

Supplementary Materials

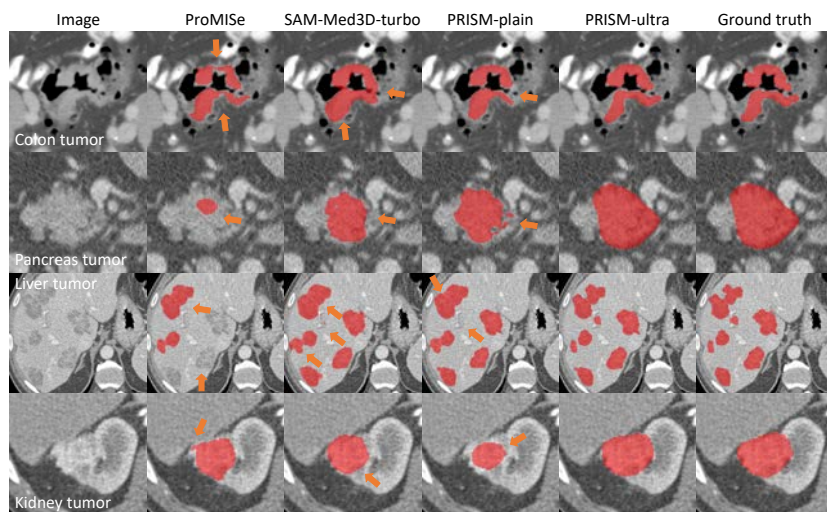


Fig. S.1: Qualitative results of four different tumor segmentation tasks. The orange arrows indicate the major defects.

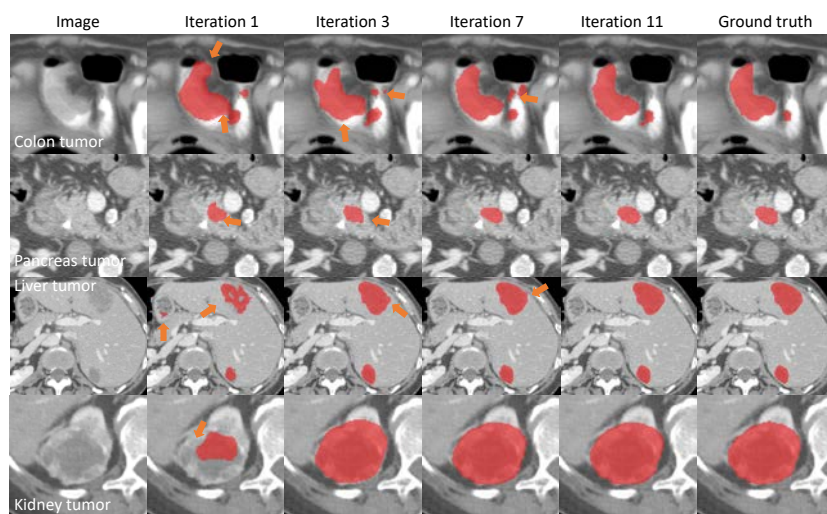


Fig. S.2: Qualitative results of PRISM-ultra across iterations. The orange arrows indicate the major defects which are corrected in the subsequent iteration.

Table S.1: Detailed settings for architecture and learning strategy in the proposed ablation studies.

Variants	Encoder type			Learning strategy	
	ViT	CNN	ViT + CNN	Confidence learning	Corrective learning
ViT encoder	✓				
CNN encoder		✓			
Hybrid			✓		
PRISM-plain			✓	✓	✓
PRISM-plain-b			✓	✓	✓
no ConL			✓		✓
PRISM-basic			✓	✓	✓
no CorL			✓	✓	
PRISM-ultra			✓	✓	✓

Table S.2: Detailed prompt setting for the proposed ablation studies. Tr. and T. represent training and test, respectively. The notation [1, 50] indicates the range from which the number of point prompts is randomly sampled during training. The term “varies” refers to the number of points used in tests, which include 1, 10, 50, and 100. “-erode” and “-dilate” denote box sizes that are 5 voxels smaller or larger in each dimension, respectively.

Variants	Detailed prompt setting				
	point	point num. (Tr. / T.)	box	box type in T.	scribble usage
ViT encoder	✓	1 / 1			
CNN encoder	✓	1 / 1			
PRISM-plain	✓	1 / 1			
PRISM-plain-b	✓	1 / 1	✓	tight	
no ConL	✓	1 / 1	✓	tight	
-plain-b-1	✓	1 / 1	✓	tight	
-plain-b-50	✓	50 / varies	✓	tight	
PRISM-basic	✓	[1, 50] / 1	✓	tight	
no CorL	✓	[1, 50] / 1	✓	tight	
-erode	✓	[1, 50] / varies	✓	undersized	
-dilate	✓	[1, 50] / varies	✓	oversized	
PRISM-ultra	✓	[1, 50] / varies	✓	tight	T.
PRISM-ultra+	✓	[1, 50] / varies	✓	tight	Tr. and T.