RP-SAM2: Refining Point Prompts for Stable Surgical Instrument Segmentation

BMVC 2025

Nuren Zhaksylyk, Ibrahim Almakky, Jay Paranjape, S. Swaroop Vedula, Shameema Sikder, Vishal M. Patel, and Mohammad Yaqub

Mohamed Bin Zayed University of Artificial Intelligence,
Abu Dhabi, United Arab Emirates

nuren.zhaksylyk@mbzuai.ac.ae

ARXIV GITHUB

MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

JOHNS HOPKINS UNIVERSITY

BioMedIA

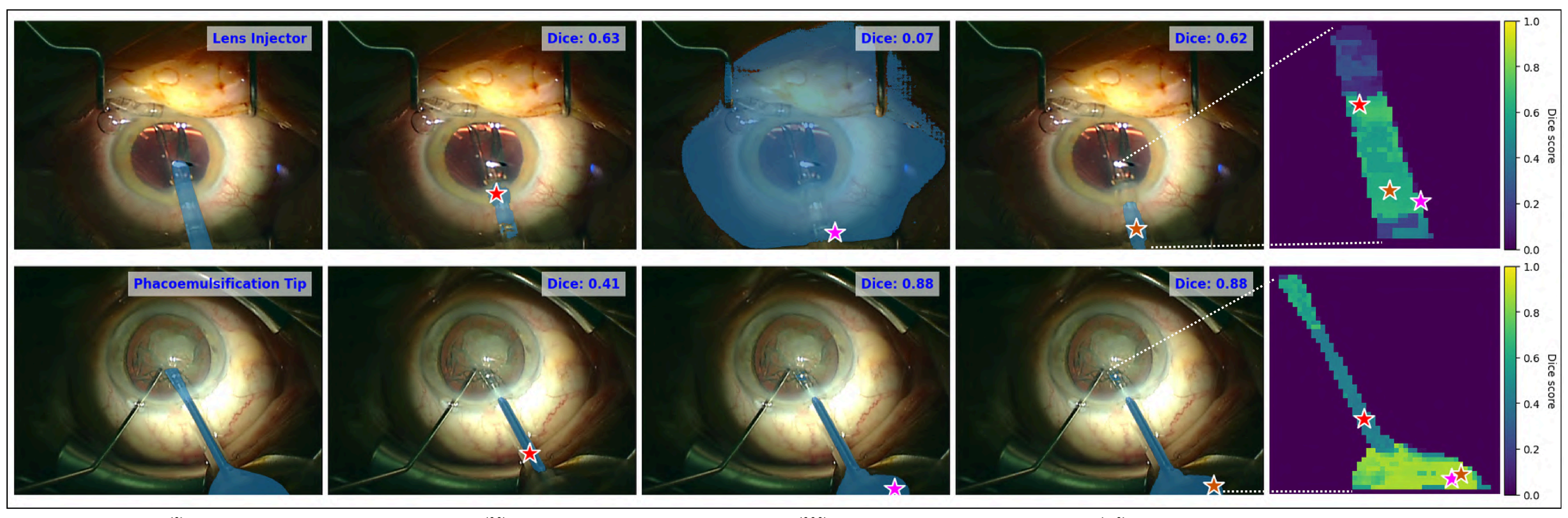
Introduction

Problem Statement: Accurate surgical instrument segmentation is critical in cataract surgery for skill assessment and workflow optimization.
Data Scarcity: Limited annotated data hinders the development of fully automatic models.
Prompt-based Model Challenges: While flexible, prompt-based methods like SAM2 are highly sensitive to point prompt placement, leading to inconsistent segmentations.
Proposed Solution: We introduce RP-SAM2, incorporating a novel shift block and a compound loss function to stabilize point prompts.

Research Gap

- A grid-based sampling of point prompts (ii, iii, vi) across surgical instruments region in cataract surgery images shows that SAM2's segmentation performance heavily depends on the point location resulting in significant fluctuations in prediction masks (v) as shown in Fig. 1.

Fig. 1. Zero-shot performance of SAM2 on surgical instrument segmentation using single-point prompts.



(i) (ii) (iii) (vi) (v)

Contributions

- We introduce a novel shift block into SAM2, creating the **R**obust **P**oint prompt SAM2 (**RP-SAM2**) model.
- We propose a **compound loss** for RP-SAM2 that enhances segmentation performance regardless of the user's point prompt location.
- Our approach eases the annotator's burden while maintaining the robust segmentation capabilities of foundation models such as SAM2.

Methodology

Fig. 2. Proposed point shifting approach

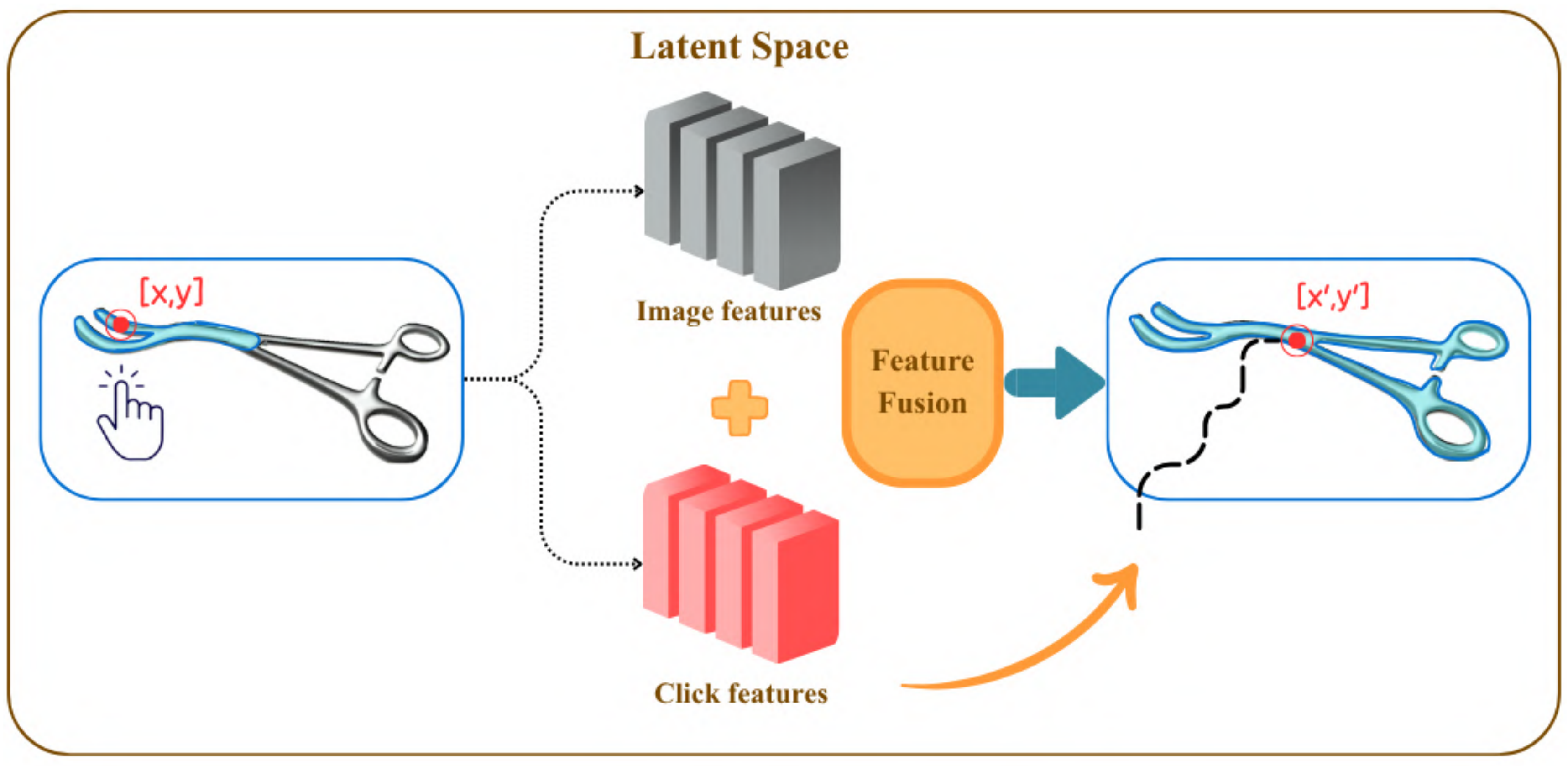
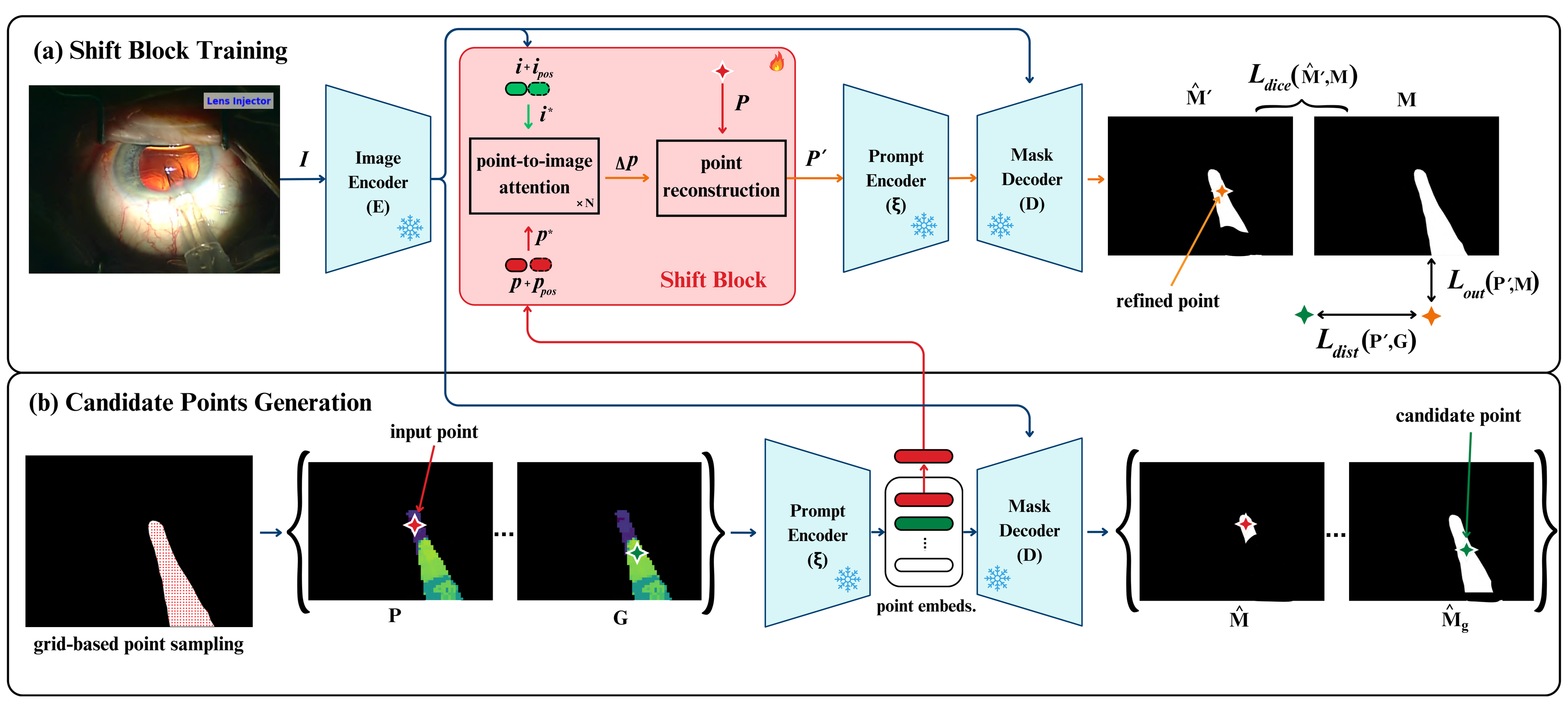


Fig. 3. Detailed RP-SAM2 architecture



Results

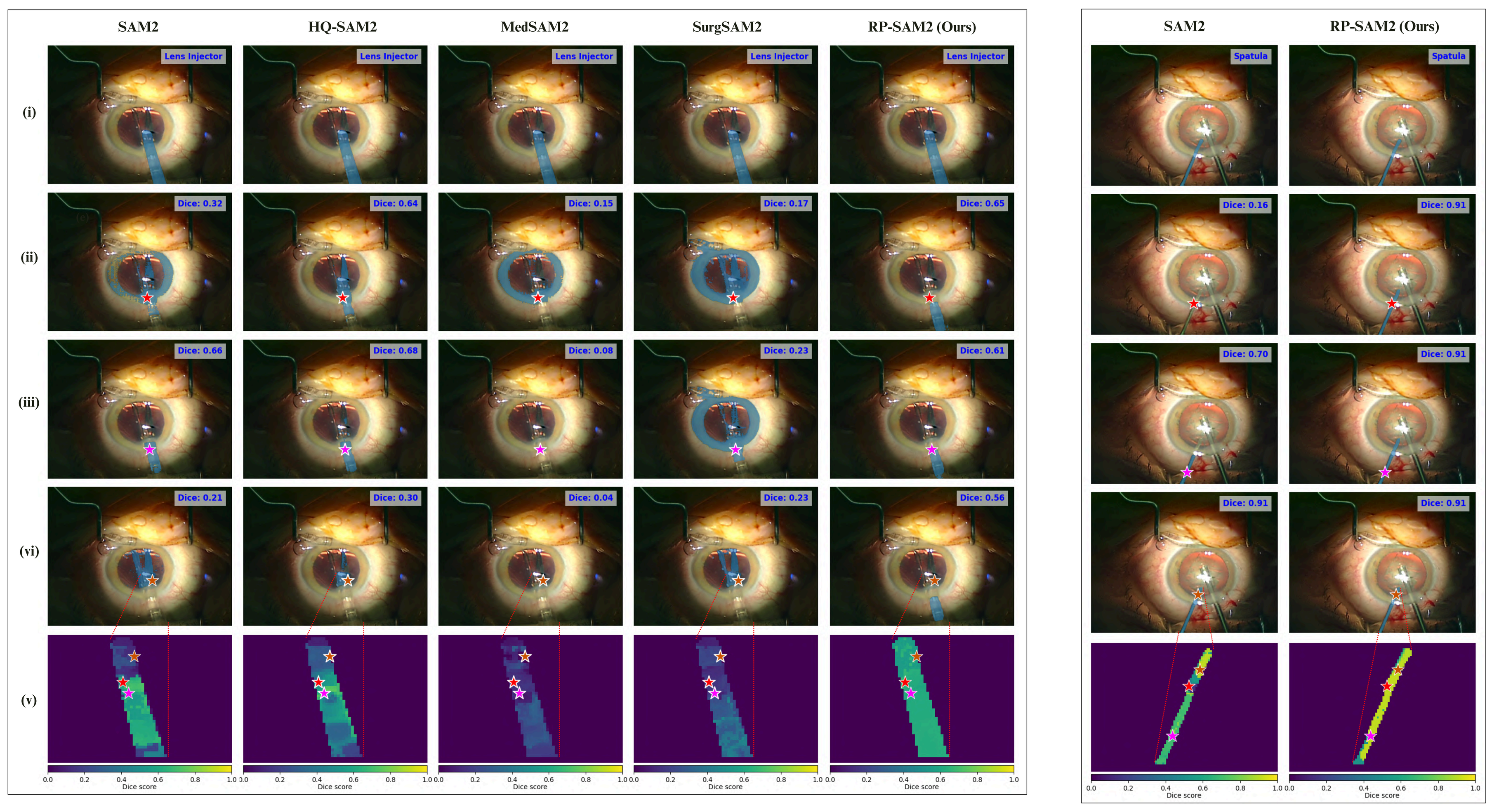
Table 1. Average performance and stability of SOTA methods on the Cataract1k test set, evaluated over 10 single-point prompts.

Classes	SAM2 [3] | 236.5M		HQ-SAM2 [1] | 5.1M		MedSAM2 [4] | 38.9M		SurgSAM2 [2] | 11.5M		RP-SAM2 (Ours) | 12.1M	
	mDSC(%) ↑	mHD95 ↓	mDSC(%) ↑	mHD95 ↓	mDSC(%) ↑	mHD95 ↓	mDSC(%) ↑	mHD95 ↓	mDSC(%) ↑	mHD95 ↓
CC	80.17±3.05	31.56±9.94	66.37±4.13	64.73±18.58	73.18±2.67	49.76±12.59	78.29±3.25	38.13±16.47	81.16±3.10	22.98±6.71
CF	73.98±1.96	47.28±15.41	51.97±3.06	106.27±27.73	61.65±3.60	52.39±13.90	76.76±2.00	43.90±15.21	73.40±3.63	51.14±16.07
G	79.98±1.79	33.74±6.92	69.91±1.72	58.27±8.03	72.89±2.27	43.9±7.86	78.75±1.49	33.57±4.75	78.29±1.18	41.45±5.97
IA	78.00±1.23	66.93±5.99	55.43±1.85	143.50±11.06	62.81±1.45	88.90±6.01	69.15±1.81	119.61±10.19	81.42±0.88	55.39±4.31
IK	89.14±9.71	80.78±27.62	75.20±7.63	59.01±17.60	82.46±9.44	39.52±18.26	87.91±8.43	31.11±24.70	91.52±4.02	72.35±19.37
KF	71.64±5.56	60.32±27.63	70.46±4.13	35.05±12.44	74.31±3.18	26.93±26.06	70.16±4.39	145.90±32.54	73.12±2.45	59.17±12.46
LI	67.52±4.34	137.26±16.96	52.60±6.18	162.47±12.72	48.65±7.83	160.68±28.09	63.03±4.67	140.01±20.66	75.41±3.41	109.19±11.11
PhT	72.94±2.25	106.85±8.01	54.84±1.70	167.62±6.35	56.78±2.39	135.68±8.61	62.45±1.53	179.61±7.79	75.33±0.75	89.28±5.10
SK	89.75±3.03	67.88±75.68	83.86±15.38	66.51±65.02	86.88±4.92	27.80±47.58	83.03±1.55	193.53±35.52	92.85±0.08	5.38±0.05
S	74.33±1.04	102.8±9.99	73.45±1.01	64.53±6.22	66.62±1.28	83.33±4.04	76.57±1.08	69.53±5.58	75.03±1.38	71.96±5.48
Avg.	77.74±3.40	73.54±20.42	65.41±4.68	92.80±18.57	68.62±3.90	70.89±17.30	74.61±3.02	99.49±17.34	79.75±2.09	57.83±8.66
W. Avg.	75.73±1.8	82.24±9.28	63.52±1.92	106.94±9.23	64.36±2.15	89.66±7.94	71.89±1.7	100.58±8.54	77.17±1.33	66.49±5.98

Table 2. Average performance and stability of RP-SAM2 in multiple points setting without shifting point prompts and shifting the first point on the Cataract1k test set.

Classes	2 points				3 points				5 points			
	no shift		shift first		no shift		shift first		no shift		shift first	
	mDSC(%) ↑	mHD95 ↓	mDSC(%) ↑	mHD95 ↓	mDSC(%) ↑	mHD95 ↓	mDSC(%) ↑	mHD95 ↓	mDSC(%) ↑	mHD95 ↓	mDSC(%) ↑	mHD95 ↓
CC	83.17±2.05	23.09±8.18	84.49±1.01	17.20±3.12	83.99±1.18	20.86±5.83	84.43±1.31	16.88±2.84	84.20±1.77	28.74±15.04	84.47±1.86	21.89±9.64
CF	78.78±3.81	34.52±13.90	77.90±2.32	37.20±11.33	82.08±1.27	28.54±7.10	80.70±3.46	34.34±17.62	82.02±1.44	33.83±6.88	81.69±2.02	33.03±7.52
G	82.78±0.90	27.96±6.45	81.60±1.07	30.96±6.17	83.47±0.70	27.77±4.37	83.00±0.92	28.85±3.54	83.12±0.66	38.24±6.38	82.92±0.88	37.89±5.59
IA	80.78±0.69	62.11±6.73	81.10±0.58	63.75±3.84	80.95±0.73	76.93±5.66	81.12±0.94	75.20±6.69	82.22±0.61	71.91±4.58	82.15±0.23	69.51±3.32
IK	92.20±2.65	84.48±31.31	92.22±2.04	85.04±30.58	90.88±4.07	111.16±32.46	90.95±1.90	116.54±30.92	85.50±4.85	126.38±29.60	87.34±2.37	119.81±32.30
KF	75.74±4.78	39.24±29.58	71.81±5.87	66.11±27.59	67.37±4.36	110.66±32.20	66.68±3.71	115.48±38.16	64.51±6.55	129.76±28.28	64.62±6.68	128.17±27.80
LI	73.36±4.41	113.94±17.86	77.63±2.57	101.18±10.52	77.96±2.71	105.5±11.12	79.35±3.12	97.12±10.45	80.31±3.45	92.11±13.89	81.66±3.40	86.15±14.67
PhT	76.93±0.67	105.78±4.83	77.27±1.19	103.84±12.65	77.79±1.29	125.44±6.24	77.51±0.90	125.62±6.00	79.91±0.65	122.49±6.02	79.62±0.50	120.74±7.73
SK	87.57±5.62	152.27±84.97	89.27±2.86	120.35±62.03	87.97±2.71	172.75±73.67	91.21±3.10	91.66±86.85	85.60±1.76	192.23±59.69	85.78±1.29	198.69±53.82
S	77.87±0.71	92.87±9.78	78.92±1.14	72.06±6.62	75.78±1.02	110.79±7.39	77.07±1.23	91.21±8.54	73.54±1.29	146.07±7.57	75.06±1.48	128.97±7.99
Avg.	80.92±2.63	73.63±21.36	81.22±2.06	69.77±17.44	80.82±2.00	89.04±18.60	81.20±2.06	79.29±21.16	80.09±2.30	98.18±17.79	80.53±2.07	94.49±17.04
W. Avg.	79.18±1.06	76.03±8.52	79.57±1.17	69.71±8.26	79.10±1.11	89.25±7.09	79.39±1.22	82.83±7.78	79.11±1.08	100.24±7.49	79.50±1.06	93.71±7.51

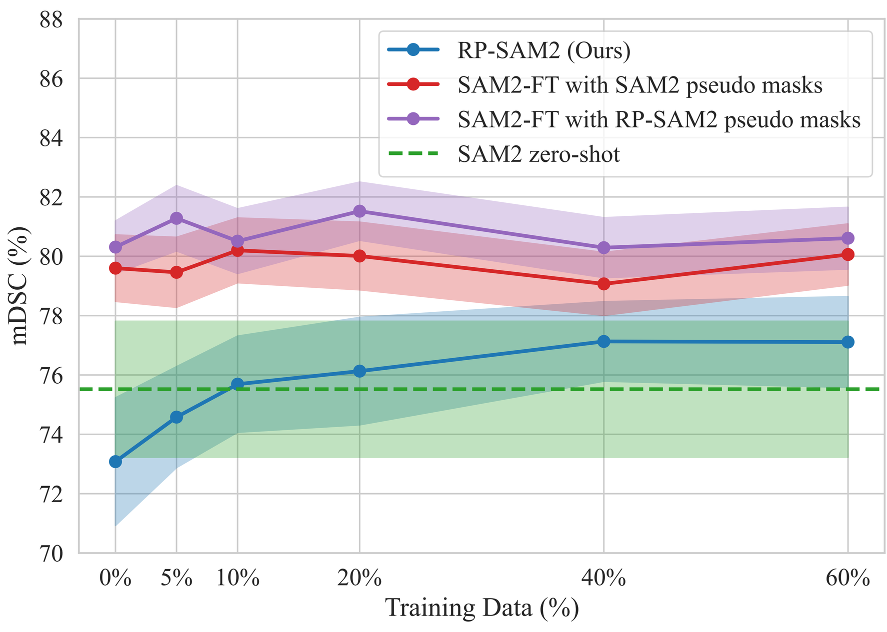
Fig. 3. An illustration of (a) Qualitative comparison with SOTA and (b) An example of strong light-reflection artefacts on the instrument.



(a) (b)

Ablation Study

Table 3. An illustration of (a) OOD performance on the CaDIS test set, (b) ablation study of loss components and the number of points sampled per object during training RP-SAM2 on Cataract1k dataset.



(a)

Method	mDSC(%) ↑	mHD95 ↓
L_{dice}	79.42±2.78	64.62±21.02
$L_{dice} + L_{dist}$	79.12±3.10	62.41±16.04
$L_{dice} + L_{dist} + L_{out}$	79.75±2.09	57.83±8.66
6 points	79.28±2.05	58.07±10.95
8 points	78.68±3.01	62.27±13.96
10 points	79.75±2.09	57.83±8.66
12 points	78.80±2.66	65.75±9.22
2 layers pt-to-img	79.03±2.02	63.23±10.42
6 layers pt-to-img	79.75±2.09	57.83±8.66

(b)

Conclusion

- RP-SAM2 is a refined version of SAM2 that is robust to user's point prompts.
- It demonstrates improved stability, robustness to noise, and generalizability.
- RP-SAM2 is positioned as a valuable tool for advancing automatic surgical workflows and reducing annotation burdens in medical image analysis.

References

[1] Ke, L., Ye, M., Danelljan, M., Tai, Y.W., Tang, C.K., Yu, F., et al.: Segment anything in high quality. Advances in Neural Information Processing Systems 36 (2024)
[2] Liu, H., Zhang, E., Wu, J., Hong, M., Jin, Y.: Surgical sam 2: Real-time segment anything in surgical video by efficient frame pruning. arXiv preprint arXiv:2408.07931 (2024)
[3] Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
[4] Zhu, J., Qi, Y., Wu, J.: Medical sam 2: Segment medical images as video via segment anything model 2. arXiv preprint arXiv:2408.00874 (2024)