Lite ENSAM: a lightweight cancer segmentation model for 3D Computed Tomography

Agnar Martin Bjørnstad[0009−0005−4207−6278],
Elias Stenhede[0009−0005−2654−4553], and Arian Ranjbar[0000−0002−0422−2255]

Faculty of Medicine, University of Oslo
Medical Technology & E-health, Akershus University Hospital,
1478 Lørenskog, Norway
arian.ranjbar@medisin.uio.no

Abstract. Accurate tumor size measurement is a cornerstone of evaluating cancer treatment response. The most widely adopted standard for this purpose is the Response Evaluation Criteria in Solid Tumors (RECIST) v1.1, which relies on measuring the longest tumor diameter in a single plane. However, volumetric measurements have been shown to provide a more reliable assessment of treatment effect. Their clinical adoption has been limited, though, due to the labor-intensive nature of manual volumetric annotation. In this paper, we present Lite ENSAM, a lightweight adaptation of the ENSAM architecture designed for efficient volumetric tumor segmentation from CT scans annotated with RECIST annotations. Lite ENSAM was submitted to the MICCAI FLARE 2025 Task 1: Pan-cancer Segmentation in CT Scans, Subtask 2, where it achieved a Dice Similarity Coefficient (DSC) of 76.06% and a Normalized Surface Dice (NSD) of 78.99%, with an average total RAM time of 50,556 MB s and an average inference time of 14.39 seconds on CPU on the public validation dataset.

Keywords: Computed Tomography · Multimodal · Tumor Segmentation

1 Introduction

Imaging plays a central role in assessing tumor response during cancer treatment, guiding oncologists in deciding whether to continue, adapt, or interrupt therapy. The most widely adopted framework for this purpose is RECIST v1.1 [5], valued for its simplicity, cost-effectiveness, and intuitiveness [7]. Under RECIST, tumor size is quantified by measuring the longest diameter in a single plane of a scan.

While RECIST remains the clinical standard, volumetric tumor measurements have been shown to provide more sensitive and prognostic information. For example, volumetric analysis has been reported to better predict overall survival in lung cancer patients [10] and to more accurately and earlier predict the pathological response to neoadjuvant chemotherapy in breast cancer than clinical assessment alone [17]. Despite these advantages, the adoption of volumetric

measurements in clinical practice has been limited by the need for labor-intensive manual annotation of tumor volumes.

Advancements in computer vision offer a path forward by enabling automated volumetric lesion segmentation, thereby reducing annotation burden and potentially improving both the speed and accuracy of tumor response assessment [15]. To advance this goal, the MICCAI FLARE 2025 challenge focuses on efficient and accurate medical image analysis under constrained computational resources. In particular, Task 1, Subtask 2 addresses the problem of converting RECIST annotations into full volumetric segmentations, bridging the gap between the widely used RECIST standard and clinically valuable volumetric measurements.

This task presents several challenges. Volumetric CT scans are inherently large, while inference must be performed under strict computational constraints. Models are limited to CPU execution with a maximum of 8 GB RAM, which requires highly memory- and compute-efficient designs. Furthermore, tumors vary widely in shape, size, and anatomical location, making robust generalization across diverse cases particularly difficult.

1.1 Related work

U-Net–style architectures have long been dominant in medical image segmentation, consistently underpinning many top-performing models in major benchmarks. The original U-Net [31] demonstrated exceptional segmentation accuracy even with limited training data, and its encoder–decoder structure became a foundational paradigm. SegResNet [27], a 3D U-Net variant, was developed for brain tumor segmentation in multimodal MRI and notably secured first place in the BraTS 2018 challenge [2]. Similarly, nnU-Net [18] emerged as a highly influential, self-configuring framework based on U-Net. It automatically adapts architecture and training strategies based on dataset characteristics. In the 2021 Kidney Tumor Segmentation (KiTS21) challenge [12], all top three solutions were U-Net–based models using nnU-Net. And in FLARE24 (whole-body cancer segmentation in CT scans), the winning solution employed a 3D U-Net architecture [16]. Collectively, these findings display the effectiveness of U-Net across imaging modalities.

ENSAM (Equivariant, Normalized, SAM in 3D) [34] is a foundation model for interactive 3D medical image segmentation. In the CVPR 2025: Foundation Models for Interactive 3D Biomedical Image Segmentation challenge[1], ENSAM performed the best among models without external pretraining, outperforming the majority of baseline methods. ENSAM adopts a 3D U-Net-inspired design based on SegResNet [28].

1.2 Objective and contribution

In this paper, we adapt the ENSAM architecture for the task of tumor segmentation in 3D CT images. In particular, we further improve performance and

[1] https://www.codabench.org/competitions/7149/

introduce a lightweight variant, named Lite ENSAM, optimized for efficient inference on CPU.

2 Method

Lite ENSAM is architecturally similar to ENSAM, consisting of a 3D U-Net, augmented with a SAM-style attention mechanism [20] linking the network's deepest layer to the embeddings of RECIST annotated tumors. The attention mechanism is implemented using a normalized transformer architecture [21] with Lie Rotational Positional Encoding (LieRE) [29]. An overview of the architecture is shown in Figure 1, and the following section provides a brief description of the model.

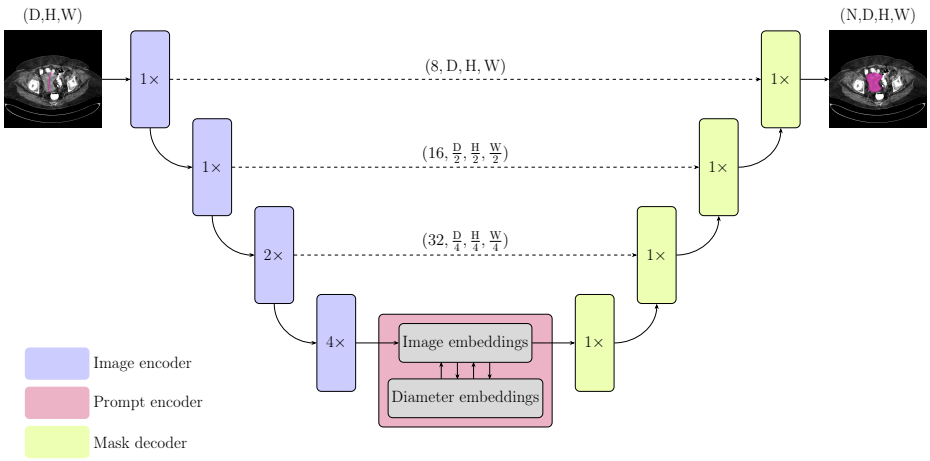

Fig. 1. Network architecture of Lite ENSAM, consisting of three main components: image encoder, prompt encoder, and mask decoder. The diameter markings are incorporated via cross-attention between image embeddings and diameter embeddings in the bottom part of the U-net.

2.1 Network architecture

Image encoder The image encoder is composed of multiple blocks that operate at progressively lower resolutions. Each block consists of an instance normalization layer, followed by two repetitions of a 3D convolutional layer and ReLU activation, and a skip connection. Downsampling between blocks is performed using a convolutional layer with stride 2.

Encoding of RECIST annotation In the MICCAI FLARE 2025 Task 1, Subtask 2, tumors are annotated according to RECIST using their longest in-plane diameter on the axial (z) slice with the largest cross-sectional area. Representing RECIST annotations as dense volumetric masks would be both information-sparse and memory-intensive, since the marking itself provides no spatial context beyond its two endpoints. Instead, we extract and encode only the diameter endpoints, which contain all available localization information.

The interaction between the diameter and image embeddings takes place at the bottleneck of the U-Net, where cross-attention with Lie Rotational Positional Encoding is applied. For details of the cross-attention mechanism and relative position encoding, we refer to the ENSAM paper [34].

Image decoder The mask decoder mirrors the image encoder, but employs a single residual block at each upsampling stage. Skip-connection activations are concatenated along the channel dimension and processed by a 3D residual block, followed by trilinear upsampling. The final layer outputs logits with the same spatial dimensions as the input, representing the predicted segmentation mask. Voxels with logits greater than zero are classified as tumor.

Modifications to ENSAM To adapt the ENSAM model for this competition, several modifications were required. First, the parsing and processing pipelines had to be updated. ENSAM was originally designed for interactive medical image segmentation, where it received a bounding box or user click for each object of interest to indicate its location. Based on these inputs, the model produced object segmentations and subsequently accepted refinement clicks, each pointing to the region of largest error for a given object. This interactive process allowed the model to iteratively improve its segmentation. In contrast, the current competition provides CT volumes together with diameter markings for each tumor. Consequently, ENSAM was modified to accept diameter points as an input modality and to operate in a non-interactive setting, with the refinement stage removed.

As inference is performed on a CPU with a maximum of 8 GB RAM, the model must operate under substantially more restrictive memory and computational constraints. To accommodate this, Lite ENSAM was designed to be more efficient in both memory footprint and compute. Specifically, the number of output filters in the image decoder and the embedding dimension were reduced by half, and the maximum patch size of the input volume to the decoder was limited to one-quarter of that used in the original ENSAM.

Finally, the post-processing step was re-implemented to be more memory- and compute-efficient, which was essential since the original implementation exceeded the available memory for certain inputs. To address this, the number of full-volume allocations was reduced by restructuring the computation of intermediate tensors, thereby reducing both memory footprint and computational overhead. Additionally, to verify the presence of all classes in the final prediction, we replaced torch.unique with torch.bincount, which avoids sorting and

is therefore faster. On CPU, these optimizations yielded a 30× speedup of the post-processing stage on the validation set, reducing its average runtime to 0.2 seconds on the hardware specified in Table 1.

2.2 Pre-processing

All volumes were min-max normalized to range between 0 and 1 during training and inference. Besides normalization, preprocessing was limited to cropping, resizing and flipping volumes, ensuring training can run within the specified VRAM limits. A maximum size of $128 \times 128 \times 128$ was used during training and inference. Anisotropy was not explicitly considered, and unlabelled or pseudolabelled data were not used for training.

2.3 Post-processing

Each volume may contain multiple annotated tumors, all of which are included in the ground truth segmentation. To reduce the impact of low-confidence predictions for individual tumors, the model is required to predict all annotated tumors in the final output. Specifically, if all logits for a given tumor are below zero, the logits within the sphere defined by its RECIST diameter are incrementally increased by an exponential offset until at least one voxel is classified as belonging to that tumor.

3 Experiments

3.1 Dataset and evaluation measures

Data was provided by the MICCAI FLARE 2025 challenge, and 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 [33], KiTS [14, 11, 13], autoPET [8, 9], TotalSegmentator [35], and AbdomenCT-1K [25], FLARE 2023 [24], DeepLesion [37], COVID-19-CT-Seg-Benchmark [26], COVID-19-20 [32], CHOS [19], LNDB [30], and LIDC [1]. The training set includes more than 10,000 abdomen CT scans, where 2,200 CT scans with partial labels and 1,800 CT scans are 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 [38], nnU-Net [18], MedSAM [22, 23], and Slicer Plugins [6, 23].

The evaluation metrics encompass two accuracy measures: DSC and NSD, alongside two efficiency measures: running time and area under the CPU memory-time curve. These metrics collectively contribute to the ranking computation. Furthermore, the running time and CPU 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 hardware.

Component	Specification
System	Debian 12
CPU	Intel(R) Core(TM) i9-14900KF
RAM	2×48 GB; 4800 MT/s
GPU	NVIDIA GeForce RTX 5090 32 GB
CUDA version	12.8
Programming language	Python 3.12
Deep learning framework	PyTorch 2.7.0, Torchvision 0.22.0

Training protocols The provided dataset[2] was used exclusively for model training and model selection; no partial- or unlabeled data were used. The samples were categorized by source, and uniformly sampled across categories to mitigate class imbalance. To concentrate computational resources on relevant structures, training volumes were randomly cropped around annotated regions with a variable margin of 1–64 voxels. Volumes exceeding a predefined size threshold were downscaled via max pooling to fit within GPU memory. Given the variability in sample dimensions, the batch size was fixed at 1. To ensure compatibility with the network architecture, zero-padding was applied so that all dimensions were divisible by 8. Each axis was randomly flipped with a probability of 50%. An overview of training configuration is provided in Table 2. Consistent with ENSAM, Lite ENSAM was trained using a single GPU.

4 Results and discussion

4.1 Quantitative results on validation set

Quantitative results are shown in Table 3. On the public validation dataset, our method achieved a DSC of 76.06% and an NSD of 78.99%.

4.2 Qualitative results on validation set

Figure 2 presents predictions for five representative validation cases, corresponding to the 5th, 25th, 50th, 75th, and 95th percentiles of the Dice score relative to the ground truth. The slice indices are selected to match the plane of the first RECIST marker. In the cases with the lowest Dice score, the model seems to undersegment the tumors.

[2] https://huggingface.co/datasets/FLARE-MedFM/FLARE-Task1-PancancerRECIST-to-3D

Table 2. Training protocols.

Parameter	Specification
Batch size	1
Patch size	Variable
Total epochs	75
Optimizer	Muon and AdamW
Initial learning rate (lr)	0.002
Training time	35h
Loss function	Soft Dice + 2·BCE
Number of model parameters	1.3M
Number of flops	29G
CO_2eq	0.214 Kg

Table 3. Quantitative evaluation results.

Methods	Public Validation		Online Validation		Testing	
	DSC(%)	NSD(%)	DSC(%)	NSD(%)	DSC(%)	NSD (%)
Lite ENSAM (ours)	76.06 ± 16.31	78.99 ± 19.05	-	-	-	-

Table 4. Quantitative evaluation of segmentation efficiency in terms of the running them and RAM use. Total RAM use denotes the area under the RAM-Time curve.

Case ID	Image Size	Running Time (s)	Max RAM (MB)	Total RAM (MB s)
0001	(55, 512, 512)	17.22	3,994	55,847
0011	(100, 512, 512)	14.80	4,761	52,143
0021	(152, 512, 512)	14.25	5,026	51,087
0031	(215, 512, 512)	13.60	4,686	47,405
0041	(157, 512, 512)	14.27	4,689	50,048
0051	(171, 512, 512)	14.17	4,397	49,788
0061	(147, 512, 512)	14.49	5,208	52,709
0071	(240, 512, 512)	14.95	5,563	54,723
0081	(103, 512, 512)	13.18	4,559	46,316
0091	(137, 512, 512)	13.00	4,541	45,498
Average		14.39	4,742	50,556

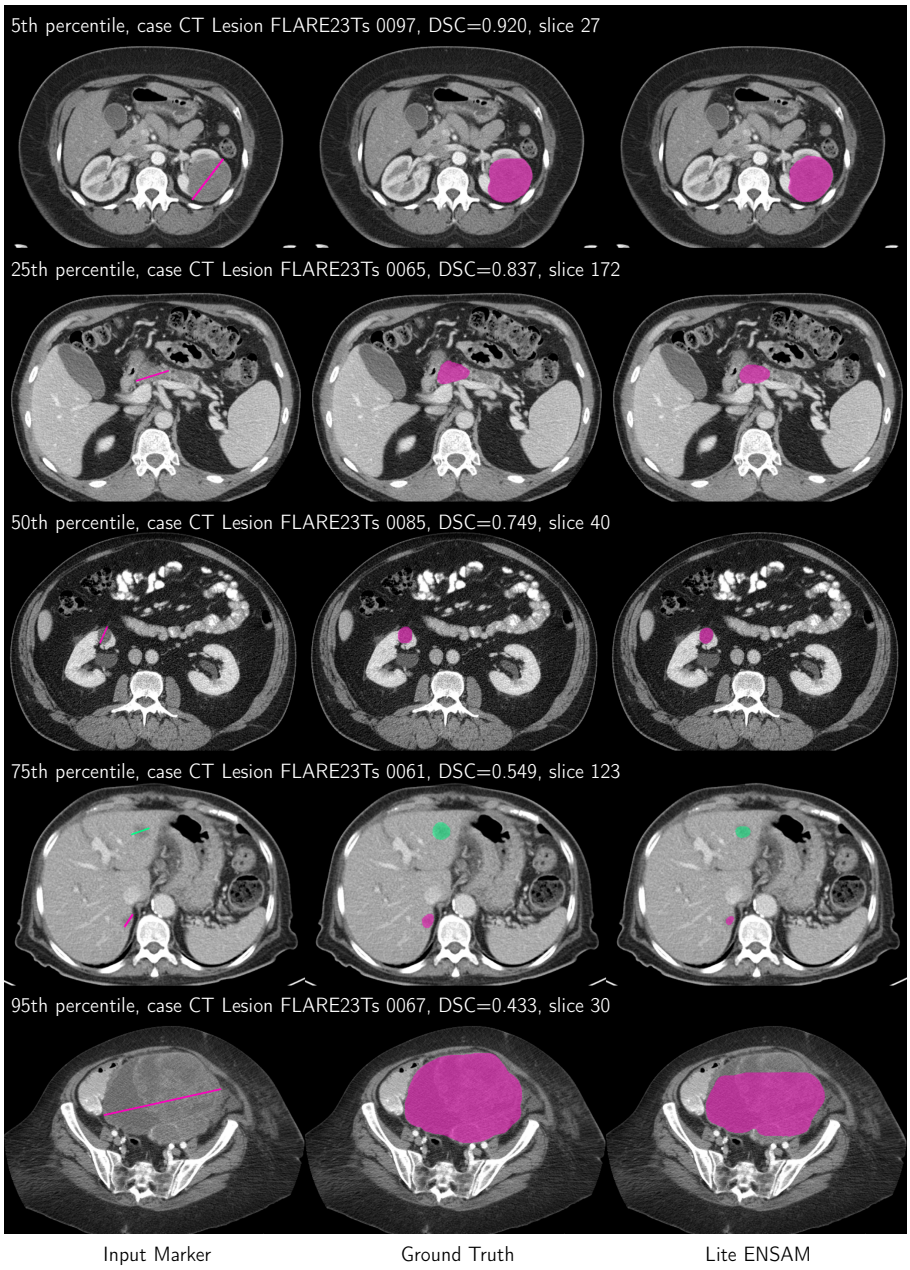

Fig. 2. Example slices from five volumes in the validation set. The volumes were chosen based on their Dice scores, corresponding to the 5th, 25th, 50th, 75th, and 95th percentiles. For each volume, the slice aligned with an input marker from class 1 was selected.

4.3 Segmentation efficiency results on validation set

Table 4 shows the inference efficiency on a selection of validation cases. For all the cases, there is a significant margin to the maximum RAM threshold of 8GB. Inference on healthy CT scans were not performed as they do not have any diameter markings.

4.4 Limitation and future work

While Lite ENSAM demonstrates competitive performance under strict resource constraints, several limitations remain that suggest directions for future work. First, the model was trained with constrained compute and exclusively on the provided training data. Training on a larger and more diverse dataset would likely improve generalization and segmentation accuracy. Second, the impact of additional data augmentation other than axis flips was not explored. Such augmentations could further enhance robustness to variations in tumor appearance. Finally, to strengthen its clinical relevance, future work should extend evaluation beyond CT to other imaging modalities widely used in treatment response assessment, such as Magnetic Resonance Imaging. This would increase the model's applicability in clinical practice and enable validation of its performance across modalities.

5 Conclusion

In this paper, we introduced Lite ENSAM, a lightweight version of the ENSAM architecture for tumor segmentation in 3D CT scans with RECIST annotations. Lite ENSAM was specifically developed to enable inference on CPU under a strict memory constraint of 8 GB RAM. On the public validation dataset, it achieved a DSC of 76.06% and an NSD of 78.99%. This work demonstrates that the ENSAM architecture can be effectively adapted to low-compute, low-memory environments while maintaining competitive segmentation performance, and further validates its effectiveness when trained with diameter-based annotations.

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 [36] for hosting the challenge platform. The authors express their appreciation to Novartis Norge AS and Akershus University Hospital for funding this work.

Disclosure of Interests

The authors declare no competing interests relevant to this work.

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)
2. Bakas, S. *et al.*: Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, (2019). https://doi.org/10.48550/arXiv.1811.02629. arXiv:1811.02629 [cs].
3. Bilic, P. *et al.*: The Liver Tumor Segmentation Benchmark (LiTS). Medical Image Analysis **84**, 102680 (2023)
4. Clark, K. *et al.*: The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of Digital Imaging **26**(6), 1045–1057 (2013)
5. Eisenhauer, E. *et al.*: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). European Journal of Cancer **45**(2), 228–247 (2009). https://doi.org/10.1016/j.ejca.2008.10.026
6. 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)
7. Fournier, L. *et al.*: Twenty Years On: RECIST as a Biomarker of Response in Solid Tumours an EORTC Imaging Group – ESOI Joint Paper. Frontiers in Oncology **11** (2022). https://doi.org/10.3389/fonc.2021.800547
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)
9. Gatidis, S. *et al.*: The autopet challenge: towards fully automated lesion segmentation in oncologic pet/ct imaging. Nature Machine Intelligence (2023)
10. Hayes, S., Pietanza, M., O'Driscoll, D., Zheng, J., Moskowitz, C., Kris, M., Ginsberg, M.: Comparison of CT volumetric measurement with RECIST response in patients with lung cancer. European journal of radiology **85**(3), 524–533 (2016). https://doi.org/10.1016/j.ejrad.2015.12.019
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)
12. Heller, N. *et al.*: The KiTS21 Challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase CT, (2023). https://doi.org/10.48550/arXiv.2307.01984. arXiv:2307.01984 [cs].
13. Heller, N. *et al.*: The kits21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase CT. arXiv preprint arXiv:2307.01984 (2023)
14. Heller, N. *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 **67**, 101821 (2021)
15. Hering, A. *et al.*: Improving assessment of lesions in longitudinal CT scans: a bi-institutional reader study on an AI-assisted registration and volumetric segmenta-

tion workflow. International Journal of Computer Assisted Radiology and Surgery **19**(9), 1689–1697 (2024). https://doi.org/10.1007/s11548-024-03181-4
16. Huang, Z.: Efficient Whole-Body Tumor Segmentation with a 5.6M Parameter 3D U-Net. (2024)
17. Hylton, N.M. *et al.*: Locally Advanced Breast Cancer: MR Imaging for Prediction of Response to Neoadjuvant Chemotherapy—Results from ACRIN 6657/I-SPY TRIAL. Radiology **263**(3), 663–672 (2012). https://doi.org/10.1148/radiol.12110748
18. 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)
19. Kavur, A.E. *et al.*: CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021)
20. Kirillov, A. *et al.*: Segment Anything. In: Proceedings of the International Conference on Computer Vision, pp. 4015–4026 (2023)
21. Loshchilov, I., Hsieh, C.-P., Sun, S., Ginsburg, B.: nGPT: Normalized Transformer with Representation Learning on the Hypersphere. In: Proceedings of the International Conference on Learning Representations (ICLR) (2025)
22. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024)
23. 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)
24. 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)
25. Ma, J. *et al.*: AbdomenCT-1K: Is Abdominal Organ Segmentation A Solved Problem? IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(10), 6695–6714 (2022)
26. Ma, J. *et al.*: Towards Data-Efficient Learning: A Benchmark for COVID-19 CT Lung and Infection Segmentation. Medical Physics **48**(3), 1197–1210 (2021)
27. Myronenko, A.: 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, pp. 311–320. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-11726-9_28
28. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: International MICCAI brainlesion workshop, pp. 311–320 (2018)
29. Ostmeier, S., Axelrod, B., Moseley, M.E., Chaudhari, A., Langlotz, C.: LieRE: Generalizing Rotary Position Encodings, (2025). https://doi.org/10.48550/arXiv.2406.10322. arXiv:2406.10322.
30. 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)
31. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, (2015). https://doi.org/10.48550/arXiv.1505.04597. arXiv:1505.04597 [cs].
32. 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)
33. Simpson, A.L. *et al.*: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019)
34. Stenhede, E., Bjørnstad, A.M., Ranjbar, A.: ENSAM: an efficient foundation model for interactive segmentation of 3D medical images. In: (2025)
35. Wasserthal, J. *et al.*: TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
36. 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)
37. 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)
38. 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)

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

Requirements	Answer
A meaningful title	Yes
The number of authors ($\leq$6)	3
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 1
Pre-processing	Page 5
Strategies to use the partial label	Page 6
Strategies to use the unlabeled images.	Page 6
Strategies to improve model inference	Page 4 and 5
Post-processing	Page 5
The dataset and evaluation metric section are presented	Page 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 2
Limitations and future work are presented	Yes
Reference format is consistent.	Yes