Segment anything model for medical image analysis: An experimental study

Yu Jiening
Yu Jiening@umac.mo

Medical Image Analysis, 2023, 89: 102918.

ELSEVIER

Medical Image Analysis
Volume 89, October 2023, 102918

Segment anything model for medical image analysis: An experimental study

Maciej A. Mazurowski [a b c d], Haoyu Dong [b], Hanxue Gu [b], Jichen Yang [b], Nicholas Konz [b], Yixin Zhang [b]

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.media.2023.102918

Get rights and content

Highlights

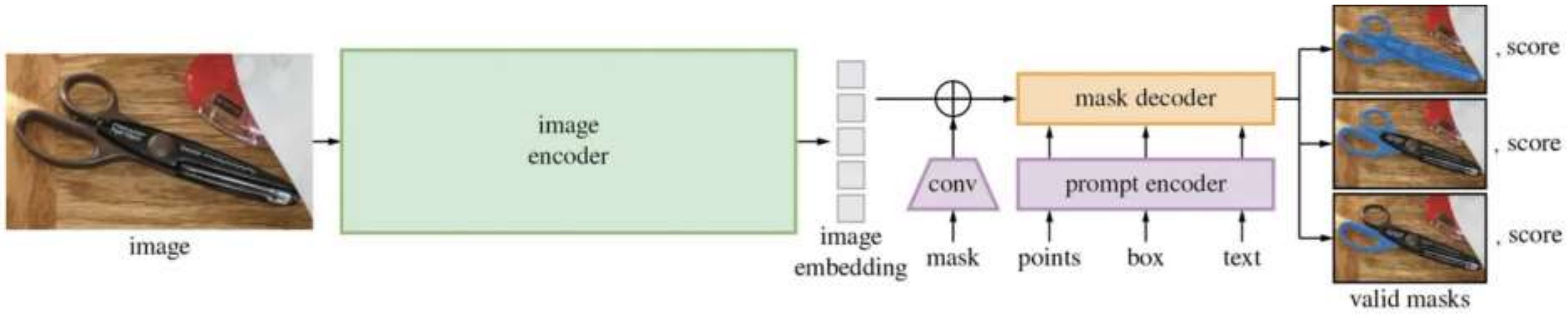
- Segment Anything Model (SAM) is a new algorithm for interactive image segmentation.
- Performance of SAM varies widely on the 19 evaluated medical imaging

1.Introduction

What is SAM（Segment Anything Model）？

Segment Anything Model (SAM) is designed to segment an object of interest in an image given certain prompts provided by a user.

high performance in the zero-shot learning regime.

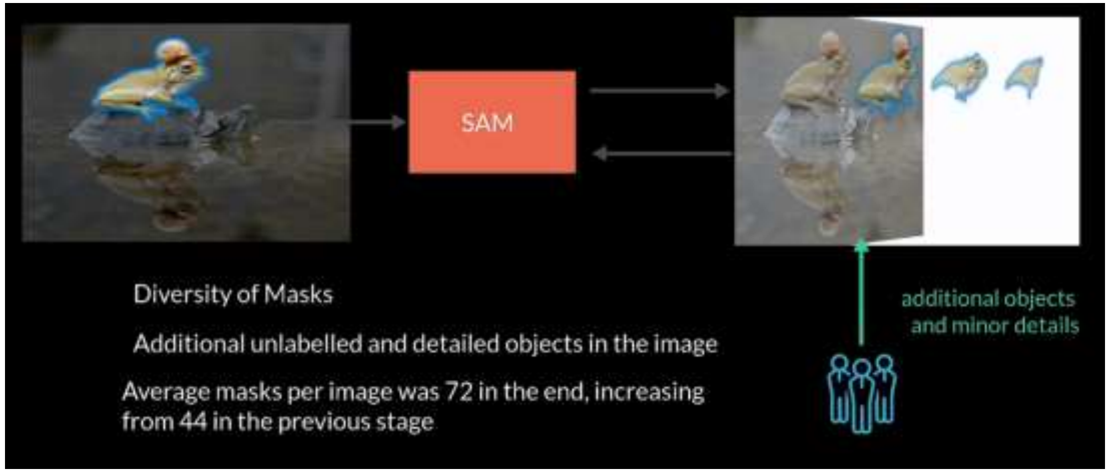


消除模糊性

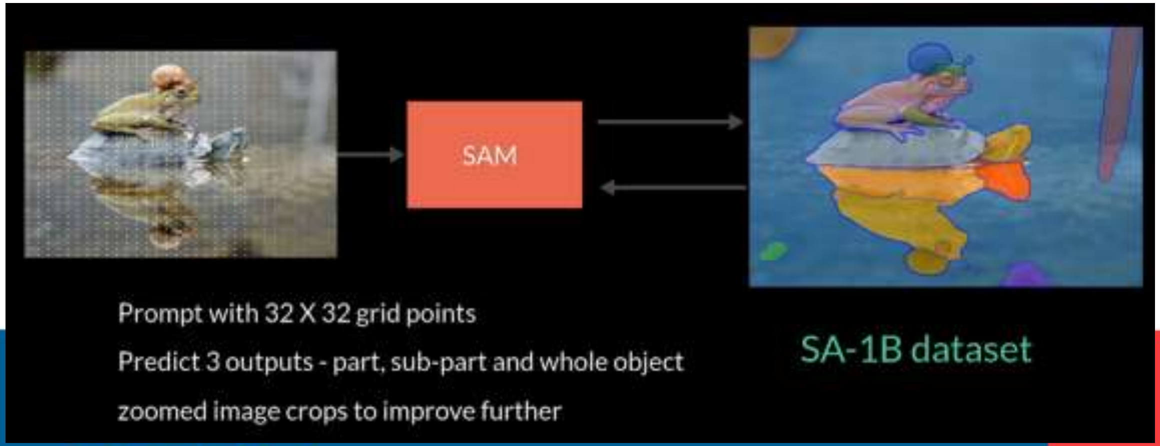
Development of the SA-1B image dataset

1.Assisted-Manul Stage

2.Semi-automatic Stage



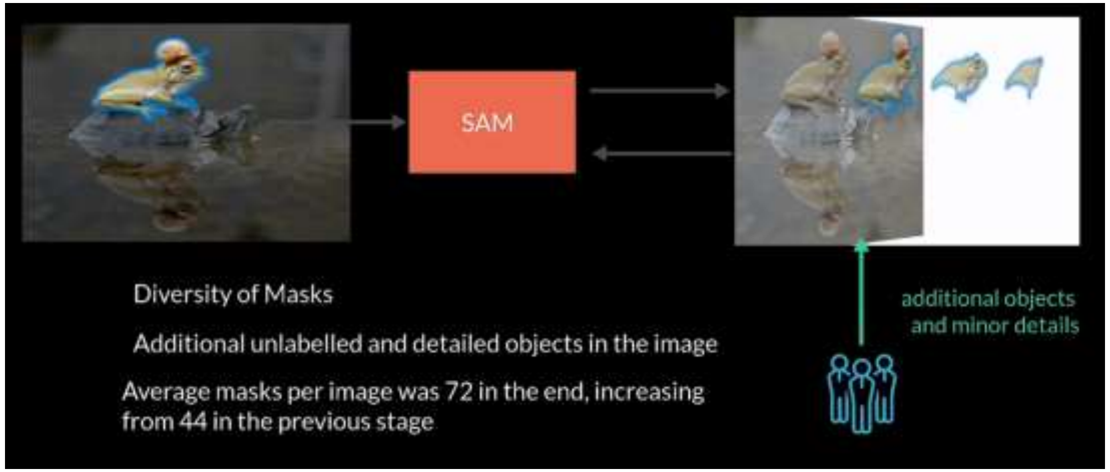
3.Fully automatic Stage



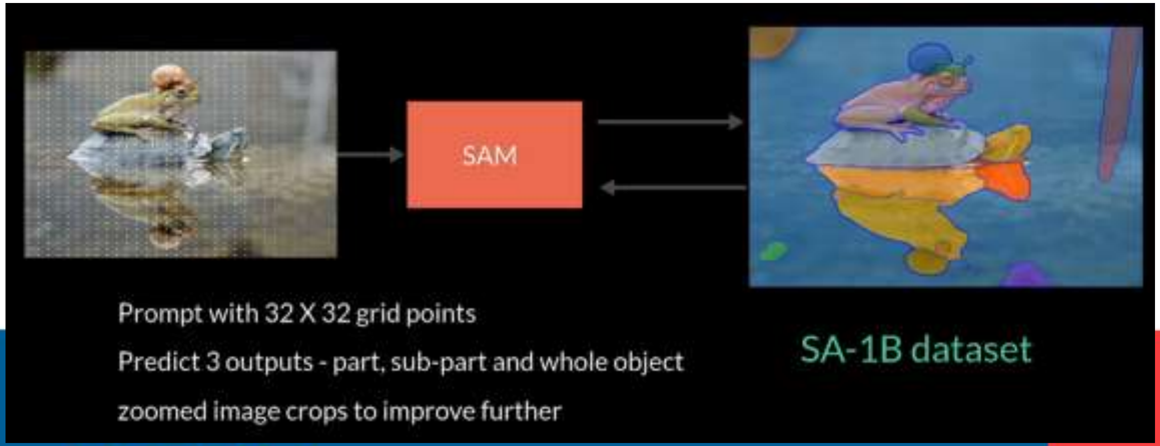
Development of the SA-1B image dataset

1.Assisted-Manul Stage

2.Semi-automatic Stage



3.Fully automatic Stage



All masks are fully automatically
generated by SAM

How to segment medical images with SAM?

- Semi-automated annotation
- SAM assisting other segmentation models
- New medical image foundation segmentation models

2.Method

2.1Datasets

Abbreviated dataset name	Full dataset name and citation	Modality	Num.classes	Object(s) of interest	Num.masks
MRI-Spine	Spinal Cord Grey MatterSegmentation Challenge(Prados et al., 2017)	MRI	2	Gray matter,spinal cord	551
MRI-Heart	Medical Segmentation Decathlon(Simpson et al., 2019)	MRI	1	Heart	1,301
MRI-Prostate	Initiative for CollaborativeComputer Vision Benchmarking(Lemaître et al., 2015)	MRI	1	Prostate	893
MRI-Brain	The Multimodal Brain Tumor ImageSegmentation Benchmark (BraTS)(Menze et al., 2014)	MRI	3	GD-enhancing tumor,Peritumoral edema,necrotic and non-enhancing tumor core	12,591
MRI-Breast	Duke Breast Cancer MRI:Breast + FGT Segmentation(Saha et al., 2018; Hu et al., 2022)	MRI	2	Breast, fibrog-landular tissue	503
Xray-Chest	Montgomery County and ShenzhenChest X-ray Datasets(Jaeger et al., 2014)	X-ray	1	Chest	704
Xray-Hip	X-ray Images of the Hip Joints(Gut, 2021)	X-ray	2	Ilium, femur	140
US-Breast	Dataset of Breast Ultrasound Images(Al-Dhabyani et al., 2020)	Ultrasound	1	Breast	647
US-Kidney	CT2US for Kidney Segmentation(Song et al., 2022)	Ultrasound	1	Kidney	4,586
US-Muscle	Transverse MusculoskeletalUltrasound Image Segmentations(Marzola et al., 2021)	Ultrasound	1	Muscle	4,044
US-Nerve	Ultrasound Nerve Segmentation Identify (Anna et al., 2016)	Ultrasound	1	Nerve	2,323
US-Ovarian-Tumor	Multi-Modality Ovarian TumorUltrasound (MMOTU)(Zhao et al., 2022)	Ultrasound	1	Ovarian tumor	1,469
CT-Colon	Medical Segmentation Decathlon(Simpson et al., 2019)	CT	1	Colon cancerprimaries	1,285
CT-HepaticVessel	Medical Segmentation Decathlon(Simpson et al., 2019)	CT	1	Vessels, tumors	13,046
CT-Pancreas	Medical Segmentation Decathlon(Simpson et al., 2019)	CT	1	parenchymaand mass	8,792
CT-Spleen	Medical Segmentation Decathlon(Simpson et al., 2019)	CT	1	spleen	1,051
CT-Liver	The Liver TumorSegmentation Benchmark (LiTS)(Bilic et al., 2023)	CT	1	Liver	5,501
CT-Organ	CT Volumes with MultipleOrgan Segmentations (CT-ORG)(Rister et al., 2019)	CT	5	Liver, bladder, lungs,kidney, bone	4,776
PET-Whole-Body	A FDG-PET/CT datasetwith annotated tumor lesions(Gatidis et al., 2022)	PET/CT	1	Lesion	1,015

- 19 publicly available medical imaging datasets, includes planar X-rays(平面x射线), magnetic resonance images (MRIs)（磁共振图像）, computed tomography (CT) images（计算机断层扫描）, ultrasound (US) images（超声图像）, and positron emission tomography (PET) images（正电子发射断层扫描图像）.

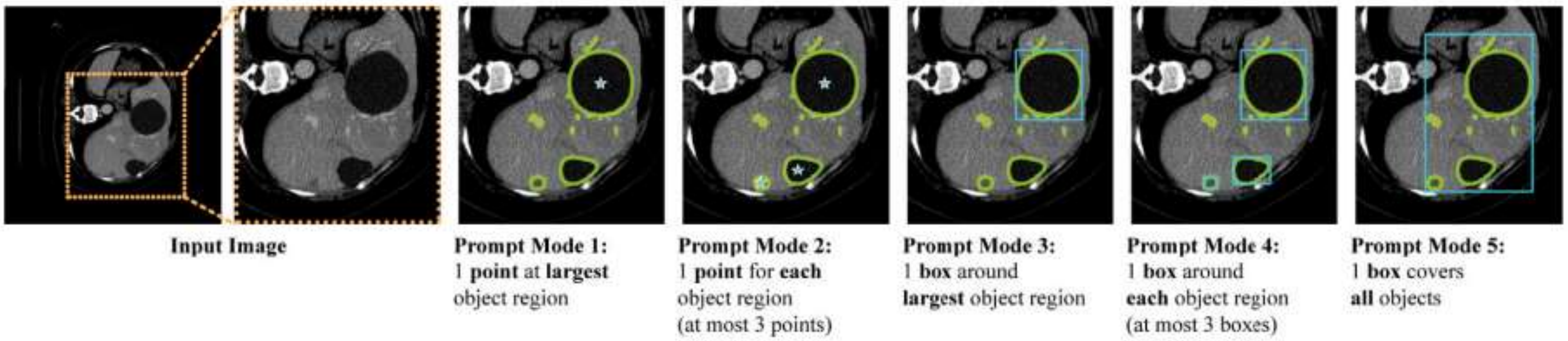
2.2. Data pre-processing

- Normalized: dividing all pixel values by the maximum value in that image and multiplying all pixels by 255.

2.3. Experiments

Prompting strategies

Non-iterative prompts (generated prior to SAM being applied)



five prompting modes

Iterative prompts(generated after seeing the model’s predictions)

Algorithm 1 Prompt Point Generation Scheme　　提示点生成方案

Input: Image $X \in \mathbb{R}^{H \times W}$, ground truth mask $M \in \{0,1\}^{H \times W}$, Segment Anything Model SAM, prompt count N, closest-zero-pixel distance function $d = \texttt{distanceTransform()}$.

将第一个提示点 p1 初始化为掩膜前景中离背景最远的点

Initialize first prompt point p_1 as the point within the mask foreground farthest from the background:

$\mathcal{P} = \underset{(i,j)}{\operatorname{argmax}}(d[(i,j),(k,l)])$ for all $(i,j),(k,l)$ such that $M_{ij} = 1$, $M_{kl} = 0$.　　从公式中得到一组符合上述条件的点

Choose randomly if multiple points satisfy this:　　随意选择一点，作为第一个输入

$p_1 = \texttt{random\_choice}(\mathcal{P})$

Predict mask $Y_1 = \mathrm{SAM}(X, p_1)$　将提示点的坐标输入SAM 得到预测分数最高的一个 mask Y1

Get prediction error region $E_1 = Y_1 \cup M - Y_1 \cap M$　　获取预测误差区域

Subsequent prompt points are those farthest from the boundary of iteratively-updating error region E_n:　　随后的提示点是距离迭代更新误差区域边界最远的点 En

for $n = 2, \ldots, N$ **do**

　$\mathcal{P}_n = \underset{(i,j)}{\operatorname{argmax}}(d[(i,j),(k,l)])$ for all $(i,j),(k,l)$ such that $[E_{n-1}]_{ij} = 1$, $[E_{n-1}]_{kl} = 0$.　**多个提示点是在每次SAM输出的预测 mask 的基础上迭代输入的过程**

　$p_n = \texttt{random\_choice}(\mathcal{P}_n)$

　$Y_n = \mathrm{SAM}(X, p_n)$

　$E_n = Y_n \cup M - Y_n \cap M$

end for

return Prompt points $p_1, \ldots, p_N \in \mathbb{N}^2$

每次迭代得到一个En，从中获得一个点作为输入，更新误差预测区域

Prompt ambiguity.

- Prompts can be ambiguous, so SAM provides multiple outputs to remove the ambiguity.
- The user can select from multiple outputs the one that is closest to the desired object.

Variability in prompt placement

- Observe the performance of SAM when prompt points are at different locations.

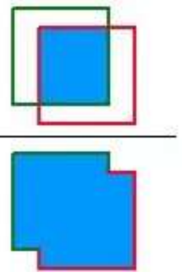
2.3.2. Comparison with other methods

- compared SAM with three interactive segmentation methods, namely **RITM** (Sofiiuk et al., 2021), **SimpleClick** (Liu et al., 2022),and **FocalClick** (Chen et al., 2022).

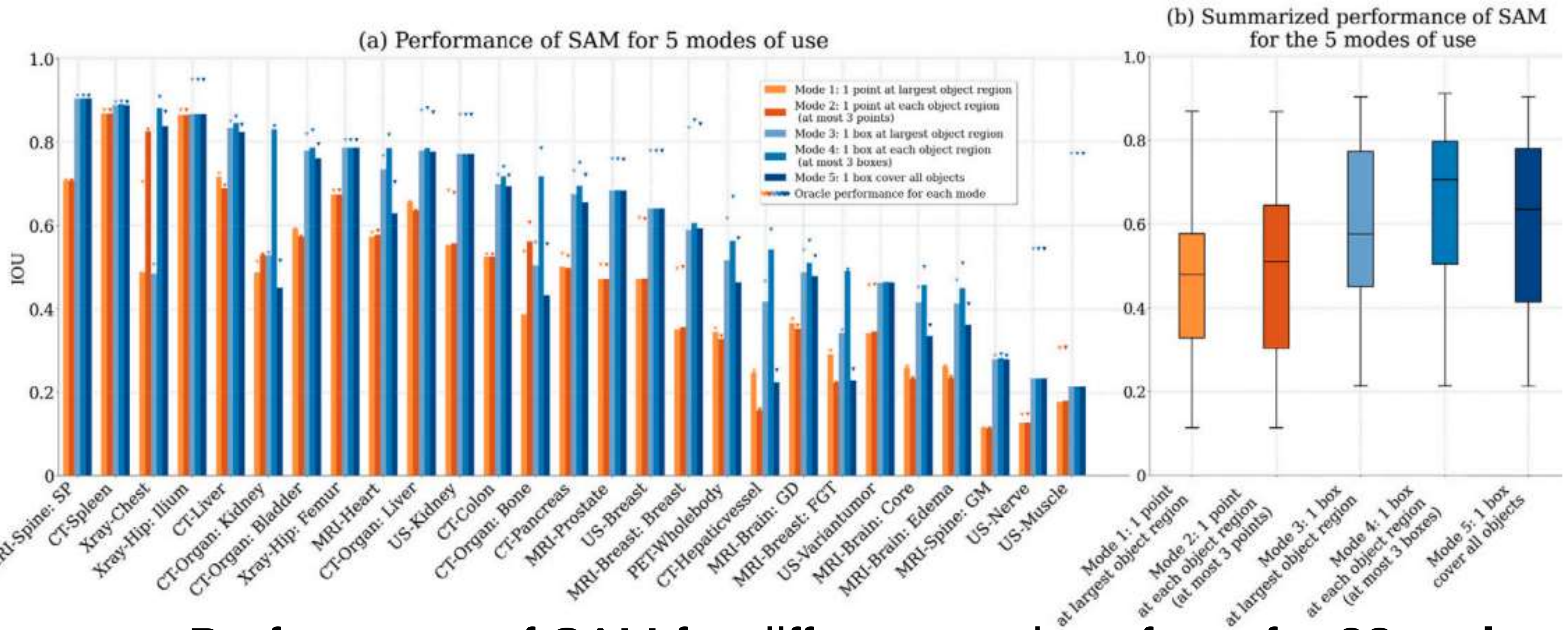
2.3.3. Performance evaluation metric

- Accuracy of generated masks、IoU、confidence interval

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$



3.Results

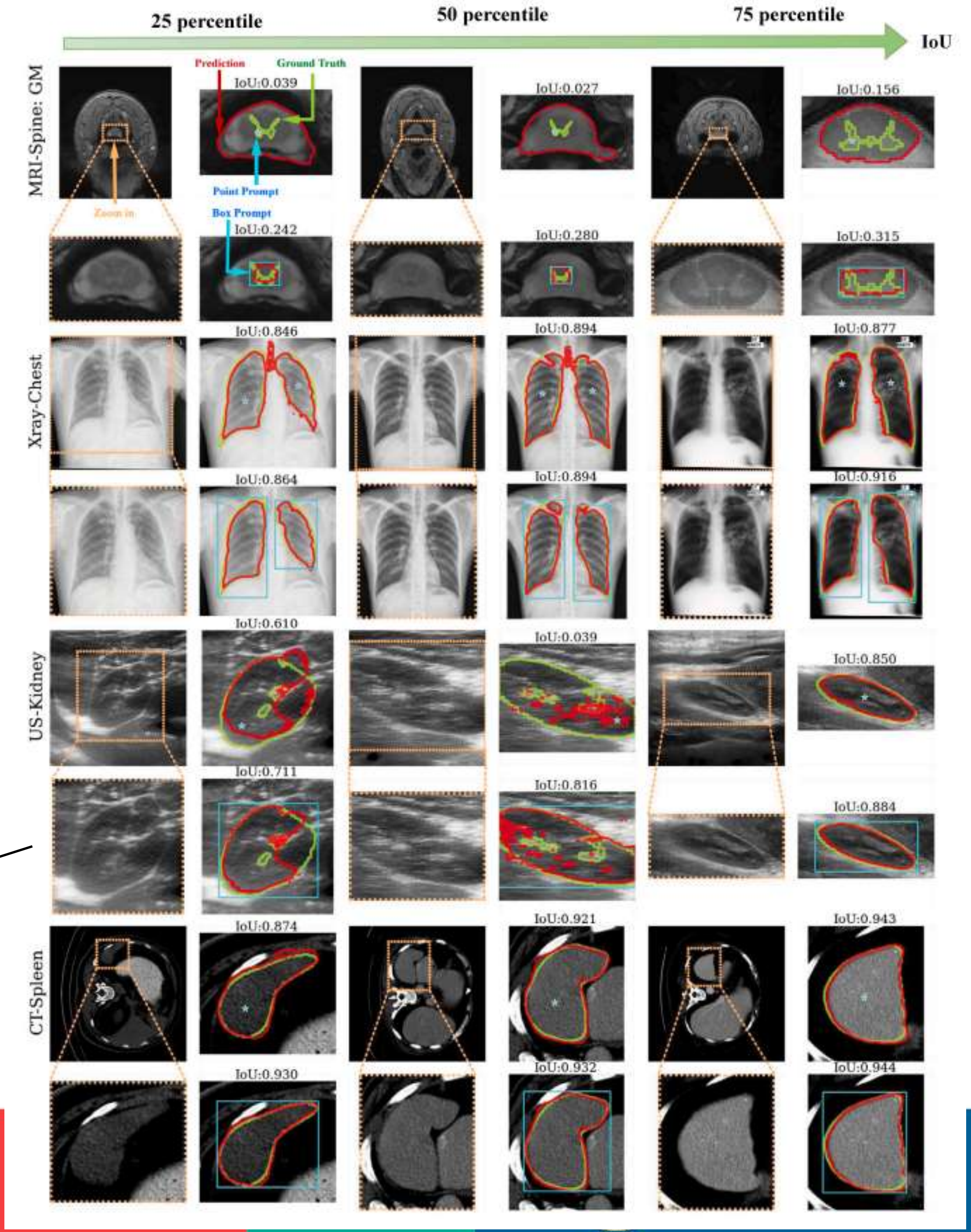


Performance of SAM for different modes of use for **28 tasks**

- The performance of SAM varies greatly across datasets, with IoU ranging from **0.9118 to 0.1136**.
- **Box prompts** are significantly **better** than point prompts.

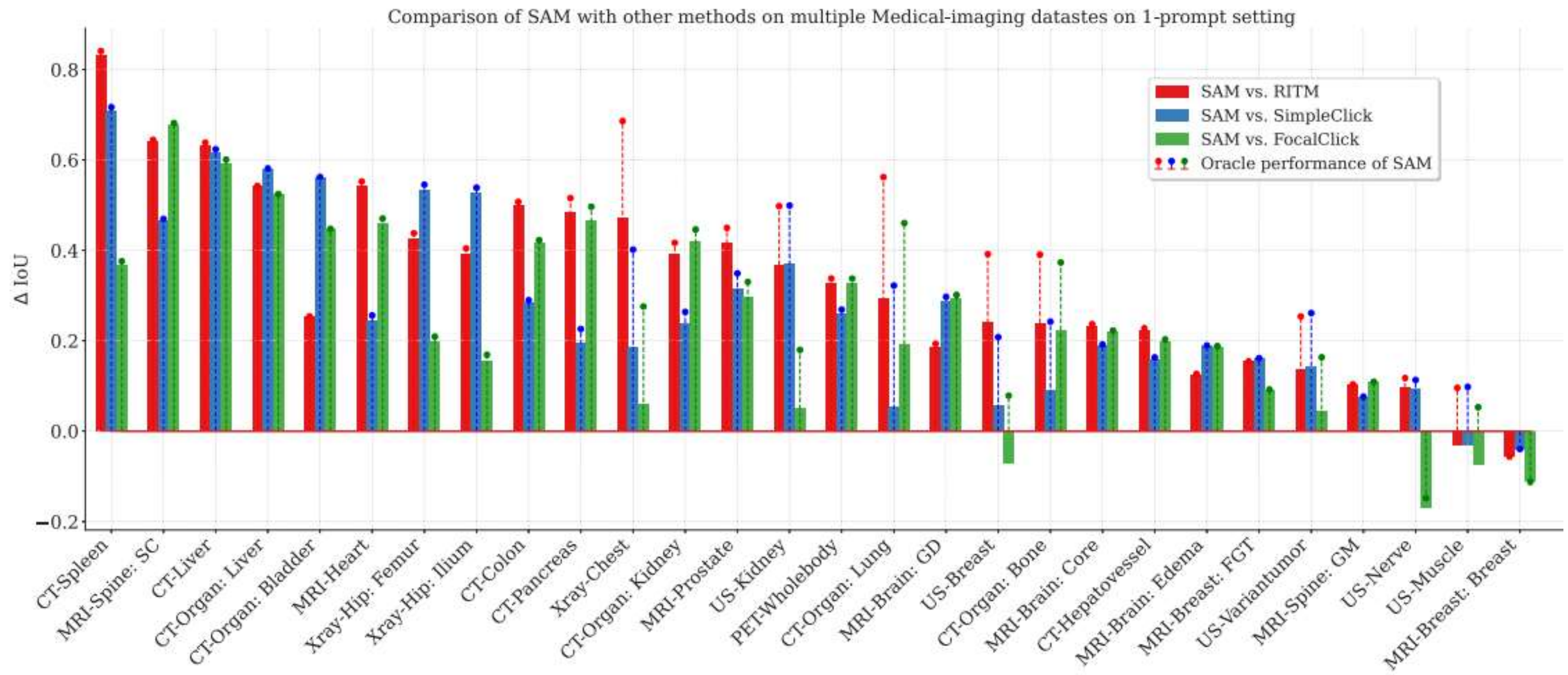
Mode 2 (a point for each object part) and Mode 4 (a box around each object part) for 4 selected datasets

- High variability in SAM performance



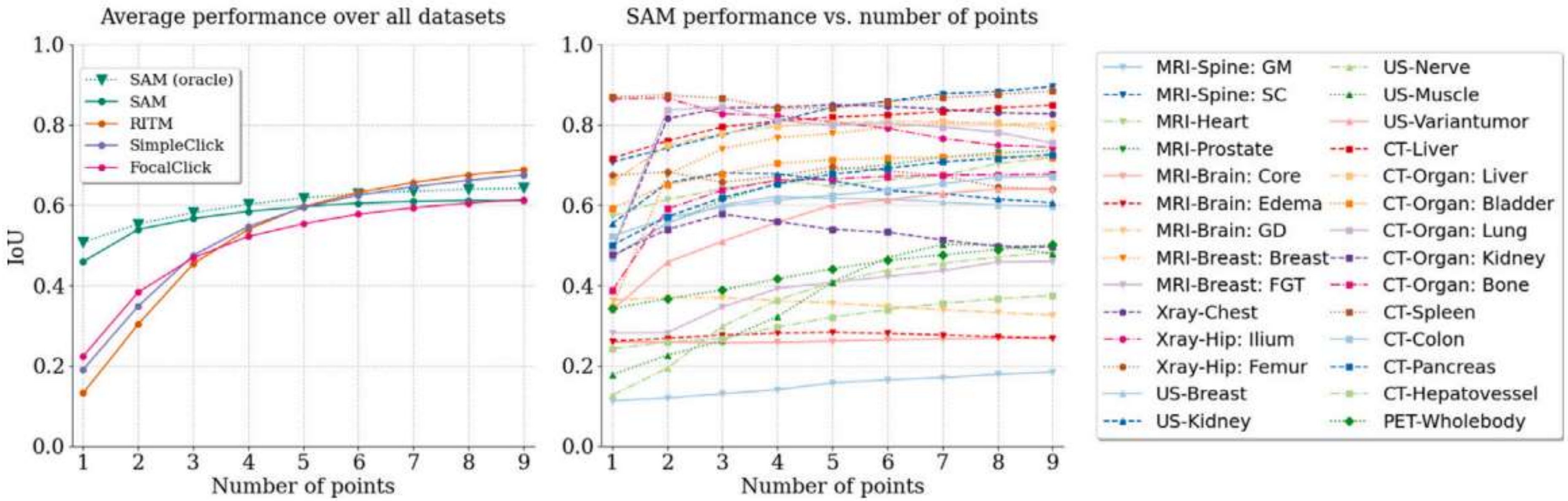
Green: ground truth
Red: prediction of SAM
Blue: box prompt

3.2. Comparing SAM to **other interactive segmentation methods**



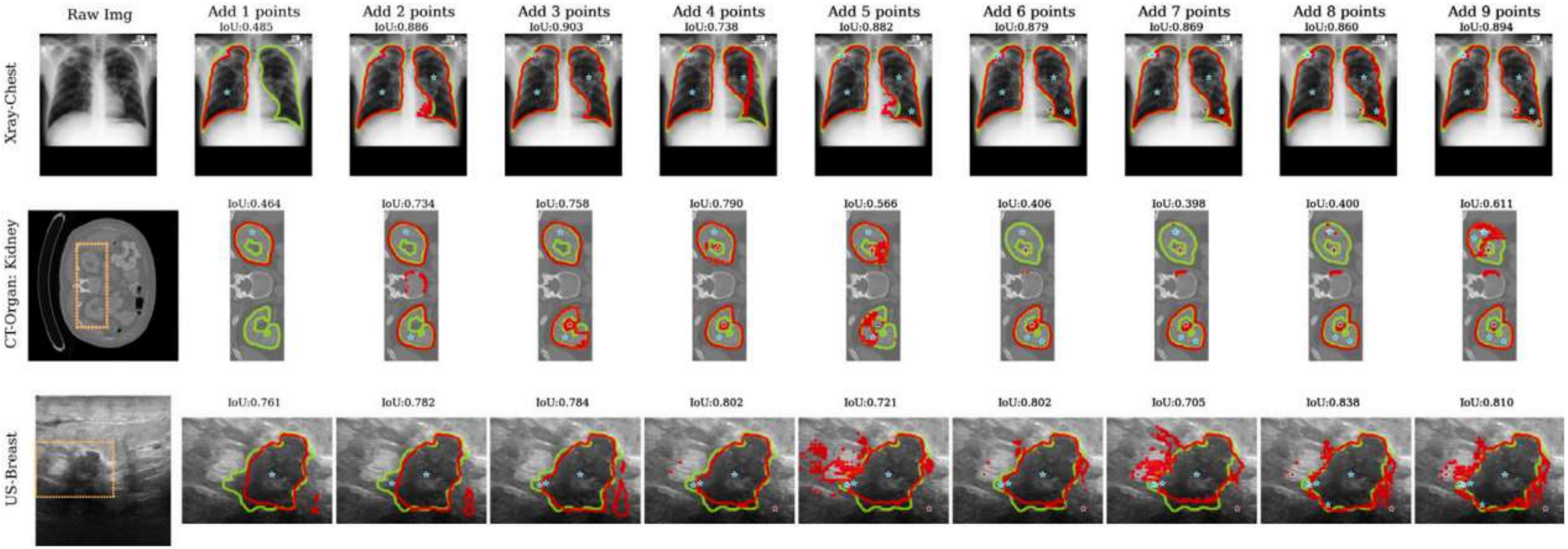
- SAM performed better than other methods on **24** tasks
- If **mode 3** is used, SAM outperforms the other methods in **all tasks**

3.3. Performance of SAM and other methods for iterative segmentation



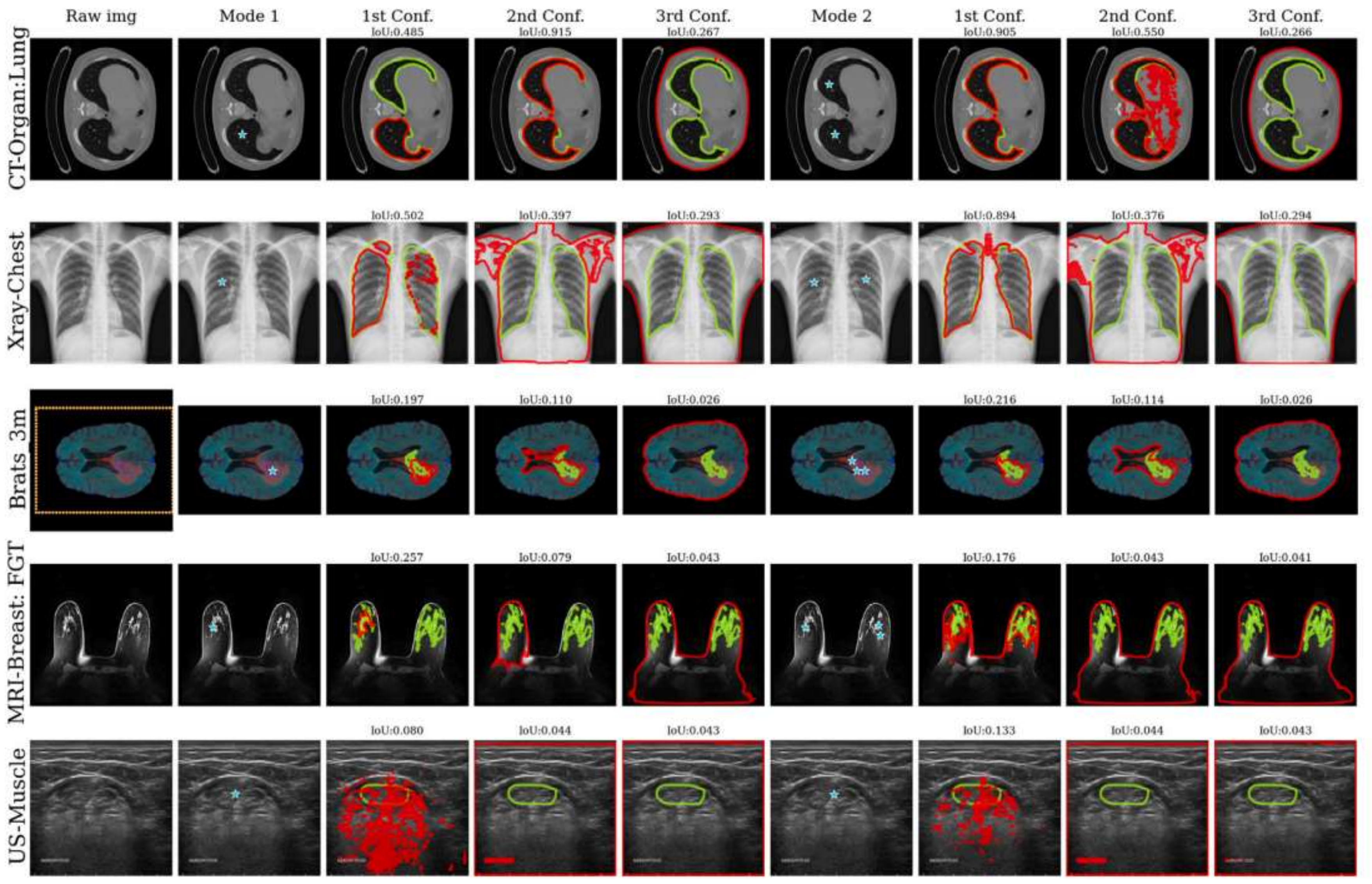
- SAM outperforms other methods in an interactive prompt setup.

- Prompt points increase, but SAM performance does not increase

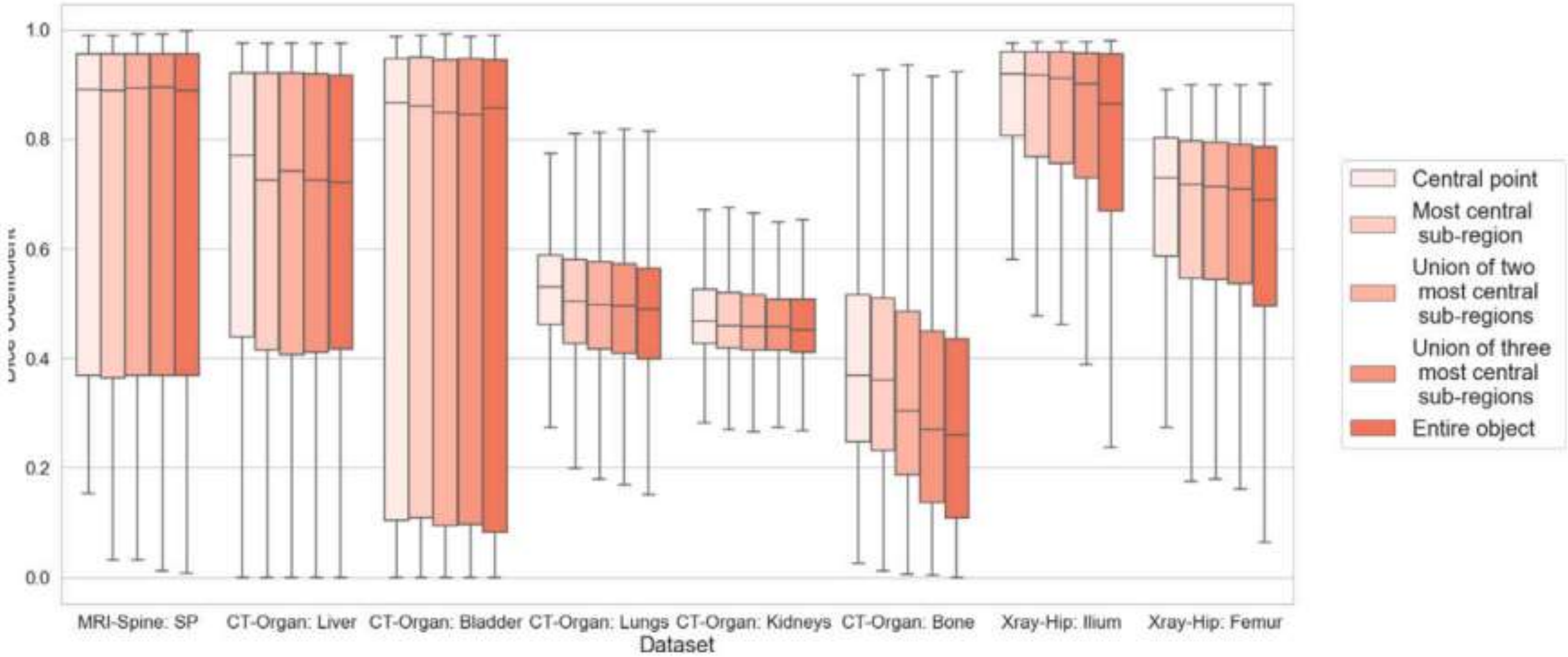


3.4. Performance of SAM in the presence of ambiguity of prompts

- SAM performance is better when confidence is high

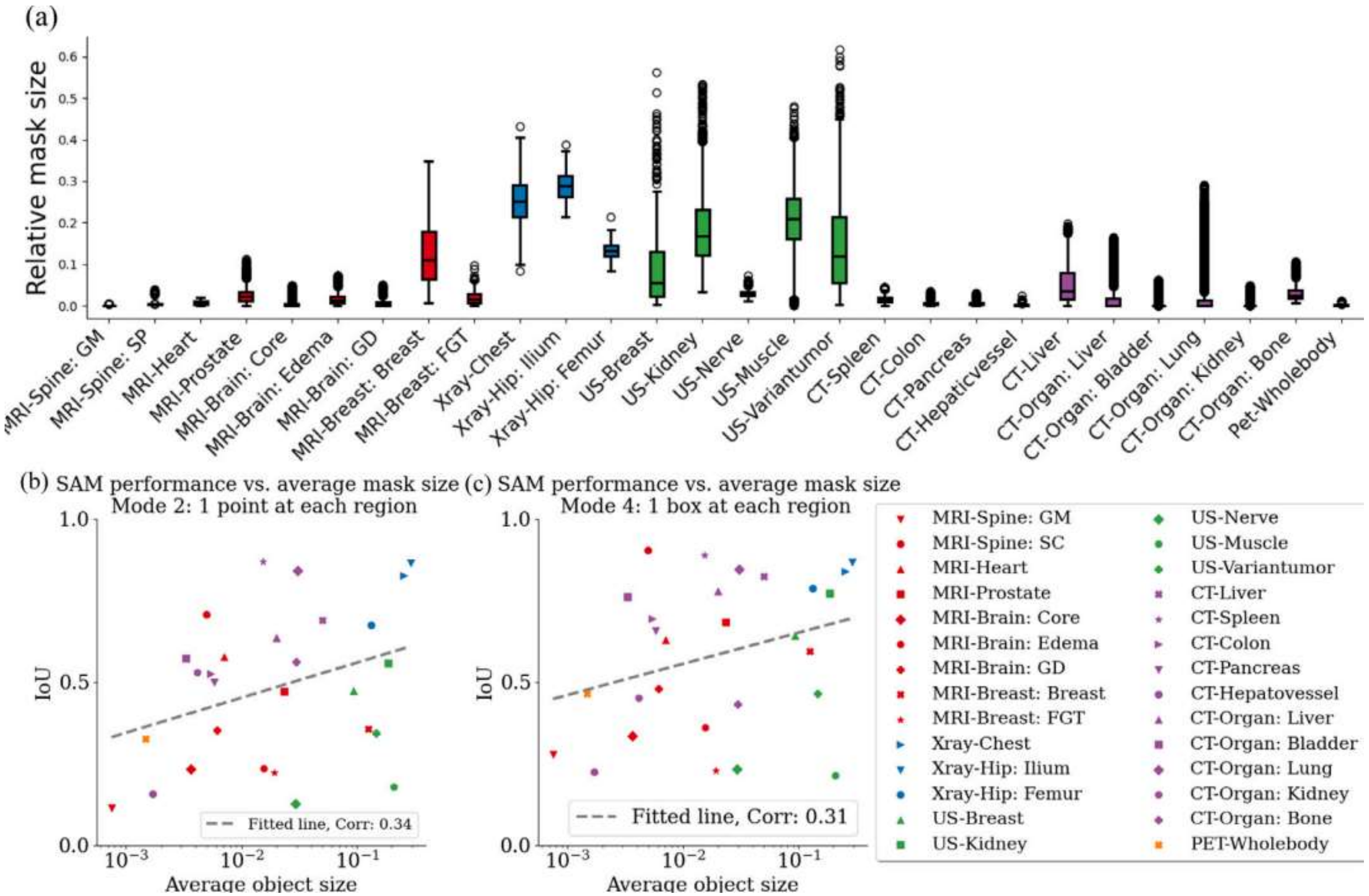


3.5. Performance of SAM given different locations of the prompt within the image



- For some tasks, there was no degradation in SAM performance when the **prompt was not centered**
- For other tasks, SAM performance degrades

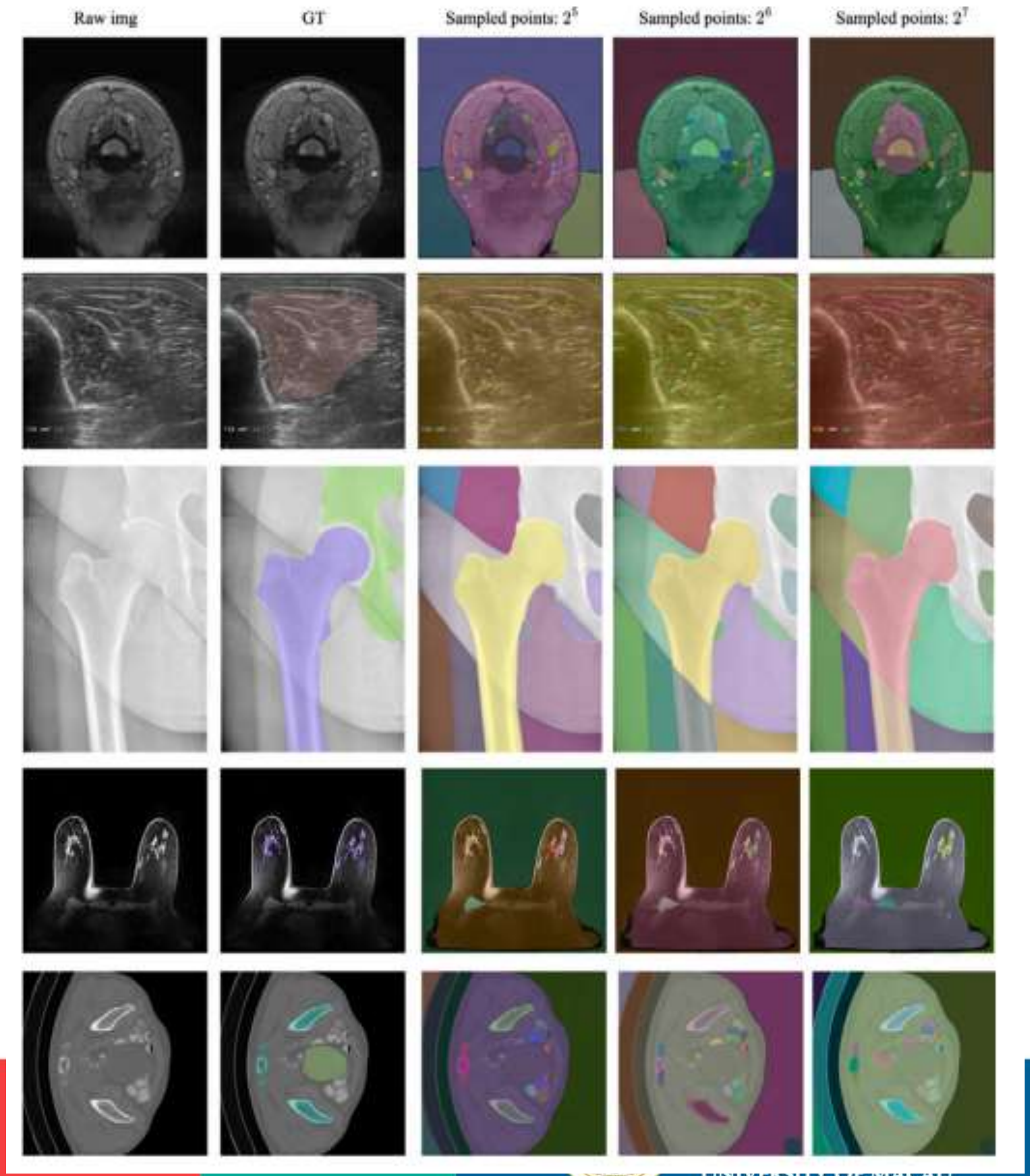
3.6. Performance of SAM for objects of different sizes



- Situations where SAM performance improves for **larger objects**

3.7. Segment-everything mode for medical images

- A example of using SAM on medical images
- Divide the image into many different areas
- Dependent on the number of prompts
- It is not perfect.



4. Conclusions and discussion

- SAM's **accuracy for zero-shot** medical image segmentation is **moderate on average** and varies significantly across different datasets and different images within a dataset.
- The model performs best with **box prompts**, particularly when one box is provided for **each separate part of the object** of interest.
- SAM outperforms RITM, SimpleClick, and FocalClick in the vast majority of the evaluated settings **where a single non-iterative prompt point is provided.**

- In the setting where **multiple iteratively-refined point prompts** are provided, SAM obtains very limited benefit from additional point prompts, except for objects with multiple parts.
- We find a small but non-statistically significant correlation between the **average object size** in a dataset and **SAM performance.**

Thank You!