SAM-Med3D

Yu Jiening
Yu Jiening@umac.mo

um 澳大

arXiv preprint arXiv, 2023.10.23

SAM-Med3D

Haoyu Wang[1,2*] Sizheng Guo[1*] Jin Ye[1*] Zhongying Deng[1*]
Junlong Cheng[1] Tianbin Li[1] Jianpin Chen[1] Yanzhou Su[1] Ziyan Huang[1,2]
Yiqing Shen[1] Bin Fu[3] Shaoting Zhang[1] Junjun He[1†] Yu Qiao[1†]

[1]Shanghai AI Laboratory
[2]Shanghai Jiao Tong University
[3]ShenZhen Key Lab of Computer Vision and Pattern Recognition,
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
{yejin, hejunjun, litianbin, zhangshaoting, qiaoyu}@pjlab.org.cn

Abstract

Although the Segment Anything Model (SAM) has demonstrated impressive performance in 2D natural image segmentation, its application to 3D volumetric medical images reveals significant shortcomings, namely suboptimal performance and unstable prediction, necessitating an excessive number of prompt points to attain the desired outcomes. These issues can hardly be addressed by fine-tuning SAM on medical data because the original 2D structure of SAM neglects 3D spatial information. In this paper, we introduce SAM-Med3D, the most comprehensive study to modify SAM for 3D medical images. Our approach is characterized by its comprehensiveness in two primary aspects: firstly, by comprehensively reformulating SAM to a thorough 3D architecture trained on a comprehensively processed large-scale volumetric medical dataset; and secondly, by providing a comprehensive evaluation of its performance. Specifically, we train SAM-Med3D with over 131K 3D masks and 247 categories. Our SAM-Med3D excels at capturing 3D spatial information, exhibiting competitive performance with significantly fewer prompt points than the top-performing fine-tuned SAM in the medical domain. We then evaluate its capabilities across 15 datasets and analyze it from multiple perspectives, including anatomical structures, modalities, targets, and generalization abilities. Our approach, compared with SAM, showcases pronouncedly enhanced efficiency and broad segmentation capabilities for 3D volumetric medical images. Our code is released at https://github.com/uni-medical/SAM-Med3D.

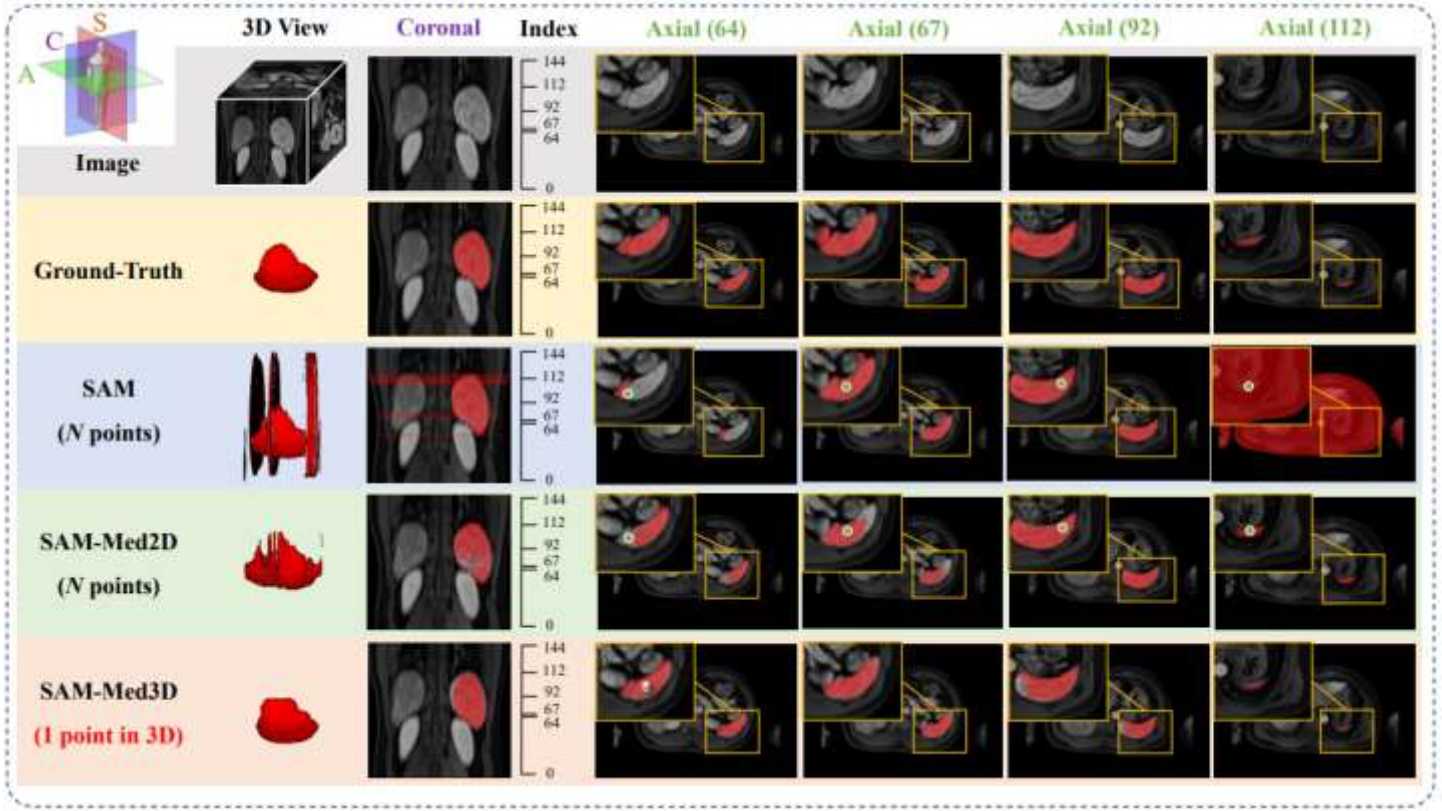
1 Introduction

Medical image analysis has become an indispensable cornerstone of modern healthcare, aiding diagnosis, treatment planning, and further medical research [39, 42, 17]. One of the most significant challenges in this realm is the precise segmentation of volumetric medical images [24]. Although numerous methods have demonstrated commendable effectiveness across a spectrum of targets [25, 26, 47], prevailing segmentation techniques exhibit a tendency of specialization towards specific organs

Catalogue

- Introduction
- Datasets
- Method
- Experiments
- Conclusion

1. Introduction



- Current medical image segmentation models all use a **slice-by-slice** method to segment 3D images, ignoring the **3D spatial information** between the slices, and therefore perform poorly on 3D medical images.

Model	Dataset Size	Category	Image Encoder	Prompt Encoder	Mask Decoder
MedLSAM [23]	~ 25K masks	~ 50	❄ 2D	❄ 2D	❄ 2D
SAM^{Med} [38]	-	-	❄ 2D	❄ 2D	❄ 2D
SAM3D [2]	2K masks	14	❄ 2D	-	🔥 **3D**
MA-SAM [3]	5 datasets	$\leq$ 13	❄ 2D + 🔥 Adapter	-	🔥 2D
MSA [45]	12K masks	15	❄ 2D + 🔥 Adapter	🔥 2D	🔥 2D
3DSAM-A [11]	$\leq$ 1K masks	4	❄ 2D + 🔥 Adapter	🔥 **3D**	🔥 **3D**
SAM-Med3D	131K masks	247	🔥 **3D**	🔥 **3D**	🔥 **3D**

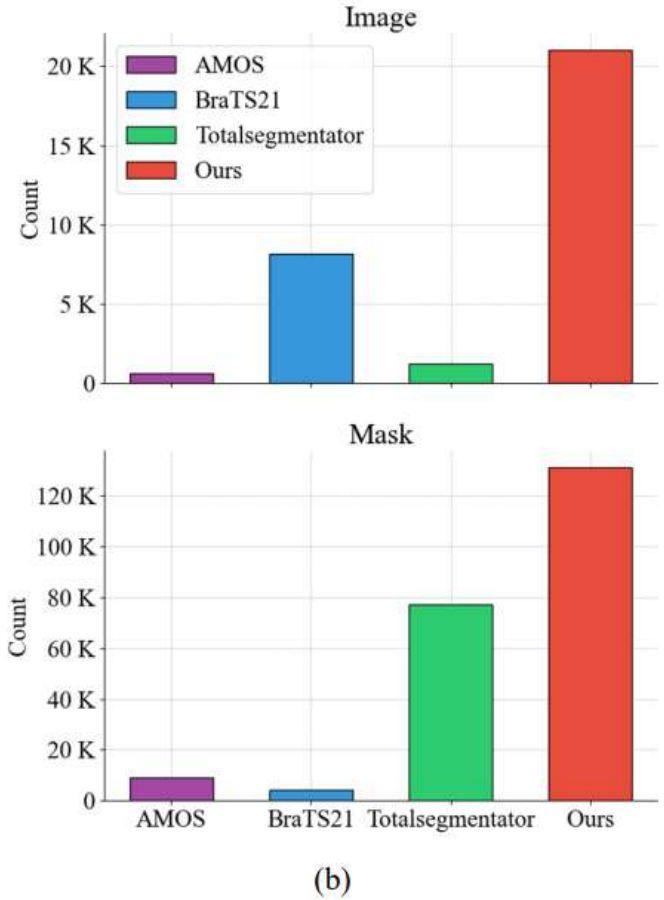
- Some researchers have trained 3D adapters by freezing 2D layers so that the model can learn from 3D images. However, these approaches have two limitations:
- (1) **Limited data size**: Their models are trained only on a limited data size and a limited number of target types.
- (2) **Inherent 2D architecture**: their models always adhere to 2D design paradigms (e.g., frozen 2D encoders), limiting the ability to fully model 3D spatial information.

2.Datasets

2.1 Training

- The dataset contains **21K medical images** and **131K masks** which is probably the largest volumetric medical image segmentation dataset to date. The dataset covers **27 modalities (CT and 26 MRI sequences)** and **7 anatomical structures.**

(a)

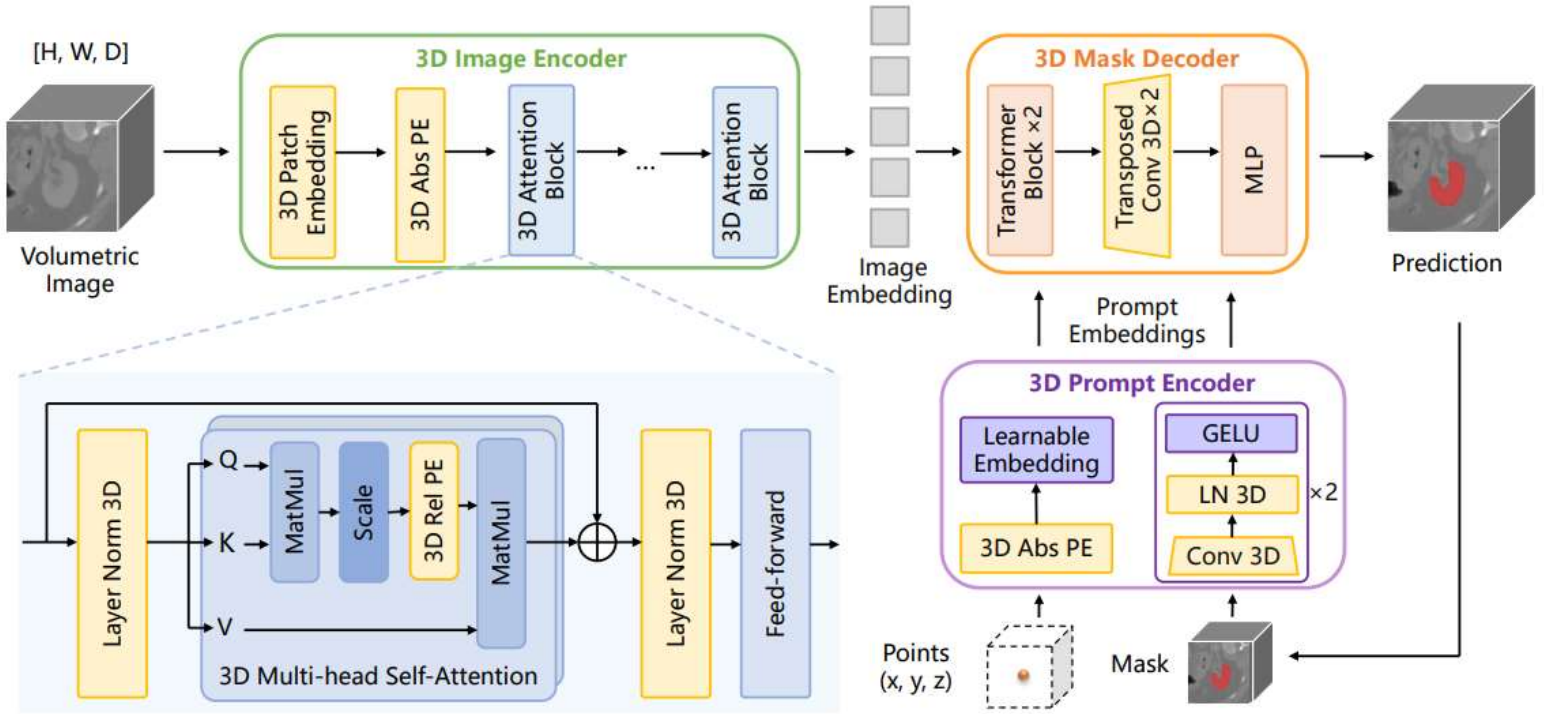


(b)

2.2 Testing

- Selected **13 public benchmark datasets** to scrutinize various clinical scenarios, and incorporated 2 additional datasets from the MICCAI 2023 Challenge to validate the performance of different models.
- This validation set encompasses **seven crucial anatomical structures**, such as thorax and abdominal organs, brain structures, bones, and more.
- It also includes **five types of lesions** that hold significant interest in the medical field, and a range of volumetric modalities, including CT, US (Ultrasound), and eight MRI sequences.

3.Method



- The modified 3D architecture of our SAM-Med3D. The original 2D components are transformed into their 3D counterparts, encompassing a **3D image encoder, 3D prompt encoder, and 3D mask decoder**. 3D convolution, 3D positional encoding (PE) and 3D layer norm are employed to construct the 3D model.

4.Experiments

- Our method is implemented in PyTorch and trained on 8 NVIDIA Tesla A100 GPUs, each with 80GB memory. We use the Adam optimizer with an initial learning rate of 8e-4 and train for a total of 800 epochs.

4.1 Quantitative comparison of different methods on our evaluation dataset

Model	Prompt	Resolution	T_{inf} (s)	Overall Dice
SAM	N points	$1024\times1024\times N$	13	17.01
SAM-Med2D	N points	$256\times256\times N$	4	42.75
SAM-Med3D	1 point	$128\times128\times128$	2	49.91
SAM	$3N$ points	$1024\times1024\times N$	19	31.86
SAM-Med2D	$3N$ points	$256\times256\times N$	7	54.61
SAM-Med3D	3 points	$128\times128\times128$	3	56.38
SAM	$5N$ points	$1024\times1024\times N$	25	44.72
SAM-Med2D	$5N$ points	$256\times256\times N$	10	55.10
SAM-Med3D	5 points	$128\times128\times128$	4	58.57
SAM-Med3D	10 points	$128\times128\times128$	6	60.94

$$\frac{2*\text{预测正确的结果}}{\text{真实结果}+\text{预测结果}}$$

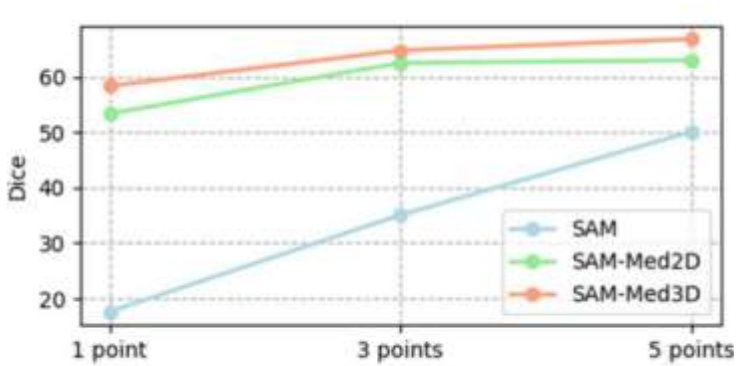
- Our experiments reveal that SAM-Med2D, which is SAM fine-tuned with medical domain knowledge, clearly outperforms SAM

4.2 Evaluation on Different Anatomical Structures

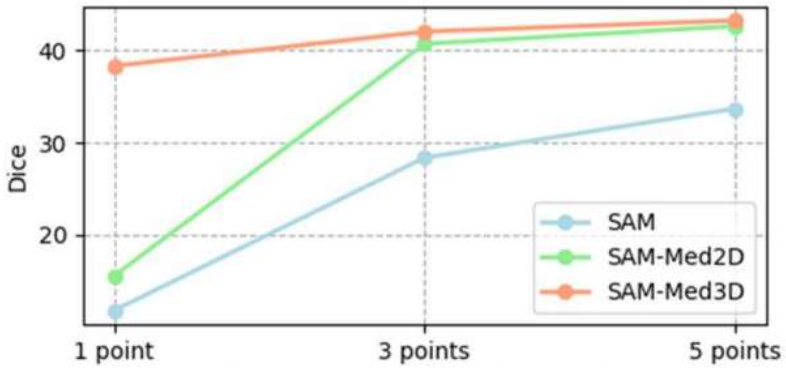
Model	Prompt	Anatomical Structure						Lesion 病变	
		A&T	Bone	Brain	Cardiac	Gland	Muscle	Seen	Unseen
SAM	N points	17.19	22.32	17.68	2.82	11.62	3.50	12.03	8.88
SAM-Med2D	N points	46.79	47.52	19.24	32.23	43.55	35.57	26.08	44.87
SAM-Med3D	1 point	46.80	54.77	34.48	46.51	57.28	53.28	42.02	40.53
SAM	$3N$ points	28.81	47.60	35.63	4.84	20.53	7.32	25.42	14.21
SAM-Med2D	$3N$ points	53.65	63.36	37.71	34.77	49.64	52.84	43.79	46.87
SAM-Med3D	3 points	52.01	63.64	37.82	49.51	60.94	61.10	47.36	44.92
SAM	$5N$ points	39.57	65.87	37.68	8.03	33.25	15.06	39.95	23.22
SAM-Med2D	$5N$ points	54.14	63.78	38.81	34.83	49.92	53.24	45.15	47.66
SAM-Med3D	5 points	53.60	66.54	39.08	51.37	62.66	64.22	48.75	45.71
SAM-Med3D	10 points	55.81	69.13	40.71	52.86	65.01	67.28	50.52	48.44

- A&T represents Abdominal and Thorax targets. N denotes the count of slices containing the target object.
- As the number of prompt points increases, our SAM-Med3D maintains a leading position in the segmentation of most anatomical structures.

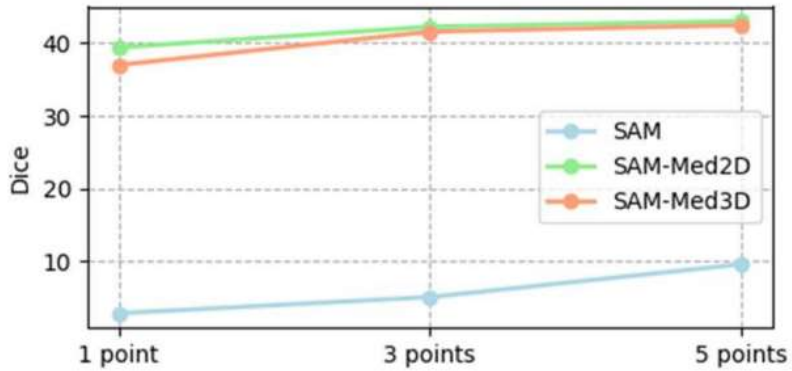
4.3 Evaluation on Different Modalities



(a) Performance on CT Images



(b) Performance on MRI Images

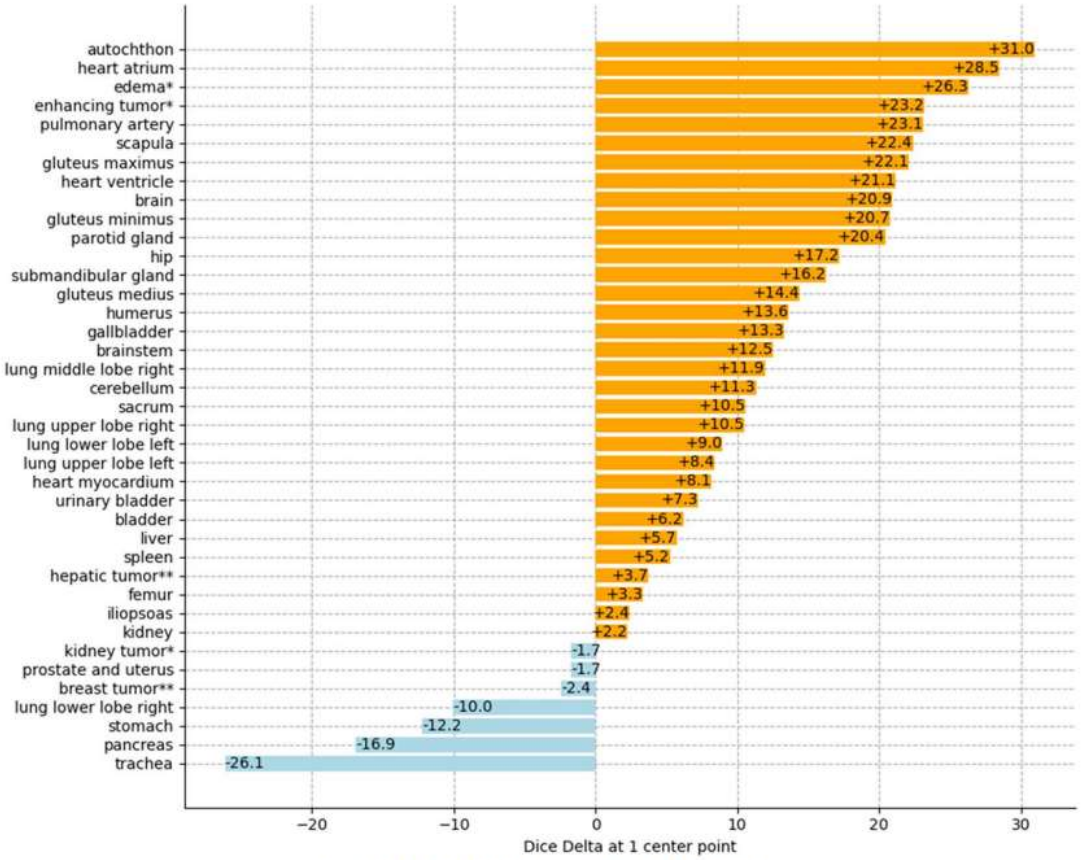


(c) Performance on US Images

- Even after the adaptation to medical images, the first-click performance of 2D SAM methods on MRI images remains sub-optimal when compared to 3D methods that holistically work on the 3D images.

4.4 Evaluation on Major Organs and Lesions

- Presents the comparison between our SAM-Med3D using 1 point and SAM-Med2D using N points per volume.



(d) SAM-Med3D *vs.* SAM-Med2D

4.5 Evaluation on Transferability

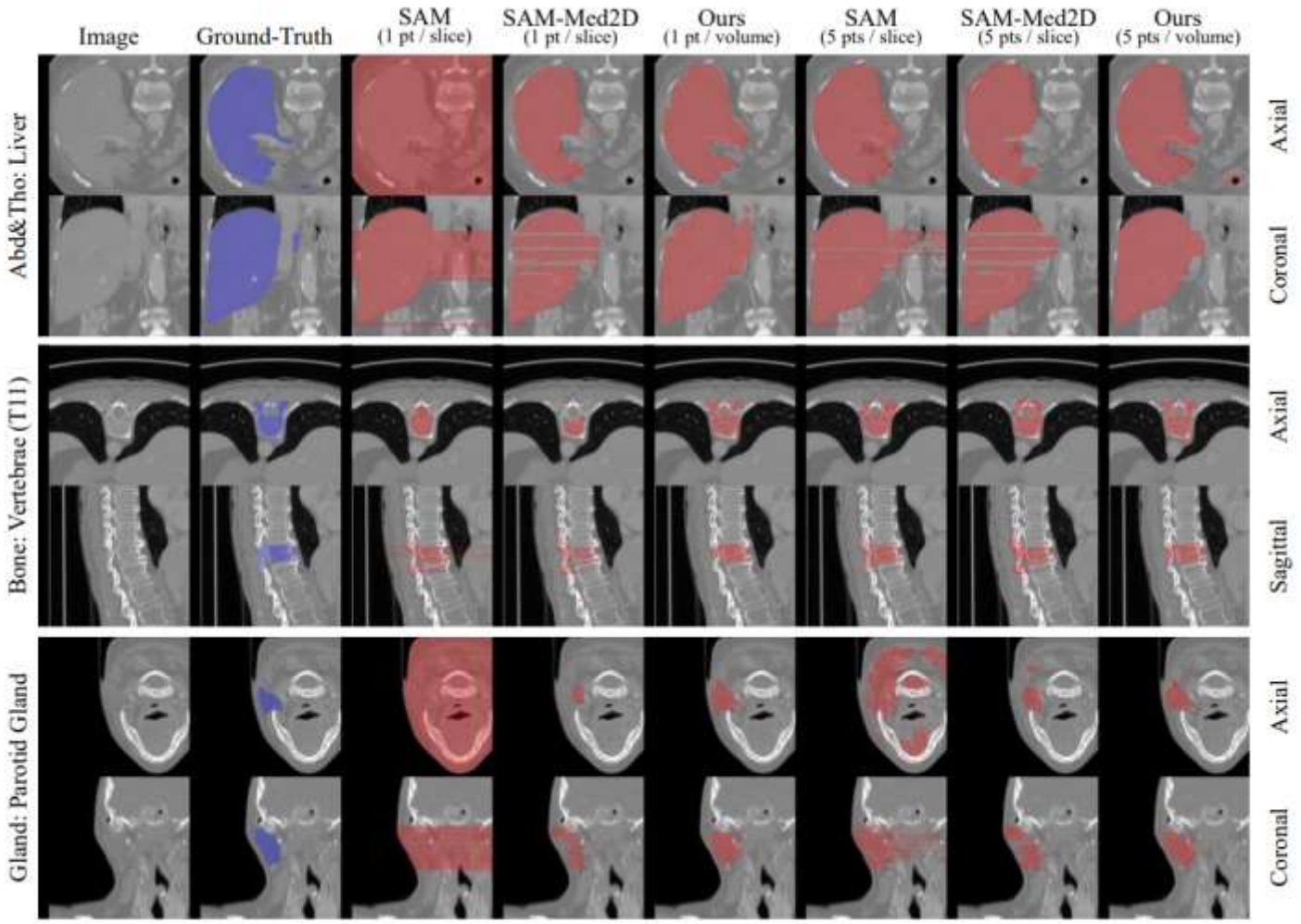
Dataset	UNETR	
	w/o pre-train	w/ pre-train
AMOS [19]	76.29	81.92
Totalsegmentator [44]	82.67	85.17
Average	79.48	83.55

UNETR: Transformers for 3D Medical Image Segmentation

- Conduct a test of transferability on two frequently-used benchmarks for 3D medical image segmentation based on UNETR.

4.6 Visualization

- Abd&Tho denotes Abdominal and Thorax.
- SAM-Med3D requires significantly fewer prompts.
- SAM-Med3D exhibits better inter-slice consistency.



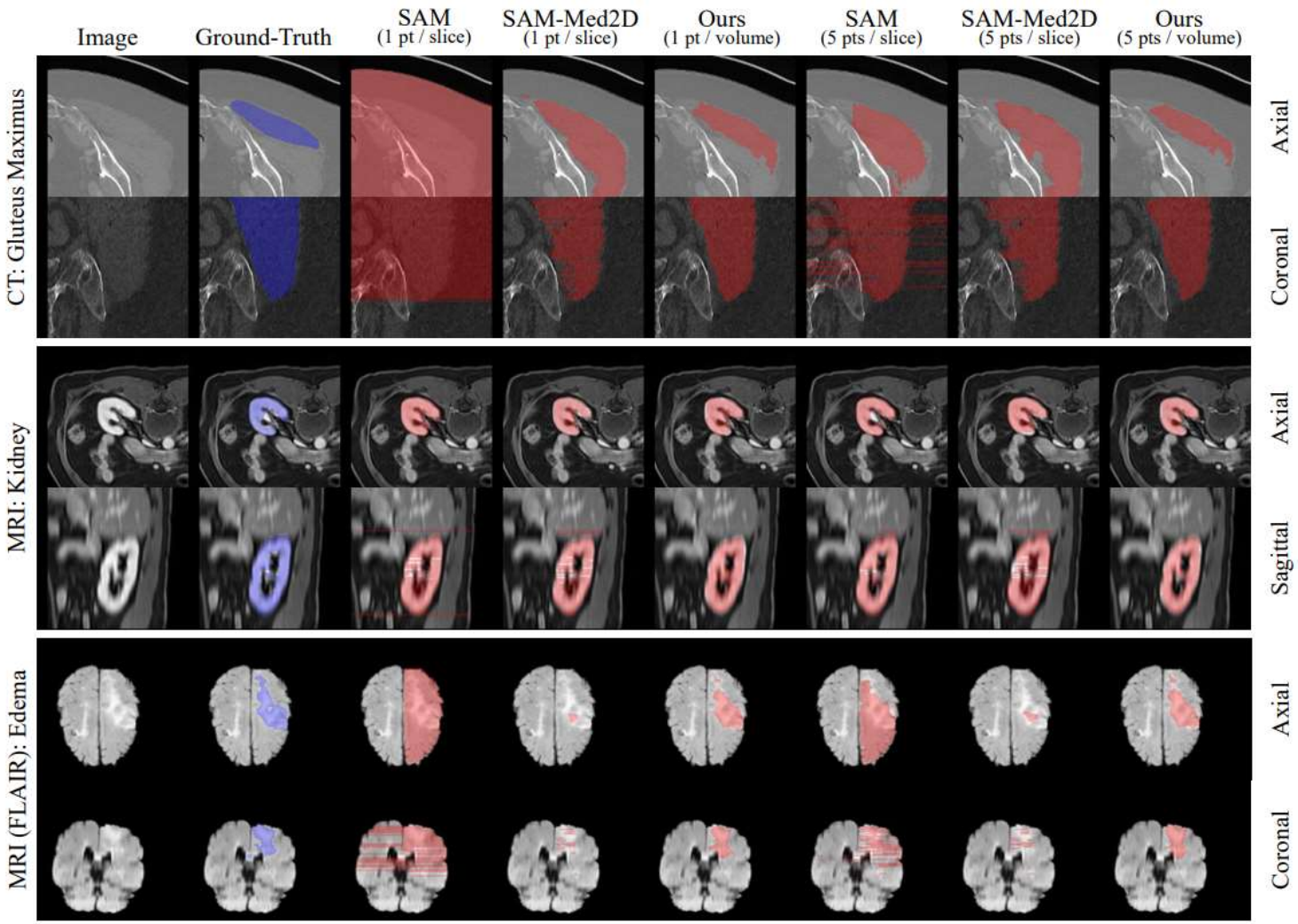


Image
Ground-Truth
SAM (1 pt / slice)
SAM-Med2D (1 pt / slice)
Ours (1 pt / volume)
SAM (5 pts / slice)
SAM-Med2D (5 pts / slice)
Ours (5 pts / volume)
CT: Gluteus Maximus
MRI: Kidney
MRI (FLAIR): Edema
Axial
Coronal
Axial
Sagittal
Axial
Coronal

5. Conclusion

- Present **SAM-Med3D**, a holistic 3D SAM model for volumetric medical image segmentation.
- SAM-Med3D achieves a **32.90%** improvement than SAM when provided with 1 point per volume.
- For various anatomical structures like bone, heart and muscle, our SAM-Med3D outperforms other methods with a clear margin when limited prompt is provided.

Thank You!