Neumours: Implicit Neural Tumour Representations for Differentiable Medical Rendering

1st Kyle Lukaszek
School of Computer Science, CCMPS
University of Guelph
Guelph, Canada
klukasze@uoguelph.ca

2nd Yuyao Wu
Master of Data Science (MDS)
University of Guelph
Guelph, Canada
ywu58@uoguelph.ca

3rd Kasra Fard
School of Engineering
University of Guelph
Guelph, Canada
rahnamak@uoguelph.ca

Abstract—**Medical image analysis faces a fundamental disconnect: deep learning models excel at diagnostic tasks but operate in offline pipelines, separate from the interactive visualization that professionals require. We propose to bridge this gap by replacing heavy convolutional backbones with a compact Implicit Neural Representation (INR) embedded within a real-time graphics pipeline. Unlike distillation-based approaches, our architecture is trained directly on multi-modal MRI data using a coordinate-based Multi-Layer Perceptron (MLP) augmented with Fourier positional encodings, which takes spatial coordinates and voxel intensities as input to predict segmentation probabilities. To balance neural evaluation cost with responsive interaction, we implement a dynamic volumetric caching strategy using the Slang shading language: rather than evaluating the INR per-ray at runtime, we compute the implicit function into a voxel grid in a single pass. The rendering pipeline is implemented with automatic differentiation support, establishing infrastructure for future clinician-guided refinement. We evaluate on the BraTS 2023+ Adult Glioma dataset, achieving 0.823 Whole Tumour Dice (92% of nnU-Net) with 118× fewer parameters and 20× faster inference. Our results demonstrate that pure INRs can provide clinically useful segmentations while enabling interactive visualization infeasible with traditional pipelines.**

Index Terms—**Computer Graphics, Machine Learning, Medical Rendering**

I. INTRODUCTION

Accurate and timely tumour segmentation from multi-modal Magnetic Resonance Imaging (mpMRI) scans (typically including T1c, T1n, T2w, and FLAIR modalities) is essential in modern clinical workflows, guiding treatment planning and assessment using criteria like RANO [16]. While manual segmentation is notoriously challenging due to complexity and high inter-rater variability, deep learning methods have achieved state-of-the-art diagnostic accuracy. In particular, CNN-based architectures like **nnU-Net** [12] have consistently demonstrated superior performance in challenges such as **BraTS** [15, 18, 3, 2], often outperforming even newer and more complex models, solidifying their status as the definitive baseline for medical segmentation [11].

Despite this computational success, a fundamental gap persists in clinical practice. These high-performing segmentation algorithms operate offline, generating discrete outputs like voxel grids or explicit meshes. This not only introduces latency into the clinical workflow, but also requires storage for high-resolution volumes. While recent methods like Marching Neu-rons have advanced the state-of-the-art by enabling analytic mesh extraction from neural representations to avoid sampling artifacts [23], the extraction step itself remains a bottleneck. We propose to bypass this discrete conversion entirely by rendering the learned representation directly.

To address this, we turn to Implicit Neural Representations (INRs), which model an input's properties (normally a image/scene) as a continuous function [26]. While INRs have often been used to compress existing representations, we propose utilizing them as a primary segmentation model. We introduce a pure INR architecture that maps spatial coordinates (x, y, z) and local mpMRI intensity directly to segmentation labels. To capture high-frequency anatomical details, our network uses Fourier-feature positional encodings [25], allowing the model to learn complex boundaries without the massive parameter count of a full CNNs model.

We implement this pipeline using the Slang shading language [9] and its automatic differentiation extension, SLANG.D [4]. However, rendering INRs via per-pixel inference is computationally expensive. We therefore employ a **dynamic caching strategy**: the INR acts as the continuous "source of truth" which is sampled into a volume texture for visualization. This decouples inference from rendering, ensuring interactive framerates (60+ FPS). The rendering pipeline is implemented with full automatic differentiation support, enabling future extensions where clinician annotations could propagate gradients back to update the INR.

We leverage the clinically relevant multi-modal **Adult Glioma Dataset** from the **BraTS 2023+** challenges for training, and the **University of Missouri Post-operative Glioma Dataset (MU-Glioma-Post)** [29] for testing. We benchmark our pure INR against the industry gold-standard **nnU-Net**. While we aim for competitive segmentation accuracy (Dice), our primary contribution is the shift in computational paradigm: moving from static, offline predictions, to a live, differentiable representation that supports rapid updates.

II. BACKGROUND AND RELATED WORK

A. State-of-the-Art in Brain Tumour Segmentation

The automated segmentation of brain tumours, particularly gliomas assessed via the multi-modal Brain Tumor Image Segmentation Benchmark (BraTS), has largely been dominated by

deep learning methodologies, offering efficiency and superior performance compared to conventional approaches [7, 1]. The foundation of this field is the U-Net architecture, noted for its encoder-decoder architecture and essential skip connections that preserve fine-grained spatial information crucial for accurate delineation in medical images [28].

Since 2018, the nnU-Net framework has represented a paradigm shift in this domain, demonstrating that the self-configuring methodology, determining optimal preprocessing, network architecture, and training schedules, is more critical for achieving state-of-the-art performance than novel architecture design alone [12, 11]. Consequently, nnU-Net has served as the foundational benchmark and the backbone for winning solutions in numerous BraTS iterations (2020, 2021, 2022, 2023) [7]. While research continues to explore new models, including Transformer-based networks like SwinUNETR [8], and State Space Models such as SegMamba [27], competitive solutions often rely on ensemble approaches combining the complementary strengths of these architectures with refined data augmentation techniques [7].

These ensemble approaches are exemplified by the introduction of VISTA3D, a unified 3D foundation segmentation model designed to integrate both highly accurate automatic segmentation and interactive editing capabilities [10]. Notably, VISTA3D utilizes the SegResNet CNN model, a U-Net type architecture, as its shared backbone, underlining the continued reliance on established convolutional architectures [10]. Due to its inherent robustness, consistency, and proven ability to deliver highly competitive results, often matching or exceeding newer models in stringent benchmarks [11], the nnU-Net framework is retained in this work as the established and meticulously validated segmentation approach, providing the perfect benchmark for our approach.

B. Neural Implicit Representations for Medical Imaging

Implicit Neural Representations (INRs), or Neural Fields, have emerged as a paradigm shift from discrete grid-based data structures. An INR parametrizes a signal, whether visual, auditory, or geometric, as a continuous function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, typically approximated by some Multilayer Perceptron (MLP) as popularized by Xie et. al [26]. While initially popularized for novel view synthesis via Neural Radiance Fields (NeRF) [20], the utility of INRs has rapidly expanded to medical imaging due to their resolution independence and extreme memory efficiency. Not long after NeRF, Tancik et al. demonstrated that INRs trained as **Fourier Feature Networks** [25] can capture high frequency information, and specifically demonstrated mpMRI reconstruction as a proof-of-concept.

In the medical domain, INRs were initially applied to reconstruction specific tasks, such as accelerating MRI acquisition or sparse-view reconstruction. Neural Sparse Voxel Fields (NSVF) [17] demonstrated that combining sparse voxel structures with implicit functions could significantly speed up inference while retaining high fidelity. More recently, the focus has shifted toward semantic segmentation specifically. Stolt-Ansó et al. introduced Neural Implicit Segmentation Functions (NISF) [24], showing that INRs could act as continuous segmentation priors capable of handling complex topologies without the discretization artifacts inherent in voxel grids.

A major challenge in utilizing INRs is visualizing them. Most existing pipelines rely on extracting an explicit mesh from the implicit field using algorithms like Marching Cubes, or more recently, analytic approaches such as *Marching Neurons* [23] or Sharp & Jacobson's range-analysis techniques in *Spelunking the Deep* [22]. However, mesh extraction can be computationally expensive and often introduce some topological errors when handling complex geometry or segmentation bounds. On the other hand, evaluating an INR for every sample along a ray during rendering is often prohibitively slow. To balance these constraints, our work uses the INR as a continuous "source of truth" that is discretized to the resolution of the MRI scan being evaluated. This approach preserves the ability to query the segmentation at arbitrary resolution (super-resolution) while leveraging the reliable speed of standard 3D voxel rendering.

C. Differentiable Rendering and Automatic Differentiation in Graphics

Automatic Differentiation (AD) is the algorithmic backbone of deep learning, allowing for the precise evaluation of derivatives through complex graphs [5, 19]. When applied to computer graphics, this yields what is called **Differentiable Rendering**, where image pixels are differentiable with respect to scene parameters.

In medical imaging, methods like *DiffUS* [6] and *DiffVox* [21] have applied differentiable rendering to ultrasound and X-ray simulation, respectively. To bring these capabilities to general MRI visualization, we leverage **Slang** [9], a shading language with native AD support [4]. By utilizing Slang and its Python interop layer, SlangPy [14], we embed our model within our runtime to allow gradients to flow back from scene parameters, enabling optimization loops directly within the visualization pipeline.

III. METHODOLOGY

Our approach bridges offline medical image segmentation with real-time interactive rendering through a three-stage pipeline. First, we implement an interactive visualization framework using the Slang shading language. Second, we train an INR directly on mpMRI data. Third, we connect these components through a dynamic caching strategy. The INR acts at the super-resolution source of truth, sampled into a voxel grid for rendering. This decouples neural inference from already costly raymarching, ensuring interactive framerates.

A. Stage 1: Interactive Multi-Modal Rendering Framework

We first validate that volumetric rendering of segmentation masks is tractable at interactive rates. The renderer, implemented in Slang, raymarches through an axis-aligned bounding box enclosing the brain volume.
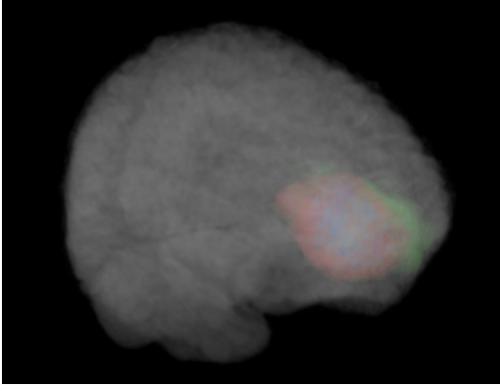
Fig. 1. FLAIR + ground truth visualized using our differentiable renderer

For each sample along the ray, we accumulate color C and transmittance T via standard volume rendering:

$$\alpha_i = 1 - \exp(-\sigma_i \cdot \Delta t) \qquad (1)$$
$$C \leftarrow C + \alpha_i \cdot T \cdot c_i \qquad (2)$$
$$T \leftarrow T \cdot (1 - \alpha_i) \qquad (3)$$

where Δt is the step size, σ_i the opacity (derived from intensity or label lookup), and c_i the color at sample i. Rays terminate when $T < 0.01$

MRI intensities are trilinearly interpolated, windowed, and gamma-corrected before contributing to c_i. Segmentation labels use nearest-neighbor sampling to preserve hard edges, with class colours and opacities from a fixed lookup table. The four modalities (T1n, T1c, T2w, T2FLAIR) can be toggled and weighted independently.

Performance exceeds 30FPS at 1080p on commodity hardware, confirming the pipeline can support interactive clinical review.

B. Stage 2: Training the INR

Our architecture departs from standard coordinate-based MLPs by splitting the input pathway into two specialized heads before fusion. Figure 2 demonstrates our simple architecture.

Spatial Head: Normalized coordinates $(x, y, z) \in [-1, 1]^3$ pass through Fourier feature encoding [25]:

$$\gamma(\mathbf{x}) = [sin(\pi k \mathbf{x}), cos(\pi k \mathbf{x})]_{k=1}^{K}$$

where $K = 64$ frequencies. The raw coordinates concatenate with these features, yielding a $3 + 6K = 387$-dimensional input. A single linear layer projects this to half the hidden dimension.

Modality Head: rather than sampling a single voxel intensity, we extract a $3 \times 3 \times 3$ patch centered at each query coordinate across all four MRI modalities. A 3D convolution with valid padding collapses this $(3, 3, 3, 4)$ tensor to a vector, capturing local texture and inter-modality relationships. This head produces the other half of the hidden dimension.
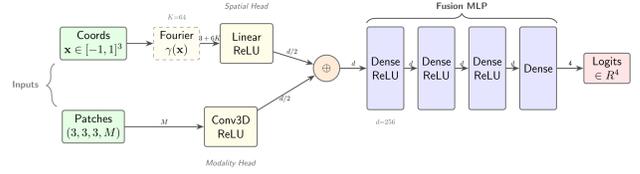

Fig. 2. Visual diagram of the simple model architecture

Fusion and Prediction: The two embeddings concatenate and pass through a 4-layer MLP with ReLU activations, outputting class logits for background, necrotic core, edema, and enhancing tumour.

Loss Function: We combine weighted cross-entropy with soft Dice:

$$\mathcal{L} = (1 - \lambda) \cdot \mathcal{L}_{CE} + \lambda \cdot (1 - \text{Dice}_{\text{mean}})$$

with $\lambda = 0.5$. Class weights $(2, 1, 1, 1)$ upweight background to discourage over-segmentation, which is the dominant failure mode for coordinate-based models lacking spatial inductive bias.

Sampling Strategy: Each training step samples $16,384$ coordinates across the volume. A foreground ratio parameter biases sampling toward tumour voxels, addressing the severe class imbalance that BraTS is famous for. Gradient accumulation over 4 micro-batches of $4,096$ coordinates enables large effective batch sizes within our memory constraints.

Optimization: We train two variants to compare optimizer dynamics. The first uses the classic AdamW optimizer, and the second uses Muon, a second-order optimizer that applies Newton-Schulz orthogonalization to weight matrices, while falling back to AdamW for non-matrix parameters (biases, conv kernels) [13]. Both use cosine decay from 10^{-3} to 5×10^{-4} over $10,000$ steps with 100 warmup steps and gradient clipping at norm 1.0.

C. Stage 3: Renderer Integration

1) Plugging in the INR: Volume Caching: Rather than evaluating the INR per-ray at render time, we compute the full segmentation volume in a single inference pass and upload it to GPU memory as a 3D texture. This decouples neural inference from raymarching: the INR runs once per case ($\sim$2.8s), producing a voxel grid that the renderer samples at interactive rates. The volume is packed into a GPU buffer for direct access during raymarching.

2) End-to-End Differentiability: Our use of Slang's autodiff enables gradients to flow from scene parameters back through to network weights. Every core rendering function is tagged as `[Differentiable]`, establishing infrastructure for future annotation-driven model refinement. We leave implementation of the interactive update loop to future work to constrain project scope.

IV. Evaluation

A. Datasets

We train on **BraTS 2023+ Adult Glioma** [15, 18, 3, 2] and test on **MU-Glioma-Post** [29], a deliberate domain shift from pre-operative to post-treatment imaging.

BraTS 2023+ provides 1,251 pre-operative glioma cases from 19 institutions. Each case contains four co-registered MRI sequences (T1n, T1c, T2w, T2-FLAIR) at $1mm^3$ isotropic resolution. Ground truth annotations mark three nested regions: necrotic core (NCR), peritumoral edema (ED), and enhancing tumour (ET). Annotations were refined by board-certified neuroradiologists with 15+ years of glioma identification experience. We use a 90/10 train/validation split.

MU-Glioma-Post offers 594 longitudinal timepoints across 203 patients, all acquired *after* surgical resection and therapy. Same four MRI modalities, same $1mm^3$ resolution, but the imaging context differs substantially. The post-op scans include resection cavities, treatment effects, and mixed 1.5T/3T acquisitions from a single institution. This contrasts sharply with BraTS's multi-site, pre-operative, 3T data.

This pairing tests generalization under realistic conditions: can a model trained on pristine pre-operative imaging handle the messier reality of post-treatment follow ups? **MU-Glioma-Post** remains fully held out, with no cases appearing during training or validation.

B. Evaluation Metrics

1) Segmentation Quality (Dice Similarity Coefficient): To quantify diagnostic accuracy, we employ the Dice Similarity Coefficient (DSC). For a predicted segmentation mask P and the ground truth expert annotation G, the DSC is defined as:

$$DSC(P, G) = \frac{2|P \cap G|}{|P| + |G|} \quad (4)$$

We utilize this metric to evaluate both **Class Accuracy** (specific tumour sub-regions like the enhancing tumour and necrotic core) and **Total Coverage Accuracy** (the union of all tumour sub-regions, often referred to as Whole Tumour).

In the context of early screening for gliomas, these metrics are critical for two reasons. *Total Coverage* validates the model's sensitivity, ensuring the system successfully flags the full spatial extent of the tumour, minimizing the risk of false negatives during a rapid review. Meanwhile, *Class Accuracy* measures the fidelity of internal boundaries (e.g., distinguishing the enhancing core from edema). High fidelity here is critical for workflow efficiency. It provides the clinician with a granular starting point, requiring them to only refine minor edge cases rather than manually segmenting complex internal structures from scratch.

2) Inference Latency and Model Performance: A critical requirement for clinical adoption is inference speed compatible with interactive workflows. We measure end-to-end latency from volume load to complete segmentation prediction on consumer hardware (Apple M4 Pro, 24GB unified memory) using Metal Performance Shaders (MPS) for GPU acceleration.

Our INR architecture aims to process full-resolution BraTS volumes ($240 \times 240 \times 155$ voxels, approximately 8.9 million unique points) in 3 ± 0.5 seconds per case. This latency includes: (1) loading four MRI modalities from disk, (2) padding and pre-processing for the 3D convolution head, (3) chunked dense inference across the depth axis, and (4) argmax prediction assembly.

The chunked inference strategy (Section III-C) processes 4–8 axial slices per forward pass, balancing memory consumption against recompilation overhead. This yields approximately 3×10^6 voxels per second throughput. While not yet achieving real-time framerates for continuous interaction, this latency is perfect for case-by-case review workflows where a clinician can open a case and have it ready for analysis within seconds after the MRI data has been processed.

C. Denoising

We evaluate our model with and without denoising applied to each predicted slice. This is done to compare whether or not denoising negatively impacts the model's predicted tumour boundaries, or has a positive impact on our results at little to no cost. The denoising task is cheap, and was measured to take around 0.28 ± 0.08 seconds per volume but could still be further optimized.

D. Hyperparameter Selection

We performed grid search over key hyperparameters using 10-fold cross-validation on BraTS 2023. Table I summarizes the final configuration.

TABLE I
SELECTED HYPERPARAMETERS

Parameter	Value
Fourier frequencies	64
Hidden dimensions	[256, 256, 256, 256]
Learning rate	$10^{-3} \to 5 \times 10^{-4}$ (cosine decay)
Warmup steps	100
Training steps	10,000
Batch size	16,384 (4×4,096 accumulation)
Dice weight (λ)	0.5
Class weights	[2.0, 1.0, 1.0, 1.0]
Foreground sampling ratio	0.1
Gradient clip norm	1.0

Key findings from ablations: (1) 64 Fourier frequencies outperformed 32 and 128, balancing expressiveness against overfitting; (2) four hidden layers of 256 units matched deeper/wider variants at lower parameter cost; (3) upweighting background ($2\times$) reduced over-segmentation; (4) low foreground ratio (0.1) improved boundary precision by exposing the model to more background content during training.

V. Results

A. Training Performance

Table II shows that AdamW achieves marginally superior training metrics with a mean Dice of 0.905 compared to Muon's 0.888. The largest gap appears in the necrotic segmentation (0.816 vs 0.777), the smallest and most challenging

TABLE II
TRAINING METRICS AT BEST CHECKPOINT (STEP 8,993)

Region	Dice		CE	
	AdamW	Muon	AdamW	Muon
Background (Class 0)	0.995	0.994	0.007	0.008
Necrotic Core (Class 1)	0.816	0.777	0.913	1.047
Peritumoral Edema (Class 2)	0.898	0.881	0.265	0.297
Enhancing Tumor (Class 3)	0.910	0.900	0.279	0.314
Mean	0.905	0.888	0.366	0.416
Combined Loss	0.072 / 0.083			

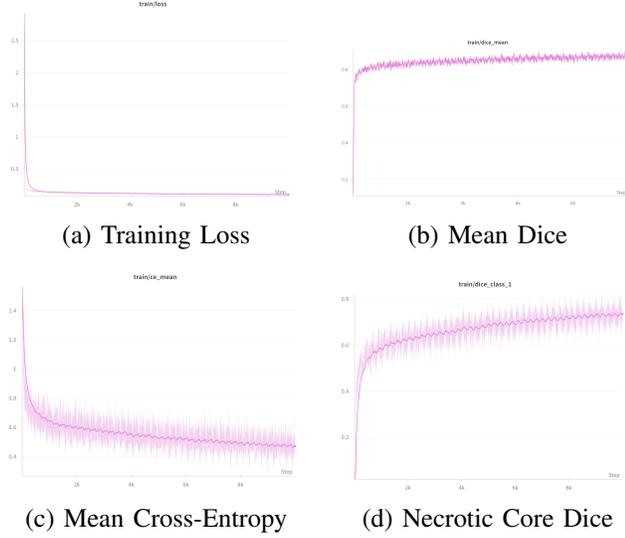tumour sub-region, which also exhibits the highest cross-entropy loss for both optimizers.

(a) Training Loss

(b) Mean Dice

(c) Mean Cross-Entropy

(d) Necrotic Core Dice

Fig. 3. AdamW Training Curves
*EMA=0.90 Training curves for 10000 steps using the AdamW optimizer.

(a) Training Loss

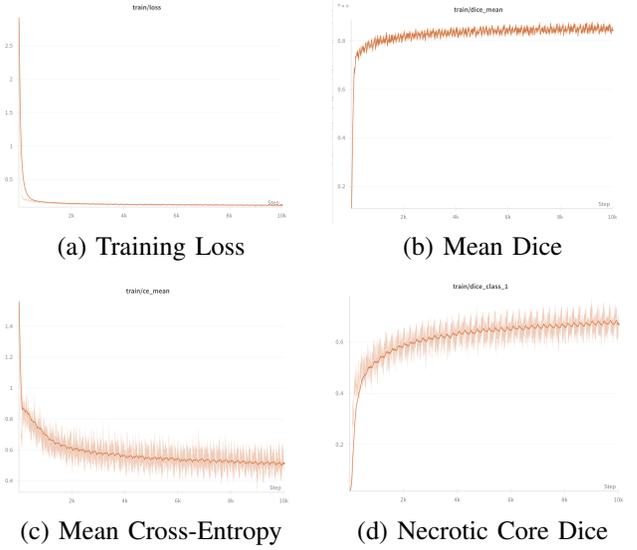(b) Mean Dice

(c) Mean Cross-Entropy

(d) Necrotic Core Dice

Fig. 4. Muon Training Curves
*EMA=0.90 Training curves for 10000 steps using the Muon optimizer.
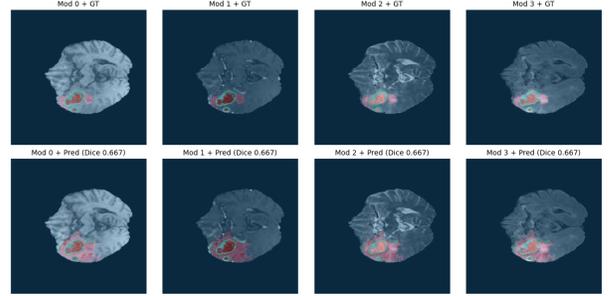
Fig. 5. BraTS 2023 Validation Fold Side-by-Side Comparison (Case 1229)

The similarity between the two optimizers can be observed in figures 3 and 4, where the only noticeable difference appears early on in the mean cross-entropy plot (c). One note of interest is that both AdamW and Muon shared the same best step at step $8,993$, though Muon was still slightly worse. In the future, we would like to train for longer, since we can see that the cross-entropy and mean dice were still very slowly ascending for both optimizers. For the rest of our results section, we will only be focusing on our AdamW trained model, as it is undoubtedly the best of the two.

B. Validation Segmentation Accuracy

Table III compares our INR against nnU-Net on the BraTS 2023 validation split ($n = 126$). We report both raw predictions and results after morphological denoising (connected component filtering and hole filling).

nnU-Net achieves superior accuracy across all tumour regions, with Whole Tumour Dice of 0.892 compared to our raw 0.767. However, post-processing substantially improves our results: denoised Whole Tumour Dice reaches 0.823 (92% of nnU-Net). The gap is most pronounced for necrotic core (0.789 vs 0.378), where nnU-Net's convolutional inductive bias pro-

vides substantial advantage for detecting small, irregularly-shaped regions.

TABLE III
VALIDATION RESULTS (BRATS 2023, $n = 126$)

Region	nnU-Net	INR	INR (Denoised)
Necrotic Core	0.789 ± 0.160	0.397 ± 0.253	0.378 ± 0.269
Peritumoral Edema	0.698 ± 0.281	0.647 ± 0.207	0.718 ± 0.193
Enhancing Tumor	0.803 ± 0.206	0.722 ± 0.233	0.778 ± 0.207
Whole Tumor	0.892 ± 0.119	0.767 ± 0.176	0.823 ± 0.138

Our denoised INR achieves competitive performance on edema (0.718 vs 0.698, exceeding nnU-Net) and enhancing tumour (0.778 vs 0.803, 97% of nnU-Net). The model appears to not be overfit on the training data, as it does a respectful job classifying the withheld validation data relative to the training results. This suggests the Fourier-feature coordinate encoding captures sufficient spatial information for these larger structures, while the lack of local connectivity priors produces noisy predictions that benefit from post-processing. We can see this in figure 5 where the model successfully isolates the enhancing tumour, but overpredicts the surrounding impacted regions.
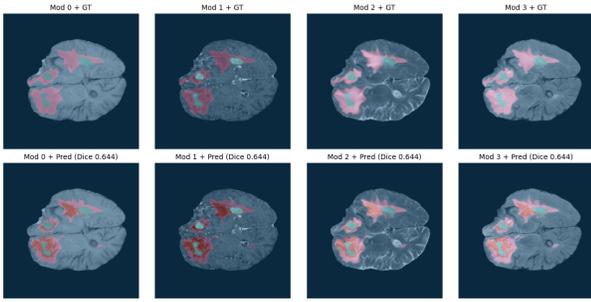
Fig. 6. MU-Glioma-Post Side-by-Side Comparison (Patient 47, Timepoint 1)

C. Domain Shift to Post-Operative Imaging

We use the MU-Glioma-Post test set ($n = 594$) to evaluate generalization to a substantially different imaging context: post-surgical, longitudinal follow up-scans from a single institution. As expected, performance degrades across all tumour classes (Table IV).

TABLE IV
TEST RESULTS (MU-GLIOMA-POST, $n = 594$)

Region	nnU-Net	INR (Raw)	INR (Denoised)
Necrotic Core	0.693 ± 0.238	$0.102 \pm 0.186^*$	$0.103 \pm 0.211^*$
Peritumoral Edema	0.275 ± 0.364	0.591 ± 0.241	$\mathbf{0.652 \pm 0.235}$
Enhancing Tumor	0.475 ± 0.321	0.491 ± 0.311	0.492 ± 0.355
Whole Tumor	$\mathbf{0.783 \pm 0.213}$	0.653 ± 0.235	0.693 ± 0.221

*59.4% of cases have no necrotic core present.

The most dramatic decline occurs in necrotic core detection (0.102 ± 0.186). This is anticipated: post-operative patients have undergone surgical resection, and 59.4% of test cases contain no necrotic core in the ground truth. nnU-Net maintains strong NCR detection (0.693) due to its convolutional priors, while our coordinate-based model struggles without explicit spatial connectivity.

Notably, our denoised INR *exceeds* nnU-Net on peritumoral edema (0.652 vs 0.275), a $2.4\times$ improvement. Both models show comparable performance on enhancing tumour (0.492 vs 0.475). This suggests Fourier features capture diffuse edema patterns effectively, while nnU-Net's localized receptive fields may be less suited to the irregular post-operative edema distributions.

Overall, nnU-Net maintains higher Whole Tumour coverage (0.783 vs 0.693), but our INR provides a $\sim20\times$ inference speedup (~2.8s vs ~54s per case) with competitive or superior performance on specific tumour classes.

D. Model Efficiency

TABLE V
MODEL COMPARISON

Model	Parameters	Inference	Whole Tumor	Hardware
nnU-Net 3D	~31M	53.9s	0.892	MPS GPU
Our INR	0.26M	2.8s	0.823	MPS GPU

Parameter reduction: $118\times$. Speedup: $19\times$. INR results with denoising.
*Inference time evaluated on M4 Pro using Torch MPS and JAX-Metal.

Our architecture achieves $118\times$ parameter reduction compared to nnU-Net, resulting in a nearly $250\times$ decrease in model memory size, and a $20.73\times$ speedup to inference time when tested on a modern Macbook Pro M4 Pro GPU.

Our model's compact size enables deployment on most consumer hardware. The entire parameter set is so small that it fits comfortably in L2 cache on modern processors or directly within GPU memory, without taking up much precious VRAM. This eliminates model related memory bandwidth issues, and only really leaves us disk read/write optimization to worry about aside from other avenues such as model quantization (which is not explored in this paper).

E. Rendering Performance

Our differentiable raymarcher maintains display refresh rate (60+ FPS) during interactive volume exploration with step sizes up to 0.02 on our testing hardware (M4 Pro). The rendering pipeline itself introduces negligible latency, and the only blocking operation occurs during model load, when INR inference must complete before the segmentation volume is available for visualization. Once cached, the volume renders at interactive rates identical to the ground truth segmentation masks. Loading the model + performing inference to produce our discretized segmentations and caching them is roughly in-line with the inference times seen earlier in our results.
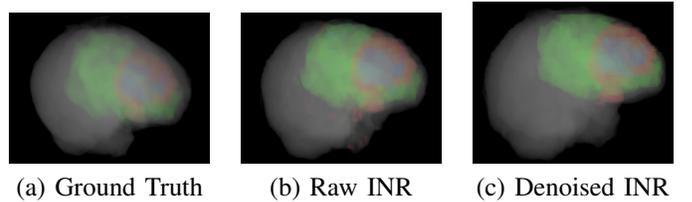

(a) Ground Truth   (b) Raw INR   (c) Denoised INR

Fig. 7. Comparison between rendered volumes with ground truth segmentations, and INR segmentations (raw and denoised)

In figure **7(b)** we can observe that there are stray speckles (noise) within the volume, but in **7(c)** the noise is completely eliminated and the predictions closely resemble the ground truth segmentations.

F. Limitations

Several limitations warrant discussion here. First, absolute segmentation accuracy lags behind nnU-Net (0.823 vs 0.892 Whole Tumour Dice after denoising), with the most significant gap in necrotic core detection. Our goal is not to match these results but to demonstrate that pure INR architectures can provide clinically useful baseline segmentations suitable for interactive refinement, with dramatically faster inference.

Second, while the ~2.8-second inference latency provides a $20\times$ speedup over nnU-Net, it does not yet support real-time interaction during volume manipulation. The dynamic caching strategy addresses this by decoupling inference from rendering, but sub-second updates would require additional optimization.

Third, the domain shift results highlight that models trained exclusively on pre-operative BraTS data require fine-tuning or

domain adaptation for post-treatment imaging contexts. The post-operative setting introduces artifacts (resection cavities, treatment effects) absent from training data.

VI. CONCLUSION

We presented an Implicit Neural Representation for brain tumour segmentation achieving 0.823 Whole Tumour Dice (92% of nnU-Net) with $118\times$ fewer parameters and $20\times$ faster inference on consumer hardware. The architecture combines Fourier-encoded spatial coordinates with a lightweight 3D convolution for local texture, proving that compact coordinate-based models can learn clinically useful tumour boundaries from multi-modal MRI.

We embed this model within a Slang-based real-time ray-tracer with full automatic differentiation support, establishing infrastructure for future clinician-in-the-loop refinement. Future work includes: (1) Hausdorff distance loss modelling, (2) sub-second inference via quantization, or architecture search/refinement, (3) implementing annotation-driven gradient updates through the renderer, (4) domain adaptation for post-operative imaging, and (5) clinical usability evaluation.

REFERENCES

[1] Claudia Takyi Ankomah et al. *How We Won BraTS-SSA 2025: Brain Tumor Segmentation in the Sub-Saharan African Population Using Segmentation-Aware Data Augmentation and Model Ensembling*. 2025. eprint: arXiv:2510.03568.

[2] Ujjwal Baid et al. *The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification*. 2021. eprint: arXiv:2107.02314.

[3] Spyridon Bakas et al. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features". In: *Scientific Data* 4.1 (Sept. 2017). ISSN: 2052-4463. DOI: 10.1038/sdata.2017.117. URL: http://dx.doi.org/10.1038/sdata.2017.117.

[4] Sai Bangaru et al. "SLANG.D: Fast, Modular and Differentiable Shader Programming". In: *ACM Transactions on Graphics (SIGGRAPH Asia)* 42.6 (Dec. 2023), pp. 1–28. DOI: 10.1145/3618353.

[5] Atilim Baydin et al. "Automatic differentiation in machine learning: A survey". In: *Journal of Machine Learning Research* 18 (Apr. 2018), pp. 1–43.

[6] Noe Bertramo, Gabriel Duguey, and Vivek Gopalakrishnan. *DiffUS: Differentiable Ultrasound Rendering from Volumetric Imaging*. 2025. eprint: arXiv:2508.06768.

[7] André Ferreira et al. *How we won BraTS 2023 Adult Glioma challenge? Just faking it! Enhanced Synthetic Data Augmentation and Model Ensemble for brain tumour segmentation*. 2024. eprint: arXiv:2402.17317.

[8] Ali Hatamizadeh et al. *Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images*. 2022. arXiv: 2201.01266 [eess.IV]. URL: https://arxiv.org/abs/2201.01266.

[9] Yong He, Kayvon Fatahalian, and Theresa Foley. "Slang: language mechanisms for extensible real-time shading systems". In: *ACM Trans. Graph.* 37.4 (July 2018). ISSN: 0730-0301. DOI: 10.1145/3197517.3201380. URL: https://doi.org/10.1145/3197517.3201380.

[10] Yufan He et al. *VISTA3D: A Unified Segmentation Foundation Model For 3D Medical Imaging*. 2024. arXiv: 2406.05285 [cs.CV]. URL: https://arxiv.org/abs/2406.05285.

[11] Fabian Isensee et al. *nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation*. 2024. eprint: arXiv:2404.09556.

[12] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* 18.2 (Dec. 2020), pp. 203–211. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01008-z. URL: http://dx.doi.org/10.1038/s41592-020-01008-z.

[13] Keller Jordan et al. *Muon: An optimizer for hidden layers in neural networks*. 2024. URL: https://kellerjordan.github.io/posts/muon/.

[14] Simon Kallweit et al. *SlangPy*. Version 0.38.0. https://github.com/shader-slang/slangpy. 2025.

[15] Alexandros Karargyris et al. "Federated benchmarking of medical artificial intelligence with MedPerf". In: *Nature Machine Intelligence* 5.7 (July 2023), pp. 799–810. ISSN: 2522-5839. DOI: 10.1038/s42256-023-00652-2. URL: http://dx.doi.org/10.1038/s42256-023-00652-2.

[16] Hongwei Bran Li et al. *The Brain Tumor Segmentation (BraTS) Challenge 2023: Brain MR Image Synthesis for Tumor Segmentation (BraSyn)*. 2024. arXiv: 2305.09011 [eess.IV]. URL: https://arxiv.org/abs/2305.09011.

[17] Lingjie Liu et al. *Neural Sparse Voxel Fields*. 2020. eprint: arXiv:2007.11571.

[18] Bjoern H. Menze et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)". In: *IEEE Transactions on Medical Imaging* 34.10 (2015), pp. 1993–2024. DOI: 10.1109/TMI.2014.2377694.

[19] Bart van Merriënboer et al. *Automatic differentiation in ML: Where we are and where we should be going*. 2019. arXiv: 1810.11530 [cs.LG]. URL: https://arxiv.org/abs/1810.11530.

[20] Ben Mildenhall et al. *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*. 2020. eprint: arXiv:2003.08934.

[21] Mohammadhossein Momeni et al. *Differentiable Voxel-based X-ray Rendering Improves Sparse-View 3D CBCT Reconstruction*. 2024. eprint: arXiv:2411.19224.

[22] Nicholas Sharp and Alec Jacobson. "Spelunking the Deep: Guaranteed Queries on General Neural Implicit Surfaces via Range Analysis". In: *ACM Trans. Graph.* 41.4 (July 2022). ISSN: 0730-0301. DOI: 10.1145/3528223.3530155. URL: https://doi.org/10.1145/3528223.3530155.

[23] Christian Stippel et al. *Marching Neurons: Accurate Surface Extraction for Neural Implicit Shapes*. 2025. arXiv: 2509.21007 [cs.GR]. URL: https://arxiv.org/abs/2509.21007.

[24] Nil Stolt-Ansó et al. *NISF: Neural Implicit Segmentation Functions*. 2023. arXiv: 2309.08643 [eess.IV]. URL: https://arxiv.org/abs/2309.08643.

[25] Matthew Tancik et al. *Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains*. 2020. arXiv: 2006.10739 [cs.CV]. URL: https://arxiv.org/abs/2006.10739.

[26] Yiheng Xie et al. *Neural Fields in Visual Computing and Beyond*. 2022. arXiv: 2111.11426 [cs.CV]. URL: https://arxiv.org/abs/2111.11426.

[27] Zhaohu Xing et al. *SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation*. 2024. arXiv: 2401.13560 [cs.CV]. URL: https://arxiv.org/abs/2401.13560.

[28] Guoping Xu et al. *Is the medical image segmentation problem solved? A survey of current developments and future directions*. 2025. arXiv: 2508.20139 [eess.IV]. URL: https://arxiv.org/abs/2508.20139.

[29] Dhemesh Yaseen et al. *University of Missouri Postoperative Glioma Dataset*. 2025. DOI: 10.7937/7K9K-3C83. URL: https://www.cancerimagingarchive.net/collection/mu-glioma-post/.