SAM-Med2D

Minwoo Jung
mai-lab.net, *Medical Artificial Intelligence Laboratory*
Electrical and Electronic Engineering
Yonsei University

Medical Artificial
Intelligence Laboratory
At Yonsei University

Segmentation

SAM-Med2D

ArXiv

2023

arXiv:2308.16184v1 [cs.CV] 30 Aug 2023

SAM-Med2D

Junlong Cheng[1,2*] Jin Ye[2] Zhongying Deng[2] Jianpin Chen[2] Tianbin Li[2]
Haoyu Wang[2] Yanzhou Su[2] Ziyan Huang[2] Jilong Chen[1] Lei Jiang[1]
Hui Sun[2] Junjun He[2] Shaoting Zhang[2] Min Zhu[1,†] Yu Qiao[2,†]

[1]Sichuan University
[2]Shanghai AI Laboratory
chengjunlong@scu.stu.edu.cn
{yejin, hejunjun, litianbin, zhangshaoting, qiaoyu}@pjlab.org.cn

Abstract

The Segment Anything Model (SAM) represents a state-of-the-art research advancement in natural image segmentation, achieving impressive results with input prompts such as points and bounding boxes. However, our evaluation and recent research indicate that directly applying the pretrained SAM to medical image segmentation does not yield satisfactory performance. This limitation primarily arises from significant domain gap between natural images and medical images. To bridge this gap, we introduce SAM-Med2D, the most comprehensive studies on applying SAM to medical 2D images. Its comprehensiveness manifests in three aspects: the comprehensive analysis on collecting the largest medical data, the most comprehensive studies on various fine-tuning options, the most comprehensive evaluation on the performance. Specifically, we first collect and curate approximately 4.6M images and 19.7M masks from public and private datasets, constructing a large-scale medical image segmentation dataset encompassing various modalities and objects. Then, we comprehensively fine-tune SAM on this dataset and turn it into SAM-Med2D. Unlike previous methods that only adopt bounding box or point prompts as interactive segmentation approach, we adapt SAM to medical image segmentation through more comprehensive prompts involving bounding boxes, points, and masks. We additionally fine-tune the encoder and decoder of the original SAM to obtain a well-performed SAM-Med2D, leading to the most comprehensive fine-tuning strategies to date. Finally, we conducted a comprehensive evaluation and analysis to investigate the performance of SAM-Med2D in medical image segmentation across various modalities, anatomical structures, and organs. Concurrently, we validated the generalization capability of SAM-Med2D on 9 datasets from MICCAI 2023 challenge. Overall, our approach demonstrated significantly superior performance and generalization capability compared to SAM. Our codes can be found at https://github.com/uni-medical/SAM-Med2D.

1 Introduction

Medical image segmentation plays a crucial role in the analysis of medical images by identifying

Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., ... & Qiao, Y. (2023). Sam-med2d. *arXiv preprint arXiv:2308.16184*.

Segmentation

SAM-Med2D

Background & Goal

Background

- Existing models are limited to specific modalities, organs, or lesions due to characteristics of medical imagery
- Due to the significant domain gap between natural images and medical images, SAM struggles to generalize to multi-modal and multi-object medical datasets

Goal

- To transfer SAM from natural images to medical images

Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., ... & Qiao, Y. (2023). Sam-med2d. *arXiv preprint arXiv:2308.16184*.

SAM-Med2D

Methods

Incorporation of Medical Knowledge into SAM

- Collected and curated the largest medical image segmentation dataset
- 3D datasets: Normalized the intensity values & extracted all slice images and corresponding masks
- 2D datasets: Checked whether pixel values were within the range [0, 255] & saved in PNG format & excluded masks where the target area constituted less than 0.153% of the total image

→ Obtained approximately 4.6M images and 19.7M masks

Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., ... & Qiao, Y. (2023). Sam-med2d. *arXiv preprint arXiv:2308.16184*.

Segmentation

SAM-Med2D

Methods

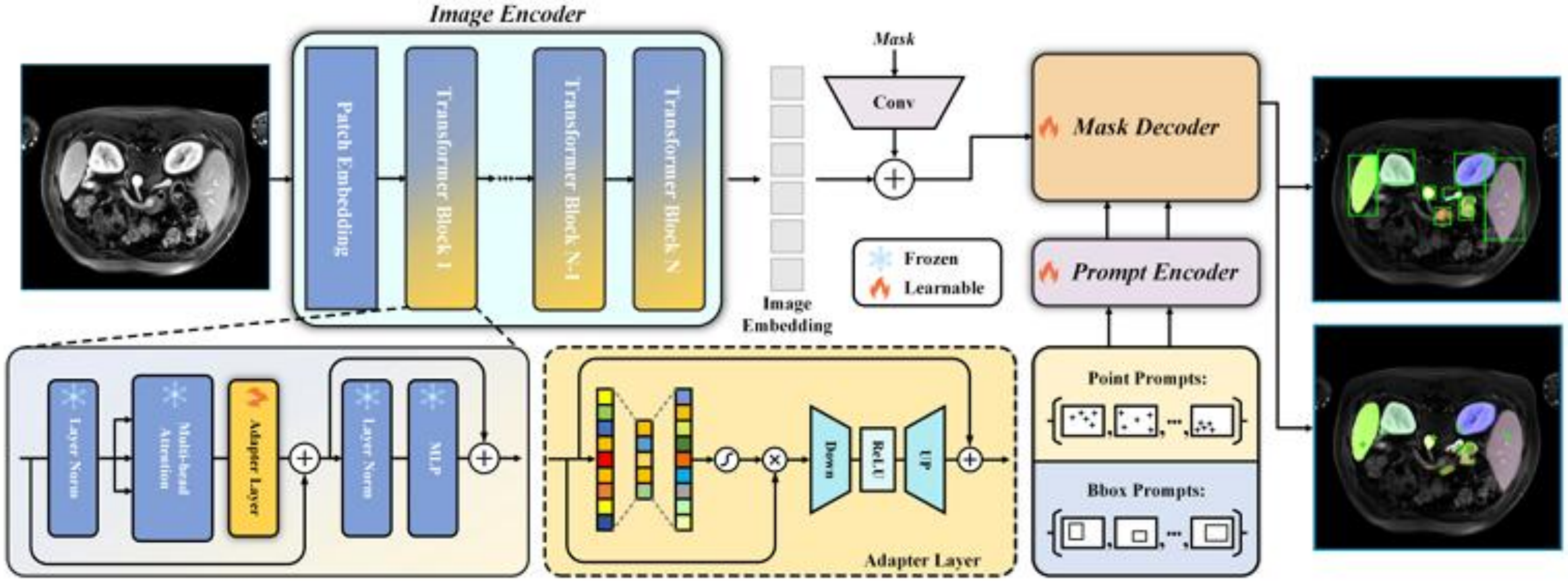


Fig 1. Overview of SAM-Med2D

Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., ... & Qiao, Y. (2023). Sam-med2d. *arXiv preprint arXiv:2308.16184*.

SAM-Med2D

Methods

Adapting Image Encoder

- Freeze all parameters of the original image encoder and deploy an adapter for each Transformer block
- Adapt the image encoder along both channel and spatial dimensions

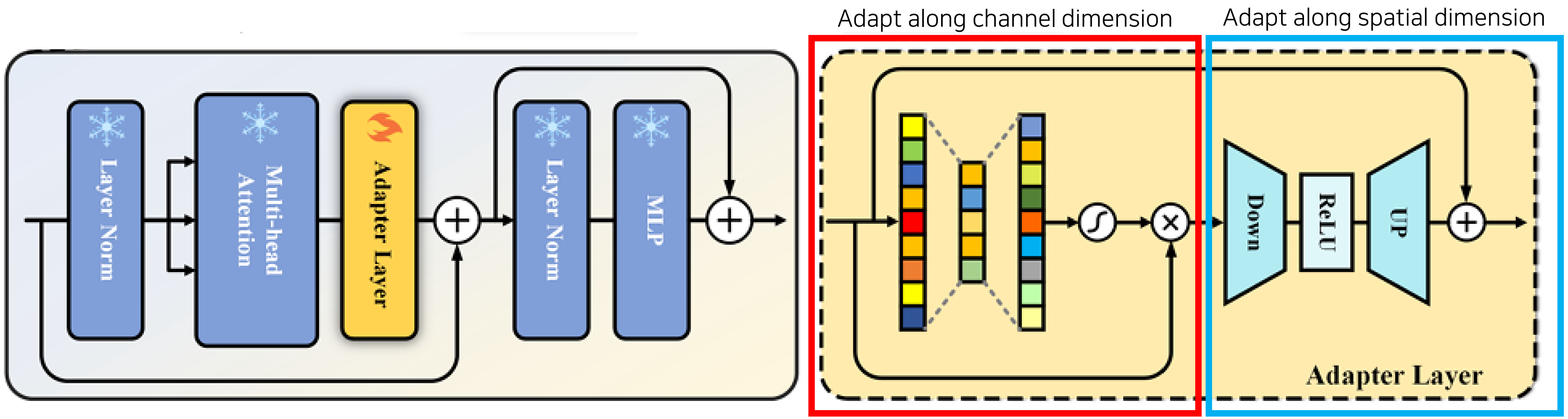


Fig 2. Overview of Image Encoder

Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., ... & Qiao, Y. (2023). Sam-med2d. *arXiv preprint arXiv:2308.16184*.

Segmentation

SAM-Med2D

Methods

Prompt Encoder

- Point mode: Represented as the sum of two learned embeddings indicating its foreground/background positions
- Bounding-box mode: Uses the positional encoding of its top-left and bottom-right corners
- Mask mode: Uses the low-resolution feature map generated after the first iteration as the mask prompt

Mask Decoder

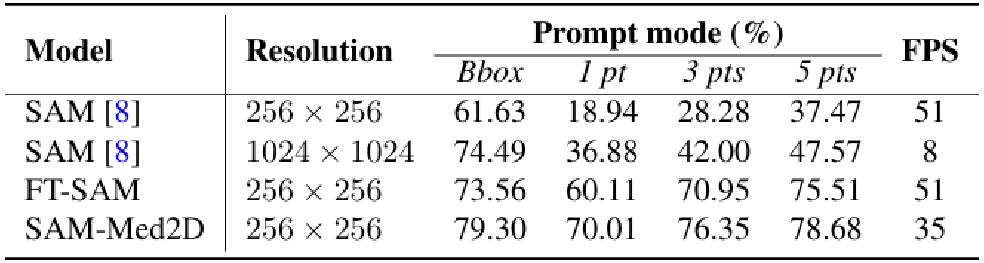
- Didn’t make any changes to the mask decoder structure and kept updating its parameters during training
- To make the model ambiguity-aware, prompt predicts multiple masks & Compute the loss through the prediction mask with the highest IoU score and propagate the gradient

Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., ... & Qiao, Y. (2023). Sam-med2d. *arXiv preprint arXiv:2308.16184*.

Segmentation

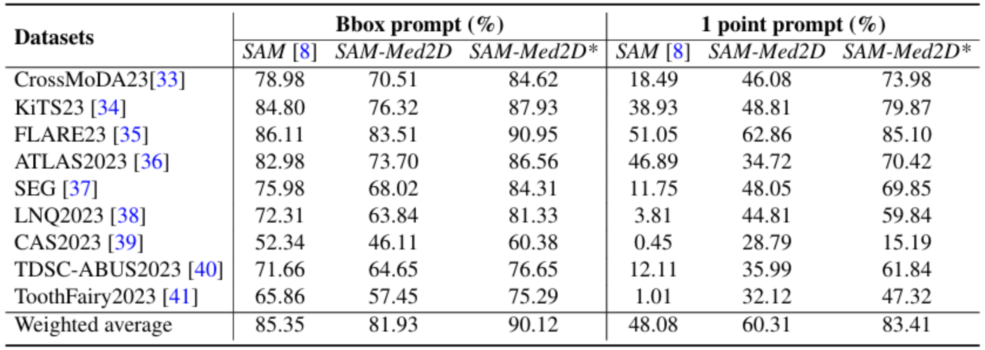
SAM-Med2D

Results



Model	Resolution	Prompt mode (%)				FPS
		Bbox	*1 pt*	*3 pts*	*5 pts*	
SAM [8]	256 × 256	61.63	18.94	28.28	37.47	51
SAM [8]	1024 × 1024	74.49	36.88	42.00	47.57	8
FT-SAM	256 × 256	73.56	60.11	70.95	75.51	51
SAM-Med2D	256 × 256	79.30	70.01	76.35	78.68	35

Table 1. Quantitative comparison of different methods



Datasets	Bbox prompt (%)			1 point prompt (%)		
	SAM [8]	*SAM-Med2D*	*SAM-Med2D**	*SAM* [8]	*SAM-Med2D*	*SAM-Med2D**
CrossMoDA23[33]	78.98	70.51	84.62	18.49	46.08	73.98
KiTS23 [34]	84.80	76.32	87.93	38.93	48.81	79.87
FLARE23 [35]	86.11	83.51	90.95	51.05	62.86	85.10
ATLAS2023 [36]	82.98	73.70	86.56	46.89	34.72	70.42
SEG [37]	75.98	68.02	84.31	11.75	48.05	69.85
LNQ2023 [38]	72.31	63.84	81.33	3.81	44.81	59.84
CAS2023 [39]	52.34	46.11	60.38	0.45	28.79	15.19
TDSC-ABUS2023 [40]	71.66	64.65	76.65	12.11	35.99	61.84
ToothFairy2023 [41]	65.86	57.45	75.29	1.01	32.12	47.32
Weighted average	85.35	81.93	90.12	48.08	60.31	83.41

Table 2. Generalization validation on 9 MICCAI2023 datasets

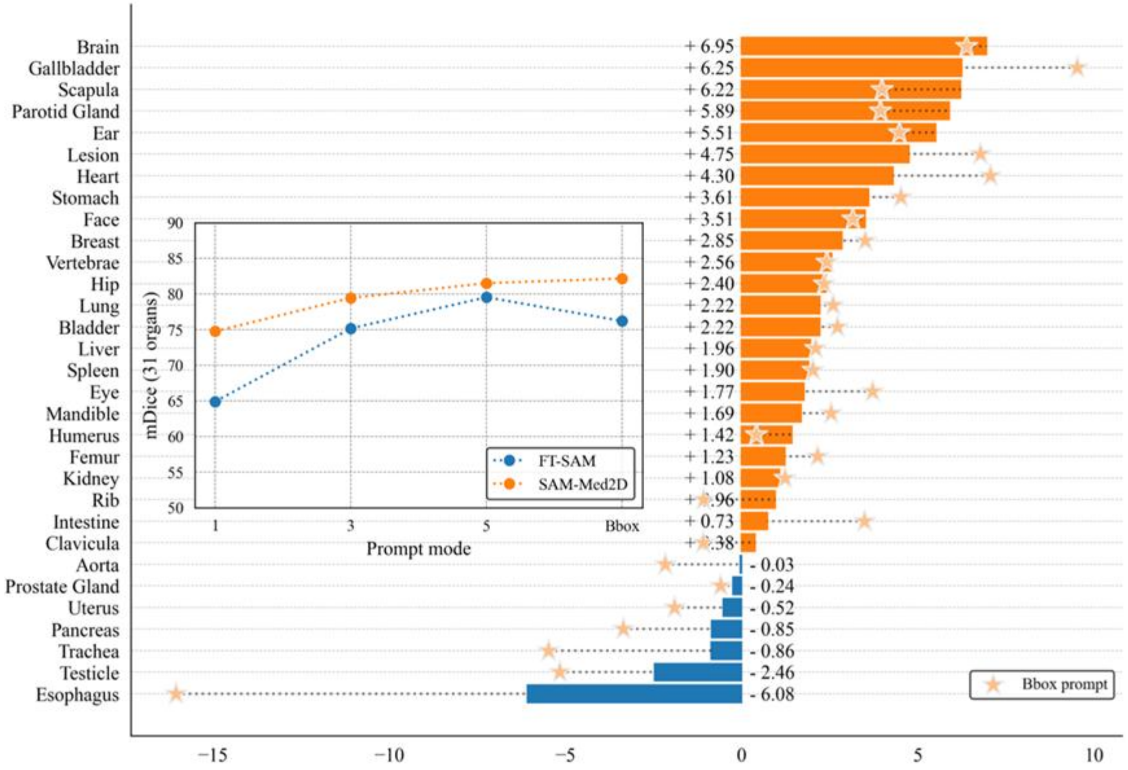


Fig 3. Comparison of segmentation performance between FT-SAM and SAM-Med2D across 31 organs

Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., … & Qiao, Y. (2023). Sam-med2d. *arXiv preprint arXiv:2308.16184*.

Segmentation

SAM-Med2D

Implications & Limitations

Implications

- Fine-tuned the SAM on a large-scale medical image dataset to adapt it to the medical image domain
- SAM-Med2D achieved satisfactory performance improvements and generalization capabilities

Limitations

- For complex shapes, small size or low contrast objects, there is still room for improvement
- Natural language can serve as another common form of user interaction, but there is a lack of relevant datasets

Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., ... & Qiao, Y. (2023). Sam-med2d. *arXiv preprint arXiv:2308.16184*.