VP-SAM: Taming Segment Anything Model for Video Polyp Segmentation via Disentanglement and Spatio-temporal Side Network

Zhixue Fang[1], Yuzhi Liu[1], Huisi Wu[1*], and Jing Qin[2]

[1] College of Computer Science and Software Engineering, Shenzhen University

[2] Centre for Smart Health, The Hong Kong Polytechnic University

CONTENTS

01 Introduction

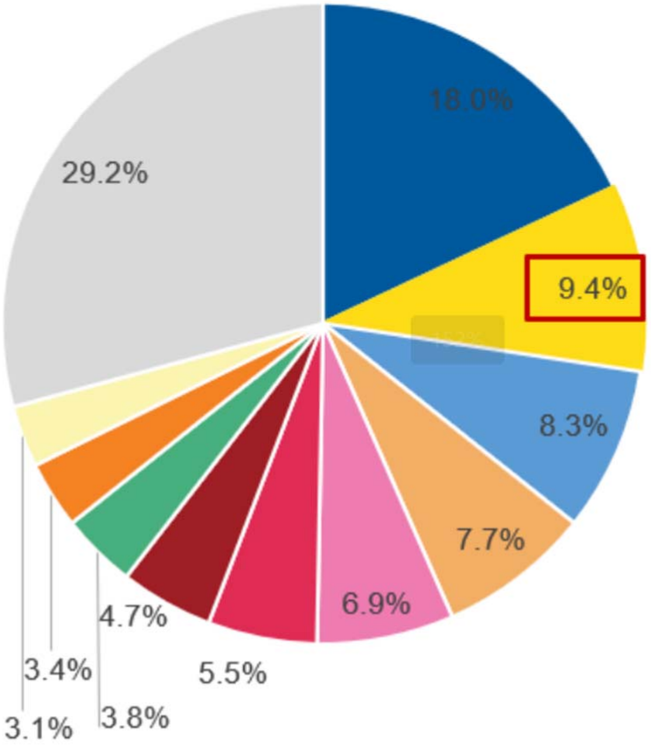
02 Methodology

03 Experiments

04 Conclusion

Background

- Massive cancer-related deaths due to colorectal cancer



🟦 Lung cancer

🟨 Colorectal cancer

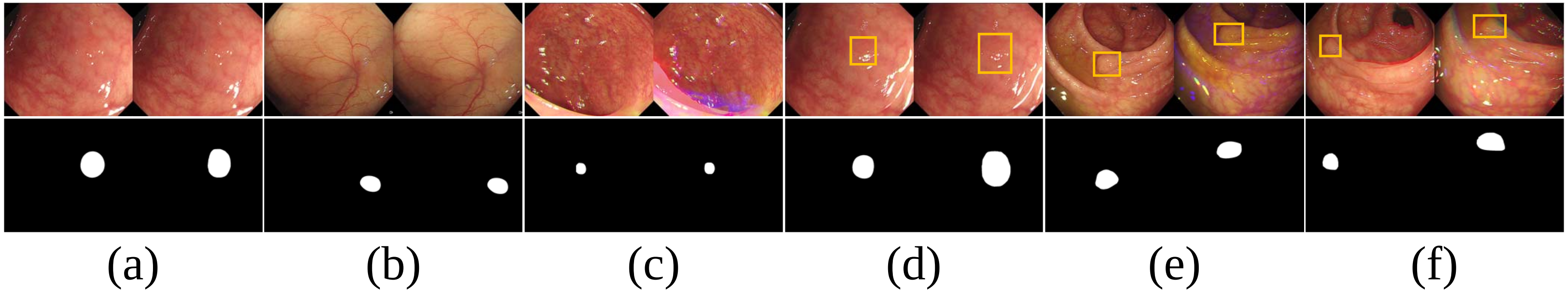
🟦 Liver cancer

🟧 Stomach cancer

🟪 Breast cancer

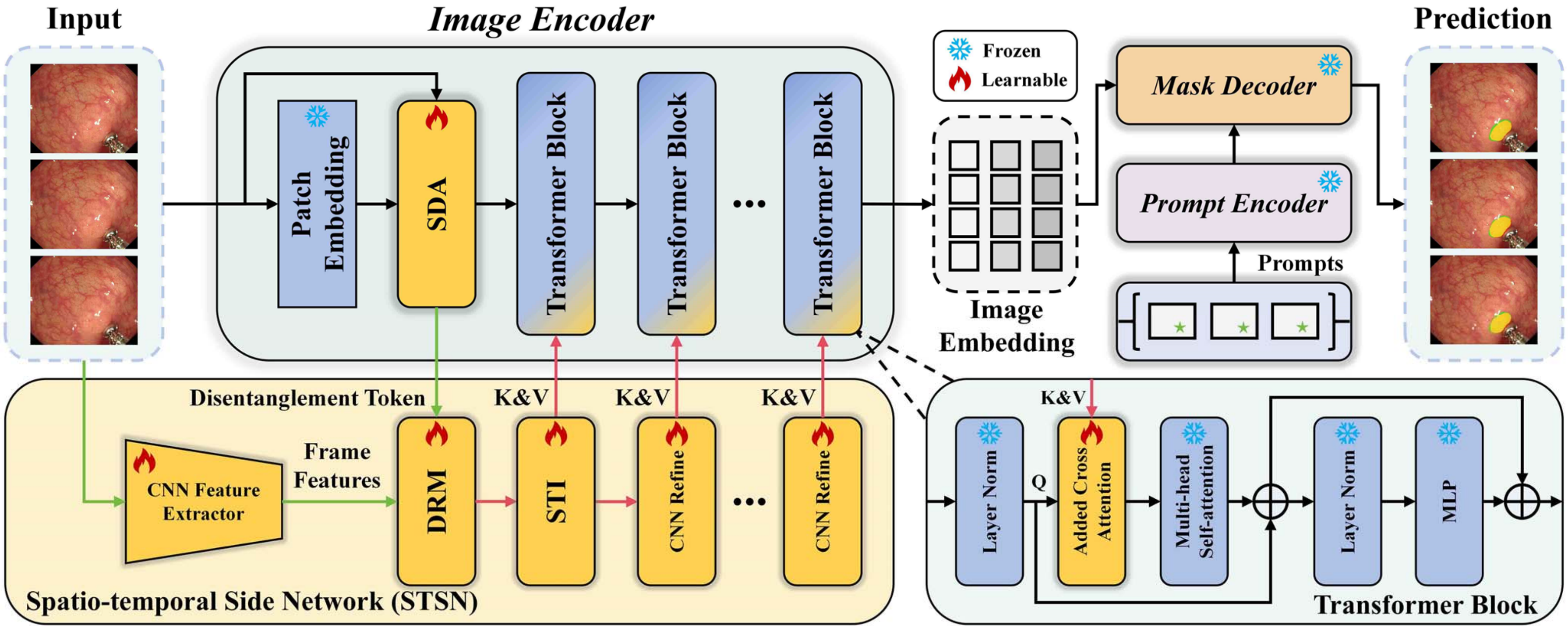
Challenges

- Low-contrast (a-c)
- Dramatic frame-to-frame variations (d-f)



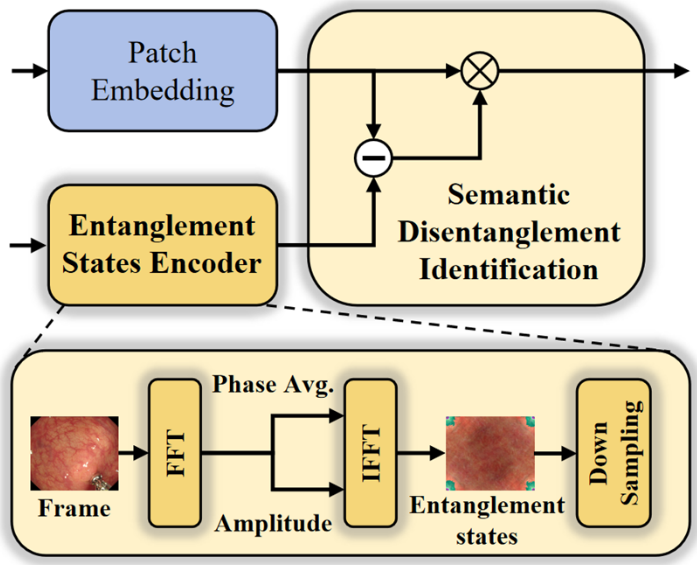
Method

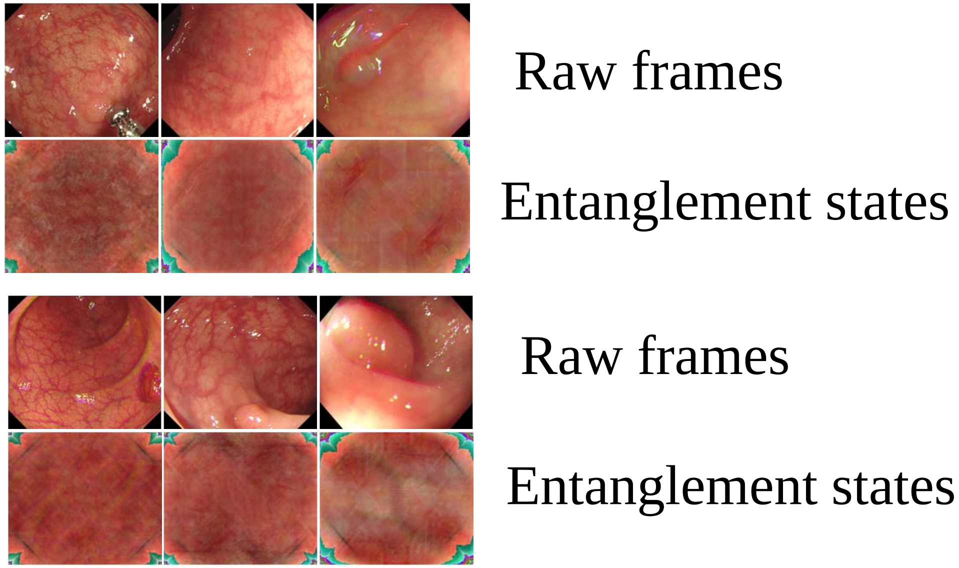
- Semantic disentanglement adapter
- Spatio-temporal side network



Method

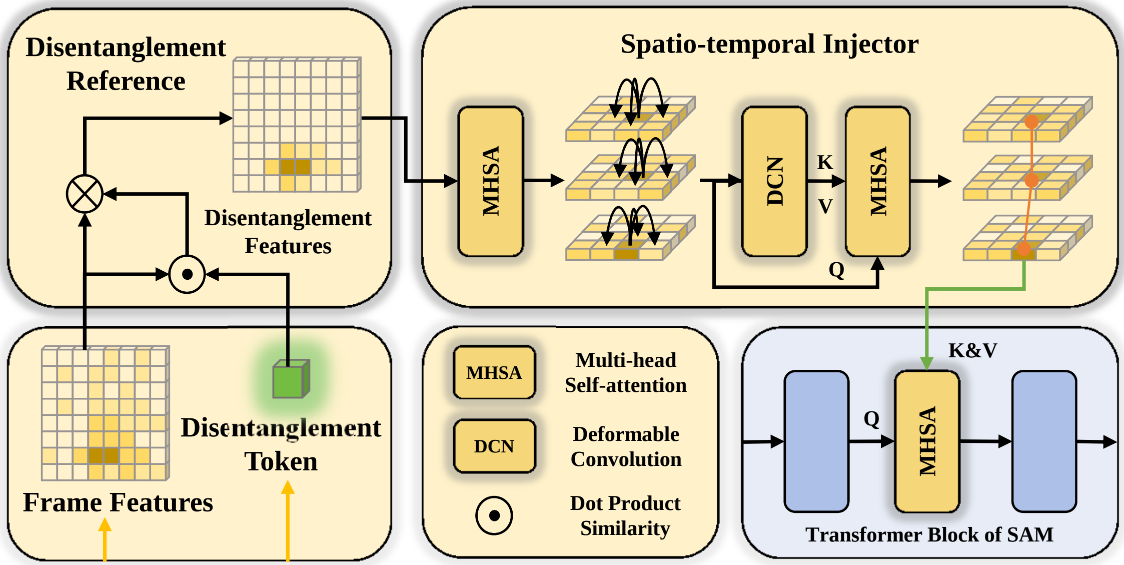
- Resolve low-contrast distractions





Method

- Supplement spatio-temporal information

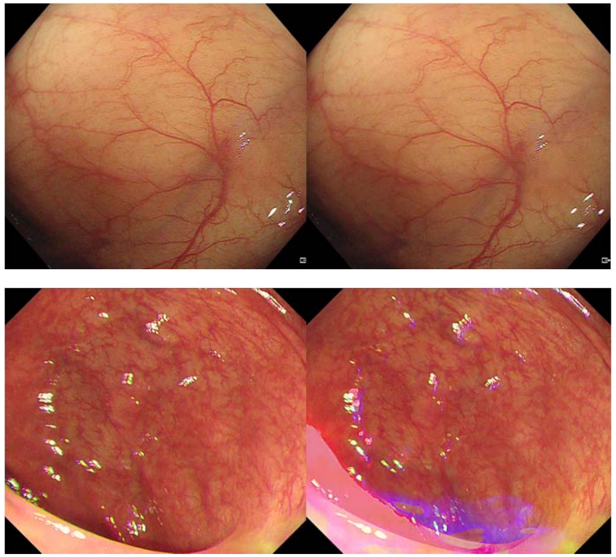


Motion Modeling

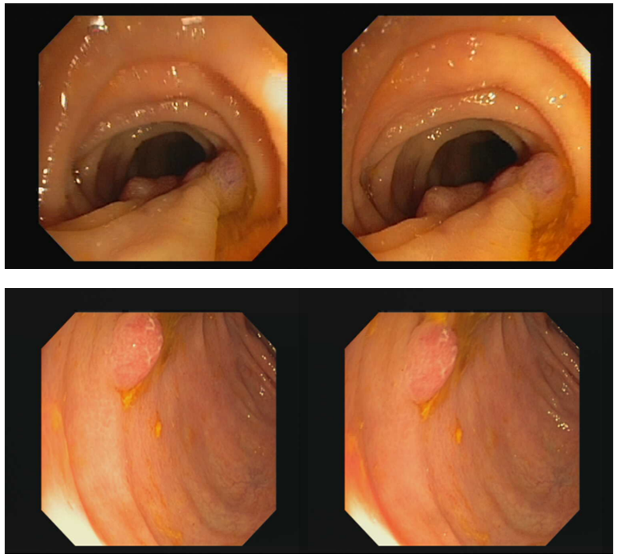
Added Cross Attention

Datasets

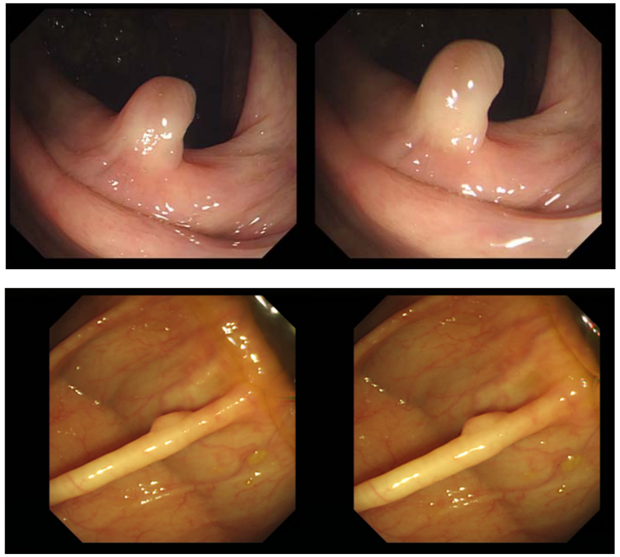
- SUN-SEG、CVC-612、CVC-300



SUN-SEG Dataset[1]



CVC-612 Dataset[2]



CVC-300 Dataset[3]

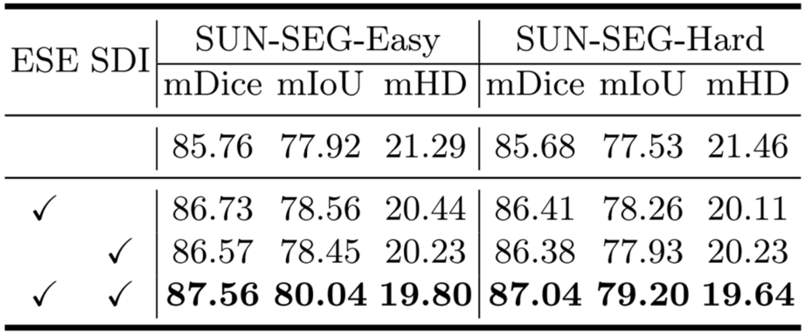
[1]. Ji G P, Xiao G, Chou Y C, et al. Video polyp segmentation: A deep learning perspective[J]. Machine Intelligence Research, 2022, 19(6): 531-549.

[2]. Bernal J, Sánchez F J, Fernández-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians[J]. Computerized medical imaging and graphics, 2015, 43: 99-111.

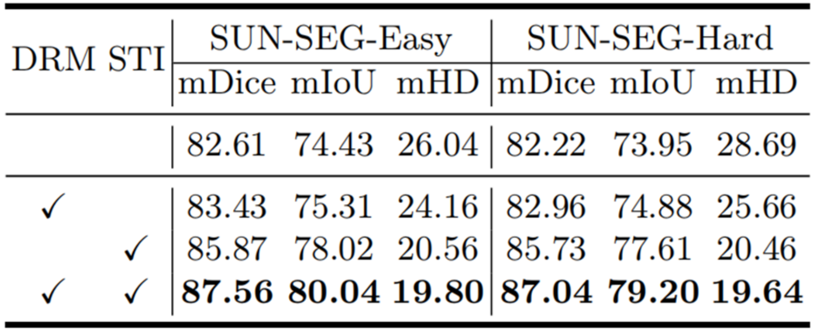
[3]. Bernal J, Sánchez J, Vilarino F. Towards automatic polyp detection with a polyp appearance model[J]. Pattern Recognition, 2012, 45(9): 3166-3182.

Ablation Experiments

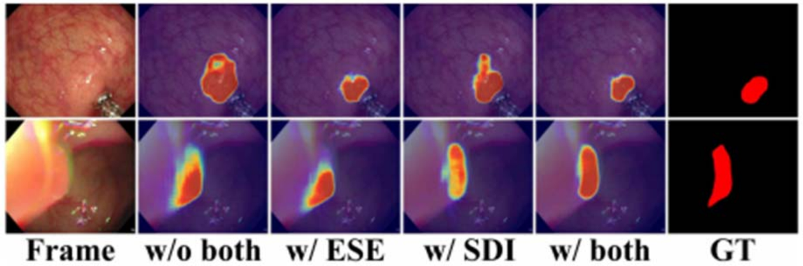
- Effectiveness of our modules

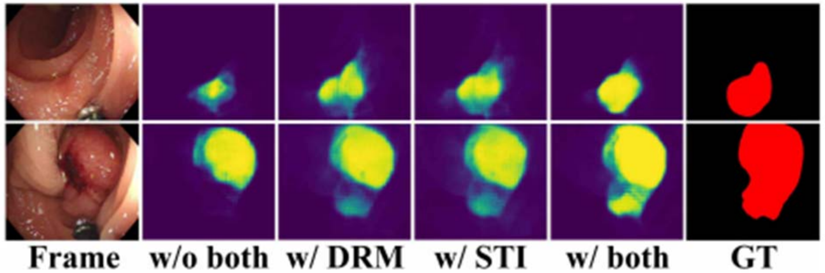


ESE	SDI	SUN-SEG-Easy			SUN-SEG-Hard		
		mDice	mIoU	mHD	mDice	mIoU	mHD
		85.76	77.92	21.29	85.68	77.53	21.46
✓		86.73	78.56	20.44	86.41	78.26	20.11
	✓	86.57	78.45	20.23	86.38	77.93	20.23
✓	✓	**87.56**	**80.04**	**19.80**	**87.04**	**79.20**	**19.64**



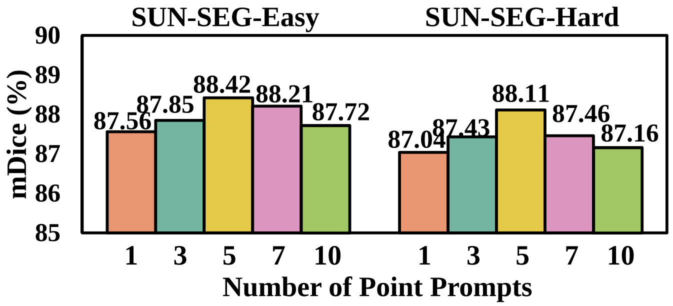
DRM	STI	SUN-SEG-Easy			SUN-SEG-Hard		
		mDice	mIoU	mHD	mDice	mIoU	mHD
		82.61	74.43	26.04	82.22	73.95	28.69
✓		83.43	75.31	24.16	82.96	74.88	25.66
	✓	85.87	78.02	20.56	85.73	77.61	20.46
✓	✓	**87.56**	**80.04**	**19.80**	**87.04**	**79.20**	**19.64**

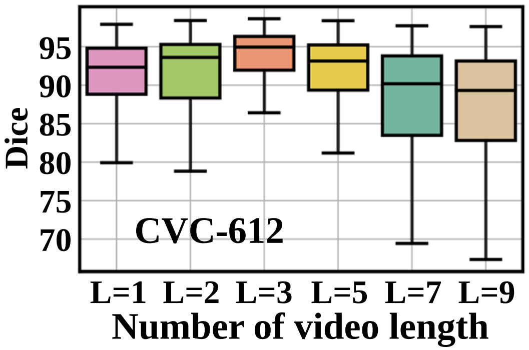


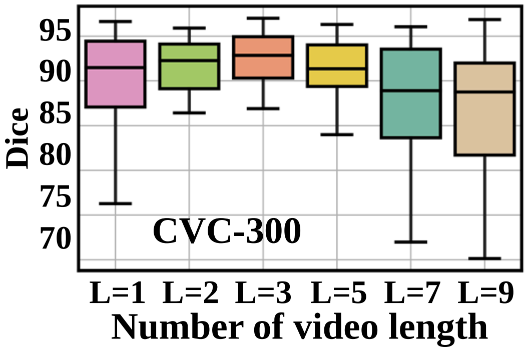


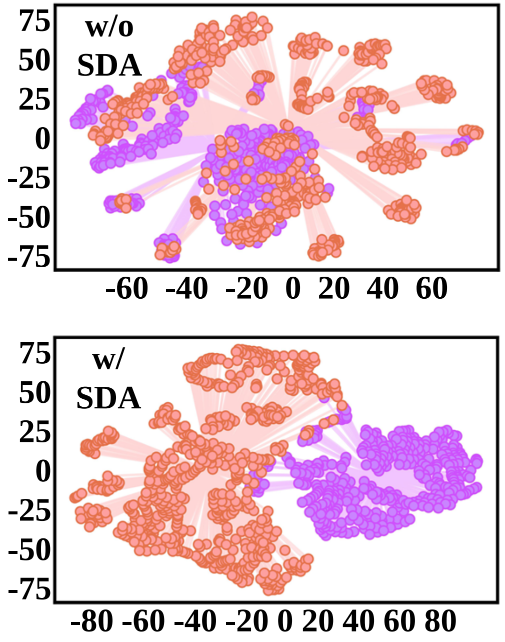
Ablation Experiments

- Effectiveness of our modules



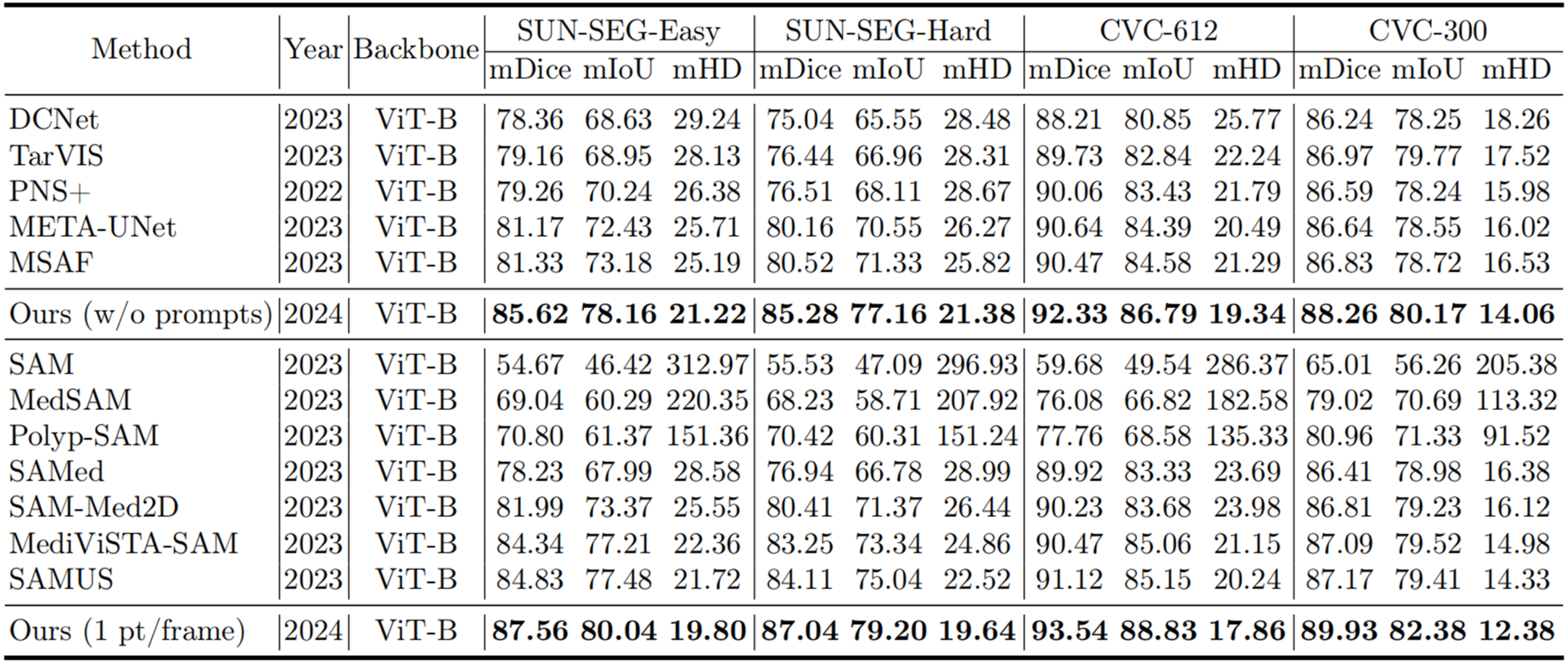






Comparison

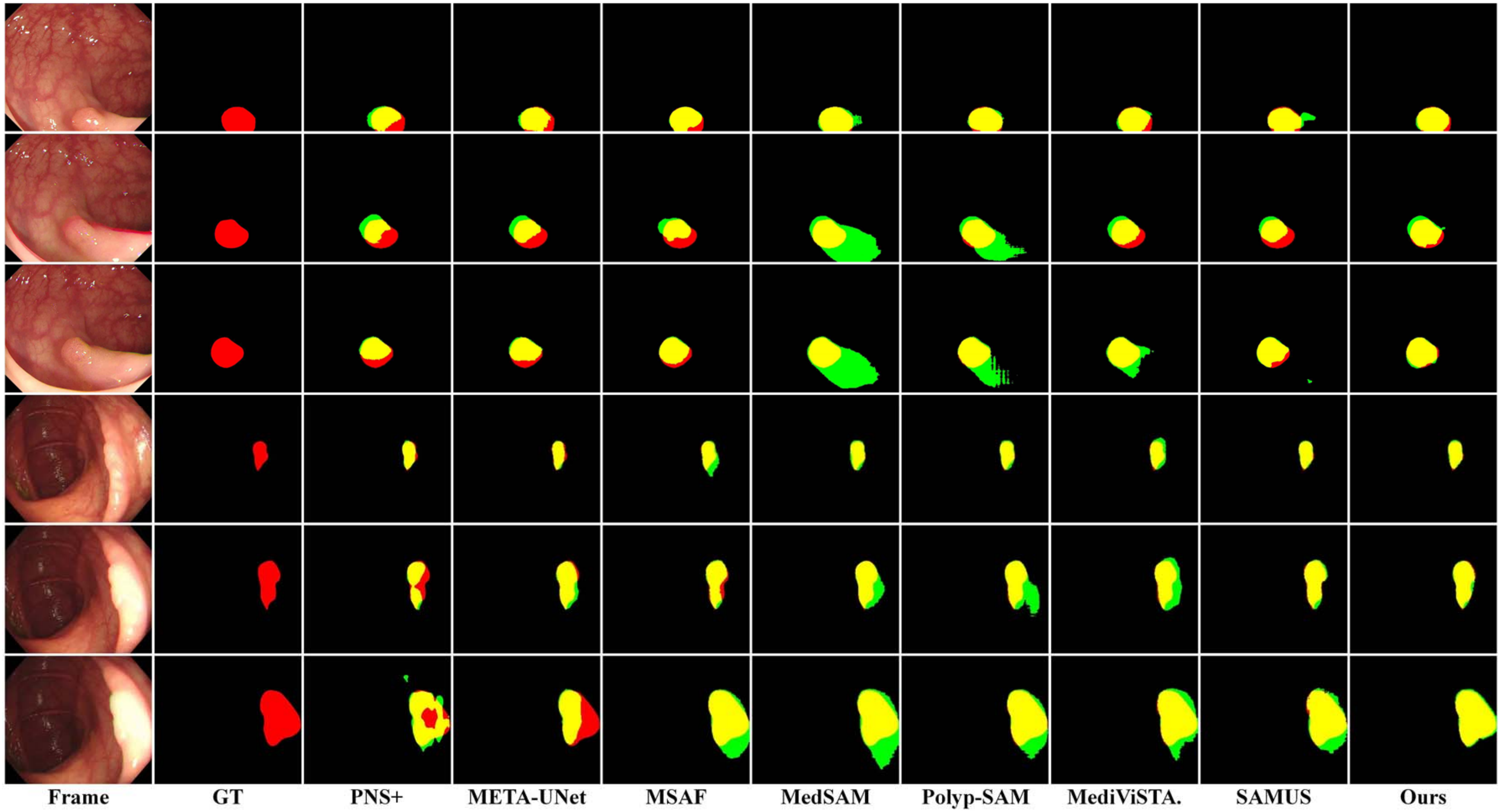
- Best performance



Method	Year	Backbone	SUN-SEG-Easy			SUN-SEG-Hard			CVC-612			CVC-300		
			mDice	mIoU	mHD	mDice	mIoU	mHD	mDice	mIoU	mHD	mDice	mIoU	mHD
DCNet	2023	ViT-B	78.36	68.63	29.24	75.04	65.55	28.48	88.21	80.85	25.77	86.24	78.25	18.26
TarVIS	2023	ViT-B	79.16	68.95	28.13	76.44	66.96	28.31	89.73	82.84	22.24	86.97	79.77	17.52
PNS+	2022	ViT-B	79.26	70.24	26.38	76.51	68.11	28.67	90.06	83.43	21.79	86.59	78.24	15.98
META-UNet	2023	ViT-B	81.17	72.43	25.71	80.16	70.55	26.27	90.64	84.39	20.49	86.64	78.55	16.02
MSAF	2023	ViT-B	81.33	73.18	25.19	80.52	71.33	25.82	90.47	84.58	21.29	86.83	78.72	16.53
Ours (w/o prompts)	2024	ViT-B	**85.62**	**78.16**	**21.22**	**85.28**	**77.16**	**21.38**	**92.33**	**86.79**	**19.34**	**88.26**	**80.17**	**14.06**
SAM	2023	ViT-B	54.67	46.42	312.97	55.53	47.09	296.93	59.68	49.54	286.37	65.01	56.26	205.38
MedSAM	2023	ViT-B	69.04	60.29	220.35	68.23	58.71	207.92	76.08	66.82	182.58	79.02	70.69	113.32
Polyp-SAM	2023	ViT-B	70.80	61.37	151.36	70.42	60.31	151.24	77.76	68.58	135.33	80.96	71.33	91.52
SAMed	2023	ViT-B	78.23	67.99	28.58	76.94	66.78	28.99	89.92	83.33	23.69	86.41	78.98	16.38
SAM-Med2D	2023	ViT-B	81.99	73.37	25.55	80.41	71.37	26.44	90.23	83.68	23.98	86.81	79.23	16.12
MediViSTA-SAM	2023	ViT-B	84.34	77.21	22.36	83.25	73.34	24.86	90.47	85.06	21.15	87.09	79.52	14.98
SAMUS	2023	ViT-B	84.83	77.48	21.72	84.11	75.04	22.52	91.12	85.15	20.24	87.17	79.41	14.33
Ours (1 pt/frame)	2024	ViT-B	**87.56**	**80.04**	**19.80**	**87.04**	**79.20**	**19.64**	**93.54**	**88.83**	**17.86**	**89.93**	**82.38**	**12.38**

Comparison

- Visualization results



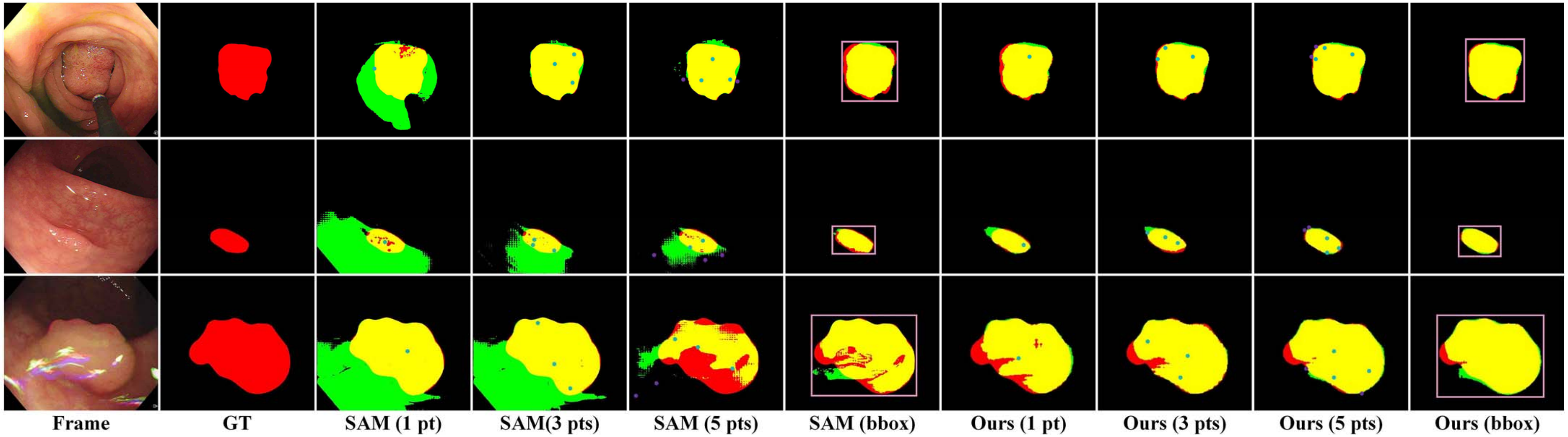
False Negative

True Positive

False Positive

Comparison

- Visualization results



True Positive   False Negative   False Positive

Conclusion

- A novel method adapted from SAM
- SDA and STSN modules
- State-of-the-art performance

Future Work

- More efficient spatio-temporal information

THANK YOU