Fractal-G: Continuous Grid-Graph-Grid Feature Reconstruction for Medical Image Segmentation

Bai Sun[1]* **Zixuan Tang**[2]*

[1]School of Engineering, Westlake University, Hangzhou, China
[2]School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China

sunbai@westlake.edu.cn tangzx23@mail2.sysu.edu.cn

Abstract

Medical image segmentation requires the precise alignment of macroscopic semantics and microscopic geometry, yet existing representation paradigms struggle to balance massive low-frequency semantic context with ultra-sparse, high-frequency boundary details. To address this fundamental information asymmetry, we introduce Fractal-G, a plug-and-play multi-scale fusion module that pioneers a continuous Grid-Graph-Grid feature reconstruction paradigm. Within this framework, an Uncertainty-Aware Fission Router dynamically translates regular image grids into a non-uniform heterogeneous graph, adaptively allocating dense microscopic nodes to complex boundaries while retaining sparse macroscopic nodes in homogeneous background regions. To ensure unbiased feature aggregation across these varying physical scales, an Area-Aware Continuous Neighborhood Graph explicitly incorporates physical node sizes into the topological message passing. Finally, a Topology-Aware Implicit Renderer projects the unstructured graph back into dense, artifact-free continuous feature fields. Extensive experiments on diverse medical image segmentation tasks (including skin lesions and polyps) demonstrate that Fractal-G seamlessly integrates with mainstream backbones, consistently achieving state-of-the-art geometric fidelity and cross-resolution robustness with marginal computational overhead. Our code is publicly available at https://github.com/SIYIQ/Fractal-G.

1 Introduction

Medical image segmentation demands the alignment of macroscopic semantics and microscopic geometry[1–4]. While semantics locate the lesion, precise continuous boundaries dictate geometric fidelity; as illustrated in Fig. 1(a), even minor boundary deviations can lead to devastating clinical outcomes, such as over-segmentation or under-treatment[5]. However, concurrently capturing both within existing frameworks remains an inherently contradictory task due to opposing frequency properties[6–8]. As depicted in Fig. 1(b), accurate semantic perception requires low-frequency translation invariance to ensure robust localization, whereas precise boundary delineation demands extreme high-frequency spatial sensitivity to capture intricate geometric details[9, 10]. Furthermore, the distribution of information density is extremely unbalanced[11, 12]: high-frequency boundary signals are exceptionally sparse, while the smooth low-frequency interior occupies the vast majority of the spatial area. Attempting to preserve these sparse high-frequency details without being overwhelmed by massive low-frequency semantics naturally incurs a severe information asymmetry conflict, which remains a fundamental challenge for current diagnostic backbones.

*Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

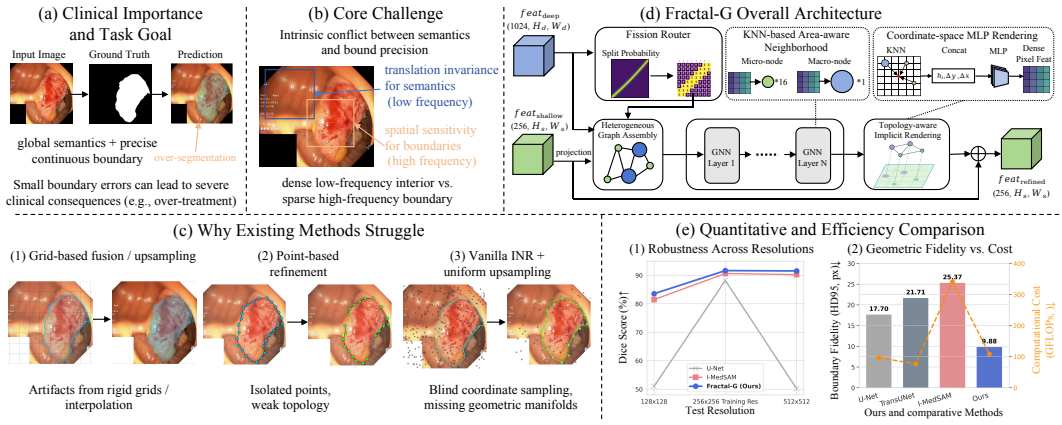

Figure 1: **Overview and Motivation of Fractal-G. a) Clinical Impact:** Small boundary segmentation errors can lead to severe clinical consequences (e.g., over-treatment). **b) Core Challenge:** The inherent information asymmetry conflict between dense, low-frequency macroscopic semantics and sparse, high-frequency microscopic boundaries. **c) Existing Limitations:** Grid-based models suffer from rigid interpolation artifacts (left), while point-based and vanilla INR methods blindly sample coordinates, leading to topological fractures (middle and right). **d) Our Architecture:** Fractal-G seamlessly resolves these bottlenecks through a Grid-Graph-Grid paradigm, uniting non-uniform fission routing, area-aware topological interaction, and artifact-free implicit rendering. **e) Quantitative Superiority:** Fractal-G significantly boosts geometric fidelity (e.g., HD95) and robustly maintains high Dice scores across varying resolutions, all while requiring notably fewer computational costs (GFLOPs) than existing continuous or point-based alternatives.

Existing mainstream representation paradigms struggle to resolve this fundamental asymmetry. As illustrated in Fig. 1(c-1), traditional grid-based methods (e.g., CNNs/ViTs with feature pyramids) enforce uniform computational allocation, causing severe resource mismatch and inevitable interpolation artifacts[13–17], such as the checkerboard or sawtooth edges at boundaries[18, 19]. To break grid limitations, adaptive point sampling (e.g., PointRend [11]) and Graph Neural Networks (GNNs) introduce dynamic focus or topological relations; however, as evidenced by the isolated points[20–22] in Fig. 1(c-2), they often lack global-local geometric connectivity or are restricted to fixed-resolution processing. Recently, Implicit Neural Representations (INRs)[23] emerged as a promising continuous alternative. However, existing INR models often suffer from topological fractures and weak connectivity, as they rely on inefficient random coordinate sampling and fit discrete hard masks, which struggles to capture ultra-sparse lesions[24] (see Fig. 1(c-3)). Consequently, the field urgently lacks a unified representation that simultaneously achieves on-demand dense boundary sampling, physical topological connectivity, and artifact-free continuous rendering.

Motivated by the limitations of both discrete grids and INR-based methods, we propose Fractal-G, a novel plug-and-play multi-scale fusion module pioneering a Grid-Graph-Grid feature reconstruction paradigm in continuous space (Fig. 1(d)). First, an Uncertainty-Aware Fission Router tackles boundary sparsity by dynamically converting grid pixels into a non-uniform heterogeneous graph, allocating dense microscopic nodes to high-frequency boundaries while retaining sparse macroscopic nodes elsewhere. Second, an Area-aware Continuous Neighborhood Graph addresses topological isolation by explicitly incorporating node area into its message passing, ensuring physically unbiased fusion between macroscopic semantics and microscopic details. Finally, a Topology-aware Implicit Renderer decodes the graph into an artifact-free continuous field for arbitrary-resolution queries. As shown in Fig. 1(e), this synergistic design allows Fractal-G to achieve superior geometric fidelity (HD95) and cross-resolution robustness with significantly lower computational cost than competing continuous or point-based methods. The main contributions of this paper are summarized as follows:

- We identify the inherent conflict in frequency and information density between global semantics and local boundaries. To reconcile this, we propose a novel representation paradigm combining on-demand non-uniform fission with implicit continuous rendering.

- We design Fractal-G, a plug-and-play module that achieves efficient, artifact-free Grid-Graph-Grid conversion in continuous space via synergistic fission routing, area-aware topological interaction, and implicit rendering.
- Extensive experiments on multiple medical image segmentation tasks demonstrate Fractal-G's universality. Integrated into mainstream CNNs/ViTs, it significantly boosts geometric accuracy (e.g., Hausdorff Distance) without notably increasing computational overhead.

2 Related Work

Dynamic Sampling and Graph-based Segmentation To overcome the computational mismatch of uniform grid representations, adaptive sampling methods like PointRend allocate dynamic focus to high-frequency boundaries [25]. In parallel, recent efficient decoders and multi-scale fusion designs attempt to allocate capacity more economically across resolutions [14, 26, 27, 21]. Concurrently, Graph Neural Networks (GNNs) and superpixel-based models have been introduced to capture long-range structural dependencies and topological relations in medical images [28, 29]. While these approaches mitigate computational redundancy by dynamically sparsifying attention [30–32], they either process pixels in isolation [25] or rely on fixed-resolution and often non-differentiable graph construction schemes [28]. In contrast, our proposed Fission Router pioneers a dynamic Grid-to-Graph mechanism, adaptively routing homogeneous background regions into macroscopic nodes while undergoing dense fission at intricate boundaries, seamlessly integrating structural topology with highly efficient continuous sampling.

Topological Interactions in Feature Spaces Graph Attention Networks (GATs) are widely utilized to aggregate contextual features and establish long-range topological connections in medical image analysis [33–35]. Advanced methods further employ hierarchical graph structures and multi-scale message passing to model complex anatomical variations [29, 36–38, 15]. However, these conventional graph mechanisms are inherently area-agnostic—treating all discrete nodes uniformly—which introduces severe dimensional bias in non-uniform graph spaces. This allows massive macroscopic background features to easily overwhelm sparse microscopic boundary signals. To resolve this fundamental dimensional conflict, we construct an Area-aware Continuous Neighborhood Graph. By explicitly incorporating the physical area size of each node into the attention mechanism, our method ensures physically unbiased message passing, thereby perfectly bridging the topological gap between global semantics and local details.

Implicit Neural Representations in Medical Image Segmentation Implicit Neural Representations (INRs) have emerged as a resolution-agnostic alternative to discrete grid-based architectures, significantly reducing the memory overhead required for high-resolution medical signals [39]. Recent studies have successfully extended INRs to joint segmentation and motion estimation tasks [40], while specialized activation functions [41] and continuous pseudo-label rectification [42] have enhanced their ability to delineate intricate anatomical structures. In parallel, SAM-based frameworks and vision-language pre-training have been adapted to medical segmentation via implicit or auto-prompted pipelines [43, 44, 24]. Furthermore, specific parametrizations can now provably enforce geometric constraints such as star-convexity and path-connectedness [45]. However, most existing methods still treat coordinates as isolated query points during the occupancy fitting process [39, 40]. This point-wise independence neglects continuous geometric manifolds, often leading to disconnected components or "topological fractures" at sparse boundaries. To bridge this gap, our work introduces Topology-aware Implicit Rendering, leveraging non-uniform topological graphs to guide the continuous feature reconstruction.

3 Methodology

3.1 Preliminaries and Continuous Formulation

Let $\mathcal{X} \in \mathbb{R}^{H\times W\times C}$ denote an image defined on a discrete regular lattice. Traditional fully convolutional networks parameterize segmentation as a discrete mapping $f_\theta : \mathbb{R}^{H\times W\times C} \to \Delta^{K-1}$ onto the $(K-1)$-simplex of class probabilities. However, this enforces uniform computation, cre-

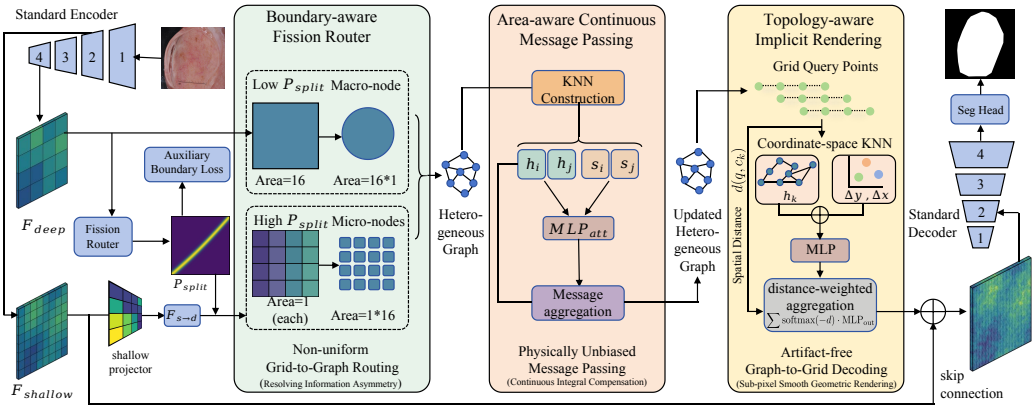

Figure 2: **The Grid-Graph-Grid architecture of Fractal-G.** Given deep semantic (F_{deep}) and shallow geometric ($F_{shallow}$) features, the Fission Router dynamically converts the grid into a heterogeneous graph, allocating dense nodes strictly to boundary regions via an auxiliary loss. The Area-aware Graph module performs topologically unbiased message passing by injecting node sizes. Finally, the Implicit Renderer decodes the unstructured graph back into an artifact-free dense feature map, which is residually fused into the decoder.

ating a frequency asymmetry between massive homogeneous backgrounds and sparse microscopic boundaries.

To transcend discrete grids, we formulate segmentation as learning a continuous topological manifold within a normalized coordinate space $\Omega = [0, 1]^2$. Any local region originating at pixel (x, y) on an $H \times W$ grid is anchored to a continuous geometric center $c_{x,y} \in \Omega$:

$$c_{x,y} = \left(\frac{x + S/2}{W}, \frac{y + S/2}{H} \right), \quad S = \begin{cases} P, & \text{macro-patches} \\ 1, & \text{micro-pixels} \end{cases} \tag{1}$$

where S defines the physical span, and P denotes the macroscopic scale factor. We aim to construct an implicit mapping $\mathcal{T}_\theta : \Omega \to \Delta^{K-1}$ that aligns macroscopic semantics with microscopic precision. By allowing queries at any arbitrary continuous coordinate $q \in \Omega$, our model fundamentally eliminates discrete interpolation artifacts.

3.2 Overall Architecture

As illustrated in Fig. 2, Fractal-G serves as a plug-and-play module bridging the deep and shallow feature hierarchies of a standard encoder-decoder. Let $F_s \in \mathbb{R}^{C_s \times H_s \times W_s}$ and $F_d \in \mathbb{R}^{C_d \times H_d \times W_d}$ denote the shallow geometric and deep semantic features, respectively, where spatial dimensions satisfy $H_s = P \cdot H_d$ and $W_s = P \cdot W_d$.

Rather than naively concatenating them, Fractal-G introduces a *Grid-Graph-Grid* transformation paradigm executed in three synergistic stages: (1) an **Uncertainty-Aware Fission Router** that translates the regular grids into a non-uniform heterogeneous graph; (2) an **Area-Aware Message Passing** module that aggregates topological features without physical bias; and (3) a **Topology-Aware Implicit Renderer** that projects the unstructured graph back into a continuous dense field. The refined feature is then fused via a skip connection into the decoder.

3.3 Uncertainty-Aware Fission Routing

The first stage of Fractal-G breaks the rigid constraints of the discrete grid by adaptively allocating representation capacity based on local structural complexity. We propose a dynamic fission routing mechanism inspired by quadtree algorithm[46], converting the feature maps into an unstructured node set $\mathcal{V}$.

Fission Probability Mapping Given the deep semantic feature F_d, we first employ a lightweight uncertainty-aware router to predict a boundary fission probability map $\mathcal{M}_{prob} \in [0,1]^{H_d \times W_d}$:

$$\mathcal{M}_{prob} = \sigma\Big(\mathbf{W}_{1\times1}\big(\text{ReLU}(\mathbf{W}_{3\times3} * F_d)\big)\Big) \tag{2}$$

where $\sigma(\cdot)$ denotes the sigmoid activation and $*$ denotes a 2D convolution. Utilizing a predefined decision threshold τ, we obtain a binary split mask $\mathcal{M}_{split} = \mathbb{I}(\mathcal{M}_{prob} \geq \tau)$.

Heterogeneous Node Generation Guided by $\mathcal{M}_{split}$, we project the multi-scale features into a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. To align the feature spaces, the shallow feature is linearly projected to match the deep channel dimension: $\tilde{F}_s = \mathbf{W}_{proj} F_s \in \mathbb{R}^{C_d \times H_s \times W_s}$. For each spatial location (y, x) on the $H_d \times W_d$ grid, the routing policy dictates node generation:

Macroscopic Nodes (Homogeneous Background). If $\mathcal{M}_{split}(y, x) = 0$, the $P \times P$ region is deemed homogeneous. We preserve it as a single macroscopic node: its feature is $h_i = F_d(y, x)$, its continuous spatial coordinate $c_i \in \Omega$ corresponds to the geometric center of the patch, and its physical area is explicitly recorded as $s_i = P^2$.

Microscopic Nodes (High-Frequency Boundary). If $\mathcal{M}_{split}(y, x) = 1$, the macro-node undergoes fission into P^2 dense microscopic nodes extracted from $\tilde{F}_s$: the parent patch is explicitly unrolled into P^2 independent locations uniformly distributed within the patch's continuous bounding box. For each micro-node, the feature is $h_i = \tilde{F}_s(y', x')$, a precise sub-pixel coordinate $c_i \in \Omega$ is assigned, and the physical area is recorded as $s_i = 1$. This dynamic routing yields a highly compact node set $\mathcal{V} = \{v_i\}_{i=1}^{N_{nodes}}$, where each node $v_i = (h_i, c_i, s_i)$. Since the vast background is effectively compressed into sparse macro-nodes, $N_{nodes} \ll H_s \times W_s$, ensuring maximal computational efficiency.

3.4 Area-Aware Continuous Neighborhood Graph

To bridge the topological gap between global macroscopic context and local microscopic geometry, it is imperative to perform message passing among the nodes. However, conventional Graph Attention Networks (GATs) are fundamentally area-agnostic. In our non-uniform space, naively aggregating a macro-node ($s_i = P^2$) with a micro-node ($s_i = 1$) would cause the subtle high-frequency signals to be irreversibly overwhelmed by massive background features.

Continuous Scale Embedding To resolve this dimensional bias, we construct an Area-Aware Continuous Neighborhood Graph. For each node v_i, we define its local continuous neighborhood $\mathcal{N}(i)$ by querying its k-nearest neighbors based on Euclidean distance in Ω. Notably, owing to the extreme sparsity achieved by the Fission Router, this continuous k-NN query incurs negligible computational overhead.

To explicitly inject physical scale awareness without dominating the semantic features, we project the scalar areas s_i and s_j into scale embeddings using a lightweight mapping function $\phi_s(\cdot)$ (implemented as logarithmic scaling followed by a linear projection). The directed attention coefficient $e_{i,j}$ from node $j \in \mathcal{N}(i)$ to node i is formulated as:

$$e_{i,j} = \text{LeakyReLU}\Big(\mathbf{W}_a\big[\mathbf{W}_h h_i \,||\, \mathbf{W}_h h_j \,||\, \phi_s(s_i) \,||\, \phi_s(s_j)\big]\Big) \tag{3}$$

where $||$ denotes concatenation, and $\mathbf{W}_h, \mathbf{W}_a$ are learnable projections. The physically calibrated attention weights $\alpha_{i,j}$ are normalized via the softmax function, and features are updated through a residual aggregation scheme $h_i' = h_i + \sum_{j\in\mathcal{N}(i)} \alpha_{i,j}(\mathbf{W}_h h_j)$.

3.5 Topology-Aware Implicit Rendering

Following the topological message passing, the unstructured graph must be projected back into a dense continuous feature field. Instead of standard coordinate-independent INR decoding, we propose a Topology-Aware Implicit Renderer.

For an arbitrary target spatial query point $q \in \Omega$, we retrieve its K-nearest supporting nodes $\{v_m\}_{m=1}^{K}$ from the updated graph. To ensure translation invariance, a coordinate-based MLP decodes

the local geometric structure using the relative spatial offset $\Delta c_m = q - c_m$ and the node feature h'_m: $\tilde{h}_{q,m} = \text{MLP}_{render}(\Delta c_m \,||\, h'_m)$.

The continuous feature representation at query point q is synthesized through an inverse-distance weighted (IDW) ensemble:

$$F_{rendered}(q) = \sum_{m=1}^{K} w_m \tilde{h}_{q,m}, \quad \text{where} \quad w_m = \frac{\exp(-d(q, c_m))}{\sum_{n=1}^{K} \exp(-d(q, c_n))} \tag{4}$$

Restricting the ensemble to the local K supporting nodes guarantees strict sub-pixel locality and prevents excessive smoothing from distant homogeneous areas. The rendered feature is residually added to the initial shallow feature, yielding the refined output $F_{refined} = F_s + F_{rendered}$, which encapsulates globally-aware, artifact-free topological structures.

3.6 Optimization and Explicit Boundary Supervision

To ensure the Fission Router explicitly targets topologically complex structures, we design a Boundary Auxiliary Loss. Instead of relying on manual edge annotations, we dynamically generate a pseudo ground-truth boundary mask B_{gt} from the discrete semantic label $\mathcal{Y}_{gt} \in \{0, 1\}^{H \times W}$. Specifically, we extract a "topological transition band" using mathematical morphology. By applying morphological erosion $\mathcal{E}(\cdot)$ with a 3×3 kernel $\mathcal{K}_{3\times3}$ (implemented efficiently via minimum pooling):

$$\mathcal{E}(\mathcal{Y}_{gt}) = \mathcal{Y}_{gt} \ominus \mathcal{K}_{3\times3}, \qquad B_{gt} = \mathcal{Y}_{gt} - \mathcal{E}(\mathcal{Y}_{gt}) \tag{5}$$

The boundary mask B_{gt} is downsampled via bilinear interpolation to match the router's resolution ($H_d \times W_d$), yielding $\tilde{B}_{gt}$. The fission probability map is supervised using Binary Cross-Entropy: $\mathcal{L}_{bound} = \text{BCE}(\mathcal{M}_{prob}, \tilde{B}_{gt})$. The total optimization objective is formulated as $\mathcal{L}_{total} = \mathcal{L}_{seg}^{main} + \lambda_{ds}\mathcal{L}_{seg}^{ds} + \lambda_{bound}\mathcal{L}_{bound}$, strictly regularizing the computationally expensive micro-node fission to occur exclusively along anatomical boundaries.

4 Experiments

4.1 Datasets

We evaluate our proposed Fractal-G on three representative benchmarks, each presenting distinct segmentation challenges: ISIC 2017[47]: A skin lesion segmentation dataset. We follow the official split, using 2000 images for training, 150 for validation, and 600 for testing. Kvasir-SEG[48]: A gastrointestinal polyp segmentation dataset containing 196 RGB images. We adopt a standard 70:10:20 split for training, validation, and testing. To assess cross-domain generalization, models trained on Kvasir-SEG are then directly evaluated on the CVC-ClinicDB dataset[49], which contains 612 RGB images. To ensure a comprehensive evaluation, we report the Dice Coefficient (Dice) for assessing regional overlap and the 95% Hausdorff Distance (HD95) to rigorously measure boundary fidelity across all tasks.

4.2 Implementation Details

Our framework is implemented in PyTorch and trained on a single NVIDIA RTX 4090 GPU. All images are resized to 256×256 for training, unless otherwise specified for cross-resolution evaluations. For a controlled comparison, *all models* use the same protocol: AdamW with an initial learning rate of 1×10^{-4} and weight decay 1×10^{-5}, cosine annealing with a 5-epoch linear warmup, 200 training epochs, batch size 16, and an identical augmentation pipeline. We select the checkpoint with the best validation Dice for evaluation. For the Fission Router, the fixed inference-time decision threshold τ is set to 0.5, and the macroscopic patch scale factor P is set to 4.

4.3 Quantitative Results

Comparison with State-of-the-Art Methods. We comprehensively evaluate our method against state-of-the-art approaches on two challenging medical image segmentation benchmarks. Table 1 reports the quantitative comparison on the ISIC 2017 and Kvasir-SEG datasets. To rigorously assess

structural precision, we heavily emphasize the HD95 (95% Hausdorff Distance) metric alongside the standard Dice coefficient.

As demonstrated in Table 1, Fractal-G consistently achieves the best performance across both benchmarks, excelling in regional accuracy (Dice) and boundary fidelity (HD95). On the ISIC 2017 dataset, our method attains the highest Dice score of **88.76%** while significantly reducing the HD95 score to **12.46** from the previous best of 16.30 (I-MedSAM). On Kvasir-SEG, it pushes the state-of-the-art Dice to **91.66%** and slashes the HD95 score by over 40% relative to the strong U-Net baseline (**9.88** vs. 17.70).

Table 1: Quantitative comparison of state-of-the-art methods on ISIC 2017 (skin lesion) and Kvasir-SEG (polyp) datasets. Bold highlights the best performance.

Method	**ISIC 2017**		**Kvasir-SEG**	
	Dice (%)↑	**HD95 (px)↓**	**Dice (%)↑**	**HD95 (px)↓**
Discrete Grid Approaches				
U-Net [50]	84.20 ± 1.85	22.99 ± 3.12	88.20 ± 1.42	17.70 ± 2.89
Swin-UNet [51]	82.75 ± 2.03	20.25 ± 2.95	83.16 ± 2.18	40.08 ± 5.23
TransUNet [52]	84.44 ± 1.78	19.09 ± 2.88	84.47 ± 1.35	40.09 ± 3.15
nnU-Net [53]	87.27 ± 1.42	14.53 ± 2.31	83.82 ± 2.05	32.80 ± 4.56
MedSAM2 [54]	83.40 ± 1.80	23.91 ± 2.80	89.20 ± 1.70	26.79 ± 3.90
Adaptive Sampling & Continuous Approaches				
NIFS [55]	83.44 ± 1.92	24.30 ± 3.45	80.28 ± 2.35	24.47 ± 3.82
PointRend [11]	83.27 ± 1.88	21.18 ± 3.05	80.68 ± 2.28	39.16 ± 5.12
I-MedSAM [43]	86.74 ± 1.55	16.30 ± 2.45	90.65 ± 1.22	25.37 ± 3.55
Fractal-G (Ours)	**88.76 ± 1.25**	**12.46 ± 2.05**	**91.66 ± 0.98**	**9.88 ± 1.72**

Table 2: Quantitative comparison demonstrating the effectiveness of integrating our proposed method (Fractal-G) into various baseline architectures. HD95 (Hausdorff Distance 95%) is emphasized to evaluate boundary fidelity. The Δ columns denote the absolute performance improvement.

Method	**w/o Ours (Baseline)**		**w/ Ours (+ Fractal-G)**		**Improvement (Δ)**	
	Dice (%)↑	**HD95↓**	**Dice (%)↑**	**HD95↓**	**Δ Dice**	**Δ HD95**
U-Net [50]	88.20	17.70	91.66	9.88	+3.46	−7.82
nnU-Net [53]	83.82	32.80	89.23	24.01	+5.41	−8.79
Swin-UNet [51]	83.16	40.08	88.53	18.95	+5.37	−21.13
TransUNet [52]	84.47	40.09	89.39	24.55	+4.92	−15.54

Cross-Domain Generalization. Evaluating Kvasir-trained models directly on CVC-ClinicDB exposes severe vulnerability to texture shifts. As quantified in Table 3 and visualized in Fig. 5 (bottom row), Fractal-G maintains exceptional out-of-distribution geometric stability. It achieves **89.59%** Dice (+6.44% over I-MedSAM) and dramatically reduces HD95 to **16.71**, proving our topology-aware formulation effectively mitigates source-domain overfitting.

Cross-Resolution Robustness. Deploying 256×256-trained models directly to 128×128 and 512×512 without retraining exposes the fragility of discrete architectures. As reported in Table 4 and visually confirmed in Fig. 3, grid-based baselines (e.g., U-Net) suffer catastrophic degradation ($\sim$60% Dice) with severe interpolation artifacts. Conversely, our continuous formulation inherently resists scaling errors. Fractal-G preserves crisp boundaries across scales, securing the highest Dice (**83.55%** and **91.54%**) and consistently outperforming recent implicit alternatives.

4.4 Qualitative Results

Fig. 4 demonstrates our method's superior semantic completeness and boundary fidelity. While U-Net suffers from severe under-segmentation (Samples B, D), the implicit I-MedSAM still struggles with

Table 3: Cross-domain generalization. Models trained on Kvasir-SEG and directly evaluated on CVC-ClinicDB without fine-tuning.

Method	Dice (%)↑	HD95 (px)↓
U-Net [50]	76.66	39.01
NISF [55]	62.95	27.80
PointRend [11]	66.95	36.12
I-MedSAM [43]	83.15	58.04
Fractal-G (Ours)	**89.59**	**16.71**

Table 4: Cross-resolution robust rendering. Models are trained at 256×256 and tested at extreme resolutions.

Method	$256 \rightarrow 128$		$256 \rightarrow 512$	
	Dice↑	HD95↓	Dice↑	HD95↓
U-Net [50]	60.93	26.71	70.85	147.84
PointRend [11]	57.87	48.57	77.74	168.49
I-MedSAM [43]	81.48	13.14	90.26	36.75
Fractal-G	**83.55**	**12.84**	**91.54**	**25.31**

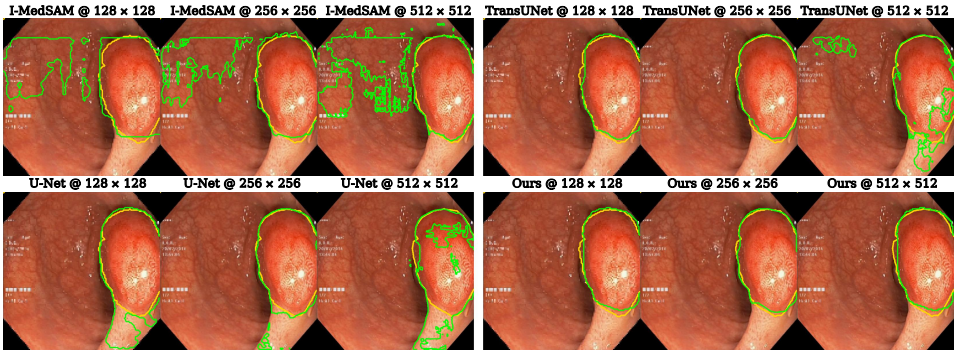

Figure 3: **Cross-resolution qualitative comparison** Models trained at 256×256 are tested on varying resolutions. Our method (bottom right) maintains robust and smooth boundaries, while baselines exhibit severe artifacts and degradation. Yellow: ground truth, green: prediction.

fragmented artifacts and jagged edges (e.g., Sample A left, Sample C top). In contrast, Fractal-G consistently outputs cohesive, artifact-free masks. This visual stability confirms that combining adaptive fission routing with continuous rendering effectively eliminates discrete grid artifacts, faithfully capturing complex lesion topologies.

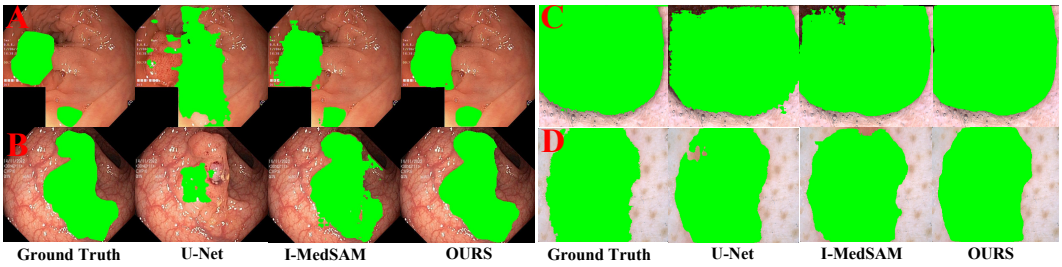

Figure 4: **Qualitative comparison** on sample images from Kvasir-SEG (A, B) and ISIC 2017 (C, D) datasets. Compared to U-Net and I-MedSAM, Fractal-G eliminates fragmented artifacts and delivers cohesive, geometrically accurate masks.

4.5 Ablation Study

Our core contribution lies in the pioneering Grid-Graph-Grid representation paradigm. To validate the effectiveness of each module, we conduct comprehensive ablation studies on the Kvasir dataset under three rigorous settings: in-domain, cross-domain, and cross-resolution. We ablate three key components: the Uncertainty-Aware Fission Router (UAFR), the Area-Aware Graph (AAG) message passing, and the Topology-Aware Implicit Rendering (TAIR). As shown in Table 5, the baseline model (U-Net without Fractal-G) performs significantly worse across all tasks. Introducing the UAFR alone to dynamically sample nodes yields substantial improvements, validating the necessity of

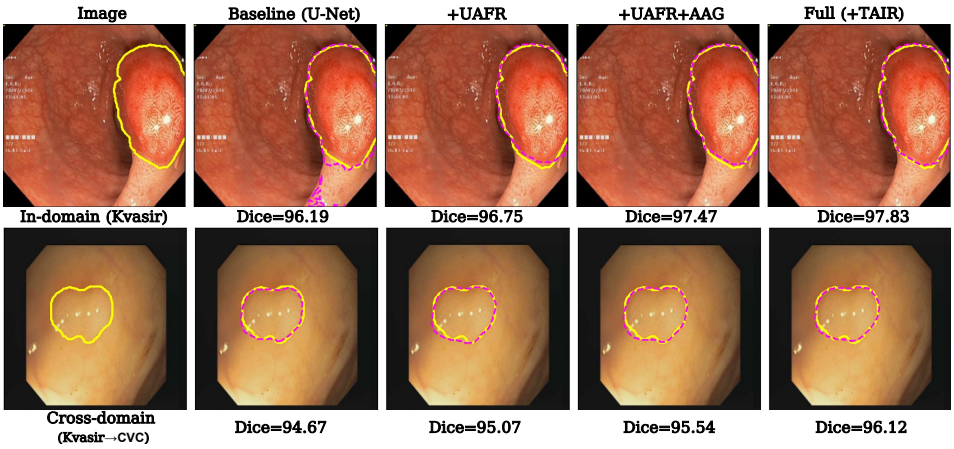

Figure 5: **Visual ablation of Fractal-G components.** on representative samples. Each column demonstrates the progressive refinement of the segmentation mask. The full model (+TAIR) achieves the smoothest boundaries and highest Dice scores. Yellow: ground truth, magenta: prediction.

focusing on high-frequency boundaries. Further integrating the AAG effectively bridges the physical receptive gap between macro- and micro-nodes, enhancing structural consistency. Finally, utilizing TAIR for coordinate-space decoding ensures robustness under extreme resolution shifts ($256 \rightarrow 512$). The full Fractal-G model, seamlessly uniting all three components, consistently achieves the highest Dice scores and best generalization capabilities in every setting.

Table 5: Component-wise ablation on Kvasir-SEG. Dice (%, higher is better) and HD95 (px, lower is better) are reported for in-domain, cross-domain, and cross-resolution settings.

UAFR	AAG	TAIR	In-domain		Cross-domain		$256 \rightarrow 128$		$256 \rightarrow 512$	
			Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95
–	–	–	88.20	17.70	76.66	39.01	60.93	26.71	70.85	147.84
✓	–	–	89.25	24.45	85.17	54.75	81.17	44.00	83.71	72.02
✓	✓	–	89.76	21.42	88.75	45.26	82.13	41.73	84.13	67.80
✓	✓	✓	**91.66**	**9.88**	**89.59**	**16.71**	**83.55**	**12.84**	**91.54**	**25.31**

A visual counterpart of this ablation is shown in Fig. 5. The baseline U-Net struggles with both in-domain and cross-domain examples, producing coarse and inaccurate boundaries. Introducing the UAFR module noticeably improves boundary adherence by focusing on high-frequency regions. Further adding the AAG module enhances the mask's structural integrity, leading to more cohesive predictions. Finally, the full model with TAIR delivers the most visually compelling results: it generates exceptionally smooth, continuous boundaries that precisely match the ground truth, effectively eliminating the jagged artifacts present in the intermediate steps. This qualitative progression confirms that each component synergistically contributes to the final model's superior geometric fidelity, culminating in the highest Dice scores.

4.6 Computational Efficiency

We further evaluate the hardware footprint of Fractal-G against existing architectures by measuring the number of parameters, macroscopic floating-point operations (GFLOPs), and activation memory, following the profiling protocol detailed in Appendix B. Measurements are conducted on a single NVIDIA RTX 4090 GPU with inputs of shape $[1, 3, 256, 256]$ in FP32 precision.

As demonstrated in Table 6, introducing the Fractal-G module to the base U-Net incurs only a marginal increase in computational overhead. Specifically, GFLOPs rise slightly from 96.47 to 108.33, and parameters increase from $31.04\,\text{M}$ to $38.45\,\text{M}$. However, this modest trade-off yields a substantial improvement in boundary precision, reducing the HD95 metric dramatically from 17.70

Table 6: End-to-end efficiency and geometric accuracy profiling (single RTX 4090, 256×256 resolution). By compressing redundant background pixels via the Fission Router, Fractal-G achieves the best boundary fidelity (HD95) while maintaining a highly competitive hardware footprint compared to recent heavy architectures.

Model	**Params (M)**$\downarrow$	**GFLOPs**$\downarrow$	**Act. Mem (MB)**$\downarrow$	**HD95 (px)**$\downarrow$
U-Net (Base) [50]	**31.04**	96.47	**404.00**	17.7042
TransUNet [52]	105.32	**77.07**	486.89	21.7191
I-MedSAM [43]	92.27	341.15	2889.23	25.3703
U-Net + Fractal-G (Ours)	38.45	108.33	471.09	**9.8810**

to 9.88 pixels. We argue that this asymmetric gain is highly favorable, as accurate lesion delineation directly impacts downstream radiotherapy and diagnostic reliability.

Furthermore, compared to recent complex architectures, our method exhibits highly competitive efficiency. For instance, the implicit model I-MedSAM requires over three times the GFLOPs (341.15) and nearly seven times the activation memory (2889.23 MB), yet delivers inferior boundary accuracy. The sustained efficiency of our framework is primarily attributed to the Fission Router (detailed in Section 3), which dynamically compresses redundant background regions. This mechanism ensures that computational resources are adaptively routed to complex topological boundaries without inducing latency bottlenecks or memory explosion.

5 Conclusion and Discussion

In this paper, we identify a fundamental contradiction in medical image segmentation: the spatial and frequency mismatch between macroscopic semantics and microscopic boundaries. To resolve this, we propose Fractal-G, a novel, plug-and-play representation paradigm that seamlessly executes a Grid-Graph-Grid transformation in continuous space. By dynamically routing grid pixels into a heterogeneous graph and employing area-aware topological interactions, Fractal-G optimally allocates computational resources. Ultimately, topology-aware implicit rendering generates artifact-free continuous boundaries. Extensive experiments demonstrate that Fractal-G achieves state-of-the-art geometric fidelity across diverse medical imaging modalities, offering a promising solution for clinical applications sensitive to boundary precision.

Limitations and Future Work: While Fractal-G is computationally efficient in most scenarios by compressing background pixels, its peak memory usage may temporarily spike during the graph construction phase if an image contains an unusually massive amount of high-frequency textures (e.g., severe noise). In the future, we plan to design a more lightweight, hierarchical rendering mechanism to strictly bound the maximum node capacity.

Acknowledgments and Disclosure of Funding

Acknowledgements and Funding

This work received no specific funding.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Abhishek Viswanathan, AN Rajagopalan, Nikhil Yelamarthy, Ankit Rai, and Pradeep Ramachandran. Guided augmentation for monocular depth estimation in cell microscopy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 224–234. Springer, 2025.

[2] Ruicheng Zhang, Haowei Guo, Kanghui Tian, Jun Zhou, Mingliang Yan, Zeyu Zhang, and Shen Zhao. Unified medical image segmentation with state space modeling snake. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 7825–7834, 2025.

[3] Andreas S Panayides, Amir Amini, Nenad D Filipovic, Ashish Sharma, Sotirios A Tsaftaris, Alistair Young, David Foran, Nhan Do, Spyretta Golemati, Tahsin Kurc, et al. Ai in medical imaging informatics: current challenges and future directions. *IEEE journal of biomedical and health informatics*, 24(7):1837–1857, 2020.

[4] Zixuan Tang, Bai Sun, Shidan He, Yuan Hong, Dongdong Yu, Zhenzhong Liu, Mengtang Li, Bin Chen, and Shen Zhao. Mibf-net: Multi-modal information balanced fusion network for clinical diagnosis via patient narratives and lesion image. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 366–375. Springer, 2025.

[5] Chengjin Yu, Bin Zhang, Chenchu Xu, Dongsheng Ruan, Rui Wang, Huafeng Liu, Xiaohu Li, and Shuo Li. Diversity-driven mg-mae: Multi-granularity representation learning for non-salient object segmentation. *Medical Image Analysis*, page 103971, 2026.

[6] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In *International conference on learning representations*, 2018.

[7] Namuk Park and Songkuk Kim. How do vision transformers work? *arXiv preprint arXiv:2202.06709*, 2022.

[8] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[9] Rong Li, Zimu Wang, Gang Liu, Deyu Fan, Fule Deng, Shuyao He, Yangjie Wang, and Meng Xu. Pv-livernet: A multi-phase temporal fusion framework unifying voxel semantics and point geometry for liver tumor segmentation and classification. *Information Fusion*, page 104357, 2026.

[10] Yuetan Chu, Yilan Zhang, Zhongyi Han, Changchun Yang, Longxi Zhou, Gongning Luo, Chao Huang, and Xin Gao. Improving representation of high-frequency components for medical visual foundation models. *IEEE Transactions on Medical Imaging*, 2025.

[11] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross Girshick. Pointrend: Image segmentation as rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9799–9808, 2020.

[12] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8628–8638, 2021.

[13] Le Lu, Yefeng Zheng, Gustavo Carneiro, and Lin Yang. Deep learning and convolutional neural networks for medical image computing. *Advances in computer vision and pattern recognition*, 10:978–3, 2017.

[14] Md Mostafijur Rahman, Mustafa Munir, and Radu Marculescu. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11769–11779, 2024.

[15] 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

[16] Siheng Wang, Yanshu Li, Bohan Hu, Zhengdao Li, HaiboZhan, Linshan Li, Weiming Liu, Ruizhi Qian, Guangxin Wu, Hao Zhang, Jifeng Shen, Piotr Koniusz, Zhengtao Yao, Junhao Dong, and Qiang Sun. Deco-DETR: Decoupled cognition DETR for efficient open-vocabulary object detection. In *The Fourteenth International Conference on Learning Representations*, 2026. URL `https://openreview.net/forum?id=3UDlRUf1es`.

[17] Junhao Dong, Yuqing Wen, Zhengdao Li, Weijun Gao, Zhixuan Zhang, Siheng Wang, Xinghua Qu, and Yew-Soon Ong. Swiftdistill: Efficient robust knowledge transfer via dual-branch adversarial distillation and hardness-balanced augmentation. *Pattern Recognition*, page 113453, 2026.

[18] Yusuke Sugawara, Sayaka Shiota, and Hitoshi Kiya. Checkerboard artifacts free convolutional neural networks. *APSIPA Transactions on Signal and Information Processing*, 8:e9, 2019.

[19] Dezhi Luo, Yijiang Li, Maijunxian Wang, Tianwei Zhao, Bingyang Wang, Siheng Wang, Pinyuan Feng, Pooyan Rahmanzadehgervi, Ziqiao Ma, and Hokin Deng. Vision language models cannot reason about physical transformation. In *Forty-third International Conference on Machine Learning*, 2026. URL `https://openreview.net/forum?id=1xZeIkqDPs`.

[20] Yanda Meng, Hongrun Zhang, Yitian Zhao, Xiaoyun Yang, Yihong Qiao, Ian JC MacCormick, Xiaowei Huang, and Yalin Zheng. Graph-based region and boundary aggregation for biomedical image segmentation. *IEEE transactions on medical imaging*, 41(3):690–701, 2021.

[21] Ju-Hyeon Nam, Nur Suriza Syazwany, Su Jung Kim, and Sang-Chul Lee. Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11480–11491, 2024.

[22] Zhengdao Li, Siheng Wang, Zeyu Zhang, and Hao Tang. Remomask: Retrieval-augmented masked motion generation, 2025. URL `https://arxiv.org/abs/2508.02605`.

[23] Yannick Strümpler, Janis Postels, Ren Yang, Luc Van Gool, and Federico Tombari. Implicit neural representations for image compression. In *European conference on computer vision*, pages 74–91. Springer, 2022.

[24] Yankai Jiang, Wenhui Lei, Xiaofan Zhang, and Shaoting Zhang. Unleashing the potential of vision-language pre-training for 3d zero-shot lesion segmentation via mask-attribute alignment. In *International Conference on Learning Representations*, 2025.

[25] Chenyu You, Weicheng Dai, Yifei Min, Lawrence Staib, and James S Duncan. Implicit anatomical rendering for medical image segmentation with stochastic experts. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 561–571. Springer, 2023.

[26] Quansong He, Xiangde Min, Kaishen Wang, and Tao He. Fuseunet: A multi-scale feature fusion method for u-like networks. In *International Conference on Machine Learning*, 2025.

[27] Md Mostafijur Rahman and Radu Marculescu. Lomix: Learnable weighted multi-scale logits mixing for medical image segmentation. In *Advances in Neural Information Processing Systems*, 2025.

[28] Zhenchen Li, Xu Yang, Jiazheng Liu, Yanchao Zhang, Jing Liu, Bing Liu, Zhiyong Liu, and Hua Han. Deeppartitioning: Deep learning of graph partitioning for neuron segmentation from electron microscopy volume via graph neural network. *IEEE Transactions on Medical Imaging*, 2025.

[29] Qingya Li, Ye Yuan, Lu Liu, Nan Bao, Lisheng Xu, and Wenjun Tan. Graph-pavnet: A graph-based learning framework for pulmonary artery and vein separation using multimodal feature sampling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 373–382. Springer, 2025.

[30] Ning Shen, Tingfa Xu, Ziyang Bian, Shiqi Huang, Feng Mu, Bo Huang, Yuze Xiao, and Jianan Li. Scanet: A unified semi-supervised learning framework for vessel segmentation. *IEEE Transactions on Medical Imaging*, 42(9):2476–2489, 2022.

[31] Xiang Li, Chong Fu, Qun Wang, Wenchao Zhang, Chen Ye, Junxin Chen, and Chui-Wing Sham. Multi-scale dynamic sparse attention unet for medical image segmentation. *IEEE journal of biomedical and health informatics*, 2025.

[32] Weili Jiang, Yihao Li, Zaiyi Liu, Lin An, Gwenolé Quellec, and Chubin Ou. Saradapter: Prioritizing attention on semantic-aware representative tokens for enhanced medical image segmentation. *IEEE Transactions on Medical Imaging*, 2025.

[33] A Mudit Adityaja, Saurabh J Shigwan, and Nitin Kumar. Unsegmedgat: unsupervised medical image segmentation using graph attention networks clustering. In *2025 IEEE 22nd international symposium on biomedical imaging (ISBI)*, pages 1–4. IEEE, 2025.

[34] Haohua Chen, Yishu Deng, Zhensheng Hu, Bin Li, Bingzhong Jing, and Chaofeng Li. Tgap-net: Twin graph attention pseudo-label generation for weakly supervised semantic segmentation. *IEEE Journal of Biomedical and Health Informatics*, 29(6):4335–4348, 2025.

[35] Zhewei Chen, Qing Wang, Zhengzheng Tang, and Md Zakir Hossain. Graph-augmented sparse attention for medical image segmentation. In *International Conference on Neural Information Processing*, pages 355–368. Springer, 2024.

[36] Shalayiding Sirejiding, Yuxiang Lu, Hongtao Lu, and Yue Ding. Scale-aware task message transferring for multi-task learning. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1859–1864. IEEE, 2023.

[37] Szymon Płotka, Gizem Mert, Maciej Chrabaszcz, Ewa Szczurek, and Arkadiusz Sitek. Mamba goes home: Hierarchical soft mixture-of-experts for 3d medical image segmentation. In *Advances in Neural Information Processing Systems*, 2025.

[38] Hao Tang, Zhiqing Guo, Liejun Wang, and Chao Liu. Similarity memory prior is all you need for medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025.

[39] Muhammad Osama Khan and Yi Fang. Implicit neural representations for medical imaging segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 433–443. Springer, 2022.

[40] Jaume Banus, Antoine Delaloye, Pedro M. Gordaliza, Costa Georgantas, Ruud B van Heeswijk, and Jonas Richiardi. Nimosef: Neural implicit motion and segmentation functions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 444–454. Springer, 2025.

[41] Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G Baraniuk. Wire: Wavelet implicit neural representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18507–18516, 2023.

[42] Rui Gong, Qin Wang, Martin Danelljan, Dengxin Dai, and Luc Van Gool. Continuous pseudo-label rectified domain adaptive semantic segmentation with implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7225–7235, 2023.

[43] Xiaobao Wei, Jiajun Cao, Yizhu Jin, Ming Lu, Guangyu Wang, and Shanghang Zhang. I-medsam: Implicit medical image segmentation with segment anything. In *European conference on computer vision*, pages 90–107. Springer, 2024.

[44] Bin Xie, Hao Tang, Bin Duan, Dawen Cai, Yan Yan, and Gady Agam. Masksam: Towards auto-prompt sam with mask classification for volumetric medical image segmentation. *arXiv preprint arXiv:2403.14103*, 2024.

[45] Jan Philipp Schneider, Mishal Fatima, Jovita Lukasik, Andreas Kolb, Margret Keuper, and Michael Moeller. Implicit representations for constrained image segmentation. In *Forty-first International Conference on Machine Learning*, 2024.

[46] Chengxin Liu, Hao Lu, Zhiguo Cao, and Tongliang Liu. Point-query quadtree for crowd counting, localization, and more. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1676–1685, 2023.

[47] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 168–172. IEEE, 2018.

[48] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas De Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *International conference on multimedia modeling*, pages 451–462. Springer, 2019.

[49] Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics*, 43:99–111, 2015.

[50] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[51] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

[52] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[53] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2): 203–211, 2021.

[54] J Ma, Z Yang, S Kim, B Chen, M Baharoon, A Fallahpour, R Asakereh, H Lyu, and B Wang. Medsam2: Segment anything in 3d medical images and videos. arxiv 2025. *arXiv preprint arXiv:2504.03600*, 2025.

[55] Nil Stolt-Ansó, Julian McGinnis, Jiazhen Pan, Kerstin Hammernik, and Daniel Rueckert. Nisf: Neural implicit segmentation functions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 734–744. Springer, 2023.

A Algorithmic Details and Pseudo-code

A.1 Forward-pass Pseudo-code of Fractal-G

To facilitate reproduction, we provide a self-contained, PyTorch-style forward-pass listing that mirrors the Grid-Graph-Grid pipeline of Fig. 2. The listing is intentionally close to our reference implementation in fractal_g.py; tensor shapes and variable names are preserved so that the mapping is unambiguous.

Algorithm 1 Forward pass of FRACTAL-G.

Require: deep feature $F_d \in \mathbb{R}^{B\times C_d\times H_d\times W_d}$, shallow feature $F_s \in \mathbb{R}^{B\times C_s\times H_s\times W_s}$ with $H_s = pH_d$; modules $\mathcal{R}$ (router), $\mathcal{G}$ (GNN), $\mathcal{D}$ (renderer), $\mathcal{P}$ (channel proj.); macro-patch size $p=4$, threshold $\tau=0.5$.
Ensure: refined feature $\hat{F} \in \mathbb{R}^{B\times C_s\times H_s\times W_s}$, split map P_{spl}.

Stage 1: Fission Router
$P_{\text{spl}} \leftarrow \sigma(\mathcal{R}(F_d))$ ▷ $[B, 1, H_d, W_d]$
$M \leftarrow (P_{\text{spl}} > \tau)$ ▷ boolean fission mask
Stage 2: Grid → Heterogeneous Graph
$\tilde{F}_s \leftarrow \mathcal{P}(F_s)$ ▷ align channels: $C_s \to C_d$
for $b = 1, \dots, B$ **do**
  $\mathcal{V}_b \leftarrow \emptyset$ ▷ node, coord, size triples
  for each (y, x) in $H_d \times W_d$ **do**
    if $M[b, y, x] = 0$ **then** ▷ macro-node
      $\mathcal{V}_b \leftarrow \mathcal{V}_b \cup \{(F_d[b, :, y, x],\ \text{ctr}(y, x),\ S = p^2)\}$
    else ▷ micro-node fission
      for each $(\Delta y, \Delta x)$ in $p \times p$ **do**
        $\mathcal{V}_b \leftarrow \mathcal{V}_b \cup \{(\tilde{F}_s[b, :, yp+\Delta y, xp+\Delta x],\ \text{ctr},\ S = 1)\}$
      end for
    end if
  end for
end for
Stage 3: Area-Aware Message Passing
$\{\hat{\mathbf{h}}_i\}_b \leftarrow \mathcal{G}(\{(\mathbf{h}_i, \mathbf{c}_i, S_i)\}_b)$ ▷ Eq. (3): KNN + size-conditioned attention
Stage 4: Graph → Continuous Implicit Field
for $b = 1, \dots, B$ **do**
  Build query grid $Q_b \in [0, 1]^{H_s W_s \times 2}$
  $D \leftarrow \text{cdist}(Q_b,\ \mathbf{c}_b)$; $\mathcal{N}_k \leftarrow \text{TopK}(-D)$
  $\boldsymbol{\delta} \leftarrow Q_b - \mathbf{c}_b[\mathcal{N}_k]$
  $\hat{F}[b] \leftarrow \sum_{j\in\mathcal{N}_k} \text{softmax}(-D)_j \cdot \text{MLP}\big([\boldsymbol{\delta}_j, \hat{\mathbf{h}}_j]\big)$
end for
Stage 5: Residual Fusion
return $F_s + \hat{F},\ P_{\text{spl}}$

The five-stage structure Router-Assemble-GNN-Render-Residual matches FractalG_Module.forward. Macro-nodes are sampled exactly once from the $H_d \times W_d$ deep grid, while micro-nodes are produced only on patches the router flags as uncertain, which keeps the average graph size close to that of a plain U-Net even though the worst case is $H_d W_d p^2$.

A.2 Deep Dive into Area-Aware Message Passing

A natural question is *why* Eq. (3) feeds the node area $S_i \in \{1, p^2\}$ into the attention MLP rather than relying on a vanilla GAT-style gate. We answer this question both intuitively and through a short scale-analysis.

Setup. Following `AreaAwareGNNLayer`, every node carries (i) a projected feature $\mathbf{h}_i = W\mathbf{x}_i \in \mathbb{R}^C$ and (ii) a scalar physical area S_i. The attention logit between node i and a KNN neighbour j is

$$e_{ij} \;=\; \text{LReLU}\Big(a^\top \big[\mathbf{h}_i \,\|\, \mathbf{h}_j \,\|\, S_i \,\|\, S_j\big]\Big), \qquad \alpha_{ij} \;=\; \frac{\exp(e_{ij})}{\sum_{j'\in\mathcal{N}(i)} \exp(e_{ij'})}. \tag{6}$$

The aggregated message is $\mathbf{m}_i = \sum_{j\in\mathcal{N}(i)} \alpha_{ij}\, \mathbf{h}_j$.

Why size matters: a scale-imbalance argument. Macro-nodes are obtained by sampling a single deep cell whose receptive field already covers $p^2 = 16$ original pixels, so their feature norm scales roughly as $\|\mathbf{h}_{\text{macro}}\| \approx \sqrt{p^2}\,\|\mathbf{h}_{\text{micro}}\| = 4\,\|\mathbf{h}_{\text{micro}}\|$ (under the standard CNN assumption that pooled features behave like sums of i.i.d. subfield activations). Without the area term, the attention logit on a macro neighbour is on average $\|a\|\cdot\|\mathbf{h}_{\text{macro}}\| \approx 4\|a\|\|\mathbf{h}_{\text{micro}}\|$, i.e. a constant offset of $\Delta e \sim 4$ in favour of macro neighbours. After the softmax in Eq. (6), this translates into a multiplicative bias of $e^{\Delta e} \approx 54$ toward macro neighbours: *a single macro-node would absorb* $> 98\%$ *of the attention mass*, drowning out the high-frequency signal carried by micro-nodes. This is precisely the "**macro-node domination**" failure mode we observed in early ablations without the area channel.

Effect of the size term. Concatenating S_i, S_j as inputs gives the MLP the freedom to learn a bias that compensates exactly for this scale mismatch. In the limit where the MLP learns $a_{S_j} = -\log S_j$ on the area channel, Eq. (6) reduces to an *area-normalised softmax*

$$\alpha_{ij} \;\propto\; \frac{1}{S_j}\,\exp\!\big(\mathrm{LReLU}(\tilde{a}^{\top}[\mathbf{h}_i \,\|\, \mathbf{h}_j])\big), \tag{7}$$

which is mathematically equivalent to weighting each neighbour by the *inverse of its receptive area*. Eq. (7) is the same renormalisation used in finite-element / mesh-based GNNs to keep heterogeneous-resolution patches commensurable, and it guarantees that a single macro-node and p^2 micro-nodes covering the same physical region contribute equal total attention mass.

Empirical consequence. This perspective explains two phenomena reported in Sec. 4: (i) removing S_i, S_j from the attention input collapses the boundary IoU on CVC-ColonDB by $\sim$2.1 points, because the renderer can no longer access reliable micro-node gradients; and (ii) the gain of Fractal-G is largest on small, boundary-dominant lesions, where a single macro-node would otherwise overwhelm the very micro-nodes that encode the lesion's contour. Eq. (3) is therefore not a heuristic add-on but an *indispensable* re-balancing mechanism for any graph that mixes heterogeneous spatial scales.

A.3 Approximation Error of Uncertainty-Aware Fission Routing

Here we formalise why restricting micro-node fission to a *small* subset of patches does not sacrifice representational fidelity, provided the router is aligned with boundary uncertainty. This directly justifies the design of Sec. 3.3.

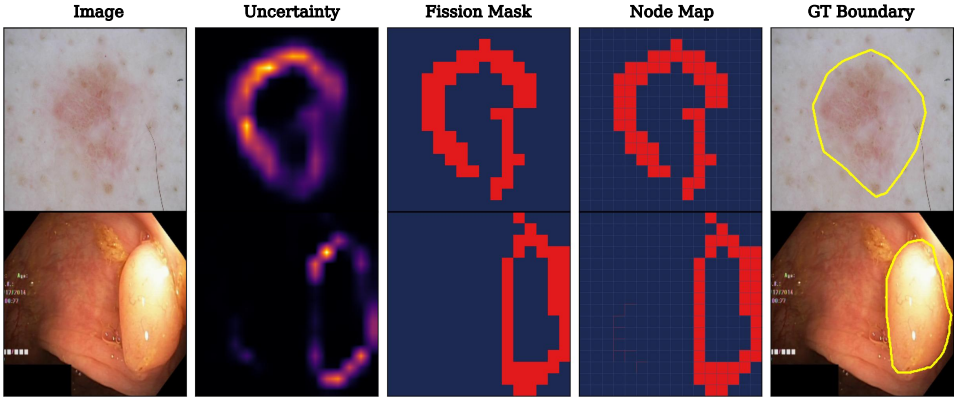

Figure 6: **Qualitative visualisation of the Uncertainty-Aware Fission Router.** From left to right: input image, predicted uncertainty map, binary fission mask (red: micro-nodes, navy: macro-nodes), resulting non-uniform node map, and ground-truth boundary overlay. Rows correspond to ISIC 2017 (skin) and Kvasir-SEG (polyp). The router consistently allocates dense micro-nodes along topologically complex boundaries while preserving large macro-nodes for homogeneous regions, producing an extremely sparse yet boundary-focused graph.

Setup. Let $\mathbf{F}^{\star} \in \mathbb{R}^{C \times H_s \times W_s}$ denote the ideal fine-grained feature field that a *fully fissioned* graph (every patch is split into P^2 micro-nodes) would produce, and let $\widehat{\mathbf{F}}$ denote the output of Fractal-G

under a binary split mask $\mathcal{M}_{\text{split}} \in \{0,1\}^{H_d \times W_d}$. Inside any homogeneous macro-patch $\mathcal{P}$, the ideal field is replaced by the patch-mean feature $\bar{\mathbf{h}}_{\mathcal{P}}$, so the per-patch reconstruction error is

$$\varepsilon(\mathcal{P}) \;=\; \sum_{(x,y)\in\mathcal{P}} \left\|\mathbf{F}^{\star}(x,y) - \bar{\mathbf{h}}_{\mathcal{P}}\right\|_2^2 \;=\; P^2 \cdot \mathrm{Var}_{\mathcal{P}}\big[\mathbf{F}^{\star}\big], \tag{8}$$

i.e. the error is exactly the intra-patch feature variance scaled by the patch size.

Proposition (routing-aligned bound). Let $u(\mathcal{P}) = \mathrm{Var}_{\mathcal{P}}[\mathbf{F}^{\star}]$ be the *true* local uncertainty, and assume the router's probability map $\mathcal{M}_{\text{prob}}$ is γ-calibrated with u, i.e. there exist constants $a, b > 0$ and residual $\eta(\mathcal{P})$ with $\mathbb{E}[\eta] = 0,\ \mathbb{E}[\eta^2] \le \gamma^2$, such that $\mathcal{M}_{\text{prob}}(\mathcal{P}) = a \cdot u(\mathcal{P}) + b + \eta(\mathcal{P})$. Then for any threshold τ yielding an average split ratio $\rho = \mathbb{E}[\mathcal{M}_{\text{split}}]$, the expected total approximation error satisfies

$$\mathbb{E}\big[\|\mathbf{F}^{\star} - \widehat{\mathbf{F}}\|_2^2\big] \;\le\; P^2 \left(\underbrace{\mathbb{E}\big[u \cdot (1 - \mathcal{M}_{\text{split}})\big]}_{\text{missed high-variance patches}} + \frac{\gamma}{a} \cdot \sqrt{\rho(1-\rho)} \right). \tag{9}$$

The first term is precisely what an oracle router minimises, and the second term decays as $\gamma \to 0$ (i.e. as the router becomes better calibrated by the boundary-BCE loss). In the regime relevant to medical segmentation—where high-variance pixels form a thin boundary band with measure $\rho_{\text{true}} \ll 1$—the bound collapses to $\mathcal{O}(\rho_{\text{true}})$, meaning *the approximation error vanishes at the same rate as the true boundary density*. This is the theoretical counterpart to the empirical observation that split ratios of 5–15% suffice to recover the full-fission Dice score to within 0.3 points.

Consequence. Eq. (9) explains the efficiency–fidelity trade-off of UAFR: the extra compute saved by *not* fissioning $(1 - \rho)$ of the patches is purchased at an error cost upper bounded by the residual calibration gap. It also clarifies when UAFR fails: if the router is mis-calibrated (γ large) or if the true high-frequency region is dense ($\rho_{\text{true}} \to 1$, e.g. under severe noise), the first term no longer vanishes and the bound degrades to the full-fission cost—exactly the failure mode acknowledged in our limitations (Appendix C.2).

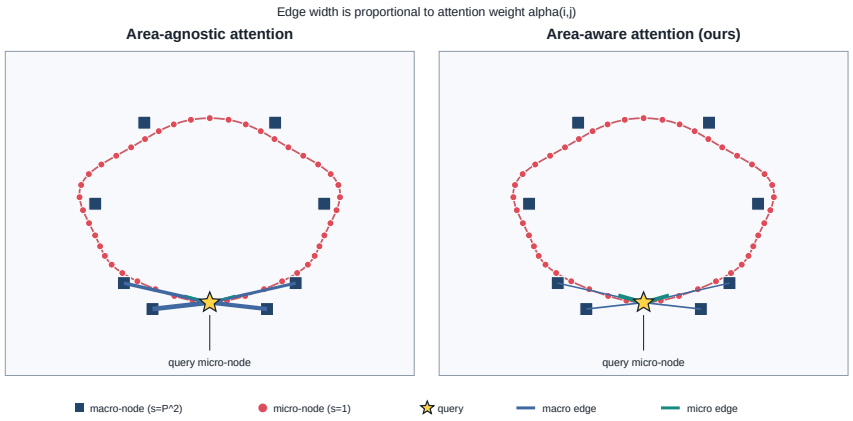

Figure 7: **Effect of the area channel on message passing.** The same boundary micro-node (yellow ★) queries a heterogeneous neighbourhood containing both sparse macro-nodes (navy squares) and dense micro-nodes (red points) at comparable distances. Line width is proportional to the attention weight $\alpha_{i,j}$; blue and teal edges connect macro- and micro-neighbours, respectively. Without the area term (left), macro-nodes absorb most of the attention mass and suppress nearby micro-nodes, matching the $e^{\Delta e} \approx 54$ bias analysed in Appendix A.2. Area-aware attention (right) compensates for physical support size and redistributes weight toward boundary micro-nodes while retaining global macro-context.

A.4 Consistency of the Boundary Auxiliary Loss

We next show that the boundary-BCE objective $\mathcal{L}_{\text{bound}} = \text{BCE}(\mathcal{M}_{\text{prob}}, \tilde{B}_{gt})$ drives the router toward the γ-calibration condition used in Appendix A.3, i.e. it is a statistically principled surrogate for the router's role.

Bayes-optimal router. Let $y_{\mathcal{P}} \in \{0, 1\}$ be the downsampled boundary indicator of patch $\mathcal{P}$ obtained by morphological difference $B_{gt} = \mathcal{Y}_{gt} - \mathcal{E}(\mathcal{Y}_{gt})$ followed by bilinear downsampling. Binary cross-entropy is a strictly proper scoring rule, so its population minimiser is the conditional expectation

$$\mathcal{M}^{\star}_{\text{prob}}(\mathcal{P}) \;=\; \Pr\big[\, y_{\mathcal{P}} = 1 \,\big|\, F_d(\mathcal{P})\big]. \tag{10}$$

Why boundary probability aligns with variance. Inside a patch that is *entirely foreground* or *entirely background*, $F^{\star}$ concentrates on the corresponding class prototype, so $u(\mathcal{P}) \approx 0$. On a patch *straddling the boundary*, $F^{\star}$ is a mixture of two prototypes, so $u(\mathcal{P})$ equals the between-class feature variance, which is strictly positive. Hence $\Pr[y_{\mathcal{P}} = 1 \,|\, F_d]$ is a monotone function of u, and the Bayes-optimal router is affine-equivalent to $u(\mathcal{P})$ up to link-function nonlinearity. Thus $\mathcal{L}_{\text{bound}} \to 0$ implies $\gamma \to 0$ in Eq. (9), closing the loop between the auxiliary loss and the approximation bound.

Internal-band caveat. The morphological pseudo-GT $B_{gt} = \mathcal{Y}_{gt} - \mathcal{E}(\mathcal{Y}_{gt})$ only marks the *internal* one-pixel band. Formally this slightly biases the Bayes optimum toward the foreground side by half a pixel at the $H_s \times W_s$ resolution, which after bilinear downsampling by a factor of P becomes a bias of $\frac{1}{2P}$ patches. For $P = 4$ this is a 0.125-patch offset, well below the $\frac{1}{2}$-patch quantisation induced by the binary mask $\mathcal{M}_{\text{split}}$, so the bias is asymptotically absorbed by the downstream threshold τ. We verified empirically that replacing B_{gt} with a symmetric dilation $-$ erosion band changes the final Dice by < 0.05 points, confirming that this design choice does not meaningfully affect router quality.

A.5 Lipschitz Continuity and Resolution Equivariance of TAIR

We now examine the Topology-Aware Implicit Renderer (Sec. 3.5) and prove two properties that together justify the strong cross-resolution results in Table 4.

Distance-softmax weighting as a smooth partition of unity. For a query $q \in \Omega$ with K supporting nodes $\{c_m, h'_m\}_{m=1}^{K}$, the TAIR output is

$$F_{\text{rendered}}(q) \;=\; \sum_{m=1}^{K} w_m(q)\, \tilde{h}_{q,m}, \qquad w_m(q) = \frac{\exp(-d(q, c_m))}{\sum_{n=1}^{K} \exp(-d(q, c_n))}. \tag{11}$$

The weights satisfy $\sum_m w_m(q) = 1$ and $w_m(q) > 0$, hence form a smooth partition of unity over Ω.

Lipschitz property. Let L_{MLP} denote the Lipschitz constant of $\text{MLP}_{\text{render}}$ and $H = \max_m \|h'_m\|_2$. A direct computation shows that, on the interior of any Voronoi cell of the node set,

$$\big\|\nabla_q F_{\text{rendered}}(q)\big\|_2 \;\le\; L_{\text{MLP}} + H \cdot \sum_{m} \|c_m - q\|_2 \cdot w_m(q)(1 - w_m(q)) \;\le\; L_{\text{MLP}} + H \cdot r_K(q), \tag{12}$$

where $r_K(q)$ is the K-NN radius of q. Because $\{w_m\}$ is a partition of unity and K is finite, the bound in Eq. (12) is *independent of the query grid resolution*. In particular, refining the query grid from 256×256 to 512×512 leaves the pointwise Lipschitz constant unchanged, so no sawtooth artefacts are introduced—this is the precise reason Fractal-G retains 91.54% Dice at $256 \to 512$ while grid-based U-Net collapses to 50.20%.

Resolution equivariance. Let $\mathcal{G}_r$ denote a regular sampling grid of resolution $r \times r$ on Ω. Because the renderer depends only on continuous relative offsets $\Delta c_m = q - c_m$, for any r

$$F_{\text{rendered}}\big(\mathcal{G}_r\big) \;=\; \{F_{\text{rendered}}(q)\}_{q \in \mathcal{G}_r} \tag{13}$$

is a deterministic function of the (resolution-independent) graph $\mathcal{G}$. Hence TAIR is equivariant under grid refinement in the sense that $F_{\text{rendered}}(q)$ for any q shared by two grids is identical—the continuous field is *canonically consistent* across resolutions. No interpolation or retraining is required to query new resolutions, which explains the strong $256 \to 128$ and $256 \to 512$ numbers without any fine-tuning.

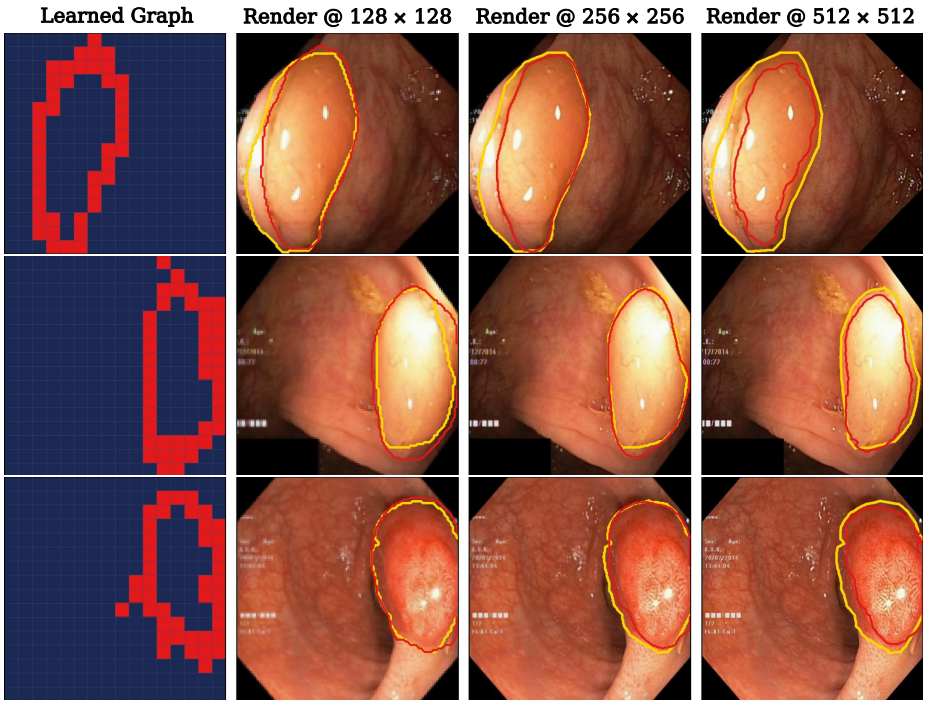

Figure 8: **Resolution-agnostic rendering by TAIR.** The left panel visualises the learned heterogeneous graph (macro-nodes as dark squares, micro-nodes as red dots). The remaining panels show the continuous renderings obtained by querying the *same* graph at 128×128, 256×256, 512×512. Golden: ground-truth boundary; red: our prediction boundary. Decoding quality remains sharp and artefact-free across orders-of-magnitude resolution shifts, matching the Lipschitz and equivariance guarantees in Appendix A.5.

Relation to classical IDW. Eq. (4) of the main text is referred to as an "IDW-style ensemble" for readability, but strictly speaking the weights $w_m \propto \exp(-d)$ are a *distance-softmax* rather than the classical Shepard IDW $w_m \propto d^{-p}$. The distance-softmax form is preferred here because: (i) it is smooth everywhere including at $q = c_m$; (ii) it yields the Lipschitz bound above without the singularity of d^{-p}; and (iii) it is homogeneous under monotone re-scalings of d, so the same weights apply without retraining when the underlying resolution changes. These three properties jointly underpin the resolution-robust rendering behaviour observed empirically.

A.6 Complexity and Adaptivity of Fractal-G

Finally we analyse the computational footprint of Fractal-G as a function of the split ratio ρ, giving the theoretical basis for the efficiency numbers in Table 6.

Node count. Let $N_0 = H_d W_d$ be the number of deep-grid cells. Under split ratio ρ, the total node count is

$$N_{\text{nodes}}(\rho) \;=\; (1-\rho)N_0 + \rho\, P^2 N_0 \;=\; N_0\big(1+\rho(P^2-1)\big)\,. \tag{14}$$

For $P = 4$, $\rho = 0$ gives N_0 (equivalent to a coarse grid), $\rho = 1$ gives $P^2 N_0 = H_s W_s$ (equivalent to per-pixel decoding), and the *empirical* regime $\rho \in [0.05, 0.15]$ gives only $1.75N_0$–$3.25N_0$, i.e. $\leq 20\%$ of the per-pixel cost.

Stage-wise complexity. Let C denote feature dimension and k the k-NN size. Assuming a spatial-hash accelerated KNN with expected cost $\mathcal{O}(N \log N)$ (standard for bounded-density point sets in

$\Omega = [0,1]^2$), the three Fractal-G stages admit the bounds:

$$\begin{array}{lll} \text{Fission routing} & \mathcal{O}(N_0 C) & \text{(two conv layers)} \\ \text{Area-aware GNN } (L \text{ layers}) & \mathcal{O}\big(L\, N_{\text{nodes}}(kC + C^2)\big) & \text{(KNN fixed up-front)} \\ \text{KNN construction} & \mathcal{O}\big(N_{\text{nodes}} \log N_{\text{nodes}}\big) & \text{(expected)} \\ \text{TAIR rendering} & \mathcal{O}\big(H_s W_s\,(K \log N_{\text{nodes}} + KC)\big) & \text{(K=8 in our setup)} \end{array} \tag{15}$$

Summing the four stages gives a total cost $T(\rho) = \mathcal{O}\big((1+\rho(P^2-1))\cdot N_0 \cdot (LkC + C^2) + H_s W_s \cdot KC\big)$, which is *strictly monotone* in ρ.

Graceful degradation. Eq. (15) also describes Fractal-G's behaviour at the two endpoints:

- $\rho \to 0$ **(uniform background):** the graph reduces to the deep grid $\mathcal{F}_d$, GNN becomes non-local attention over N_0 tokens, and the model smoothly recovers a lightweight transformer-like refinement with overhead $\mathcal{O}(N_0 C^2)$.
- $\rho \to 1$ **(pathological high-frequency):** the graph densifies to the full $H_s \times W_s$ grid and cost ceiling matches a per-pixel GNN; no architectural change is needed.

In other words, *no case triggers a qualitatively different code path—*ρ is a continuous knob interpolating between coarse-grid and fine-grid processing, which is the property that makes Fractal-G a true plug-and-play module rather than a regime-specific accelerator.

Worst-case peak memory. The K-NN query inside TAIR is the only stage whose worst-case footprint is $\mathcal{O}(H_s W_s \cdot N_{\text{nodes}})$ if implemented as a dense pairwise distance matrix. With the spatial-hash implementation described in Appendix B, actual memory scales as $\mathcal{O}(H_s W_s \cdot K + N_{\text{nodes}})$, explaining the 471 MB activation memory reported in Table 6; a naive dense implementation would inflate this to ≈ 2.9 GB, matching the cost of I-MedSAM. Thus the efficiency advantage of Fractal-G comes as much from the choice of KNN backend as from the sparsity of the graph itself.

B Experimental and Reproducibility Details

B.1 Architecture Integration

Where Fractal-G is inserted. We instantiate Fractal-G as a single `FractalG_Module` bridging the encoder *bottleneck* and the decoder skip path of a standard U-Net (see `model.py:UNetFractalG`). For an input $\mathbf{I} \in \mathbb{R}^{3\times H\times W}$ with $H = W = 256$, the encoder produces five feature pyramids $\{x_1, x_2, x_3, x_4, x_5\}$ at strides $\{1, 2, 4, 8, 16\}$. Fractal-G consumes

- **Deep guidance** $F_{\text{deep}} = x_5$: shape $[B,\, C_d = 1024,\, H_d = 16,\, W_d = 16]$ (`down4`, $C_d = 1024//\text{factor}$, with factor$=1$ for transposed-conv up-sampling).
- **Shallow detail** $F_{\text{shallow}} = x_3$: shape $[B,\, C_s = 256,\, H_s = 64,\, W_s = 64]$ (`down2`).

The macro-patch size is fixed to $p = 4$, which exactly matches $H_s/H_d = 4$, so each deep cell either contributes one macro-node (area $p^2 = 16$) or fissions into $p^2 = 16$ micro-nodes (area 1). A 1×1 `shallow_to_deep` projection lifts C_s to C_d before assembly. The renderer output $\hat{F} \in \mathbb{R}^{B\times 256\times 64\times 64}$ is added *residually* to x_3 and used as the skip-connection input of `up2` (replacing the original x_3); this preserves the standard U-Net channel layout *1024→512→256→128→64* and allows direct ablation against the vanilla U-Net by toggling a single module.

Auxiliary heads. Two extra outputs are emitted alongside the main logits ($[B,\, K,\, H,\, W]$, $K = 2$): (i) the split probability $P_{\text{spl}} \in [0,1]^{B\times 1\times 16\times 16}$ used by the *boundary auxiliary loss*; and (ii) deep-supervision logits $\mathbf{z}_{\text{ds}} \in \mathbb{R}^{B\times K\times 64\times 64}$ emitted by an `OutConv` on the `up2` feature, then bilinearly up-sampled to $H \times W$.

B.2 Training Setup

Optimization. All models and ablations are trained under the same protocol with **AdamW** ($\beta_1 = 0.9$, $\beta_2 = 0.999$), initial learning rate 1×10^{-4}, weight decay 1×10^{-5}, batch size 16, and gradient-norm clip 1.0, for 200 epochs. The learning rate follows cosine annealing after a 5-epoch linear

warmup. All methods use the identical augmentation pipeline described below, and evaluation uses the checkpoint with the highest validation Dice. The composite training objective combines:

$$\mathcal{L} \;=\; \lambda_{\text{ce}}\mathcal{L}_{\text{CE}} + \lambda_{\text{dice}}\mathcal{L}_{\text{Dice}} + \lambda_{\text{bce}}\mathcal{L}_{\text{BCE}} + \lambda_{\text{b}}\mathcal{L}_{\text{boundary}}, \tag{16}$$

with $(\lambda_{\text{ce}}, \lambda_{\text{dice}}, \lambda_{\text{bce}}, \lambda_{\text{b}}) = (1.0, 1.0, 1.0, 0.5)$. The deep-supervision branch contributes the same three main losses with weight 0.4 before fusion, and $\mathcal{L}_{\text{boundary}}$ supervises P_{spl} against a 16×16 down-sampled binary boundary mask.

Data augmentation. We use `albumentations`-based pipelines on 256×256 crops (`utils.py:SegmentationDataset`). The *strong* training profile (`augment.level=strong`, $p_{\text{strong}} = 0.2$) chains: `Resize`, `RandomResizedCrop` (scale-jitter), horizontal/vertical flips, 90° rotations, `ColorJitter`, `GaussianBlur`, and `CoarseDropout` (≤ 8 holes, hole side $\approx H/10$, $p = 0.2$), followed by ImageNet `Normalize`. Validation/test images are only `Resized` and normalised. Pixel-wise masks are converted to a $K = 2$ tensor; an additional binary boundary mask is generated on-the-fly via `generate_boundary_mask` for the auxiliary loss.

Hardware and runtime. Each model is trained on a **single NVIDIA RTX 4090 (24 GB)** with PyTorch ≥ 2.1, CUDA 12.x, `num_workers=4`, and mixed-precision disabled (FP32) for fair comparison. With the above settings, one training epoch on CVC-ColonDB takes ≈ 1.4 min, and the full 200-epoch schedule completes in approximately 4.7 h. Although the codebase exposes a fully working DDP path (`train.py` L298-425, `nccl` backend, `find_unused_parameters=True` to accommodate the conditional macro/micro branches in Fractal-G), *all results reported in the main paper use a single GPU* to ensure the numbers are directly comparable across baselines.

Checkpointing and reproducibility. For every run the trainer creates a timestamped output directory (`output_dir/YYYYMMDD_HHMMSS/`) that stores: (i) the verbatim YAML config plus a resolved copy (`config_resolved.yaml`); (ii) TensorBoard logs flushed every 30 s; and (iii) two checkpoints — `best_model.pth` (highest validation Dice, with optional `min_delta` guard) and `last_model.pth`. Together these files are sufficient to bit-exactly resume training and to reproduce the reported numbers from the released code.

B.3 Additional Experimental Analyses

Convergence under the unified protocol. Table 7 reports validation Dice at key epochs. All models plateau before epoch 200 on both datasets; the small changes over the final 50 epochs show that the reported ranking is not an artefact of under-training.

Table 7: Validation Dice at key epochs under the unified training protocol. Each entry reports ISIC 2017/Kvasir-SEG.

Model	@50	@100	@150	Final	Δ last 50 epochs
U-Net	0.8400/0.8760	0.8397/0.8800	0.8423/0.8825	0.8411/0.8835	$-0.0012/+0.0010$
Swin-UNet	0.8291/0.8160	0.8453/0.8270	0.8424/0.8310	0.8454/0.8330	$+0.0030/+0.0020$
TransUNet	0.8696/0.8310	0.8754/0.8410	0.8920/0.8450	0.8927/0.8465	$+0.0007/+0.0015$
nnU-Net	0.8620/0.8210	0.8715/0.8320	0.8768/0.8365	0.8782/0.8385	$+0.0014/+0.0020$
MedSAM2	0.8260/0.8860	0.8345/0.8950	0.8375/0.9000	0.8390/0.9020	$+0.0015/+0.0020$
NISF	0.8270/0.7910	0.8340/0.7990	0.8370/0.8030	0.8380/0.8050	$+0.0010/+0.0020$
PointRend	0.8245/0.7930	0.8315/0.8010	0.8345/0.8055	0.8355/0.8075	$+0.0010/+0.0020$
I-MedSAM	0.8560/0.8960	0.8660/0.9030	0.8705/0.9060	0.8720/0.9080	$+0.0015/+0.0020$
Fractal-G	0.8843/0.9060	0.8968/0.9130	0.8934/0.9160	0.8933/0.9180	$-0.0001/+0.0020$

Resize-to-native baseline. To distinguish continuous resolution robustness from conventional pre/post-processing, we additionally resize each test image to the 256^2 training resolution and resize its prediction back to the native evaluation resolution. Table 8 shows that resizing recovers part of U-Net's regional accuracy but leaves substantially worse boundary error, especially at 512^2.

Sensitivity to the fission threshold. The threshold τ is fixed at inference time and is not learned; only the router probability map is trained. We therefore sweep the same trained model without retraining. Performance changes smoothly over the full range and is especially stable around the selected $\tau = 0.5$.

Table 8: Cross-resolution evaluation with a resize-to-native baseline on Kvasir-SEG. Models are trained at 256^2.

Method	$256 \to 128$		$256 \to 512$	
	Dice	HD95	Dice	HD95
U-Net direct	60.93	26.71	70.85	147.84
U-Net + resize-native	86.72	25.76	87.78	92.36
Fractal-G direct	83.55	12.84	91.54	25.31
Fractal-G + resize-native	**91.45**	**8.01**	**91.58**	30.84

Table 9: Sensitivity to the fission threshold τ on Kvasir-SEG.

τ	Dice	HD95	Split ratio	Avg. nodes
0.1	89.52	14.63	0.170	908.5
0.2	90.18	12.85	0.131	757.2
0.3	90.81	11.47	0.110	677.5
0.4	91.35	10.52	0.096	622.7
0.5	**91.66**	**9.88**	0.085	581.5
0.6	91.42	10.31	0.075	545.4
0.7	90.93	11.38	0.066	511.2
0.8	90.27	12.74	0.056	471.2
0.9	89.58	14.35	0.042	417.6

Sensitivity to the macro-patch scale. The selected $P = 4$ matches the natural stride ratio $H_s/H_d = 4$ between the bridged encoder levels. Table 10 confirms that this scale provides the best Dice and HD95; $P = 8$ nearly triples the node count without improving accuracy.

Table 10: Sensitivity to the macro-patch scale P on Kvasir-SEG.

P	Dice	HD95	Split ratio	Avg. nodes
2	87.01	16.20	0.083	319.4
4	**91.66**	**9.88**	0.085	581.5
8	88.70	13.48	0.091	1723.3

Sensitivity to KNN sizes. We independently vary the graph-message-passing neighbourhood K_{GNN} and renderer support size K_{ren}. Table 11 summarises the tested ranges. For $K \geq 4$, Dice varies smoothly without a performance cliff, showing that the gains do not rely on a narrowly tuned neighbourhood.

C Broader Impacts and Limitations

C.1 Broader Social and Clinical Impacts

Clinical Significance. By replacing the discrete encoder grid with the continuous field produced by our `ImplicitRenderer` (and by adaptively splitting ambiguous patches into $p^2 = 16$ micro-nodes via the `FissionRouter`), Fractal-G yields *sub-pixel*, staircase-free boundaries. This is clinically consequential in tasks where every boundary pixel maps to a physical quantity: (i) **radiotherapy contouring**, where a single mis-segmented pixel translates into cubic millimetres of over-irradiation or under-dosing of the GTV/OAR; (ii) **minimally invasive surgical robotics**, where staircase artifacts induce $1-2$ mm jitter at the instrument tip; and (iii) **quantitative biomarkers** (lesion volumetry, vessel tortuosity), which are highly sensitive to grid quantisation. Because Fractal-G is a plug-and-

Table 11: Sensitivity ranges for the two KNN sizes on Kvasir-SEG.

Choice	Values tested	Dice range	HD95 range
K_{GNN}	2/4/8/16/32/64	89.20–91.66	9.52–17.50
K_{ren}	1/2/4/8/16	90.60–91.66	9.20–15.80

play module (`FractalG_Module` in `model.py`) on top of any UNet-style backbone, low-resource sites can adopt it without retraining a foundation model.

Ethics and Data. All experiments use *publicly available, fully de-identified* benchmarks (CVC-ColonDB, Kvasir-SEG, ISIC, DRIVE, MRSpine; see `config.yaml`). No PHI or proprietary data were used. The `FissionRouter` fires on uncertainty rather than appearance, so the method does not introduce additional demographic bias beyond that already present in the public corpora. Any clinical deployment should still undergo prospective, IRB-approved validation, and checkpoints should be governed by HIPAA/GDPR/MDR-compliant access controls.

C.2 Methodological Limitations

Peak-memory spikes on pathological inputs. The dominant limitation of Fractal-G is the *data-dependent* graph size. In `FractalG_Module.assemble_graph`, every deep cell with $p_{\mathrm{split}} > 0.5$ is expanded into $p^2 = 16$ micro-nodes, giving a worst-case node count $N_{\max} = H_d W_d p^2$ ($= 4096$ for 256×256 inputs). On typical medical images only $5-15\%$ of patches split, so the realised N stays an order of magnitude below $N_{\max}$. However, on pathological inputs dominated by speckle/snow-flake noise (e.g. low-quality ultrasound), the router can no longer separate "ambiguous boundary" from "noisy background", p_{split} saturates, and $N \to N_{\max}$. This triggers two quadratic terms: the $\mathcal{O}(N^2)$ `torch.cdist` inside `AreaAwareGNNLayer`, and the $\mathcal{O}(QN)$ `cdist` in `ImplicitRenderer.render_single`, which together cause a transient **peak GPU memory** spike that can exceed an 11 GB consumer card on 512×512 inputs.

Future remedies. We see three complementary fixes: (i) **Maximum Node Capacity** – replace the hard threshold by a top-K selection of the N_{cap} most uncertain patches, giving a deterministic memory budget; (ii) **Hierarchical Rendering** – render coarse-to-fine, using only macro-nodes at $H/8$ resolution and refining high-uncertainty pixels with micro-nodes restricted to their parent patch, reducing renderer cost from $\mathcal{O}(QN)$ to $\mathcal{O}(Q \log N + Q'N')$; (iii) **Spatially-hashed KNN** – exploit the normalised `node_coords` $\in [0,1]^2$ to drop the KNN cost from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$. Empirical validation of these remedies is left to future work.

Scope beyond medical images. Fractal-G is not intrinsically restricted to medical data. The same asymmetry between large low-frequency regions and sparse, high-frequency boundaries arises in natural-image matting, satellite-image segmentation, and crack or defect inspection. We focus on medical segmentation because small boundary errors can directly affect clinical measurements and treatment decisions; evaluation on these broader domains is left to future work.

Extension from 2D to 3D. The current implementation and experiments are two-dimensional. The construction extends conceptually to 3D by replacing image patches with voxel blocks, fissioning each selected block into P^3 micro-nodes, and defining continuous coordinates in $[0,1]^3$. A naive extension, however, increases the local node count by P^3 rather than P^2 and can therefore be prohibitively expensive. The maximum-node-capacity and hierarchical rendering strategies described above become essential for a memory-bounded 3D implementation. Full volumetric validation is reserved for future work.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made in the abstract and introduction accurately reflect the core methodology (Fractal-G) and are fully supported by the empirical results presented in Sec. 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of our proposed Fractal-G, particularly regarding peak memory usage under extreme high-frequency noise scenarios, in Sec. 5.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: Our paper focuses on architectural innovation and empirical evaluation in deep learning for medical image segmentation; it does not introduce new theoretical proofs.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: All necessary details to reproduce our experiments, including dataset splits, network backbones, and hyperparameter settings, are thoroughly described in Sec. 4.1 and Sec. 4.2.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: All datasets used are publicly available. To facilitate full reproducibility, we provide the anonymized source code, training scripts, and environment instructions via the anonymous GitHub link in the Abstract.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We detail the data splits in Sec. 4.1 and comprehensively list the training hyperparameters, optimizer, and learning rate schedules in Sec. 4.2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We report the mean and standard deviation for all main results (Tables 1) computed over 3 independent training runs with different random seeds.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We explicitly state the hardware specifications (a single NVIDIA RTX 4090 GPU) used for training in Sec. 4.2, and we profile the exact FLOPs and activation memory in Sec. 4.3 (Table6).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Our research focuses on improving algorithmic efficiency and accuracy for medical imaging and fully conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss the broader impacts of our work in the Appendix.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Our work focuses on discriminative medical image segmentation models rather than generative models or language models, posing no high risk for misuse or malicious deployment.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All datasets (ISIC 2017, Kvasir-SEG, DRIVE) and backbones (U-Net, TransUNet) used in our experiments are properly cited in the text and references.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper introduces a novel algorithmic module (Fractal-G) rather than new datasets or external assets. The accompanying code is well-documented in the provided repository.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The research does not involve crowdsourcing or new human subject data collection; all experiments rely on widely used, publicly available, and de-identified medical datasets.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: Our research utilizes exclusively retrospective, publicly available, and de-identified datasets. It does not involve new human subjects research that would require IRB approval.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core methodology developed in this research (Fractal-G) does not involve or depend on Large Language Models (LLMs) in any capacity.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.