CAT: Coordinating Anatomical-Textual Prompts for Multi-Organ and Tumor Segmentation

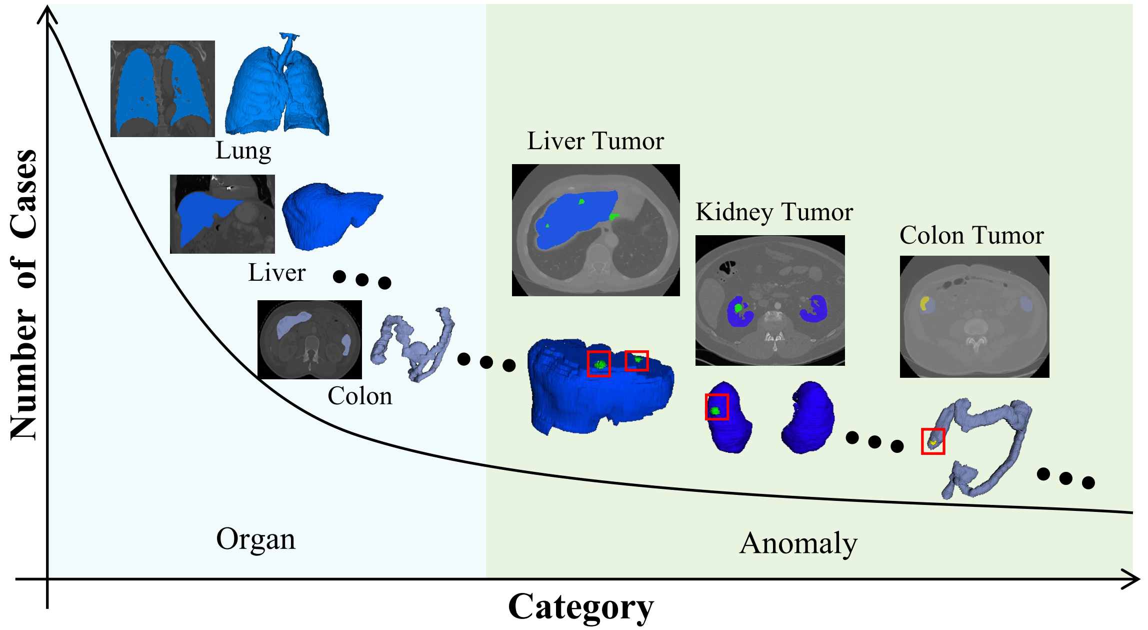
Zhongzhen Huang[1,2], Yankai Jiang[2], Rongzhao Zhang[2]

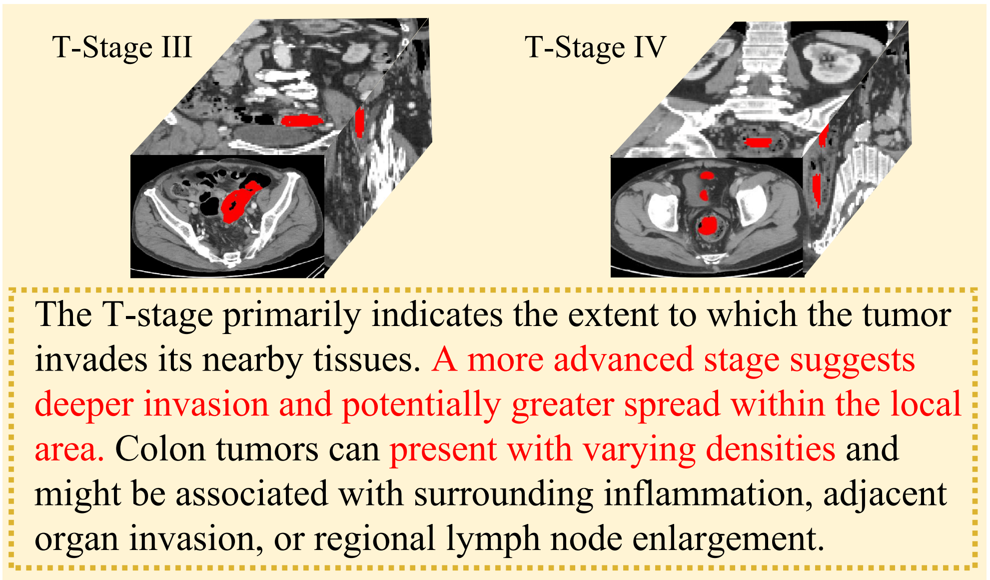
Shaoting Zhang[1,2], Xiaofan Zhang[1,2]

[1]Qing Yuan Research Institute, Shanghai Jiao Tong University [2]Shanghai AI Laboratory

Motivation

- Textual-prompted methods utilize textual representations from referred text phrases to guide the segmentation process. Data scarcity due to long-tailed distribution hinders the effective learning of alignments between textual and visual representations.
- Visual prompts provides a more intuitive and direct method to enhance the segmentation process but fail to convey the general concept of each object, leading to a performance drop when confronted with various scenarios in medical domains, especially for tumors.

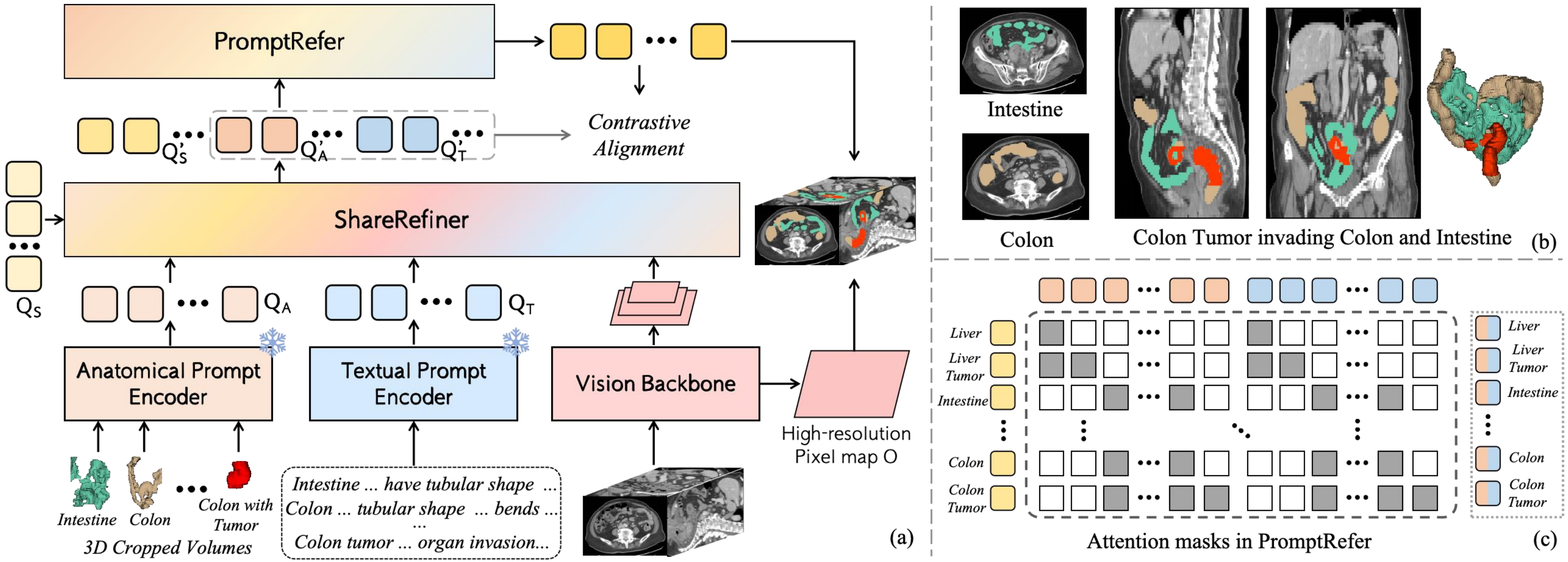




Colon Tumor in Different Cancer Stages

Method

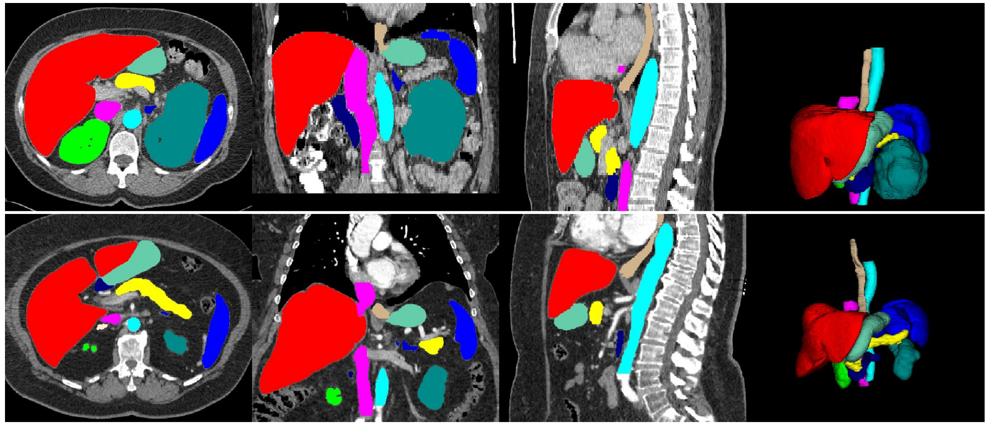
- Coordinating Anatomical-Textual Prompts (CAT)
- **Dual-perspective prompting scheme**: employ the cropped volumes derived from the anatomical structure and enhance the textual prompts with more comprehensive knowledge
- **ShareRefiner**: refine segmentation queries and prompt queries
- **PromptRefer**: updates segmentation queries by integrating both types of prompt queries



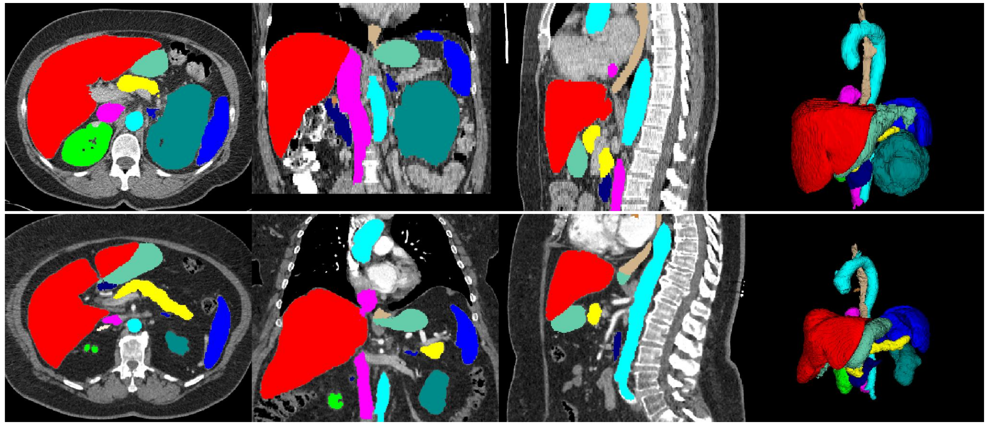
Experiments-Organ Segmentation

Methods	Liv.	R_Kid.	Spl.	Pan.	Aor.	IVC	RAG	LAG	Gal.	Eso.	Sto.	Duo.	L_Kid.	Avg.
SAM [22]	86.0	87.6	84.5	53.4	77.5	44.5	19.4	33.9	52.4	35.2	68.0	44.4	82.6	59.2
MedSAM [17]	93.0	90.0	89.1	73.5	82.5	76.5	36.0	48.7	56.4	64.7	84.0	53.9	89.7	72.2
SAM-Med2D [19]	91.4	83.7	83.9	58.8	60.6	18.6	10.6	27.1	32.9	28.1	72.9	45.4	86.0	53.8
SAM-Med3D [40]	85.4	84.2	84.7	46.9	60.4	44.5	32.6	35.3	56.0	32.6	46.9	27.4	84.9	55.5
SegVol [39]	83.9	71.7	75.9	69.4	83.1	80.3	42.1	49.7	55.6	69.6	81.1	55.6	75.1	68.7
CT-SAM3D [41]	95.6	95.0	96.1	83.6	**94.5**	**91.8**	**78.4**	**82.5**	**88.4**	**82.9**	**92.3**	73.2	94.8	**88.4**
Universal† [13]	97.4	95.5	96.4	73.7	84.9	84.4	72.9	73.4	86.0	76.8	88.5	**74.5**	96.9	84.7
ZePT* [12]	96.7	95.6	96.6	84.3	90.0	84.4	67.2	66.8	79.6	74.2	85.2	59.1	97.2	82.8
CAT	**97.7**	**96.3**	**97.1**	**89.2**	90.5	88.0	73.6	74.3	83.0	80.1	88.2	73.4	**97.3**	86.8

Organ segmentation performance on FLARE22. The results(%) are evaluated by DSC. Liv.-Liver, R_Kid.-Right Kidney, Spl.-Spleen, Pan.-Pancreas, Aor.-Aorta, IVC-Inferior Vena Cava, RAG-Right Adrenal Gland, LAG-Left Adrenal Gland, Gal.-Gallbladder, Eso.-Esophagus, Sto.-Stomach, Duo.-Duodenum, L_Kid.-Inferior Vena Cava.



Ground Truth

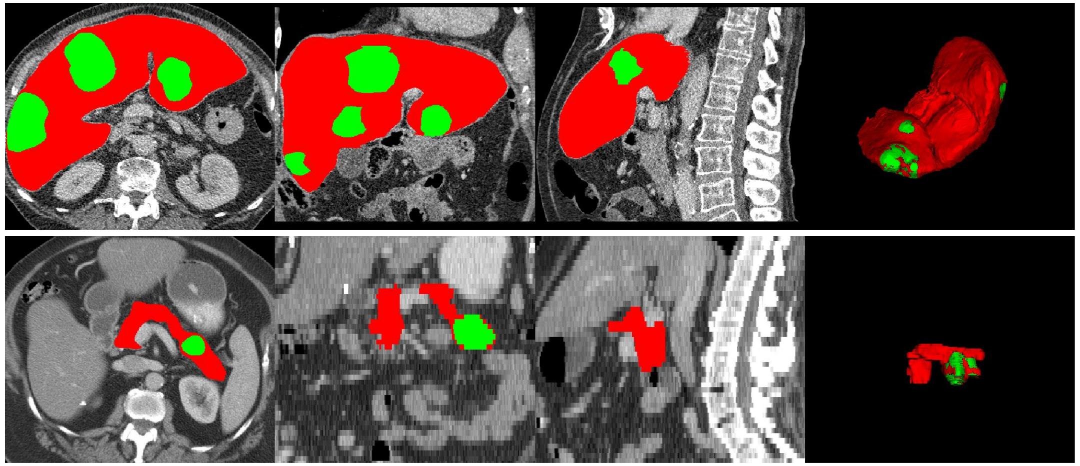


CAT

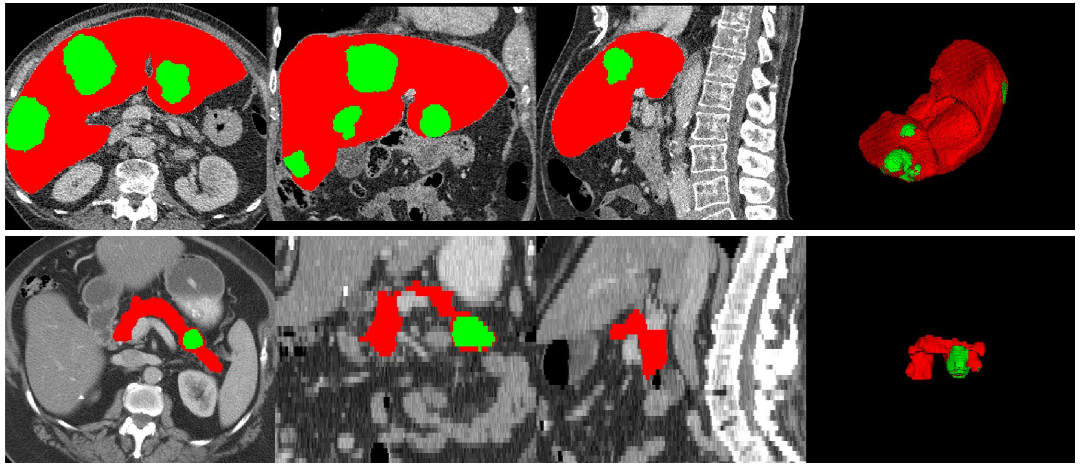
Experiments-Tumor Segmentation

Method	MSD Dataset (Tumor in Abdomen)								In-house Data (Colon Tumor)					
	Liver		Pancreas		Hepatic Vessel		Colon		T1	T2	T3	T4	Avg.	
	DSC↑	HD95↓	DSC↑	HD95↓	DSC↑	HD95↓	DSC↑	HD95↓	DSC↑				DSC↑	HD95↓
nnUNet* [62]	66.42	42.29	43.50	25.80	66.90	47.59	41.41	153.06	19.51	45.06	44.87	45.54	43.00	150.48
Swin UNETR* [48]	68.67	42.54	41.77	22.87	63.32	44.02	39.35	161.26	21.40	33.32	46.11	52.72	45.92	168.25
SAM-Med3D† [40]	44.78	-	40.05	-	44.86	-	39.23	-	34.28	42.65	50.20	42.65	47.11	-
SegVol† [39]	66.20	-	46.36	-	68.57	-	**60.63**	-	**36.93**	42.63	**60.17**	49.83	50.28	-
Universal† [13]	65.68	63.31	45.72	16.58	66.31	51.47	42.26	115.40	7.11	43.28	46.52	53.08	47.14	140.28
ZePT* [12]	68.58	43.23	44.39	19.47	68.12	33.94	40.38	113.07	23.87	34.64	50.81	51.09	46.28	155.83
CAT	**72.73**	**34.64**	**49.67**	**15.56**	**70.11**	**33.44**	48.31	**108.26**	30.62	**45.61**	55.85	**57.37**	**53.35**	**80.96**

Segmentation performance (%) of tumors on MSD and In-house dataset.

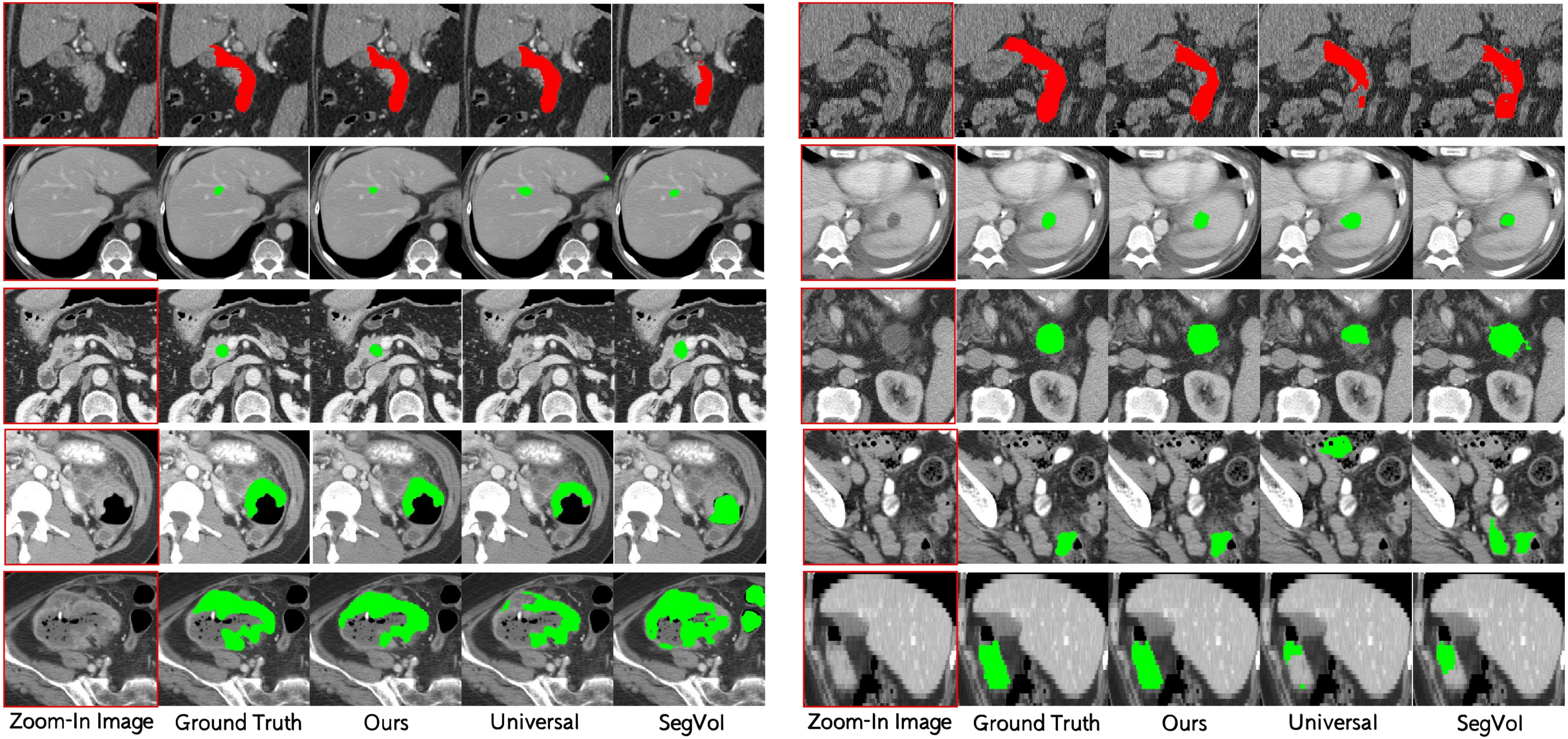


Ground Truth



CAT

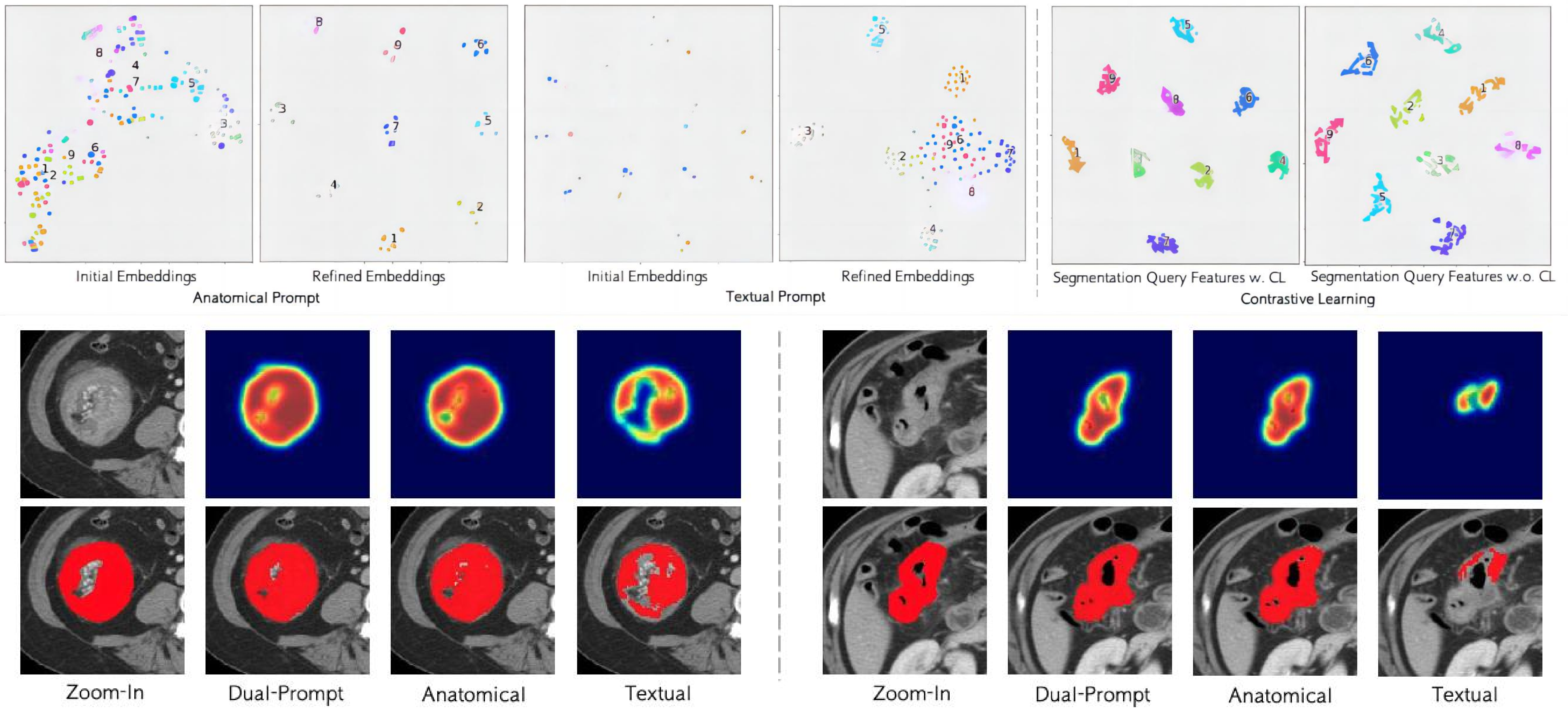
Experiments-Qualitative Comparison



Experiments-Ablation Study

Variant				Organ (%)					Tumor (%)				
AP	TP	Hard	Mask	Pan.	RAG	LAG	Eso.	Duo.	Liver	Pancreas	HepVes.	Colon	T4
				78.18	69.63	69.08	76.99	54.39	66.37	42.05	62.20	39.85	51.17
✓				83.55	72.80	71.65	79.31	60.45	64.82	45.08	68.72	43.84	53.91
	✓			80.62	71.02	70.34	72.81	57.31	69.13	44.31	65.18	40.16	52.32
✓	✓			83.50	72.71	69.96	78.99	64.08	72.49	44.55	69.40	44.50	55.84
	✓		✓	86.74	72.41	69.00	77.68	59.99	69.12	43.23	67.75	41.32	54.33
✓	✓	✓		87.36	**74.46**	74.02	75.39	70.80	72.64	48.49	69.02	47.29	53.67
✓	✓		✓	88.49	73.24	74.51	**80.76**	70.26	72.18	46.46	69.97	46.65	**58.49**
✓	✓	✓	✓	88.28	74.42	72.50	79.30	71.26	70.95	45.52	69.51	46.07	56.41
✓	✓	✓	✓	**89.24**	73.69	**74.63**	80.10	**73.46**	**72.73**	**49.67**	**70.11**	**48.31**	57.37

Experiments-Visualization



T-SNE visualization of the distribution of Features and Heatmaps.

Conclusion

- A promising attempt towards comprehensive medical segmentation via coordinating anatomical-textual prompts.
- Apart from performing generic organ segmentation, CAT can identify varying tumors without human interaction.
- To effectively integrate two prompt modalities into a single model, we design ShareRefiner to refine latent prompt queries with different strategies and introduce PromptRefer with specific attention masks to assign prompts to segmentation queries for mask prediction.
- Extensive experiments indicate that the coordination of these two prompt modalities yields competitive performance on organ and tumor segmentation benchmarks. Further studies revealed the robust generalization capabilities to segment tumors in different cancer stages.

Thanks for Listening !

Code Link: https://github.com/zongzi3zz/CAT

Contact: huangzhongzhen@sjtu.edu.cn