PE-MED: Prompt Enhancement for Interactive Medical Image Segmentation

Ao Chang^{1,2,3,4*}, Xing Tao^{1,2,3*}, Xin Yang^{1,2,3}, Yuhao Huang^{1,2,3}, Xinrui Zhou^{1,2,3}, Jiajun Zeng^{1,2,3}, Ruobing Huang^{1,2,3}, and Dong Ni^{1,2,3}(\boxtimes)

¹National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Health Science Center, Shenzhen University, China nidong@szu.edu.cn

 $^2\mathrm{Medical}$ Ultrasound Image Computing (MUSIC) Lab, Shenzhen University, China

³Marshall Laboratory of Biomedical Engineering, Shenzhen University, China ⁴Shenzhen RayShape Medical Technology Co., Ltd, China

Abstract. Interactive medical image segmentation refers to the accurate segmentation of the target of interest through interaction (e.g., click) between the user and the image. It has been widely studied in recent years as it is less dependent on abundant annotated data and more flexible than fully automated segmentation. However, current studies have not fully explored user-provided prompt information (e.g., points), including the knowledge mined in one interaction, and the relationship between multiple interactions. Thus, in this paper, we introduce a novel framework equipped with prompt enhancement, called PE-MED, for interactive medical image segmentation. First, we introduce a Self-Loop strategy to generate warm initial segmentation results based on the first prompt. It can prevent the highly unfavorable scenarios, such as encountering a blank mask as the initial input after the first interaction. Second, we propose a novel Prompt Attention Learning Module (PALM) to mine useful prompt information in one interaction, enhancing the responsiveness of the network to user clicks. Last, we build a Time Series Information Propagation (TSIP) mechanism to extract the temporal relationships between multiple interactions and increase the model stability. Comparative experiments with other state-of-the-art (SOTA) medical image segmentation algorithms show that our method exhibits better segmentation accuracy and stability.

Keywords: Interactive Segmentation · Prompt Learning.

1 Introduction

Medical image segmentation is a pivotal aspect of research in medical image analysis, intended for the extraction of specific targets or regions in medical images for further analysis and diagnosis [22]. Traditional segmentation methods heavily rely on image processing techniques and machine learning algorithms,

^{*} Ao Chang and Xing Tao contribute equally to this work.

2 Chang et al.

which can be computationally intensive, time-consuming, and require a high level of expertise. Deep learning-based methods have achieved state-of-the-art (SOTA) performance in automatic segmentation of medical images [17]. However, most current automatic methods lack the learning of informative prompts, resulting in inaccurate and inflexible segmentation.

Interactive segmentation methods offer a promising solution to these challenges, utilizing limited user guidance to extract the target object and combining this with image features to yield finely segmented results [21]. While user interaction leads to more precise segmentation results, the interaction process should be efficient and time-saving to reduce the burden on the user.

Traditional interactive methods for image segmentation use low-level features like edge information or color distribution, such as GraphCuts and Random Walk [9,2,6,7]. These methods often require multiple user interactions and are time-consuming to produce satisfactory results as low-level features may not always distinguish the desired object from the background. To reduce user interactions and improve segmentation accuracy, machine learning techniques are leveraged. For instance, GrabCut [16] uses a Gaussian mixture model to estimate foreground and background distributions. The initial results can be obtained via a user-provided bounding box, and refined through additional interactions.

Recently, **deep learning-based methods** have achieved SOTA performance in medical image segmentation [5,4,8,11], thanks to the neural networks for automatically capturing high-level semantic features [17]. Thus, deep models have been proposed to integrate with interactive methods for medical image segmentation. There are two main streams of current approaches, as introduced below:

-The first type of method ignores prompt learning during training [3,20,19,14]. These methods require pre-training a semantic segmentation network, followed by fine-tuning the predicted mask through user interactions. However, they may not be suitable for multi-class segmentation tasks that require accurate delineation of different targets. Besides, they cannot ensure the quality of feature extraction when dealing with unfamiliar data patterns or categories, causing a performance drop.

-The second type of method involves prompt learning during training. Prompts are leveraged to guide the learning of deep models in these methods, including iSegFormer [13], Segment Anything Model (SAM) [12,10], etc. During testing, users need to interactively click on the foreground or background to achieve an accurate target segmentation. Such prompt learning techniques have the potential to make interactive segmentation more flexible, accurate, and general to complex scenarios. However, it is still challenging to deeply mine sparse prompts information and improve network response to user-provided prompts.

In this study, we propose a novel interactive approach with prompt enhancement to improve medical image segmentation performance, named PE-MED. Our contribution is three-fold. First, we employ a simple yet effective self-loop method to address the issue of insufficient information during the first interaction. Second, we propose a novel Prompt Attention Learning Module (PALM) that explores the relationship between user interactions and image features to



Fig. 1: Overview of our proposed interactive segmentation framework. The green and red dots in the Click Prompt represent positive and negative prompt points, respectively. Sequential processing through four Transformer blocks (block1-4) enables the generation of multi-level features at $\{1/2^n, n = 2, 3, 4, 5\}$ of the original image resolution. The fusion module is composed of a convolutional layer and a normalization layer. The decoder module is a simple multilayer perceptron.

extract essential interaction details, enhancing the network's response to user input. Third, we introduce a Time Series Information Propagation (TSIP) mechanism to model the continuity between multiple interactions for improving stability. Extensive experiments validated that, compared with the SOTA methods, our PE-MED can achieve accurate results with less user interaction.

2 Methodology

Fig. 1 shows the schematic view of our proposed method. We propose the PE-MED, a novel iterative refinement framework equipped with prompt-enhanced modules, for medical image segmentation. PE-MED consists of two stages: 1) The initial stage (see Time = 1) serves as the foundation for the subsequent clicks, 2) The main stage represents the subsequent clicks after the initial one. In the first stage, we first introduce a Self-Loop strategy to obtain a good initialized mask. Then, the original image, previous mask and prompts (positive and negative points) are taken as inputs, and transmitted to the PALM and TSIP for prompt enhancement. Last, the network will output the refined segmentation iteratively.

2.1 Self-Loop Strategy for Warm Start

For interactive segmentation methods, the initial segmentation generated by the user's first click plays a fundamental role in the subsequent interactions. Most





Fig. 2: Overview of our proposed PALM.

of the current method lacks the optimization of the first interaction, resulting in poor initialization with inadequate prompt information. This will have a negative impact on the subsequent network module optimization and even segmentation failure. To address this issue, inspired by [23,24], we introduce a Self-Loop strategy to obtain a warm start for iterative segmentation (left-upper yellow block in Fig. 1). Specifically, after the user's first click, fully empty masks are sent to the network, obtaining the rough segmentation (M0). Then, the coarse prediction will enter the loop, and output the information-enhanced mask (M1) without any interaction. Compared to M0, M1 contains richer prompt information, and makes the learning of subsequent modules easier.

2.2 PALM for Prompt Feature Enhancement

Extracting rich information from sparse user hints is a challenging task. In our study, we developed PALM to effectively leverage and enhance the prompt information. As shown in Fig. 2, PALM consists of two modules: PALM-I (see Fig. 2(a)) and PALM-O (see Fig. 2(b)). Specifically, PALM-I primarily focuses on augmenting the intrinsic features of prompts, whereas PALM-O enhances the interplay between prompts and images.

PALM takes four inputs, including the image I_{image} , positive and negative prompts (I_{pos} and I_{neg}), and previous mask (I_{prev}). After the users input is converted to a disk map following a click, I_{pos} and I_{neg} are encoded and utilized in the PALM module. As shown in Fig. 2(a), in the PALM-I part, we combine I_{pos} , I_{neg} , and I_{prev} to form the concatenated feature map M_{pos} , M_{global} , and M_{neg} using patch embeddings. The Cross Attention Module (Fig. 2(c)) is then applied to enhance these feature maps, resulting in augmented features $\hat{M}_{pos} = \theta_c(M_{neg}, M_{pos}, M_{pos})$ and $\hat{M}_{neg} = \theta_c(M_{pos}, M_{neg}, M_{neg})$, with $\theta_c(Q, K, V) = Softmax(QK^T/d_k)V$. Simultaneously, the Self Attention Module improves M_{global} , yielding the global feature map \hat{M}_{global} :

$$\hat{M}_{qlobal} = \theta_s(M_{qlobal}) = \theta_c(M_{qlobal}, M_{qlobal}, M_{qlobal}), \tag{1}$$

After obtaining enriched interaction and global information, we compute the prompt feature E_p by summing the features resulting from cross-attentive mechanisms using \hat{M}_{pos} and \hat{M}_{neg} as query vectors, E_p can be defined as:

$$E_p = \theta_c(\hat{M}_{pos}, \hat{M}_{global}, \hat{M}_{global}) + \theta_c(\hat{M}_{neg}, \hat{M}_{global}, \hat{M}_{global}),$$
(2)

Besides, we propose the PALM-O to bridge the information gap between prompt and image features. It takes E_p and the image features F_I (after Transformer Block 1) as input. Then, with the normalized layer (*norm*), the enhanced mixture feature EP_{infor} can be illustrated as:

$$EP_{infor} = F_I + norm(\theta_c(F_I, E_p, E_p)) + E_p + norm(\theta_c(E_p, F_I, F_I)).$$
(3)

2.3 TSIP Mechanism for Stability Enhancement among Prompts

Most of the existing interactive segmentation algorithms often ignore the relationship between consecutive interactions. It may lead to poor segmentation stability, especially when dealing with multiple interactions. To overcome this limitation, we propose to integrate TSIP mechanism into the framework to enhance the stability among multiple prompts.

The TSIP mechanism enables the extraction of dynamic and continuous interaction information, with the entire network functioning as a cohesive unit to convey temporal information. Specifically, the previous network output serves as a candidate memory for the current moment using a simple multilayer perceptron (MLP), while the current output serves as a candidate memory for the next moment. This mechanism is mathematically represented by:

$$O_t = F(I_{input}) + Sigmoid(\theta(O_{t-1})), \tag{4}$$

where F refers to PE-MED without TSIP, θ refers a MLP. I_{input} denotes the input image, click prompts and previous mask, O_{t-1} and O_t represent the output of the previous and current network, respectively.

3 Experimental Results

Materials and Implementation Details. We validate our proposed framework on the public multi-organ dataset (named **Synapse** [1]) proposed in the 2015 MICCAI Multi-Atlas Abdomen Labeling Challenge. **Synapse** includes 30 cases with a total of 3779 axial abdomen 2D images where each CT volume

Mathada	$\mathrm{DSC}(\%)\uparrow$								
Methous	AVG	Aorta	Gallbladder	Kidney(L)	Kidney(R)	Liver	Pancreas	Spleen	Stomach
U-Net [15]	76.85	89.07	69.72	77.77	68.60	93.43	53.98	86.67	75.58
TransUnet [5]	77.48	87.23	63.13	81.87	77.02	94.08	55.86	85.08	75.62
Swin-Unet [4]	79.13	85.47	66.53	83.28	79.61	94.29	56.58	90.66	76.60
HiFormer [8]	80.69	87.03	68.61	84.23	78.37	94.07	60.77	90.44	82.03
GrabCut [16]	42.46	27.87	0.00	67.46	59.87	72.39	3.42	41.08	33.85
iSegFormer(2D)-p1 [13]	74.62	82.88	58.08	75.05	72.23	78.46	62.89	79.13	64.67
iSegFormer(2D)-p3 [13]	86.43	89.07	70.53	89.60	86.92	89.73	77.25	88.71	83.22
iSegFormer(2D)-p5 [13]	89.74	91.43	79.14	91.56	90.61	92.20	82.42	90.57	88.61
SAM-p1 [12]	75.33	88.98	52.74	87.05	85.98	72.41	41.19	78.06	64.44
SAM-p3 [12]	78.61	87.93	52.22	86.76	86.04	81.59	48.27	79.29	71.21
SAM-p5 [12]	79.64	87.37	53.26	86.71	85.96	84.26	51.18	81.23	74.53
Ours-p1	80.76	82.67	57.05	86.46	85.85	84.31	67.97	87.96	73.86
Ours-p3	90.51	89.43	77.99	91.57	91.75	94.53	83.45	94.54	90.66
Ours-p5	92.76	91.68	84.81	92.88	92.88	96.00	87.52	95.43	94.15

Table 1: Method comparison on **Synapse** dataset. $-p^*$ denotes the number of point prompts.

Table 2: Method comparison on **OL12** dataset, evaluated by DSC.

Method	Point:1	Point:2	Point:3	Point:5	Point:10	NoC@85	NoC@90
iSegFormer(2D) [13]	47.03(20.95)	58.27(17.66)	65.87(14.37)	76.17(9.59)	85.83(4.89)	7.51(2.19)	9.56(1.13)
SAM [12]	51.07(34.85)	56.90(32.96)	58.49(32.79)	61.31(32.23)	57.28(33.31)	6.83(4.12)	7.67(3.80)
Ours	86.49(12.25)	91.35(7.41)	92.91(5.65)	94.24(3.97)	95.18(2.97)	1.65(1.45)	2.39(2.37)

involves 85-198 slices. The dataset is divided randomly into 24 cases for training, and 6 cases for testing. Eight organs are annotated by experts, including Aorta, Gallbladder, Left Kidney, Right Kidney, Liver, Pancreas, Spleen and Stomach. To further test the performance of our proposed PE-MED, we built another dataset including different modalities (CT&MRI) and 12 common Organs/Lesions, named **OL12**. We randomly split **OL12** into 5050, 2041 and 4912 images for training, validation and testing at the case level.

We implemented our framework using PyTorch using one NVIDIA 3090 GPU with 24 GB of memory. The input image sizes are 224×224 and 256×256 in **Synapse** and **OL12** datasets, respectively. The models are trained for 100 epochs using a batch size of 128 and 64 for each dataset, using the normalized focal loss function [18]. We optimized our model using the Adam optimizer, starting with a learning rate of 5×10^{-3} , and reducing the learning rate by a rate factor of 0.6 every 20 epochs. The dice score coefficient (DSC) was adopted to quantitatively evaluate segmentation performance. We also evaluated the methods using *number of clicks* (NoC@†) metric, to measure the number of interactions required to achieve a predefined DSC (†).

Quantitative and Qualitative Analysis. For Synapse dataset, we compared the proposed PE-MED with the fully supervised image segmentation methods, including TransUnet [5], Swin-Unet [4], and HiFormer [8]. Moreover, we also evaluated the performance of PE-MED with interactive segmentation methods such as GrabCut [16], iSegFormer (2D version) [13], and SAM [12].

Methods	$\mathrm{DSC}(\%)\uparrow$								
	AVG	Aorta	Gallbladder	$\operatorname{Kidney}(L)$	Kidney(R)	Liver	Pancreas	Spleen	Stomach
Baseline	89.11	90.25	74.17	89.14	88.40	93.71	78.03	92.00	91.25
Baseline-SL	91.70	93.12	73.02	91.77	91.92	95.13	83.79	94.45	93.11
Baseline-I	91.84	93.79	73.00	93.07	92.10	95.40	82.69	94.93	91.70
Baseline-O	91.22	92.29	71.63	91.84	91.91	95.06	83.12	93.62	92.56
Baseline-IO	92.24	91.15	83.00	92.15	92.54	95.46	87.35	94.85	93.84
Baseline-T	91.01	91.90	72.62	91.17	91.15	94.79	82.65	93.73	93.18
Ours	92.76	91.68	84.81	92.88	92.88	96.00	87.52	95.43	94.15

Table 3: Ablation study under the setting of five point prompts.

For **OL12** dataset, we only compared PE-MED with two SOTA interactive segmentation methods, i.e., iSegFormer [13] and SAM [12]. The ablation study was conducted on **Synapse** dataset, by comparing the PE-MED with different components, including Self-Loop (Baseline-SL), PALM-I (BaseLine-I), PALM-O (BaseLine-O), PALM (BaseLine-IO), and TSIP (BaseLine-T). Quantitative results are presented in Table 1-3, with best results shown in Blue.

For Table 1, it can be observed that our method achieves an average DSC of 80.76% with a single point prompt, slightly higher than the current SOTA fully automated segmentation algorithm, i.e., HiFormer (80.69%). With more prompts (-p5), our method achieve a DSC of 92.76%, outperforming all the reported fully automatic and interactive segmentation methods. It is also notable that our proposed method shows the best performance compared to interactive methods under the same number of prompts. Specifically, the DSC are 6.14%/5.43% (p1), 4.08%/11.90% (p3), 3.02%/13.15% (p5) higher than iSegFormer and SAM, respectively. Results on **OL12** dataset are reported in Table 2. It can be seen that iSegFormer and SAM struggle to obtain satisfactory DSC even with 5 or 10 point prompts (<86%). However, our method achieved the DSC of 86.49% even with one user interaction, and can outperform iSegFormer with 5&10 prompts by about 20%&10% DSC, respectively. The last two columns (NoC@[†]) reveals that our method only require average 1.65 and 2.39 to reach a DSC of 85% and 90%, respectively. While the other two methods require more user interaction to satisfy the corresponding requirements.

Results of ablation study can be found in Table 3. Experiments validated that by adding our proposed modules separately on the basis of baseline, the DSC performance can be improved. Simultaneously, integrating all the modules (**Ours**) can further enhance the performance of the network. It can also be observed that PE-MED achieves the highest DSC for six out of the eight organs, with only a slight deviation from the best results for the remaining two organs. Specifically, for the most challenging organ, i.e., *Gallbladder* with DSC of 74.17% in Baseline, our proposed PE-MED can increase it by 10.64%.

Fig. 3 depicts the visualization results of different methods (rows 1-2), demonstrating our method as the closest to the Ground truth. Row 3 of Fig. 3 presents the visualization results for various numbers of interactions. Furthermore, in-



Fig. 3: Segmentation performance on the **Synapse** and **OL12** dataset. The red rectangles highlight organ regions where the superiority of *Ours* is evident.



Fig. 4: DSC variation tendency curve on Synapse dataset.

creasing the number of interactions in our methodology leads to a progressive improvement in performance. We also test the number of clicks and the results are shown in Fig. 4. In the left figure, we observe the rising trend of DSC for different organs and the average value, from clicks 1 to 10. Specifically, see the *Average* curve, clicks 1-5 gain a higher improvement than clicks 6-10. Thus, we consider 5 clicks a suitable choice, since it is a trade-off between the NoC and DSC performance. In the right figure, compared with *Ours* and the ablation results, the contribution of our proposed prompt enhancement techniques can be validated. Besides, it can be found that *Ours* outperforms *iSegFormer* at every click (red and blue curves), which further illustrates the power of our method.

4 Conclusion

In this work, we introduce an interactive framework for medical image segmentation, named PE-MED. Via the click user prompts, PE-MED can progressive optimize the segmentation results. We proposed three techniques for enhancing the prompt information, including 1) Self-Loop strategy for providing warm initialization at the first interaction, 2) PALM for feature aggregation at one click, and 3) TSIP for temporal modeling among multiple interactions. Extensive experiments on two large datasets validate that PE-MED is general and efficient, achieving the best DSC results among all the strong competitors. In the future, we will extend a 3D version PE-MED to directly handle the volumetric data.

Acknowledgement. This work was supported by the grant from National Natural Science Foundation of China (Nos.62171290, 62101343, 62101342), Shenzhen-Hong Kong Joint Research Program (No.SGDX20201103095613036), Shenzhen Science and Technology Innovations Committee (No.20200812143441001), and Guangdong Basic and Applied Basic Research Foundation (No.2023A1515012960).

References

- MICCAI 2015 Multi-Atlas Abdomen Labeling Challenge. Synapse multi-organ segmentation dataset. https://www.synapse.org/#!Synapse:syn3193805/wiki/ 217789 (2015), Accessed: 2023-07-10
- Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In: Proceedings eighth IEEE international conference on computer vision. ICCV 2001. vol. 1, pp. 105–112. IEEE (2001)
- Bredell, G., Tanner, C., Konukoglu, E.: Iterative interaction training for segmentation editing networks. In: Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9. pp. 363–370. Springer (2018)
- Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: Proceedings of the European Conference on Computer Vision Workshops(ECCVW) (2022)
- Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
- Criminisi, A., Sharp, T., Blake, A.: Geos: Geodesic image segmentation. In: Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I 10. pp. 99–112. Springer (2008)
- Grady, L.: Random walks for image segmentation. IEEE transactions on pattern analysis and machine intelligence 28(11), 1768–1783 (2006)
- Heidari, M., Kazerouni, A., Soltany, M., Azad, R., Aghdam, E.K., Cohen-Adad, J., Merhof, D.: Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6202–6212 (2023)
- Hu, Y., Soltoggio, A., Lock, R., Carter, S.: A fully convolutional two-stream fusion network for interactive image segmentation. Neural Networks 109, 31–42 (2019)
- Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., et al.: Segment anything model for medical images? arXiv preprint arXiv:2304.14660 (2023)
- Huang, Y., Yang, X., Zou, Y., Chen, C., Wang, J., Dou, H., Ravikumar, N., Frangi, A.F., Zhou, J., Ni, D.: Flip learning: Erase to segment. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Con-

10 Chang et al.

ference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 493–502. Springer (2021)

- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)
- Liu, Q., Xu, Z., Jiao, Y., Niethammer, M.: isegformer: interactive segmentation via transformers with application to 3d knee mr images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 464–474. Springer (2022)
- Liu, W., Ma, C., Yang, Y., Xie, W., Zhang, Y.: Transforming the interactive segmentation for medical imaging. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 704–713. Springer (2022)
- Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
- Rother, C., Kolmogorov, V., Blake, A.: "GrabCut" interactive foreground extraction using iterated graph cuts. ACM transactions on graphics (TOG) 23(3), 309– 314 (2004)
- 17. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual review of biomedical engineering **19**, 221–248 (2017)
- Sofiiuk, K., Petrov, I.A., Konushin, A.: Reviving iterative training with mask guidance for interactive segmentation. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 3141–3145. IEEE (2022)
- Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., et al.: Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE transactions on medical imaging 37(7), 1562–1573 (2018)
- 20. Wang, G., Zuluaga, M.A., Li, W., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., et al.: Deepigeos: a deep interactive geodesic framework for medical image segmentation. IEEE transactions on pattern analysis and machine intelligence 41(7), 1559–1572 (2018)
- Zhou, L., Wang, Y., Chen, D., Zeng, W., Zhang, Q., Yang, J.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis 63, 101693 (2020)
- 22. Zhou, L., Wang, S., Zhang, Q., Shen, D.: A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE 109(5), 820–838 (2021)
- Zhou, W., Tao, X., Wei, Z., Lin, L.: Automatic segmentation of 3d prostate mr images with iterative localization refinement. Digital Signal Processing 98, 102649 (2020)
- Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., Yap, P.T., Shen, D.: Multitask learning for segmentation and classification of tumors in 3d automated breast ultrasound images. Medical Image Analysis 70, 101918 (2021)