

Unified Prompt-Visual Interactive Segmentation of Clinical Target Volume in CT for Nasopharyngeal Carcinoma with Prior Anatomical Information

Supplementary Material

Table 1. Ablation study combining different prompts. † indicate SAM-RT outperforms other methods with p -value < 0.05 .

Prompt Methods					DSC (%) ↑		ASD (mm) ↓	
Q_1	Q_5	Q_{10}	B	M	mean ± std.	med.	mean ± std.	med.
✓					80.07 ± 15.41 [†]	80.23	6.81 ± 3.49	6.98
	✓				83.59 ± 14.02 [†]	84.55	6.41 ± 3.55 [†]	6.37
		✓			85.57 ± 12.68 [†]	84.73	5.99 ± 2.38 [†]	6.05
			✓		86.66 ± 10.09 [†]	87.34	5.74 ± 3.50	5.97
				✓	82.43 ± 14.55	81.95	8.84 ± 4.58 [†]	8.50
		✓	✓		88.45 ± 13.69 [†]	89.52	4.41 ± 3.35 [†]	4.87
		✓		✓	87.97 ± 11.32 [†]	88.05	5.54 ± 2.32 [†]	5.21
		✓	✓	✓	91.05 ± 13.54	90.65	2.19 ± 1.47	1.85

Table 2. Ablation study using different ViT.

SAM Backbone	DSC (%) ↑		ASD (mm) ↓		HSD (mm) ↓	
	mean ± std.	med.	mean ± std.	med.	mean ± std.	med.
ViT-L	84.22 ± 14.97 [†]	84.59	5.61 ± 3.19 [†]	6.05	24.68 ± 13.02 [†]	24.11
ViT-B	88.67 ± 15.69 [†]	88.43	3.95 ± 2.05 [†]	4.35	20.41 ± 12.66 [†]	20.05
ViT-H	91.05 ± 13.54	90.65	2.19 ± 1.47	1.85	15.57 ± 10.38	14.90

Table 3. Ablation study on the LoRA and SeqLoRA applied to different projection layers.

Projection Layer	DSC (%) ↑		ASD (mm) ↓		HSD (mm) ↓	
	mean ± std.	med.	mean ± std.	med.	mean ± std.	med.
Q	81.75 ± 19.56 [†]	82.30	4.55 ± 2.93 [†]	5.12	20.74 ± 12.85 [†]	21.49
Q+V	91.05 ± 13.54	90.65	2.19 ± 1.47	1.85	15.57 ± 10.38	14.90
Q+V+K	77.43 ± 20.56 [†]	78.32	5.98 ± 3.61	6.08	22.48 ± 16.19 [†]	23.25
Q+V+K+O	60.71 ± 26.33 [†]	61.25	7.67 ± 4.32 [†]	7.29	28.54 ± 19.53 [†]	29.12

Table 4. Ablation study on the rank size of LoRA and SeqLoRA.

Rank Size	DSC (%) ↑		ASD (mm) ↓		HSD (mm) ↓	
	mean ± std.	med.	mean ± std.	med.	mean ± std.	med.
1	78.65 ± 15.97 [†]	78.57	5.41 ± 3.17 [†]	5.55	18.23 ± 17.29 [†]	17.95
4	91.05 ± 13.54	90.65	2.19 ± 1.47	1.85	15.57 ± 10.38	14.90
16	70.12 ± 17.34 [†]	71.29	6.52 ± 2.97 [†]	6.98	22.64 ± 13.27 [†]	22.51
64	64.37 ± 19.64 [†]	65.69	7.12 ± 4.52 [†]	7.59	26.64 ± 18.99 [†]	26.17

Table 5. Quantitative evaluation of segmentation performance on public datasets S_G , S_O and clinical dataset S_C^f for SAM-RT.

Datasets	DSC (%) \uparrow		ASD (mm) \downarrow		HSD (mm) \downarrow	
	mean \pm std.	med.	mean \pm std.	med.	mean \pm std.	med.
S_C^f : 20 OARs + 1 GTV	75.49 \pm 10.62	75.91	4.27 \pm 2.70	4.55	20.59 \pm 16.74	19.35
S_G : 45 OARs + 2 GTVs	72.43 \pm 15.22	73.12	7.07 \pm 4.29	7.23	25.59 \pm 12.17	26.41
S_O : 30 OARs	76.25 \pm 18.54	76.90	6.21 \pm 4.37	6.55	23.17 \pm 17.24	23.50

Table 6. Training of State-of-the-Art Models.

Models	Fine-Tuning	Optimizer	Rank Size
nnUNet, UNETR, UNetGTV, DDNN, SI-Net	No	AdamW optimizer: $(\beta_1, \beta_2, lr) = (0.9, 0.999, 0.001)$	N/A
SAM, SAMed, SAM-Med2D, SAM-Med3D	LoRA		4
Models	Loss Function		
DDNN, SI-Net, SAM, SAMed, SAM-Med2D	2D Dice loss [21]		
nnUNet, UNETR, UNetGTV, SAM-Med3D	3D Dice loss [21]		

Table 7. Prompt Strategies.

Prompt Strategies	Detailed Descriptions
Point prompt	CTV boundary points are identified using K-Medoids sampling with a ground-truth mask, generating foreground points without extra clicks.
Bounding box prompt	Replicates the largest CTV mask's bounding box across all axial planes with a CTV mask.
Mask prompt	Marks five CTV centroids with a 20-pixel radius, setting inside pixels to 1 and outside pixels to 0.

Table 8. Notations and definitions

Notations	Definitions	Notations	Definitions
\mathcal{F}^c	Decision function for CTV task	\mathcal{T}_c	CTV contouring task
\mathcal{F}^s	Decision function for natural image task	\mathcal{T}_G	GTV and OAR contouring task
\mathcal{F}^G	Decision function for GTV and OAR task	S_G	Segrap2023 public dataset
\mathcal{D}_c	Target domain for CTV contouring task	S_O	Han-Seg public dataset
\mathcal{D}_s	Source domain for natural image	S_C^f	Clinical dataset for GTV and OAR task
\mathcal{D}_G	Target domain for GTV and OAR task	S_C	Clinical dataset for CTV task