Crop and Segment: Efficient Whole body pan-cancer segmentation

Yinyin Luo[†], Yue Liu[†], Wenbin Liu, Jingheng Dai, Xunliang Xiao, and Gang Fang[*]

Institute of Computing Science and Technology, Guangzhou University, Guangzhou, 510006, China
[†]Co-first authors
[*]Corresponding author
gangf@gzhu.edu.cn

Abstract. Despite significant advancements in deep learning models for medical segmentation, the detection and segmentation of tumors, particularly through whole-body scans, remain challenging. To address this issue, we explored the application of nnUNet for whole-body tumor segmentation in CT scans, proposing a more precise cropping strategy and introducing an organ-interference segmentation approach to effectively enhance segmentation efficiency. Experiments on the MICCAI FLARE 2024 dataset demonstrated significant improvements in both segmentation accuracy and efficiency. Our method achieved an average organ Dice Similarity Coefficient (DSC) of 10.47% and a Normalized Surface Dice (NSD) of 7.98% on the public validation set. In the FLARE 2024 Task 1 online validation, the method achieved an average organ Dice Similarity Coefficient (DSC) of 17.08%, a Normalized Surface Dice (NSD) of 7.42%, an average inference time of 8 seconds and the area under GPU memory-time curve is 18.07s and 45688 MB. The code is available at https://github.com/lay-john/FLARE24-Task1.

Keywords: Semi-supervised · Deep learning · Tumour segmentation .

1 Introduction

In recent years, deep learning models have significantly advanced the field of medical image analysis, particularly in medical segmentation. Notable models such as U-Net [20], which introduced a novel architecture for biomedical image segmentation, and its extended versions like nnU-Net [10], have demonstrated substantial improvements in various segmentation tasks. These advancements have facilitated enhanced performance in segmenting organs and pathological regions across numerous medical imaging modalities.However, despite these breakthroughs, the accurate detection and segmentation of tumors in whole-body scans continue to pose significant challenges. Tumors can vary greatly in shape, size, and appearance across different anatomical regions, making their segmentation inherently complex. Additionally, whole-body scans involve high-resolution

3D CT images that are both computationally intensive and resource-demanding. This complexity is exacerbated by the need to process large volumes of data while maintaining high accuracy and speed.Recent studies have highlighted the difficulties in extending segmentation models to whole-body scans. For instance, the work by Liu et al. (2018) on the DeepMedic model [12] demonstrated improved performance in brain tumor segmentation, but challenges remain in generalizing these methods to whole-body tumor detection. Similarly, the work by Zhou et al. (2019) on 3D U-Net [28] showed promising results for organ segmentation but highlighted the need for further improvements in handling the variability of tumors across the body.To address these challenges, various approaches have been explored. For example, the study by Yang et al. (2021) proposed a multi-scale approach to handle the variability in tumor sizes and shapes [26]. Despite these advancements, achieving accurate and efficient whole-body tumor segmentation remains a critical and ongoing challenge in the field of medical imaging.Among these methods, it is highly noted that nnUNet has consistently demonstrated outstanding performance in different medical image segmentation tasks.

Our study builds on these advancements by leveraging the nnUNet framework, known for its robust and flexible design in medical image segmentation. We propose novel strategies to enhance the efficiency and accuracy of whole-body tumor segmentation, including more precise cropping methods and refined post-processing techniques. This approach aims to address the inherent challenges associated with high-resolution 3D CT scans, ultimately improving the performance and applicability of segmentation models in clinical practice.To tackle these challenges, we focused on leveraging the nnUNet framework, a state-of-the-art model known for its robustness and flexibility in medical image segmentation. nnUNet has shown promising results in various medical imaging tasks.

Specifically, we propose a cropping strategy based on the largest connected component, which retains the largest connected region in CT images by evaluating voxel intensity relationships. The boundaries for cropping are then determined based on the content of the retained largest connected component, thereby achieving a more precise delineation of the Region of Interest (ROI). Furthermore, During our experiments, we observed that most tumor labels in the dataset were only partially annotated. Due to the complexity of the whole-body regions, the trained model frequently misclassified most organ areas as tumors, even though the actual tumor regions in each CT scan occupy only a small portion. To address this, we introduced an organ-interference segmentation approach. By jointly training the model on both tumors and organs, we enforced the model to focus more on the tumor regions during training, thus helping it to eliminate most of the misclassified organ areas without significantly impacting the accuracy of tumor region predictions. The detailed methodology is described in Section 4.1.

To summarize, this paper presents a precise cropping strategy based on the largest connected component and introduced an organ-interference segmentation approach to address MICCAI FLARE 2024 Challenge Task 1. The main contributions are as follows: (1) To achieve more accurate cropping regions and

improve inference speed, we propose a precise cropping strategy based on the largest connected component, which enhances both inference speed and segmentation accuracy. (2) To reduce false positives in tumor segmentation, we further introduce an organ-interference segmentation approach. Experimental results demonstrate the effectiveness of this approach in improving model performance.

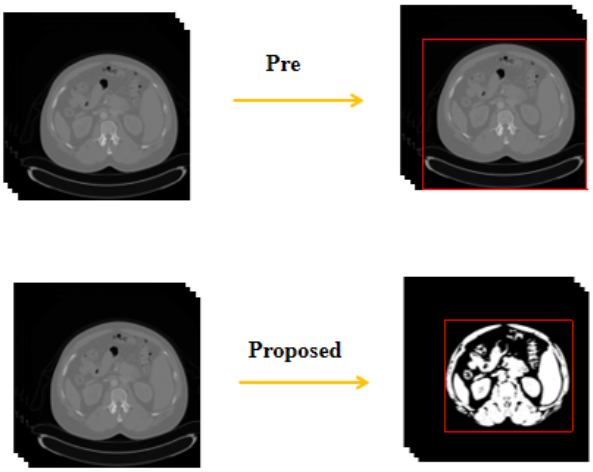

Fig. 1. "Pre" refers to the previous default cropping method, while "Proposed" denotes the cropping strategy based on the largest connected component that we have introduced in this study.

2 Method

2.1 Preprocessing

- **Resample and normalization:** We resample the pixel spacing to (1.6, 1.2859, 1.2859) for all cases, and clip the pixel value based on the Hounsfield units to [-160, 240], and normalize all the cases in [0, 1] to ensure data stability and consistency.
- **Data augmentation:** In order to prevent the model from over-fitting, data augmentation is used in this study. The augmentation approaches of nnU-Net methodology have been utilized.

2.2 Proposed Method

Specifically, the primary focus of this work is on preprocessing and organ-interference. Therefore, we utilize the default nnUNet architecture as our model for training. We use the provided FLARE 2024 Task 1 dataset, which comprises over 10,000 CT scans covering various whole-body cancer types. The dataset includes 5,000 partially annotated scans and 5,490 unannotated scans. In the partially annotated dataset, only the primary lesions are labeled, while other lesions, such as metastatic ones, may remain unlabeled. For our model training, we use only the 5,000 partially annotated scans.The first contribution of this paper is the proposal of a cropping strategy based on the largest connected component. This strategy retains the largest connected region in CT images by evaluating voxel intensity relationships. The cropping boundaries are then determined based on the content of the retained largest connected component, resulting in more precise and smaller cropping regions. Another contribution is the introduction of an organ-interference segmentation approach. By jointly training the model on both tumors and organs, we enforced the model to focus more on the tumor regions during training, thus helping it to eliminate most of the misclassified organ areas without significantly impacting the accuracy of tumor region predictions.

A cropping strategy based on the largest connected component: First, we obtain a mask by thresholding the original CT voxel values greater than 50. Then, the largest connected component is preserved within this mask, resulting in a refined mask containing only the largest connected region. Based on this refined mask, we determine the bounding box for cropping and subsequently perform the cropping operation.As shown in Fig. 1., it can be seen that compared with previous cropping methods, our cropping strategy can significantly reduce the cropped area without affecting our subsequent segmentation.

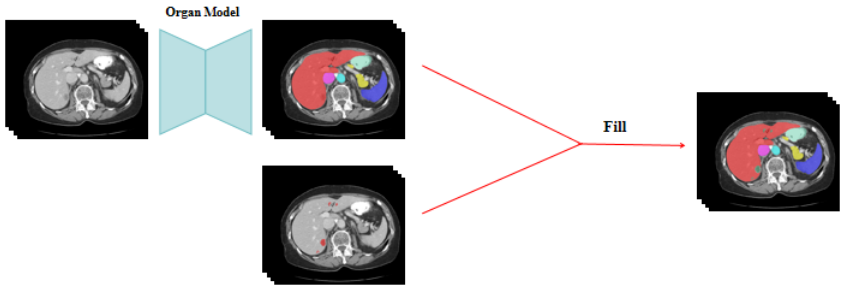

Fig. 2. Generate final organ and tumor hybrid labels using organ models and original labels.

An organ-interference segmentation approach: We utilized the abdominal CT data from FLARE 2022 [16] to train an organ model. This model

was then applied to generate labels for abdominal organs on 5,000 annotated CT scans, resulting in pseudo-labels for the abdominal organs. The tumor labels from the provided annotations were then integrated into the generated pseudo-labels, forming the final training labels, as illustrated in Fig. 2. During training, these final labels were sampled with a certain probability during data loading. If a label was selected, we extracted a patch centered on the selected label. We increased the probability of selecting tumor labels by approximately five times compared to other labels, while the probabilities for the remaining organ labels were kept equal. The training process employed a weighted DiceLoss and CELoss. For the weight w, we set the tumor regions assigned a weight five times higher than the organ regions. This weighting scheme was designed to direct the model's focus toward more accurate tumor segmentation. Our loss function formula is as follows:

$$L_{Dice}(y, \hat{y}, w) = \sum_{i}^{c_0} w_i \left(1 - \frac{2\sum_{j=1}^{N} y_j^i \hat{y}_j^i}{\sum_{j=1}^{N} y_j^i + \hat{y}_j^i}\right) \tag{1}$$

$$L_{CE}(y, \hat{y}, w) = \sum_{i}^{c_0} w_i \left(-\frac{1}{N}\sum_{j=1}^{N}\left(y_j^i \log(\hat{y_j^i}) + (1 - y_j^i)\log(1 - \hat{y_j^i})\right)\right) \tag{2}$$

$$L_{Seg} = \alpha_{dc} \cdot L_{Dice}(y, \hat{y}, w) + \alpha_{ce} \cdot L_{CE}(y, \hat{y}, w) \tag{3}$$

Where the y and $\hat{y}$ mean the ground truth and the predicted probability, respectively, and N is the number of pixels. α_{dc} and α_{ce} are the hyperparameters to balance the contribution of DiceLoss and CELoss. α_{dc} and α_{ce} are set to 0.5 in this study.

Fig. 3. Processing of inferred labels.

2.3 Post-processing

We remove the organ labels from the inferred results and only retain the tumor labels, as shown in Fig. 3. In addition, no post-processing operations are performed.

3 Experiments

3.1 Dataset and evaluation measures

The segmentation targets cover various lesions. The training dataset is curated from more than 50 medical centers under the license permission, including TCIA [3], LiTS [2], MSD [22], KiTS [7,9,8], autoPET [6,5], TotalSegmentator [23], and AbdomenCT-1K [18], FLARE 2023 [17], DeepLesion [25], COVID-19-CT-Seg-Benchmark [15], COVID-19-20 [21], CHOS [11], LNDB [19], and LIDC [1]. The training set includes 4000 abdomen CT scans where 2200 CT scans with partial labels and 1800 CT scans without labels. The validation and testing sets include 100 and 400 CT scans, respectively, which cover various abdominal cancer types, such as liver cancer, kidney cancer, pancreas cancer, colon cancer, gastric cancer, and so on. The lesion annotation process used ITK-SNAP [27], nnU-Net [10], MedSAM [13], and Slicer Plugins [4,14].

The evaluation metrics encompass two accuracy measures—Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD)—alongside two efficiency measures—running time and area under the GPU memory-time curve. These metrics collectively contribute to the ranking computation. Furthermore, the running time and GPU memory consumption are considered within tolerances of 45 seconds and 4 GB, respectively.

3.2 Implementation details

Environment settings The development environments and requirements are presented in Table 1.

Table 1. Development environments and requirements.

System	Ubuntu 22.04 LTS or Windows 10
CPU	Intel(R) Core(TM) i9-10900X CPU@3.70GHz
RAM	4×32GB; 2933MT/s
GPU (number and type)	NVIDIA GeForce RTX™3090 24G
CUDA version	12.1
Programming language	Python 3.9.16
Deep learning framework	torch 2.1.0, torchvision 0.16.0
Specific dependencies	nnU-Net 1.7.0
Code	https://github.com/lay-john/FLARE24-Task1

Training protocols During the training phase, we set the batch size to 2 and randomly select all samples within each epoch. For each sample, we perform random patch cropping with patch sizes of (96, 128, 160). As for the optimizer, we utilize AdamW with a learning rate of 1e-2 and a weight decay of 1e-5. The learning rate updating follows the default mechanism of AdamW. Additional details are presented in Table 2.

Table 2. Training protocols.

Network initialization	
Batch size	2
Patch size	96×128×160
Total epochs	500
Optimizer	AdamW with weight decay(μ = 1e -5)
Initial learning rate (lr)	0.01
Lr decay schedule	halved by 200 epochs
Training time	42.5 hours
Loss function	DiceLoss and CELoss
Number of model parameters	30.8M[1]
Number of flops	838.6116 KG[2]
CO_2eq	3.91908 Kg[3]

4 Results and discussion

4.1 Quantitative results on validation set

To conduct a more comprehensive ablation study of our proposed method, we performed quantitative experiments, as shown in Table 3.

Table 3. Result in Public Validation, Online Validation and Final Testing.

Methods	Public Validation		Online Validation		Testing	
	DSC(%)	NSD(%)	DSC(%)	NSD(%)	DSC(%)	NSD (%)
Algorithm	10.47	7.98	17.08	7.42	32.73	20.58

To visually demonstrate the impact of our method on inference speed, we conducted quantitative experiments on inference speed, as shown in Table 4. The length of the step is [7/8, 7/8, 7/8] times the window width for each axis.

Table 4. Overview of Ablation Experiment Results.

Target	Base		Proposed	
	DSC(%)	NSD(%)	DSC(%)	NSD(%)
Tumour	3.97	1.98	10.47	7.98

To visually demonstrate the impact of our method on inference speed, we conducted quantitative experiments on inference speed, as shown in Table 5. The length of the step is [7/8, 7/8, 7/8] times the window width for each axis.

Table 5. Overview of Ablation Experiment Results on inference speed. Time is measured in seconds.

Target	Average time(s)
Base	9.89
Proposed	8.09

4.2 Qualitative results on validation set

In this section, we show the two good segmentation cases and two bad segmentation cases.

Good segmentation cases:Fig. 4 presents examples of good segmentation results. It can be observed that the segmentation performance of our method is nearly comparable to that of manual annotations. In comparison to the baseline, while the baseline can also segment the tumor regions, it exhibits inferior boundary delineation for tumors. This improved boundary accuracy is one of the key reasons why our method outperforms the baseline.

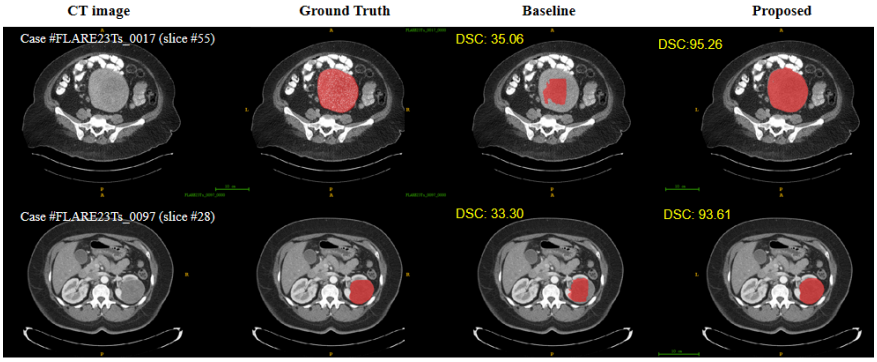

Fig. 4. Good segmentation cases from public validation set.

Bad segmentation cases: Fig. 5 presents examples of bad segmentation results.In the case of LNDb-0303, we observe that there is a small tumor region in the Ground Truth that our method failed to segment. Although the baseline was able to segment this region, it also incorrectly classified large additional areas as tumors, resulting in suboptimal segmentation performance. For LNDb-0312, where there is no tumor region in the Ground Truth, both our method and the baseline misclassified other regions as tumors. These two cases highlight that our method struggles with CT scans where there is either a very small tumor region or no tumor at all. This could be due to the model's inherent insensitivity to such regions or potentially incomplete training. Due to the summer break and other factors, the final model was trained on a single 3060 GPU for only 250 epochs. Another possible explanation is the absence of organs in these regions, leading to no organ interference, which might have contributed to the poor, or even highly inaccurate, segmentation performance in these areas.

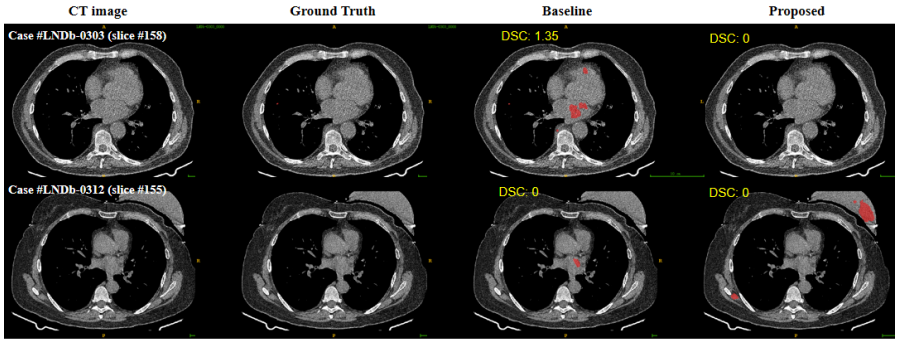

Fig. 5. Bad segmentation cases from public validation set.

4.3 Segmentation efficiency results on validation set

We quantitatively evaluate the segmentation efficiency of our model, as shown in Table 6. We also tested our model the false positive rate on the healthy CT scans and concluded that our method has the false positive rate of 0.875 on provided healthy CT scans.

4.4 Results on final testing set

4.5 Limitation and future work

In this study, we incorporated only abdominal organs to interfere with tumor segmentation. As a result, our method demonstrated good overall performance in tumor segmentation within the abdominal region. However, the performance

Table 6. Quantitative evaluation of segmentation efficiency in terms of the running them and GPU memory consumption. Total GPU denotes the area under GPU Memory-Time curve. Evaluation GPU platform: NVIDIA QUADRO RTX5000 (16G).

Case ID	Image Size	Running Time (s)	Max GPU (MB)	Total GPU (MB)
0001	(512, 512, 55)	6	4008	14146
0051	(512, 512, 100)	12	4660	43015
0017	(512, 512, 150)	12	4738	43835
0019	(512, 512, 215)	9	4194	26961
0099	(512, 512, 334)	7.5	4386	24930
0063	(512, 512, 448)	10.5	4589	30123
0048	(512, 512, 499)	9	4566	30352
0029	(512, 512, 554)	16.5	5142	60602

in other regions was less satisfactory. In future work, we plan to explore the inclusion of additional organ labels to enhance tumor segmentation in non-abdominal areas. Furthermore, the accuracy of our model in delineating tumor boundaries requires further improvement, and we will also focus on refining boundary segmentation in future studies.

5 Conclusion

To tackle the challenging task of whole-body pan-cancer segmentation and improve the overall performance, this paper proposes a tumor segmentation method based on organ interference. To enhance inference speed, we also introduce a cropping strategy based on the largest connected component. Both quantitative and qualitative results demonstrate that the proposed method effectively and flexibly learns tumor information from the dataset. We validated our approach on the MICCAI FLARE 2024 challenge dataset, proving its efficacy in whole-body pan-cancer segmentation.

Acknowledgements The authors of this paper declare that the segmentation method they implemented for participation in the FLARE 2024 challenge has not used any pre-trained models nor additional datasets other than those provided by the organizers. The proposed solution is fully automatic without any manual intervention. We thank all data owners for making the CT scans publicly available and CodaLab [24] for hosting the challenge platform.

Disclosure of Interests

The authors declare no competing interests.

References

1. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al.: The

lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. Medical Physics **38**(2), 915–931 (2011) 6

2. Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G.E.H., Chartrand, G., Lohöfer, F., Holch, J.W., Sommer, W., Hofmann, F., Hostettler, A., Lev-Cohain, N., Drozdzal, M., Amitai, M.M., Vivanti, R., Sosna, J., Ezhov, I., Sekuboyina, A., Navarro, F., Kofler, F., Paetzold, J.C., Shit, S., Hu, X., Lipková, J., Rempfler, M., Piraud, M., Kirschke, J., Wiestler, B., Zhang, Z., Hülsemeyer, C., Beetz, M., Ettlinger, F., Antonelli, M., Bae, W., Bellver, M., Bi, L., Chen, H., Chlebus, G., Dam, E.B., Dou, Q., Fu, C.W., Georgescu, B., i Nieto, X.G., Gruen, F., Han, X., Heng, P.A., Hesser, J., Moltz, J.H., Igel, C., Isensee, F., Jäger, P., Jia, F., Kaluva, K.C., Khened, M., Kim, I., Kim, J.H., Kim, S., Kohl, S., Konopczynski, T., Kori, A., Krishnamurthi, G., Li, F., Li, H., Li, J., Li, X., Lowengrub, J., Ma, J., Maier-Hein, K., Maninis, K.K., Meine, H., Merhof, D., Pai, A., Perslev, M., Petersen, J., Pont-Tuset, J., Qi, J., Qi, X., Rippel, O., Roth, K., Sarasua, I., Schenk, A., Shen, Z., Torres, J., Wachinger, C., Wang, C., Weninger, L., Wu, J., Xu, D., Yang, X., Yu, S.C.H., Yuan, Y., Yue, M., Zhang, L., Cardoso, J., Bakas, S., Braren, R., Heinemann, V., Pal, C., Tang, A., Kadoury, S., Soler, L., van Ginneken, B., Greenspan, H., Joskowicz, L., Menze, B.: The liver tumor segmentation benchmark (lits). Medical Image Analysis **84**, 102680 (2023) 6
3. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F.: The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of Digital Imaging **26**(6), 1045–1057 (2013) 6
4. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al.: 3d slicer as an image computing platform for the quantitative imaging network. Magnetic Resonance Imaging **30**(9), 1323–1341 (2012) 6
5. Gatidis, S., Früh, M., Fabritius, M., Gu, S., Nikolaou, K., Fougère, C.L., Ye, J., He, J., Peng, Y., Bi, L., Ma, J., Wang, B., Zhang, J., Huang, Y., Heiliger, L., Marinov, Z., Stiefelhagen, R., Egger, J., Kleesiek, J., Sibille, L., Xiang, L., Bendazolli, S., Astaraki, M., Schölkopf, B., Ingrisch, M., Cyran, C., Küstner, T.: The autopet challenge: towards fully automated lesion segmentation in oncologic pet/ct imaging. Nature Machine Intelligence (2023) 6
6. Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, C., Schölkopf, B., Küstner, T., Cyran, C., Rubin, D.: A whole-body fdg-pet/ct dataset with manually annotated tumor lesions. Scientific Data **9**(1), 601 (2022) 6
7. Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., Yao, G., Gao, Y., Zhang, Y., Wang, Y., Hou, F., Yang, J., Xiong, G., Tian, J., Zhong, C., Ma, J., Rickman, J., Dean, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Kaluzniak, H., Raza, S., Rosenberg, J., Moore, K., Walczak, E., Rengel, Z., Edgerton, Z., Vasdev, R., Peterson, M., McSweeney, S., Peterson, S., Kalapara, A., Sathianathen, N., Papanikolopoulos, N., Weight, C.: The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. Medical Image Analysis **67**, 101821 (2021) 6
8. Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Zhao, Z., Chen, H., Wang, L., Golts, A., Khapun, D., Shats, D., Shoshan, Y., Gilboa-Solomon, F., George, Y., Yang, X., Zhang, J., Zhang, J., Xia, Y., Wu, M., Liu, Z., Walczak, E., McSweeney,

S., Vasdev, R., Hornung, C., Solaiman, R., Schoephoerster, J., Abernathy, B., Wu, D., Abdulkadir, S., Byun, B., Spriggs, J., Struyk, G., Austin, A., Simpson, B., Hagstrom, M., Virnig, S., French, J., Venkatesh, N., Chan, S., Moore, K., Jacobsen, A., Austin, S., Austin, M., Regmi, S., Papanikolopoulos, N., Weight, C.: The kits21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase ct. arXiv preprint arXiv:2307.01984 (2023) 6

9. Heller, N., McSweeney, S., Peterson, M.T., Peterson, S., Rickman, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Rosenberg, J., et al.: An international challenge to use artificial intelligence to define the state-of-the-art in kidney and kidney tumor segmentation in ct imaging. American Society of Clinical Oncology **38**(6), 626–626 (2020) 6
10. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021) 1, 6
11. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., Baydar, B., Lachinov, D., Han, S., Pauli, J., Isensee, F., Perkonigg, M., Sathish, R., Rajan, R., Sheet, D., Dovletov, G., Speck, O., Nürnberger, A., Maier-Hein, K.H., Bozdağı Akar, G., Ünal, G., Dicle, O., Selver, M.A.: CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021) 6
12. Li, X., Xu, D., Zhang, X., et al.: Deepmedic: A deep learning approach to brain tumor segmentation. Medical Image Analysis **30**, 183–193 (2016) 2
13. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024) 6
14. Ma, J., Kim, S., Li, F., Baharoon, M., Asakereh, R., Lyu, H., Wang, B.: Segment anything in medical images and videos: Benchmark and deployment. arXiv preprint arXiv:2408.03322 (2024) 6
15. Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z., Cao, T., Zhu, Y., Nie, Z., Yang, X.: Towards data-efficient learning: A benchmark for covid-19 ct lung and infection segmentation. Medical Physics **48**(3), 1197–1210 (2021) 6
16. Ma, J., Zhang, Y., Gu, S., Ge, C., Ma, S., Young, A., Zhu, C., Meng, K., Yang, X., Huang, Z., Zhang, F., Liu, W., Pan, Y., Huang, S., Wang, J., Sun, M., Xu, W., Jia, D., Choi, J.W., Alves, N., de Wilde, B., Koehler, G., Wu, Y., Wiesenfarth, M., Zhu, Q., Dong, G., He, J., the FLARE Challenge Consortium, Wang, B.: Unleashing the strengths of unlabeled data in pan-cancer abdominal organ quantification: the flare22 challenge. Lancet Digital Health (2024) 4
17. Ma, J., Zhang, Y., Gu, S., Ge, C., Wang, E., Zhou, Q., Huang, Z., Lyu, P., He, J., Wang, B.: Automatic organ and pan-cancer segmentation in abdomen ct: the flare 2023 challenge. arXiv preprint arXiv:2408.12534 (2024) 6
18. Ma, J., Zhang, Y., Gu, S., Zhu, C., Ge, C., Zhang, Y., An, X., Wang, C., Wang, Q., Liu, X., Cao, S., Zhang, Q., Liu, S., Wang, Y., Li, Y., He, J., Yang, X.: Abdomenct-1k: Is abdominal organ segmentation a solved problem? IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(10), 6695–6714 (2022) 6
19. Pedrosa, J., Aresta, G., Ferreira, C., Atwal, G., Phoulady, H.A., Chen, X., Chen, R., Li, J., Wang, L., Galdran, A., et al.: Lndb challenge on automatic lung cancer patient management. Medical Image Analysis **70**, 102027 (2021) 6
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241 (2015) 1

21. Roth, H.R., Xu, Z., Tor-Díez, C., Jacob, R.S., Zember, J., Molto, J., Li, W., Xu, S., Turkbey, B., Turkbey, E., et al.: Rapid artificial intelligence solutions in a pandemic—the covid-19-20 lung ct lesion segmentation challenge. Medical Image Analysis **82**, 102605 (2022) 6
22. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M., Golia-Pernicka, J., Heckers, S.H., Jarnagin, W.R., McHugo, M.K., Napel, S., Vorontsov, E., Maier-Hein, L., Cardoso, M.J.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019) 6
23. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023) 6
24. Xu, Z., Escalera, S., Pavão, A., Richard, M., Tu, W.W., Yao, Q., Zhao, H., Guyon, I.: Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. Patterns **3**(7), 100543 (2022) 10
25. Yan, K., Wang, X., Lu, L., Summers, R.M.: Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. Journal of Medical Imaging **5**(3), 036501–036501 (2018) 6
26. Yang, G., Yu, S., Yang, J., Zhang, Y., et al.: Multi-scale cnn for tumor detection in medical images. International Journal of Computer Assisted Radiology and Surgery **16**(6), 1137–1147 (2021) 2
27. Yushkevich, P.A., Gao, Y., Gerig, G.: Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 3342–3345 (2016) 6
28. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 84–92 (2016) 2

Table 7. Checklist Table. Please fill out this checklist table in the answer column.

Requirements	Answer
A meaningful title	Yes
The number of authors ($\leq$6)	6
Author affiliations and ORCID	Yes
Corresponding author email is presented	Yes
Validation scores are presented in the abstract	Yes
Introduction includes at least three parts: background, related work, and motivation	Yes
A pipeline/network figure is provided	1 2 3
Pre-processing	3
Strategies to improve model inference	4 5
Post-processing	6
The dataset and evaluation metric section are presented	7
Environment setting table is provided	1
Training protocol table is provided	2
Ablation study	7 8
Efficiency evaluation results are provided	3
Visualized segmentation example is provided	4 5
Limitation and future work are presented	Yes
Reference format is consistent.	Yes