

Iterative Residual Slice Learning based 2.5D Segment Anything Model for Promptable Medical Image Segmentation — Supplementary Material

A More Experimental Results

Experiments on Synapse dataset. Table 1 and Table 2 compare the performance of IRSL-2.5DSAM with other mainstream medical image segmentation methods on the Synapse multi-organ segmentation dataset, using Intersection over Union (IoU) and 95% Hausdorff Distance (HD95) as evaluation metrics respectively. As shown in Table 1, the average IoU of IRSL-2.5DSAM reaches 77.01%, representing an improvement of 5.35% compared to Med-SA. For multiple organs, such as the aorta and kidneys, the IoU values of IRSL-2.5DSAM surpass those of other methods. Table 2 indicates that IRSL-2.5DSAM achieves the lowest average HD95 of 9.10, significantly lower than most comparison methods, demonstrating that its segmentation results are closer to the ground truth and exhibit higher precision. We also visualize the segmentation results of the comparative experiments in Figure 1. For regions that are challenging to distinguish and segment, IRSL-2.5DSAM demonstrates a relatively accurate segmentation capability. The overall experimental results indicate that IRSL-2.5DSAM outperforms mainstream segmentation methods based on 2D, 2.5D, and SAM architectures.

Impact of slice numbers. In 3D medical image segmentation tasks, the use of adjacent slices to guide the segmentation of the current slice can provide valuable context information. Introducing more adjacent slices offers the model richer contextual information, helping it better understand the spatial relationships between slices. This is particularly beneficial in cases where anatomical structures are complex or boundaries are unclear. However, excessive adjacent slices increase the computational load, raising both computational cost and memory requirements. Therefore, investigating the impact of the number of adjacent slices on segmentation performance is crucial. We set different numbers of adjacent slices (3, 5, 7, and 9) to investigate the impact of slice quantity on medical image segmentation tasks. As shown in Table 3, we compute the Dice coefficients for each organ and the average Dice score. The model achieves the best performance in the segmentation of multiple organs when using 5 adjacent slices, with the highest average Dice score of 84.78%. However, when the number of slices is increased to 7 or 9, the model performance decreases, suggesting that more slices do not always lead to better results.

Table 1: Quantitative comparison results (in terms of IoU) between state-of-the-art methods and our proposed method (IRSL-2.5DSAM) on Synapse multi-organ segmentation dataset. The highest performance is highlighted in bold.

	Methods	Aorta	Gallbladder	Kidney (L)	Kidney (R)	Liver	Pancreas	Spleen	Stomach	Average
2D-based	U-Net	76.36	42.07	71.86	63.34	86.31	36.75	72.92	53.71	62.92
	HiFormer	73.78	41.53	74.31	66.73	88.65	38.26	79.44	63.13	65.73
	TransUnet	75.84	46.75	72.07	70.89	87.44	35.72	71.65	54.01	64.30
	TransDeepLab	66.43	42.62	71.92	65.41	86.27	38.42	76.70	59.97	63.47
	SwinUnet	62.75	40.87	67.83	56.91	88.29	33.03	68.44	57.55	59.46
	DAE-Former	77.28	49.06	76.20	66.54	90.05	44.69	82.33	69.55	69.46
	MissFormer	60.08	49.06	71.96	68.02	88.43	39.56	83.38	63.36	65.48
2.5D-based	CSAM-Net	76.79	59.52	69.43	74.41	74.68	47.29	76.31	59.26	67.21
	CSA-Net	65.97	55.15	60.24	57.96	70.01	42.61	73.62	52.87	59.80
	GLCSA	74.24	52.25	69.06	60.47	72.14	43.56	69.31	50.84	61.48
	CAT-Net	75.19	62.39	72.43	75.29	75.78	54.93	76.79	62.14	69.37
SAM-based	SAM	78.80	38.82	78.72	72.15	88.23	47.54	85.35	59.04	68.58
	SAMed	77.32	39.12	72.78	71.63	88.79	47.36	80.29	62.50	67.47
	Med-SA	79.28	49.92	80.82	78.66	85.47	50.88	83.64	64.63	71.66
	SAM-Med	81.38	44.22	83.13	81.07	82.41	42.59	82.58	67.70	70.64
	IRSL-2.5DSAM	83.85	62.61	87.19	84.57	84.34	50.94	86.26	76.35	77.01

Table 2: Quantitative comparison results (in terms of HD95) between state-of-the-art methods and our proposed method (IRSL-2.5DSAM) on Synapse multi-organ segmentation dataset. The best performance is highlighted in bold.

	Methods	Aorta	Gallbladder	Kidney(L)	Kidney(R)	Liver	Pancreas	Spleen	Stomach	Average
2D-based	U-Net	25.04	33.68	50.62	71.28	42.71	14.59	53.56	22.65	39.27
	HiFormer	16.97	28.56	48.64	26.73	25.41	12.30	29.86	13.49	25.25
	TransUnet	17.21	24.62	29.32	68.10	35.74	17.13	36.58	18.49	30.90
	TransDeepLab	13.36	28.04	41.17	60.61	23.17	13.94	25.97	16.96	27.90
	SwinUnet	23.03	32.90	31.24	43.19	12.55	14.73	26.77	16.73	25.14
	DAE-Former	13.85	27.12	57.36	55.03	21.17	10.90	31.95	15.06	29.06
2.5D-based	MissFormer	28.39	13.14	54.89	41.44	16.39	15.75	9.11	20.76	24.98
	CSAM-Net	15.92	29.41	33.44	36.28	36.07	10.20	22.37	16.04	24.97
	CSA-Net	20.36	25.61	40.28	49.41	44.59	11.24	36.56	17.50	30.69
	GLCSA	14.61	24.27	35.56	40.81	47.45	13.01	39.28	22.34	29.67
	CAT-Net	16.31	17.16	28.33	30.94	30.55	10.03	19.52	18.80	21.46
SAM-based	SAM	6.94	13.17	20.27	28.25	15.59	17.38	7.58	12.97	15.27
	SAMed	12.38	21.94	32.42	23.62	8.54	16.16	21.75	18.93	19.47
	Med-SA	5.26	17.28	3.08	6.43	14.83	20.27	7.40	19.32	11.73
	SAM-Med	3.14	32.93	4.37	5.64	9.21	27.78	6.20	21.26	13.82
	IRSL-2.5DSAM	2.90	15.66	1.50	2.94	9.29	23.61	5.00	11.87	9.10

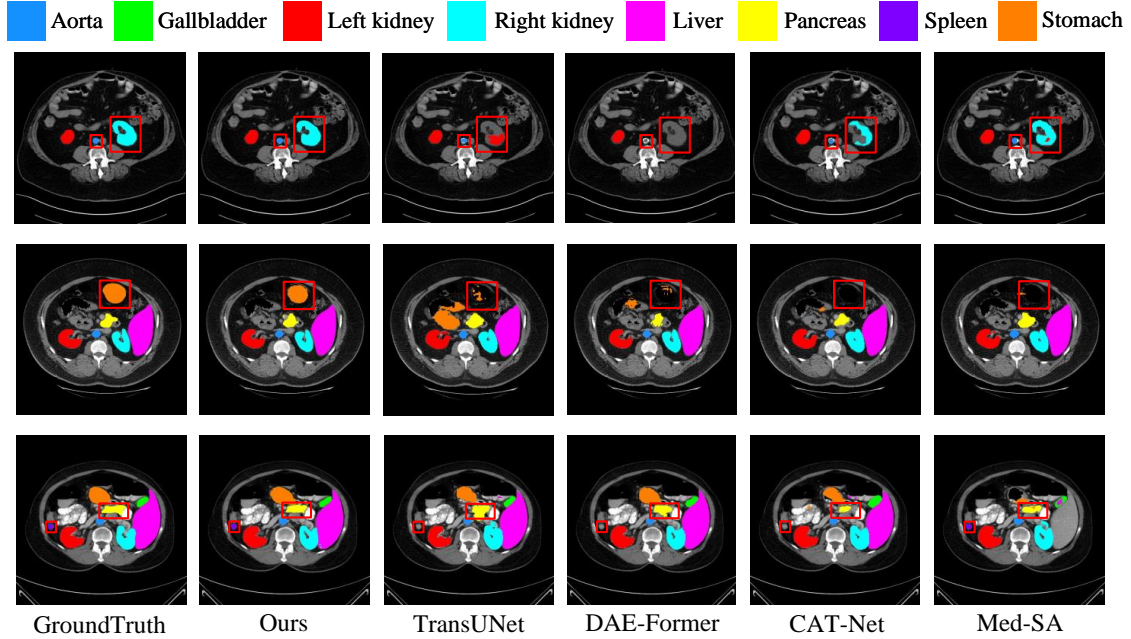


Figure 1: Visual comparison between our proposed IRSL-2.5DSAM and SOTA segmentation methods, including TransUnet, DAE-Former, CAT-Net and Med-SA.

Table 3: Comparison of the segmentation performance across different organs with varying numbers of slices.

Slice	Aorta	Gallbladder	Kidney (L)	Kidney (R)	Liver	Pancreas	Spleen	Stomach	Average
3	89.99	67.20	91.91	90.23	92.86	62.94	91.18	81.34	83.46
5	91.70	77.31	92.26	90.66	91.56	61.78	91.24	81.72	84.78
7	91.02	72.50	91.51	90.44	90.85	64.50	91.10	80.44	84.05
9	91.46	66.68	90.50	90.65	93.33	64.40	91.20	76.30	83.07