MedAGI

MICCAI 2023 Vancouver CANADA

MICCAI 2023 1st International Workshop on Foundation Models for General Medical AI

SAM Meets Robotic Surgery: An Empirical Study on Generalization, Robustness and Adaptation

An Wang[1*], Mobarakol Islam[2*], Mengya Xu[3], Yang Zhang[1,4], Hongliang Ren[1,3**]

[1] SHIAE, Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China

[2] WEISS, Department of Medical Physics and Biomedical Engineering, University College London, London, UK

[3] Department of Biomedical Engineering, National University of Singapore, Singapore

[4] School of Mechanical Engineering, Hubei University of Technology, Wuhan, China

The Segment Anything Model (SAM) serves as a fundamental model for semantic segmentation and demonstrates remarkable generalization capabilities across a wide range of downstream scenarios. In this empirical study, we examine SAM's robustness and zero-shot generalizability in the field of robotic surgery, as well as propose SurgicalSAM with Low-rank Adaptation (LoRA) as the finetuning strategy.

Experimental Settings

Datasets: two classical datasets in endoscopic surgical instrument segmentation, i.e., EndoVis17 and EndoVis18

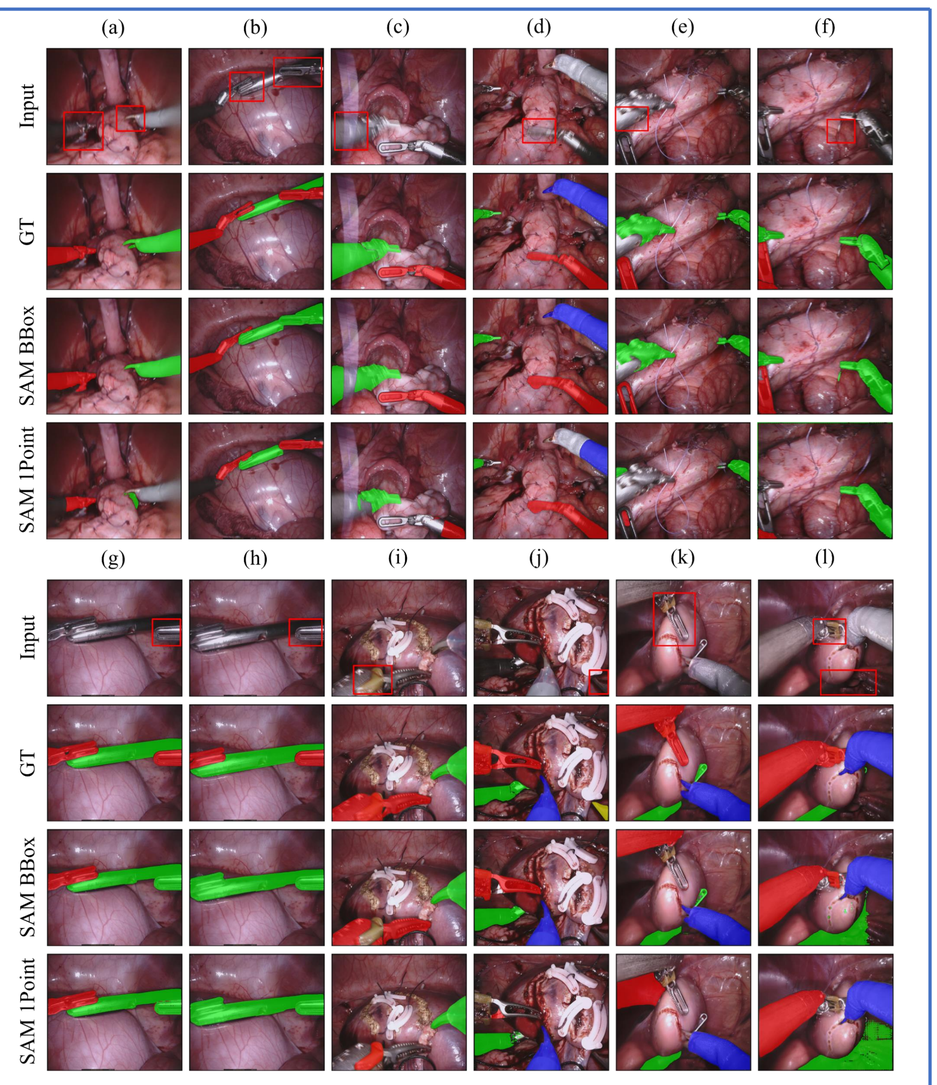
Prompts: Acquire box and point prompts based on the mask labels, manually and automatically.

Metrics: The IoU and Dice metrics from the EndoVis17 challenge

Promptable Segmentation of Surgical Instruments

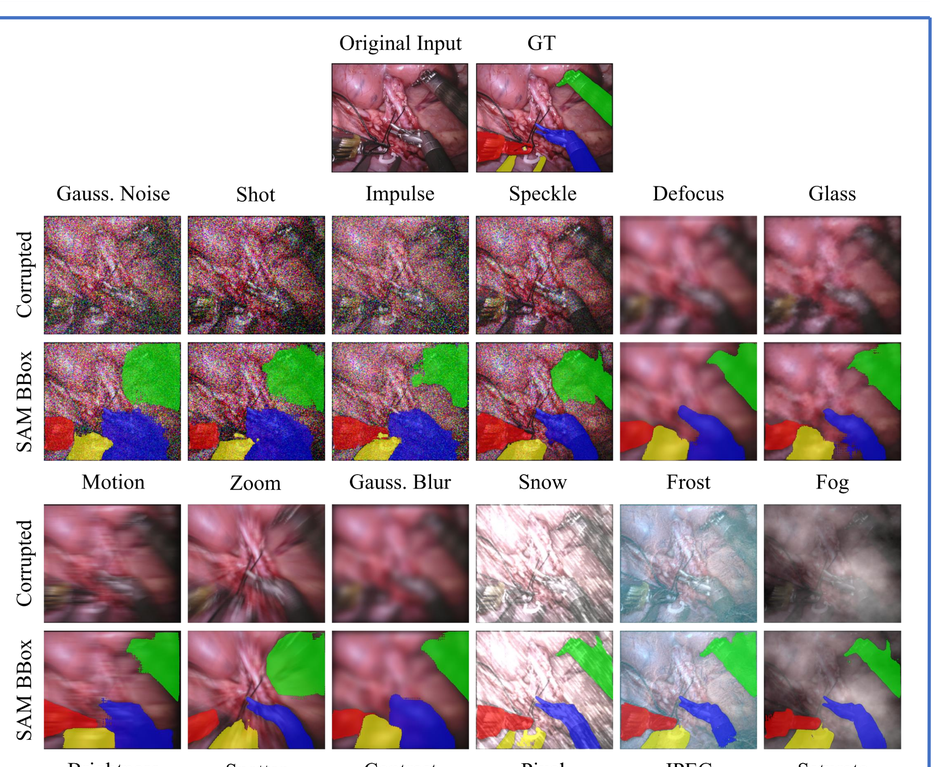
Type	Method	Pub/Year(20-)	Arch.	EndoVis17		EndoVis18	
				Binary IoU	Instrument IoU	Binary IoU	Instrument IoU
Single-Task	Vanilla UNet	MICCAI15	UNet	75.44	15.80	68.89	-
	TernausNet	ICMLA18	UNet	83.60	35.27	-	46.22
	MF-TAPNet	MICCAI19	UNet	87.56	37.35	-	67.87
	Islam et al.	RA-L19	-	84.50	-	-	-
	ISINet	MICCAI21	Res50	-	55.62	-	73.03
	Wang et al.	MICCAI22	UNet	-	-	58.12	-
Multi-Task	ST-MTL	MedIA21	-	83.49	-	-	-
	AP-MTL	ICRA20	-	88.75	-	-	-
	S-MTL	RA-L22	-	-	-	-	43.54
	TraSeTR	ICRA22	Res50 + Trfm	-	60.40	-	76.20
	S3Net	WACV23	Res50	-	72.54	-	75.81
Prompt-based	SAM 1 Point	arxiv23	ViT_h	53.88	55.96*	57.12	54.30*
	SAM Box	arxiv23	ViT_h	**89.19**	**88.20***	**89.35**	**81.09***

* Categorical information directly inherits from associated prompts.

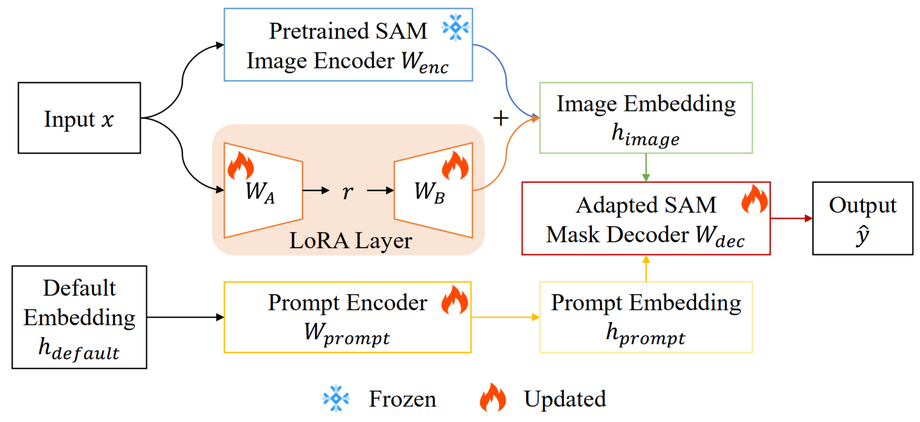


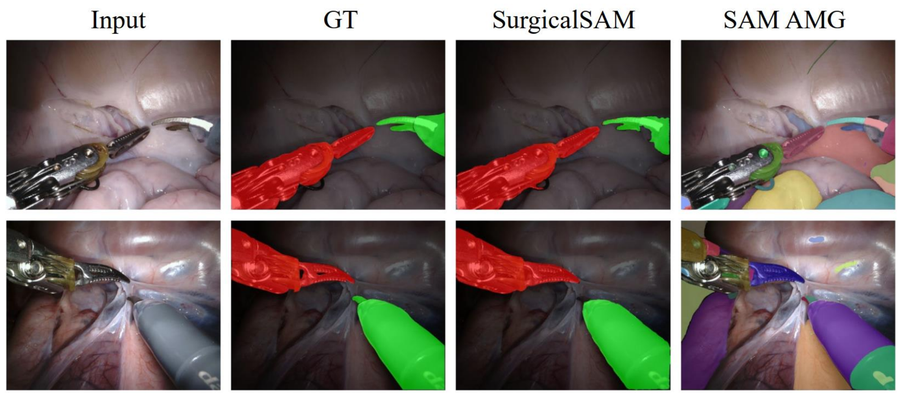
Robustness of SAM under Data Corruption

Task	Severity	Noise				Blur					Weather					Digital			
		Gaussian	Shot	Impulse	Speckle	Defocus	Glass	Motion	Zoom	Gaussian	Snow	Frost	Fog	Bright	Spatter	Contrast	Pixel	JPEG	Saturate
Binary	0	89.35																	
	1	77.69	80.18	80.43	83.28	82.01	80.53	82.99	80.30	85.40	84.08	83.12	85.38	87.43	86.69	85.76	81.12	58.77	86.64
	2	73.92	76.07	76.15	81.65	80.21	79.20	80.22	77.55	81.69	80.69	80.34	84.65	87.27	84.21	84.90	79.32	56.04	84.85
	3	69.21	71.74	73.02	77.74	76.96	72.64	75.50	75.27	78.31	79.58	78.90	83.62	87.23	82.50	83.36	73.81	56.25	86.84
	4	63.80	65.41	67.29	75.28	73.79	72.38	69.60	73.22	75.23	76.33	78.38	82.28	87.06	83.12	77.12	70.82	57.59	83.21
	5	57.07	60.61	61.61	71.83	69.85	69.59	66.25	71.58	66.96	77.66	76.82	78.84	86.43	79.62	66.58	68.55	56.77	81.26
Instrument	0	81.09																	
	1	69.51	71.83	72.25	74.82	73.64	72.13	74.33	71.41	76.79	75.40	74.42	76.82	79.16	78.24	77.17	72.94	54.86	78.27
	2	66.06	68.09	68.53	73.19	71.74	71.02	71.46	68.85	73.15	72.13	71.65	76.14	79.00	75.54	76.22	71.55	52.23	76.61
	3	62.01	64.44	65.89	69.75	68.74	64.97	67.13	67.12	70.08	70.97	70.21	75.01	78.90	73.70	74.67	66.83	51.63	78.39
	4	57.28	59.12	61.03	67.82	65.87	64.87	62.15	65.18	67.23	68.43	69.79	73.73	78.73	74.24	69.48	63.99	51.88	74.91
	5	51.56	55.16	55.86	64.76	62.43	62.23	59.26	63.96	60.60	69.33	68.32	70.45	78.19	70.72	61.14	61.79	51.01	73.35

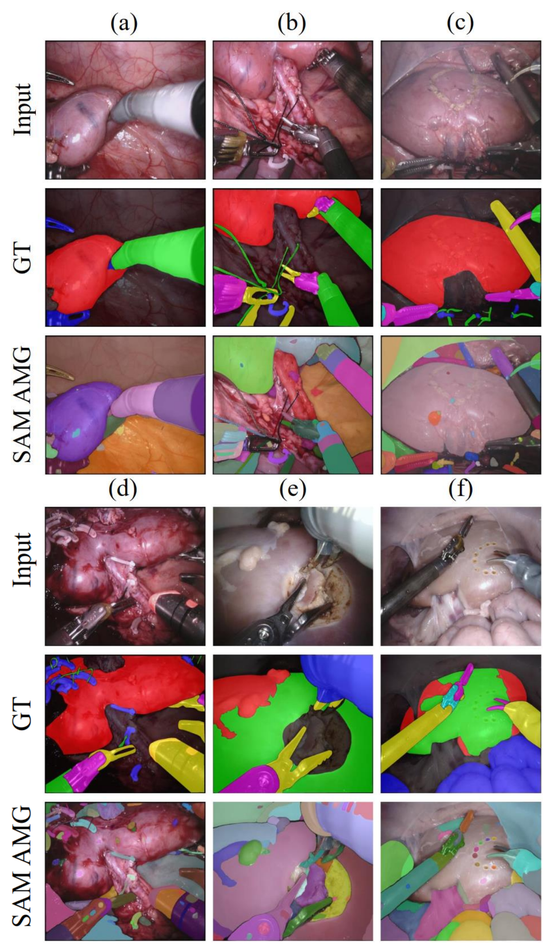


SurgicalSAM: Parameter-efficient Finetuning with Low-rank Adaptation





Automatic Mask Generation (AMG)



Summary

- We extensively explore SAM under different scenarios, including prompted and unprompted cases, bounding box and points-based prompt approaches, as well as the ability to generalize under corruptions and perturbations at five severity levels.
- The results indicate that SAM is deficient in segmenting entire surgical instruments with point-based prompts or unprompted settings; it also fails to precisely identify and segment when faced with complex scenarios, such as tool overlapping and the existence of blood, reflection, blur, and shade.
- Finetuning SAM in a parameter-efficient manner helps improve its performance in surgical scenes, yet there is still significant work left in the context of surgical instrument segmentation.

Multiple openings for Ph.D./Postdoc/RA to perform research on Medical Robotics Perception & AI at CUHK, Hong Kong. Check details at http://www.labren.org/mm/.