# A Lightweight and Effective nnU-Net Framework for Whole-Body Pan-Cancer Segmentation

Hui Meng[1], Lusha Li[1], Pengyu Zhu[1], Yinlong Wang[1], Hongyu Liu[1], and Zhe Chen[1]

School of Intelligent Science and Technology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
huimeng@ucas.ac.cn

**Abstract.** Despite recent advancements medical image segmentation, achieving robust pan-cancer segmentation across diverse anatomical regions remains a significant challenge. To address challenges such as tumor diversity, data heterogeneity, and high computational demands, we explored a more streamlined and efficient approach for whole-body tumor segmentation based on nnU-Net. Specifically, we designed a lightweight 4-layer network architecture, optimized patch size and spacing during preprocessing, and developed a fine-tuned training strategy with extended epochs. Our method achieved an average score of 53.48% and 43.84% for the lesion DSC and NSD on the MICCAI FLARE 2025 challenge public validation dataset and the average running time and area under GPU memory-time curve are 17.8s and 64115MB, respectively.

**Keywords:** Pan-Cancer Segmentation · nnU-Net · Deep learning.

## 1 Introduction

Accurate whole-body cancer segmentation in 3D CT scans is essential for precise oncology diagnosis, treatment planning, and therapy monitoring. While recent advances in medical image segmentation—particularly for abdominal cancers through initiatives like FLARE—have shown promise, robust pan-cancer segmentation across diverse anatomical regions remains a significant challenge. FLARE 2025 Task 1 addresses this gap with an expansive dataset of over 10,000 CT scans spanning heterogeneous cancer types, requiring automated methods capable of delineating tumors throughout the entire body. This task faces multifaceted challenges:

- Tumor diversity: Lesions vary widely in morphology and location. Metastatic tumors often appear irregular and diffusely distributed, while primary tumors exhibit organ-specific characteristics (e.g., lung nodules versus liver masses), complicating the development of a unified segmentation approach.
- Data inconsistency: Multi-source training data suffers from discrepancies in voxel spacing and scan dimensions. Public datasets' organ-centric focus further widens the domain gap, hindering generalization to whole-body contexts.

- Dataset partial labeling: Only primary lesions are marked in each case, while other lesions (e.g., metastatic lesions) may remain unlabeled. Consequently, models may struggle to identify lesions beyond the primary tumor.
- Computational demands: Whole-body CT volumes—comprising thousands of slices—pose substantial processing burdens, making real-time analysis challenging.

Current pan-cancer segmentation methods primarily build upon SAM and nnU-Net architectures.While SAM offers zero-shot segmentation capability and can be adapted to medical image segmentation through fine-tuning on medical datasets [26,30], these approaches require computationally intensive training with massive parameters on 3D CT volumes. Moreover, they rely on manual input of points, boxes, or text prompts to generate target masks, leading to a semi-automated workflow. Some methods have improved upon SAM to achieve pan-cancer segmentation. Lin et al. [14] enhanced local features and introduced a task-indicator prompt encoder, extending the SAM-based model into a pan-cancer automatic segmentation model. Gao et al. [6] integrated the strong feature extraction capability of SAM, the powerful auto-configuration design capability of nnU-Net, and dynamic convolution to enhance the model's performance in pan-cancer segmentation. Despite these innovations, SAM-based pan-cancer adaptations exhibit poor cross-cancer generalization and suffer from high GPU memory consumption and prolonged inference times during deployment.

nnU-Net [12] has emerged as the most widely used method in medical image segmentation due to its self-configuring pipeline and robust performance in organ-specific tasks. Researchers have adapted this framework for pan-cancer segmentation through key innovations. Bai et al. [2] treated all lesions as a single class and directly training a universal lesion segmentation model using large image spacing and input volumes. Luo et al. [15] designed a cropping strategy based on the maximum connected component and an organ-interference segmentation method to address the issue where training models often misclassify most organ regions as tumors.

Given its robustness, flexibility, and efficiency in medical image segmentation, nnU-Net is adopted as the basic framework in this study. We propose several methods to enhance the efficiency and accuracy of pan-cancer segmentation:

- Lightweight Model Architecture: A streamlined 4-layer nnU-Net variant is proposed, significantly reducing model complexity and memory footprint while effectively preserving tumor segmentation capability.
- Data-Adaptive Preprocessing Optimization: Reconstruction of target patch size and spacing during preprocessing substantially increases the anatomical coverage per patch, enhancing contextual information capture.
- Training Strategy Optimization: Fine-tuning the training epochs ensures sufficient learning without overfitting, improving the model's generalization performance.

# 2 Method

## 2.1 Preprocessing

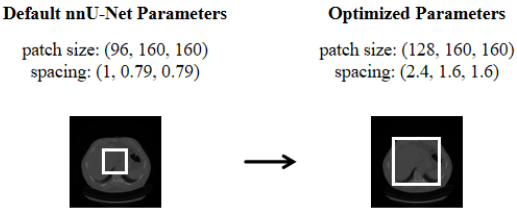


**Fig. 1.** Data-adaptive preprocessing with optimized patch size and spacing to capture extended contextual information.

- Statistical analysis: Flare2025 Task 1 comprises 10,328 annotated CT volumes aggregated from multiple public datasets. Among these, the annotations for the DeepLesion5K subset were generated by MedSAM2, while the remaining 5,328 cases originate from various other datasets. Most of these datasets are single-tumor annotation datasets, with 7 of them involving lung tumors or lesions—and the lung-related datasets also constitute the largest proportion of samples. The lesion volumes exhibit substantial variability, ranging from approximately 11 $mm^3$ (the smallest lesion from the LIDC-IDRI dataset) to 7,375,883 $mm^3$ (the largest lesion from the KiTS23 dataset). Note that this analysis excludes multi-lesion datasets such as PSMA.
- Data Selection: Due to the critical impact of annotation accuracy on segmentation performance, we exclusively utilized 5,328 rigorously verified cases for training.
- Resampling: Figure 1 demonstrates our optimization of the sampling method. The default nnU-Net parameters specify a patch size of 96×160×160 and a normalized spacing of 1.0×0.79×0.79 $mm^3$. We adjusted these to a patch size of 128×160×160 and a normalized spacing of 2.4×1.6×1.6 $mm^3$. This modification expanded the spatial coverage from 96×126×126 $mm^3$ to 307×256×256 $mm^3$, while simultaneously reducing GPU memory consumption and accelerating training speed.
- Normalization: Hounsfield Unit (HU) values were clipped at the 0.05% and 99.5% percentiles computed exclusively from foreground voxels across the entire training dataset, followed by z-score normalization using the global mean and standard deviation of the foreground intensities.

## 2.2 Proposed Method

Figure 2 illustrates the architecture of our model. Our method builds upon the state-of-the-art nnU-Net framework, streamlining it to a 4-layer architec-

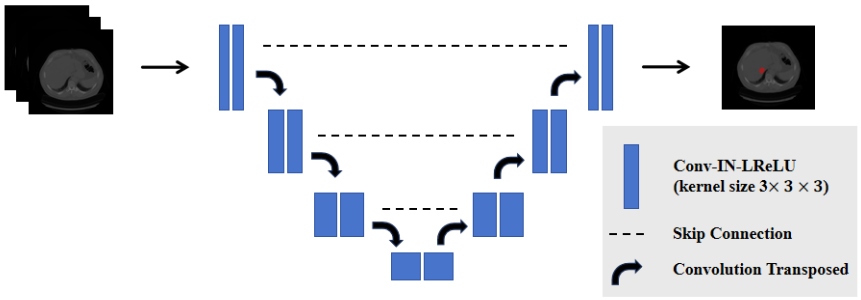


**Fig. 2.** A streamlined 4-layer network architecture that significantly reduces computational complexity and memory footprint.

ture, significantly reducing GPU memory consumption and computational time. While deeper networks may marginally improve accuracy, they are prone to overfitting with diminishing returns, especially when dealing with partially labeled data. During preprocessing, we resampled the CT scans to a resolution of $2.4\times1.6\times1.6$ mm$^3$. At this lower resolution, the high-resolution feature extraction stages in deeper networks contribute minimally to final segmentation quality. The 4-layer architecture optimally aligns with low-resolution feature extraction requirements, eliminating redundant computations while maintaining an ideal efficiency-accuracy balance.

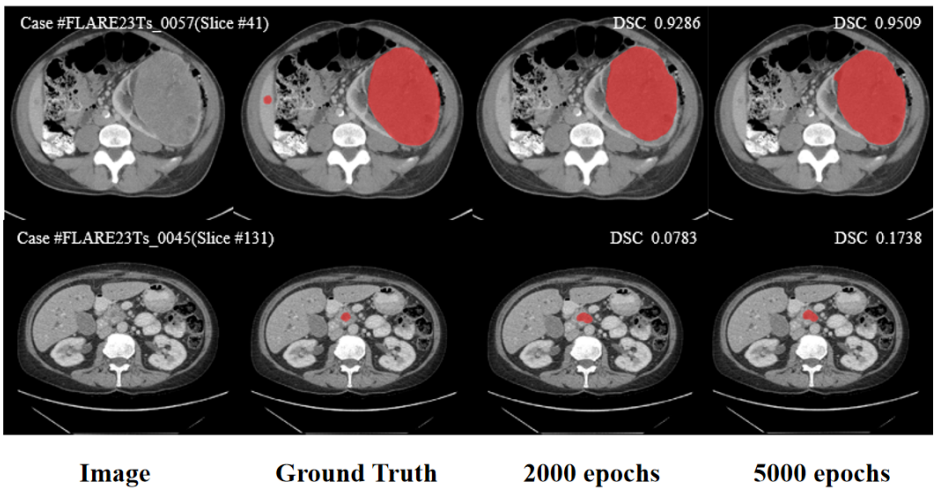


**Fig. 3.** Comparison chart of segmentation results between training for 2000 epochs and training for 5000 epochs.

We employ a fine-tuned approach to determine the number of training epochs. The initial training is set to 2,000 epochs; however, we observe that this proves insufficient, as all metrics indicate further potential for improvement. Consequently, the number of epochs is systematically increased, with validation performed every 500 additional epochs. The training curves and validation metrics are closely monitored to prevent overfitting. Based on continual monitoring, 5,000 epochs are identified as the optimal setting, achieving the best performance without signs of overfitting. Figure 3 demonstrates superior segmentation results from 5,000 training epochs versus 2,000 epochs.

Loss function: we use the summation between Dice loss and cross-entropy loss because compound loss functions have proven robust in various medical image segmentation tasks [16].

## 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 [4], LiTS [3], MSD [24], KiTS [9,11,10], autoPET [8,7], TotalSegmentator [25], and AbdomenCT-1K [21], FLARE 2023 [20], DeepLesion [28], COVID-19-CT-Seg-Benchmark [18], COVID-19-20 [23], CHOS [13], LNDB [22], and LIDC [1]. The training set includes more than 10000 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 [29], nnU-Net [12], MedSAM [17,19], and Slicer Plugins [5,19].

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.

**Training protocols** Our training protocols followed the default settings of nnU-Net.

No specific strategy is employed for processing unlabeled images and partial labels. We use only the 5,328 annotated cases for training, without utilizing

**Table 1.** Development environments and requirements.

| | |
|---|---|
| System | Ubuntu 20.04.6 LTS |
| CPU | 2x AMD EPYC 7542 32-Core Processor |
| RAM | 512GB |
| GPU (number and type) | 8x NVIDIA GeForce RTX 4090 |
| CUDA version | 12.6 |
| Programming language | Python 3.10 |
| Deep learning framework | Pytorch (Torch 2.7.1) |
| Specific dependencies | nnU-Net 2.6.2 |
| Code | |

unlabeled data or labels generated by MedSAM2. The default data augmentation techniques and patch sampling strategies from nnU-Net are employed, with no additional modifications.

**Table 2.** Training protocols.

| | |
|---|---|
| Network initialization | He |
| Batch size | 2 |
| Patch size | 128×160×160 |
| Total epochs | 5000 |
| Optimizer | SGD |
| Initial learning rate (lr) | 0.001 |
| Lr decay schedule | poly |
| Training time | 62 hours |
| Loss function | Dice plus CE |
| Number of model parameters | 45.2M[1] |
| Number of flops | 1665.8G[2] |
| $CO_2$eq | 34.14KG[3] |

## 4 Results and discussion

### 4.1 Quantitative results on validation set

Quantitative results are shown in Table 3. On the public validation set, our method achieved a mean Dice of 53.48% and NSD of 43.84%. However, no Dice or NSD metrics were obtained for the hidden validation and test sets.

### 4.2 Qualitative results on validation set

Figure 4 shows two examples with good segmentation results and two examples with bad segmentation results in the validation set.The first two rows demon-

**Table 3.** Quantitative evaluation results.

| Methods | Public Validation | | Online Validation | | Testing | |
|---|---|---|---|---|---|---|
| | DSC(%) | NSD(%) | DSC(%) | NSD(%) | DSC(%) | NSD (%) |
| Algorithm1 | 53.48 ± 38.06 | 43.84 ± 35.29 | - | - | | |

**Table 4.** 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.

| Case ID | Image Size | Running Time (s) | Max GPU (MB) | Total GPU (MB) |
|---|---|---|---|---|
| 0001 | (512, 512, 55) | 21.18 | 6904 | 71904 |
| 0051 | (512, 512, 100) | 13.75 | 7068 | 49261 |
| 0017 | (512, 512, 150) | 13.94 | 6878 | 48888 |
| 0019 | (512, 512, 215) | 14.06 | 6888 | 49187 |
| 0099 | (512, 512, 334) | 25.92 | 6920 | 111241 |
| 0063 | (512, 512, 448) | 14.03 | 6874 | 48860 |
| 0048 | (512, 512, 499) | 20.53 | 6869 | 69476 |
| 0029 | (512, 512, 554) | 18.74 | 6888 | 64100 |

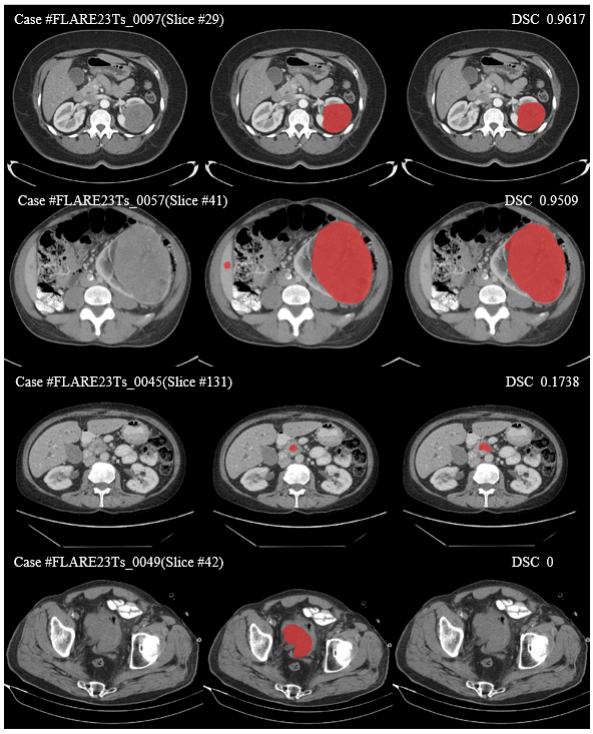


**Fig. 4.** Two examples with good segmentation results and two examples with bad segmentation results in the validation set.

strate successful examples where the tumor regions are accurately segmented, achieving Dice scores above 95%. The latter two rows show failure examples with extremely low segmentation precision, including instances where the Dice score drops to zero. When the tumor region is particularly small, the model may misclassify other areas as tumors (as shown in Row 3), leading to suboptimal segmentation. Additionally, when low contrast exists between tumor regions and surrounding tissues, the model may completely fail to segment the tumor areas (as seen in Row 4), mistakenly classifying the image as healthy and resulting in extremely poor segmentation performance.

### 4.3 Segmentation efficiency results on validation set

Table 4 shows running times and VRAM utilization for the eight representative cases from the public validation set. Average runtime is around 18 seconds per scan, with peak GPU usage below 7.1 GB, showing efficient computational performance for segmentation tasks.

### 4.4 Results on final testing set

This is a placeholder. We will send you the testing results during MICCAI.

### 4.5 Limitation and future work

Our method is exclusively trained on labeled datas without employing any semi-supervised techniques , which constrains the model's generalization capability. We use a streamlined four-layer nnU-Net for training; however, by increasing the number of training epochs, we gain little advantage in terms of training parameters, training duration, and computational load. Additionally, the model still struggles to accurately segment extremely small tumor regions as well as tumors with low contrast and poor resolution, indicating considerable room for improvement in segmentation accuracy.

## 5 Conclusion

In this study, we propose a lightweight and modified nnU-Net framework for automatic whole-body tumor segmentation in CT images. By reducing the network depth to four layers, optimizing patch size and spacing during preprocessing, and carefully extending training epochs to 5,000, we developed a model that balances computational efficiency and segmentation performance. We validated the method on the MICCAI FLARE 2025 Challenge dataset, and both quantitative and qualitative results demonstrate its significant efficacy in the whole-body pan-cancer segmentation task.

**Acknowledgements** The authors of this paper declare that the segmentation method they implemented for participation in the FLARE 2025 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 [27] for hosting the challenge platform. This work was supported by the National Natural Science Foundation of China under grant numbers 62303127, 62273009.

## 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) 5
2. Bai, X., Chen, Z., Zhang, S., Liu, Q., Liu, X., Xia, Y.: Reframing universal lesion segmentation as a large-organ lesion segmentation task. In: Ma, J., Wang, B. (eds.) Fast, Low-Resource, Accurate Robust Organ and Pan-cancer Segmentation. pp. 23–33. Springer Nature Switzerland, Cham (2026) 2
3. 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) 5
4. 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) 5

5. 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) 5
6. Gao, R., Li, X., Wang, J., Zheng, S.: Pan-cancer segmentation in ct scans based on dynamic convolution with nnsam. In: Ma, J., Wang, B. (eds.) Fast, Low-Resource, Accurate Robust Organ and Pan-cancer Segmentation. pp. 34–42. Springer Nature Switzerland, Cham (2026) 2
7. 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) 5
8. 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) 5
9. 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) 5
10. 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) 5
11. 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) 5
12. 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) 2, 5
13. 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) 5

14. Lin, X., Wang, Z., Shang, C., Shi, J., Yan, Z., Yu, L.: Segment any cancer in ct scans through equipping sam with cross-slice interaction and indicator prompt. In: Ma, J., Wang, B. (eds.) Fast, Low-Resource, Accurate Robust Organ and Pan-cancer Segmentation. pp. 54–64. Springer Nature Switzerland, Cham (2026) 2
15. Luo, Y., Liu, Y., Liu, W., Dai, J., Xiao, X., Fang, G.: Crop and segment: Efficient whole body pan-cancer segmentation. In: Ma, J., Wang, B. (eds.) Fast, Low-Resource, Accurate Robust Organ and Pan-cancer Segmentation. pp. 1–12. Springer Nature Switzerland, Cham (2026) 2
16. Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L.: Loss odyssey in medical image segmentation. Medical Image Analysis **71**, 102035 (2021) 5
17. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024) 5
18. 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) 5
19. Ma, J., Yang, Z., Kim, S., Chen, B., Baharoon, M., Fallahpour, A., Asakereh, R., Lyu, H., Wang, B.: Medsam2: Segment anything in 3d medical images and videos. arXiv preprint arXiv:2504.03600 (2025) 5
20. 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) 5
21. 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) 5
22. 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) 5
23. 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) 5
24. 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) 5
25. 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) 5
26. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation (2023), https://arxiv.org/abs/2304.12620 2
27. 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) 9

28. 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) 5
29. 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) 5
30. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation (2023), https://arxiv.org/abs/2304.13785 2

**Table 5.** Checklist Table. Please fill out this checklist table in the answer column.

| Requirements | Answer |
|---|---|
| A meaningful title | Yes |
| The number of authors (≤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 | Figure 2 |
| Pre-processing | Page 3 |
| Strategies to use the partial label | Page 3 |
| Strategies to use the unlabeled images. | Page 3 |
| Strategies to improve model inference | Page 3,4 |
| Post-processing | no |
| The dataset and evaluation metric section are presented | Page 4,5 |
| Environment setting table is provided | Table 1 |
| Training protocol table is provided | Table 2 |
| Ablation study | no |
| Efficiency evaluation results are provided | Table 4 |
| Visualized segmentation example is provided | Figure 4 |
| Limitation and future work are presented | Yes |
| Reference format is consistent. | Yes |