Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn


# Supplementary materials for

Chuyun SHEN, Wenhao LI, Qisen XU, Bin HU, Bo JIN, Haibin CAI, Fengping ZHU, Yuxin LI, Xiangfeng WANG, 2023. Interactive medical image segmentation with self-adaptive confidence calibration. *Front Inform Technol Electron Eng*, 24(9):1332-1348. https://doi.org/10.1631/FITEE.2200299

## 1 More related works

Image segmentation is a fundamental problem in computer vision or image processing that has been widely studied. Deep learning (DL) has further promoted the development of automatic segmentation algorithms. Convolutional neural network (CNN) type methods are typical DL algorithms for image segmentation, e.g., fully convolutional networks (FCNs) (Shelhamer et al., 2017) and DeepLab (Chen et al., 2018). U-Net (Ronneberger et al., 2015), which is considered an evolutionary variant of FCN, becomes one of the state-of-the-art methods and performs better in medical image segmentation. Medical image segmentation is the key to modern auxiliary diagnosis and treatment response evaluation. A series of related works have been published with progressive performance for medical image segmentation (Milletari et al., 2016; Kamnitsas et al., 2017a; Li et al., 2017). In the following, we will review the development of interactive image segmentation methods and discuss the uncertainty estimation for image segmentation relevant to our proposed algorithm.

Traditional interactive image segmentation methods: The classical random walk (Grady, 2006) can create a weight map with pixels as vertices and segment the image based on user interactions. GrabCut (Rother et al., 2004) and GraphCut (Boykov and Jolly, 2001) were designed to associate image segmentation with the maximum flow and minimum cut algorithms on graphs, respectively, while GeoS (Criminisi et al., 2008) was proposed to evaluate the similarity between pixels using the geodesic distance. These traditional methods aim to utilize additional expert interaction information to modify the segmentation performance further.

DL-based interactive image segmentation methods: Xu et al. (2016) segmented images based on CNN interactively. DeepCut (Rajchl et al., 2017) and ScribbleSup (Lin et al., 2016) both employ weakly supervised expert hints to establish interactive image segmentation methods. DeepIGeoS (Wang et al., 2019) employs the geodesic distance metric to construct a hint map. The interactive segmentation process can be considered a sequential iterative process. It becomes natural to introduce the reinforcement learning (RL) framework to model the interactive segmentation process. Polygon-RNN (Castrejón et al., 2017) fundamentally segments each target as a polygon and iteratively chooses the polygon vertices through a recurrent neural network (RNN). Polygon-RNN++ (Acuna et al., 2018) employs almost the same idea as Polygon-RNN, but it learns to choose vertices by RL. SeedNet (Lee and Song, 2018) trains an expert interaction generation RL model that obtains newly simulated interaction information at each interaction step. IteR-MRL (Liao et al., 2020) and BS-IRIS (Ma et al., 2021) both model the dynamic interaction process as a Markov decision process (MDP) and employ multi-agent RL (MARL) models to segment images. Some researchers aimed to reduce the annotation cost of interactive image segmentation. IFSL (Feng et al., 2021) introduces interactive learning into the few-shot learning strategy and addresses the annotation burden of medical image segmentation models. IOG (Zhang et al., 2020) uses a practical inside-outside guidance approach to minimize the labeling cost. It is difficult for these interactive methods to effectively utilize experts' short- and long-term interaction information simultaneously, thus necessitating error correction operations.

Uncertainty estimation for image segmentation: Uncertainty estimation is helpful in the context of deployed machine learning systems because it can detect when a neural network is likely to make an incorrect prediction or when the input may be out of distribution. Traditionally, many of the works were inspired by Bayesian statistics, or the Bayesian neural network (BNN) (MacKay, 1992; Neal, 1995). Unfortunately, Bayesian inference is computationally intractable in practice, so much effort has been put into developing approximations of BNNs that are easier to train. Recent efforts to approximate BNNs in this area include Monte-Carlo dropout (Gal and Ghahramani, 2016), multiplicative normalizing flows (Louizos and Welling, 2017), and stochastic batch normalization (Atanov et al., 2018). These methods are capable of producing uncertainty estimates, although with varying degrees of success. The main disadvantage of these BNN approximations is that they require sampling to generate the output distributions. As such, uncertainty estimates are often time-consuming or resource-intensive to produce, requiring 10 to 100 forward passes through a neural network to produce useful uncertainty estimates at inference time. An alternative to BNNs is the ensemble method (Dietterich, 2000; Kamnitsas et al., 2017b; Lakshminarayanan et al., 2017; Mehrtash et al., 2018, 2020), which uses a frequentist approach for uncertainty estimation by training many models and observing the variance in their predictions. However, this technique is still resource-intensive, because it requires inference from multiple models to produce the uncertainty estimate. A promising alternative to sampling-based methods is to have the neural network learn what its uncertainty should be for any given input, i.e., learning-based uncertainty estimation or confidence learning.

## 2 More visualizations

Fig. S1 presents a visualization of the segmentation process of our proposed method. Our proposed method models the whole interaction process. Fig. S1 shows the current interaction step and its previous and next interaction steps. At each step, the second column is the confidence map. We find that the confidence values of object edges are always lower than those in other regions, and these regions will receive more "punishment" when rewards are generated. We observe that MECCA can gradually correct the edges around the user clicks (the red regions).

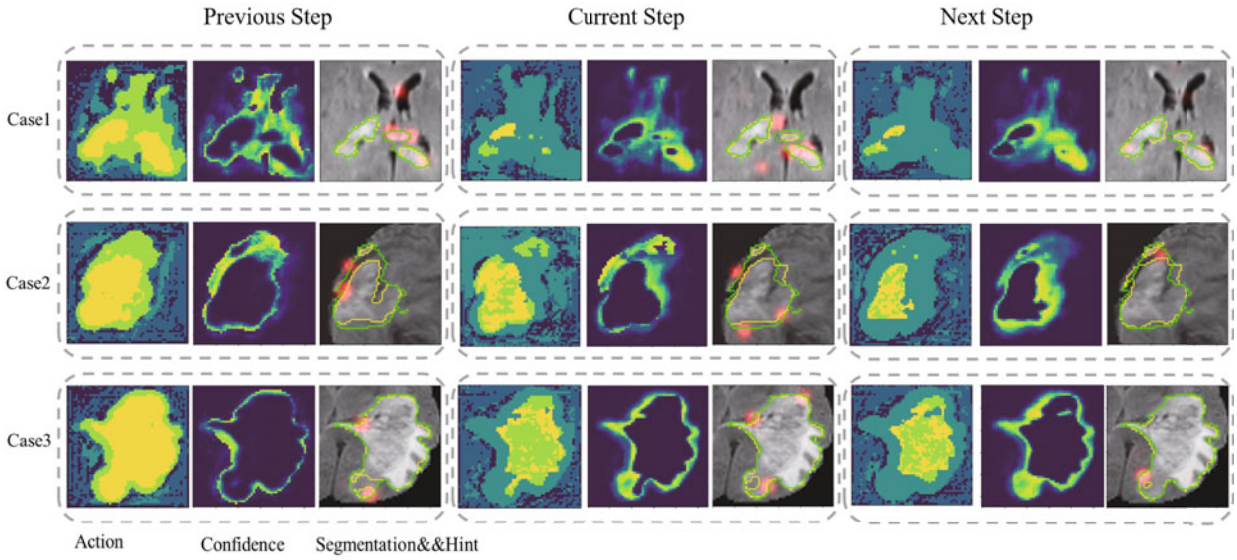



**Fig. S1 MECCA segmentation process. At each step, the first column is the action map, the second column is the confidence map, and the third column is the segmentation result with the user's hint information**

Fig. S2 shows the qualitative segmentation results of our method trained with different data sizes. It shows that the model trained with 12.5% labeled data can capture only the main region of the tumor, but

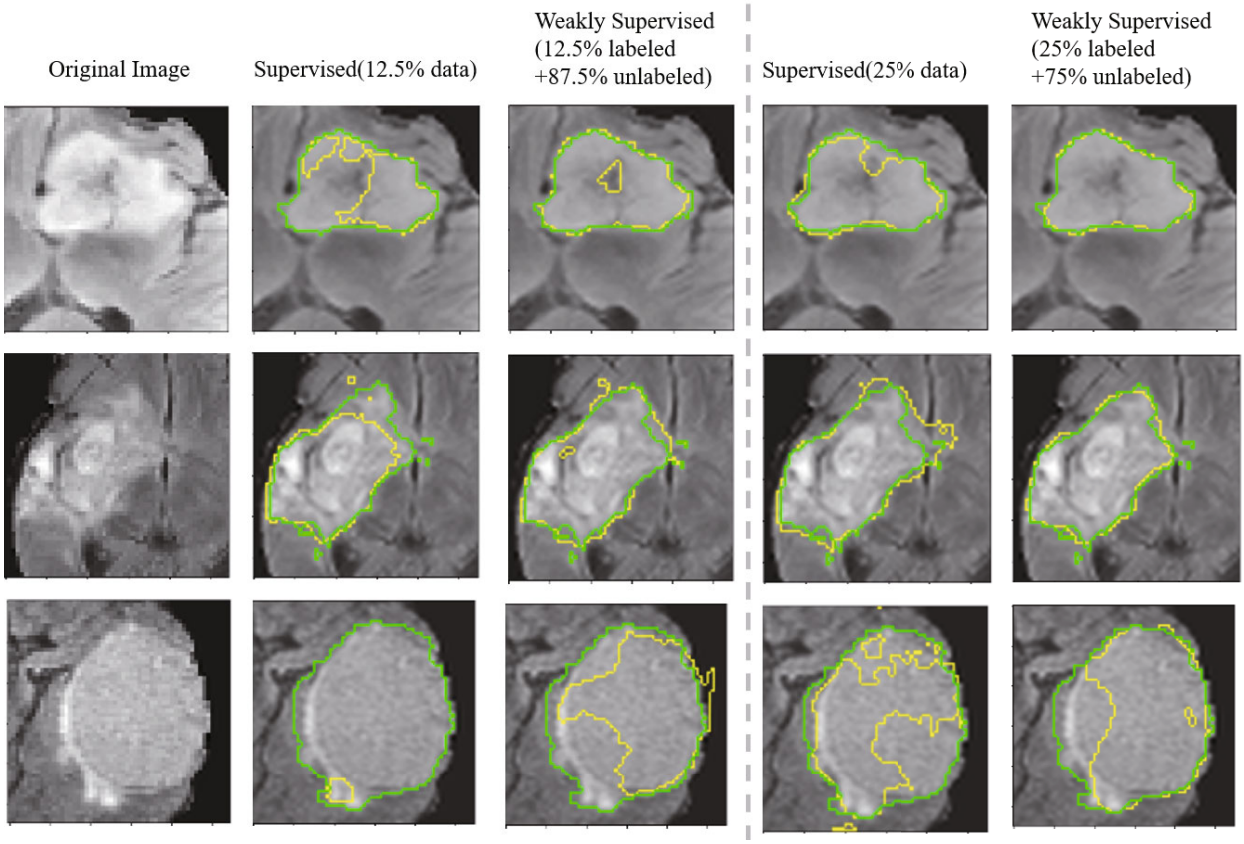



**Fig. S2 Qualitative segmentation results of MECCA for the BraTS2015 validation set. The first column shows the cropped original images. The second to fifth columns respectively show the results of supervised learning using only 12.5% labeled data, weakly supervised learning using 12.5% labeled and 87.5% unlabeled images, supervised learning using only 25% labeled data, and weakly supervised learning using 25% labeled and 75% unlabeled images. Both supervised learning and weakly supervised learning are based on our proposed method**

the model is unable to distinguish the infiltration areas of the tumor. For instance, the boundaries of tumors in the figure are difficult to distinguish because they are more similar to the healthy regions. In this case, the models trained with fewer data tend to ignore these boundary regions of the tumor, while the model trained with both labeled and unlabeled data can determine smoother and more accurate boundaries. This phenomenon occurs mainly because the distribution of the training set is not consistent with that of the validation set. The main advantage of the model trained with both labeled and unlabeled data is that it can minimize the gap between training and validation data.

## 3 Robustness of MECCA

We select the best- and worst-performing samples from the test set for more analysis. The results are shown in Table S1. The improvement of Dice on easy samples through interaction is not considerable, but the Dice on difficult samples through interaction can be twice as much. However, significantly more interactions are required for difficult samples to achieve the same segmentation accuracy as easy samples. Moreover, it is very difficult and time-consuming to accurately mark edge points in a real scene, and it is less practical to ask users to click the accurate edge points. To verify MECCA's tolerance to inaccurate edge points, we conduct a robustness study with the same settings (Table S2). Specifically, during each interaction in the training phase, random noises are added to simulated edge points. The noise range is $\pm 2$ voxels in the three directions of $x$, $y$, and $z$. In this way, the edge points used by the algorithm will be randomly selected from 64 ($4 \times 4 \times 4$) voxels within the real edge point neighborhood. This disturbance radius can cover the edge

ambiguity area in most cases. In the testing phase, we adopt the same disturbance operation, and the final results are shown in Table S2.

**Table S1 Dice of our method which varies with the number of interactions under different cases**

| Case | Dice (%) | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| Easy | 93.92 | 94.05 | 95.15 | 95.38 | 95.91 | 96.25 |
| Difficult | 39.89 | 41.45 | 47.38 | 65.78 | 77.79 | 79.37 |

**Table S2 MECCA's tolerance for inaccurate interaction point**

| Interaction type | Dice (%) | ASSD (pixel) |
|---|---|---|
| Disturbed | $88.75 \pm 9.72$ | $1.11 \pm 0.23$ |
| Non-disturbed | $90.29 \pm 5.07$ | $1.50 \pm 0.33$ |

Table S2 shows that the perturbed edge points will make the Dice value of the algorithm drop and have a more considerable variance. However, the maximum value of Dice exceeds that at the accurate edge points. The performance at perturbed edge points can even exceed that at accurate edge points in terms of the ASSD value. We believe there may be two reasons for the above phenomenon. First, MECCA uses adaptive confidence calibration to improve the information misunderstanding of the iterative algorithm, but it also makes the algorithm more "conservative." The perturbation at the edge points results in better or more "radical" segmentation effect of our method. Second, perturbing the interactive information during the training phase can enable the RL algorithm to explore the environment (medical images). State-of-the-art RL algorithms generally impose entropy constraints on the policy, enhance the randomness of the policy, and encourage exploration. Moreover, our perturbation of interactive information will indirectly affect the policy of the algorithm. In a word, MECCA has good robustness to inaccurate edge points.

## 4 Comparison of baseline responses to the same user interaction

In this section, we show how different methods respond to the same user interaction according to the same initial segmentation, especially for difficult cases. We select some images with poor performance in baselines or MECCA as the research objects. Specifically, the generation mechanism of the same user interaction is described as follows. A total of 45 hint points are randomly selected from the intersection area of the boundary of the foreground object and the error region of the initialized segmentation. Then, these 45 points are allocated to the five interaction steps of all methods according to the combination of 25, 5, 5, 5, and 5. The number of hint point used in each interaction step is the same as that of all methods during training and testing. The results are shown in Figs. S3–S5. It can be seen from the results that MECCA can use the hint point information stably in all cases. The interactive misunderstanding phenomenon arises in other methods. These methods either ignore the expert's correction information or are even adversely affected by correction information, as shown in Fig. S3.

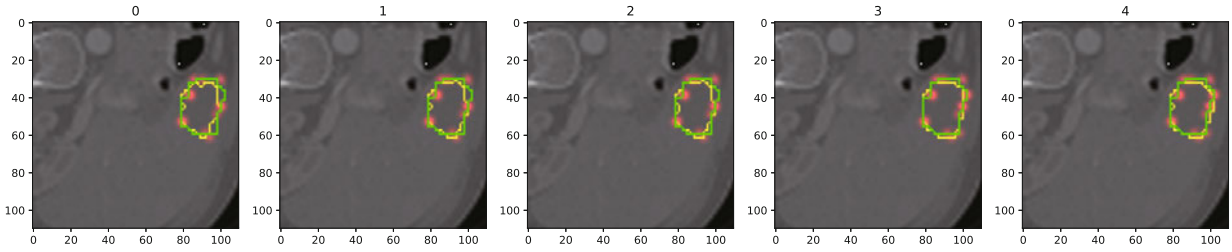
### References

Acuna D, Ling H, Kar A, et al., 2018. Efficient interactive annotation of segmentation datasets with polygon-RNN++. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.859-868. https://doi.org/10.1109/CVPR.2018.00096

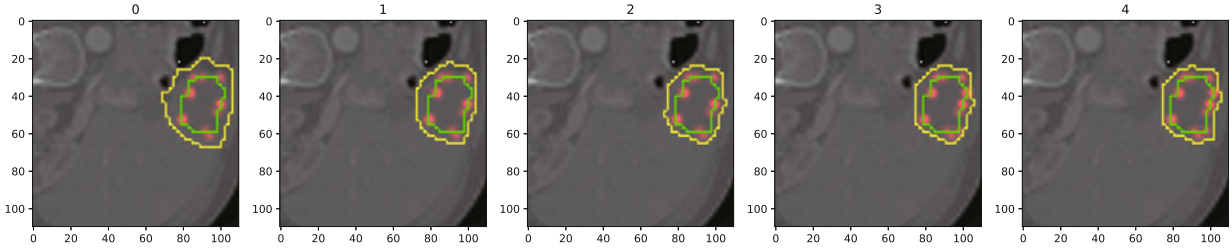
Atanov A, Ashukha A, Molchanov D, et al., 2019. Uncertainty estimation via stochastic batch normalization. 16th Int Symp on Neural Networks, p.261-269. https://doi.org/10.1007/978-3-030-22796-8_28

Boykov YY, Jolly MP, 2001. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. Proc 8th IEEE Conf on Computer Vision, p.105-112. https://doi.org/10.1109/ICCV.2001.937505

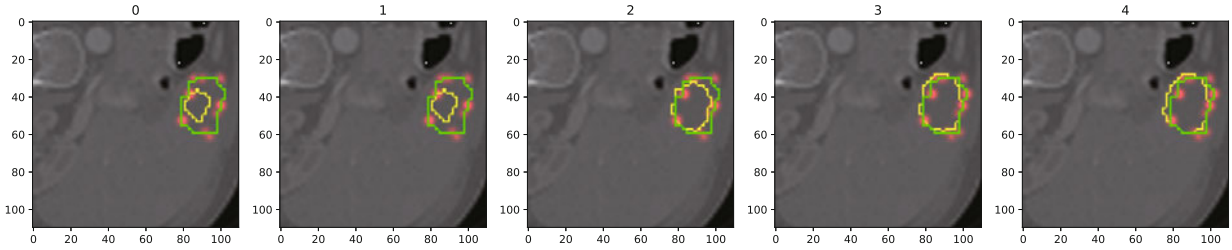
Castrejón L, Kundu K, Urtasun R, et al., 2017. Annotating object instances with a Polygon-RNN. IEEE Conf on Computer Vision and Pattern Recognition, p.4485-4493. https://doi.org/10.1109/CVPR.2017.477



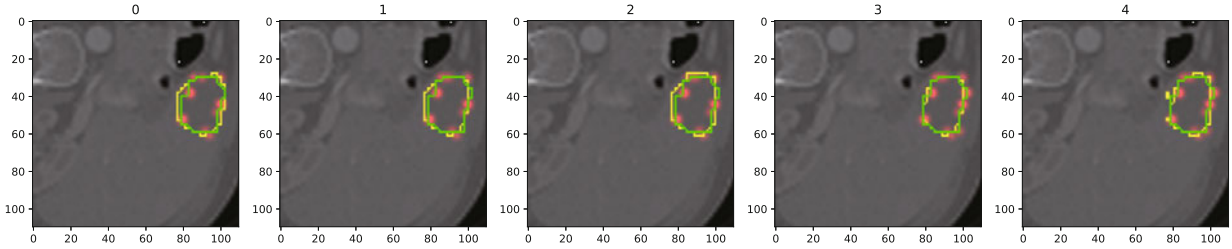
(a) DeepIGeoS



(b) InterCNN



(c) IteR-MRL



(d) MECCA

**Fig. S3 Results of different methods' responses to the same user interactions according to the same initial segmentation on the 4/10 testing instance and 7/30 channel for the Liver dataset in Medical Segmentation Decathlon**


Chen LC, Papandreou G, Kokkinos I, et al., 2018. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Patt Anal Mach Intell*, 40(4):834-848. https://doi.org/10.1109/TPAMI.2017.2699184

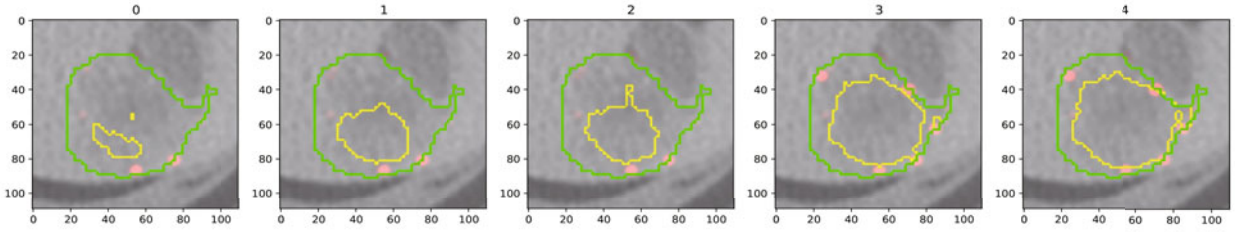
Criminisi A, Sharp T, Blake A, 2008. GeoS: geodesic image segmentation. European Conf on Computer Vision, p.99-112. https://doi.org/10.1007/978-3-540-88682-2_9

Dietterich TG, 2000. Ensemble methods in machine learning. Int Workshop on Multiple Classifier Systems, p.1-15. https://doi.org/10.1007/3-540-45014-9_1

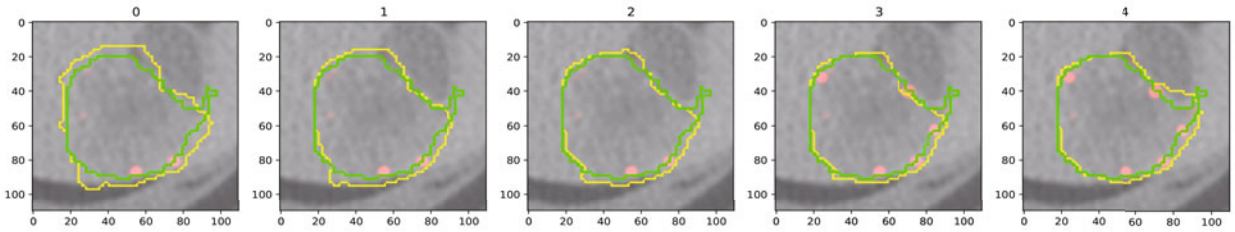
Feng RW, Zheng XS, Gao TX, et al., 2021. Interactive few-shot learning: limited supervision, better medical image segmentation. *IEEE Trans Med Imag*, 40(10):2575-2588. https://doi.org/10.1109/TMI.2021.3060551

Gal Y, Ghahramani Z, 2016. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. Proc 33rd Int Conf on Machine Learning, p.1050-1059. https://doi.org/10.5555/3045390.3045502

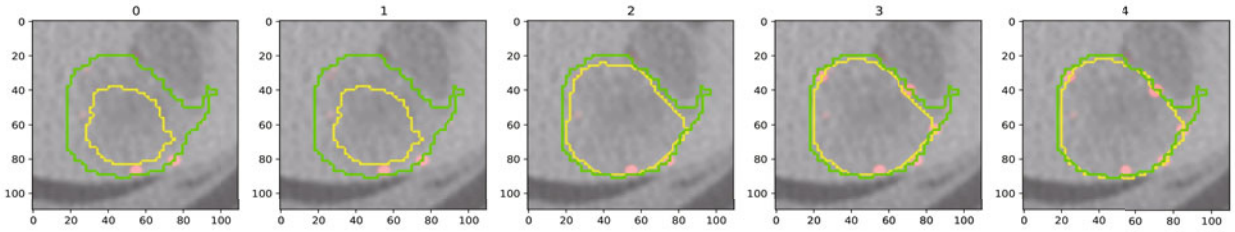
Grady L, 2006. Random walks for image segmentation. *IEEE Trans Patt Anal Mach Intell*, 28(11):1768-1783. https://doi.org/10.1109/TPAMI.2006.233



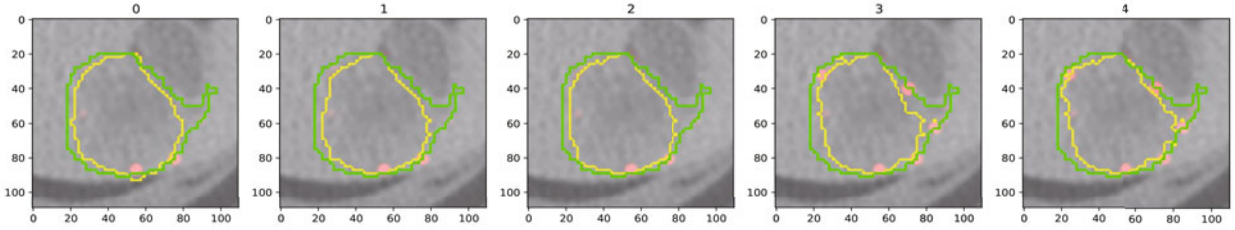
(a) DeepIGeoS



(b) InterCNN



(c) IteR-MRL



(d) MECCA

**Fig. S4 Results of different methods' responses to the same user interactions according to the same initial segmentation on the 6/10 testing instance and 18/30 channel for the Liver dataset in Medical Segmentation Decathlon**


Kamnitsas K, Ledig C, Newcombe VFJ, et al., 2017a. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*, 36:61-78. https://doi.org/10.1016/j.media.2016.10.004

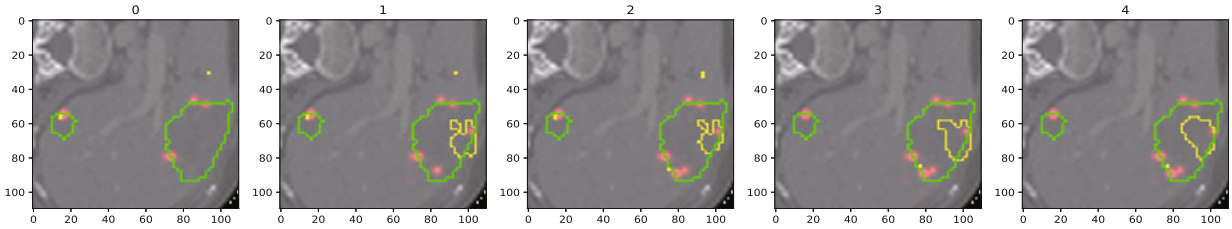
Kamnitsas K, Bai W, Ferrante E, et al., 2017b. Ensembles of multiple models and architectures for robust brain tumour segmentation. Int MICCAI Brain Lesion Workshop, p.450-462. https://doi.org/10.1007/978-3-319-75238-9_38

Lakshminarayanan B, Pritzel A, Blundell C, 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. Proc 31st Int Conf on Neural Information Processing System, p.6405-6416. https://dl.acm.org/doi/10.5555/3295222.3295387

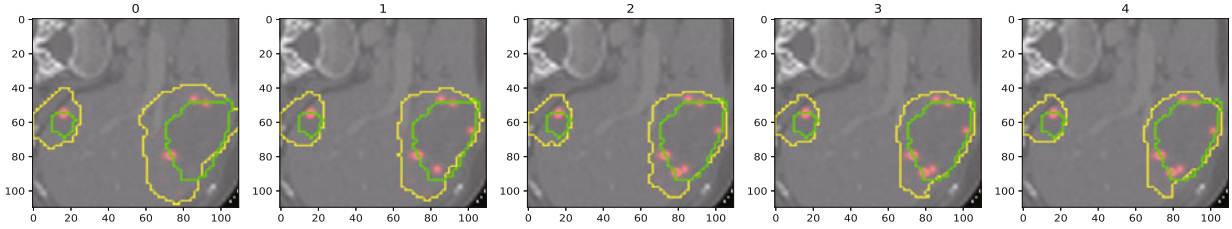
Lee KM, Song G, 2018. SeedNet: automatic seed generation with deep reinforcement learning for robust interactive segmentation. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.1760-1768. https://doi.org/10.1109/CVPR.2018.00189

Li W, Wang G, Fidon L, et al., 2017. On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task. Int Conf on Information Processing in Medical Imaging, p.348-360. https://doi.org/10.1007/978-3-319-59050-9_28

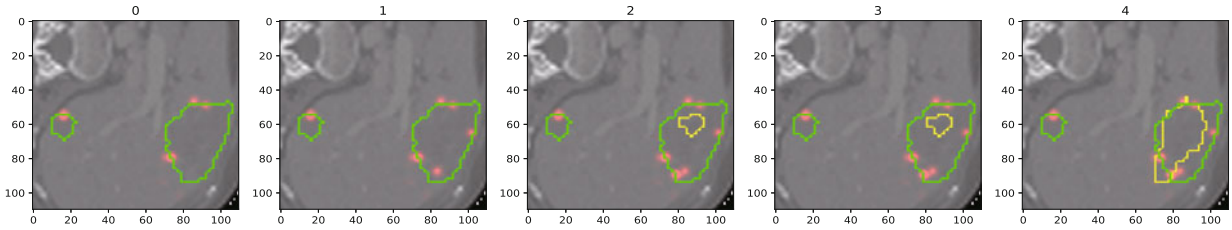
Liao X, Li WH, Xu QS, et al., 2020. Iteratively-refined interactive 3D medical image segmentation with multi-agent reinforcement learning. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.9394-9402. https://doi.org/10.1109/CVPR42600.2020.00941



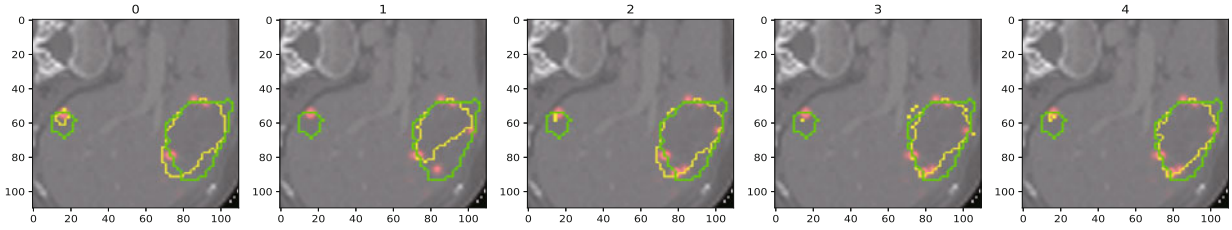
(a) DeepIGeoS



(b) InterCNN



(c) IteR-MRL



(d) MECCA

**Fig. S5 Results of different methods' responses to the same user interactions according to the same initial segmentation on the 7/10 testing instance and 10/30 channel for the Liver dataset in Medical Segmentation Decathlon**


Lin D, Dai JF, Jia JY, et al., 2016. ScribbleSup: scribble-supervised convolutional networks for semantic segmentation. IEEE Conf on Computer Vision and Pattern Recognition, p.3159-3167. https://doi.org/10.1109/CVPR.2016.344

Louizos C, Welling M, 2017. Multiplicative normalizing flows for variational Bayesian neural networks. Proc 34$^{th}$ Int Conf on Machine Learning, p.2218-2227.

Ma CF, Xu QS, Wang XF, et al., 2021. Boundary-aware supervoxel-level iteratively refined interactive 3D image segmentation with multi-agent reinforcement learning. *IEEE Trans Med Imag*, 40(10):2563-2574. https://doi.org/10.1109/TMI.2020.3048477

MacKay DJC, 1992. A practical Bayesian framework for backpropagation networks. *Neur Comput*, 4(3):448-472. https://doi.org/10.1162/neco.1992.4.3.448

Mehrtash A, Ghafoorian M, Pernelle G, et al., 2018. Automatic needle segmentation and localization in MRI with 3-D convolutional neural networks: application to MRI-targeted prostate biopsy. *IEEE Trans Med Imag*, 38(4):1026-1036. https://doi.org/10.1109/TMI.2018.2876796

Mehrtash A, Wells WM, Tempany CM, et al., 2020. Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE Trans Med Imag*, 39(12):3868-3878. https://doi.org/10.1109/TMI.2020.3006437

Milletari F, Navab N, Ahmadi SA, 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation. 4$^{th}$ Int Conf on 3D Vision, p.565-571. https://doi.org/10.1109/3DV.2016.79

Neal RM, 1995. Bayesian Learning for Neural Networks. University of Toronto, Canada.

Rajchl M, Lee MCH, Oktay O, et al., 2017. DeepCut: object segmentation from bounding box annotations using convolutional neural networks. *IEEE Trans Med Imag*, 36(2):674-683. https://doi.org/10.1109/TMI.2016.2621185

Ronneberger O, Fischer P, Brox T, 2015. U-Net: convolutional networks for biomedical image segmentation. Proc 18$^{\text{th}}$ Int Conf on Medical Image Computing and Computer-Assisted Intervention, p.234-241. https://doi.org/10.1007/978-3-319-24574-4_28

Rother C, Kolmogorov V, Blake A, 2004. "GrabCut": interactive foreground extraction using iterated graph cuts. *ACM Trans Graph*, 23(3):309-314. https://doi.org/10.1145/1015706.1015720

Shelhamer E, Long J, Darrell T, 2017. Fully convolutional networks for semantic segmentation. *IEEE Trans Patt Anal Mach Intell*, 39(4):640-651. https://doi.org/10.1109/TPAMI.2016.2572683

Wang GT, Zuluaga MA, Li WQ, et al., 2019. DeepIGeoS: a deep interactive geodesic framework for medical image segmentation. *IEEE Trans Patt Anal Mach Intell*, 41(7):1559-1572. https://doi.org/10.1109/TPAMI.2018.2840695

Xu N, Price B, Cohen S, et al., 2016. Deep interactive object selection. IEEE Conf on Computer Vision and Pattern Recognition, p.373-381. https://doi.org/10.1109/CVPR.2016.47

Zhang SY, Liew JH, Wei YC, et al., 2020. Interactive object segmentation with inside-outside guidance. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.12231-12241. https://doi.org/10.1109/CVPR42600.2020.01225