

OCTN: Neural OCT Representations for Robot-Guided Precision Intervention

Ravi Prakash, *Student Member, IEEE*, Ryan P. McNabb *Member, IEEE*, Patrick J. Codd, and Shan Lin

[Project Website](#)

Abstract—Optical coherence tomography (OCT) offers compact, contactless, micron-scale imaging suitable for intraoperative guidance, but native OCT volumes are discretely sampled, anisotropic, and currently inefficient for downstream geometric reasoning and robot integration. We present OCTN (pronounced “octane”), an implicit neural representation framework that converts volumetric OCT scans into a continuous, differentiable, and spatially faithful tissue-intensity field. OCTN uses a two-stage hybrid training strategy that combines supervision from acquired voxels with inter-slice interpolations, preserving B-scan fidelity while improving continuity in sparsely sampled regions. For versatility, we first show that OCTN enables fast volumetric reasoning by storing the learned tissue representation natively on the GPU, supporting intensity-based spatial queries with up to $43\times$ speedup over conventional CPU processing. We then demonstrate OCTN-enabled OCT-guided robotic laser surgery where the continuous tissue representation supports implicit surface discovery and surface-constrained path planning via multiple optimization strategies, including Newton- and SGD-based optimization. Next, OCTN enables reconstruction of dense volumetric structure from sparsely acquired B-scans, while reducing acquisition time by $4\times$ and preserving clinically relevant structures. Across the newly generated *Duke TissueOCT* dataset and public OCT datasets, OCTN achieves robust high-fidelity reconstruction with PSNR > 30 dB with training time < 10 s and preserves surface consistency within $10\mu\text{m}$ Chamfer distance relative to baseline reconstruction. The *TissueOCT* dataset and code are publicly available at raprakashvi.github.io/octn.

Index Terms—Optical Coherence Tomography, Implicit Neural Representation, Robotic Surgery, Image Reconstruction, 3D Reconstruction

I. INTRODUCTION

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible. *Manuscript currently under review.*

Aspects of the technology described in this manuscript are the subject of the following patent application: R. Prakash, R. P. McNabb, P. J. Codd, and S. Lin, “Systems and methods for neural OCT tissue representations for robot-guided intervention,” U.S. Provisional Patent Application No. 64/045,412, patent pending.

Corresponding author: Ravi Prakash.

Ravi Prakash, Ryan P. McNabb, and Patrick J. Codd are with Duke University, Durham, NC 27705 USA (e-mail: first-name.lastname@duke.edu).

Shan Lin is with Arizona State University, Phoenix, AZ 85287 USA (e-mail: shan.lin.2@asu.edu).

OPTICAL Coherence Tomography (OCT) is a contactless, non-ionizing sensing modality suited for high-resolution, near-real-time surgical imaging [1]. OCT allows micron-scale visualization of surface and shallow subsurface tissue features within a compact form factor. These properties make OCT attractive for intraoperative imaging and robot-guided intervention [2] over conventional counterparts. Although currently intraoperative ultrasound (IOUS) is widely used, it requires acoustic contact and provides limited spatial resolution (~ 0.5 mm) compared to OCT. Since its inception, OCT has established clinical use in ophthalmology [3]–[5] and has more recently been explored in neurosurgery [6]–[8], breast imaging [9], [10], dermatology [11], [12], and urology [13]. However, broader adoption remains limited by a restricted field of view (FoV), anisotropic scaling, and a computationally cumbersome discrete volume representation.

Recent advances in robot-mounted OCT [13]–[18] have enabled a larger field of view (FoV), surface-aware high-resolution imaging [19], extending utility. Despite this progress, native OCT volumes remain discretely sampled with anisotropic resolution and sampling, making them cumbersome to use directly for downstream computation. As a result, geometric reasoning and planning over OCT data remain inefficient and slow. Therefore, devising an isotropic and queryable representation of OCT data can overcome these limitations. Implicit neural representation (INR) is a Multi-Layer Perceptron (MLP) based method to parameterize the tissue volume by learning a mapping from spatial coordinates to intensity [20]. This results in a continuous, isotropic, and queryable neural encoding of the volume.

In recent years, INR have gained prominence in medical imaging [20] for reconstruction [21], [22], super-resolution [23], [24], segmentation, and registration [24], [25]. Unlike INR for natural 3D scenes, which often rely on multi-view image capture [26], medical imaging modalities are typically reconstructed from sparsely sampled parallel single-view cross-sectional slices. Methods such as CuNeRF [23] have adapted INRs for CT and MRI scans by learning a mapping from 3D spatial coordinates to voxel intensity, yielding a spatially correct continuous volume representation. When available, auxiliary information such as attenuation can further augment input representation and enhance reconstruction quality [27]. Expanding these methods to OCT requires handling sensor-specific idiosyncrasies and anisotropies under tight computational constraints for relatively small tissue volumes. Prior

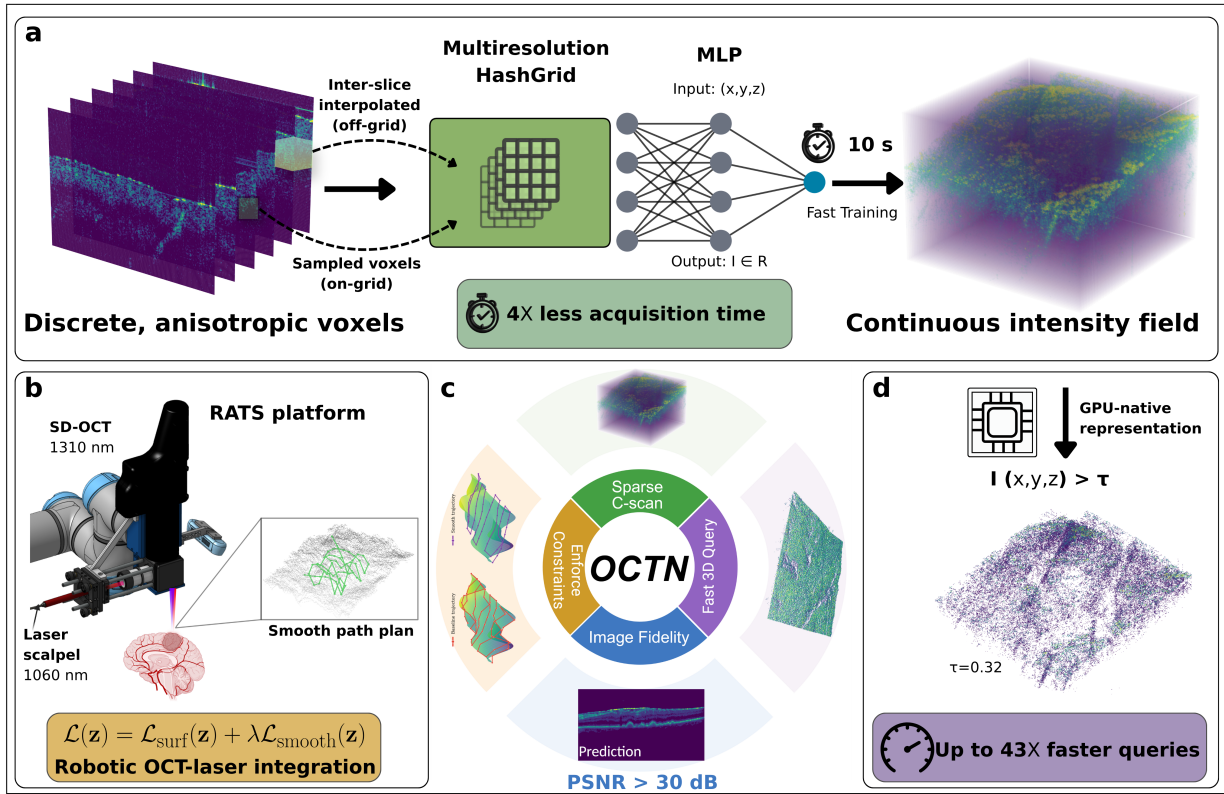


Fig. 1. **System overview.** **a)** OCTN maps discrete, anisotropic OCT voxels to a continuous tissue-intensity field using on-grid samples and off-grid inter-slice supervision. Its multiresolution hash-grid and MLP enable fast training, arbitrary spatial queries, and $4\times$ sparse acquisition. **b)** OCTN supports smooth, surface-aware robotic laser path planning. **c)** OCTN balances sparse C-scan reconstruction, image fidelity, fast 3D querying, and surface constraints. **d)** Its GPU-native representation enables threshold-based tissue queries with up to $43\times$ faster spatial reasoning.

works have addressed OCT sampling anisotropy either through additional sensing modalities such as scanning laser ophthalmoscopy [24] or through physics-informed modeling of A-scan formation [21]. Despite these advances, existing INR-based OCT methods rely on specialized hardware, multimodal sensing, and prolonged training time, limiting their practicality for broader intraoperative deployment.

In this work, we present **OCTN**, an INR-based method for learning a continuous representation of OCT volumes under anisotropic sampling and speckle noise (Fig. 1a). OCTN enables constraint-based optimization for robot path planning, Fig. 1b, with high-fidelity image reconstruction Fig. 1c, supports fast volumetric queries Fig. 1d, while recovering spatially accurate geometry from a sparsely sampled volume. OCTN enhances the existing OCT experience in terms of resolution, speed, and robotic guidance. Overall, we propose:

- **Hybrid framework for continuous OCT representation:** We introduce a two-stage training strategy combining discrete and interpolated voxels to learn a continuous, isotropic OCT representation.
- **Fast query and robot path planning:** We demonstrate OCTN enables $19\times$ – $43\times$ faster volumetric queries than conventional CPU-based processing through a GPU-native implementation. Furthermore, its continuous representation supports constrained surface-aware path planning using Newton- and Stochastic Gradient Descent (SGD)-based optimization separately, on an OCT-guided robotic laser

platform [18] Fig. 1b.

- **Sparse volume reconstruction:** We show that OCTN supports reconstruction from sparsely sampled B-scan stacks, enabling up to $4\times$ reduction in B-scan density within clinical constraints. For a critically sampled volume, this corresponded to reducing acquisition time from 24 s to 6 s.
- **TissueOCT Benchmarking:** We release an open-source Duke TissueOCT dataset and comprehensive set of global and local metrics to benchmark learned B-scan fidelity, sparse to dense volume reconstruction quality, visualization speed for volumetric queries, and surface aware path planning over realistic tissue samples. The dataset consists of critically sampled OCT volumes from *ex vivo* tissues at varying noise and artifact levels, as seen in Table I, together with calibration parameters for 3D reconstruction.

II. MATERIALS AND METHODS

Here we describe our core methods and evaluation pipelines. This section is broadly divided into three broad themes: **i)** OCTN Core (dataset Section II-A, OCTN framework Section II-B, training pipeline Section II-C, & speed optimization Section II-D), **ii)** evaluation schemes and protocols (evaluation metrics Section II-E and sparse volume reconstruction Section II-F), **iii)** downstream robotic laser based tissue ablation by integrating OCTN for path planning Section II-H).

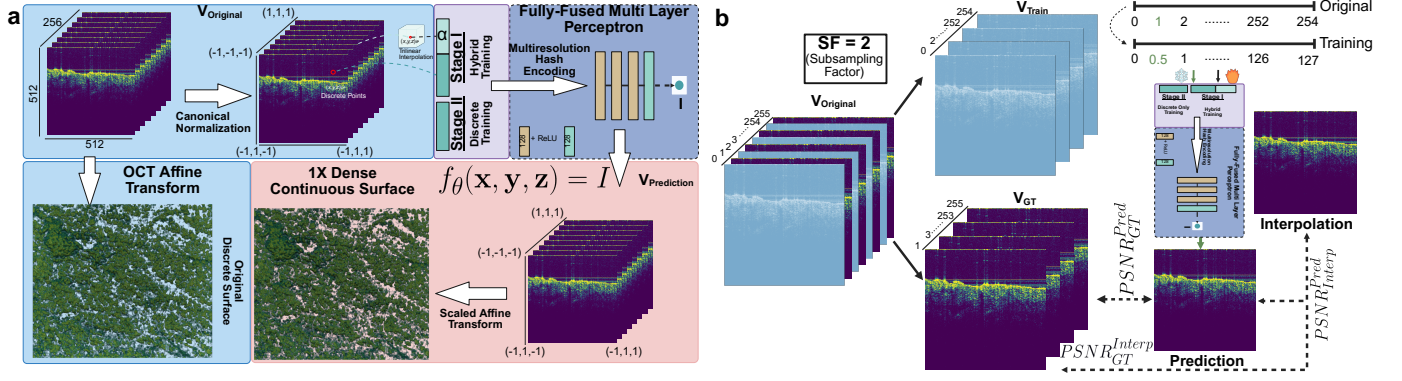


Fig. 2. **OCTN Framework and sparse reconstruction.** a) OCTN converts an anisotropically sampled OCT C-scan, V_{Original} , into a continuous neural volumetric representation. Volume is mapped to a canonical coordinate domain $[-1, 1]^3$, and an INR $f_{\theta}(x, y, z) \rightarrow I$ is trained using a two-stage strategy. Stage I combines acquired discrete voxels with interpolated inter-slice in ratio (α) to encourage off-grid continuity, while Stage II performs discrete-only refinement. The learned field can be queried at arbitrary resolution, enabling dense surface extraction and continuous volume rendering. b) Sparse reconstruction protocol shown for $SF = 2$. The subsampled training volume, V_{Train} , is used to train OCTN, and held-out B-scans form V_{GT} for evaluation. OCTN predictions, V_{Pred} , are compared with ground truth and with trilinear interpolation (baseline) to quantify reconstruction quality at unobserved slice locations.

A. Duke TissueOCT dataset

We present the **Duke TissueOCT dataset** to benchmark OCTN performance for volumetric reconstruction, sparse acquisition, and 3D tissue representation. The dataset, as described in Table I, contains OCT volumes of heterogeneous *ex vivo* animal tissues. Data were acquired using a low-cost spectral-domain (SD)-OCT system with a 1310 nm central wavelength and a 110 mm focal length (Lumedita Inc, NC, United States). Each rectangular volume had dimensions of $[D, H, W] = [256, 512, 512]$, where D is the number of B-scans and H, W denote the B-scan height and width. The voxel spacing was $14.59 \mu\text{m}$ axially, $14 \mu\text{m}$ laterally, and $28 \mu\text{m}$ between adjacent B-scans, corresponding to a critically sampled volume spanning $7.17 \times 7.17 \times 7.47 \text{ mm}^3$ with tissue-dependent axial penetration of 2–4 mm. Porcine muscle samples were acquired under two noise-suppression settings, with a subset containing artificial dye artifacts from cancer tissue-marking dye. Bovine and chicken tissue samples were acquired under a low-speckle-noise setting. As INR is susceptible to noise and prone to overfitting to easy-to-learn structures, the TissueOCT dataset contains a diversity in noise level, tissue type, and optical contrast to pose it as fair benchmarking.

B. OCTN: Continuous Neural Representation of Volumes

Given an OCT C-scan, $[D, H, W]$, OCTN learns a continuous neural intensity field that maps spatial coordinates to voxel intensity. Voxel intensities are normalized between $[0, 1]$, and coordinates are mapped to the canonical frame $[-1, 1]^3$, providing a resolution agnostic representation across volumes. OCTN uses a two-stage training strategy comprised of a hybrid and discrete training phase, further detailed in Section II-C. The hybrid phase learns an isotropic continuous representation for smooth sub-voxel queries by utilizing a mixture of discrete voxels from the acquired volume and interpolated counterparts generated through trilinear interpolation during training time, Fig. 2a. During the later discrete training phase, the emphasis is on assuring high-fidelity reconstruction on the native OCT

TABLE I
DUKE TISSUEOCT SAMPLE DISTRIBUTION.

Tissue	Volumes
Porcine muscle, high SNR	5
Porcine muscle, low SNR	11
Chicken breast	12
Bovine tissue	11
Porcine muscle with dye	7
Total	46

grid. Having a fraction of interpolated voxels mixed with discrete counterparts prevents overfitting on native voxels, and improves overall robustness of representation. The resultant neural tissue representation is given by Eq. (1).

$$f_{\theta}(x, y, z) = I \quad (1)$$

where (x, y, z) denotes a canonical 3D coordinate and I denotes the predicted OCT intensity. Unlike a discrete voxel array, the formulation supports intensity evaluation at arbitrary spatial coordinates, including unobserved locations between acquired B-scans.

C. Hybrid Discrete-Interpolated Training

First, a discrete OCT volume $\mathbf{V} \in \mathbb{R}^{D \times H \times W}$, is mapped to canonical coordinates through Eq. (2).

$$x_i = \frac{2i}{W-1} - 1, \quad y_j = \frac{2j}{H-1} - 1, \quad z_k = \frac{2k}{D-1} - 1 \quad (2)$$

This is followed by the division of the training process into hybrid and discrete phases based on r_{hyb} and r_{disc} ratios for total iterations, respectively. For a batch of size B , α denotes the fraction of interpolated samples during the hybrid phase, detailed in Algorithm 1. Each batch during the hybrid phase contains a sample distribution given by $C_{\text{interp}} = \lfloor \alpha B \rfloor$ and $C_{\text{disc}} = \lfloor B - C_{\text{interp}} \rfloor$ respectively. With speed in mind, a fully fused multi-layer perceptron (MLP) is used to learn the volume [28], [29]. Voxel coordinates in canonical space undergo projection to high-dimensional space through multiresolution

Algorithm 1 OCTN training procedure

Require: OCT volume $\mathbf{V}$, neural field f_θ , iterations T , hybrid iterations T_{hyb} , batch size B , fraction α

- 1: Initialize θ , optimizer, and learning-rate scheduler
- 2: **for** $t = 1$ to T **do**
- 3: **if** $t \leq T_{\text{hyb}}$ **then**
- 4: $C_{\text{interp}} \leftarrow \lfloor \alpha B \rfloor$, $C_{\text{disc}} \leftarrow B - C_{\text{interp}}$
- 5: Sample $\mathbf{c}_{\text{interp}} \sim \mathcal{U}([-1, 1]^3)$ and $\mathbf{c}_{\text{disc}}$ from the native grid
- 6: Get I_{interp} by trilinear interpolation and I_{disc} by voxel lookup in $\mathbf{V}$
- 7: $\mathbf{c} \leftarrow \mathbf{c}_{\text{interp}} \cup \mathbf{c}_{\text{disc}}$, $I \leftarrow I_{\text{interp}} \cup I_{\text{disc}}$
- 8: **else**
- 9: Sample $\mathbf{c}_{\text{disc}}$ from the native grid and obtain I_{disc} by voxel lookup
- 10: $\mathbf{c} \leftarrow \mathbf{c}_{\text{disc}}$, $I \leftarrow I_{\text{disc}}$
- 11: **end if**
- 12: $\hat{I} \leftarrow f_\theta(\mathbf{c})$
- 13: $\mathcal{L}_{\text{rec}} \leftarrow \mathcal{L}(\hat{I}, I)$
- 14: Update θ by backpropagation
- 15: **end for**
- 16: **return** f_θ

hash encoding to accurately represent high-frequency tissue features [30]. Specifically, the hashgrid encoder used has 24 levels, with 2 *features/level* and a hash table of size 2^{20} with base resolution of 16, following standard guidelines [28]. Encoded features are passed to MLP with 3 hidden layers with 128 neurons per layer and ReLU activation, mapped to a linear output head to regress intensity. The ADAM optimizer is used during training with mean squared error loss.

D. GPU-first Implementation

With Tiny-CUDA-NN as backbone, OCTN achieves orders of magnitude training speed gain over contemporary methods [23], due to GPU-optimized kernels that fuse multiple operations into a single CUDA kernel. This removes memory bandwidth bottlenecks, allowing OCTN to run efficiently even on a standard NVIDIA GPU, expanding access, unlike popular methods [23], [24]. OCTN has been successfully tested on NVIDIA 3070 Ti, 4080 Super, and 5090 GPUs. The result is a continuous neural intensity field, Eq. (1), that maps canonical coordinates $(x, y, z) \in [-1, 1]^3$ to voxel intensity I .

All metrics and comparisons reported are for a single NVIDIA 4080 Super GPU. During training, the full OCT volume is loaded onto the GPU to minimize data transfer between the CPU and GPU. All training samples are generated on the fly from the GPU-resident volume, avoiding dataloader overhead. Rather than traversing the full voxel grid at each iteration [26], training used only $N_{\text{batch}} * B$ sampled voxels per iteration, reducing per-iteration computation. Both interpolated and discrete training samples were generated directly on the GPU using CUDA-accelerated trilinear interpolation and direct tensor indexing. Together, on-the-fly GPU sampling and reduced host-device communication alongside CUDA-optimized kernels enabled low-latency training and inference on a single

consumer-grade GPU. These steps differentiate OCTN from previously reported sparse-view INR solutions in terms of training speed by providing realistic learned representation within seconds and reducing the time barrier for integration with simulation and robotic platforms.

E. Evaluation Metrics

As OCT volumes contain device-dependent signal noise, attenuation, and artifacts, a diverse set of metrics is needed to adequately capture the reconstruction quality. We propose a set of local metrics, alongside global counterparts, suited for imaging modalities such as OCT, where the noise is inherent and accompanied by the global features. The local metrics quantify OCTN performance on regions of interest, while the global metrics focus on average performance over the entire image (including speckle noise), aligning with commonly reported metrics in literature for comparison.

1) 2D image-based metrics: Signal-to-noise ratio (SNR): In an OCT image, the tissue regions are treated as signal, and air regions above the tissue surface are treated as noise [31]. SNR is computed *per* image; therefore, a signal and noise pair in the ground-truth image is used, and similarly for the neural prediction and interpolation-generated images. In this study, representative regions [32] with sharp intensity transitions are used for evaluation.

Peak signal to noise ratio (PSNR): PSNR is computed over the same manually selected regions of interest used for SNR. **Edge preservation index (EPI):** EPI evaluates how well local edge structures are preserved relative to the ground-truth image [32]. Unlike the preceding metrics, EPI is a reference-based metric. Image-based metrics are evaluated at two scales: globally over the entire generated images and locally within selected regions. In contrast, 3D metrics are reported only at the global scale.

2) 3D surface similarity metrics: Spatial consistency is evaluated by comparing tissue surfaces extracted from reconstructed and ground-truth volumes as pointcloud. The tissue surface is defined as the air-tissue interface, which is assumed to correspond to the largest axial intensity transition within each A-scan. For an OCT intensity field $f(x, y, z)$, with z denoting the axial direction, the surface location is defined as

$$z^*(x, y) = \arg \max_z |\nabla_z f(x, y, z)|, \quad (3)$$

where ∇_z denotes the axial intensity gradient. For baseline native discrete OCT volumes, ∇_z is approximated using a first-order finite difference along the axial direction.

Ground-truth and predicted surfaces are represented as point clouds, denoted by $\mathbf{G}$ and $\mathbf{P}$, respectively. For computational efficiency, both point clouds are uniformly downsampled to at most $M = 100,000$ points. The 3D surface metrics are defined as follows.

Chamfer distance (CD): CD quantifies the average bidirectional nearest-neighbor distance between two point clouds [33]. It is defined as

$$d_{\text{CD}}(\mathbf{G}, \mathbf{P}) = \frac{1}{2N_G} \sum_{i=1}^{N_G} \min_j \|\mathbf{g}_i - \mathbf{p}_j\|_2 + \frac{1}{2N_P} \sum_{j=1}^{N_P} \min_i \|\mathbf{p}_j - \mathbf{g}_i\|_2 \quad (4)$$

Hausdorff distance (HD): HD measures the maximum nearest-neighbor discrepancy between two point clouds. It is defined as

$$d_H(\mathbf{G}, \mathbf{P}) = \max \left\{ \max_i \min_j \|\mathbf{g}_i - \mathbf{p}_j\|_2, \max_j \min_i \|\mathbf{p}_j - \mathbf{g}_i\|_2 \right\} \quad (5)$$

where $\|\cdot\|_2$ denotes the Euclidean norm. Lower CD and HD values indicate greater geometric agreement, with HD capturing the worst-case surface deviation. As with the image-based metrics, both surface metrics can be affected by OCT speckle noise and a lack of distinguishing features.

F. Intermediate and Sparse Reconstruction Protocol

To evaluate whether OCTN learns an isotropic continuous representation, the trained field was queried at fractional locations between adjacent acquired B-scans. For a volume with D B-scans, anchor slices were selected around the volume center: $(\frac{D}{2}, \frac{D}{2} + 1)$ for even D , and $(\frac{D-1}{2}, \frac{D+1}{2})$ for odd D . Intermediate slices were rendered at 0.25 cross-sectional intervals between the anchors. Since physical OCT measurements are not available at these fractional locations, trilinear interpolation of the acquired volume was used to generate proxy reference images. This evaluation was performed on a representative critically sampled TissueOCT volume Section II-A.

The protocol was then extended to sparse reconstruction from lower-density B-scan stacks. This evaluates whether OCTN can recover held-out cross-sections from sparsely acquired volumes, similar to interpolation without the need for any other method. This would help reduce acquisition burden without requiring specialized high-speed OCT hardware [2]. A subsampling factor, SF , was used to define the B-scan sparsity. For each SF , the original volume was split into acquired training B-scans and held-out ground-truth test B-scans. OCTN was trained (Algorithm 1) using only the acquired B-scans (Fig. 2b). After training, the checkpoint was frozen and queried at the held-out B-scan locations. The reconstructed B-scans were then compared directly with corresponding ground-truth OCT measurements for fair evaluation.

G. Intensity-based Volumetric Query Benchmark

With image and volume reconstruction tasks in place, we evaluated OCTN for robotic tasks. A simple yet common problem involved volumetric queries for an intensity-based extraction task. Each query consisted of loading the volume, filtering voxels within a specified intensity range, and applying a geometric mask to retain points within a predefined region. Three intensity ranges $[0.2 - 0.4]$, $[0.4 - 0.6]$, $[0.6 - 1.0]$, were used to evaluate low-, intermediate-, and high-intensity structures, respectively. Intensity range $[0.6 - 1.0]$ preserved high contrast tissue surface features. For lower and upper intensity bounds, I_{low} and I_{high} , the queried set was given

$$\mathcal{Q}_{[I_{low}, I_{high}]} = \{(x, y, z) \in \mathbb{R}^3 \mid I_l \leq f(x, y, z) \leq I_h\}. \quad (6)$$

Baseline OCT ground-truth volumes were stored in Neuroimaging Informatics Technology Initiative (NIFTI) file format, while for OCTN the learned volume was queried.

H. Surface-Constrained Robotic Raster Path Planning

The continuous representation learned by OCTN enables new image-guided robotic frontiers through surface-aware path planning for OCT-guided robotic surgery. Consider tumor ablation with a precision robotic laser scalpel, where the laser is applied over a target region using a raster trajectory. A key requirement is to maintain a fixed focal offset from the tissue surface while preserving smooth robot motion. We therefore formulate raster execution as a surface-constrained path planning problem. Given a tissue volume, a raster path is first defined as $G = \{(x_i, y_i)\}_{i=1}^N$, where each (x_i, y_i) denotes a planned lateral location. The objective is to estimate the corresponding surface depth z_i such that each point $p_i = (x_i, y_i, z_i)$ lies on, or at a prescribed offset from the surface. We describe the steps involved in the optimization process.

1) Surface discovery: Tissue geometry is defined as an intensity level set,

$$S_\tau = \{(x, y, z) \mid f_\theta(x, y, z) = \tau\}, \quad (7)$$

where τ denotes the target surface intensity level. In this study, τ is estimated from a dense-volume surface extracted using the axial-gradient method. Specifically, τ is assigned as the median intensity of the extracted surface points, providing a simple data-driven threshold for querying the learned field.

2) Task setup: A benefit of OCTN representation is the continuity of voxel space, enabling gradient-based optimization of raster points onto the tissue surface. However, direct optimization is sensitive to initialization. Since the raster path G is defined only in the lateral plane, each depth value z_i must be initialized near the surface. Points initialized in air have near-zero intensity gradients and can therefore fail to move toward the tissue boundary.

We address this using a two-step strategy, and demonstrate using two separate methods, Newton, and SGD, as described below with varying benefits. First, a coarse axial search initializes each point near the surface by sampling N_{warmup} candidate depths and selecting the location with the maximum finite-difference gradient magnitude. Second, the initialized points are refined using gradient-based optimization.

$$z_i^{(0)} = \arg \max_{z \in \mathcal{Z}} \left| \frac{\partial f_\theta(x_i, y_i, z)}{\partial z} \right|. \quad (8)$$

3) Newton refinement: To obtain sub-voxel surface estimates, the coarse depth initialization is refined using Newton's method. Surface discovery for G is formulated as a root-finding problem along the axial direction:

$$f_\theta(G_i, z) - \tau = 0 \quad (9)$$

Starting from coarse estimates, each depth value is updated as

$$z_i^{(t+1)} = z_i^{(t)} - \frac{f_\theta(G_i, z_i^{(t)}) - \tau}{\frac{\partial f_\theta}{\partial z}(G_i, z_i^{(t)})}. \quad (10)$$

In implementation, the axial gradient is approximated using a central finite difference method:

$$\frac{\partial f_\theta}{\partial z}(G_i, z_i) \approx \frac{f_\theta(G_i, z_i + \epsilon) - f_\theta(G_i, z_i - \epsilon)}{2\epsilon}, \quad (11)$$

where ϵ is set to half of the OCT axial resolution in canonical coordinates. Safeguards are applied to improve stability. Updates are skipped when the gradient magnitude is too small, Newton step is clamped, and refinement is restricted to a neighborhood around the coarse estimate.

4) *SGD refinement*: For uneven tissue surfaces, a smooth raster trajectory is desired to reduce robot jerk while maintaining contact with the neural iso-surface. We therefore use stochastic gradient descent-based optimization to jointly refine the raster depths and enforce axial smoothness. The surface loss penalizes deviation from the target iso-intensity level:

$$\mathcal{L}_{\text{surf}}(\mathbf{z}) = \sum_{i=1}^N (f_{\theta}(G_i, z_i) - \tau)^2, \quad (12)$$

where $\mathbf{z} = [z_1, \dots, z_N]^T$.

A second-order smoothness term penalizes abrupt depth changes along each raster line:

$$\mathcal{L}_{\text{smooth}}(\mathbf{z}) = \sum_{i=2}^{N-1} v_i (z_{i+1} - 2z_i + z_{i-1})^2, \quad (13)$$

where v_i is a binary mask that prevents smoothing across raster-line turns, preserving sharp corners.

Together, the joint objective for the fine adjustment of raster points is given by

$$\mathcal{L}(\mathbf{z}) = \mathcal{L}_{\text{surf}}(\mathbf{z}) + \lambda \mathcal{L}_{\text{smooth}}(\mathbf{z}), \quad (14)$$

where λ is the smoothness weight. Setting $\lambda = 0$ removes smoothness regularization. During optimization, each update is clamped to remain within N_{voxels} of the coarse initialization, preventing points from drifting away from the local surface.

5) *Baseline*: Baseline method utilizes the raw OCT voxel volume $I(D, H, W)$. For each C-scan, an initial surface is estimated through identifying the highest axial intensity change as described above. With $\mathcal{S} = \{\mathbf{s}_j\}_{j=1}^N$, $\mathbf{s}_j = (x_j, y_j, z_j)$ surface obtained, each raster query G is assigned a corresponding z_i from the nearest surface point in the XY plane using the k-nearest neighbor (kNN) algorithm. Due to the discrete nature of volume, depth precision is limited by axial resolution, and doing a kNN search is computationally expensive as the number of raster query points increases. Timing comparisons include only the core operations required by each method; volume or model loading and auxiliary outlier checking/rejection were excluded. In practical deployment, OCTN's speed advantage compounds over repeated high-volume spatial query tasks.

III. EXPERIMENTS AND RESULTS

In this section, we evaluate OCTN as a continuous volumetric representation for OCT data. We demonstrate that its ability to support spatially consistent geometric reasoning and efficient volumetric querying while allowing sparse-acquisition reconstruction as a route for faster imaging, and finally quantify cross-sectional reconstruction quality alongside comparison and ablation studies.

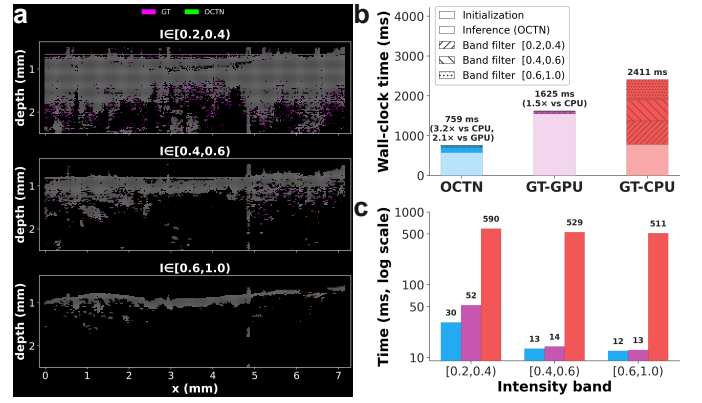


Fig. 3. **Intensity-based volumetric query.** a) Baseline and OCTN tissue volumes across intensity ranges, with better agreement under reduced speckle. b) End-to-end runtime shows $2\times$ – $3\times$ OCTN speedup. c) Query-only runtime shows larger gains at higher voxel counts.

A. OCTN Enables Fast Volumetric Query

Imaging tissues at high resolution entails the computational cost of querying tens of millions of voxels. Conventional lookup table-based processing is discrete and time-cost increases as volume size grows [34]. In contrast, the OCTN learned volumes can be queried efficiently on the GPU. Here we benchmark the intensity based volumetric query-speed for a porcine muscle tissue obtained from the TissueOCT dataset Eq. (1), as described in Section II-G. Reported query times included initialization, voxel filtering, geometric masking, and GPU-to-CPU transfer.

At native resolution, the corresponding OCT volume contains $256 \times 512 \times 512 = 67,108,864$ voxels. Our method generates qualitatively identical cross-sectional images at the tested volumetric conditions, Fig. 3a. This is further backed by Table II with point clouds generated through the learned method staying below $42 \mu\text{m}$ chamfer distance, showing excellent reproducibility of volumes and within the acceptable clinical threshold of $< 0.5 \text{ mm}$ for error. Hausdorff distance decreases as the points become sparse, reducing outliers in volumes. It takes OCTN 0.158 s to infer $\sim 67 \text{ M}$ points with additional 0.0243 s to build the uniform grid at native resolution and convert from canonical space to physical coordinate space, Fig. 3b, showing OCTN to have $2\times$ speed gain over GPU and $3\times$ speed gain over CPU during the complete process. Isolating the time needed to perform volumetric queries, OCTN achieves the target at $\sim 19\times$, $\sim 40\times$, $\sim 43\times$ faster than CPU-based methods. For fair comparison, the native OCT volume was transferred to GPU before evaluation. The gains over GT-GPU increase with an increase in total points queried. This is explained by an increase in GPU utilization in conventional search with an increase in net points, whereas OCTN inference batch GPU utilization remains unchanged.

B. Evaluation of OCTN-guided Robot Path Planning

We further demonstrate novel tasks towards robotic integration and guidance enabled by OCTN for the task of OCT-guided tissue resection. As described in Section II-H,

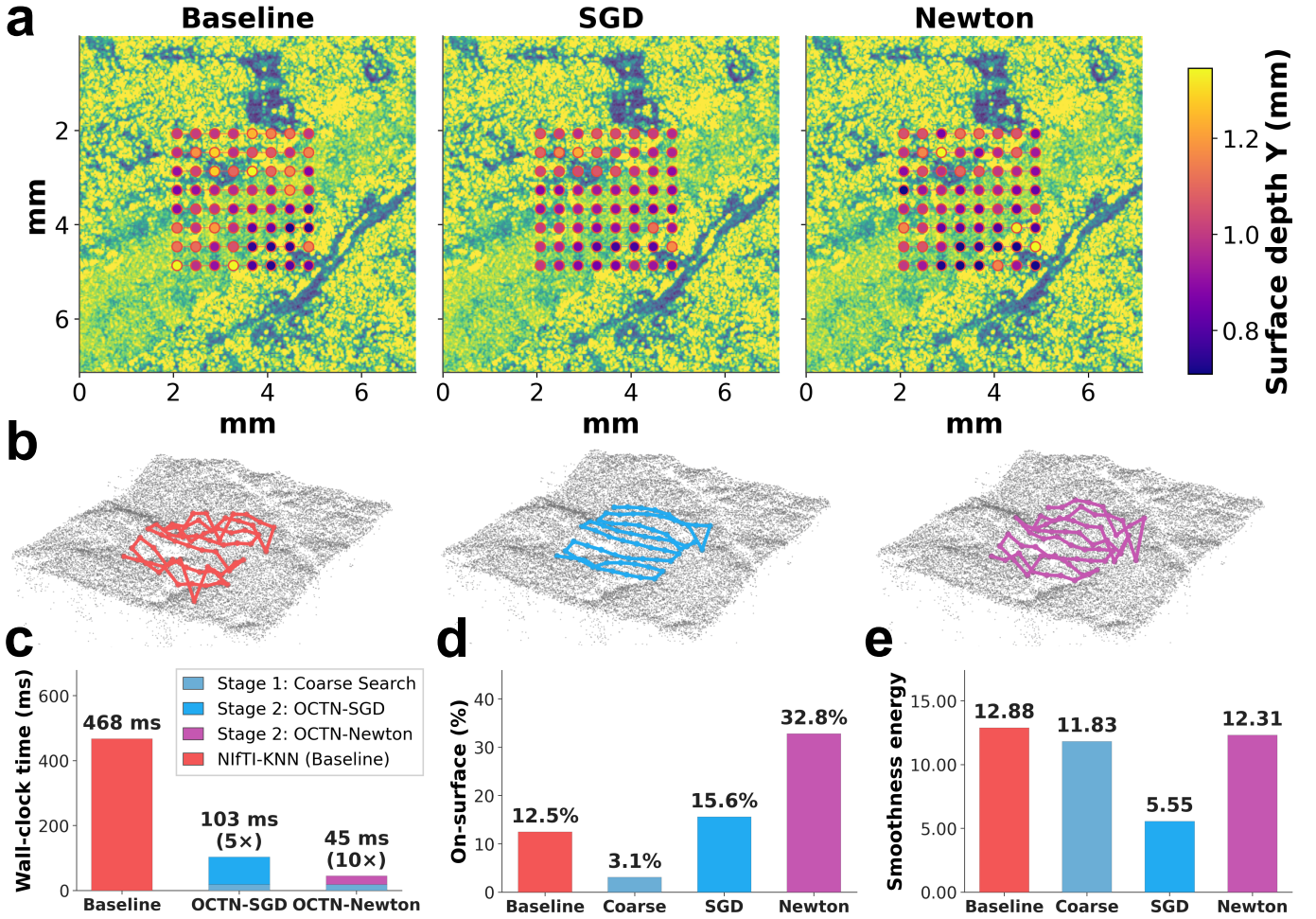


Fig. 4. **OCTN-enabled path planning.** **a)** Porcine tissue MIP with a $3 \text{ mm} \times 3 \text{ mm}$ raster pattern; waypoints are colored by surface depth. **b)** Tissue surface with optimized raster paths. SGD enforces smoothness and reduces jerk, unlike Baseline and Newton. Methods are compared by **c)** path generation time, **d)** surface-hit rate within 5% of target τ , and **e)** depth smoothness energy $\Sigma(\Delta^2 y_i)^2 \text{ (mm}^2\text{)}$.

we use the RATS platform, an image-guided robotic laser surgery system with coaxial OCT imaging and laser ablation [18], shown in Fig. 1b. Here, we evaluated OCTN for two separate tasks: i) ablate a tumor $3 \times 3 \text{ mm}^2$ in a raster pattern (**Newton method**) and ii) ensure smoothness of robot motion (**stochastic gradient descent**), as described in Section II-H. The overall trajectory is represented in Fig. 4a with a depth-based color map of raster points over MIP *en face* image of the porcine tissue surface. The optimized raster path on the implicitly estimated tissue surface is shown in Fig. 4b. Compared to baseline kNN-based search, Newton method found the surface for the raster path 10× faster, while placing 32.8% points on the surface, compared to 12.5% for baseline, Fig. 4c and Fig. 4d. Given relatively rough tissue surfaces and OCT speckle noise, this creates difficulties in generating smooth robot motion. The SGD-based method achieves the smoothest path with 56.9% reduction over baseline, Fig. 4e, while requiring $\sim 5\times$ lower computational speed and 15.6% points on the surface.

TABLE II
VOLUMETRIC QUERY ACCURACY ACROSS OCT INTENSITY BANDS.

Band	Chamfer (mm)↓	Hausdorff (mm)↓
[0.2, 0.4]	0.0423	1.572
[0.4, 0.6]	0.0264	1.451
[0.6, 1.0]	0.0158	0.963

C. Sparse Reconstruction for Faster Volume Acquisition

We next evaluate whether OCTN can recover dense volumetric structure from sparsely acquired B-scans, thereby reducing the number of samples required per C-scan, as per Section II-F. Volumes from TissueOCT data were uniformly subsampled with $SF \in \{2, 4, 8, 16, 32\}$, and OCTN was trained using only the retained training volume part. Reconstructed B-scans were evaluated against held-out ground truth B-scans. Trilinear interpolation was used as baseline.

OCTN consistently matched or exceeded trilinear interpolation on held-out B-scans. For qualitative comparison Fig. 5a shows representative slices selected near the midpoint between neighboring acquired B-scans, where interpolation is least favored but the evaluation is most informative. At aggressive subsampling levels, beginning at $SF = 8\times$, with inter B-

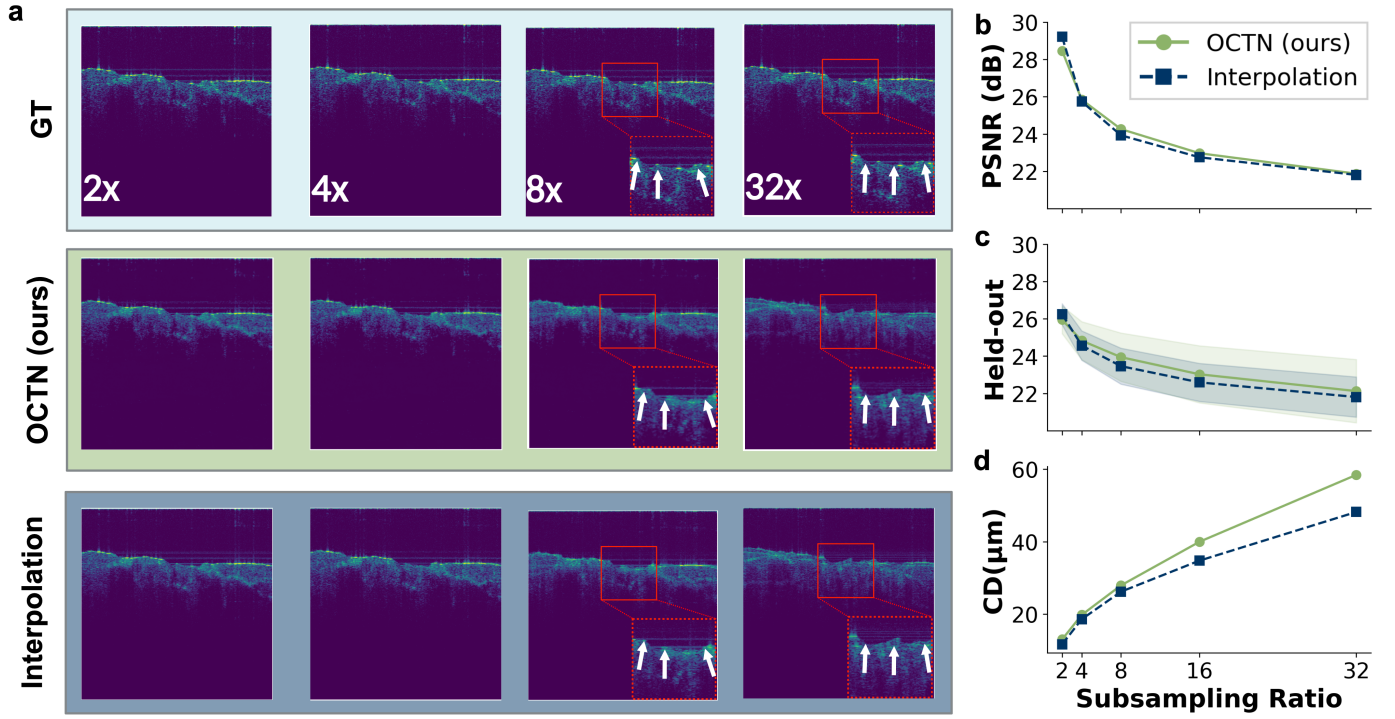


Fig. 5. **Sparse volume reconstruction.** **a)** Qualitative B-scan comparison across SF for OCTN and interpolation. Red boxes and white arrows mark interpolation-induced surface artifacts visible in $8\times$ and beyond. **b)** Global PSNR at native resolution $[256, 512, 512]$. **c)** Held-out B-scan performance reported as mean $\pm 1\sigma$ over each volume. **d)** Chamfer distance (CD) quantifies reconstruction quality and spatial-feature preservation.

scan distance of $224\ \mu\text{m}$, visible blurring of fine tissue features is observed (highlighted in red), with substantial degradation seen at $SF = 32\times$. This is expected: at $SF = 8\times$, only 32 of 256 B-scans are used for training, and at $SF = 32\times$, only 8 of 256 B-scans are available at $896\ \mu\text{m}$ spacing between acquired data. Consistent with sampling limits, neither OCTN nor interpolation can recover high-frequency structure that is absent from the acquired data. We evaluated the trained neural representation on the full grid at native resolution, $[256, 512, 512]$, as shown in Fig. 5b, and at held-out dataset in Fig. 5c. As a baseline, trilinear interpolation is used to upsample the subsampled input volume from $[256/SF, 512, 512]$ to $[256, 512, 512]$. OCTN demonstrates improved performance over standard method without the extra step needed for interpolation. Furthermore, as seen in Fig. 5d, volumes reconstructed through OCTN shows closeness to ground truth data. OCTN prediction provides smoother representation even for conditions with higher SF .

D. Reconstruction of OCT B-scans

Finally, the benefits of robot integration and sparse reconstruction cannot be achieved at the cost of reduction in B-scan reconstruction quality. We assess cross-dataset generalization by evaluating OCTN over public swept-source and spectral-domain datasets (Duke Srinivasan [35], Harvard GDP [36], Noor Eye [37] and our TissueOCT dataset). Two variants of TissueOCT dataset volumes are generated with smoothing (Gaussian and Median with filter of 3) to gently smooth the noise. For qualitative comparison, B-scan midway through the volume was chosen for evaluation. Here, $\alpha = 0.20$ was

chosen throughout. Local metrics are calculated on the regions of interest, here defined by intensity change and air-tissue boundary. The results Fig. 6 show OCTN is able to reconstruct volumes at high rendering quality. As seen in Table III, rendered images maintain the native SNR, highlighting the ability to learn high-frequency image features. For datasets containing high speckle noise (TissueOCT), local PSNR and SSIM show gains for $R1$ and $R2$ but experience degradation in performance at the air-tissue boundary ($R3$). Regions were selected based on local intensity-feature variation to highlight areas where reconstruction is more challenging. TissueOCT Gaussian and Median show higher global metrics with degraded performance locally, suggesting inflation due to the smoothing of speckle noise. Presence of EPI of ~ 1 shows the preserves edge well relative to the ground truth. We report PSNR, SSIM, and EPI relative to regions in ground truth.

Furthermore, the two-stage training helps in lowering the loss and improving PSNR values, Fig. 7a, achieving acceptable PSNR (30 dB) in less than 10 s on average. Reconstruction quality generally tends to improve with a reduction in total voxels in a dataset, Fig. 7b. Higher variance in performance stems from varied C-scan dimensions and noise levels within datasets. To highlight the effect of speckle noise on rendering performance beyond TissueOCT dataset, a noise-suppressing step is applied on the Srinivasan dataset. Results, as seen in Fig. 7c, highlight improvement of 6.5 dB and 3.3 dB in PSNR while 0.12 and 0.002 in SSIM, respectively, for TissueOCT (Porcine) and Srinivasan datasets, showing effectiveness.

E. Real World Ex vivo Ablation

TABLE III
GLOBAL AND LOCAL RECONSTRUCTION PERFORMANCE ACROSS REPRESENTATIVE DATASETS. $\Delta m = m_{\text{ROI}} - m_{\text{global}}$.

Dataset	Index	Global		ROI	ROI		Δ ROI		SNR _{GT} ↑	SNR _{Pred} ↑	EPI ~ 1
		PSNR ↑	SSIM ↑		PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑			
Srinivasan	24	44.64 / 0.996		R1	44.81 / 0.997		+0.17 / +0.001		9.42	9.41	0.998
				R2	43.39 / 0.992		-1.25 / -0.004		3.32	3.32	0.977
				R3	43.60 / 0.999		-1.04 / +0.003		6.62	6.62	0.992
Harvard	100	45.05 / 0.995		R1	45.24 / 0.997		+0.19 / +0.002		4.87	4.86	0.990
				R2	46.09 / 0.999		+1.04 / +0.004		8.47	8.48	1.000
				R3	45.77 / 0.999		+0.72 / +0.004		9.67	9.67	0.989
Noor Eye	12	49.00 / 0.997		R1	47.96 / 0.998		-1.04 / +0.001		11.96	11.98	1.048
				R2	49.00 / 0.997		-0.00 / +0.000		9.84	9.86	1.058
				R3	50.05 / 1.000		+1.05 / +0.003		10.44	10.46	1.016
TissueOCT	128	35.61 / 0.833		R1	40.34 / 0.997		+4.73 / +0.164		16.56	18.66	0.994
				R2	36.99 / 0.988		+1.38 / +0.155		13.56	15.64	0.981
				R3	29.40 / 0.988		-6.21 / +0.155		17.65	19.72	0.897
TissueOCT Gaussian	128	38.34 / 0.870		R1	37.43 / 0.985		-0.91 / +0.115		19.72	22.56	0.955
				R2	36.65 / 0.968		-1.69 / +0.098		16.39	19.20	0.958
				R3	33.60 / 0.992		-4.74 / +0.122		20.66	23.52	0.979
TissueOCT Median	128	39.04 / 0.955		R1	36.80 / 0.984		-2.24 / +0.029		22.75	25.63	1.001
				R2	34.74 / 0.961		-4.30 / +0.006		19.42	22.25	0.992
				R3	34.61 / 0.995		-4.43 / +0.040		23.57	26.48	0.965

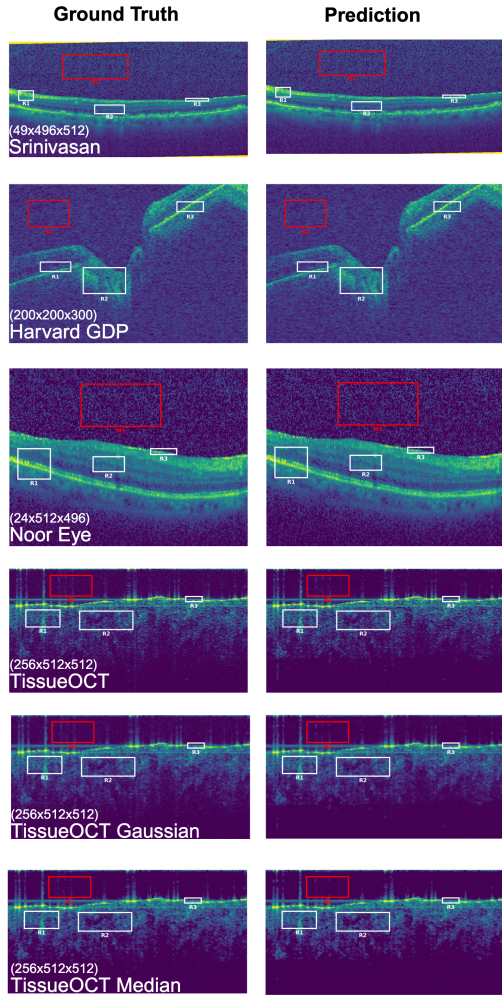


Fig. 6. **B-scan reconstruction:** OCTN performance over Srinivasan [35], Harvard GDP [36], Noor Eye [37], and variants of proposed TissueOCT data. For local metric estimation, regions marked as noise are shown in red, and signal in white, with results in Table III.

We perform real-world tissue ablation experiments to translate the properties of OCTN to a physical robotic surgery

platform as described in Section II-H. RATS laser surgery platform is used, Fig. 8a, [18] with the task of tissue ablation in a $3\text{mm} \times 3\text{mm}$ region. Porcine tissue surface was captured at native resolution, as seen in Fig. 8b. OCTN was used to learn a neural representation of volume within 1000 iterations at α 0.20, and raster waypoints were generated, using Newton, SGD ($\lambda = 10$ tuned based on tissue noise), and baseline methods. Using methods described in [18], a coordinate transformation matrix from OCT physical frame to Robot end effector (EE) frame was obtained, T_{OCT}^{EE} , allowing conversion of T_{Tissue}^{World} to the frame. Qualitatively, ablation was observed, as seen in Fig. 8c with discoloration and concurrently in, as seen in Fig. 8d in the thermal image. During implementation, OCTN helped by reducing the computational cost of generating waypoints and reducing-jerk robot motion Fig. 8e.

F. Ablation Studies

We test the robustness of the proposed methods through ablation studies. First, the effect of α in the two-stage training strategy is evaluated through global PSNR/SSIM metric on volume at native discrete pixel locations, and at a fractional B-scan at position 128.5. The value is chosen around the center of the volume. As shown in Table IV, with increasing α , the global discrete metrics decrease while the fractional metrics increase. The ideal value of $\alpha = 0.20$ provides highest PSNR at discrete locations with marginal degradation of image quality at fractional locations. The gain of 4.33 dB in PSNR and 0.161 in SSIM on average from $\alpha = 0.0$ to $\alpha = 0.20$ shows the effect that marginal involvement of interpolated voxels during training can have. For comparison, we chose CuNeRF [23], an INR method for CT and MRI which follows a similar input and output space. The hyperparameters were author-reported, with minor modifications made to training and evaluation batch size pertaining to the GPU in use. As seen in Table V, OCTN consistently performs better than CuNeRF with the final reconstruction losing sharpness for high intensity features. The extended training time > 6 hrs makes it difficult for use with OCT, where each volume represents a minuscule

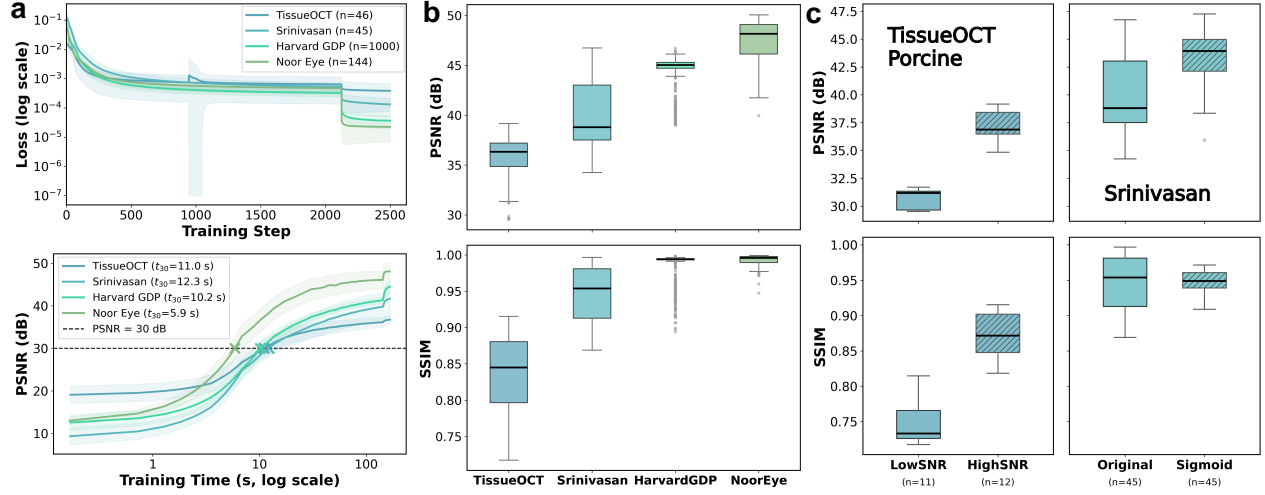


Fig. 7. **OCTN performance across datasets.** **a)** Training loss and PSNR, showing two-stage training reaches > 30 dB PSNR within seconds. PSNR is computed on 16 uniformly spaced B-scans from 10 random volumes. **b)** Global PSNR and SSIM distributions. **c)** Speckle-noise reduction improves PSNR and SSIM as seen in TissueOCT (left) and Srinivasan (right) volumes.

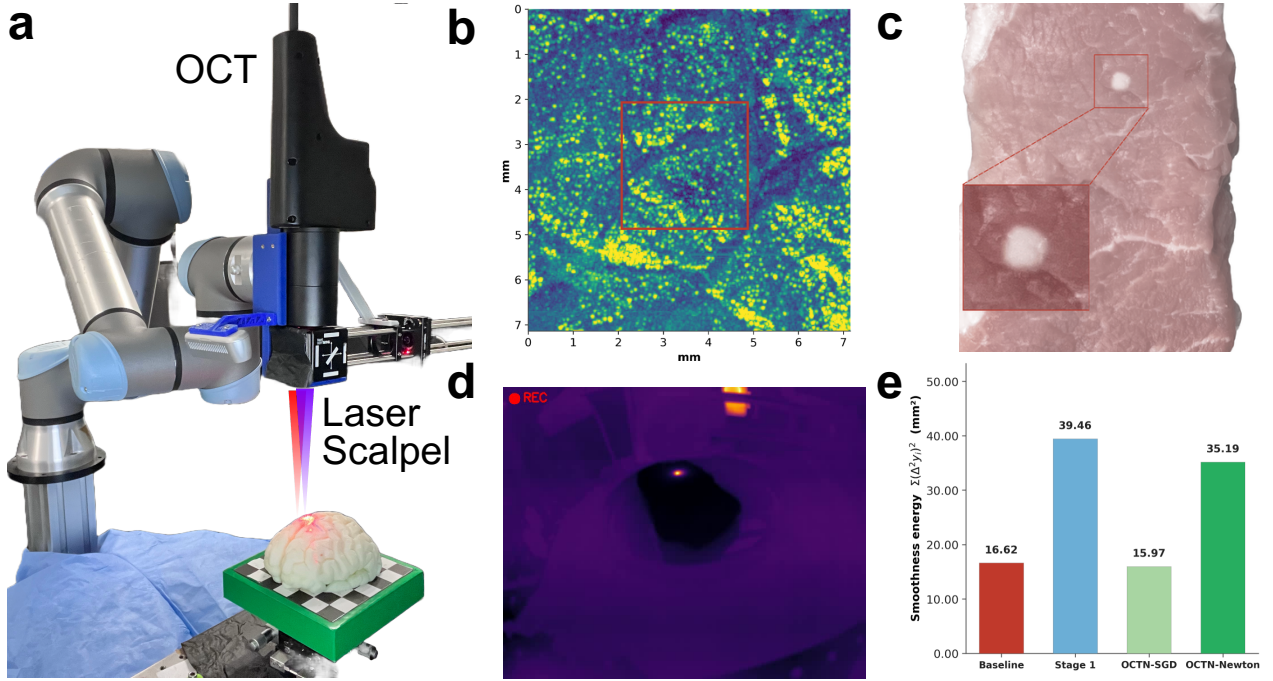


Fig. 8. **Porcine tissue ablation demonstration.** **a)** RATS robotic laser platform integrating OCT and a laser scalpel. **b)** *En face* MIP of porcine with target region marked in red. **c)** Post-ablation tissue showing localized coagulation and discoloration, with corresponding **d)** thermal image. **e)** Smoothness comparison, showing SGD achieves the lowest smoothness energy in the physical experiment.

area of $7 \times 7 \text{ mm}^2$. Reported results were evaluated on uniformly sampled B-scans for both the methods. CuNeRF may find benefit with modalities such as CT and MRI with volumes representing larger tissue regions, though OCTN might show an advantage in terms of speed.

IV. DISCUSSION

Widespread adoption of OCT for intraoperative surgical guidance is hindered by imaging anisotropy, acquisition speed, and computational complexity involved in data handling. Here, we present OCTN, an isotropic continuous representation of OCT volumes allowing up to $4\times$ reduction in acquisition speed

TABLE IV
ABLATION OF OCTN TWO-STAGE TRAINING ON TISSUEOCT. METRICS ARE REPORTED FOR THE FULL VOLUME AND FRACTIONAL B-SCAN 128.5.

Strategy	α	Volume PSNR/SSIM $\uparrow$	128.5 PSNR/SSIM $\uparrow$
Discrete only	0.00	35.26 ± 1.44 / 0.87 ± 0.03	22.86 ± 3.19 / 0.506 ± 0.192
OCTN, proposed	0.20	35.35 ± 2.87 / 0.83 ± 0.06	27.19 ± 3.12 / 0.667 ± 0.160
Low-interp.	0.40	34.93 ± 3.30 / 0.81 ± 0.08	28.04 ± 3.00 / 0.698 ± 0.144
Moderate-interp.	0.60	34.32 ± 3.50 / 0.79 ± 0.09	28.55 ± 2.89 / 0.713 ± 0.134
High-interp.	0.80	33.62 ± 3.48 / 0.80 ± 0.03	28.79 ± 2.85 / 0.723 ± 0.128
Interp. only	1.00	32.74 ± 3.34 / 0.75 ± 0.10	28.20 ± 2.63 / 0.716 ± 0.120

TABLE V
COMPARISON OF OCTN AND CUNERF [23] ($n = 3$): ~ 171 S/2.5K STEPS VERSUS ~ 394 MIN/250K STEPS.

Dataset	Method	PSNR $\uparrow$	SSIM $\uparrow$
Srinivasan [35]	CuNeRF	12.98 $\pm$ 0.96	0.38 $\pm$ 0.09
	OCTN	41.99 $\pm$ 5.33	0.97 $\pm$ 0.025
Harvard GDP [36]	CuNeRF	24.70 $\pm$ 0.78	0.48 $\pm$ 0.06
	OCTN	44.55 $\pm$ 1.23	0.99 $\pm$ 0.01
Noor Eye [37]	CuNeRF	19.62 $\pm$ 1.95	0.30 $\pm$ 0.038
	OCTN	45.95 $\pm$ 1.27	0.99 $\pm$ 0.00
TissueOCT (ours)	CuNeRF	23.82 $\pm$ 0.93	0.69 $\pm$ 0.01
	OCTN	35.09 $\pm$ 0.32	0.84 $\pm$ 0.02

and fast volumetric query (~ 67 M voxels in 0.158 s). With optimized two-stage training strategy, OCTN allows learning tissue volumes within seconds (> 30 PSNR in < 10 s) and permit sparse reconstruction of tissue volumes. In contrast with recent INR works in OCT utilizing external sensors [24] or theoretical model [21] of OCT, OCTN is hardware agnostic, and develops a direct map between spatial coordinates and voxel intensity by learning a continuous smooth volume. When coupled with a dexterous, precise laser scalpel, OCTN enables smooth tissue ablation by optimizing with sub-voxel accuracy at near-real-time speed. a

Overall, we evaluated OCTN on two categories of tasks: i) pertaining to current clinical diagnostics and ii) enabling integration with robotic hardware for guiding tools. For the first part, the core contribution is the hybrid training strategy. This allows reconstruction of B-scans in plane of choice, as desired by the clinicians. The use of local image-based and 3D surface similarity metrics perform holistic evaluation while the global metrics allow easy comparison with popular methods on multiple datasets. Downstream robotic integration and fast queries are the byproducts of the learned continuous volume residing in the GPU. This allowed us to perform intensity-based volumetric queries at $3.3\times$ faster than on CPU and about $2.1\times$ faster than on GPU (includes initialization and filtering). This speed gain is visible for queries with higher computational load. For the task of tissue ablation using the RATS [18] surgical system, we successfully generate a raster path with over 39% of waypoints on the tissue surface (**Newton method**) and lower smoothness energy (**SGD method**) over baseline by formulating the surface discovery as an optimization problem with OCTN allowing easy gradient calculation. This is finally demonstrated on *ex vivo* porcine tissue resulting in translation of OCTN to real world tasks.

Despite benefits, OCTN has a few limitations. For high-resolution tissue volumes (> 250 M voxels) obtained from the latest SS-OCTs, we observe degradation in reconstruction quality, amounting to a pseudo-Gaussian blur effect for an MLP of fixed size. Background noise is detrimental to the final learned quality, and there is no explicit method of handling it. Possible preprocessing to suppress background noise, or a specialized loss function during training, can assist in overcoming the challenge. Another issue stems from the choice of learning architecture used. Fully-fused MLP limits the backpropagation of first-order gradients during optimization with recent fixes suggesting potential solutions. Once solved, we anticipate OCTN to be the representation of choice for OCT data handling and integration.

Future work will extend OCTN toward dynamic surgical guidance by integrating clinical segmentation and diagnostic models, such as SAM2, with temporal inputs for deforming tissue. In ocular imaging, coupling OCTN with scanning laser ophthalmoscopy may improve reconstruction [24], while sequential OCT registration could support temporally consistent tissue representations [19]. Deriving signed distance functions from OCTN volumes may further enable faster geometric queries and planning [38], [39], including guidance for needles and scalpels [40]. Together, these directions position OCTN as a robot-ready OCT representation for multimodal surgical guidance, moving OCT beyond ophthalmic workflows toward deformable soft-tissue surgery, tumor resection, and broader sensor-fusion-driven [41] robotic intervention.

V. CONCLUSION

We present OCTN, an INR-based method for generating isotropic, continuous OCT volumes for enabling tasks in geometrical reasoning, robotics and imaging. OCTN proposes a two-stage training strategy to utilize native discrete and interpolated voxels to convert discrete OCT C-scans into continuous neural counterparts agnostic of the OCT sensor. With GPU-native data handling, OCTN allows enabling fast volumetric query, sparse volume reconstruction and easy out-of-the-box integration with surgical robotic platforms for path planning and tissue ablation with minimal loss in reconstruction quality. Future work would extend OCTN to include spatial and temporal sensor information on for wider clinical utility. The adjoining TissueOCT dataset and codebase are released to allow one-step conversion of OCT data to super-charged OCTN representation for public use.

REFERENCES

- [1] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, *et al.*, "Optical coherence tomography," *science*, vol. 254, no. 5035, pp. 1178–1181, 1991.
- [2] M. Draelos, P. Ortiz, R. Qian, C. Viehland, R. McNabb, K. Hauser, A. N. Kuo, and J. A. Izatt, "Contactless optical coherence tomography of the eyes of freestanding individuals with a robotic scanner," *Nature biomedical engineering*, vol. 5, no. 7, pp. 726–736, 2021.
- [3] C. A. Toth, D. G. Narayan, S. A. Boppart, M. R. Hee, J. G. Fujimoto, R. Birngruber, C. P. Cain, C. D. DiCarlo, and W. P. Roach, "A comparison of retinal morphology viewed by optical coherence tomography and by light microscopy," *Archives of ophthalmology*, vol. 115, no. 11, pp. 1425–1428, 1997.
- [4] M. Wojtkowski, T. Bajraszewski, I. Gorczyńska, P. Targowski, A. Kowalczyk, W. Wasilewski, and C. Radzewicz, "Ophthalmic imaging by spectral optical coherence tomography," *American journal of ophthalmology*, vol. 138, no. 3, pp. 412–419, 2004.
- [5] A. C. Tan, G. S. Tan, A. K. Denniston, P. A. Keane, M. Ang, D. Milea, U. Chakravarthy, and C. M. G. Cheung, "An overview of the clinical applications of optical coherence tomography angiography," *Eye*, vol. 32, no. 2, pp. 262–286, 2018.
- [6] C. Kut, K. L. Chaichana, J. Xi, S. M. Raza, X. Ye, E. R. McVeigh, F. J. Rodriguez, A. Quiñones-Hinojosa, and X. Li, "Detection of human brain cancer infiltration *ex vivo* and *in vivo* using quantitative optical coherence tomography," *Science translational medicine*, vol. 7, no. 292, pp. 292ra100–292ra100, 2015.
- [7] I. Dolganova, P. Aleksandrova, P. Nikitin, A. Alekseeva, N. Chernomyrdin, G. Musina, S. Beshplav, I. Reshetov, A. Potapov, V. Kurllov, *et al.*, "Capability of physically reasonable oct-based differentiation between intact brain tissues, human brain gliomas of different who grades, and glioma model 101.8 from rats," *Biomedical optics express*, vol. 11, no. 11, pp. 6780–6798, 2020.

- [8] Z. S. Navabi, R. Peters, B. Gulner, A. Cherkkil, E. Ko, F. Dadashi, J. O. Brien, M. Feldkamp, and S. B. Kodandaramaiah, "Computer vision-guided rapid and precise automated cranial microsurgies in mice," *Science Advances*, vol. 11, no. 15, p. eadt9693, 2025.
- [9] S. A. Boppart, W. Luo, D. L. Marks, and K. W. Singletary, "Optical coherence tomography: feasibility for basic research and image-guided surgery of breast cancer," *Breast cancer research and treatment*, vol. 84, no. 2, pp. 85–97, 2004.
- [10] R. Bareja, D. Mojahed, H. Hibshoosh, and C. Hendon, "Classifying breast cancer in ultrahigh-resolution optical coherence tomography images using convolutional neural networks," *Applied Optics*, vol. 61, no. 15, pp. 4458–4462, 2022.
- [11] R. Dey, S. Alexandrov, P. Owens, J. Kelly, S. Phelan, and M. Leahy, "Skin cancer margin detection using nanosensitive optical coherence tomography and a comparative study with confocal microscopy," *Biomedical Optics Express*, vol. 13, no. 11, pp. 5654–5666, 2022.
- [12] F. Latriglia, J. Ogien, C. Tavernier, S. Fischman, M. Suppa, J.-L. Perrot, and A. Dubois, "Line-field confocal optical coherence tomography (lc-oct) for skin imaging in dermatology," *Life*, vol. 13, no. 12, p. 2268, 2023.
- [13] X. Ma, M. Moradi, X. Ma, Q. Tang, M. Levi, Y. Chen, and H. K. Zhang, "Large area kidney imaging for pre-transplant evaluation using real-time robotic optical coherence tomography," *Communications Engineering*, vol. 3, no. 1, p. 122, 2024.
- [14] R. Prakash, V. Wang, Z. Chen, L. Bridgeman, and P. J. Codd, "Portable dual sensor large area visualization system for robotic laser surgery," in *2025 International Symposium on Medical Robotics (ISMR)*. IEEE, 2025, pp. 58–65.
- [15] G. Ma and M. Draelos, "Geometry-aware volumetric data stitching using local surface mapping and robot optical coherence tomography," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 3788–3794.
- [16] G. Ma, M. McCloud, Y. Tian, A. Narawane, H. Shi, R. Trout, R. P. McNabb, A. N. Kuo, and M. Draelos, "Robotics and optical coherence tomography: current works and future perspectives," *Biomedical Optics Express*, vol. 16, no. 2, pp. 578–602, 2025.
- [17] S. Lotz, M. Göb, S. Böttger, L. Ha-Wissel, J. Hundt, F. Ernst, and R. Huber, "Large area robotically assisted optical coherence tomography (lara-oct)," *Biomedical Optics Express*, vol. 15, no. 6, pp. 3993–4009, 2024.
- [18] R. Prakash, V. Y. Wang, A. Mishra, D. Yuliarti, P. Zhong, R. P. McNabb, P. J. Codd, and L. J. Bridgeman, "See, plan, cut: Mpc-based autonomous volumetric robotic laser surgery with oct guidance," *arXiv preprint arXiv:2511.17777*, 2025.
- [19] R. Fang, P. Zhang, T. Zhang, Z. Yan, D. Kim, E. Sun, R. Kuranov, J. Kweon, A. S. Huang, and H. F. Zhang, "Robotic optical coherence tomography with expanded three-dimensional field-of-view using point-cloud-based volumetric montaging," *IEEE transactions on medical imaging*, 2025.
- [20] A. Molaei, A. Aminimehr, A. Tavakoli, A. Kazerouni, B. Azad, R. Azad, and D. Merhof, "Implicit neural representation in medical imaging: A comparative survey," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2381–2391.
- [21] Y. Li, C. Zhang, T. Huang, Y. Fan, G. Ning, and H. Liao, "Computational multi-angle optical coherence tomography using implicit neural representation," *Optics & Laser Technology*, vol. 184, p. 112551, 2025.
- [22] S. Saha, Z. Liang, S. Lin, J. Lu, M. Yip, and S. Liu, "Based: Bundle-adjusting surgical endoscopic dynamic video reconstruction using neural radiance fields," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2025, pp. 3003–3012.
- [23] Z. Chen, L. Yang, J.-H. Lai, and X. Xie, "Cunerf: Cube-based neural radiance field for zero-shot medical image arbitrary-scale super resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 21 185–21 195.
- [24] B. Kahrs, J. Andresen, F. Falta, M. Santarossa, H. Handels, and T. Kepp, "Don't mind the gaps: Implicit neural representations for resolution-agnostic retinal oct analysis," *arXiv preprint arXiv:2601.02447*, 2026.
- [25] S. Komeda, N. Tateno, Y. Liu, R. Morishita, X. Wang, I. Abd El-Sadek, A. Furukawa, S. Matsusaka, S. Makita, and Y. Yasuno, "High-speed full-field swept-source dynamic optical coherence tomography enabled by neural-network-based image generation," *Biomedical Optics Express*, vol. 17, no. 5, pp. 2242–2261, 2026.
- [26] S. Liu, S. Lin, J. Lu, A. Supikov, and M. Yip, "Baa-ngp: Bundle-adjusting accelerated neural graphics primitives," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 850–857.
- [27] L. Zhou, C. Fang, B. Morovati, Y. Liu, S. Han, Y. Xu, and H. Yu, "nerf: Leveraging attenuation priors in neural radiance field for 3d computed tomography reconstruction," *arXiv preprint arXiv:2412.05322*, 2024.
- [28] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM transactions on graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.
- [29] T. Müller, "tiny-cuda-nn," 4 2021. [Online]. Available: <https://github.com/NVlabs/tiny-cuda-nn>
- [30] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Trans. Graph.*, vol. 41, no. 4, pp. 102:1–102:15, July 2022. [Online]. Available: <https://doi.org/10.1145/3528223.3530127>
- [31] M. Wu, W. Chen, Q. Chen, and H. Park, "Noise reduction for sd-oct using a structure-preserving domain transfer approach," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3460–3472, 2021.
- [32] Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, "Speckle noise reduction in optical coherence tomography images based on edge-sensitive cgan," *Biomedical optics express*, vol. 9, no. 11, pp. 5129–5146, 2018.
- [33] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.
- [34] Y. Liang, H. Vo, J. Kong, and F. Wang, "Ispeed: an efficient in-memory based spatial query system for large-scale 3d data with complex structures," in *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2017, pp. 1–10.
- [35] P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, and S. Farsiu, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomedical optics express*, vol. 5, no. 10, pp. 3568–3577, 2014.
- [36] Y. Luo, M. Shi, Y. Tian, T. Elze, and M. Wang, "Harvard glaucoma detection and progression: A multimodal multitask dataset and generalization-reinforced semi-supervised learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 20 471–20 482.
- [37] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular oct classification using a multi-scale convolutional neural network ensemble," *IEEE transactions on medical imaging*, vol. 37, no. 4, pp. 1024–1034, 2017.
- [38] G. Ma, R. Prakash, B. Mann, W. Ross, and P. Codd, "3d laser-and-tissue agnostic data-driven method for robotic laser surgical planning," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 8446–8453.
- [39] V. Y. Wang, R. Prakash, S. R. Oca, E. J. LoCicero, P. J. Codd, and L. J. Bridgeman, "Sampling-based model predictive control for volumetric ablation in robotic laser surgery," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 4527–4533.
- [40] G. Ma, R. Prakash, B. Schleupner, J. Everitt, A. Mishra, J. Chen, B. Mann, B. Chen, L. Bridgeman, P. Zhong, *et al.*, "Tumormap: A laser-based surgical platform for 3d tumor mapping and fully-automated tumor resection," *arXiv preprint arXiv:2511.05723*, 2025.
- [41] B. Chen, R. Kwiatkowski, C. Vondrick, and H. Lipson, "Fully body visual self-modeling of robot morphologies," *Science Robotics*, vol. 7, no. 68, p. eabn1944, 2022.