

# Learning Consistency- and Discrepancy-Context for 2D Organ Segmentation

Lei Li<sup>1\*</sup>, Sheng Lian<sup>2,3,4\*</sup>, Zhiming Luo<sup>2(✉)</sup>, Shaozi Li<sup>2(✉)</sup>,  
Beizhan Wang<sup>1</sup>, and Shuo Li<sup>3,4</sup>

<sup>1</sup> Department of Software Engineering, Xiamen University, Xiamen, Fujian, China

<sup>2</sup> Department of Artificial Intelligence, Xiamen University, Xiamen, Fujian, China

<sup>3</sup> Digital Image Group (DIG), London, ON, Canada

<sup>4</sup> School of Biomedical Engineering, Western University, London, ON, Canada

✉Correspondences: {zhiming.luo, szlig}@xmu.edu.cn

**Abstract.** Recently, CNN-based methods lead tremendous progress in segmenting abdominal organs (*e.g.*, kidney, liver, and pancreas) and anomaly tumors in CT scans. Although 3D CNN-based methods can significantly improve accuracy by using 3D volume as input, they need more computational cost and may not satisfy the efficiency requirement for many practical applications. In this study, we mainly aim at improving the 2D segmentation by leveraging the consistency- and- discrepancy-context information from adjacent slices. Specifically, the consistency context mainly considers that the prediction variance of two adjacent slices needs to follow the variance in the ground truth. The discrepancy-context assumes the label difference of adjacent slices usually occurs in the edge area of organs. To fully utilize the above context information, we further devise a two-stage 2.5D segmentation framework based on the U-Net that takes three adjacent slices as input. In the first stage, we encourage the predictions of the three slices following the consistency context. In the second stage, we refine the segmentation result by adopting the prediction discrepancy area of adjacent slices as an extra input. Experimental results on several challenging datasets demonstrate the effectiveness of our proposed methods. Moreover, the adjacent-slice context information considered in this study can be effortlessly incorporated into other segmentation frameworks without extra testing overhead.

**Keywords:** Organ segmentation · Consistency context · Discrepancy context.

## 1 Introduction

Segmenting abdominal organs (*e.g.*, kidney, liver, and pancreas) and anomaly tumors in CT scans is a critical prerequisite for many clinical scenarios, including computer-aided diagnosis, radiotherapy planning, and computer-assisted surgery overtreatment [11,15,20,8,22,12,17]. Apart from artifact and low contrast in the

---

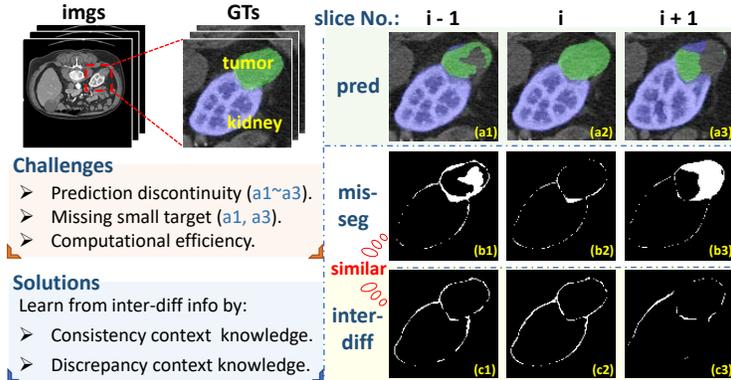
\* L. Li and S. Lian — Equal contribution

original CT scan, this segmentation task still needs to deal with challenges from different aspects, such as complex contextual associations, class imbalance, and intricate tissue shapes/sizes. Due to these challenges, a vanilla U-Net [16] model usually suffers the issue of discontinuous prediction, missing the small target, as shown in Fig. 1.

The development of CNN-based methods has significantly boosted the performance of this task. Existing CNN-based methods can be categorized as 2D-based [4,19,23] and 3D-based [10,6,5]. The 2D-based methods are more efficient since they process each slice independently while not utilizing the context information from the adjacent slice, resulting in relatively low performance. On the other hand, the 3D-based approaches are more accurate by using 3D volume as input. Meanwhile, they need more computational cost and may not satisfy the efficiency requirements for many practical applications. Therefore, it raises a question: *Is it possible to utilize the contextual information for improving segmentation performance while still maintaining the efficiency as in 2D-based models?* The work in [7] has tried to tackle this problem by training the segmentation model in a 2.5D manner, in which several adjacent slices are used as input. However, [7] simply uses more input slices and fails to exploit the inter-slice knowledge.

In Fig. 1, we illustrate the predictions from a vanilla U-Net. Moreover, we also compute the mis-segmented regions and the inter-slice difference (*inter-diff*) regions of the ground truth. Notice that, the *inter-diff* in the ground truth of two adjacent slices (formulated in Eq. 1) is roughly located in the boundary regions of the target tissues. By observing the mis-segmented regions, we can find that a large variance occurs between the predictions of adjacent slices, but the *inter-diff* of two adjacent slices are very similar. Besides, there are also some errors that happen around the boundary of tissues. Therefore, we could improve the segmentation accuracy by reducing the prediction variance in adjacent slices and leveraging the information of *inter-diff* regions.

In this paper, for accurate segmentation of abdominal organs in CT scans, we propose to leverage the *inter-diff* information from two aspects: (1) **Consistency-context** mainly considers that the prediction variance of two adjacent slices needs to follow the variance in the ground truth (GT). (2) **Discrepancy-context** assumes the label difference of adjacent slices usually occurs in the organ’s boundary area with more segment difficulty. To fully utilize the context information, we further devise a two-stage 2.5D segmentation framework that takes three adjacent slices as input. In the first stage, we encourage the predictions of the three slices to follow the consistency-context by leveraging it as an additional supervision signal. In the second stage, we further refine the segmentation results by adopting the prediction discrepancy area of adjacent slices as an extra input. In this way, our model can better capture the inter-slice context information of the CT scans, especially the organ/tumor edge regions with more segment difficulty, thus improving segmentation accuracy and maintaining the high efficiency. Besides, the proposed context information can be easily incorporated into other promising segmentation models (e.g., DeepLabv3 [3], PSPNet [24]). Experiments on three different organ datasets demonstrate our model’s: (a) Effectiveness of



**Fig. 1.** Schematic diagram of the results of adjacent slices predicted by U-Net model. *Mis-seg*: Regions of inconsistency between predictions and GTs. *Inter-diff*: The inter-slice differences between adjacent GT slices. The inter-diff regions have a high degree of overlapping with the mis-seg regions, which are basically located in tissues’ edge regions. These regions are difficult for segmentation models and can act as critical information for model learning. (Best viewed in color)

different components. (b) Generalization ability on different organ datasets. (c) Portability to different segmentation models.

## 2 Methodology

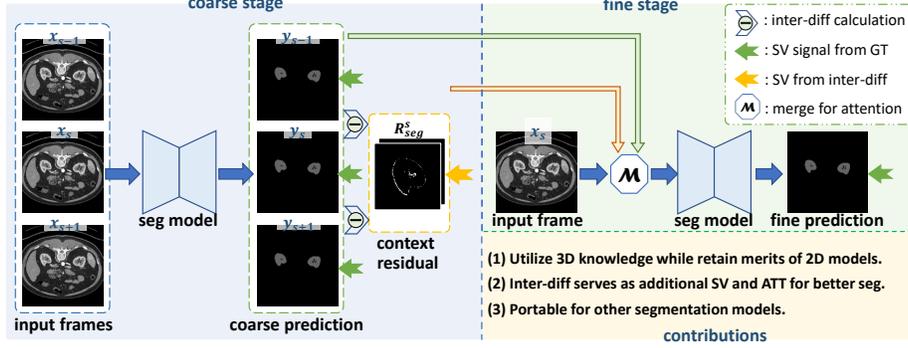
For robust segmentation of organ and anomaly tumor in 3D CT scan, our model adopts a two-stage 2.5D framework, which benefits from inter-slice difference (inter-diff) knowledge from two aspects: (1) Consistency context similarity between the predictions and GTs (Sec. 2.1), and (2) Discrepancy context (Sec. 2.2). The overall pipeline of our model is illustrated in Fig. 2.

Firstly, we will introduce the inter-diff, consistency context, and discrepancy context used in our model. We denote a 3D organ volume, the corresponding one-hot prediction and GT as  $\mathbf{X} \in \mathbf{R}^{N \times H \times W}$ ,  $\mathbf{Y}_{seg} \in \mathbf{R}^{N \times C \times H \times W}$ ,  $\mathbf{Y}_{GT} \in \mathbf{R}^{N \times C \times H \times W}$ , where  $N$  is the number of slices,  $C$  is the number of categories,  $H$  and  $W$  are height and width of each slice, respectively.

**Inter-diff:** The inter-diff in prediction for a slice  $t$  is the prediction difference between itself and the adjacent slice. In the same manner, we can get the inter-diff of the GT.

$$\mathcal{R}_{seg}^t = \mathbf{Y}_{seg}^t - \mathbf{Y}_{seg}^{t-1}, \quad \mathcal{R}_{GT}^t = \mathbf{Y}_{GT}^t - \mathbf{Y}_{GT}^{t-1}. \quad (1)$$

**Consistency context.** We assume that the prediction’s inter-diff of adjacent slices needs to follow the one in GT, and this knowledge act as an additional supervision signal in our model. This context constraint can be measured by the



**Fig. 2.** The framework of our model, which adopts a 2.5D two-stage architecture. Abbreviations: SV = supervision, ATT = attention. Best viewed in color.

L1 distance between the L2-norm of  $\mathcal{R}_{seg}^t$  and  $\mathcal{R}_{GT}^t$ , which goes as:

$$\mathcal{D}^t = \left| \|\mathcal{R}_{seg}^t\|^2 - \|\mathcal{R}_{GT}^t\|^2 \right|. \quad (2)$$

**Discrepancy context.** Our model leverages the fact that inter-diff usually occurs in tissues' edge regions which are hard to segment. We denote such knowledge as discrepancy context, and utilize it as the attention guidance for refinement.

## 2.1 Consistency context-based organ segmentation

As illustrated in Fig. 2, in the coarse segmentation stage, our model takes the target slice ( $x_t$ ) and its adjacent slices ( $x_{t-1}$  &  $x_{t+1}$ ) as input, and output the predicted segmentation masks of the three slices. For each slice, we calculate DiceCE loss ( $\mathcal{L}_{DCE}$ ) to measure the prediction accuracy, which goes as:

$$\mathcal{L}_{DCE} = \mathcal{L}_{Dice} + \mathcal{L}_{CE}. \quad (3)$$

Specifically,  $\mathcal{L}_{Dice} = -\sum_{i=1}^n (1 - 2 \cdot \frac{t_i p_i}{t_i + p_i})$  and  $\mathcal{L}_{CE} = -\sum_{i=1}^n t_i \log(p_i)$ , where  $t_i$  is the GT label, and  $p_i$  is the predicted probability for the  $i^{th}$  class.

Apart from the DiceCE loss for the three input slices, we further use a Contextual Residual loss ( $\mathcal{L}_{CR}$ ) to encourage the consistency context of the ( $x_{t-1}$ ,  $x_t$ ) and ( $x_t$ ,  $x_{t+1}$ ) to be minimum. The Contextual Residual loss ( $\mathcal{L}_{CR}$ ) goes as:

$$\mathcal{L}_{CR} = \mathcal{D}^t + \mathcal{D}^{t+1} \quad (4)$$

Finally, the overall loss function for the first coarse stage is

$$\mathcal{L}_{coarse} = \mathcal{L}_{DCE}(x_{t-1}) + \mathcal{L}_{DCE}(x_t) + \mathcal{L}_{DCE}(x_{t+1}) + \lambda_{CR} \mathcal{L}_{CR}, \quad (5)$$

where  $\lambda_{CR}$  is a hyper-parameter and is set to 0.05 in this study.

## 2.2 Segmentation refinement with discrepancy context knowledge

After the coarse stage, we further propose to utilize the discrepancy context to further rectify the segmentation results. As shown in Fig. 1, the edge regions of organ & tumor are with higher difficulty, and the inter-diff mask ( $M_t$ ) is with a high degree of overlap with the edge regions. Therefore, we leverage the discrepancy context as extra attention guidance. The discrepancy context used in this stage is formulated as the prediction difference regions in the adjacent slices from the coarse stage’s outputs, which goes as:

$$M^t = \mathbb{1}(L_{seg}^t \neq L_{seg}^{t-1}) + \mathbb{1}(L_{seg}^t \neq L_{seg}^{t+1}) \quad (6)$$

where  $L_{seg}^{t-1}$ ,  $L_{seg}^t$  and  $L_{seg}^{t+1}$  are the predicted labels.

Then, we leverage  $M^t$  as the attention information for refining the result  $L_{seg}^t$  of the slice  $X_t$ . In this study, we directly concatenate the  $X_t$  with  $M^t$  and the coarse segmentation result  $Y_{seg}^t$  as an augment input for a 2D U-Net model, and compute the refined segmentation results. In this stage, we train the model by the DiceCE loss.

## 3 Datasets and Implementation Details

**Datasets.** We conduct the experiments to evaluate the performance of our method on three different organ CT datasets: KiTS ( $\mathcal{D}_K$ , for kidney)[9], LiTS ( $\mathcal{D}_L$ , for liver)[1], and Pancreas ( $\mathcal{D}_P$ )[18]. The statistical details of the datasets are summarized in Table 1.

**Table 1.** The statistics of datasets adopted in our study.

Datasets	Labeled Organs	#Volumes		
		training	testing	total
KiTS ( $\mathcal{D}_K$ )	Kidney	168	42	210
LiTS ( $\mathcal{D}_L$ )	Liver	104	26	130
Panc ( $\mathcal{D}_P$ )	Pancreas	225	57	282

**Implementation Details.** We implement our model with the PyTorch library [14] in a device with an NVIDIA 2080TI GPU. The image intensity is windowed by [-160, 240] for  $\mathcal{D}_K$ , [-100, 400] for  $\mathcal{D}_L$  and [-100, 240] for  $\mathcal{D}_P$ . Each slice is resized to  $512 \times 512$  and then randomly cropped to  $256 \times 256$  for training. We adopt random flip for data augmentation. For the U-Net model, we adopt the ResNeXT50\_32x4d [21] as the encoder’s backbone, which could be replaced by other segmentation backbones such as VGG-16 and ResNet-50. For both stages, we set the number of training epochs and batch-size to 12 and 8, respectively. The Adam [13] is used for optimization with an initial learning rate of 1e-4.

**Evaluation metrics:** We adopt Dice-Sørensen Coefficient (DSC) as the evaluation metrics, which goes as  $DSC(\mathcal{P}, \mathcal{G}) = \frac{2 \times |\mathcal{P} \cap \mathcal{G}|}{|\mathcal{P}| + |\mathcal{G}|}$ , where  $\mathcal{P}$  is the binary prediction and  $\mathcal{G}$  is the ground-truth.

**Table 2.** The segmentation accuracy of our model and the baseline models on three organ segmentation datasets (in DSC (%)). Kid., Liv., and Panc. are short for kidney, liver, and pancreas, respectively.

Model	$\mathcal{D}_K$			$\mathcal{D}_L$			$\mathcal{D}_P$		
	mean	Kid.	tumor	mean	Liv.	tumor	mean	Panc.	tumor
2D U-Net	78.64	94.90	62.37	72.38	94.39	50.37	58.86	76.27	41.45
2.5D U-Net-1	78.71	94.47	62.95	74.64	95.07	54.21	59.54	76.56	42.51
2.5D U-Net-3	79.46	95.06	63.85	75.29	95.07	55.50	59.76	76.72	42.80
Ours(coarse stage)	81.06	95.26	66.86	75.80	95.44	56.16	60.05	77.03	43.07
Ours(fine stage)	<b>81.78</b>	<b>96.57</b>	<b>66.98</b>	<b>79.63</b>	<b>95.62</b>	<b>63.63</b>	<b>60.49</b>	<b>77.45</b>	<b>43.53</b>

## 4 Experimental results and analysis

### 4.1 The effectiveness of our proposed method

We first evaluate the effectiveness of the proposed model by comparing it with three baseline models. Specifically, we choose U-Net in vanilla version and two 2.5D variants as baselines. (1) *2D U-Net* is a vanilla 2D U-Net that only uses one slice as input. (2) *2.5D U-Net-1* is a 2.5D U-Net that takes three adjacent slices as input, while only predict the segmentation result of the middle slice for training and testing. (3) *2.5D U-Net-3* takes three adjacent slices as input, while computing the segmentation results for all three slices for training and testing. The DSC results on three different organ datasets ( $\mathcal{D}_K$ ,  $\mathcal{D}_L$ ,  $\mathcal{D}_P$ ) are reported in Table 2. From the table, we can make the following two conclusions.

**(1) More input slices and labels are beneficial for the accuracy.** In Table 2, we observe that 2.5D-based model can consistently improve the performance on three datasets. For example, the mean DSC on  $\mathcal{D}_L$  is improved from 72.38% to 74.64% when switching the model from 2D U-Net to 2.5D U-Net-1. Besides, with the supervision of more adjacent slices, the 2.5D U-Net-3 can further increase the mean DSC from 74.64% to 75.29% compared with 2.5D U-Net-1. These results demonstrate that the 2.5D architecture with more input slices and labels can jointly improve the segmentation performance.

**(2) The consistency- and discrepancy- knowledge are effective for the segmentation.** From the last two rows in Table 2, we can find that our model outperforms the above baseline models consistently in both coarse stage and fine stage. For example, with consistency context, the result in the coarse stage outperforms the best baseline model (2.5D U-Net-3) by 3.13% and 2.32% in the tumor DSC and mean DSC on  $\mathcal{D}_K$ , respectively. In addition, the discrepancy context leveraged in the fine stage further boosts the accuracy by 0.72%, 3.83%, and 0.44% in mean DSC of  $\mathcal{D}_K$ ,  $\mathcal{D}_L$ ,  $\mathcal{D}_P$ , respectively. Note that the tumor DSC on  $\mathcal{D}_L$  witnesses a remarkable gain from 56.16% to 63.63%. These results verify that incorporating consistency- and discrepancy- knowledge can improve model’s segmentation performance, especially for relatively smaller targets like tumors.

## 4.2 Portability with different segmentation models

In this section, we evaluate the portability of our proposed context information with several different promising models, including U-Net [16], PSPNet [24], DeepLabv3 [2]. Notice that all the comparing models are implemented in the same 2.5D two-stage training manner. In Table 3, we report the accuracy without and after using the proposed context, and can have following observations.

(1) **Both contexts can increase the accuracy of different models.** As shown in the left part of Table 3, our model can bring consistent performance gains for the coarse stage of all the three segmentation models (1.61/12.87/2.57 points gains in mean DSC by using U-Net, PSPNet, DeepLabV3, respectively). In the right part of the table, we can observe similar results for all the three models at the fine stage (2.00/11.75/2.68 points gains in mean DSC over U-Net, PSPNet, DeepLabV3, respectively). These experiments demonstrate the portability of the proposed consistency- and discrepancy-context that can be effortlessly incorporated into different models.

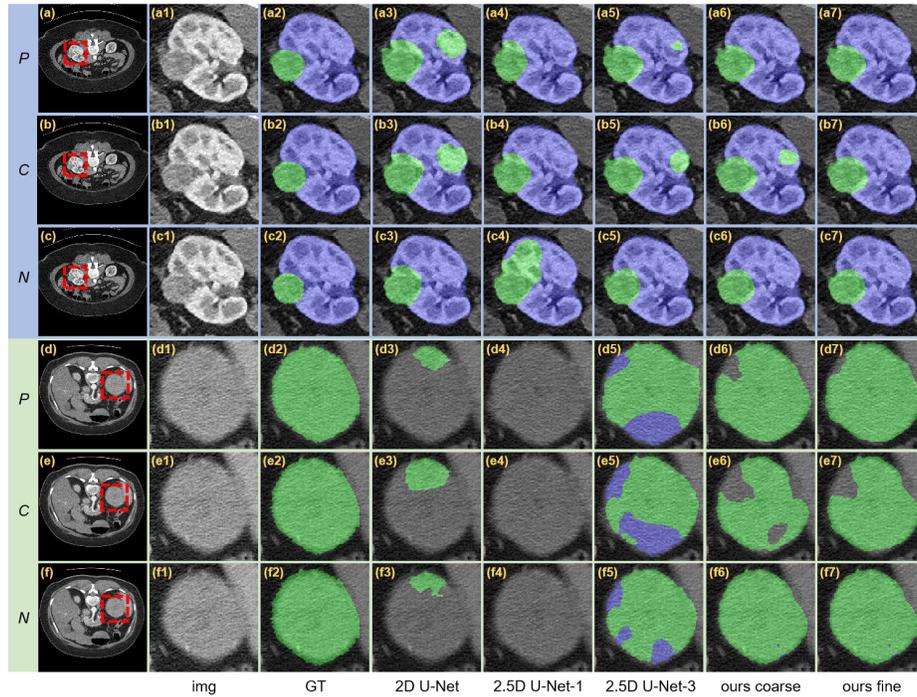
(2) **The context information has a more positive impact on small targets.** In Table 3, we can find that tumor DSCs achieve relatively more improvement, e.g., 64.84% to 68.74% for DeepLabv3’s coarse stage and 64.90% to 69.31% for DeepLabv3’s fine stage. These results suggest that our proposed method has a more positive impact on smaller targets, which can be critical for clinical scenarios since accurate recognition of the small tumor is essential.

**Table 3.** Experimental results (in DSC) of different segmentation models without and after using the proposed context on the KiTS dataset.

Method	Coarse stage			Fine stage		
	mean(%)	kidney(%)	tumor(%)	mean(%)	kidney(%)	tumor(%)
U-Net [16]	79.46	95.06	63.85	79.78	95.06	64.50
U-Net+ours	81.06	95.26	66.86	81.78	96.57	66.98
PSPNet [24]	52.69	72.78	32.59	54.30	72.55	36.04
PSPNet+ours	65.55	86.48	44.62	66.04	87.12	44.96
DeepLabv3 [2]	79.00	93.16	64.84	79.37	93.84	64.90
DeepLabv3+ours	81.57	94.39	68.74	82.05	94.78	69.31

## 4.3 Qualitative results

Fig. 3 depicts the qualitative results of our model against baselines on examples from  $\mathcal{D}_K$ . Although the original adjacent slices are similar, the baseline models give inconsistent predictions, especially for false tumor regions. The 2D UNet is prone to over-segment the tumor regions (a3, b3) and missing kidney regions (d3, e3, f3). The 2.5D U-Net-1 has better prediction in (a4, b4), but over-segment tumor in (c4), and can not recognize tumors in (d4, e4, f4). The 2.5D U-Net-3 could alleviate the over-segmentation in (a5, b5, c5), but errors still exist. In



**Fig. 3.** The qualitative segmentation results of two representative adjacent slices. Blue and Green colors show the predictions of organs and tumors. P=previous, C=current, N=Next. Best viewed in color.

addition, 2.5D U-Net-3 locate tumor regions of (d, e, f) better, but all have false-positive regions of kidney. In (b6), our coarse model also suffers the same issue but is corrected by our fine model (b7). Comparing the above results with (d6, e6, f6) and (d7, e7, f7), our coarse model can also help locate the tumor regions and eliminate the false positive region of kidney. Besides, our fine model further recovers the missing tumor areas ((d6) to (d7), (e6) to (e7)).

## 5 Conclusion

In this study, we aim to boost the 2D segmentation model’s accuracy by leveraging the inter-slice context information while maintaining the efficiency of 2D models. Specifically, our model adopts a 2.5D coarse-to-fine architecture, which benefits from the inter-slice context knowledge from two aspects: (1) Consistency context similarity between the predictions and GTs for additional supervision, and (2) Discrepancy context for attention guidance. The experiments demonstrate that our method achieves considerable improvement in different datasets and can be incorporated into other promising segmentation models. For future work, we will further investigate the incorporation of a more extensive range

of contextual information and the organ’s anatomical priors to further improve accuracy.

## Acknowledgement

This work is supported by the National Nature Science Foundation of China (No. 61876159, 61806172, 62076116 & U1705286), the China Postdoctoral Science Foundation Grant (No. 2019M652257), the Guiding Project of Science and Technology Department of Fujian Province (No. 2019Y0018).

## References

1. Bilic, P., Christ, P.F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C.W., Han, X., Heng, P.A., Hesser, J., et al.: The liver tumor segmentation benchmark (lits). arXiv (2019)
2. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv (2017)
3. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: ECCV (2018)
4. da Cruz, L.B., Araújo, J.D.L., Ferreira, J.L., Diniz, J.O.B., Silva, A.C., de Almeida, J.D.S., de Paiva, A.C., Gattass, M.: Kidney segmentation from computed tomography images using deep neural network. *Computers in Biology and Medicine* (2020)
5. Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., Heng, P.A.: 3d deeply supervised network for automatic liver segmentation from ct volumes. In: MICCAI (2016)
6. Haghighi, M., Warfield, S.K., Kurugol, S.: Automatic renal segmentation in dcmri using convolutional neural networks. In: ISBI (2018)
7. Han, X.: Automatic liver lesion segmentation using a deep convolutional neural network method. arXiv (2017)
8. Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., et al.: The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. *Medical Image Analysis* (2021)
9. Heller, N., Sathianathan, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., et al.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv (2019)
10. Hou, X., Xie, C., Li, F., Wang, J., Lv, C., Xie, G., Nan, Y.: A triple-stage self-guided network for kidney tumor segmentation. In: ISBI (2020)
11. Howe, R.D., Matsuoka, Y.: Robotics for surgery. *Annual Review of Biomedical Engineering* (1999)
12. Khalifa, F., Elnakib, A., Beache, G.M., Gimel’farb, G., El-Ghar, M.A., Ouseph, R., Sokhadze, G., Manning, S., McClure, P., El-Baz, A.: 3d kidney segmentation from ct images using a level set approach guided by a novel stochastic speed function. In: MICCAI (2011)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv (2014)
14. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: NeurIPS (2019)

15. Pekar, V., McNutt, T.R., Kaus, M.R.: Automated model-based organ delineation for radiotherapy planning in prostatic region. *IJROBP* (2004)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *MICCAI* (2015)
17. Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B., Summers, R.M.: Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: *MICCAI* (2015)
18. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv* (2019)
19. Thong, W., Kadoury, S., Piché, N., Pal, C.J.: Convolutional networks for kidney segmentation in contrast-enhanced ct scans. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* (2018)
20. Van Ginneken, B., Schaefer-Prokop, C.M., Prokop, M.: Computer-aided diagnosis: how to move from the laboratory to the clinic. *Radiology* (2011)
21. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: *CVPR* (2017)
22. Yang, D., Xu, D., Zhou, S.K., Georgescu, B., Chen, M., Grbic, S., Metaxas, D., Comaniciu, D.: Automatic liver segmentation using an adversarial image-to-image network. In: *MICCAI* (2017)
23. Yu, Z., Pang, S., Du, A., Orgun, M.A., Wang, Y., Lin, H.: Fine-grained tumor segmentation on computed tomography slices by leveraging bottom-up and top-down strategies. In: *Medical Imaging 2020: Image Processing* (2020)
24. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: *CVPR* (2017)