A Results of 5-fold Cross-validation

In the experiment section, we evaluate our APAUNet using 5-fold cross validation on MSD-Liver and MSD-Pancreas, here we present the complete results in Tab. 1. The mean results of our APAUNet on the Liver dataset are 95.56, 71.99 and 83.78 respectively, and the standard deviations are 0.35, 0.81 and 0.35 respectively.

Similarly, on the Pancreas dataset, our APAUNet achieves the mean results of 82.29, 54.93 and 68.61 respectively and the standard deviations are 1.73, 1.64 and 1.63 respectively. The standard deviations are slightly higher than those on the Liver dataset, since the target size of the Pancreas dataset is much smaller than that of the Liver dataset. Overall, the results of experiments demonstrate the effectiveness and stability of our APAUNet.

Table 1. The complete segmentation results of 5-fold validation on the Liver and Pancreas Tumour datasets.

Method	liver			Pancreas		
	Organ	Cancer	Avg	Organ	Cancer	Avg
fold-1	96.10	72.50	**84.30**	83.05	55.21	69.13
fold-2	**96.23**	71.45	83.84	80.16	53.20	66.68
fold-3	94.87	70.66	82.77	**83.40**	**56.66**	**70.03**
fold-4	95.48	72.66	84.07	81.88	54.30	68.09
fold-5	95.12	**72.69**	83.91	82.94	55.27	69.11
Mean	95.56	71.99	83.78	82.29	54.93	68.61
STD	±0.35	±0.81	±0.35	±1.73	±1.64	±1.63

B Detailed Computational Analysis

Here we show the computational cost analysis in Tab. 2. The inference time is measured by a sliding window inference with patch size of 96 on GPU. Our APAUNet has a similar training GPU memory with HFA-Net/UNETR and a similar inference GPU with UTNet/CoTr, and achieves the lowest inference time.

C Effect of Each Strategy

In this section, we conduct several ablation studies by adding our strategies step by step. The results are shown in Tab. 3. It can be observed that all the strategies are necessary for the final performance. Combining all the proposed strategies, our model achieves the new state-of-the-art performance. In addition, our projection strategy has lower memory consumption despite it needs replicate the 3D fused features three times.

Table 2. Computational analysis of our APAUNet and other methods. Parameters (M), training/inference GPU consumption (G) and inference time (s).

Method	Param.(M)	Training GPU(G)	Inference GPU(G)	Inference time(s)
UNet3+	94.67	30.37	14.33	39.65
HFA-Net	40.7	12.02	8.54	24.80
UTNet	**35.26**	9.72	6.01	18.16
UNETR	92.24	13.78	**3.87**	13.46
CoTr	41.86	**8.69**	6.91	21.77
APAUNet	76.94	11.74	6.42	**13.36**

Table 3. Ablation experiments under different settings. OP - orthogonal projection, DH - dimension hybridization, MRF - multi-resolution fusion.

Settings	OP	DH	MRF	GPU (G)	Liver			Pancreas		
					Organ	Cancer	Avg	Organ	Cancer	Avg
1	✓	×	×	11.75	87.42	63.88	75.65	75.40	46.28	60.84
2	×	×	✓	13.02	90.15	62.41	76.28	76.80	47.13	61.97
3	✓	✓	×	11.72	94.81	69.32	82.07	80.30	53.92	67.11
4	✓	×	✓	11.71	92.76	67.88	80.32	77.83	49.03	63.43
APAUNet	✓	✓	✓	11.74	**96.10**	**72.50**	**84.30**	**83.05**	**55.21**	**69.13**

D Weighted Importance of Projection

As we mentioned in the main paper, adding learnable weights could lead to a performance boost. In this part, we present the detailed results of the learnable weights on the Liver and Pancreas datasets. Tab. 4 shows the weights learned after the training process on Liver and Pancreas dataset, respectively. It can be observed from the results that the weights of the shallow encoders (1-3) in all directions are relatively balanced, while the encoders of deep level may selectively emphasize a certain projection axis, e.g., 0.63 of Encoder-4 on *axial*-axis on the Liver dataset and 0.46 of Encoder-4 on *sagittal*-axis on the Pancreas dataset. We conjecture that the encoder path is mainly used for feature extraction and analysis, thus the importance weight will not be particularly inclined to a certain projection axis.

While on the decoder path, the phenomenons of importance selection are more obvious. Most decoders will tend to choose one or two of the projection dimensions, e.g., *sagittal*-axis of Decoder-2 and *axial*-axis of Decoder-3 on the Liver dataset, *axial*-axis of Decoder-1 and Decoder-2 on the Pancreas dataset. We analyze that the decoder path is used to synthesize multi-scale features to obtain the segmentation prediction, thus it may highlight more salient features and filter the redundant information.

Table 4. The detailed results of learned importance weights.

Dataset	Level	Sagittal	Axial	Coronal
Liver	Encoder-1	0.28	0.37	0.35
	Encoder-2	0.31	0.34	0.35
	Encoder-3	0.31	0.30	0.38
	Encoder-4	0.27	0.63	0.10
	Encoder-5	0.20	0.39	0.39
	Decoder-1	0.36	0.30	0.34
	Decoder-2	0.70	0.16	0.14
	Decoder-3	0.09	0.78	0.13
	Decoder-4	0.30	0.42	0.28
Pancreas	Encoder-1	0.33	0.34	0.33
	Encoder-2	0.22	0.38	0.40
	Encoder-3	0.30	0.37	0.33
	Encoder-4	0.46	0.24	0.30
	Encoder-5	0.18	0.44	0.38
	Decoder-1	0.17	0.52	0.31
	Decoder-2	0.22	0.67	0.11
	Decoder-3	0.31	0.18	0.51
	Decoder-4	0.27	0.34	0.37

E More Visualization Results

In this part, we demonstrate more visualization results of our APAUNet on MSD-Liver, MSD-Pancreas and BTCV in Fig. 1, Fig. 2 and Fig. 3.

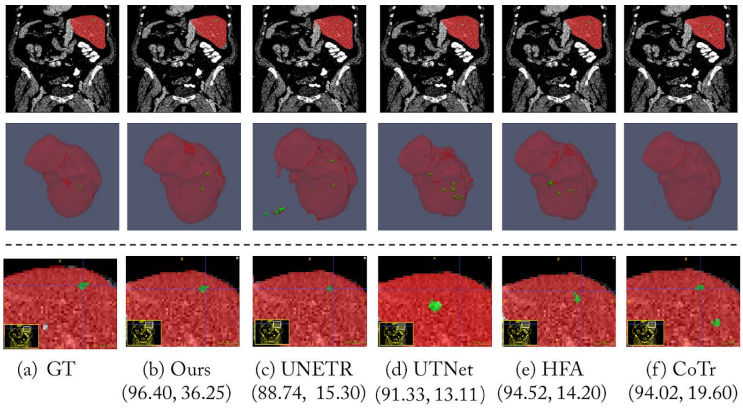

Fig. 1. More visualization results of our APAUNet and comparison methods on Liver.

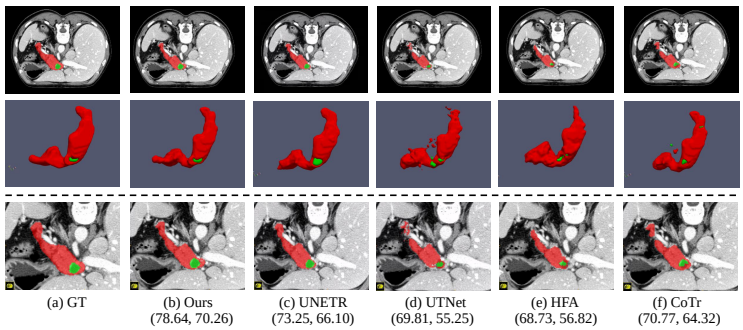

Fig. 2. More visualization results of our APAUNet and comparison methods on Pancreas.

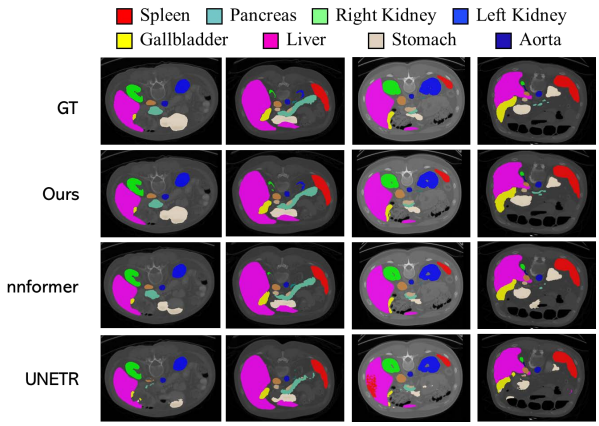

Fig. 3. More visualization results of our APAUNet and comparison methods on BTCV.