Supplementary Material for Continuous Adaptation for Interactive Segmentation Using Teacher-Student Architecture

Barsegh Atanyan[1], Levon Khachatryan[1], Shant Navasardyan[1], Yunchao Wei[2], Humphrey Shi[1,3]
[1]Picsart AI Research (PAIR), [2]BJTU, [3]Georgia Tech
https://github.com/Picsart-AI-Research/Interactive-Segmentation-with-Continuous-Adapation

1. Implementation Details

We have conducted our experiments with FocalClick SegFormerB3-S2 [3, 9] as the off-the-shelf pre-trained interactive segmentation model. To update student model parameters, we have used Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, batch size 1 and learning rate 5×10^{-5} (for DRIONS-DB [1] and DAVIS [5] learning rate is 10^{-4} and 10^{-5} correspondingly). We set $\gamma = 2$ in both L_C and L_I. In case of continual adaptation a new optimizer is used for each new dataset D_l. We set $\alpha = 0.999$ for exponential moving average (EMA) updates of the teacher model parameters. Also, during the adaptation on each dataset D_l parameter change regularizer L_R uses the teacher model parameters obtained after adapting on dataset D_{l-1} as initial parameters.

To verify that the proposed approach does not depend on the off-the-shelf pre-trained interactive segmentation model and deteriorates catastrophic forgetting we use RITM [7] with HRNet-18 backbone [2]. The student model is updated using Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, batch size 1 and learning rate 5×10^{-6} (for DRIONS-DB [1] and DAVIS [5] learning rate is 10^{-5} and 10^{-6} correspondingly). The teacher update rule is the same.

We use a GeForce RTX 2080 for our experiments.

2. More Qualitative Results

More qualitative results are provided to demonstrate the effectiveness of the proposed method to tackle catastrophic forgetting in continual adaptation. Figures 1 and 2 compare adaptation results of the baseline and our methods to show that it takes fewer click to achieve a higher IOU from our method. Figures 3, 4, 5 and 6 demonstrate the effectiveness of the proposed method for adaptation on new datasets comparing with SOTA interactive segmentation method FocalClick [3].

References

[1] Enrique J Carmona, Mariano Rincón, Julián García-Feijoó, and José M Martínez-de-la Casa. Identification of the optic nerve head with genetic algorithms. *Artificial intelligence in medicine*, 43(3):243–259, 2008. 1, 2, 3

[2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 1

[3] Xi Chen, Zhiyan Zhao, Yilei Zhang, Manni Duan, Donglian Qi, and Hengshuang Zhao. Focalclick: towards practical interactive image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1300–1309, 2022. 1, 4, 5, 6, 7

[4] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001. 3

[5] Federico Perazzi, Jordi Pont-Tuset, Brian McWilliams, Luc Van Gool, Markus Gross, and Alexander Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 724–732, 2016. 1

[6] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "grabcut" interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)*, 23(3):309–314, 2004. 2

[7] Konstantin Sofiiuk, Ilya A Petrov, and Anton Konushin. Reviving iterative training with mask guidance for interactive segmentation. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3141–3145. IEEE, 2022. 1

[8] Xiaolu Sun, C Mario Christoudias, and Pascal Fua. Free-shape polygonal object localization. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13*, pages 317–332. Springer, 2014. 2, 3

[9] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021. 1

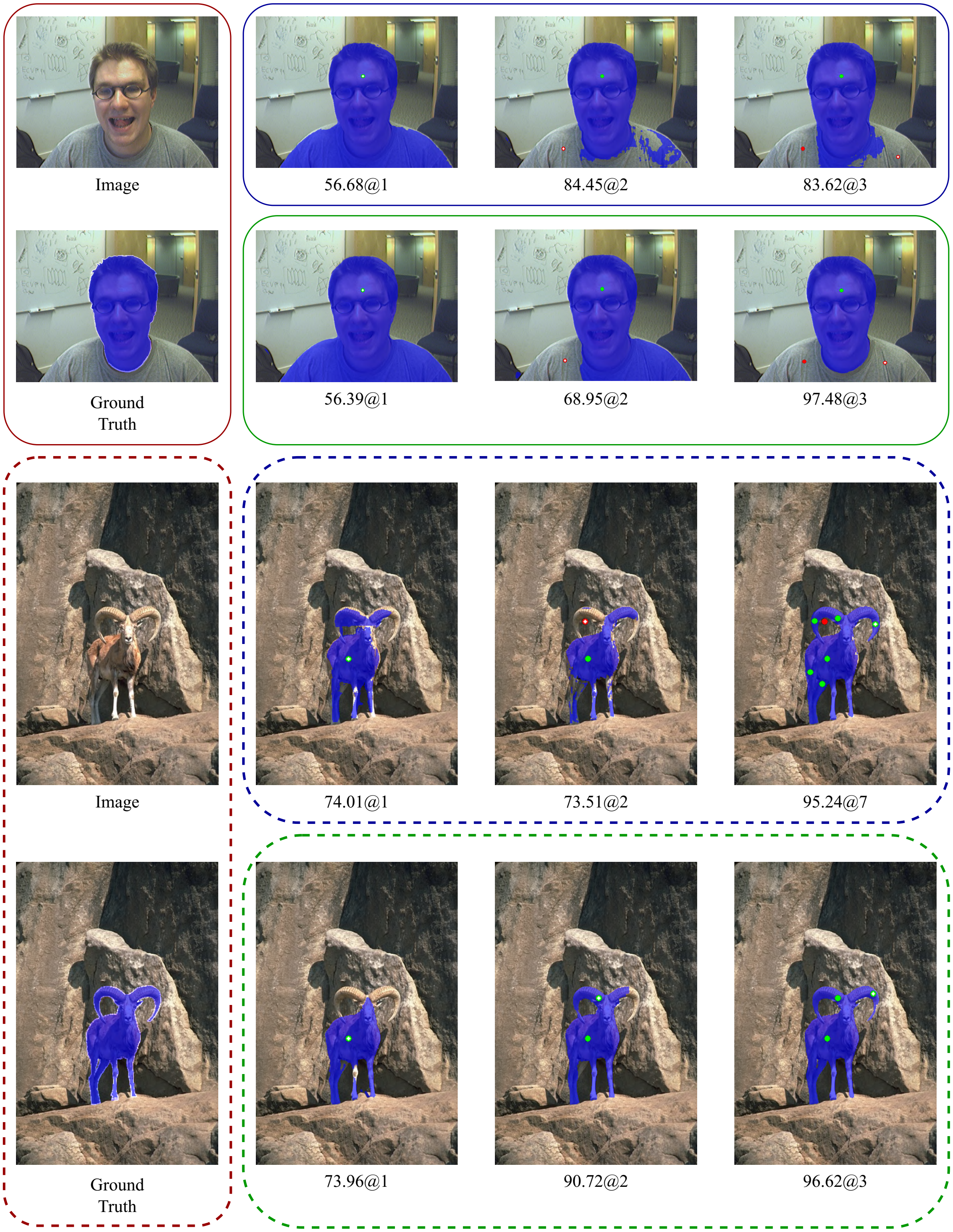

Figure 1. Comparison between the baseline and our method. Images are taken from $GrabCut$ dataset. To illustrate the effectiveness of the proposed method to decrease catastrophic forgetting, the baseline and our method have been continuously adapted on DRIONS-DB [1] $\rightarrow$ Rooftop [8] $\rightarrow$ GrabCut [6]. Green and red points represent positive and negative clicks correspondingly. Blue horizontal boxes include results of the baseline method. Green horizontal boxes include results for the proposed teacher-student approach. IOU@Number of Clicks is reported. Red vertical boxes include the image and the ground truth.

Figure 2. Comparison between the baseline and our method. Images are taken from *Berkeley* dataset. To illustrate the effectiveness of the proposed method to decrease catastrophic forgetting, the baseline and our method have been continuously adapted on DRIONS-DB [1] → Rooftop [8] → Berkeley [4]. Green and red points represent positive and negative clicks correspondingly. Blue horizontal boxes include results of the baseline method. Green horizontal boxes include results for the proposed teacher-student approach. IOU@Number of Click is reported. Red vertical boxes include the image and the ground truth.

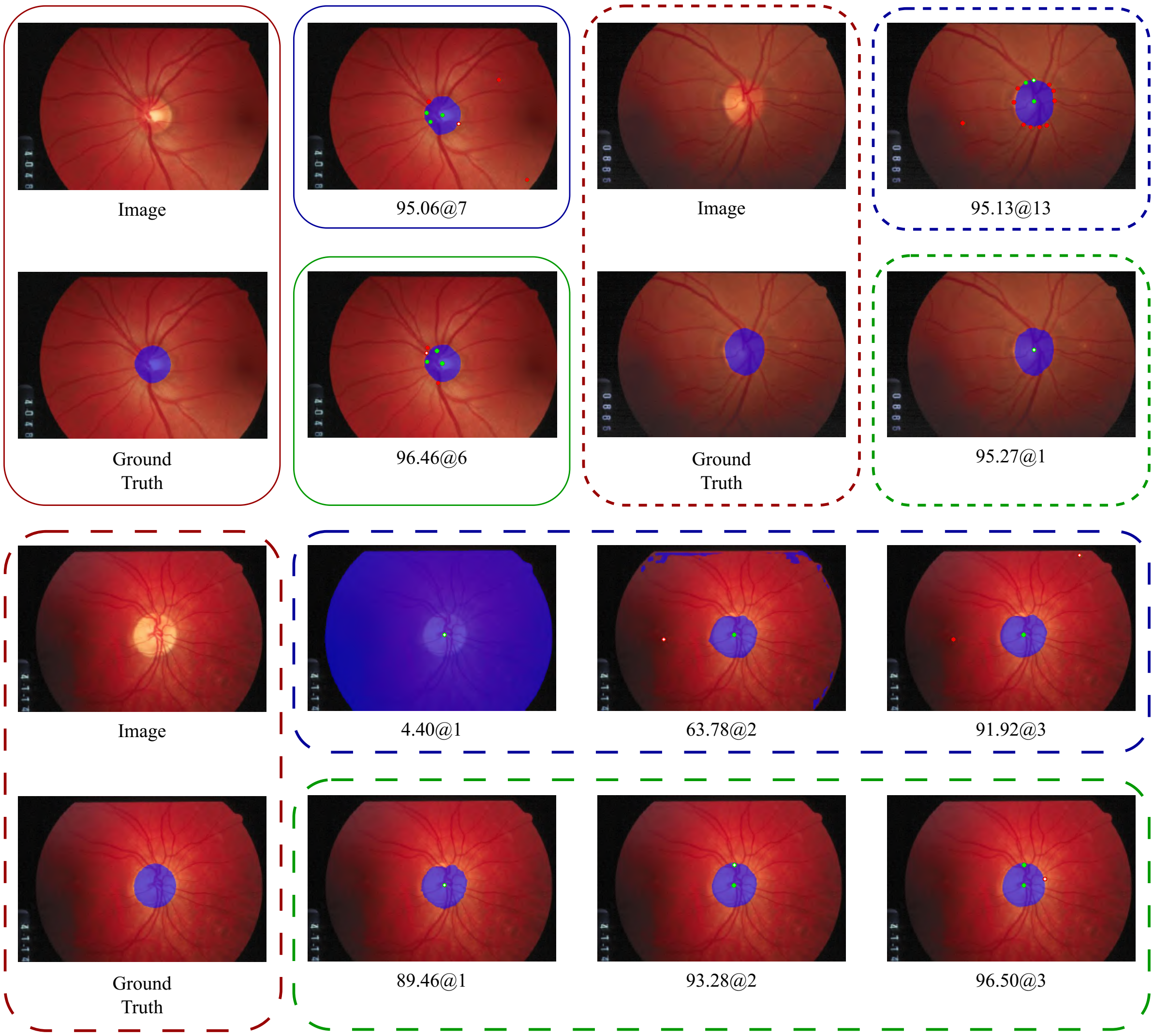

Figure 3. Comparison between FocalClick [3] and our method. Images are taken from $DRIONS-DB$ dataset. Green and red points represent positive and negative clicks correspondingly. Blue horizontal boxes include results of FocalClick. Green horizontal boxes include results for the proposed teacher-student approach. IOU@Number of Click is reported. Red vertical boxes include the image and the ground truth.

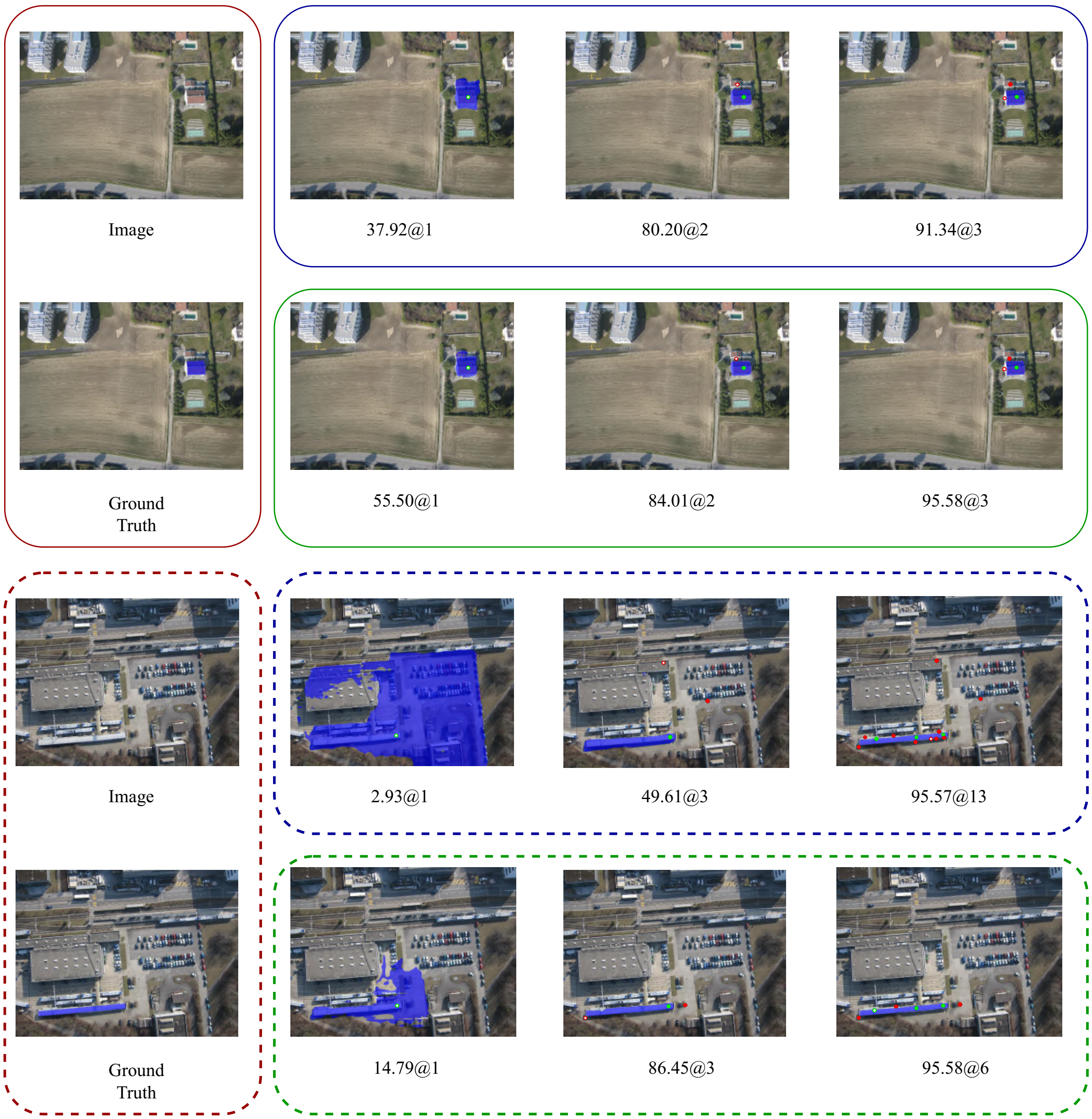

Figure 4. Comparison between FocalClick [3] and our method. Images are taken from $Rooftop$ dataset. Green and red points represent positive and negative clicks correspondingly. Blue horizontal boxes include results of FocalClick. Green horizontal boxes include results for the proposed teacher-student approach. IOU@Number of Click is reported. Red vertical boxes include the image and the ground truth.

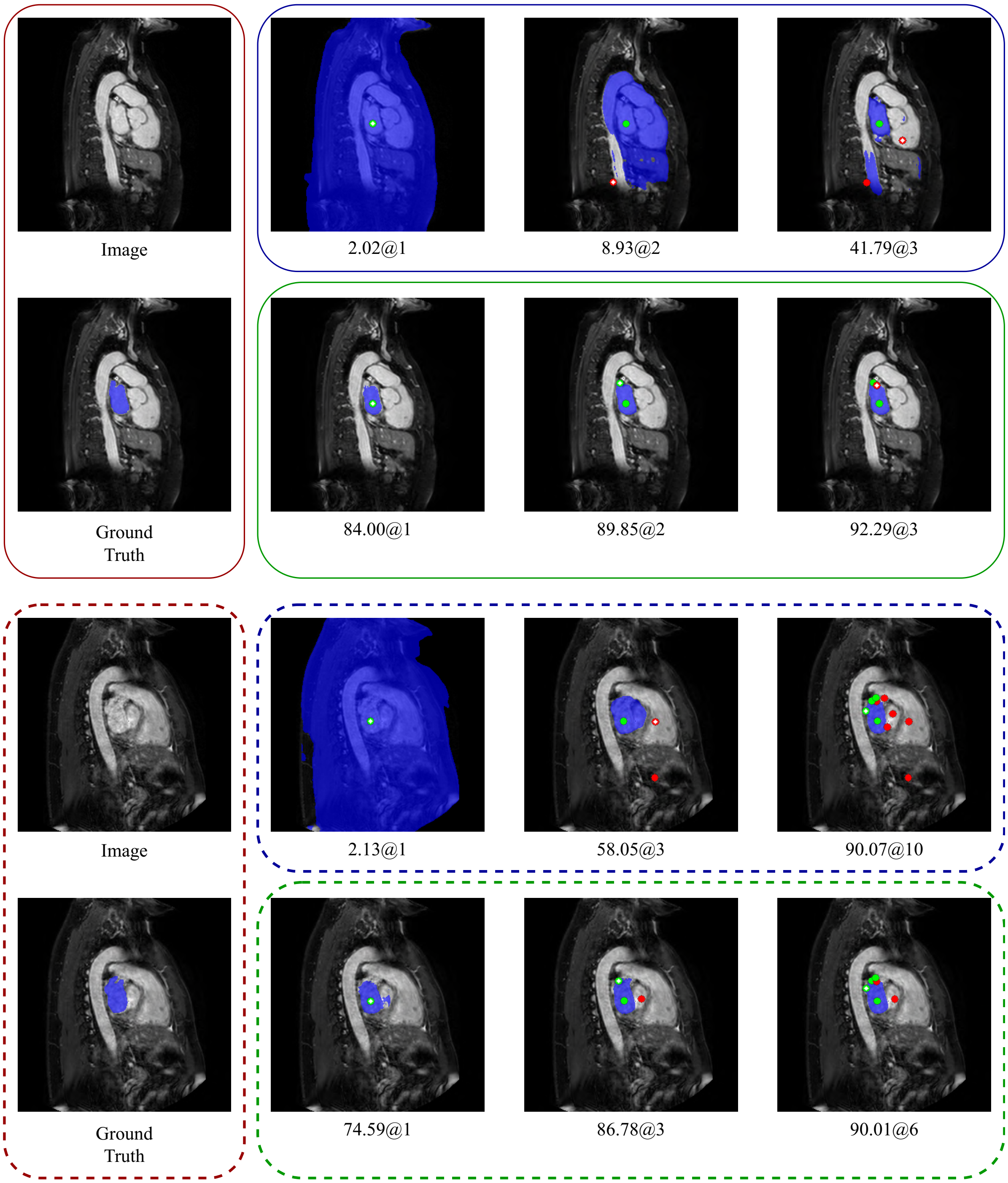

Figure 5. Comparison between FocalClick [3] and our method. Images are taken from $Heart$ dataset. Green and red points represent positive and negative clicks correspondingly. Blue horizontal boxes include results of FocalClick. Green horizontal boxes include results for the proposed teacher-student approach. IOU@Number of Click is reported. Red vertical boxes include the image and the ground truth.

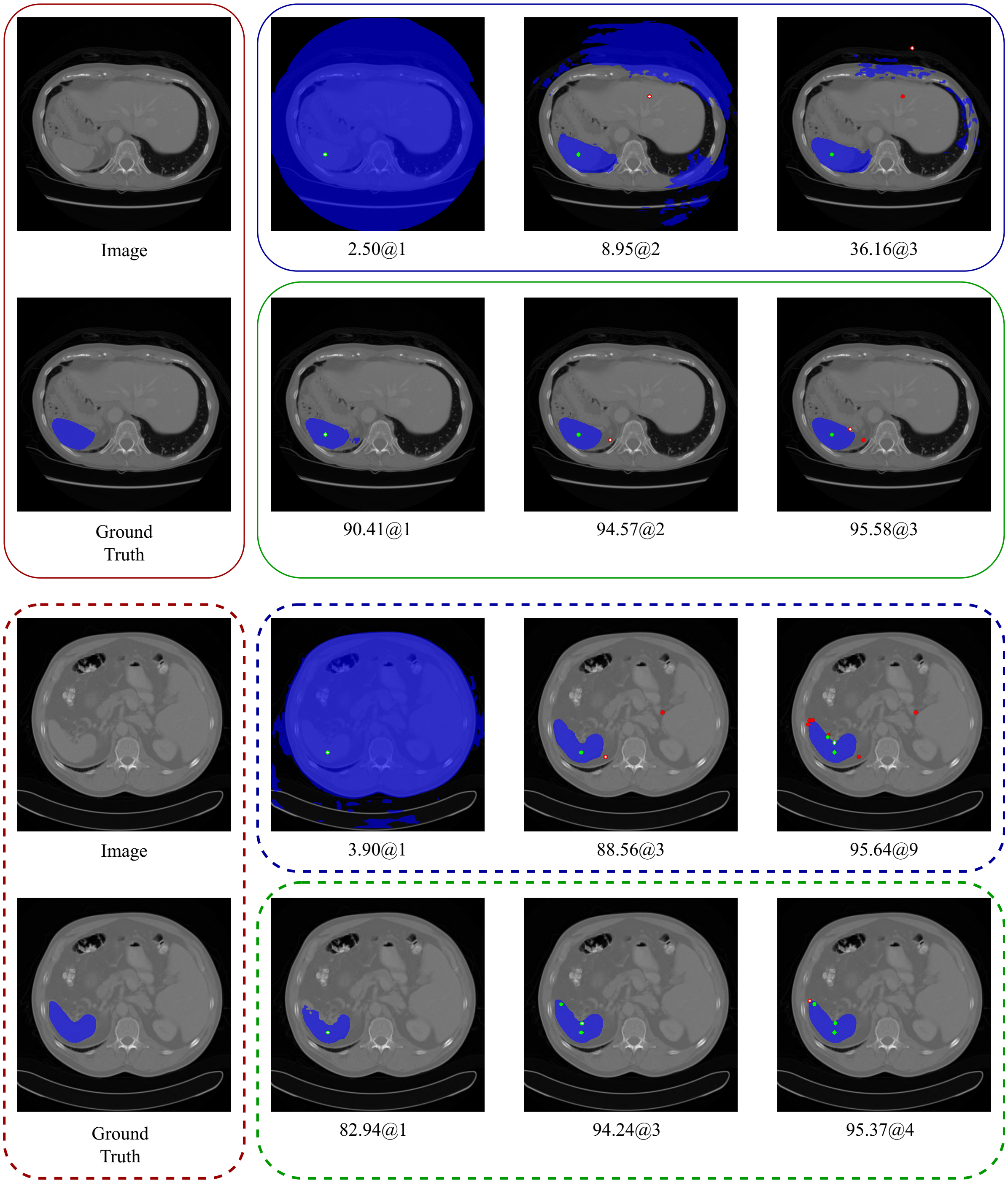

Figure 6. Comparison between FocalClick [3] and our method. Images are taken from *Spleen* dataset. Green and red points represent positive and negative clicks correspondingly. Blue horizontal boxes include results of FocalClick. Green horizontal boxes include results for the proposed teacher-student approach. IOU@Number of Click is reported. Red vertical boxes include the image and the ground truth.