上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Advancing Generalizable Tumor Segmentation with Anomaly-Aware Open-Vocabulary Attention Maps and Frozen Foundation Diffusion Models

Yankai Jiang[1], Peng Zhang[2], Donglin Yang[3], Yuan Tian[1], Hai Lin[2], Xiaosong Wang[1]

[1]Shanghai AI Laboratory [2]Zhejiang University [3]The University of British Columbia

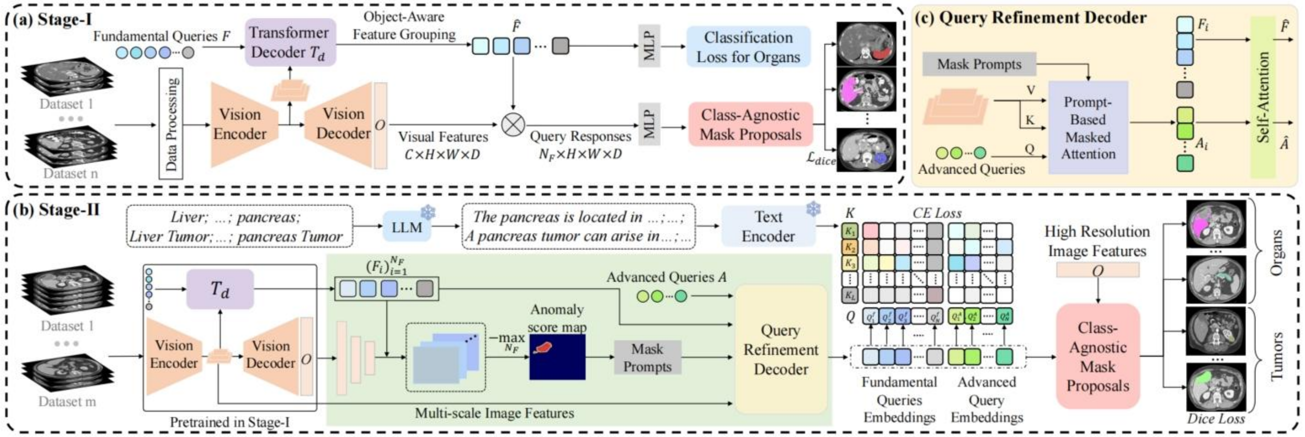
Background

Task Definition：

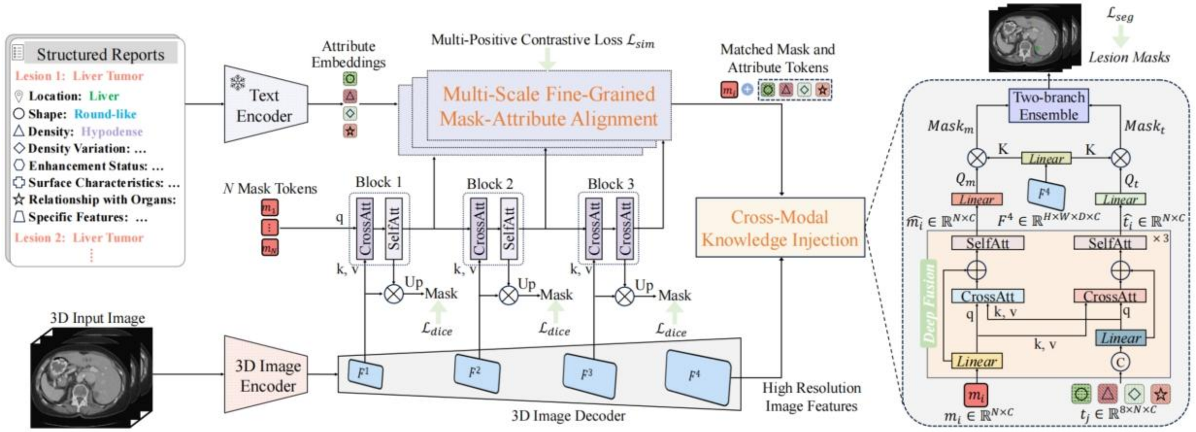
- We investigate generalizable tumor segmentation, aiming to train a single model that can not only segment tumor types seen during training but also generalize in a zero-shot fashion to unseen tumor categories, thereby enabling tumor segmentation across diverse anatomical regions.
- Existing methods face limitations related to segmentation quality, scalability, and the range of applicable imaging modalities.

Design deficiencies of existing methods

1. Vision-Language Models for zero-shot 3D medical image segmentation

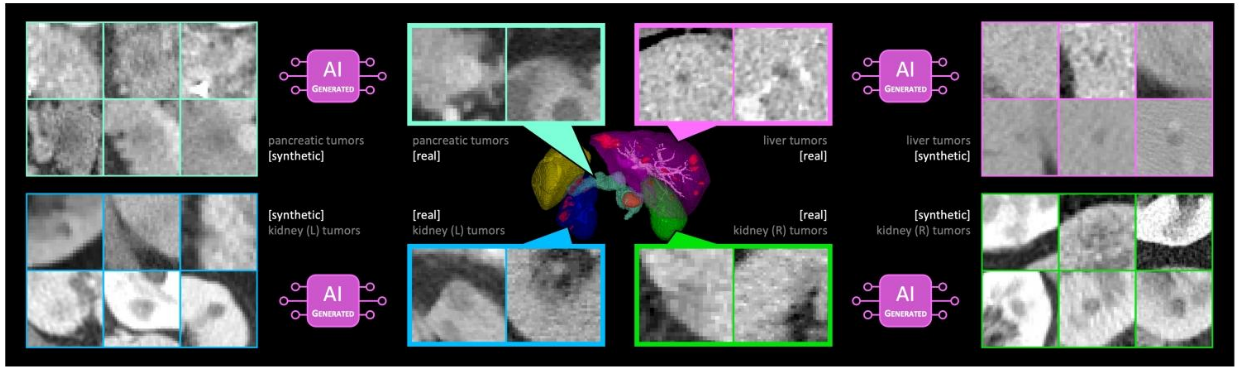


ZePT (CVPR 2024)

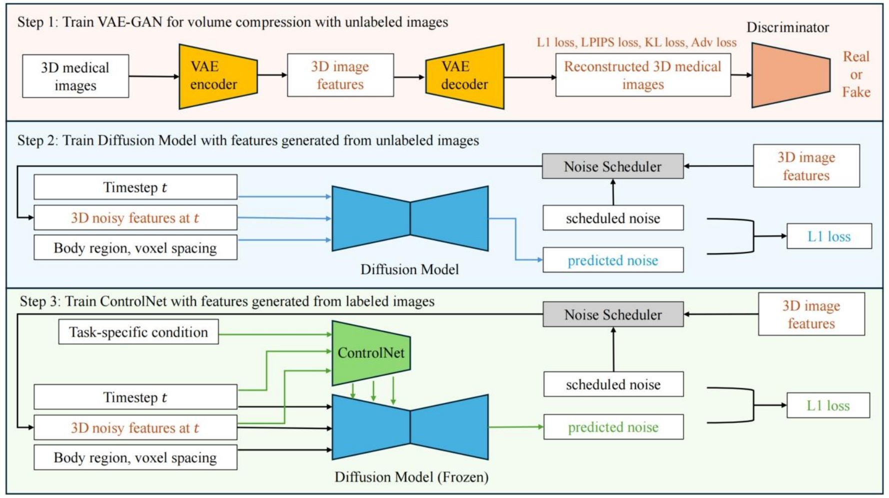


Malenia (ICLR 2025)

2. Tumor synthesis for label-free tumor segmentation



Towards Generalizable Tumor Synthesis
Chen, et al. CVPR 2024



MAISI (Guo, et al.)

Motivation

We take an innovative approach towards generalizable tumor segmentation.

Key insights:

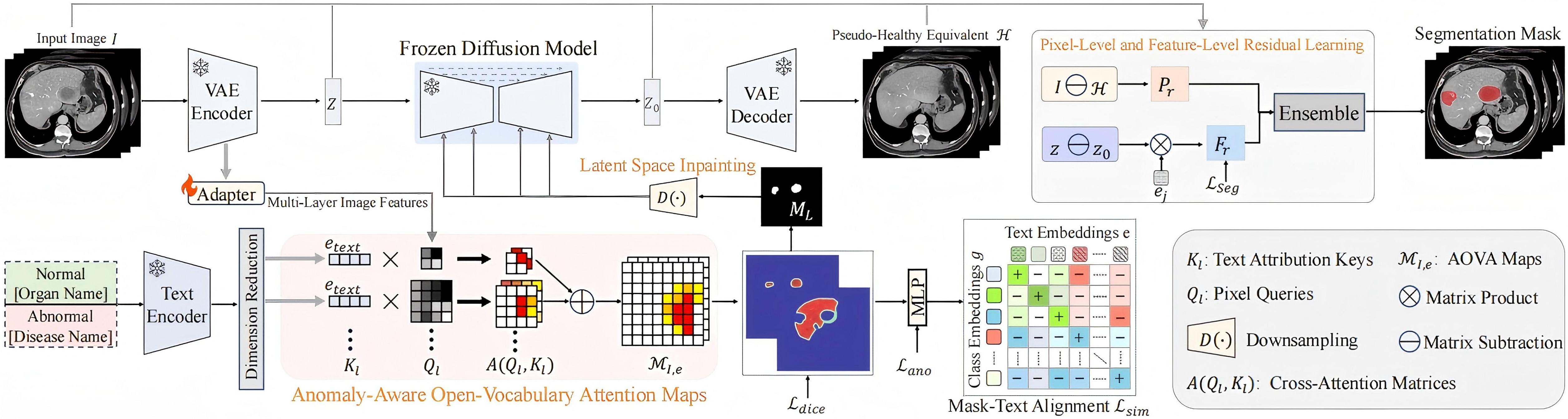
- Despite the challenges in simulating tumors, a medical foundation diffusion model (MFDM) trained on large-scale data is capable of understanding rich, diverse anatomical structures and organ-specific knowledge. Moreover, this valuable knowledge is already embedded within its internal representations.

- While directly synthesizing a wide variety of tumor types may be challenging, synthesizing healthy organs is relatively feasible with medical foundation diffusion models. Then the tumor region can then be obtained by computing the discrepancies between the input image and the synthetic image.

Challenges:

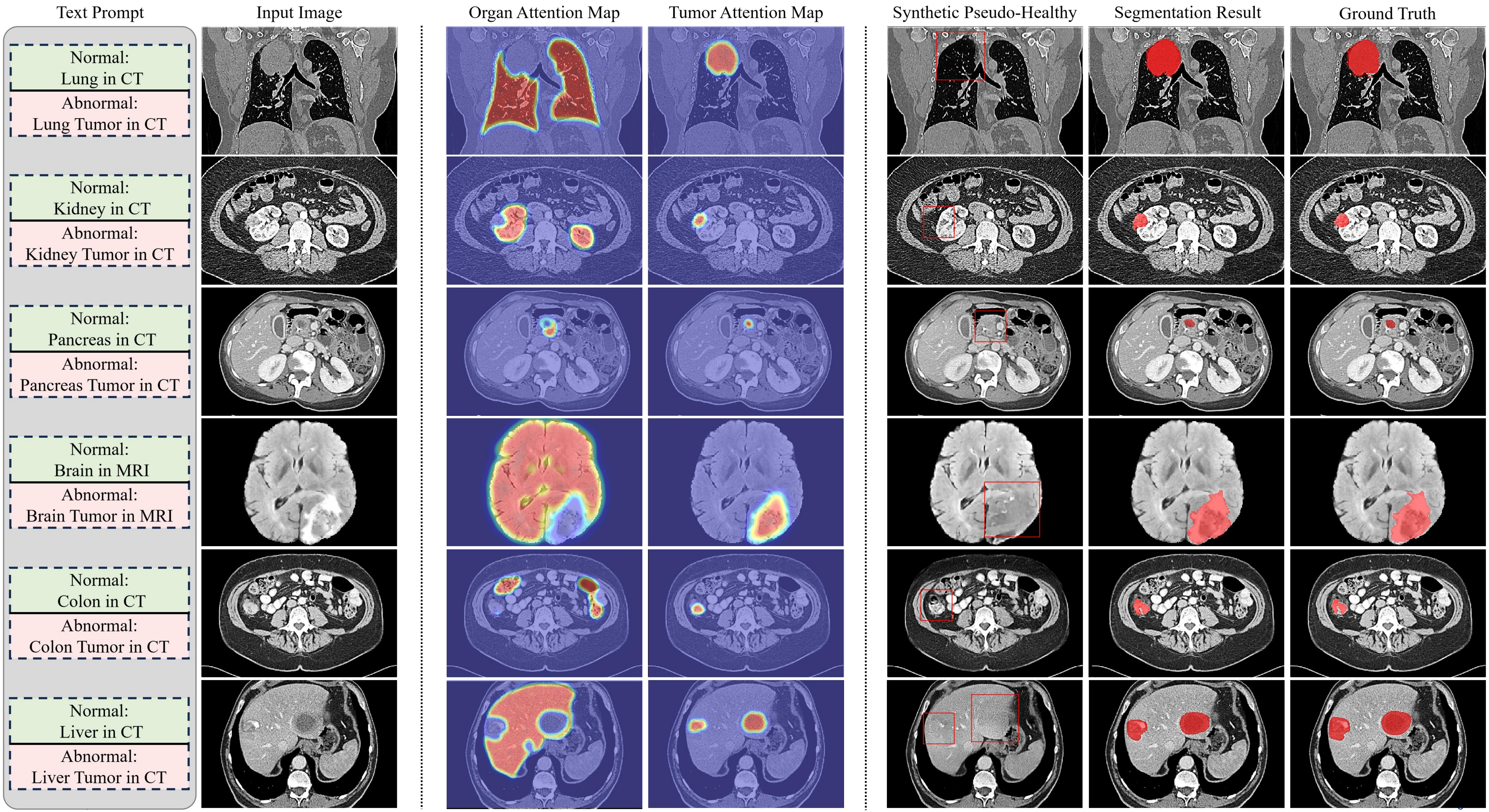
How to enable the model to detect diseased organs in a zero-shot manner, generate their corresponding healthy counterparts, and then achieve generalizable tumor Segmentation?

Method



Key Techniques:

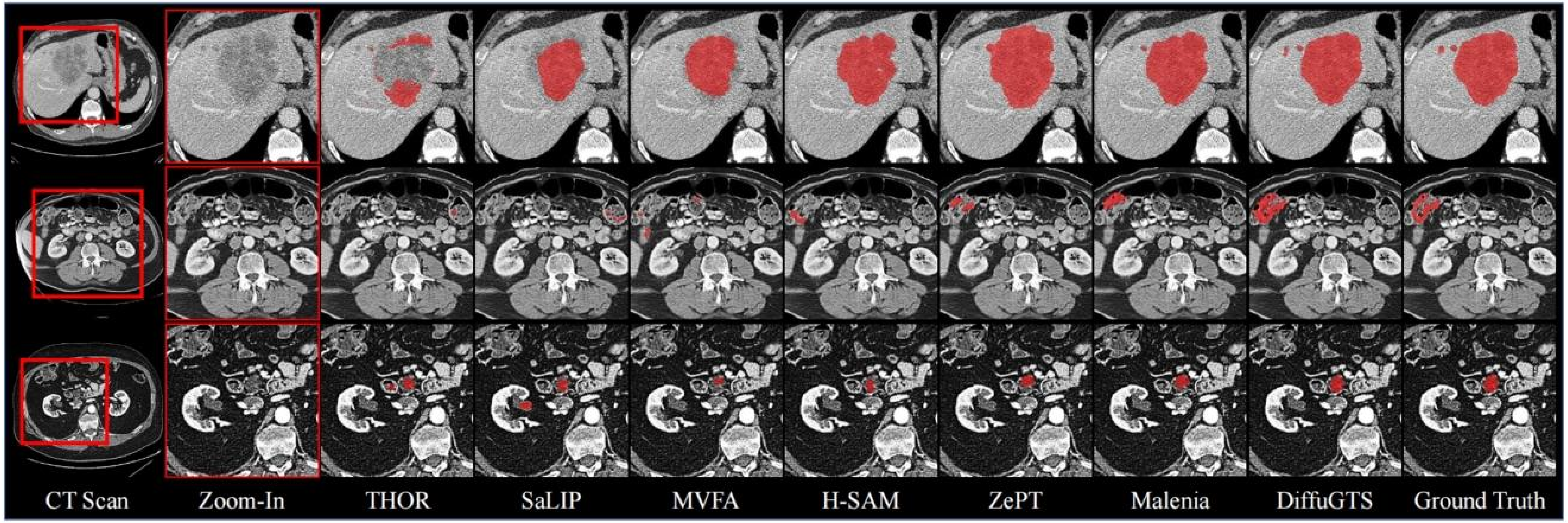
- Anomaly-aware open-vocabulary attention (AOVA) maps.
- Text-driven region-level anomaly detection.
- Mask refinement with frozen diffusion model.
- Training-free latent-space inpainting.
- Pixel-level and feature-level residual learning.



(a) Text Prompts and Input Images (b) Anomaly-Aware Open-Vocabulary Attention Maps (c) Mask Refinement with Frozen Diffusion Model

Results

Method Type	Method	MSD Dataset										KiTS23 Dataset	
		Pancreas Tumor		Lung Tumor		Liver Tumor		Colon Tumor		Hepatic Vessel Tumor		Kidney Tumor	
		DSC↑	NSD↑	DSC↑	NSD↑	DSC↑	NSD↑	DSC↑	NSD↑	DSC↑	NSD↑	DSC↑	NSD↑
SAM-based Methods	SAM 2 [47]	26.65	38.77	15.39	15.96	42.75	50.22	13.08	21.40	38.92	49.08	36.48	42.59
	SaLIP [2]	31.28	44.33	20.05	20.77	48.39	56.90	19.33	27.02	44.18	55.84	39.11	45.24
	H-SAM [10]	35.19	50.02	25.36	26.11	53.03	60.44	23.02	30.67	51.85	61.24	45.57	52.16
3D Zero-Shot Lesion Segmentation Methods	ZePT [22]	39.40	54.76	30.02	31.23	59.16	68.72	33.85	42.31	55.83	65.72	48.75	54.91
	Malenia [23]	40.26	55.82	32.75	33.92	59.83	70.08	34.72	42.59	59.71	69.98	55.37	61.16
Medical Anomaly Detection Methods	DDPM-MAD [41]	28.52	40.18	25.70	26.81	43.54	51.03	13.97	21.84	39.57	49.62	36.55	42.72
	MVFA [20]	32.77	45.63	24.49	25.07	49.25	57.72	20.44	27.87	45.28	56.14	40.81	46.29
	THOR [5]	29.45	41.06	27.73	28.98	47.68	56.40	18.36	26.51	41.33	52.05	38.39	44.64
	DiffuGTS	**43.61**	**58.48**	**42.94**	**44.01**	**63.23**	**73.58**	**38.72**	**45.60**	**62.76**	**72.35**	**59.80**	**65.99**



上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Thanks!