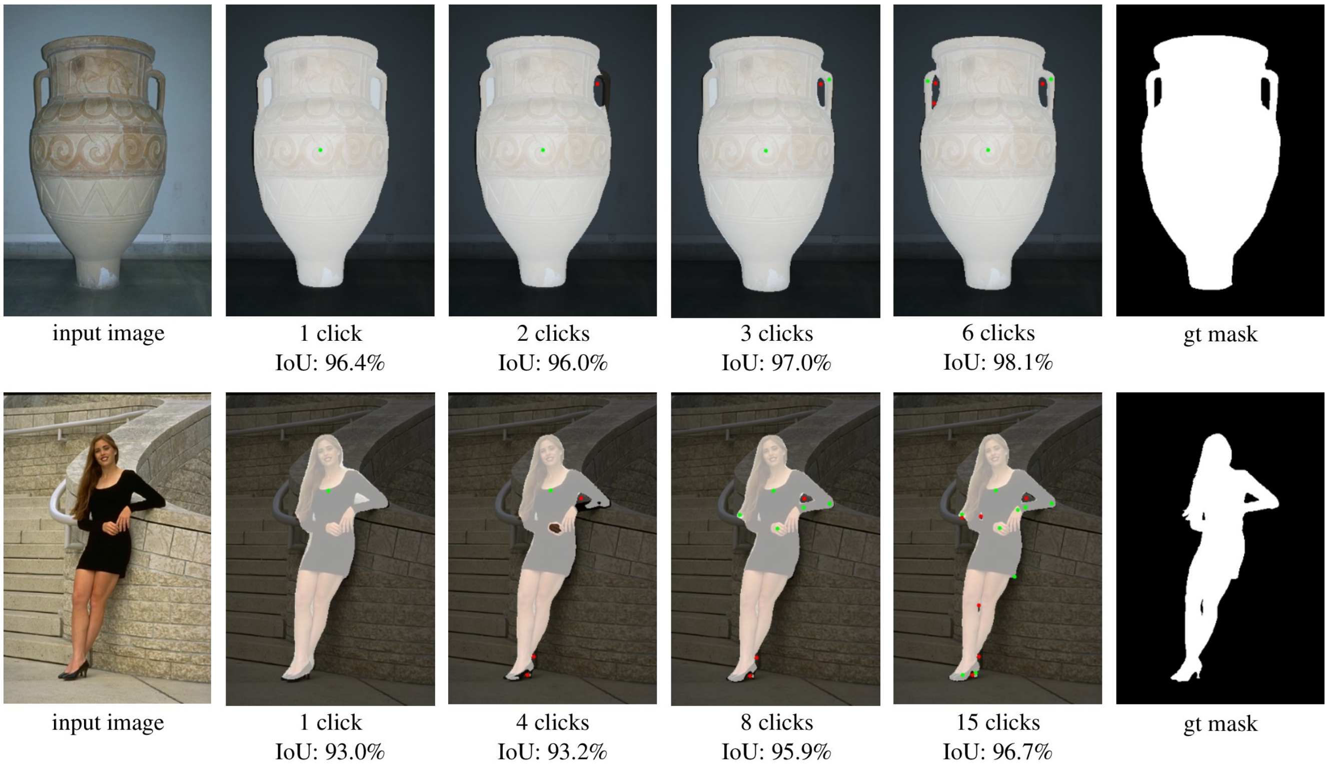

Figure 3. Examples of good convergence of the proposed f-BRS-B method with ResNet-50 backbone on Berkeley dataset.

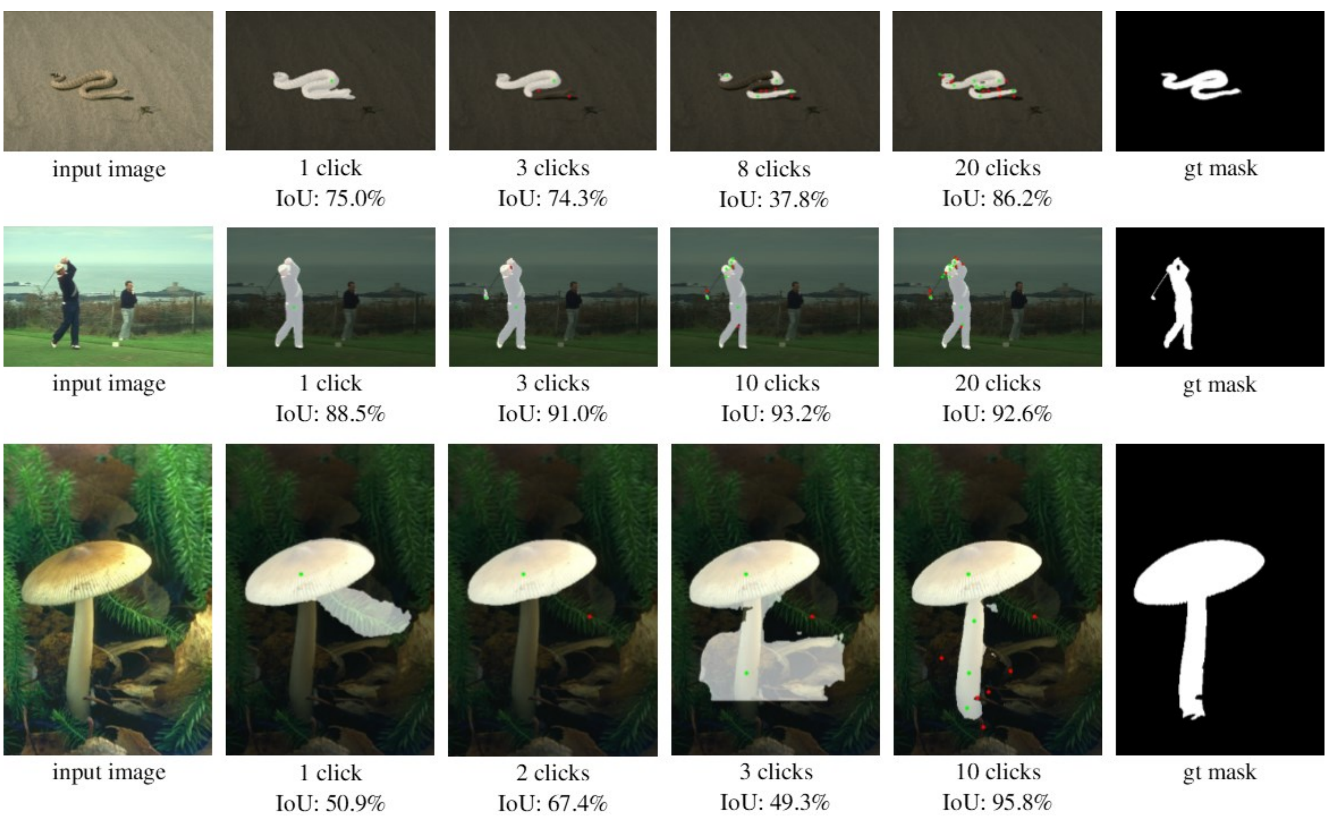

Figure 4. Some challenging examples from Berkeley dataset.

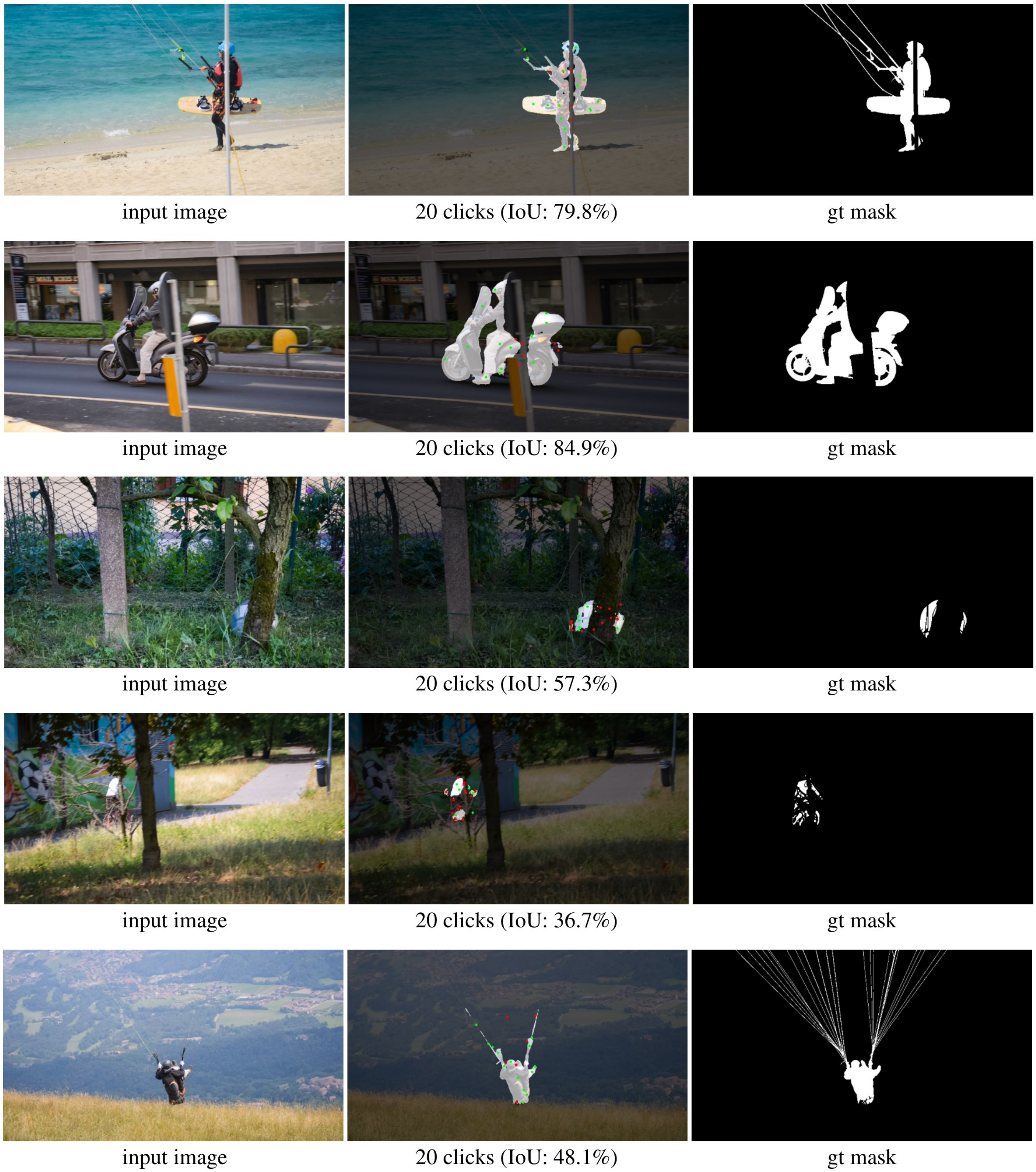

Figure 5. Some of the worst examples from DAVIS dataset.