

# Appendix of “Volumetric Axial Disentanglement Enabling Advancing in Medical Image Segmentation”

## A Appendix Overview

This supplementary material presents additional results to complement the main paper. We begin by detailing the datasets and implementation, followed by a detailed presentation of the benchmarking results and ablation experiments results, including a supplementary ablation study. Additionally, we offer further qualitative results. Our code will be released after accepted.

## B Details of datasets and implementation

### B.1 Detailed information of datasets

In this section, we provide detailed descriptions of the FLARE2021 [Ma *et al.*, 2022], OIMHS [Ye *et al.*, 2023], and SegTHOR [Lambert *et al.*, 2020] datasets, along with the data used in the experiments, summarized in Table 1.

**FLARE2021.** A publicly accessible CT dataset focusing on abdominal organ segmentation, comprises 361 training instances, 50 validation instances, and 100 test instances across four categories: liver, spleen, pancreas, and kidney. The spatial resolution ranges from 0.61mm to 0.98mm in plane and 0.5mm to 7.5mm through plane. We utilize the 361 publicly labeled instances for our experiments.

**OIMHS.** A dataset pertains to the segmentation of retinal 3D OCT and encompasses 125 sequences, each containing 19 to 73 scans. It categorizes four types of retinal conditions: the retina, choroid, macular hole, and macular edema, with a spacing range from 10.7um to 14.0um in plane and 7.0um to 40.0um through plane. We resample all 125 publicly available sequences for optimized training efficacy.

**SegTHOR.** Dedicated to the segmentation of thoracic organs at risk (OARs) in CT scans, SegTHOR includes 60 CT images, split into 40 for training and 20 for testing. It features organs like the heart, trachea, main arteries, and esophagus, each presenting unique spatial and appearance characteristics. The dataset’s spacing ranges from 0.90mm to 1.37mm in plane and 2.0mm to 2.5mm through plane. We utilize the 40 publicly labeled data for our experimentation.

### B.2 Data preprocessing & Implementation detail

The data preprocessing pipeline includes the following steps: Initially, the intensity values of the datasets are clipped to standardized ranges: FLARE2021 to  $[-125, 275]$ , OIMHS to

$[0, 300]$ , and SegTHOR to  $[-1000, 300]$ . Subsequently, min-max normalization  $(X - X_1)/(X_{99} - X_1)$  is applied to each volume, scaling the data to the  $[0, 1]$  range, where  $X_p$  denotes the  $p$ -th percentile of intensity in  $X$ . During training, volumes of size  $96 \times 96 \times 96$  are randomly cropped and augmented through random rotation, scaling, and intensity shifts to enhance model generalization. Experiments are conducted under uniform software and hardware conditions, specifically on servers with two NVIDIA GeForce RTX 4090 GPUs and 128GB of memory, using Python 3.9, PyTorch 2.0.0, and MONAI 0.9.0 within the DDP framework. We further comprehensively summarize the training parameters in Table 2.

### B.3 Evaluation metrics

In this study, we utilize IoU/mIoU, Dice, and HD95 as evaluation metrics to comprehensively assess segmentation performance. IoU measures overall accuracy in multi-class segmentation tasks, providing a robust assessment of the overlap between predicted and ground truth. Dice is particularly effective for medical image segmentation and excels in evaluating small target regions by emphasizing precise delineation. HD95 quantifies boundary accuracy, offering sensitivity to boundary detail errors, and is crucial for capturing intricate anatomical structures and variations.

## C Additional ablation study

### C.1 Parameter initialization selection.

We conduct an additional ablation experiment on the initialization of trainable parameters  $W$  and  $b$  within the SAT’s adaptive weighting function  $\mathcal{F}(x)$ , aiming to assess the impact of different initialization configurations on optimizing feature distribution time-axial disentanglement. As depicted in Table 6, the optimal result is achieved when the weighting parameter  $W$  is initialized to 1 and the bias  $b$  to 0. This initial setting allows the model to maintain the original distribution of axial features during the early training phases, preventing premature loss or distortion of critical information. As training progresses, these parameters are automatically adjusted via gradient descent to weight axial information optimally.

### C.2 Group effects on spatial redistribution.

We evaluated the impact of varying group numbers within the spatial redistribution on the FLARE2021 dataset, with re-

Table 1: Complete detailed information of three publicly available Datasets used in the paper.

Datasets	FLARE 2021	OIMHS	SegTHOR
Imaging Modality	Multi-Contrast CT	OCT	Multi-Contrast CT
Anatomical Region	Abdomen	Eye	Chest
Sample Size	361	125	40
Dimensions	$512 \times 512 \times \{37 - 751\}$	$512 \times 512 \times \{19 - 73\}$	$256 \times 256 \times 256$
Resolution	$\{0.61 - 0.98\} \text{mm} \times \{0.61 - 0.98\} \text{mm} \times \{0.50 - 7.50\} \text{mm}$	$\{10.7 - 14.0\} \mu\text{m} \times \{10.7 - 14.0\} \mu\text{m} \times \{7.0 - 40.0\} \mu\text{m}$	$\{0.90 - 1.37\} \text{mm} \times \{0.90 - 1.37\} \text{mm} \times \{2.00 - 2.50\} \text{mm}$
Anatomical Label	Spleen, Kidney, Liver, Pancreas	Retinal, Choroid, Macular hole, Macular edema	Esophagus, Heart, Trachea, Aorta

Table 2: Hyperparameters of training scenarios on three public datasets.

Training Steps	80000
Batch Size	2
AdamW <sub><math>\epsilon</math></sub>	$1e-8$
AdamW <sub><math>\beta</math></sub>	(0.9, 0.999)
Peak Learning Rate	$1e-4$
Learning Rate Scheduler	ReduceLROnPlateau
Factor & Patience	0.9, 10
Data Augmentation	Intensity Shift, Rotation, Scaling
Cropped Foreground	✓
Intensity Offset	0.1
Rotation Degree	( $-30^\circ$ , $+30^\circ$ )
Scaling Factor	0.1

sults presented in Table 7. It reallocates and supplements specific axial spatio-temporal information within feature maps, enhancing the model’s ability of feature representation and achieving more accurate segmentation. An appropriate number of groups can balance the information load within each group and the model’s overall efficiency in capturing supplementary time-axial features. Too many groups can lead to excessive information fragmentation, while too few may fail to capture sufficient detail and variability. Testing with 4, 8, 16, and 24 groups revealed that setting the group number to 8 yielded the best results across various metrics.

## D Detailed results of benchmarking and ablation studies

This section provides more detailed results of the benchmarking and ablation experiments. Tables 3, 4, and 5 detail performance metrics of the benchmarking results across all categories for the FLARE2021, OIMHS, and SegTHOR datasets, respectively. Tables 8, 9, 10, 13, 14, present detailed results for each category in the ablation studies. The benchmarking results further demonstrate effective improvements in segmentation accuracy and boundary precision with the integration of PaR. The ablation study explores the optimal hyperparameter settings and analyzes the effectiveness and robustness of PaR by examining fusion strategies, time-axial information density, disentanglement along different axes, self-attention heads, and preliminary feature extraction layer kernel sizes.

## E Qualitative results

In this section, we present additional qualitative results to show the impact of integrating the PaR module with benchmark models on the FLARE2021, OIMHS, and SegTHOR

datasets. Figures 1, 2, and 3 illustrate comparative visualizations of performance transformation for 11 previous segmentation methods incorporating PaR post-decoder on the FLARE2021, OIMHS, and SegTHOR test sets, respectively. These visualizations exemplify the enhancement in segmentation accuracy and boundary delineation facilitated by PaR, highlighting its efficacy in refining spatial resolution and structural detail through the axial disentangling and attention enhancement of time-axial features. The results consistently showcase improved boundary clarity and overall segmentation precision, underscoring the significant performance gains achieved by the PaR module across diverse datasets and baseline models.

## References

- [Lambert *et al.*, 2020] Zoé Lambert, Caroline Petitjean, Bernard Dubray, and Su Kuan. Segthor: Segmentation of thoracic organs at risk in ct images. In *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2020.
- [Ma *et al.*, 2022] Jun Ma, Yao Zhang, Song Gu, Xingle An, Zhihe Wang, Cheng Ge, Congcong Wang, Fan Zhang, Yu Wang, Yinan Xu, et al. Fast and low-gpu-memory abdomen ct organ segmentation: the flare challenge. *Medical Image Analysis*, 82:102616, 2022.
- [Ye *et al.*, 2023] Xin Ye, Shucheng He, Xiaxing Zhong, Jiafeng Yu, Shangchao Yang, Yingjiao Shen, Yiqi Chen, Yaqi Wang, Xingru Huang, and Lijun Shen. Oimhs: An optical coherence tomography image dataset based on macular hole manual segmentation. *Scientific Data*, 10(1):769, 2023.

Table 3: Detailed results for each category in the benchmarking experiments on the FLARE2021 dataset show the performance transformation of 11 previous segmentation models integrating PaR post-decoder. The metrics for all models with integrated PaR are highlighted in bold.

Method	Average			Liver			Kidney			Spleen			Pancreas		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
3D U-Net +PaR	87.92 <b>89.20</b>	93.08 <b>93.89</b>	16.31 <b>2.49</b>	95.63 <b>96.36</b>	97.75 <b>98.14</b>	9.14 <b>1.67</b>	92.16 <b>93.64</b>	95.89 <b>96.70</b>	17.05 <b>2.30</b>	95.33 <b>95.61</b>	97.60 <b>97.74</b>	21.11 <b>1.33</b>	68.55 <b>71.19</b>	81.08 <b>82.96</b>	17.96 <b>4.67</b>
RAUNet +PaR	87.93 <b>88.32</b>	93.08 <b>93.28</b>	27.37 <b>26.71</b>	95.38 <b>96.49</b>	97.63 <b>98.21</b>	60.58 <b>1.58</b>	92.73 <b>93.11</b>	96.21 <b>96.42</b>	2.48 <b>2.28</b>	95.17 <b>95.92</b>	97.51 <b>97.91</b>	<b>9.38</b> 11.60	<b>68.42</b> 67.76	<b>80.99</b> 80.59	<b>37.04</b> 91.38
ResUNet +PaR	87.38 <b>87.83</b>	92.56 <b>92.94</b>	30.15 <b>11.80</b>	<b>96.29</b> 96.01	<b>98.10</b> 97.95	<b>1.80</b> 15.04	<b>93.14</b> 92.83	<b>96.43</b> 96.26	<b>2.55</b> 10.56	<b>95.78</b> 95.47	<b>97.84</b> 97.67	<b>3.53</b> 15.20	64.30 <b>67.02</b>	77.88 <b>79.86</b>	112.73 <b>6.40</b>
SegResNet +PaR	86.24 <b>88.02</b>	91.81 <b>93.13</b>	3.22 <b>2.78</b>	95.37 <b>96.14</b>	97.63 <b>98.03</b>	1.91 <b>1.68</b>	91.66 <b>93.06</b>	95.63 <b>96.39</b>	2.71 <b>2.28</b>	<b>94.76</b> 94.52	<b>97.31</b> 97.13	<b>1.18</b> 1.79	63.14 <b>68.37</b>	76.69 <b>80.95</b>	7.09 <b>5.37</b>
V-Net +PaR	84.13 <b>85.99</b>	89.89 <b>91.49</b>	12.93 <b>5.98</b>	94.66 <b>95.28</b>	97.24 <b>97.57</b>	14.32 <b>10.83</b>	91.93 <b>92.39</b>	95.77 <b>96.03</b>	2.88 <b>2.62</b>	94.18 <b>95.43</b>	96.96 <b>97.65</b>	2.26 <b>1.22</b>	55.76 <b>60.86</b>	69.56 <b>74.69</b>	32.28 <b>9.23</b>
UNETR +PaR	84.74 <b>85.87</b>	90.70 <b>91.48</b>	4.63 <b>3.64</b>	95.23 <b>95.37</b>	97.55 <b>97.62</b>	2.26 <b>2.48</b>	91.36 <b>93.10</b>	95.44 <b>96.40</b>	3.01 <b>2.58</b>	94.12 <b>94.92</b>	96.92 <b>97.38</b>	4.32 <b>1.42</b>	58.26 <b>60.11</b>	72.89 <b>74.49</b>	8.94 <b>8.06</b>
Swin UNETR +PaR	88.28 <b>89.47</b>	93.23 <b>94.04</b>	3.25 <b>2.61</b>	96.42 <b>96.53</b>	98.17 <b>98.23</b>	1.63 <b>1.54</b>	93.03 <b>94.06</b>	96.36 <b>96.92</b>	4.39 <b>2.18</b>	95.81 <b>96.05</b>	97.85 <b>97.98</b>	1.28 <b>1.15</b>	67.85 <b>71.24</b>	80.55 <b>83.02</b>	5.72 <b>5.58</b>
nnFormer +PaR	85.50 <b>88.83</b>	91.43 <b>93.69</b>	5.41 <b>2.35</b>	94.99 <b>96.20</b>	97.42 <b>98.06</b>	2.10 <b>1.66</b>	89.69 <b>92.82</b>	94.52 <b>96.27</b>	3.16 <b>2.23</b>	94.84 <b>95.79</b>	97.35 <b>97.84</b>	<b>1.16</b> 1.17	62.47 <b>70.51</b>	76.43 <b>82.61</b>	15.24 <b>4.33</b>
TransBTS +PaR	87.63 <b>88.36</b>	92.84 <b>93.27</b>	3.54 <b>2.90</b>	96.27 <b>96.29</b>	98.09 <b>98.11</b>	<b>1.58</b> 1.69	92.15 <b>93.67</b>	95.89 <b>96.71</b>	3.87 <b>2.12</b>	95.39 <b>95.89</b>	97.63 <b>97.90</b>	1.20 <b>1.13</b>	66.70 <b>67.58</b>	79.76 <b>80.37</b>	7.52 <b>6.65</b>
MultiResUNet +PaR	85.92 <b>86.76</b>	91.35 <b>91.85</b>	9.04 <b>3.68</b>	95.22 <b>95.85</b>	97.54 <b>97.88</b>	3.50 <b>1.95</b>	93.07 <b>93.88</b>	96.39 <b>96.82</b>	2.63 <b>2.23</b>	93.95 <b>94.89</b>	96.84 <b>97.37</b>	20.07 <b>1.30</b>	61.43 <b>62.44</b>	74.63 <b>75.35</b>	9.97 <b>9.25</b>
3D UX-NET +PaR	88.40 <b>89.24</b>	93.31 <b>93.84</b>	8.85 <b>2.43</b>	96.31 <b>96.55</b>	98.12 <b>98.24</b>	11.58 <b>1.56</b>	93.76 <b>94.07</b>	96.76 <b>96.93</b>	2.35 <b>2.16</b>	95.90 <b>96.14</b>	97.90 <b>98.03</b>	1.19 <b>1.13</b>	67.62 <b>70.21</b>	80.46 <b>82.17</b>	20.31 <b>4.88</b>

Table 4: Detailed results for each category in the benchmarking experiments on the OIMHS dataset show the performance transformation of 11 previous segmentation models integrating PaR post-decoder. The metrics for all models with integrated PaR are highlighted in bold.

Method	Average			Macular Hole			Retinal			Macular Edema			Choroid		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
3D U-Net +PaR	86.60 <b>87.55</b>	92.49 <b>93.08</b>	3.40 <b>2.89</b>	76.87 <b>79.86</b>	86.79 <b>88.67</b>	8.06 <b>6.36</b>	97.55 <b>97.57</b>	98.75 <b>98.76</b>	<b>1.15</b> <b>1.15</b>	79.73 <b>80.10</b>	88.46 <b>88.71</b>	1.47 <b>1.34</b>	92.24 <b>92.67</b>	95.95 <b>96.18</b>	2.93 <b>2.71</b>
RAUNet +PaR	84.52 <b>86.20</b>	91.14 <b>92.25</b>	13.61 <b>5.37</b>	74.36 <b>78.44</b>	85.05 <b>87.76</b>	25.19 <b>15.27</b>	97.29 <b>97.43</b>	98.61 <b>98.69</b>	1.26 <b>1.20</b>	76.00 <b>78.08</b>	85.97 <b>87.39</b>	6.18 <b>1.90</b>	90.42 <b>90.84</b>	94.95 <b>95.18</b>	21.80 <b>3.13</b>
ResUNet +PaR	84.06 <b>87.07</b>	90.84 <b>92.79</b>	3.92 <b>3.33</b>	74.00 <b>78.89</b>	84.87 <b>88.06</b>	8.53 <b>7.62</b>	97.15 <b>97.45</b>	98.54 <b>98.70</b>	1.26 <b>1.20</b>	74.13 <b>80.05</b>	84.72 <b>88.66</b>	2.36 <b>1.49</b>	90.95 <b>91.87</b>	95.24 <b>95.74</b>	3.52 <b>3.00</b>
SegResNet +PaR	83.59 <b>84.82</b>	90.52 <b>91.35</b>	12.05 <b>5.07</b>	73.22 <b>75.77</b>	84.31 <b>86.02</b>	34.92 <b>8.04</b>	97.14 <b>97.18</b>	98.54 <b>98.56</b>	1.23 <b>1.20</b>	73.08 <b>75.18</b>	84.01 <b>85.48</b>	8.32 <b>7.44</b>	90.92 <b>91.14</b>	95.22 <b>95.34</b>	3.72 <b>3.61</b>
V-Net +PaR	81.00 <b>83.21</b>	88.53 <b>90.26</b>	18.13 <b>13.52</b>	65.21 <b>72.31</b>	77.88 <b>83.61</b>	59.79 <b>40.83</b>	96.84 <b>97.00</b>	98.38 <b>98.47</b>	1.32 <b>1.22</b>	71.52 <b>72.69</b>	82.93 <b>83.77</b>	7.56 <b>8.29</b>	90.41 <b>90.86</b>	94.94 <b>95.19</b>	3.85 <b>3.73</b>
UNETR +PaR	81.52 <b>83.73</b>	89.05 <b>90.55</b>	29.15 <b>7.24</b>	69.07 <b>74.31</b>	81.25 <b>84.94</b>	64.78 <b>16.19</b>	96.72 <b>97.08</b>	98.32 <b>98.51</b>	1.23 <b>1.18</b>	70.76 <b>73.46</b>	82.20 <b>84.01</b>	46.68 <b>7.91</b>	89.53 <b>90.05</b>	94.43 <b>94.73</b>	3.91 <b>3.69</b>
Swin UNETR +PaR	87.11 <b>87.92</b>	92.82 <b>93.27</b>	5.21 <b>2.92</b>	78.64 <b>80.46</b>	87.91 <b>89.02</b>	15.02 <b>6.54</b>	97.59 <b>97.60</b>	<b>98.78</b> <b>98.78</b>	1.15 <b>1.13</b>	79.88 <b>80.86</b>	88.59 <b>89.07</b>	1.79 <b>1.37</b>	92.31 <b>92.75</b>	95.99 <b>96.23</b>	2.87 <b>2.62</b>
nnFormer +PaR	80.54 <b>85.50</b>	88.29 <b>91.80</b>	25.32 <b>7.36</b>	65.97 <b>78.64</b>	78.77 <b>87.86</b>	77.19 <b>22.49</b>	96.97 <b>97.22</b>	98.45 <b>98.58</b>	1.20 <b>1.18</b>	69.56 <b>76.13</b>	81.42 <b>86.02</b>	18.96 <b>1.84</b>	89.66 <b>90.03</b>	94.51 <b>94.73</b>	<b>3.95</b> <b>3.95</b>
TransBTS +PaR	79.40 <b>83.68</b>	87.39 <b>90.55</b>	33.52 <b>21.00</b>	60.34 <b>70.63</b>	74.43 <b>82.43</b>	121.00 <b>77.84</b>	96.87 <b>97.23</b>	98.40 <b>98.59</b>	1.28 <b>1.15</b>	69.97 <b>75.69</b>	81.77 <b>85.82</b>	8.09 <b>1.57</b>	90.44 <b>91.16</b>	94.95 <b>95.35</b>	3.70 <b>3.42</b>
MultiResUNet +PaR	86.53 <b>88.22</b>	92.44 <b>93.49</b>	3.23 <b>2.85</b>	77.86 <b>80.21</b>	87.36 <b>88.91</b>	6.97 <b>6.35</b>	97.20 <b>97.70</b>	98.57 <b>98.83</b>	1.21 <b>1.15</b>	78.72 <b>81.86</b>	87.82 <b>89.80</b>	1.99 <b>1.29</b>	92.34 <b>93.12</b>	96.00 <b>96.42</b>	2.74 <b>2.61</b>
3D UX-NET +PaR	87.45 <b>88.51</b>	93.01 <b>93.66</b>	4.61 <b>2.65</b>	79.24 <b>81.42</b>	88.28 <b>89.64</b>	13.23 <b>5.92</b>	97.57 <b>97.69</b>	98.77 <b>98.83</b>	<b>1.15</b> <b>1.15</b>	80.22 <b>81.56</b>	88.77 <b>89.61</b>	1.42 <b>1.35</b>	92.76 <b>93.36</b>	96.23 <b>96.56</b>	2.62 <b>2.21</b>

Table 5: Detailed results for each category in the benchmarking experiments on the SegTHOR dataset show the performance transformation of 11 previous segmentation models integrating PaR post-decoder. The metrics for all models with integrated PaR are highlighted in bold.

Method	Average			Esophagus			Heart			Trachea			Aorta		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
3D U-Net <b>+PaR</b>	78.69 <b>82.08</b>	87.59 <b>89.82</b>	4.32 <b>2.98</b>	63.84 <b>67.54</b>	77.46 <b>80.38</b>	3.84 <b>3.78</b>	89.23 <b>90.39</b>	94.29 <b>94.94</b>	7.94 <b>2.93</b>	76.69 <b>83.19</b>	86.76 <b>90.80</b>	<b>3.00</b> <b>3.00</b>	85.01 <b>87.20</b>	91.86 <b>93.16</b>	2.51 <b>2.20</b>
RAUNet <b>+PaR</b>	79.34 <b>81.04</b>	88.13 <b>89.08</b>	14.58 <b>2.99</b>	<b>66.41</b> 64.85	<b>79.55</b> 78.27	<b>3.61</b> 3.82	88.90 <b>90.16</b>	94.10 <b>94.80</b>	25.27 <b>3.01</b>	76.59 <b>82.27</b>	86.71 <b>90.26</b>	<b>2.90</b> 2.99	85.47 <b>86.90</b>	92.16 <b>92.98</b>	26.53 <b>2.15</b>
ResUNet <b>+PaR</b>	79.61 <b>80.23</b>	88.26 <b>88.71</b>	3.22 <b>3.14</b>	66.34 <b>68.22</b>	79.44 <b>80.81</b>	3.93 <b>3.30</b>	88.32 <b>89.07</b>	93.74 <b>94.21</b>	3.50 <b>3.09</b>	76.36 <b>77.26</b>	86.57 <b>87.14</b>	<b>3.12</b> 3.55	<b>87.41</b> 86.37	<b>93.27</b> 92.68	<b>2.33</b> 2.60
SegResNet <b>+PaR</b>	77.78 <b>81.77</b>	86.99 <b>89.60</b>	3.38 <b>2.87</b>	62.21 <b>67.00</b>	76.28 <b>79.94</b>	4.29 <b>3.88</b>	88.53 <b>90.08</b>	93.89 <b>94.77</b>	3.23 <b>2.98</b>	74.65 <b>82.58</b>	85.46 <b>90.44</b>	3.66 <b>2.38</b>	85.73 <b>87.40</b>	92.31 <b>93.27</b>	2.37 <b>2.23</b>
V-Net <b>+PaR</b>	75.17 <b>76.23</b>	85.12 <b>85.81</b>	16.10 <b>11.49</b>	<b>60.46</b> 59.42	<b>74.90</b> 73.90	<b>4.61</b> 4.77	83.40 <b>88.23</b>	90.34 <b>93.72</b>	52.05 <b>34.62</b>	71.23 <b>71.90</b>	83.05 <b>83.54</b>	5.11 <b>4.16</b>	<b>85.57</b> 85.38	<b>92.21</b> 92.09	2.63 <b>2.39</b>
UNETR <b>+PaR</b>	73.76 <b>74.13</b>	84.03 <b>84.18</b>	4.71 <b>4.65</b>	<b>54.21</b> 53.41	<b>69.71</b> 68.86	6.00 <b>5.69</b>	87.83 <b>87.99</b>	93.50 <b>93.59</b>	<b>3.59</b> 3.70	69.88 <b>70.46</b>	82.14 <b>82.58</b>	<b>6.28</b> 6.52	83.13 <b>84.67</b>	90.76 <b>91.68</b>	2.95 <b>2.71</b>
Swin UNETR <b>+PaR</b>	78.19 <b>78.51</b>	87.26 <b>87.42</b>	3.87 <b>3.62</b>	<b>63.12</b> 62.18	<b>77.03</b> 76.29	<b>4.41</b> 4.58	<b>88.94</b> 88.74	<b>94.13</b> 94.01	<b>3.18</b> 3.31	73.79 <b>75.59</b>	84.89 <b>86.04</b>	5.66 <b>4.44</b>	86.90 <b>87.55</b>	92.98 <b>93.35</b>	2.24 <b>2.18</b>
nnFormer <b>+PaR</b>	77.27 <b>79.00</b>	86.65 <b>87.69</b>	5.11 <b>3.51</b>	<b>62.41</b> 61.58	<b>76.43</b> 75.66	<b>4.17</b> 5.27	87.80 <b>89.17</b>	93.48 <b>94.24</b>	10.06 <b>3.43</b>	72.70 <b>78.88</b>	84.14 <b>88.18</b>	3.85 <b>2.92</b>	86.17 <b>86.37</b>	92.55 <b>92.68</b>	<b>2.36</b> 2.40
TransBTS <b>+PaR</b>	77.70 <b>81.02</b>	86.88 <b>89.13</b>	3.84 <b>3.75</b>	61.47 <b>66.73</b>	75.63 <b>79.71</b>	4.12 <b>3.52</b>	88.60 <b>89.37</b>	93.93 <b>94.36</b>	<b>3.05</b> 3.15	74.79 <b>80.71</b>	85.50 <b>89.27</b>	<b>5.97</b> 6.14	85.97 <b>87.28</b>	92.44 <b>93.20</b>	2.24 <b>2.18</b>
MultiResUNet <b>+PaR</b>	79.87 <b>80.81</b>	88.53 <b>89.10</b>	26.75 <b>11.06</b>	67.41 <b>68.89</b>	80.38 <b>81.34</b>	7.18 <b>3.25</b>	87.54 <b>88.36</b>	93.31 <b>93.79</b>	<b>28.33</b> 35.08	<b>79.25</b> 78.63	<b>88.40</b> 88.02	<b>3.52</b> 3.54	85.29 <b>87.35</b>	92.05 <b>93.24</b>	67.98 <b>2.38</b>
3D UX-NET <b>+PaR</b>	78.30 <b>79.01</b>	87.34 <b>87.77</b>	4.69 <b>3.61</b>	63.18 <b>64.14</b>	77.12 <b>77.78</b>	<b>3.97</b> 4.25	87.81 <b>89.42</b>	93.47 <b>94.40</b>	7.38 <b>3.30</b>	<b>74.92</b> 74.69	<b>85.57</b> 85.44	5.05 <b>4.58</b>	87.29 <b>87.78</b>	93.20 <b>93.48</b>	2.35 <b>2.28</b>

Table 6: Details of the ablation study on the initialization of parameters W and b on the FLARE2021 dataset.

W	b	Average			Liver			Kidney			Spleen			Pancreas		
		IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
0	0	88.57	93.49	5.64	<b>96.39</b>	<b>98.16</b>	11.06	93.38	96.56	2.33	95.47	97.67	3.35	69.03	81.58	5.84
0	1	88.70	93.57	4.11	96.35	98.13	<b>1.61</b>	93.47	96.61	2.31	95.48	97.68	7.63	69.51	81.87	4.88
<b>1</b>	<b>0</b>	<b>89.20</b>	<b>93.89</b>	<b>2.49</b>	96.36	98.14	1.67	<b>93.64</b>	<b>96.70</b>	2.30	95.61	97.74	1.33	<b>71.19</b>	<b>82.96</b>	<b>4.67</b>
1	1	88.76	93.56	2.52	96.37	98.15	1.58	93.49	96.62	<b>2.21</b>	<b>95.85</b>	<b>97.88</b>	<b>1.13</b>	69.31	81.60	5.17

Table 7: Details on the ablation study of groups in spatial redistribution on the FLARE2021 dataset.

Groups	Average			Liver			Kidney			Spleen			Pancreas		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
4	88.73	93.57	2.70	96.23	98.07	1.71	93.28	96.50	2.36	95.95	<b>97.93</b>	1.18	69.47	81.79	5.56
<b>8</b>	<b>89.20</b>	<b>93.89</b>	<b>2.49</b>	96.36	98.14	<b>1.67</b>	93.64	96.70	2.30	95.61	97.74	1.33	<b>71.19</b>	<b>82.96</b>	<b>4.67</b>
16	88.97	93.72	7.00	96.29	98.10	9.46	<b>93.73</b>	<b>96.74</b>	<b>2.20</b>	<b>95.96</b>	<b>97.93</b>	11.53	69.89	82.11	4.79
24	88.98	93.74	2.50	<b>96.38</b>	<b>98.15</b>	1.69	93.56	96.65	2.37	95.94	97.92	<b>1.17</b>	70.05	82.24	4.78

Table 8: Details on the ablation study of key components and fusion strategies on the FLARE2021 dataset.

Method	Average			Liver			Kidney			Spleen			Pancreas		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
3D U-Net	87.92	93.08	16.31	95.63	97.75	9.14	92.16	95.89	17.05	95.33	97.60	21.11	68.55	81.08	17.96
+PDI	88.92	93.73	10.72	96.05	97.98	21.35	93.22	96.47	15.34	95.92	97.91	1.17	70.51	82.57	5.03
+SAI	88.45	93.49	20.33	95.91	97.90	22.37	92.80	96.25	2.36	94.30	97.04	51.70	70.79	82.75	4.90
+PaR-cat	88.67	93.56	8.24	95.67	97.76	13.58	<b>93.68</b>	<b>96.71</b>	<b>2.26</b>	95.32	97.59	12.39	69.99	82.19	4.73
+PaR-mult	88.58	93.48	13.37	94.45	97.04	44.64	93.60	96.68	2.37	<b>96.07</b>	<b>97.99</b>	<b>1.16</b>	70.18	82.22	5.30
<b>+PaR-add</b>	<b>89.20</b>	<b>93.89</b>	<b>2.49</b>	<b>96.36</b>	<b>98.14</b>	<b>1.67</b>	93.64	96.70	2.30	95.61	97.74	1.33	<b>71.19</b>	<b>82.96</b>	<b>4.67</b>

Table 9: Details on the ablation study of key components and fusion strategies on the OIMHS dataset.

Method	Average			Macular Hole			Retinal			Macular Edema			Choroid		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
UNETR	81.52	89.05	29.15	69.07	81.25	64.78	96.72	98.32	1.23	70.76	82.20	46.68	89.53	94.43	3.91
+PDI	82.88	89.98	22.59	71.25	82.82	77.97	<b>97.12</b>	<b>98.53</b>	1.18	73.33	83.96	7.72	89.81	94.59	<b>3.51</b>
+SAI	81.88	89.21	13.35	69.35	81.19	27.54	96.99	98.47	<b>1.16</b>	71.35	82.59	20.69	89.81	94.59	4.01
+PaR-cat	82.61	89.77	16.87	72.32	83.53	58.80	96.91	98.43	1.21	71.80	82.77	<b>3.25</b>	89.41	94.37	4.21
+PaR-mult	82.07	89.42	27.31	68.05	80.34	93.50	96.89	98.41	1.19	<b>73.65</b>	<b>84.38</b>	10.63	89.70	94.54	3.93
<b>+PaR-add</b>	<b>83.73</b>	<b>90.55</b>	<b>7.24</b>	<b>74.31</b>	<b>84.94</b>	<b>16.19</b>	97.08	98.51	1.18	73.46	84.01	7.91	<b>90.05</b>	<b>94.73</b>	3.69

Table 10: Details on the ablation study of different frame sampling scenarios on the FLARE2021 dataset.

Method	Average			Liver			Kidney			Spleen			Pancreas		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
3D U-Net	87.92	93.08	16.31	95.63	97.75	9.14	92.16	95.89	17.05	95.33	97.60	21.11	68.55	81.08	17.96
<b>+PaR</b>	<b>89.20</b>	<b>93.89</b>	<b>2.49</b>	<b>96.36</b>	<b>98.14</b>	<b>1.67</b>	<b>93.64</b>	<b>96.70</b>	<b>2.30</b>	<b>95.61</b>	<b>97.74</b>	<b>1.33</b>	<b>71.19</b>	<b>82.96</b>	<b>4.67</b>
3D U-Net(1/3)	86.27	91.76	14.52	95.69	97.79	<b>13.92</b>	92.13	95.81	<b>2.48</b>	94.32	97.03	14.97	62.96	76.40	26.73
<b>+PaR</b>	<b>87.55</b>	<b>92.64</b>	<b>8.55</b>	<b>95.74</b>	<b>97.81</b>	19.16	<b>93.73</b>	<b>96.74</b>	3.40	<b>95.20</b>	<b>97.52</b>	<b>4.28</b>	<b>65.53</b>	<b>78.51</b>	<b>7.35</b>
3D U-Net(1/2)	83.25	89.30	16.28	95.10	97.47	18.07	88.17	93.37	3.97	92.80	96.07	2.65	56.92	70.29	40.43
<b>+PaR</b>	<b>85.29</b>	<b>90.80</b>	<b>7.10</b>	<b>95.81</b>	<b>97.85</b>	<b>3.44</b>	<b>92.07</b>	<b>95.77</b>	<b>2.71</b>	<b>94.66</b>	<b>97.23</b>	<b>1.72</b>	<b>58.62</b>	<b>72.33</b>	<b>20.55</b>

Table 11: Details on the ablation study of disentanglement along different axes on the FLARE2021 dataset.

Method	Average			Liver			Kidney			Spleen			Pancreas		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
3D U-Net	87.92	93.08	16.31	95.63	97.75	9.14	92.16	95.89	17.05	95.33	97.60	21.11	68.55	81.08	17.96
+PaR(W)	<b>89.20</b>	<b>93.89</b>	<b>2.49</b>	96.36	98.14	1.67	93.64	96.70	2.30	95.61	97.74	1.33	<b>71.19</b>	<b>82.96</b>	<b>4.67</b>
+PaR(H)	88.50	93.47	8.16	95.24	97.52	24.27	93.52	96.64	2.32	95.68	97.79	1.18	69.54	81.92	4.88
+PaR(D)	89.10	93.78	<b>2.49</b>	<b>96.52</b>	<b>98.22</b>	<b>1.57</b>	<b>93.80</b>	<b>96.78</b>	<b>2.18</b>	<b>96.05</b>	<b>97.98</b>	<b>1.13</b>	70.02	82.12	5.10

Table 12: Details on the ablation study of disentanglement along different axes on the OIMHS dataset.

Method	Average			Macular Hole			Retinal			Macular Edema			Choroid		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
3D U-Net	86.60	92.49	3.40	76.87	86.79	8.06	97.55	98.75	1.15	79.73	88.46	1.47	92.24	95.95	2.93
+PaR(W)	<b>87.55</b>	<b>93.08</b>	<b>2.89</b>	<b>79.86</b>	<b>88.67</b>	<b>6.36</b>	<b>97.57</b>	<b>98.76</b>	1.15	<b>80.10</b>	<b>88.71</b>	<b>1.34</b>	92.67	96.18	<b>2.71</b>
+PaR(H)	86.96	92.72	5.12	78.91	88.06	14.86	97.48	98.72	<b>1.13</b>	79.56	88.36	1.54	91.87	95.75	2.95
+PaR(D)	87.44	93.01	2.92	79.51	88.44	6.49	97.56	<b>98.76</b>	1.15	79.95	88.61	<b>1.34</b>	<b>92.76</b>	<b>96.23</b>	<b>2.71</b>

Table 13: Details on the ablation study of self-attention heads in PDI pipeline on the FLARE2021 dataset.

Heads	Average			Liver			Kidney			Spleen			Pancreas		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
4	88.62	93.44	4.19	96.29	98.10	8.41	<b>93.64</b>	<b>96.70</b>	<b>2.19</b>	<b>95.63</b>	<b>97.76</b>	<b>1.20</b>	68.94	81.20	4.98
<b>8</b>	<b>89.20</b>	<b>93.89</b>	<b>2.49</b>	<b>96.36</b>	<b>98.14</b>	1.67	<b>93.64</b>	<b>96.70</b>	2.30	95.61	97.74	1.33	<b>71.19</b>	<b>82.96</b>	<b>4.67</b>
16	88.56	93.42	2.63	96.32	98.12	<b>1.62</b>	93.57	96.66	2.23	95.61	97.75	1.24	68.75	81.14	5.42
24	88.34	93.34	4.75	95.92	97.91	9.73	93.30	96.52	2.44	95.17	97.51	1.30	68.98	81.40	5.54

Table 14: Details on the ablation study of kernel sizes in the preliminary feature extraction layer on the FLARE2021 dataset.

Ks	Average			Liver			Kidney			Spleen			Pancreas		
	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95	IoU	Dice	HD95
3	<b>89.20</b>	<b>93.89</b>	5.06	96.29	98.10	12.19	93.70	96.73	2.24	95.64	97.76	1.26	71.17	<b>82.97</b>	<b>4.56</b>
5	88.98	93.73	10.47	96.34	98.13	10.51	93.39	96.56	14.68	95.97	<b>97.94</b>	11.72	70.20	82.28	4.98
<b>7</b>	<b>89.20</b>	<b>93.89</b>	<b>2.49</b>	96.36	98.14	1.67	93.64	96.70	2.30	95.61	97.74	1.33	<b>71.19</b>	82.96	4.67
9	89.05	93.78	5.35	96.49	98.21	1.61	93.02	96.36	13.75	<b>95.98</b>	<b>97.94</b>	<b>1.13</b>	70.72	82.62	4.90
13	89.06	93.77	5.28	<b>96.54</b>	<b>98.24</b>	<b>1.60</b>	<b>93.76</b>	<b>96.76</b>	<b>2.17</b>	95.85	97.87	12.11	70.11	82.20	5.24


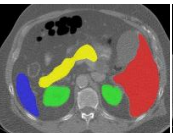
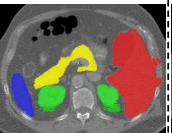
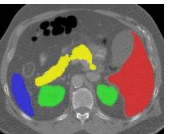
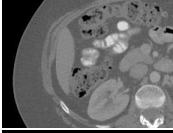
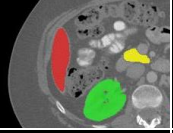
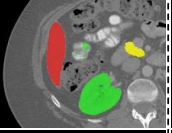

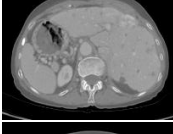
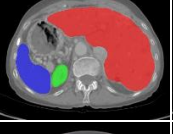
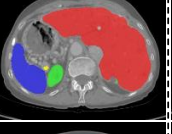
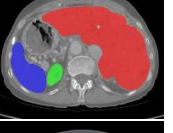
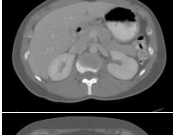
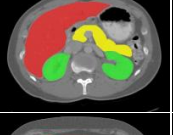
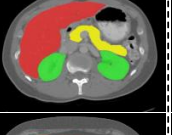
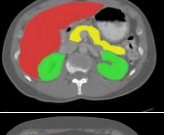
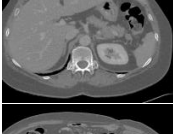
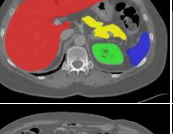
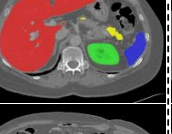
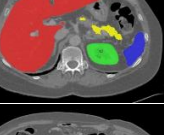
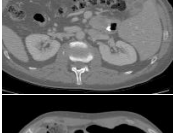
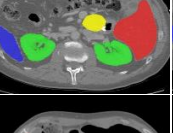
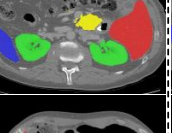
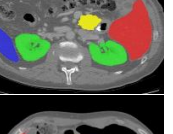
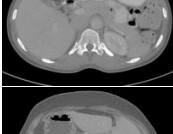
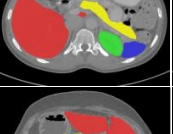
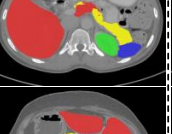
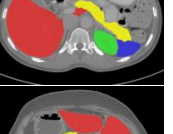
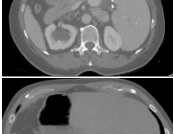
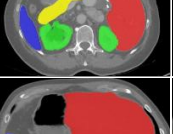
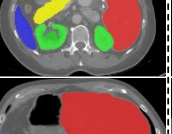
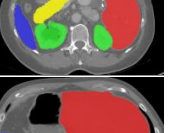
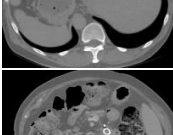
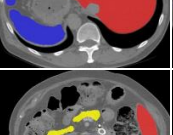
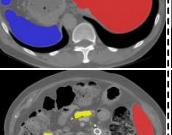
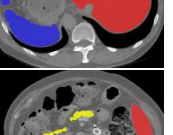
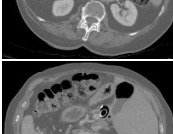
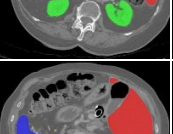
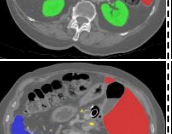
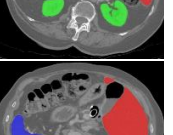
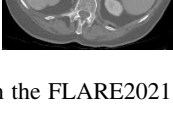
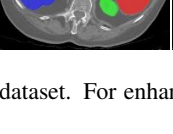
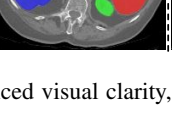

	Original Image	GT	Backbone	+PaR
3D U-Net				
RAUNet				
ResUNet				
SegResNet				
V-Net				
UNETR				
Swin UNETR				
nnFormer				
TransBTS				
MultiResUNet				
3D UX-Net				

Figure 1: Additional qualitative results on the FLARE2021 dataset. For enhanced visual clarity, the displayed images have been cropped. Please kindly zoom in for a better view.



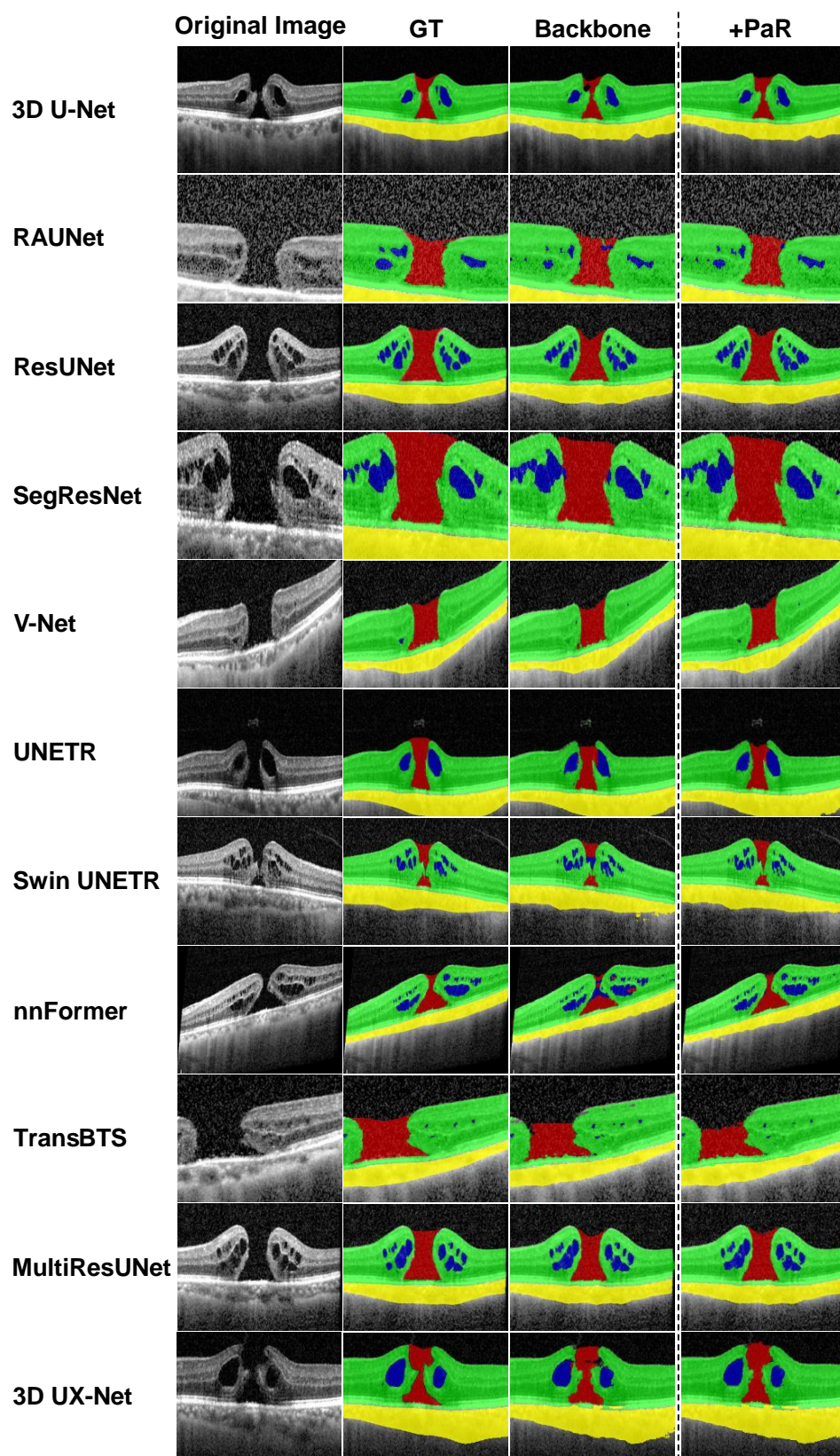


Figure 2: Additional qualitative results on the OIMHS dataset. For enhanced visual clarity, the displayed images have been cropped. Please kindly zoom in for a better view.

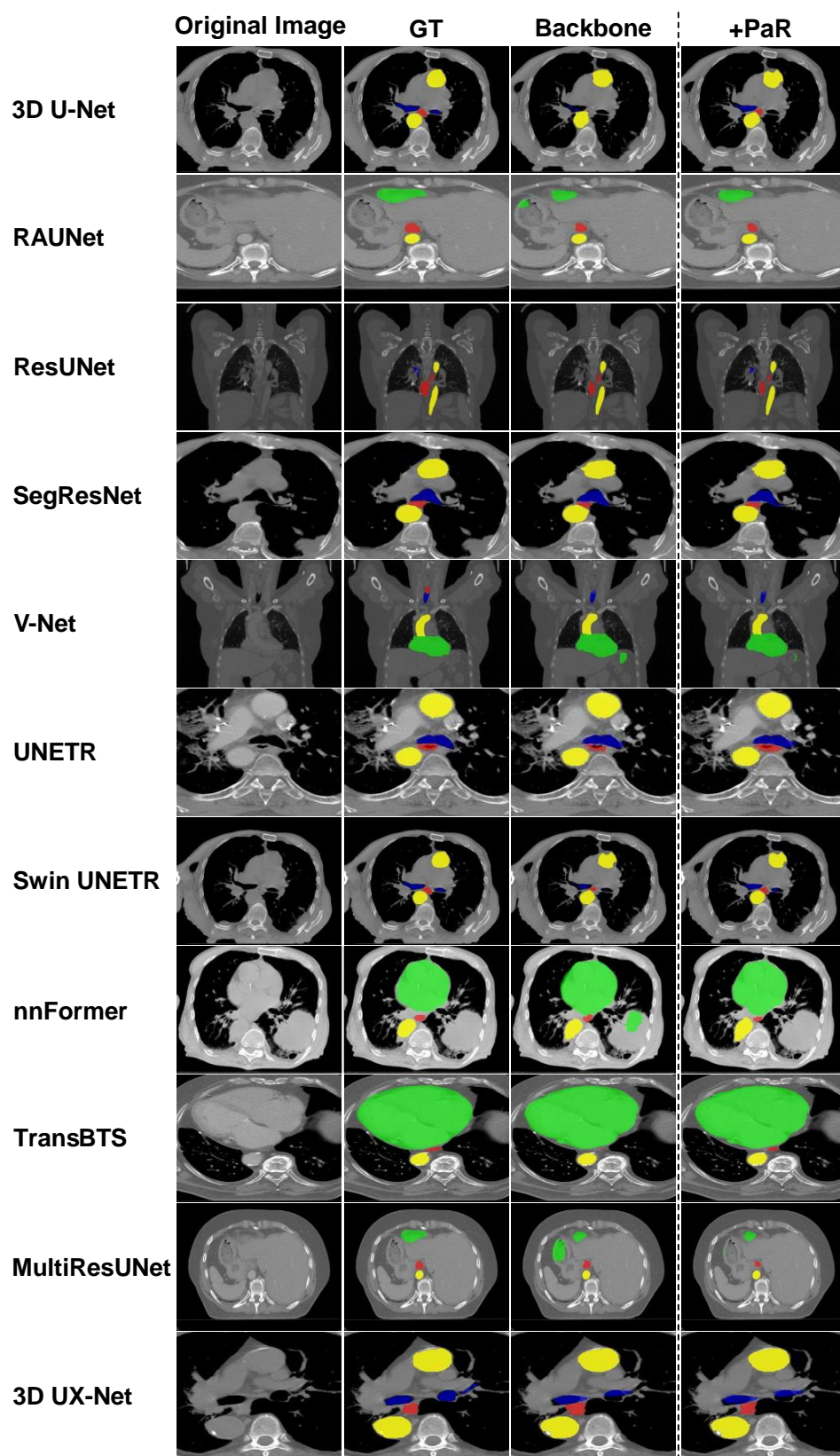


Figure 3: Additional qualitative results on the SegTHOR dataset. For enhanced visual clarity, the displayed images have been cropped. Please kindly zoom in for a better view.