SwiftMedSAM: An Ultra-Lightweight Prompt-based Universal Medical Image Segmentation Model for Highly Constrained Environments

Youngbin Kong[1,2][0009−0002−6248−4269], Kwangtai Kim[2][0009−0009−1141−5985], Seoi Jeong[2][0009−0009−3047−2889], Kyu Eun Lee[3][0000−0002−2354−3599], and Hyoun-Joong Kong[2,4,5,⋆][0000−0001−5456−4862]

[1] Interdisciplinary Program in Bioengineering, Graduate School, Seoul National University, Seoul, Republic of Korea
[2] Department of Transdisciplinary Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[3] Department of Surgery, Seoul National University Hospital and College of Medicine, Seoul, Republic of Korea
[4] Department of Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
[5] Innovative Medical Technology Research Institute, Seoul National University Hospital, Seoul, Republic of Korea
gongcop7@snu.ac.kr

Abstract. Medical image segmentation is a crucial step for accurate diagnosis and treatment planning, as it provides quantitative information about anatomical structures and pathological lesions in various clinical scenarios. However, the existing methodologies have limitations in terms of their generalizability and computational efficiency. In this study, we propose SwiftMedSAM, an ultra-lightweight prompt-based general model, to enable efficient medical image segmentation even in resource-constrained environments. Based on LiteMedSAM, we significantly reduced the model size and computational complexity through the hyperparameter optimization of the image encoder and mask decoder components. The developed model shows remarkable segmentation performance across various imaging modalities and anatomical structures while enabling real-time inference in resource-limited computing environments. The experimental results demonstrate that SwiftMedSAM outperforms the existing methodologies in terms of the trade-off between accuracy and efficiency. SwiftMedSAM achieved a validation score of 0.75 on the validation dataset. Owing to its unprecedented generalizability and low computational cost, SwiftMedSAM is expected to enable high-quality medical image analysis in resource-constrained settings, thereby contributing to advancements in precision medicine and telemedicine.

Keywords: Medical Imaging Segmentation · Segment Anything · Deep Learning

⋆ corresponding author

1 Introduction

Medical image segmentation is a critical step in accurate diagnosis and treatment planning. It provides quantitative information about anatomical structures and pathological lesions in various clinical scenarios such as computer-aided diagnosis, surgical guidance, treatment monitoring, and patient follow-up. For example, accurate identification of the location, size, and boundaries of a tumor is essential for determining cancer staging, surgical planning, and radiation therapy. However, such segmentation tasks are highly complex and time-consuming, necessitating the development of automated high-performance segmentation models [34].

Driven by advancements in deep learning techniques, innovative achievements have been made in the field of medical image analysis in recent years, with significant progress in segmentation problems. Initially, transfer-learned CNN-based models such as U-Net [49] and V-Net [43] were predominant, and subsequently, the introduction of cutting-edge models such as Vision Transformer [15] and Swin Transformer [36] led to substantial accuracy improvements. However, most existing studies have been limited by a lack of generalizability as they employ architectures and training methods tailored to specific clinical tasks or datasets [17].

Active research has been conducted in the field of segmentation foundation models for prompt-based universal image segmentation. A representative model, the Segment Anything Model (SAM), has demonstrated the ability to effectively perform various general image segmentation tasks using a single model through prompt engineering. However, SAM is a heavy model, making it impractical for use in resource-constrained environments or edge devices. To address this issue, lightweight models such as MobileSAM [61] and EfficientViT-SAM [62] have been proposed; however, they are specialized for natural image datasets rather than medical images, which presents a limitation.

In response, MedSAM was introduced, fine-tuning the existing SAM model on an unprecedented large-scale dataset comprising over one million medical image-mask pairs, achieving remarkable performance in medical image segmentation. MedSAM underwent comprehensive experimental evaluation on 86 internal and 60 external validation tasks, encompassing various anatomical structures, pathological conditions, and medical imaging modalities. The results showed that MedSAM consistently outperformed the previous SOTA segmentation model, SAM, and exhibited performance on par with or superior to specialized models [25] trained on the same imaging modality.

However, MedSAM, with 93M parameters, is an extremely large model that requires significant computational resources, making it difficult to utilize in resource-constrained computing environments. To address this limitation, LiteMedSAM, a lightweight version of the original MedSAM, was proposed. It was trained in two stages: distilling a lightweight encoder from MedSAM's large image encoder and then fine-tuning the entire pipeline with the distilled encoder. Through this process, LiteMedSAM achieved a significant reduction in model size and computational complexity compared to MedSAM, enabling faster inference in resource-constrained settings.

CVPR 2024: SEGMENT ANYTHING IN MEDICAL IMAGES ON LAPTOP Challenge focuses on developing a prompt-based general model for medical image segmentation. This challenge provides a large-scale dataset comprising over 1,000,000 image-mask pairs, including 11 medical imaging modalities and more than 20 types of cancer. The goal is to develop a prompt-based universal segmentation model that can handle various medical image segmentation tasks while being computationally lightweight enough to run on edge devices such as laptops. In this study, we used LiteMedSAM as the baseline model and optimized the hyperparameters of the image encoder and mask decoder components to develop a more lightweight SwiftMedSAM. While leveraging the large-scale dataset provided, we further reduced the model size and computational complexity to enable real-time inference, even in resource-constrained computing environments.
The developed SwiftMedSAM model is expected to have a significantly reduced model size and computational cost compared to LiteMedSAM, enabling real-time inference in even more constrained environments. Through this research, we aim to further mitigate the generalizability-efficiency trade-off of existing methods and achieve high-quality medical image segmentation under highly limited computing resources.

2 Method

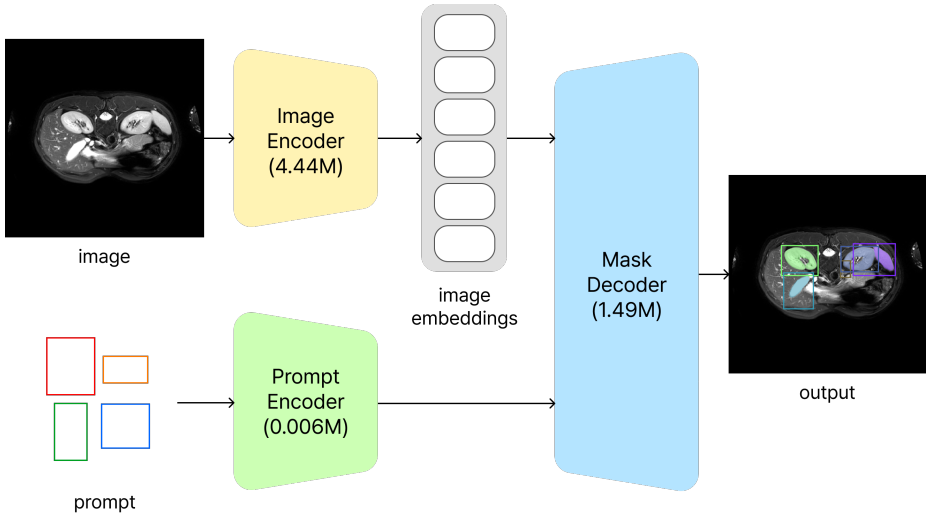

Fig. 1. Overall framework of the proposed method. The image and bounding box prompts serve as inputs to the model, passing through their respective encoders. The resulting outputs are then passed through the mask decoder to produce the segmentation results.

2.1 Preprocessing

The preprocessing strategy was inspired by MedSAM. We utilized a large-scale dataset with over one million image-mask pairs based on publicly available datasets that were used for MedSAM training. This dataset includes 11 imaging modalities (CT, MRI, endoscopy, ultrasound, etc.) and more than 30 types of cancers. The original 3D CT and MRI data, as well as grayscale images (X-ray, ultrasound, etc.) and RGB images (endoscopy, fundus, etc.), were converted to npz format for use. To structure the dataset and enable efficient management, the ground-truth masks and additional information, such as spacing, for both 2D and 3D images were stored together in a single file.

2.2 Proposed Method

image encoder and mask decoder components, using LiteMedSAM as the baseline. The backbone of the existing LiteMedSAM was maintained and the focus was on hyperparameter optimization.

In the image encoder part, the block depths were adjusted to reduce the computational load. As the block depth increases, the model capacity and computational complexity increase. Therefore, a strategy for gradually decreasing the depths were employed. SwiftMedSAMv1 applied block depths of [2, 2, 4, 2], whereas SwiftMedSAMv2 used [1, 2, 2, 2], resulting in a lighter model.

In the mask decoder, modifications were primarily made to the transformer and the IoU head components. First, the transformer depth was reduced from 2 to 1, thereby decreasing the computational cost. Additionally, the transformer's MLP dimensions were significantly reduced from the original 2048 to 1024 and further to 256. As larger MLP dimensions increase the model capacity and computational load, reducing them contributes to the lightening effect.

The number of multi-head attention units in the transformer was also cut roughly in half, from 8 to 4. While more attention heads allow for feature extraction from diverse perspectives, an excessive number risks overfitting and increases the computational complexity. Therefore, an appropriate reduction was made to reduce model capacity and computational load.

Lastly, the depth of the IoU head was lowered from 3 to 2, saving computations in this component as well. The IoU head is related to mask prediction. Although the computational share is not large, this process is worth considering for lightening purposes. The specific hyperparameter adjustment values for each model component are presented in Table 1. As a result, SwiftMedSAM has approximately 5.9M parameters, which is approximately 40% less than the original 9.8M parameters of LiteMedSAM.

Meanwhile, the provided training dataset had a significant imbalance with a substantially lower number of PET modality images. To address this issue, an additional autoPETIII [18] dataset was used to construct the final training dataset. More details on the dataset can be found in Section 3.1

By combining structural lightening strategies, SwiftMedSAM achieves real-time performance and efficiency while maintaining high segmentation accuracy.

Through pre-training on large-scale medical image data, the decrease in segmentation accuracy compared to the original model was not substantial.

Table 1. SwiftMedSAM Hyperparameters.

Component	HyperParameters	Lite MedSAM	Swift MedSAMv1	Swift MedSAMv2
Image Encoder	Block Depths	[2, 2, 6, 2]	[2, 2, 4, 2]	[1, 2, 2, 2]
Mask Decoder	Transformer Depth	2	1	1
Mask Decoder	Transformer MLP Dim	2048	1024	256
Mask Decoder	Transformer Num Heads	8	8	4
Mask Decoder	IOU Head Depth	3	2	2

2.3 Post-processing

The Swift MedSAM model proposed in this study includes a post-processing stage that converts the predicted mask to the original image size through a series of steps. This post-processing stage ensures that the mask output by the model is aligned with the original image size, thereby providing an accurate segmentation result. The post-processing stage consists of the following steps:

1. Cropping
   The predicted mask is resized to the size of the input image (256×256) for the model, the mask undergoes a cropping process to eliminate unnecessary padding areas.

2. Resizing
   The cropped mask is resized to the original size of the image. This is achieved through bilinear interpolation, upsampling the mask to the same size as the original image.

3. Sigmoid Activation Function
   A sigmoid activation function is applied to the upsampled mask, normalizing the values of each pixel between 0 and 1. This step ensures that each pixel in the mask represents the probability of belonging to the target region.

4. Binarization
   The mask that undergoes the sigmoid activation function is binarized using a threshold of 0.5. In other words, values greater than or equal to 0.5 are converted to 1, and values below 0.5 are converted to 0, generating the final mask. This process ensures that the predicted mask has a clear binary form.

3 Experiments

3.1 Dataset and evaluation measures

In the CVPR 2024: SEGMENT ANYTHING IN MEDICAL IMAGES ON LAPTOP challenge, participants could use the training and validation datasets provided by the organizers and external publicly available datasets. The datasets used in developing SwiftMedSAM are as follows: COVID-19-20 [51], AbdomenCT-1K [40], FDG-PET-CT-Lesions [18], NSCLC Radiogenomics [6], NSCLC-Radiomics [18], CT Lymph Nodes [50], NSCLC-PleuralEffusion [31], NSCLC-Lung MSD-LUNG [53,4,38], KiTS23 [20], CT-ORG [4], COVID-19-20-CTSEG [38], TotalSegmentator [58], AMOS [28], LCTSC [60], HCC-TACE-Seg [45], Adrenal-ACC-Ki67-Seg [44], MSD [4,52], ISLES [21], WMH [33], BraTS [5,42], PROMISE12 [35], MSD-Prostate [4,52], NCI-ISBI [7], Crossmoda [14], QIN-PROSTATE-Repeatability [16], CC-Tumor Heterogeneity [8], COVID-19 Radiography Database [48,10], COVID-QU-Ex [55,13,10,48], Chest Xray Masks and Labels [9,26], Chest X-Ray Images with Pneumothorax Masks, CDD-CESM [30,29], Intraretinal Cystoid Fluid [2], ps-fh-aop-2023 [37], hc18 [22,23], Breast Ultrasound Images Dataset [3], ISIC2018 [56,11,12], CholecSeg8k [24,57], Kvasir-SEG [27,46], m2caiSeg [41], PAPILA [32], IDRiD [47], NeurIPS CellSeg [39], autoPETIII.

The training dataset includes 11 imaging modalities: Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), X-ray, ultrasound, mammography, Optical Coherence Tomography (OCT), endoscopy, fundus, dermoscopy, and microscopy. A total of 1,490,576 medical image-mask pairs were used to train our model. The validation dataset contains 9 modalities and is a subset of the testing set used in this challenge.

The evaluation metrics for this challenge were divided into accuracy and efficiency. The accuracy metrics are Dice Similarity Coefficient (DSC), which measures the overlap between the ground truth and predictions, and Normalized Surface Dice (NSD), which measures the similarity between the ground truth boundary and predictions. The efficiency metric is runtime, measured using only the CPU without GPU assistance.

3.2 Implementation details

Environment settings The development environments and requirements are presented in Table 2.

Table 2. Development environments and requirements.

System	Ubuntu 20.04.6 LTS
CPU	AMD EPYC™7402X CPU@2.8GHz
RAM	8×64GB; 3200MT/s
GPU (number and type)	Four NVIDIA A100 80G
CUDA version	11.8
Programming language	Python 3.10.13
Deep learning framework	torch 2.1.0 , torchvision 0.16.0
Code	

Training protocols The training protocols of SwiftMedSAM is listed in Table 3.

Table 3. Training protocols.

Pre-trained Model	LiteMedSAM
Batch size	32
Patch size	256×256×3
Total epochs	26
Optimizer	AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Initial learning rate (lr)	0.005
Lr decay schedule	ReduceLROnPlateau
Training time	93.5 hours
Loss function	Dice Loss, Binary Cross Entrophy Loss
Number of model parameters	5.94M[6]
Number of flops	4x312T[7]
CO_2eq	16.16 Kg[8]

4 Results and discussion

4.1 Quantitative results on validation set

In this study, we conducted experiments to evaluate the performance of SwiftMedSAM. The accuracy was measured based on the 3,076 validation data images provided and compared with the baseline model, LiteMedSAM. The results are

presented in Table 4. Compared to the baseline, SwiftMedSAMv1 exhibited a 0.05% average decrease in DSC, but a 2.00% improvement in NSD. Compared to SwiftMedSAMv1, SwiftMedSAMv2 showed a 0.63% improvement in DSC and 0.73% improvement in NSD.
In particular, when examining each imaging modality, for CT images, SwiftMedSAMv1 achieved a 2.38% improvement in DSC and a 3.39% improvement in NSD compared to the baseline. For MR images, DSC improved by 3.42% and NSD improved by 4.45%. For PET images, there was a significant improvement, with DSC increasing by 13.77% and NSD by 21.79%. However, for US images, DSC decreased by 15.61% and NSD decreased by 11.69%. For X-ray images, DSC decreased by 11.34% and NSD decreased by 9.55%.
Comparing SwiftMedSAMv2 and SwiftMedSAMv1, for CT images, DSC decreased by 0.33%, whereas NSD decreased by 0.27%. For MR images, DSC decreased by 0.78% and NSD decreased by 0.56%. For PET images, DSC decreased by 5.49% and NSD decreased by 0.93%. For US images, DSC improved by 1.72%, and NSD improved by 2.14%. For X-ray images, DSC improved by 6.24%, and NSD improved by 5.88%.

Table 4. Quantitative evaluation results on validation set.

Target	Baseline		Swift MedSAMv1		Swift MedSAMv2	
	DSC(%)	NSD(%)	DSC(%)	NSD(%)	DSC(%)	NSD (%)
CT	40.71	40.27	43.09	43.66	42.76	43.39
MR	61.17	62.40	64.59	66.85	63.81	66.29
PET	55.10	29.17	68.87	50.96	63.38	50.03
US	94.77	96.81	79.16	85.12	80.88	87.26
X-Ray	75.82	80.38	64.48	70.83	70.72	76.71
Dermotology	92.47	93.85	93.88	95.30	93.43	94.80
Endoscopy	96.04	98.11	94.57	97.22	95.58	98.02
Fundus	94.81	96.41	94.10	95.79	96.13	97.69
Microscopy	61.63	65.39	69.41	75.05	66.13	73.13
Average	74.73	73.64	74.68	75.64	75.31	76.37

4.2 Qualitative results on validation set

For the comparison of qualitative results, we used publicly available datasets with ground truth annotations: CT2USforKidneySeg [54], HipXRay [19], and NSCLC-Radiomics [1], which contain ultrasound, X-ray, and CT modalities, respectively. Examples of SwiftMedSAMv2's segmentation results for these datasets are shown in Fig 2. While SwiftMedSAMv2 demonstrated refined segmentation performance on the CT2USforKidneySeg and NSCLC-Radiomics datasets, it yielded suboptimal results for some images from the HipXRay dataset with low-contrast or unclear boundaries.

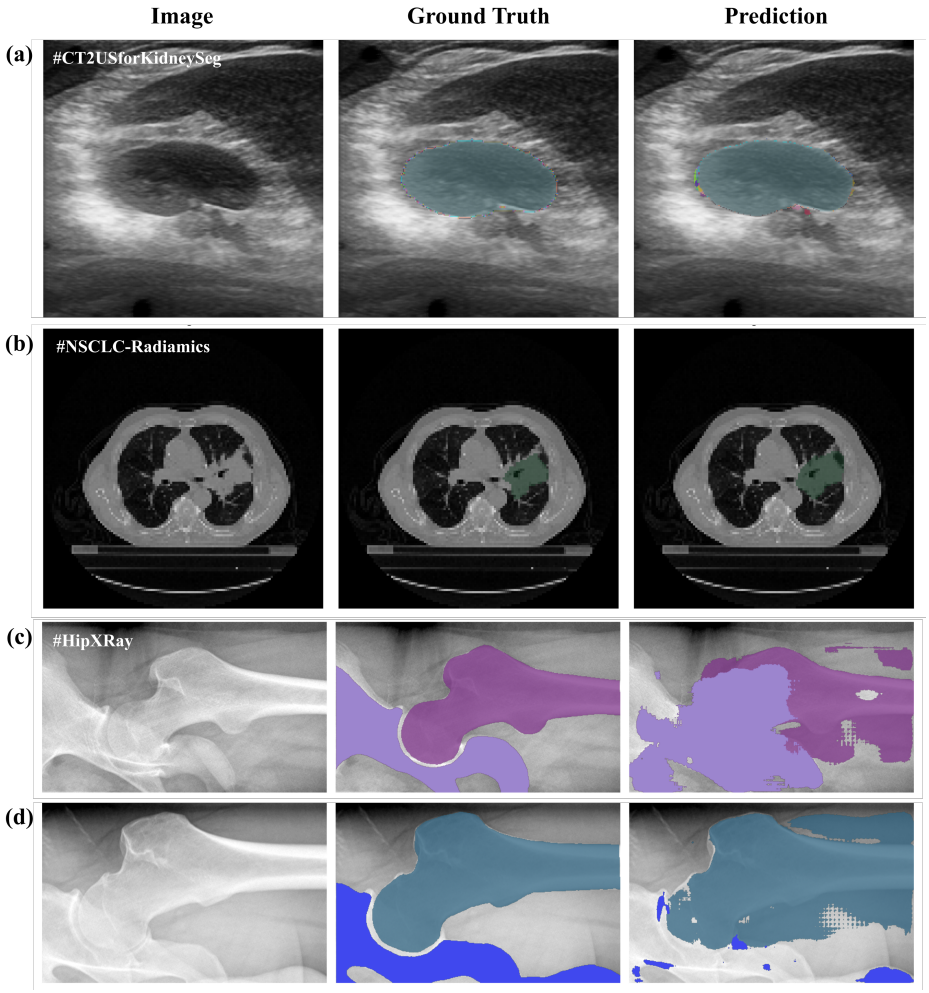

Fig. 2. Qualitative results of our SwiftMedSAMv2. (a-b) Good segmentation cases from CT2USforKidneySeg dataset and NSCLC-Radiomics dataset, respectively. Our method accurately delineates the kidney boundaries in the CT image (a) and captures the tumor region in the lung CT image (b). (c-d) Bad segmentation cases from the HipXRay dataset, where our method struggles to segment the femoral head and acetabulum regions precisely due to low contrast and complex anatomical structures.

4.3 Segmentation efficiency results on validation set

The efficiency experiments for the final model were performed on a CPU: AMD EPYC™7402X CPU@2.8GHz, RAM: 8×64GB; 3200MT/s. The specific inference time measurements for some cases are listed in Table 5. Consequently, the SwiftMedSAMv1 and SwiftMedSAMv2 models showed reduced execution times compared to the baseline model in most cases. Particularly, for the

3DBox_MR_0621 case, while the baseline model's execution time was 630.2 seconds, SwiftMedSAMv1 and SwiftMedSAMv2 took 166.8 seconds and 145.0 seconds, respectively, showing a significant reduction. Compared to SwiftMedSAMv1, SwiftMedSAMv2 exhibited better performance in some cases, and an overall slight improvement was observed. For instance, in the 3DBox_CT_0566 case, SwiftMedSAMv1 recorded 343.2 seconds, while SwiftMedSAMv2 took 331.9 seconds, demonstrating a faster execution time.

Table 5. Quantitative evaluation of efficiency in terms of running time (s).

Case ID	Size	Num. Objects	Baseline	Swift MedSAMv1	Swift MedSAMv2
3DBox_CT_0566	(287, 512, 512)	6	499.2	343.2	331.9
3DBox_CT_0888	(237, 512, 512)	6	114.1	97.4	94.1
3DBox_CT_0860	(246, 512, 512)	1	17.7	15.9	14.2
3DBox_MR_0621	(115, 400, 400)	6	630.2	166.8	145.0
3DBox_MR_0121	(64, 290, 320)	6	119.4	94.7	90.7
3DBox_MR_0179	(84, 512, 512)	1	17.7	14.7	13.3
3DBox_PET_0001	(264, 200, 200)	1	31.9	9.9	8.6
2DBox_US_0525	(256, 256, 3)	1	1.8	1.7	1.6
2DBox_X-Ray_0053	(320, 640, 3)	34	9.8	9.3	9.8
2DBox_Dermoscopy_0003	(3024, 4032, 3)	1	7.9	8.4	7.9
2DBox_Endoscopy_0086	(480, 560, 3)	1	6.1	2.7	2.6
2DBox_Fundus_0003	(2048, 2048, 3)	1	2.6	4.2	4.0
2DBox_Microscope_0008	(1536, 2040, 3)	19	19.5	18.1	18.0
2DBox_Microscope_0016	(1920, 2560, 3)	241	257.5	253.1	267.1

4.4 Results on final testing set

This is a placeholder. We will announce the testing results during CVPR (6.17-18)

4.5 Limitation and future work

SwiftMedSAM exhibited versatility in medical image segmentation despite its remarkably small model size and low computational complexity. However, there are still limitations in which segmentation errors occur when the boundaries of the structures/lesions are ambiguous. We expect that these issues can be resolved with higher-quality data and more powerful training strategies in the future. Currently, the performance is maintained at the level of the baseline model, but we aim to achieve a fast inference speed and improve the performance in the future.

5 Conclusion

In this study, we proposed SwiftMedSAM, an ultra-lightweight prompt-based model that enables real-time high-performance medical image segmentation even

in highly constrained computing environments. While maintaining the backbone of the existing SOTA model MedSAM, we introduced the lightweight LiteMedSAM as the baseline and performed a process of hyperparameter tuning to drastically reduce the model size and computational complexity. Through experiments, we verified the comparable segmentation performance and fast inference speed of SwiftMedSAM.

The proposed SwiftMedSAM demonstrates the potential for a universal prompt-based medical image segmentation model while simultaneously pursuing efficiency and generalizability. This will enable high-quality medical image analysis, even in resource-constrained environments, contributing to advancements in precision medicine and telemedicine.

Acknowledgements We thank all the data owners for making the medical images publicly available and CodaLab [59] for hosting the challenge platform.

References

1. Aerts, H.J.W.L., Wee, L., Velazquez, E.R., Leijenaar, R.T.H., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., Hoebers, F., Rietbergen, M.M., Leemans, C.R., Dekker, A., Quackenbush, J., Gillies, R.J., Lambin, P.: Data from nsclc-radiomics (version 4) [data set]. The Cancer Imaging Archive (2014). https://doi.org/10.7937/K9/TCIA.2015.PF0M9REI, https://doi.org/10.7937/K9/TCIA.2015.PF0M9REI 8
2. Ahmed, Z., Panhwar, S.Q., Baqai, A., Umrani, F.A., Ahmed, M., Khan, A.: Deep learning based automated detection of intraretinal cystoid fluid. International Journal of Imaging Systems and Technology **32**(3), 902–917 (2022). https://doi.org/https://doi.org/10.1002/ima.22662, https://onlinelibrary.wiley.com/doi/abs/10.1002/ima.22662 6
3. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in Brief **28**, 104863 (Feb 2020). https://doi.org/10.1016/j.dib.2019.104863 6
4. Antonelli, M., Reinke, A., Bakas, S., et al.: The medical segmentation decathlon. Nature Communications **13**, 4128 (2022). https://doi.org/10.1038/s41467-022-30695-9 6
5. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., Prastawa, M., Alberts, E., Lipková, J., Freymann, J.B., Kirby, J.S., Bilello, M., Fathallah-Shaykh, H.M., Wiest, R., Kirschke, J., Wiestler, B., Colen, R.R., Kotrotsou, A., LaMontagne, P., Marcus, D.S., Milchenko, M., Nazeri, A., Weber, M., Mahajan, A., Baid, U., Kwon, D., Agarwal, M., Alam, M., Albiol, A., Albiol, A., Varghese, A., Tuan, T.A., Arbel, T., Avery, A., B., P., Banerjee, S., Batchelder, T., Batmanghelich, K.N., Battistella, E., Bendszus, M., Benson, E., Bernal, J., Biros, G., Cabezas, M., Chandra, S., Chang, Y., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. CoRR **abs/1811.02629** (2018), http://arxiv.org/abs/1811.02629 6
6. Bakr, S., Gevaert, O., Echegaray, S., et al.: A radiogenomic dataset of non-small cell lung cancer. Scientific Data **5**, 180202 (2018). https://doi.org/10.1038/sdata.2018.202 6

7. Bloch, N., Madabhushi, A., Huisman, H., Freymann, J., Kirby, J., Grauer, M., Enquobahrie, A., Jaffe, C., Clarke, L., Farahani, K.: Nci-isbi 2013 challenge: Automated segmentation of prostate structures. The Cancer Imaging Archive (2015), http://doi.org/10.7937/K9/TCIA.2015.zF0vlOPv 6
8. Bowen, S.R., Yuh, W.T., Hippe, D.S., Wu, W., Partridge, S.C., Elias, S., Jia, G., Huang, Z., Sandison, G.A., Nelson, D., Knopp, M.V., Lo, S.S., Kinahan, P.E., Mayr, N.A.: Tumor radiomic heterogeneity: Multiparametric functional imaging to characterize variability and predict response following cervical cancer radiation therapy. Journal of Magnetic Resonance Imaging **47**(5), 1388–1396 (2018). https://doi.org/https://doi.org/10.1002/jmri.25874, https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.25874 6
9. Candemir, S., Jaeger, S., Palaniappan, K., Musco, J.P., Singh, R.K., Xue, Z., Karargyris, A., Antani, S., Thoma, G., McDonald, C.J.: Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Transactions on Medical Imaging **33**(2), 577–590 (2014). https://doi.org/10.1109/TMI.2013.2290491 6
10. Chowdhury, M.E.H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Emadi, N.A., Reaz, M.B.I., Islam, M.T.: Can ai help in screening viral and covid-19 pneumonia? IEEE Access **8**, 132665–132676 (2020). https://doi.org/10.1109/ACCESS.2020.3010287 6
11. Codella, N.C.F., Gutman, D.A., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N.K., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (ISIC). CoRR **abs/1710.05006** (2017), http://arxiv.org/abs/1710.05006 6
12. Codella, N.C.F., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S.W., Gutman, D.A., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M.A., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). CoRR **abs/1902.03368** (2019), http://arxiv.org/abs/1902.03368 6
13. Degerli, A., Ahishali, M., Yamac, M., et al.: Covid-19 infection map generation and detection from chest x-ray images. Health Inf Sci Syst **9**, 15 (2021). https://doi.org/10.1007/s13755-021-00146-8, https://doi.org/10.1007/s13755-021-00146-8 6
14. Dorent, R., Kujawa, A., Ivory, M., Bakas, S., Rieke, N., Joutard, S., Glocker, B., Cardoso, J., Modat, M., Batmanghelich, K., Belkov, A., Calisto, M.B., Choi, J.W., Dawant, B.M., Dong, H., Escalera, S., Fan, Y., Hansen, L., Heinrich, M.P., Joshi, S., Kashtanova, V., Kim, H.G., Kondo, S., Kruse, C.N., Lai-Yuen, S.K., Li, H., Liu, H., Ly, B., Oguz, I., Shin, H., Shirokikh, B., Su, Z., Wang, G., Wu, J., Xu, Y., Yao, K., Zhang, L., Ourselin, S., Shapey, J., Vercauteren, T.: Crossmoda 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwannoma and cochlea segmentation. Medical Image Analysis **83**, 102628 (2023). https://doi.org/https://doi.org/10.1016/j.media.2022.102628, https://www.sciencedirect.com/science/article/pii/S1361841522002560 6
15. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020) 2

16. Fedorov, A., Schwier, M., Clunie, D., et al.: An annotated test-retest collection of prostate multiparametric mri. Sci Data **5**, 180281 (2018). https://doi.org/10.1038/sdata.2018.281, https://doi.org/10.1038/sdata.2018.281 6
17. Fidon, L.: Trustworthy deep learning for medical image segmentation. arXiv preprint arXiv:2305.17456 (2023) 2
18. Gatidis, S., Hepp, T., Früh, M., et al.: A whole-body fdg-pet/ct dataset with manually annotated tumor lesions. Scientific Data **9**, 601 (2022). https://doi.org/10.1038/s41597-022-01718-3 4, 6
19. Gut, D.: X-ray images of the hip joints. Mendeley Data, V1 (2021). https://doi.org/10.17632/zm6bxzhmfz.1 8
20. 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). https://doi.org/https://doi.org/10.1016/j.media.2020.101821, https://www.sciencedirect.com/science/article/pii/S1361841520301857 6
21. Hernandez Petzsche, M.R., de la Rosa, E., Hanning, U., et al.: Isles 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. Sci Data **9**, 762 (2022). https://doi.org/10.1038/s41597-022-01875-5, https://doi.org/10.1038/s41597-022-01875-5 6
22. van den Heuvel, T.L., de Bruijn, D., de Korte, C.L., Ginneken, B.v.: Automated measurement of fetal head circumference using 2d ultrasound images. PloS one **13**(8), e0200412 (2018) 6
23. Heuvel, T.v.d., de Bruijn, D., de Korte, C.L., van Ginneken, B.: Automated measurement of fetal head circumference (2018). https://doi.org/10.5281/zenodo.xxxxxxx, https://doi.org/10.5281/zenodo.xxxxxxx 6
24. Hong, W., Kao, C., Kuo, Y., Wang, J., Chang, W., Shih, C.: Cholecseg8k: A semantic segmentation dataset for laparoscopic cholecystectomy based on cholec80. CoRR **abs/2012.12453** (2020), https://arxiv.org/abs/2012.12453 6
25. 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
26. Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R.K., Antani, S., Thoma, G., Wang, Y.X., Lu, P.X., McDonald, C.J.: Automatic tuberculosis screening using chest radiographs. IEEE Transactions on Medical Imaging **33**(2), 233–245 (2014). https://doi.org/10.1109/TMI.2013.2284099 6
27. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: Ro, Y.M., Cheng, W.H., Kim, J., Chu, W.T., Cui, P., Choi, J.W., Hu, M.C., De Neve, W. (eds.) MultiMedia Modeling. pp. 451–462. Springer International Publishing, Cham (2020) 6
28. Ji, Y., Bai, H., Ge, C., Yang, J., Zhu, Y., Zhang, R., Li, Z., Zhanng, L., Ma, W., Wan, X., et al.: Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. Advances in Neural Information Processing Systems **35**, 36722–36732 (2022) 6

29. Khaled, R., Helal, M., Alfarghaly, O., Mokhtar, O., Elkorany, A., El Kassas, H., Fahmy, A.: Categorized digital database for low energy and subtracted contrast enhanced spectral mammography images (2021). https://doi.org/10.7937/29kw-ae92, https://doi.org/10.7937/29kw-ae92 6
30. Khaled, R., Helal, M., Alfarghaly, O., et al.: Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research. Sci Data **9**, 122 (2022). https://doi.org/10.1038/s41597-022-01238-0, https://doi.org/10.1038/s41597-022-01238-0 6
31. Kiser, K.J., Barman, A., Stieb, S., et al.: Novel autosegmentation spatial similarity metrics capture the time required to correct segmentations better than traditional metrics in a thoracic cavity segmentation workflow. Journal of Digital Imaging **34**, 541–553 (2021). https://doi.org/10.1007/s10278-021-00460-3 6
32. Kovalyk, O., Morales-Sánchez, J., Verdú-Monedero, R., et al.: Papila: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. Sci Data **9**, 291 (2022). https://doi.org/10.1038/s41597-022-01388-1 6
33. Kuijf, H.J., Biesbroek, J.M., De Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., Casamitjana, A., Collins, D.L., Dadar, M., Georgiou, A., Ghafoorian, M., Jin, D., Khademi, A., Knight, J., Li, H., Llado, X., Luna, M., Mahmood, Q., McKinley, R., Mehrtash, A., Ourselin, S., Park, B.Y., Park, H., Park, S.H., Pezold, S., Puybareau, , Rittner, L., Sudre, C.H., Valverde, S., Vilaplana, V., Wiest, R., Xu, Y., Xu, Z., Zeng, G., Zhang, J., Zheng, G., Chen, C., van der Flier, W., Barkhof, F., Viergever, M.A., Biessels, G.J.: Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge. IEEE Trans Med Imaging **38**(11), 2556–2568 (Nov 2019). https://doi.org/10.1109/TMI.2019.2905770 6
34. Lei, T., Nandi, A.: Image Segmentation for Medical Analysis, pp. 199–227 (09 2022). https://doi.org/10.1002/9781119859048.ch9 2
35. Litjens, G., van Ginneken, B., Huisman, H., van de Ven, W., Hoeks, C., Barratt, D., Madabhushi, A.: Promise12: Data from the miccai grand challenge: Prostate mr image segmentation 2012. Medical Image Analysis **18**(2), 359–373 (2023). https://doi.org/10.5281/zenodo.8026660, https://doi.org/10.5281/zenodo.8026660 6
36. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021) 2
37. Lu, Y., Zhou, M., Zhi, D., Zhou, M., Jiang, X., Qiu, R., Ou, Z., Wang, H., Qiu, D., Zhong, M., Lu, X., Chen, G., Bai, J.: The jnu-ifm dataset for segmenting pubic symphysis-fetal head. Data in Brief **41**, 107904 (2022). https://doi.org/https://doi.org/10.1016/j.dib.2022.107904, https://www.sciencedirect.com/science/article/pii/S2352340922001160 6
38. Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z., et al.: Toward data-efficient learning: A benchmark for covid-19 ct lung and infection segmentation. Medical physics **48**(3), 1197–1210 (2021) 6
39. Ma, J., Xie, R., Ayyadhury, S., Ge, C., Gupta, A., Gupta, R., Gu, S., Zhang, Y., Lee, G., Kim, J., Lou, W., Li, H., Upschulte, E., Dickscheid, T., de Almeida, J.G., Wang, Y., Han, L., Yang, X., Labagnara, M., Gligorovski, V., Scheder, M., Rahi, S.J., Kempster, C., Pollitt, A., Espinosa, L., Mignot, T., Middeke, J.M., Eckardt, J.N., Li, W., Li, Z., Cai, X., Bai, B., Greenwald, N.F., Van Valen, D., Weisbart, E., Cimini, B.A., Cheung, T., Brück, O., Bader, G.D., Wang, B.: The multimodality

cell segmentation challenge: toward universal solutions. Nature Methods (Mar 2024). https://doi.org/10.1038/s41592-024-02233-6, http://dx.doi.org/10.1038/s41592-024-02233-6 6

40. 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
41. Maqbool, S., Riaz, A., Sajid, H., Hasan, O.: m2caiseg: Semantic segmentation of laparoscopic images using convolutional neural networks. arXiv preprint arXiv:2008.10134 (2020) 6
42. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, , Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694 6
43. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016) 2
44. Moawad, A.W., Ahmed, A.A., ElMohr, M., Eltaher, M., Habra, M.A., Fisher, S., Perrier, N., Zhang, M., Fuentes, D., Elsayes, K.: Voxel-level segmentation of pathologically-proven adrenocortical carcinoma with ki-67 expression (adrenal-acc-ki67-seg). Data set (2023). https://doi.org/10.7937/1FPG-VM46, https://doi.org/10.7937/1FPG-VM46 6
45. Morshid, A., Elsayes, K.M., Khalaf, A.M., Elmohr, M.M., Yu, J., Kaseb, A.O., Hassan, M., Mahvash, A., Wang, Z., Hazle, J.D., Fuentes, D.: A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. Radiology: Artificial Intelligence **1**(5), e180021 (2019). https://doi.org/10.1148/ryai.2019180021, https://doi.org/10.1148/ryai.2019180021, pMID: 31858078 6
46. Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.T., Lux, M., Schmidt, P.T., Riegler, M., Halvorsen, P.: Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection (2017), https://doi.org/10.1145/3193289 6
47. Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Meriaudeau, F.: Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research. Data **3**(3), 25 (2018). https://doi.org/10.3390/data3030025 6
48. Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S.B., Islam, M.T., Al Maadeed, S., Zughaier, S.M., Khan, M.S., Chowdhury, M.E.: Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. Computers in Biology and Medicine **132**, 104319 (2021). https://doi.org/https://doi.org/10.1016/j.

compbiomed.2021.104319, https://www.sciencedirect.com/science/article/pii/S001048252100113X 6
49. 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) 2
50. Roth, H.R., Lu, L., Seff, A., Cherry, K.M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., Summers, R.M.: A new 2.5d representation for lymph node detection using random sets of deep convolutional neural network observations. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. pp. 520–527. Springer International Publishing, Cham (2014) 6
51. Roth, H.R., Xu, Z., Tor-Díez, C., Sanchez Jacob, R., Zember, J., Molto, J., Li, W., Xu, S., Turkbey, B., Turkbey, E., Yang, D., Harouni, A., Rieke, N., Hu, S., Isensee, F., Tang, C., Yu, Q., Sölter, J., Zheng, T., Liauchuk, V., Zhou, Z., Moltz, J.H., Oliveira, B., Xia, Y., Maier-Hein, K.H., Li, Q., Husch, A., Zhang, L., Kovalev, V., Kang, L., Hering, A., Vilaça, J.L., Flores, M., Xu, D., Wood, B., Linguraru, M.G.: Rapid artificial intelligence solutions in a pandemic—the covid-19-20 lung ct lesion segmentation challenge. Medical Image Analysis **82**, 102605 (2022). https://doi.org/https://doi.org/10.1016/j.media.2022.102605, https://www.sciencedirect.com/science/article/pii/S1361841522002353 6
52. 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
53. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B.H., Ronneberger, O., Summers, R.M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M., Golia-Pernicka, J., Heckers, S., Jarnagin, W.R., McHugo, M., Napel, S., Vorontsov, E., Maier-Hein, L., Cardoso, M.J.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. CoRR **abs/1902.09063** (2019), http://arxiv.org/abs/1902.09063 6
54. Song, Y., Zheng, J., Lei, L., Ni, Z., Zhao, B., Hu, Y.: Ct2us: Cross-modal transfer learning for kidney segmentation in ultrasound images with synthesized data. Ultrasonics **122**, 106706 (2022) 8
55. Tahir, A.M., Chowdhury, M.E., Khandakar, A., Rahman, T., Qiblawey, Y., Khurshid, U., Kiranyaz, S., Ibtehaz, N., Rahman, M.S., Al-Maadeed, S., Mahmud, S., Ezeddin, M., Hameed, K., Hamid, T.: Covid-19 infection localization and severity grading from chest x-ray images. Computers in Biology and Medicine **139**, 105002 (2021). https://doi.org/https://doi.org/10.1016/j.compbiomed.2021.105002, https://www.sciencedirect.com/science/article/pii/S0010482521007964 6
56. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data **5**, 180161 (2018). https://doi.org/10.1038/sdata.2018.161 6
57. Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., de Mathelin, M., Padoy, N.: Endonet: A deep architecture for recognition tasks on laparoscopic videos. CoRR **abs/1602.03012** (2016), http://arxiv.org/abs/1602.03012 6

58. 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). https://doi.org/10.1148/ryai.230024, https://doi.org/10.1148/ryai.230024 6
59. 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) 11
60. Yang, J., Veeraraghavan, H., Armato III, S.G., Farahani, K., Kirby, J.S., Kalpathy-Kramer, J., van Elmpt, W., Dekker, A., Han, X., Feng, X., Aljabar, P., Oliveira, B., van der Heyden, B., Zamdborg, L., Lam, D., Gooding, M., Sharp, G.C.: Autosegmentation for thoracic radiation treatment planning: A grand challenge at aapm 2017. Medical Physics **45**(10), 4568–4581 (2018). https://doi.org/https://doi.org/10.1002/mp.13141, https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13141 6
61. Zhang, C., Han, D., Qiao, Y., Kim, J.U., Bae, S.H., Lee, S., Hong, C.S.: Faster segment anything: Towards lightweight sam for mobile applications. arXiv preprint arXiv:2306.14289 (2023) 2
62. Zhang, Z., Cai, H., Han, S.: Efficientvit-sam: Accelerated segment anything model without performance loss. In: CVPR Workshop: Efficient Large Vision Models (2024) 2

Table 6. Checklist Table. Please fill out this checklist table in the answer column.

Requirements	Answer
A meaningful title	Yes
The number of authors (≤6)	5
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 4
Strategies to data augmentation	(none)
Strategies to improve model inference	Page 4
Post-processing	Page 5
Environment setting table is provided	Table 2
Training protocol table is provided	Table 3
Ablation study	Page 7
Efficiency evaluation results are provided	Table 5
Visualized segmentation example is provided	Figure 2
Limitation and future work are presented	Yes
Reference format is consistent.	Yes
Main text >= 8 pages (not include references and appendix)	Yes