CROWn: A Unified Framework for Anti-Aliased Downsampling and Phase-Calibrated Fusion in 3D Medical Segmentation

Supplementary Material

1. Datasets

To validate the effectiveness of the proposed method, we conducted experiments on 15 publicly available datasets. Descriptions of these datasets are provided below and related summary in Table S1.

1) SegTHOR: SegTHOR is a thoracic CT benchmark focused on four clinically critical OARs: the heart, trachea, aorta, and esophagus. Each volume is 3D with an in-plane spacing of 0.90×0.90–1.37×1.37 mm², a slice thickness of 2.0–2.5 mm, and a matrix size of 512×512 with 147–284 axial slices. Voxel-wise annotations are provided for all four structures. In our experiments, we use all 40 publicly available labeled cases. The dataset presents notable challenges, heterogeneous appearance across organs, class imbalance, and low-contrast boundaries (especially for the esophagus), making it a strong testbed for robust 3D medical image segmentation methods.

2) FLARE 2022: FLARE 2022 is a multi-center benchmark for 3D abdominal organ segmentation from CT, covering 13 anatomies, including Liver, Right Kidney, Spleen, Pancreas, Aorta, Inferior Vena Cava, Right Adrenal Gland, Gall Bladder, Esophagus, Stomach, Duodenum and Left Kidney, designed to stress-test robustness under strong class imbalance and fuzzy boundaries. The volumes exhibit realistic clinical heterogeneity, with in-plane resolution spanning 0.64×0.64–0.77×0.77 mm², slice thickness of 2.5–5.0 mm, and a matrix of 512×512 with 71–113 axial slices per scan. Our experiments adopt all 50 publicly released, expert-annotated cases for training and validation, enabling fair comparison while reflecting the challenge's emphasis on accurate small-structure delineation (e.g., adrenal glands, duodenum) alongside large parenchymal organs.

3) SKI10: SKI10 is a public 3D knee MRI benchmark for joint tissue segmentation, providing 100 fully annotated volumes with voxel-wise labels for four structures: femur bone, femoral cartilage, tibia bone, and tibial cartilage. The scans feature anisotropic spacing with an in-plane resolution of 0.4×0.4 mm and 1.0 mm slice thickness (0.4×0.4×1.0 mm³). In our experiments, we use all 100 publicly labeled cases to enable a fair and statistically robust evaluation of segmentation methods on clinically relevant bone–cartilage delineation.

4) FeTA 2021: FeTA 2021 is a curated 3D fetal brain MRI segmentation benchmark targeting tissue-level parcellation under real-world fetal imaging conditions. It provides voxel-wise labels for seven anatomies about external cerebrospinal fluid, gray matter, white matter, ventricles, cerebellum, deep gray matter, and brainstem—on volumes standardized to 256×256×256. Image spacing spans 0.43×0.43–1.00×1.00 mm² in-plane with 0.43–1.00 mm through-plane thickness, exposing models to clinically realistic resolution variability. In our experiments, we use all 80 publicly available labeled subjects.

5) OIMHS: OIMHS is a 3D optical coherence tomography (OCT) segmentation dataset targeting macular pathology, with voxel-wise annotations for four clinically relevant structures: retina, macular hole, macular edema, and choroid. Volumes are sized 512×512×{19–79} with anisotropic spacing, spanning in-plane resolutions of 10.07×10.07 μm²–14.0×14.0 μm² and slice thicknesses of 7.0–40.0 μm, offering substantial diversity in sampling density and morphology. The benchmark uses all 125 publicly released, expert-labeled cases, enabling rigorous evaluation of 3D segmentation methods under real-world variability in pathology, scale, and contrast.

6) MSD BRATS Tumour: MSD Brain Tumour (Task01 of the Medical Segmentation Decathlon) is a multiparametric MRI benchmark for 3D glioma segmentation, requiring delineation of three subregions: enhancing tumour (ET), peritumoral oedema (ED), and necrotic core (NC). Each subject includes co-registered native T1-weighted, post-Gadolinium T1-weighted (T1-Gd), T2-weighted, and T2-FLAIR volumes, standardized to 1.0×1.0×1.0 mm³ voxel spacing with a canonical matrix of 240×240×155. Our experiments use all 484 publicly available cases.

7) MSD Lung Tumour: The MSD Lung Tumour dataset (Task06 of the Medical Segmentation Decathlon) targets automatic delineation of lung tumors in thoracic CT. It provides volumetric scans with a single anatomical label about Lung Tumour, covering in-plane resolutions from 0.60×0.60 to 0.98×0.98 mm² and slice thicknesses from 0.63 to 2.5 mm. Volumes are 512×512 with axial depth varying from 112 to 636 slices. In our experiments, we use all 63 publicly available cases.

8) MSD Pancreas Tumour: MSD Pancreas Tumour (Task07 of the Medical Segmentation Decathlon) targets fully supervised segmentation of the pancreas and associated lesions in contrast-enhanced abdominal CT, with two anatomical labels: Pancreas and PancreasTumour. The cohort exhibits substantial acquisition heterogeneity, its in-plane resolution spans 0.61×0.61–0.98×0.98 mm², slice thickness ranges 0.70–7.5 mm, and volumes are standardized to 512×512×{37–751} voxels, posing strong chal-

lenges for scale robustness and small, low-contrast tumour delineation. We utilize all 281 publicly available annotated volumes for experimentation.

9) MSD Spleen: The MSD Spleen dataset (Task09 of the Medical Segmentation Decathlon) is a 3D abdominal CT benchmark for single-organ segmentation, providing annotations for the spleen alone (label set: Spleen). It comprises 41 publicly available labeled volumes with pronounced acquisition heterogeneity: in-plane spacing ranges from 0.61 × 0.61mm²- 0.98 × 0.98mm², slice thickness from 1.5–8.0 mm, and spatial dimensions of 512×512×{31–168}. We utilize all 41 publicly available annotated volumes for experimentation.

10) MSD Colon Cancer: MSD Colon Cancer (Task 10 of the Medical Segmentation Decathlon) targets delineation of colon tumors from abdominal CT volumes. The dataset contains 126 publicly available cases with a single anatomical label about Colon Cancer. Images have a fixed in-plane matrix of 512×512 with variable depth (37–729 slices), in-plane spacing ranging from 0.54×0.54 mm² to 0.98×0.98 mm², and slice thickness from 1.3–7.5 mm, introducing notable anisotropy and inter-scan heterogeneity. In our experiments, we utilize all 126 labeled cases.

11) HOCMvalvesSeg: HOCMvalvesSeg is a 3D cardiac CT dataset tailored for multi-class segmentation in hypertrophic obstructive cardiomyopathy (HOCM) surgical planning, providing expert voxel-level labels for seven anatomies: aortic valve (AV), mitral valve (MV), aorta (AO), left atrium (LA), left ventricle (LV), myocardium, and excised myocardium. It comprises 27 patient volumes (512×512×{275–571} voxels) with a typical voxel size of 0.25×0.25×0.5 mm³, enabling precise modeling of small, clinically critical valve structures and their anatomical context. We utilize all 27 publicly available annotated volumes for experimentation.

12) CrossMoDA2021: CrossMoDA2021 is a 3D brain MRI segmentation dataset targeting two anatomies about vestibular schwannoma and the cochlea, using 105 publicly available, expertly annotated T1-contrast–enhanced (T1-CE) volumes with corresponding masks. Each study exhibits in-plane resolution (0.41×0.41 mm²) with 1.0–1.5 mm slice thickness and a canonical volume size of 512×512×{120–160}, yielding realistic through-plane anisotropy. The combination of small, clinically critical targets and heterogeneous spacing imposes strict boundary-precision and robustness requirements, making CrossMoDA2021 a compact yet demanding benchmark for high-fidelity 3D medical image segmentation across modern architectures.

13) SLIVER07: SLIVER07 is a classic benchmark for 3D liver segmentation in contrast-enhanced CT. It provides 20 publicly available volumes with manual liver annotations, covering a wide range of acquisition settings: in-plane spacing 0.58–0.81 mm, slice thickness 0.70–5.0 mm, and volumes sized 512×512×{64–394}. The dataset's heterogeneity (non-isotropic spacing, variable coverage, and contrast differences) makes it a stringent testbed for cross-scanner generalization and robust 3D modeling. Unless otherwise stated, we follow the common protocol and use all 20 labeled cases for training and evaluation.

14) COVID-19 CT Seg: COVID-19 CT Seg is a 3D chest CT dataset for semantic segmentation of COVID-19 pathology, providing voxelwise annotations for three anatomical labels: left lung, right lung, and infection (pneumonia) regions. Volumes exhibit clinical heterogeneity, with in-plane spacing ranging from 0.58–0.81 mm and slice thickness from 1.0–6.0 mm, and image matrices spanning 512×512 to 630×630, covering an axial extent of 39–418 mm. In our experiments, we used all 20 publicly available subjects to enable reproducible benchmarking of lung and lesion segmentation under varying acquisition protocols and resolutions.

15) FLARE 2021: FLARE 2021 is a 3D abdominal CT benchmark targeting automatic segmentation of four key organs about liver, kidneys, spleen, and pancreas. It provides axial volumes with a standardized in-plane matrix of 512×512 and variable depth (37–751 slices), with voxel spacings ranging from 0.61–0.98 mm in-plane and 0.5–7.5 mm through-plane, capturing real clinical variability. Anatomical labels include liver, kidney, spleen, and pancreas. In our experiments we used all 361 publicly available labeled cases.

2. Evaluation Metric

To validate the effectiveness and robustness of the proposed method, we evaluate 3D segmentations using Intersection-over-Union (IoU), Dice coefficient, Average Symmetric Surface Distance (ASSD), the 95th-percentile Hausdorff distance (HD95), and the Adjusted Rand Index (Adj-Rand). All distances are computed in physical coordinates using the dataset-specific voxel spacing; metrics are computed per volume and summarized as mean ± standard deviation. For multi-class tasks, we report per-class scores and the unweighted macro-average over foreground classes.

For class c on domain Ω with prediction $P_c \subset \Omega$ and ground truth $G_c \subset \Omega$, the overlap scores are

$$\mathrm{IoU}_c = \frac{|P_c \cap G_c|}{|P_c \cup G_c|} \tag{1}$$

$$\mathrm{Dice}_c = \frac{2|P_c \cap G_c|}{|P_c| + |G_c|}. \tag{2}$$

Boundary distances. All distances are computed in physical units using voxel spacing $\mathbf{s} = (s_x, s_y, s_z)^\top \in \mathbb{R}^3_{>0}$:

$$\mathrm{dist}_{\mathbf{s}}(x, S) = \min_{y \in S} \| (x - y) \odot \mathbf{s} \|_2 . \tag{3}$$

Table S1. Summary of dataset statistics and preprocessing for 15 public 3D benchmarks.

Dataset	SegTHOR	FLARE 2022	SKI10
Strength Cutting Range	[-1000,300]	[-1000,300]	[0,3000]
Anatomical Label	Esophagus, Heart, Trachea, Aorta	Liver, Right/Left Kidney, Spleen, Pancreas, Aorta, Inferior Vena Cava, Right/Left Adrenal Gland, Gall Bladder, Esophagus, Stomach, Duodenum	Femur Bone, Femoral Cartilage, Tibia Bone, Tibial Cartilage
Dimensions	512 × 512 × {147-395}	512 × 512 × {71-113}	–
Resolution	{0.90-1.37}mm × {0.90-1.37}mm × {2.0-2.5}mm	{0.64-0.77}mm × {0.64-0.77}mm × {2.5-5.0}mm	0.4mm × 0.4mm × 1.0mm
Sample Size	40	50	100
Train Set / Test Set	32 / 8	40 / 10	60 / 40

Dataset	FeTA 2021	OIMHS	MSD BRATS Tumour
Strength Cutting Range	[0,1000]	[0,300]	[0,3000]
Anatomical Label	External Cerebrospinal Fluid, Grey Matter, White Matter, Ventriculus, Cerebellum, Deep Grey Matter, Brainstem	Macular Hole, Retina, Macular Edema, Choroid	Enhancing Tumor, Peritumoral Edema, Necrotic Core
Dimensions	256 × 256 × 256	512 × 512 × {19-79}	240×240×155
Resolution	{0.43-1.00}mm × {0.43-1.00}mm × {0.43-1.00}mm	{10.7-14.0}μm × {10.7-14.0}μm × {7.0-40.0}μm	1.0mm × 1.0mm × 1.0mm
Sample Size	80	125	484
Train Set / Test Set	62 / 18	100 / 25	411 / 73

Dataset	MSD Lung Tumour	MSD Pancreas Tumour	MSD Spleen
Strength Cutting Range	[-1000,1000]	[-180,500]	[-150,230]
Anatomical Label	Lung Tumour	Pancreas, Pancreas Tumour	Spleen
Dimensions	512 × 512 × {112-636}	512 × 512 × {37-751}	512 × 512 × {31-168}
Resolution	{0.60-0.98}mm × {0.60-0.98}mm × {0.63-2.50}mm	{0.61-0.98}mm × {0.61-0.98}mm × {0.70-7.50}mm	{0.61-0.98}mm × {0.61-0.98}mm × {1.5-8.0}mm
Sample Size	63	281	41
Train Set / Test Set	50 / 13	243 / 38	33 / 8

Dataset	MSD Colon Cancer	HOCMvalvesSeg	CrossMoDA2021
Strength Cutting Range	[-200,230]	[0,255]	[0,1800]
Anatomical Label	Colon Cancer	Aortic Valve, Mitral Valve, Aorta, Left Atrium, Left Ventricle, Myocardium, Excised Myocardium	Vestibular Schwannoma, Cochlea
Dimensions	512 × 512 × {37-729}	512 × 512 × {275-571}	512 × 512 × {120-160}
Resolution	{0.54-0.98}mm × {0.54-0.98}mm × {1.3-7.5}mm	0.25mm × 0.25mm × 0.5mm	0.41mm × 0.41mm × {1.0-1.5}mm
Sample Size	126	27	105
Train Set / Test Set	106 / 20	22 / 5	84 / 21

Dataset	SLIVER07	COVID-19 CT Seg	FLARE 2021
Strength Cutting Range	[-200,400]	[-1025,300]	[-125,275]
Anatomical Label	Liver	Left Lung, Right Lung, Infections	Liver, Kidneys, Spleen, Pancreas
Dimensions	512 × 512 × {64-394}	{512-630}×{512-630}×{39-418}	512 × 512 × {37-151}
Resolution	{0.58-0.81}mm × {0.58-0.81}mm × {0.70-5.0}mm	{0.58-0.81}mm × {0.58-0.81}mm × {1.0-6.0}mm	{0.61-0.98}mm × {0.61-0.98}mm × {0.5-7.5}mm
Sample Size	20	20	361
Train Set / Test Set	16 / 4	16 / 4	342 / 19

Let $\mathcal{S}(P_c)$ and $\mathcal{S}(G_c)$ denote the voxelized surfaces of P_c and G_c. The average symmetric surface distance (ASSD) and the 95th-percentile Hausdorff distance (HD95) are

$$\mathrm{ASSD}_c = \frac{\sum_{p\in\mathcal{S}(P_c)} \mathrm{dist}_{\mathbf{s}}\big(p,\mathcal{S}(G_c)\big) + \sum_{g\in\mathcal{S}(G_c)} \mathrm{dist}_{\mathbf{s}}\big(g,\mathcal{S}(P_c)\big)}{|\mathcal{S}(P_c)| + |\mathcal{S}(G_c)|} \tag{4}$$

$$\mathrm{HD95}_c = \max\Big\{ Q_{0.95}\big(\{\mathrm{dist}_{\mathbf{s}}(p,\mathcal{S}(G_c))\}_{p\in\mathcal{S}(P_c)}\big), Q_{0.95}\big(\{\mathrm{dist}_{\mathbf{s}}(g,\mathcal{S}(P_c))\}_{g\in\mathcal{S}(G_c)}\big)\Big\}. \tag{5}$$

where $Q_{0.95}(\cdot)$ is the 95th percentile.

Adjusted Rand (Adj-Rand). Treating connected components (or instance IDs) as clusters, let $n_{ij} = |P_i \cap G_j|$, $a_i = \sum_j n_{ij}$, $b_j = \sum_i n_{ij}$, $n = \sum_{ij} n_{ij}$, and $\binom{x}{2} = x(x-1)/2$. The Adj-Rand is

$$AdjRand = \frac{\sum_{i,j}\binom{n_{ij}}{2} - \frac{\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}}{\binom{n}{2}}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \frac{\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}}{\binom{n}{2}}} \tag{6}$$

3. Additional Ablation Experiment Results

3.1. Effectiveness of μPCAD

3.1.1. Stage-wise placement on the encoder

We compare inserting μPCAD at a single encoder stage (tested stage by stage) versus deploying it at all encoder stages on the MSD Lung Tumour datasets. As shown in the Table S2.

Table S2. Stage-wise effect of μPCAD.

Method	MSD Lung Tumour				
	IoU	Dice	ASSD	HD95	Adj-Rand
Stage1	60.35±17.53	73.26±18.57	7.65±7.68	46.90±68.67	73.25±15.58
Stage2	59.52±18.70	72.21±20.86	5.51±7.58	23.08±47.96	72.20±20.86
Stage3	59.82±19.65	72.28±21.35	7.21±11.08	31.03±57.77	72.27±21.36
Stage4	59.01±17.70	72.15±18.83	3.00±4.24	16.08±32.67	72.14±18.84
All	60.46±16.90	73.63±16.56	3.51±4.50	14.66±32.83	73.61±16.58

3.1.2. Qualitative impact of μPCAD across backbones

Augmenting representative CNN, Transformer, ConvNeXt and Mamba baselines with μPCAD consistently improves overlap and boundary metrics across the MSD Pancreas Tumour and FLARE2022 datasets (Table S3). Across pancreatic-tumor and multi-organ CT examples, shown in Figure S1, inserting μPCAD into classical backbone produces segmentations that more sharper, more continuous boundaries with fewer leaks, holes and better preservation of small structures.

Table S3. Qualitative effect of μPCAD across backbones.

Method	MSD Pancreas				FLARE2022			
	IoU	Dice	ASSD	HD95	IoU	Dice	ASSD	HD95
3D U-Net [2]	54.14	67.99	4.44	26.36	80.17	87.87	2.40	10.98
+μPCAD	55.33	68.82	3.62	15.97	81.82	89.20	1.51	9.77
UNETR [4]	38.80	51.69	8.04	25.63	73.74	82.83	2.64	13.68
+μPCAD	50.98	64.48	3.38	11.85	79.85	87.66	2.34	9.77
3D UX-Net [5]	51.53	64.21	15.41	40.67	80.44	87.98	1.69	6.32
+μPCAD	56.91	70.11	7.97	29.71	83.10	90.06	0.95	4.15
MedNeXt [11]	53.59	67.25	2.91	11.02	78.57	86.83	1.75	6.55
+μPCAD	56.35	69.84	2.30	8.16	78.64	86.93	1.87	8.33
UNesT [15]	51.40	65.04	3.73	14.17	80.58	88.31	2.16	8.37
+μPCAD	52.74	65.61	2.96	11.38	82.25	89.46	1.01	4.19
SwinSMT [10]	49.29	62.67	10.84	31.93	78.63	86.65	3.56	15.83
+μPCAD	52.96	66.58	5.57	23.50	79.14	87.14	3.27	16.30
SegMamba [12]	54.85	67.69	5.22	17.67	81.58	88.86	1.89	9.21
+μPCAD	56.96	69.92	4.56	18.81	82.72	89.71	0.91	4.02

3.1.3. Versus classic modules

Additional visual comparison as shown in Figure S2.

3.2. Effectiveness of OCF

3.2.1. Stage-wise placement on the skip connections

As shown in Table S4, applying OCF to a single skip (tested per level) versus to all skips shows the latter is superior and more stable.

Table S4. Stage-wise effect of OCF.

Method	MSD Lung Tumour				
	IoU	Dice	ASSD	HD95	Adj-Rand
Stage1	60.25±16.51	73.54±16.27	4.26±7.02	19.41±46.74	73.52±16.29
Stage2	59.84±19.08	72.39±21.06	5.66±8.04	17.67±35.38	72.38±21.06
Stage3	59.58±15.34	73.32±14.11	7.85±9.75	46.86±69.37	73.30±14.13
Stage4	58.57±19.14	71.35±21.26	6.04±9.32	29.97±45.29	71.34±21.26
All	60.46±16.90	73.63±16.56	3.51±4.50	14.66±32.83	73.61±16.58

3.2.2. Qualitative impact of OCF across backbones

As show in Table S5 and Figure S3, adding OCF ablation on MSD Pancreas Tumour and FLARE2022 to diverse baselines on consistently improves overlap and reduces boundary errors.

3.2.3. Versus classic modules

Additional visual comparison as shown in Figure S4.

4. Efficiency Analysis

Table S6 compares model complexity and inference efficiency in terms of parameter count, FLOPs, and throughput, while Figures S5, S6, S7, and S8 further illustrate the

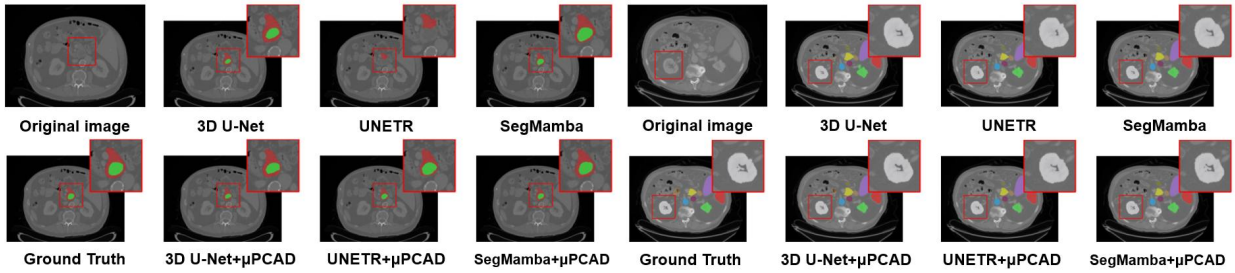

Figure S1. Qualitative effect of μPCAD as a plug-in decimator across backbones. Adding μPCAD systematically sharpens boundaries, suppresses alias-induced leakage and spurious fragments, and better preserves small structures under anisotropic sampling.

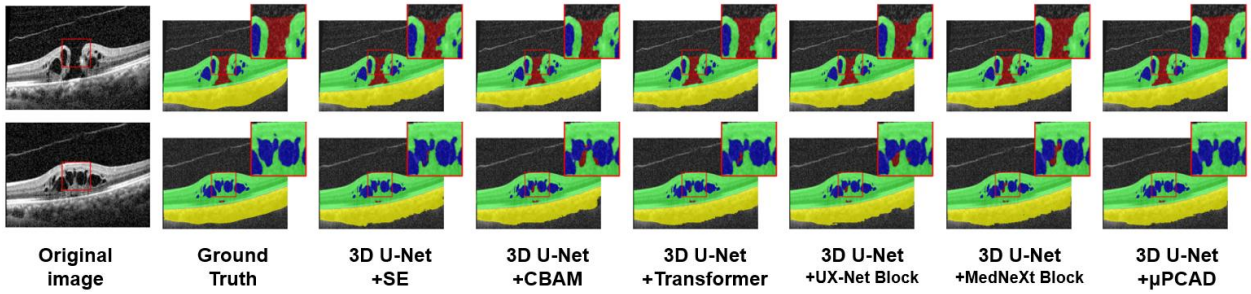

Figure S2. Additional qualitative comparison of μPCAD on the OIMHS dataset.

Table S5. Qualitative impact of OCF across backbones.

Method	MSD Pancreas				FLARE2022			
	IoU	Dice	ASSD	HD95	IoU	Dice	ASSD	HD95
3D U-Net [2]	54.14	67.99	4.44	26.36	80.17	87.87	2.40	10.98
+OCF	55.62	68.94	3.97	19.30	81.64	88.93	1.87	7.93
UNETR [4]	38.80	51.69	8.04	25.63	73.74	82.83	2.64	13.68
+OCF	44.85	57.18	8.41	27.29	76.24	84.78	2.50	9.78
3D UX-Net [5]	51.53	64.21	15.41	40.67	80.44	87.98	1.69	6.32
+OCF	53.86	67.53	3.91	17.21	81.63	88.93	1.09	4.45
MedNeXt [11]	53.49	67.25	2.91	11.02	78.57	86.83	1.75	6.55
+OCF	53.97	67.53	3.25	15.45	78.93	87.17	1.61	7.50
UNesT [15]	51.40	65.04	3.73	14.17	80.58	88.31	2.16	8.37
+OCF	54.39	67.77	2.91	11.41	82.25	89.46	1.01	4.19
SwinSMT [10]	49.29	62.67	10.84	31.93	78.63	86.65	3.56	15.83
+OCF	49.83	62.91	10.82	39.32	78.95	86.88	3.63	19.57
SegMamba [12]	54.85	67.69	5.22	17.67	81.58	88.86	1.89	9.21
+OCF	55.47	68.84	3.03	13.05	82.28	89.45	1.14	4.49

accuracy–efficiency trade-offs on the FLARE2022 dataset. FLOPs are computed with an input size of $96 \times 96 \times 96$, and throughput is measured on a single NVIDIA RTX 4090 GPU under the same input setting.

CROWn contains 23.78M parameters and requires 199.58G FLOPs, achieving an inference throughput of 27.70 samples/s. As shown in the four trade-off plots, CROWn strikes a favorable balance between segmentation accuracy and computational cost. In particular, CROWn achieves state-of-the-art Dice and the lowest HD95 among competing methods while maintaining a moderate model size and computational budget.

Compared with recent Transformer- and SSM-based methods, CROWn delivers substantially lower computational complexity while achieving better segmentation performance. Compared with lightweight CNN-based models, CROWn introduces only a moderate increase in computation but yields clear gains in accuracy and boundary quality. We attribute this favorable efficiency to the proposed alias-aware decimation and coset-fibrated skip calibration, which allocate computation more effectively to scale transitions and skip fusion, where accurate 3D boundary modeling is most critical, rather than uniformly increasing the backbone capacity.

5. Additional and Detailed Experiments Results

To facilitate reproducibility and fair cross-method comparison, we report the complete numerical results on all 15 public 3D datasets. For each dataset we provide five metrics about IoU, Dice, ASSD, HD95, and Adj-Rand, and scores are summarized as mean $\pm$ standard deviation over volumes. Across all datasets, CROWn achieves leading macro-averaged Dice and IoU, and attains the lowest or tied-lowest HD95 and ASSD on most cohorts. This aligns with the mechanism-level design: μPCAD performs alias-aware downsampling that preserves boundary-relevant high frequencies, while OCF phase-calibrates high-resolution skips prior to decoder fusion, reducing cross-scale misalignment. For ease of navigation, detailed Tables are

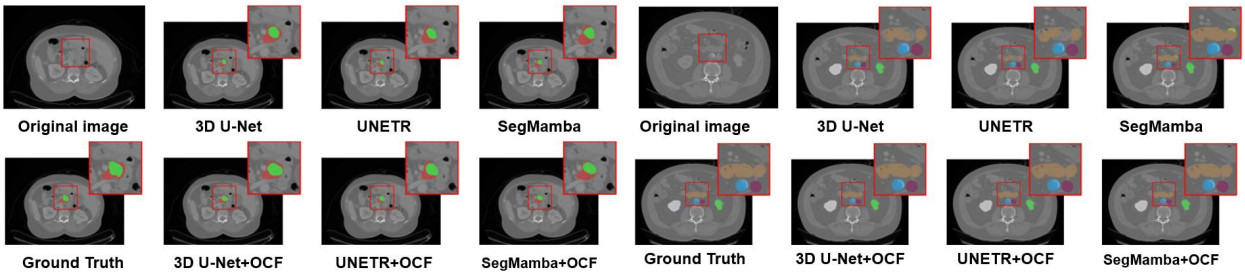

Figure S3. Qualitative impact of OCF across backbones. All backbones yield sharper, leak-free borders, fewer holes and spurious fragments, and better preservation of thin structures under with OCF, demonstrating anti-aliasing and phase-aligned cross-scale fusion.

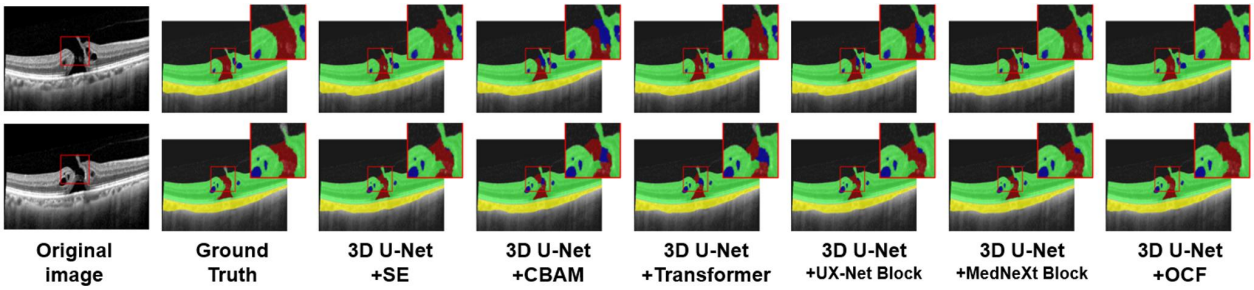

Figure S4. Additional qualitative comparison of OCF on the OIMHS dataset.

Table S6. Comparison of model complexity and inference efficiency in terms of parameters, FLOPs, and throughput.

Method	Source	Params ↓	FLOPs ↓	Throughput ↑
3D U-Net	MICCAI 2016	5.75 M	135.88 G	75.29 samples/s
V-Net	3DV 2016	45.61 M	331.10 G	38.65 samples/s
3D UX-Net	ICLR 2023	53.00 M	631.76 G	18.83 samples/s
UNETR	WACV 2022	92.62 M	82.58 G	58.96 samples/s
SwinUNETR	MICCAI 2022	61.99 M	329.46 G	20.35 samples/s
nnFormer	TIP 2023	149.10 M	224.36 G	52.60 samples/s
MedNeXt	MICCAI 2023	11.64 M	174.10 G	9.79 samples/s
SegFormer3D	CVPR 2024	4.49 M	4.97 G	231.09 samples/s
SegMamba	MICCAI 2024	64.24 M	655.62 G	19.04 samples/s
UNesT	MedIA 2023	84.88 M	258.11 G	39.78 samples/s
SwinSMT	MICCAI 2024	15.51 M	83.71 G	28.91 samples/s
VSmTrans	MedIA 2024	49.86 M	358.89 G	15.92 samples/s
SegTom	JBHI 2025	3.85 M	49.71 G	35.40 samples/s
nnWNet	CVPR 2025	52.23 M	648.26 G	11.97 samples/s
SuperLightNet	CVPR 2025	2.75 M	19.52 G	37.20 samples/s
DiffUNet	MedIA 2025	178.12 M	6333.85 G	13.62 samples/s
HiPaSNet	Nat.Com 2025	6.40 M	123.58 G	69.09 samples/s
CROWn (Ours)	–	23.78 M	199.58 G	27.70 samples/s

provided as SegTHOR (Table S7), FLARE2022 (Table (S8), SKI10 (Table S9), FeTA2021 (Table S10), OIMHS (Table S11), MSD BRATS Tumour (Table S12), MSD Lung Tumour (Table S13), MSD Pancreas Tumour (Table S14), MSD Spleen (Table S15), MSD Colon Cancer (Table S16), HOCMvalvesSeg (Table S17), CrossMoDA2021 (Table S18), SLIVER07 (Table S19), COVID-19 CT Seg (Table S20), FLARE2021 (Table S21).

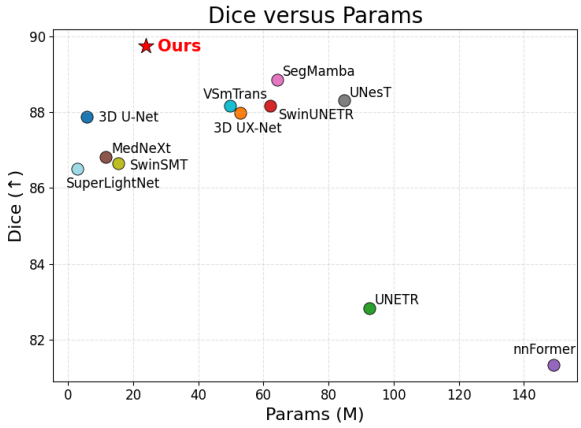

Figure S5. Dice is plotted against model size (Params, M) for representative 3D segmentation methods trained and evaluated under the same protocol.

Table S7. Results on the SegTHOR dataset. The best performance is highlighted in red , and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	83.46±3.27	90.66±2.20	1.65±1.09	2.87±0.71	90.62±2.21
V-Net [7]	3DV 2016	82.35±3.29	89.96±2.21	0.92±0.14	3.02±0.73	89.92±2.22
3D UX-Net [5]	ICLR 2023	83.36±2.85	90.66±1.87	0.91±0.11	3.21±0.77	90.61±1.87
UNETR [4]	WACV 2022	80.25±3.48	88.59±2.42	1.11±0.23	3.83±1.16	88.54±2.42
SwinUNETR [3]	MICCAI 2022	83.14±3.73	90.43±2.58	1.10±0.23	3.16±0.96	90.38±2.58
nnFormer [16]	TIP 2023	77.90±4.36	86.95±3.35	1.24±0.23	4.25±1.16	86.89±3.35
MedNeXt [11]	MICCAI 2023	81.82±3.38	89.59±2.36	1.08±0.29	3.90±1.73	89.54±2.36
SegFormer3D [9]	CVPR 2024	78.66±4.22	87.51±3.15	1.18±0.26	4.30±1.64	87.46±3.15
SegMamba [12]	MICCAI 2024	83.64±3.14	90.78±2.09	1.20±0.57	2.97±0.91	90.74±2.09
UNesT [15]	MedIA 2023	83.38±2.87	90.62±1.91	1.21±0.47	2.94±0.64	90.58±1.91
SwinSMT [10]	MICCAI 2024	82.94±3.37	90.36±2.27	1.06±0.22	3.24±0.93	90.31±2.28
VSmTrans [6]	MedIA 2024	82.85±3.26	90.34±2.12	0.96±0.12	3.09±0.81	90.30±2.12
SegTom[8]	JBHI 2025	81.77±3.48	89.59±2.37	1.13±0.53	3.29±0.73	89.54±2.38
nnWNet [17]	CVPR 2025	82.73±2.92	90.25±1.93	0.90±0.12	3.03±0.62	90.20±1.93
SuperLightNet [14]	CVPR 2025	82.68±3.82	90.15±2.57	1.03±0.23	3.03±0.77	90.11±2.57
DiffUNet [13]	MedIA 2025	83.49±3.07	90.72±2.00	0.89±0.13	2.96±0.71	90.67±2.00
HiPaSNet [1]	Nat.Com 2025	82.55±3.23	90.15±2.16	0.96±0.22	3.24±1.15	90.11±2.16
Ours	–	84.05±3.05	91.05±1.99	0.84±0.19	2.86±0.92	91.00±1.99

Table S8. Results on the FLARE2022 dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	80.17±8.88	87.87±6.79	2.40±1.71	10.98±12.35	87.81±6.81
V-Net [7]	3DV 2016	71.20±9.55	80.77±8.37	8.00±2.80	45.66±17.83	80.69±8.39
3D UX-Net [5]	ICLR 2023	80.44±9.21	87.98±6.94	1.69±1.05	6.32±5.55	87.92±6.97
UNETR [4]	WACV 2022	73.74±10.58	82.83±9.16	2.64±1.72	13.68±8.04	82.75±9.18
SwinUNETR [3]	MICCAI 2022	80.57±9.22	88.18±6.90	2.69±2.67	12.76±13.00	88.12±6.93
nnFormer [16]	TIP 2023	71.55±10.00	81.32±8.68	3.39±1.32	18.80±10.18	81.23±8.71
MedNeXt [11]	MICCAI 2023	78.57±9.49	86.83±7.20	1.75±0.99	6.55±6.02	86.76±7.22
SegFormer3D [9]	CVPR 2024	71.51±11.84	80.99±10.59	3.65±2.07	22.50±14.62	80.91±10.62
SegMamba [12]	MICCAI 2024	81.58±8.79	88.86±6.57	1.89±1.71	9.21±10.46	88.80±6.59
UNesT [15]	MedIA 2023	80.58±7.72	88.31±5.55	2.16±1.40	8.37±8.84	88.25±5.57
SwinSMT [10]	MICCAI 2024	78.63±9.85	86.65±8.04	3.56±1.58	15.83±11.76	86.59±8.06
VSmTrans [6]	MedIA 2024	80.66±7.93	88.17±6.07	2.21±1.25	6.77±6.87	88.11±6.09
SegTom[8]	JBHI 2025	75.41±10.53	84.29±8.54	2.50±1.54	10.13±7.54	84.22±8.56
nnWNet [17]	CVPR 2025	69.60±16.37	78.69±15.73	–	–	78.58±15.73
SuperLightNet [14]	CVPR 2025	78.25±9.58	86.51±7.58	2.72±1.44	15.38±11.60	86.45±7.61
DiffUNet [13]	MedIA 2025	81.44±6.74	88.92±4.83	2.31±1.17	11.53±10.07	88.87±4.85
HiPaSNet [1]	Nat.Com 2025	77.61±10.17	85.95±8.27	2.39±1.57	12.62±11.03	85.89±8.29
Ours	–	82.65±7.20	89.76±5.03	1.17±0.96	4.30±4.93	89.70±5.04

Table S9. Results on the SKI10 dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	78.46±4.29	86.73±3.23	0.94±0.40	3.51±2.24	86.19±3.41
V-Net [7]	3DV 2016	38.18±29.99	46.62±33.26	–	–	43.39±34.88
3D UX-Net [5]	ICLR 2023	78.05±8.49	86.31±7.03	0.95±0.68	3.42±2.58	85.74±7.47
UNETR [4]	WACV 2022	76.46±6.45	85.25±5.06	1.20±1.25	4.72±6.05	84.61±5.46
SwinUNETR [3]	MICCAI 2022	77.53±9.61	85.84±8.47	1.14±1.35	4.18±5.03	85.26±8.81
nnFormer [16]	TIP 2023	76.52±4.71	85.29±3.64	1.08±0.67	4.13±4.15	84.67±3.91
MedNeXt [11]	MICCAI 2023	78.67±4.75	86.87±3.50	0.93±0.46	3.41±1.88	86.34±3.71
SegFormer3D [9]	CVPR 2024	69.54±14.65	79.30±15.12	–	–	78.42±15.21
SegMamba [12]	MICCAI 2024	78.42±5.91	86.70±4.31	0.99±0.74	3.43±2.73	86.13±4.74
UNesT [15]	MedIA 2023	78.25±4.60	86.54±3.61	0.94±0.52	3.41±2.21	85.99±3.83
SwinSMT [10]	MICCAI 2024	77.78±7.54	86.16±5.95	1.03±0.84	3.81±3.21	85.58±6.44
VSmTrans [6]	MedIA 2024	73.49±12.58	82.53±11.47	2.96±4.79	12.06±18.97	81.77±12.19
SegTom[8]	JBHI 2025	78.49±3.79	86.74±2.89	0.89±0.26	3.26±1.28	86.23±3.01
nnWNet [17]	CVPR 2025	43.29±29.74	51.88±32.98	–	–	50.38±33.04
SuperLightNet [14]	CVPR 2025	78.26±6.53	86.58±4.89	0.96±0.75	3.48±2.73	85.99±5.36
DiffUNet [13]	MedIA 2025	79.03±4.45	87.13±3.37	0.87±0.38	3.11±1.83	86.62±3.55
HiPaSNet [1]	Nat.Com 2025	77.47±6.96	85.98±5.54	1.10±0.97	4.26±4.39	85.38±5.94
Ours	–	79.16±4.24	87.26±3.15	0.85±0.32	3.12±1.61	86.75±3.33

Table S10. Results on the FeTA2021 dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	78.54±3.12	87.79±2.04	0.93±0.12	2.41±0.34	87.39±1.96
V-Net [7]	3DV 2016	48.65±13.35	62.54±14.36	–	–	61.63±14.75
3D UX-Net [5]	ICLR 2023	78.70±2.91	87.90±1.89	0.91±0.09	2.33±0.31	87.50±1.82
UNETR [4]	WACV 2022	77.41±3.59	86.99±2.50	0.97±0.16	2.58±0.76	86.55±2.47
SwinUNETR [3]	MICCAI 2022	78.25±3.31	87.58±2.20	0.93±0.12	2.47±0.64	87.16±2.15
nnFormer [16]	TIP 2023	78.38±2.71	87.70±1.75	0.92±0.08	2.38±0.34	87.30±1.67
MedNeXt [11]	MICCAI 2023	78.78±2.77	87.95±1.79	0.90±0.07	2.32±0.25	87.55±1.70
SegFormer3D [9]	CVPR 2024	76.38±2.54	86.38±1.72	1.01±0.06	2.58±0.26	85.94±1.64
SegMamba [12]	MICCAI 2024	78.96±2.85	88.08±1.83	0.90±0.09	2.29±0.28	87.67±1.76
UNesT [15]	MedIA 2023	78.82±2.67	87.99±1.71	0.91±0.08	2.32±0.28	87.59±1.62
SwinSMT [10]	MICCAI 2024	77.95±3.53	87.36±2.46	0.95±0.14	2.52±0.73	86.94±2.43
VSmTrans [6]	MedIA 2024	75.90±3.13	86.06±2.10	1.12±0.30	2.82±0.62	85.59±2.12
SegTom[8]	JBHI 2025	78.71±2.55	87.92±1.64	0.91±0.07	2.36±0.26	87.52±1.56
nnWNet [17]	CVPR 2025	29.40±17.50	40.65±22.26	–	–	40.20±22.08
SuperLightNet [14]	CVPR 2025	78.23±2.81	87.60±1.83	0.93±0.08	2.41±0.26	87.19±1.75
DiffUNet [13]	MedIA 2025	78.73±3.03	87.92±1.98	0.91±0.08	2.32±0.30	87.52±1.89
HiPaSNet [1]	Nat.Com 2025	78.71±2.55	87.92±1.64	0.91±0.07	2.36±0.26	87.52±1.56
Ours	–	79.04±2.63	88.13±1.67	0.89±0.08	2.30±0.29	87.74±1.59

Table S11. Results on the OIMHS dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	87.69±2.79	93.18±1.75	0.54±0.57	2.91±5.45	92.64±1.80
V-Net [7]	3DV 2016	80.18±5.91	88.26±4.36	1.43±0.65	6.32±5.07	87.18±4.63
3D UX-Net [5]	ICLR 2023	88.37±3.33	93.55±2.13	0.49±0.55	2.87±5.44	93.06±2.17
UNETR [4]	WACV 2022	85.65±4.21	91.85±2.87	1.10±1.03	4.36±7.83	91.23±2.90
SwinUNETR [3]	MICCAI 2022	88.15±2.68	93.44±1.66	0.71±0.72	3.34±5.40	92.94±1.70
nnFormer [16]	TIP 2023	73.03±6.94	82.27±6.92	3.22±3.76	19.25±24.63	81.23±6.94
MedNeXt [11]	MICCAI 2023	87.19±3.81	92.78±2.63	0.45±0.32	2.20±1.24	92.27±2.64
SegFormer3D [9]	CVPR 2024	82.89±4.45	90.02±3.01	0.83±0.55	3.64±5.24	89.28±3.20
SegMamba [12]	MICCAI 2024	88.83±3.32	93.83±2.07	0.41±0.40	2.66±4.54	93.37±2.13
UNesT [15]	MedIA 2023	88.69±2.83	93.76±1.77	0.45±0.47	2.72±5.00	93.27±1.81
SwinSMT [10]	MICCAI 2024	88.13±3.39	93.40±2.18	0.72±0.72	2.92±5.35	92.89±2.23
VSmTrans [6]	MedIA 2024	88.56±2.63	93.70±1.63	0.68±0.59	2.62±5.10	93.20±1.68
SegTom[8]	JBHI 2025	85.07±3.64	91.49±2.46	1.20±0.98	5.19±7.62	90.86±2.49
nnWNet [17]	CVPR 2025	77.81±6.03	86.82±4.32	3.14±1.67	15.02±16.92	85.36±4.80
SuperLightNet [14]	CVPR 2025	83.85±4.43	90.62±3.23	0.66±0.31	3.50±1.29	89.97±3.24
DiffUNet [13]	MedIA 2025	88.80±3.04	93.82±1.91	0.55±0.65	3.77±7.23	93.34±1.95
HiPaSNet [1]	Nat.Com 2025	87.50±3.23	93.00±2.15	0.44±0.33	2.83±3.72	92.49±2.17
Ours	–	89.41±2.91	94.20±1.79	0.43±0.49	2.55±5.15	93.74±1.84

Table S12. Results on the MSD BRATS Tumour dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	67.24±10.34	78.74±8.70	1.41±2.27	4.93±7.14	78.56±8.74
V-Net [7]	3DV 2016	32.83±27.81	40.71±32.80	–	–	40.52±32.71
3D UX-Net [5]	ICLR 2023	65.82±11.00	77.65±9.29	1.24±0.70	4.73±3.69	77.46±9.34
UNETR [4]	WACV 2022	63.05±11.52	75.33±9.99	1.69±1.55	7.18±9.66	75.12±10.04
SwinUNETR [3]	MICCAI 2022	66.08±11.26	77.87±9.49	1.37±1.50	5.27±6.29	77.68±9.54
nnFormer [16]	TIP 2023	61.65±11.61	74.28±10.47	1.80±1.71	6.69±7.80	74.06±10.51
MedNeXt [11]	MICCAI 2023	66.71±10.63	78.39±8.92	1.31±1.14	4.90±5.43	78.21±8.96
SegFormer3D [9]	CVPR 2024	52.94±16.65	65.44±17.57	–	–	65.21±17.59
SegMamba [12]	MICCAI 2024	67.24±10.79	78.75±8.88	1.29±1.02	4.91±5.28	78.57±8.92
UNesT [15]	MedIA 2023	65.41±11.01	77.33±9.52	1.33±0.95	5.18±6.27	77.14±9.56
SwinSMT [10]	MICCAI 2024	65.04±11.03	77.05±9.41	1.33±0.86	5.25±5.93	76.85±9.45
VSmTrans [6]	MedIA 2024	60.91±12.42	73.48±10.80	1.92±1.75	8.16±8.92	73.26±10.86
SegTom[8]	JBHI 2025	65.95±10.21	77.70±8.89	1.42±1.09	5.35±5.66	77.51±8.92
nnWNet [17]	CVPR 2025	32.27±28.66	39.96±33.92	–	–	39.77±33.82
SuperLightNet [14]	CVPR 2025	67.09±10.07	78.75±8.26	1.28±1.10	4.71±4.47	78.57±8.30
DiffUNet [13]	MedIA 2025	67.29±10.33	78.77±8.53	1.41±2.23	4.74±6.87	78.59±8.56
HiPaSNet [1]	Nat.Com 2025	67.35±10.25	78.90±8.63	1.30±1.30	4.79±5.79	78.72±8.66
Ours	–	67.61±10.17	79.07±8.45	1.23±0.77	4.71±3.71	78.89±8.48

Table S13. Results on the MSD Lung Tumour dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	58.93±13.43	73.20±11.41	11.36±14.30	79.68±116.90	73.18±11.43
V-Net [7]	3DV 2016	51.61±23.26	64.29±25.07	13.85±35.13	30.44±62.55	64.28±25.07
3D UX-Net [5]	ICLR 2023	56.76±18.43	70.07±20.29	9.21±8.01	43.94±60.34	70.06±20.30
UNETR [4]	WACV 2022	44.43±22.85	57.72±24.38	28.85±40.47	129.54±121.21	57.71±2.44
SwinUNETR [3]	MICCAI 2022	59.52±15.58	73.29±13.86	8.18±10.46	51.46±84.56	73.27±13.88
nnFormer [16]	TIP 2023	50.01±20.41	63.63±22.72	12.79±13.36	49.72±60.00	63.62±22.72
MedNeXt [11]	MICCAI 2023	53.74±18.78	67.39±20.99	12.07±13.57	73.46±84.46	67.38±20.99
SegFormer3D [9]	CVPR 2024	39.58±20.10	53.50±22.58	27.69±28.81	127.71±117.65	53.49±22.58
SegMamba [12]	MICCAI 2024	57.14±17.32	70.97±16.02	8.74±10.44	44.35±73.87	70.95±16.03
UNesT [15]	MedIA 2023	57.44±19.21	70.41±21.24	5.09±7.06	20.56±35.21	70.40±21.24
SwinSMT [10]	MICCAI 2024	58.88±14.12	73.06±11.96	11.15±11.78	51.99±69.84	73.04±11.98
VSmTrans [6]	MedIA 2024	58.93±17.21	72.43±16.10	5.83±5.68	29.96±51.58	72.42±16.12
SegTom[8]	JBHI 2025	55.91±18.99	69.27±20.42	8.25±10.88	42.23±84.88	69.26±20.43
nnWNet [17]	CVPR 2025	51.08±20.35	64.67±22.30	11.54±31.79	16.47±37.34	64.66±22.30
SuperLightNet [14]	CVPR 2025	53.91±17.36	67.95±18.97	16.39±21.55	74.95±97.07	67.94±18.98
DiffUNet [13]	MedIA 2025	57.67±17.45	71.35±16.46	5.06±7.30	20.56±47.81	71.33±16.47
HiPaSNet [1]	Nat.Com 2025	53.88±19.54	67.33±21.60	11.68±14.22	46.58±66.73	67.32±21.59
Ours	–	60.46±16.90	73.63±16.56	3.51±4.50	14.66±32.83	73.61±16.58

Table S14. Results on the MSD Pancreas Tumour dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	54.14±10.57	67.99±10.20	4.44±7.15	26.36±66.14	67.94±10.21
V-Net [7]	3DV 2016	47.92±11.84	61.59±13.55	–	–	61.54±13.56
3D UX-Net [5]	ICLR 2023	51.53±12.14	64.21±13.22	15.41±34.82	40.67±82.85	64.16±13.22
UNETR [4]	WACV 2022	38.80±11.28	51.69±13.83	8.04±8.09	25.63±23.39	51.63±13.83
SwinUNETR [3]	MICCAI 2022	50.39±12.93	63.43±13.73	10.27±14.37	44.79±68.06	63.38±13.73
nnFormer [16]	TIP 2023	47.83±12.73	60.99±14.08	9.03±17.26	24.57±36.53	60.94±14.09
MedNeXt [11]	MICCAI 2023	53.49±11.19	67.25±11.41	2.91±2.36	11.02±11.57	67.21±11.41
SegFormer3D [9]	CVPR 2024	46.15±12.05	59.48±13.88	–	–	59.43±13.89
SegMamba [12]	MICCAI 2024	54.85±13.33	67.69±13.88	5.22±8.69	17.67±33.08	67.65±13.89
UNesT [15]	MedIA 2023	51.40±10.56	65.04±11.47	3.73±4.41	14.17±14.80	64.99±11.47
SwinSMT [10]	MICCAI 2024	49.29±12.03	62.67±13.04	10.84±21.30	31.93±61.52	62.62±13.05
VSmTrans [6]	MedIA 2024	55.09±11.52	68.53±11.36	8.23±10.21	42.56±50.14	68.48±11.36
SegTom[8]	JBHI 2025	49.38±11.66	62.60±12.52	-	-	62.55±12.53
nnWNet [17]	CVPR 2025	47.78±13.35	60.49±15.19	–	–	60.44±15.19
SuperLightNet [14]	CVPR 2025	53.26±11.34	67.09±11.24	3.54±2.96	15.20±21.32	67.04±11.24
DiffUNet [13]	MedIA 2025	55.27±11.67	68.46±12.15	–	–	68.41±12.15
HiPaSNet [1]	Nat.Com 2025	53.09±11.73	66.57±12.62	–	–	66.52±12.62
Ours	–	56.68±10.95	70.20±10.53	2.86±2.31	11.72±14.33	70.16±10.54

Table S15. Results on the MSD Spleen dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	91.84±1.96	95.74±1.07	0.39±0.17	1.05±0.14	95.71±1.08
V-Net [7]	3DV 2016	87.39±3.75	93.23±2.18	2.43±2.89	17.21±42.09	93.18±2.18
3D UX-Net [5]	ICLR 2023	91.37±1.69	95.48±0.93	0.42±0.56	1.18±0.34	95.45±0.93
UNETR [4]	WACV 2022	89.84±3.69	94.61±2.09	1.00±0.86	1.68±1.15	94.57±2.11
SwinUNETR [3]	MICCAI 2022	91.48±2.29	95.53±1.26	0.56±0.41	1.13±0.33	95.50±1.27
nnFormer [16]	TIP 2023	83.83±2.36	91.19±1.38	9.40±7.75	68.29±86.54	91.31±1.40
MedNeXt [11]	MICCAI 2023	90.66±1.72	95.09±0.95	0.55±0.22	1.23±0.34	95.06±0.95
SegFormer3D [9]	CVPR 2024	76.68±10.31	86.37±7.33	1.56±1.56	4.67±5.54	86.29±7.37
SegMamba [12]	MICCAI 2024	92.38±1.49	96.03±0.81	0.88±1.15	1.18±0.34	96.00±0.82
UNesT [15]	MedIA 2023	91.84±1.74	95.74±0.95	0.40±0.15	1.10±0.18	95.71±0.96
SwinSMT [10]	MICCAI 2024	90.73±2.07	95.13±1.15	0.64±0.43	1.14±0.26	95.09±1.16
VSmTrans [6]	MedIA 2024	91.96±2.71	95.79±1.49	0.94±1.19	5.79±12.29	95.76±1.51
SegTom[8]	JBHI 2025	91.03±3.14	95.28±1.74	2.43±3.79	17.65±43.14	95.24±1.75
nnWNet [17]	CVPR 2025	90.31±5.19	95.12±1.99	0.49±0.33	1.66±1.22	94.79±3.95
SuperLightNet [14]	CVPR 2025	91.10±2.47	95.32±1.36	0.54±0.26	1.23±0.34	95.29±1.38
DiffUNet [13]	MedIA 2025	92.09±2.36	95.86±1.30	0.62±0.39	1.13±0.33	95.84±1.31
HiPaSNet [1]	Nat.Com 2025	91.62±2.40	95.61±1.31	0.83±0.76	1.27±0.38	95.58±1.32
Ours	–	92.65±2.40	96.17±1.31	0.43±0.30	1.18±0.34	96.14±1.32

Table S16. Results on the MSD Colon Cancer dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	42.90±23.74	55.69±26.44	25.71±31.55	96.14±81.80	55.65±26.44
V-Net [7]	3DV 2016	23.59±19.63	34.00±26.35	38.26±24.18	145.65±63.06	33.96±26.33
3D UX-Net [5]	ICLR 2023	30.23±22.64	41.40±28.96	26.48±32.22	88.66±93.75	41.37±28.95
UNETR [4]	WACV 2022	18.91±17.25	28.35±23.94	38.09±39.65	127.05±70.47	28.31±23.93
SwinUNETR [3]	MICCAI 2022	33.84±21.32	46.41±26.36	31.69±37.44	130.48±120.22	46.38±26.36
nnFormer [16]	TIP 2023	19.31±15.06	29.95±19.35	72.48±29.75	214.26±66.76	29.89±19.34
MedNeXt [11]	MICCAI 2023	33.75±20.92	46.35±26.53	21.50±35.71	63.54±69.98	46.33±26.52
SegFormer3D [9]	CVPR 2024	30.46±21.18	42.43±26.46	26.34±35.44	77.79±75.77	42.40±26.46
SegMamba [12]	MICCAI 2024	38.99±20.36	52.61±24.08	22.93±29.98	95.38±81.48	52.57±24.07
UNesT [15]	MedIA 2023	39.03±25.08	50.76±30.01	19.44±24.82	63.78±69.03	50.74±30.00
SwinSMT [10]	MICCAI 2024	35.03±23.68	46.94±28.32	28.11±38.70	97.71±87.12	46.91±28.32
VSmTrans [6]	MedIA 2024	41.26±22.21	54.56±24.78	25.07±36.08	85.15±84.91	54.52±24.79
SegTom[8]	JBHI 2025	36.60±23.26	48.96±27.44	28.11±35.00	105.63±79.79	48.93±27.43
nnWNet [17]	CVPR 2025	27.04±22.84	37.81±26.81	31.77±49.46	99.94±85.01	37.78±26.80
SuperLightNet [14]	CVPR 2025	35.64±20.68	48.77±25.03	22.23±29.88	84.79±78.87	48.74±25.03
DiffUNet [13]	MedIA 2025	39.52±23.14	52.41±25.77	23.63±26.46	91.94±80.88	52.38±25.76
HiPaSNet [1]	Nat.Com 2025	35.33±22.89	47.58±27.62	32.14±34.13	134.45±69.48	47.55±27.62
Ours	–	43.80±25.25	55.89±28.89	25.20±37.15	75.99±78.37	55.86±28.89

Table S17. Results on the HOCMvalvesSeg dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	74.21±4.62	84.09±3.27	2.77±1.88	11.98±10.19	83.92±3.34
V-Net [7]	3DV 2016	58.49±8.94	71.90±7.29	6.22±2.55	29.67±12.78	71.56±7.35
3D UX-Net [5]	ICLR 2023	74.30±4.96	84.19±3.49	2.73±2.09	11.99±10.49	84.01±3.56
UNETR [4]	WACV 2022	70.89±5.61	81.63±4.14	3.08±1.99	13.41±10.13	81.42±4.20
SwinUNETR [3]	MICCAI 2022	73.51±4.45	83.62±3.07	2.84±2.12	12.84±10.55	83.45±3.14
nnFormer [16]	TIP 2023	68.38±5.06	79.89±3.66	3.99±1.90	17.94±10.97	79.66±3.72
MedNeXt [11]	MICCAI 2023	73.44±4.94	83.53±3.59	2.75±1.67	12.11±10.00	83.35±3.65
SegFormer3D [9]	CVPR 2024	71.74±4.46	82.48±3.16	3.15±2.08	13.14±10.35	82.27±3.22
SegMamba [12]	MICCAI 2024	74.13±4.70	84.09±3.26	2.66±1.90	11.96±10.23	83.92±3.33
UNesT [15]	MedIA 2023	73.94±4.78	83.93±3.30	2.94±2.08	12.48±10.32	83.74±3.38
SwinSMT [10]	MICCAI 2024	74.14±4.62	84.06±3.21	2.86±1.84	11.63±10.01	83.89±3.28
VSmTrans [6]	MedIA 2024	73.38±5.30	83.45±3.82	3.47±2.18	13.43±11.10	83.27±3.89
SegTom[8]	JBHI 2025	69.59±3.65	80.58±2.67	3.85±2.20	14.75±11.03	80.38±2.72
nnWNet [17]	CVPR 2025	60.94±9.61	73.51±9.17	-	-	73.19±9.25
SuperLightNet [14]	CVPR 2025	71.18±4.77	81.84±3.45	3.45±1.81	13.67±10.15	81.64±3.50
DiffUNet [13]	MedIA 2025	73.96±5.06	83.96±3.55	3.05±2.16	12.48±10.63	83.78±3.62
HiPaSNet [1]	Nat.Com 2025	71.93±4.70	82.52±3.28	3.84±2.26	14.40±10.96	82.32±3.35
Ours	–	74.84±4.51	84.63±3.11	2.82±1.82	11.56±10.02	84.46±3.18

Table S18. Results on the CrossMoDA2021 dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	67.95±6.44	79.76±4.65	0.57±0.30	1.97±1.19	79.76±4.65
V-Net [7]	3DV 2016	68.08±8.40	79.65±7.08	0.69±0.67	2.10±2.01	79.64±7.09
3D UX-Net [5]	ICLR 2023	66.79±8.80	78.76±6.61	0.77±0.53	2.60±2.78	78.75±6.62
UNETR [4]	WACV 2022	61.29±8.46	74.39±7.22	2.34±1.89	10.95±20.60	74.38±7.22
SwinUNETR [3]	MICCAI 2022	69.18±7.09	80.49±5.50	1.42±1.52	6.36±18.99	80.49±5.50
nnFormer [16]	TIP 2023	20.00±12.43	26.04±14.48	–	–	26.04±14.48
MedNeXt [11]	MICCAI 2023	67.25±8.22	78.93±6.37	0.58±0.28	1.84±0.96	78.92±6.37
SegFormer3D [9]	CVPR 2024	53.54±9.32	68.00±9.20	1.48±1.24	3.80±2.66	68.00±9.20
SegMamba [12]	MICCAI 2024	67.65±7.02	79.29±5.04	2.97±8.01	10.46±36.41	79.28±5.04
UNesT [15]	MedIA 2023	68.61±7.30	80.15±5.31	0.57±0.36	1.92±1.17	80.15±5.32
SwinSMT [10]	MICCAI 2024	68.99±6.31	80.40±4.76	1.79±1.99	10.22±27.02	80.40±4.77
VSmTrans [6]	MedIA 2024	67.32±8.22	79.20±6.30	3.62±6.40	14.83±23.70	79.19±6.30
SegTom[8]	JBHI 2025	67.58±7.36	79.42±5.31	0.72±0.52	1.95±1.16	79.42±5.32
nnWNet [17]	CVPR 2025	64.52±8.78	76.99±7.54	2.18±3.04	8.65±15.15	76.99±7.55
SuperLightNet [14]	CVPR 2025	64.09±7.78	76.38±6.28	5.05±4.64	23.96±36.43	76.38±6.28
DiffUNet [13]	MedIA 2025	68.51±6.66	80.05±4.85	0.55±0.30	1.80±1.09	80.05±4.85
HiPaSNet [1]	Nat.Com 2025	68.48±6.67	79.99±5.07	1.28±1.39	9.53±16.31	79.99±5.07
Ours	–	69.27±6.41	80.65±4.58	0.54±0.33	1.77±1.15	80.65±4.59

Table S19. Results on the SLIVER07 dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	87.46±8.10	93.10±4.70	7.56±7.33	48.23±51.25	92.33±4.99
V-Net [7]	3DV 2016	83.03±8.99	90.46±5.42	9.24±7.87	71.60±75.46	89.38±5.71
3D UX-Net [5]	ICLR 2023	90.62±4.00	95.03±2.20	4.18±3.13	26.23±23.27	94.42±2.37
UNETR [4]	WACV 2022	87.12±5.58	93.02±3.22	11.83±8.80	76.41±74.28	92.18±3.37
SwinUNETR [3]	MICCAI 2022	91.31±3.76	95.42±2.07	3.40±2.28	25.15±21.77	94.80±2.40
nnFormer [16]	TIP 2023	86.81±4.96	92.86±2.89	9.28±7.66	72.59±74.93	92.01±2.93
MedNeXt [11]	MICCAI 2023	89.88±6.28	94.55±3.60	7.97±9.61	51.30±79.70	93.94±3.74
SegFormer3D [9]	CVPR 2024	84.40±13.62	90.89±8.73	7.24±7.94	43.97±47.23	90.10±9.03
SegMamba [12]	MICCAI 2024	91.18±4.46	95.33±2.49	3.18±3.40	13.90±16.89	94.80±2.54
UNesT [15]	MedIA 2023	89.35±5.65	94.28±3.19	4.34±3.06	34.35±34.42	93.61±3.35
SwinSMT [10]	MICCAI 2024	89.89±4.38	94.62±2.45	4.14±3.12	28.22±25.37	93.97±2.56
VSmTrans [6]	MedIA 2024	91.35±2.85	95.46±1.56	4.84±2.14	27.07±20.48	94.86±1.75
SegTom[8]	JBHI 2025	83.52±12.92	90.43±8.29	9.29±10.00	58.13±65.36	89.51±8.59
nnWNet [17]	CVPR 2025	82.91±8.42	90.42±5.04	11.23±6.67	90.69±66.17	89.28±5.44
SuperLightNet [14]	CVPR 2025	89.87±4.20	94.61±2.32	3.95±2.70	16.95±17.27	93.95±2.48
DiffUNet [13]	MedIA 2025	90.24±4.63	94.81±2.58	5.24±5.04	41.17±46.41	94.19±2.70
HiPaSNet [1]	Nat.Com 2025	89.55±6.60	94.35±3.78	4.33±4.20	35.93±38.75	93.74±3.93
Ours	–	91.68±3.22	95.63±1.76	3.78±2.47	26.45±23.45	95.09±1.85

Table S20. Results on the COVID-19 CT Seg dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	71.50±15.38	81.92±11.09	8.03±5.99	51.96±46.95	80.60±11.73
V-Net [7]	3DV 2016	8.78±7.95	13.68±11.34	-	-	12.29±10.68
3D UX-Net [5]	ICLR 2023	71.44±16.52	81.75±11.98	8.39±6.17	57.31±48.43	80.42±12.67
UNETR [4]	WACV 2022	62.38±19.78	74.20±15.65	12.06±9.38	61.26±44.32	72.44±16.54
SwinUNETR [3]	MICCAI 2022	70.29±15.99	80.97±11.42	14.10±12.10	81.29±64.52	79.51±12.28
nnFormer [16]	TIP 2023	67.52±15.95	79.07±11.38	9.61±8.20	50.66±49.58	77.44±12.33
MedNeXt [11]	MICCAI 2023	70.06±17.69	80.55±13.38	7.51±5.72	34.54±29.85	79.15±14.03
SegFormer3D [9]	CVPR 2024	49.07±33.27	57.82±31.83	9.42±7.55	36.17±35.53	56.19±32.85
SegMamba [12]	MICCAI 2024	71.73±17.34	81.84±12.74	7.05±5.74	49.95±46.51	80.53±13.40
UNesT [15]	MedIA 2023	69.81±17.42	80.39±13.03	8.35±6.39	45.62±45.24	79.00±13.72
SwinSMT [10]	MICCAI 2024	68.81±17.64	79.55±13.25	8.09±6.37	46.39±46.13	78.13±13.97
VSmTrans [6]	MedIA 2024	64.31±18.80	75.53±15.00	13.64±6.70	81.06±54.27	74.02±15.73
SegTom[8]	JBHI 2025	62.88±17.83	74.49±14.08	14.07±11.28	67.64±61.43	72.91±14.88
nnWNet [17]	CVPR 2025	12.23±12.23	17.98±17.99	-	-	16.49±16.51
SuperLightNet [14]	CVPR 2025	72.62±13.88	82.87±9.94	5.71±2.51	36.32±17.91	81.64±10.44
DiffUNet [13]	MedIA 2025	71.09±15.31	81.63±11.02	9.46±5.94	60.21±39.48	80.27±11.71
HiPaSNet [1]	Nat.Com 2025	70.54±13.38	81.43±9.65	6.24±4.44	37.47±34.50	80.09±10.18
Ours	–	73.91±15.05	83.75±10.61	6.39±5.10	32.50±29.22	82.51±11.22

Table S21. Results on the FLARE2021 dataset. The best performance is highlighted in red, and the second best is in blue.

Method	Source	Average				
		IoU	Dice	ASSD	HD95	Adj-Rand
3D U-Net [2]	MICCAI 2016	88.67±2.68	93.50±1.86	1.27±1.84	5.81±14.00	93.39±1.88
V-Net [7]	3DV 2016	77.00±10.28	85.12±8.59	3.52±2.99	14.59±13.82	84.87±8.72
3D UX-Net [5]	ICLR 2023	88.97±2.97	93.66±2.11	1.21±1.25	4.47±8.20	93.55±2.13
UNETR [4]	WACV 2022	85.88±3.67	91.52±2.80	1.57±1.45	6.24±8.69	91.39±2.82
SwinUNETR [3]	MICCAI 2022	88.84±2.73	93.61±1.85	1.27±1.21	5.22±6.75	93.50±1.88
nnFormer [16]	TIP 2023	79.86±7.05	86.50±6.70	3.99±7.69	11.61±9.83	86.32±6.78
MedNeXt [11]	MICCAI 2023	89.27±2.44	93.91±1.63	0.89±0.55	2.33±0.94	93.81±1.65
SegFormer3D [9]	CVPR 2024	85.87±3.17	91.83±2.28	1.86±2.22	8.93±17.30	91.68±2.30
SegMamba [12]	MICCAI 2024	89.02±2.65	93.73±1.80	1.51±1.97	8.53±15.47	93.62±1.82
UNesT [15]	MedIA 2023	89.28±2.57	93.90±1.74	0.77±0.44	2.50±1.31	93.80±1.76
SwinSMT [10]	MICCAI 2024	88.66±2.37	93.56±1.51	1.12±0.85	3.25±2.38	93.44±1.55
VSmTrans [6]	MedIA 2024	88.25±3.55	93.17±2.52	4.46±8.27	19.65±34.21	93.05±2.57
SegTom[8]	JBHI 2025	87.66±2.72	92.88±1.89	0.87±0.39	2.94±1.30	92.75±1.91
nnWNet [17]	CVPR 2025	69.50±12.97	77.53±11.87	6.98±4.99	29.30±18.63	77.24±12.01
SuperLightNet [14]	CVPR 2025	86.91±3.06	92.30±2.19	1.70±1.67	8.34±20.88	92.17±2.22
DiffUNet [13]	MedIA 2025	88.71±2.38	93.59±1.49	1.26±1.01	4.80±9.01	93.48±1.53
HiPaSNet [1]	Nat.Com 2025	88.85±3.09	93.59±2.20	0.90±0.41	2.44±1.14	93.49±2.23
Ours	–	89.33±1.70	94.00±1.06	1.30±1.79	5.35±9.52	93.90±1.08

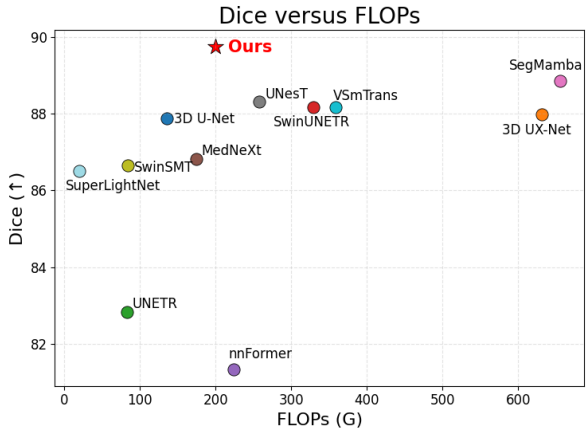

Figure S6. Dice is plotted against model compute (FLOPs, G) for representative 3D segmentation methods trained and evaluated under the same protocol.

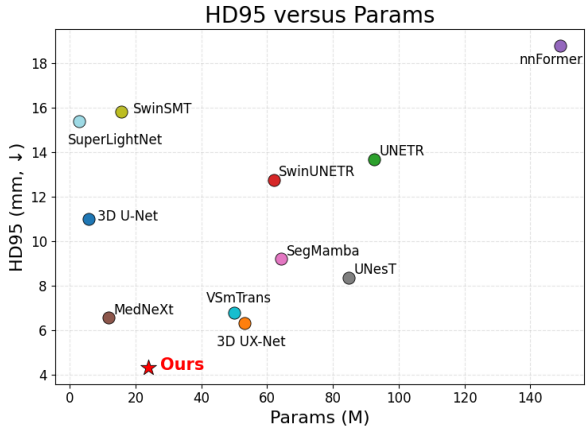

Figure S7. HD95 is plotted against model size (Params, M) for representative 3D segmentation methods trained and evaluated under the same protocol.

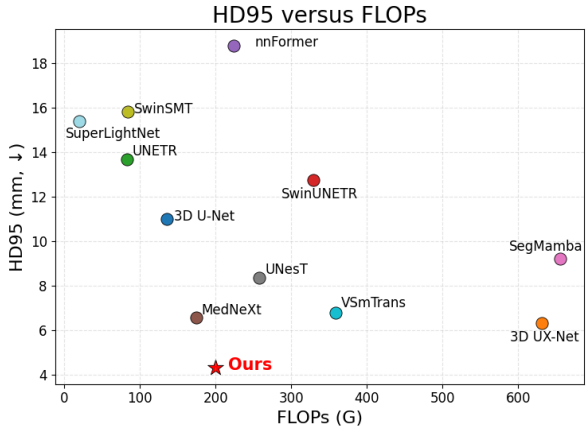

Figure S8. HD95 is plotted against model compute (FLOPs, G) for representative 3D segmentation methods trained and evaluated under the same protocol.

References

[1] Yuetan Chu, Gongning Luo, Longxi Zhou, Shaodong Cao, Guolin Ma, Xianglin Meng, Juexiao Zhou, Changchun Yang, Dexuan Xie, Dan Mu, et al. Deep learning-driven pulmonary artery and vein segmentation reveals demography-associated vasculature anatomical differences. *Nature Communications*, 16(1):2262, 2025. 7, 8, 9, 10, 11, 12, 13, 14

[2] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016. 4, 5, 7, 8, 9, 10, 11, 12, 13, 14

[3] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI brainlesion workshop*, pages 272–284. Springer, 2021. 7, 8, 9, 10, 11, 12, 13, 14

[4] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584, 2022. 4, 5, 7, 8, 9, 10, 11, 12, 13, 14

[5] Ho Hin Lee, Shunxing Bao, Yuankai Huo, and Bennett A Landman. 3d ux-net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation. In *The Eleventh International Conference on Learning Representations*, 2023. 4, 5, 7, 8, 9, 10, 11, 12, 13, 14

[6] Tiange Liu, Qingze Bai, Drew A Torigian, Yubing Tong, and Jayaram K Udupa. Vsmtrans: A hybrid paradigm integrating self-attention and convolution for 3d medical image segmentation. *Medical image analysis*, 98:103295, 2024. 7, 8, 9, 10, 11, 12, 13, 14

[7] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016. 7, 8, 9, 10, 11, 12, 13, 14

[8] Yan Pang, Yunhao Li, Jiaming Liang, Hao Chen, Ying Hu, and Qiong Wang. Segtom: A 3d volumetric medical image segmentation framework for thoracoabdominal multi-organ anatomical structures. *IEEE Journal of Biomedical and Health Informatics*, 2025. 7, 8, 9, 10, 11, 12, 13, 14

[9] Shehan Perera, Pouyan Navard, and Alper Yilmaz. Segformer3d: an efficient transformer for 3d medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4981–4988, 2024. 7, 8, 9, 10, 11, 12, 13, 14

[10] Szymon Płotka, Maciej Chrabaszcz, and Przemyslaw Biecek. Swin smt: Global sequential modeling for enhancing 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 689–698. Springer, 2024. 4, 5, 7, 8, 9, 10, 11, 12, 13, 14

[11] Saikat Roy, Gregor Koehler, Constantin Ulrich, Michael Baumgartner, Jens Petersen, Fabian Isensee, Paul F Jaeger, and Klaus H Maier-Hein. Mednext: transformer-driven scaling of convnets for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 405–415. Springer, 2023. 4, 5, 7, 8, 9, 10, 11, 12, 13, 14

[12] Zhaohu Xing, Tian Ye, Yijun Yang, Guang Liu, and Lei Zhu. Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 578–588. Springer, 2024. 4, 5, 7, 8, 9, 10, 11, 12, 13, 14

[13] Zhaohu Xing, Liang Wan, Huazhu Fu, Guang Yang, Yijun Yang, Lequan Yu, Baiying Lei, and Lei Zhu. Diff-unet: A diffusion embedded network for robust 3d medical image segmentation. *Medical Image Analysis*, page 103654, 2025. 7, 8, 9, 10, 11, 12, 13, 14

[14] Feng Yu, Jiacheng Cao, Li Liu, and Minghua Jiang. Superlightnet: Lightweight parameter aggregation network for multimodal brain tumor segmentation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5197–5206, 2025. 7, 8, 9, 10, 11, 12, 13, 14

[15] Xin Yu, Qi Yang, Yinchi Zhou, Leon Y Cai, Riqiang Gao, Ho Hin Lee, Thomas Li, Shunxing Bao, Zhoubing Xu, Thomas A Lasko, et al. Unest: local spatial representation learning with hierarchical transformer for efficient medical segmentation. *Medical Image Analysis*, 90:102939, 2023. 4, 5, 7, 8, 9, 10, 11, 12, 13, 14

[16] Hong-Yu Zhou, Jiansen Guo, Yinghao Zhang, Xiaoguang Han, Lequan Yu, Liansheng Wang, and Yizhou Yu. nnformer: Volumetric medical image segmentation via a 3d transformer. *IEEE transactions on image processing*, 32: 4036–4045, 2023. 7, 8, 9, 10, 11, 12, 13, 14

[17] Yanfeng Zhou, Lingrui Li, Le Lu, and Minfeng Xu. nnwnet: Rethinking the use of transformers in biomedical image segmentation and calling for a unified evaluation benchmark. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 20852–20862, 2025. 7, 8, 9, 10, 11, 12, 13, 14