Geometry-Aware Implicit Neural Reconstruction of Oblique Micro-Ultrasound Scans

Harrison Chojnowski CHOJNOWSKI.H@UFL.EDU
Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611, USA

Gorkem Can Ates GATES@UFL.EDU
Department of Medicine, University of Florida, Gainesville, FL 32611, USA

Wayne G. Brisbane WBRISBANE@MEDNET.UCLA.EDU
Department of Urology, University of California Los Angeles, Los Angeles, CA, USA

Wei Shao WEISHAO@UFL.EDU
Department of Medicine, University of Florida, Gainesville, FL 32611, USA

Editors: Under Review for MIDL 2026

Abstract

Micro-ultrasound is a new modality for accurate, low-cost prostate cancer imaging, but its acquisition produces oblique slices that do not align with axial MRI or histopathology. This geometric mismatch complicates interpretation and prevents direct registration to histopathology, which is necessary to map ground-truth cancer outlines onto micro-ultrasound for training machine learning models for automated cancer detection. We address this challenge with a geometry-aware reconstruction framework that converts oblique micro-ultrasound slices into axial 3D volumes. Our method includes: (i) a coordinate-based sampling scheme that uses cylindrical geometry to accurately map each voxel into Cartesian space, and (ii) a generalized implicit neural representation that models the continuous intensity field between slices, preserving high-frequency speckle texture that traditional interpolation blurs. The reconstructed volumes achieve a 9% relative SSIM improvement over a coordinate-matched trilinear baseline while maintaining ultrasound-specific texture and boundary detail. This framework produces high-quality axial micro-ultrasound volumes suitable for reliable histopathology registration and for creating pathology-informed datasets to train cancer detection models.

Keywords: micro-ultrasound, implicit neural representation, prostate, volume reconstruction

1. Introduction

Micro-ultrasound imaging operates at 29 MHz, offering three to four times higher spatial resolution than conventional ultrasound while costing roughly one-tenth as much as MRI, yet achieving comparable diagnostic performance for prostate cancer detection (Lughezzani et al., 2019; Ashouri et al., 2023; Basso Dias and Ghai, 2023; Pensa et al., 2024; Klotz et al., 2020). Despite these advantages, its clinical adoption faces key challenges. During transrectal acquisition, the probe sweeps through the prostate while rotating, producing oblique-plane images at irregular angular intervals. Unlike fixed-spacing axial modalities such as MRI, these oblique acquisitions lack a consistent anatomical frame of reference. As a result, clinicians must interpret each slice in an unfamiliar orientation, and downstream

computational tasks such as registration and segmentation become significantly more difficult. Such limitations highlight the need for methods that convert oblique micro-ultrasound slices into coherent volumetric representations without degrading the high-frequency texture essential for diagnostic interpretation.

Reconstruction techniques have been proposed to mitigate these issues by transforming oblique micro-ultrasound acquisitions into coherent 3D axial volumes. Recent work has shown that the geometry of the acquisition permits mapping oblique slices onto Cartesian grids suitable for downstream tasks such as robotic biopsy guidance (Vassallo et al., 2023). However, existing reconstruction pipelines rely primarily on linear interpolation to increase the target volume. Although geometrically straightforward, linear interpolation is fundamentally ill-suited for ultrasound data as it acts as a low-pass filter which suppresses the high-frequency patterns and subtle transitions that carry diagnostic value. The resulting volumes appear over-smoothed, with diminished edge sharpness and loss of fine anatomical detail. Conversely, nearest-neighbor strategies (Imran et al., 2024) preserve texture but produce blocky, discontinuous reconstructions with pronounced banding artifacts, weakening spatial coherence.

These current limitations show that a valid reconstruction method must go beyond geometric mapping and linear interpolation. In particular, we argue that a learned method capable of synthesizing high-frequency ultrasound texture in a manner consistent with the learned anatomical statistics of prostate tissue is necessary. Further, the method must be flexible to work through the irregularity of micro ultra-sound data. Therefore in this paper, we introduce a non-linear generalized implicit neural representation based method that retains the geometric accuracy of coordinate-based mapping while resolving the limitations of interpolation. Our contributions are as follows:

(i) We develop a coordinate-based reconstruction framework, enabling accurate voxel-by-voxel sampling using physical acquisition geometry.

(ii) We propose a generalizable implicit neural representation that replaces linear averaging with a learned continuous function, preserving texture and edge fidelity between acquired slices.

2. Related Work

Implicit Neural Representations (INRs) INRs model data as continuous functions parameterized by neural networks. For images or volumes, intensity values are represented by a model that can be queried at any spatial coordinate, enabling tasks such as arbitrary-scale super-resolution (Chen et al., 2021; Wu et al., 2022). INR methods generally fall into two categories: *generalized* and *per-scene optimization* models. Generalized approaches typically employ encoder-decoder architectures trained on large datasets, allowing unseen samples to be encoded into implicit functions without retraining (Chen et al., 2021; Wu et al., 2022). Per-scene optimization methods, in contrast, fit a separate INR for each input (Sitzmann et al., 2020; Mildenhall et al., 2021). While these can achieve high quality with smooth gradients and compact representation, they require separate optimization for each scan, making them impractical for large medical volumes when rapid processing is needed. Our work builds on the generalized approach.

Ultrasound Volume Reconstruction Traditional approaches to ultrasound volume reconstruction rely on geometric transformation and interpolation. (Imran et al., 2024) used a method for micro-ultrasound that transforms oblique-plane acquisitions to axial volumes using nearest-neighbor angle selection followed by coordinate-based sampling. (Vassallo et al., 2023) proposed a robotically controlled system using similar methods with linear interpolation. While both interpretations are computationally efficient, both approaches have their drawbacks. Nearest neighbor sampling produces visible banding which causes reduced visual clarity. Linear interpolation degrades reconstructions through blurring of high frequency features through averaging. Our work adopts a similar continuous coordinate mapping but replaces the fixed interpolators with a learned implicit representation that can synthesize plausible tissue texture in the voids between acquired slices.

Ultrasound INRs Existing INR approaches for ultrasound reconstruction address three primary acquisition modalities: *robotically controlled* scans with fixed spacing (Grutman et al., 2025; Velikova et al., 2024), *tracked freehand* scans using external positioning sensors (Chen et al., 2024a; Guo et al., 2024), and *sensorless freehand* scans lacking spatial priors (Yeung et al., 2021; Chen et al., 2025). However, a critical limitation across these categories is the reliance on per-scene optimization. This requirement renders them impractical for real time clinical applications, specifically when handling high dimensional micro-ultrasound volumes (> 300 slices). To address this, we propose a generalizable, feed-forward model that generates INRs directly, bypassing the need for individual optimization while preserving visual fidelity.

3. Methods

Micro-ultrasound images are acquired by rotating a transrectal probe through the prostate, sweeping from approximately $-90°$ to $+90°$ while capturing oblique-plane slices whose angles are recorded by an internal accelerometer. This single axis of rotation seen in Figure 1 lends itself to using cylindrical coordinates. Consequently, any spatial location within the scanned volume can be described by its lateral position x along the probe axis, radial distance r from the probe center, and angular position θ within the sweep.

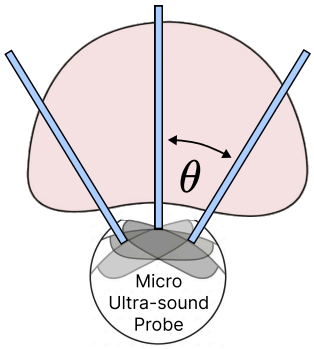

Figure 1: Diagram of transrectal micro-ultrasound slice acquisition.

To enable clinically usable 3D representations, we propose a continuous volumetric reconstruction framework that explicitly models this acquisition geometry. First, we develop a coordinate-based sampling algorithm that maps each target voxel in the reconstructed axial volume back to its corresponding location. This formulation decouples geometric mapping from the interpolation mechanism, allowing the use of advanced non-linear predictors in place of traditional interpolators. Second, we introduce a generalized INR that learns a continuous function describing the intensity field between adjacent oblique slices. Our feed-forward INR integrates naturally with the coordinate-based reconstruction algorithm,

enabling the generation of anatomically consistent structure and texture across the swept volume. Conceptually, the reconstruction process can be expressed as a function R that takes as input the set of micro-ultrasound slices, their associated angles $\boldsymbol{\theta}$, and an interpolation operator f.

3.1. Reconstruction Algorithm

Our reconstruction algorithm leverages the inherent cylindrical geometry of transrectal micro-ultrasound acquisition. Following the mapping strategy introduced by (Vassallo et al., 2023), we precompute the cylindrical coordinates (x, r, θ) associated with every voxel in the target axial volume. For each voxel located at Cartesian index (i, j, k) in the output grid, the algorithm proceeds as follows: (i) Convert to physical coordinates and compute the corresponding angle θ and radial distance r. (ii) Identify the pair of adjacent input slices I_{θ_s} and $I_{\theta_{s+1}}$ that bracket θ. (iii) Compute normalized query coordinates (x_n, r_n, z_n) for interpolation, where z_n represents the normalized position between the two slices. (iv) Query an interpolation function $f(I_{\theta_s}, I_{\theta_{s+1}}, (x_n, r_n, z_n))$ to predict the voxel intensity.

This formulation decouples the geometric reconstruction from the interpolation method. For traditional approaches, f can be trilinear interpolation. For our method, f is a learned implicit neural representation that models the continuous function between slices. Algorithm 1 in Appendix A presents the complete reconstruction procedure.

3.2. Model Architecture

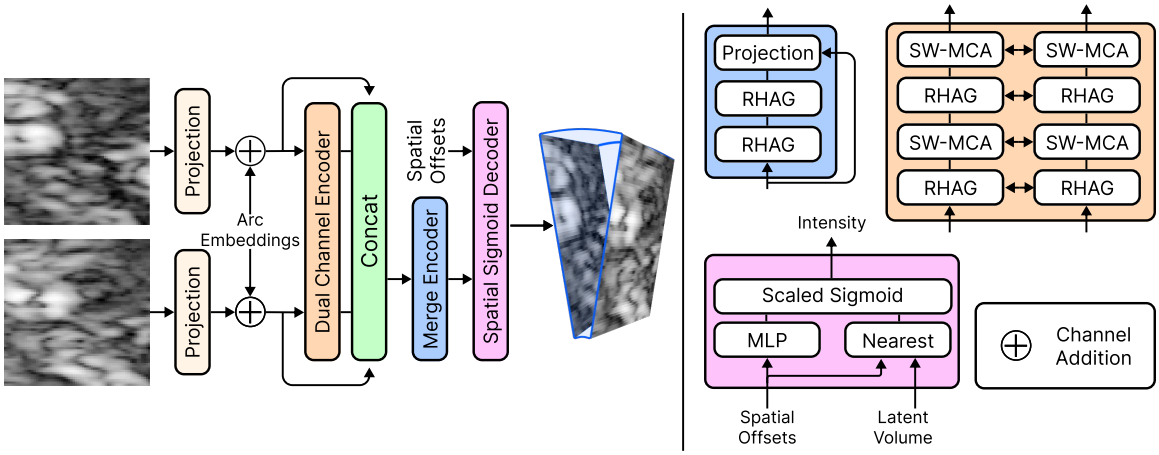

Figure 2: Overview of the proposed architecture. The model fuses features of adjacent slices to intensity values at arbitrary locations between the slices.

Given two micro-ultrasound oblique plane slices acquired at angles θ_L and θ_R, our model learns an implicit neural field representing the spatial region between them. Directly encoding full-resolution slices (1372×962) is infeasible due to memory constraints, so we instead extract corresponding patches $\mathbf{I}_L, \mathbf{I}_R \in \mathbb{R}^{p_h \times p_w \times 1}$ from each slice, where p_h and p_w

denote the patch height and width. Conditioned on these patches and query coordinates $\mathbf{x} = \langle x, r, \theta \rangle$ within the patch region, the model predicts the scalar intensity at $\mathbf{x}$, as illustrated by the blue shaded region in Fig. 2.

3.2.1. Encoder

The two image patches $\mathbf{I}_L$ and $\mathbf{I}_R$ are first projected to an embedding dimension C. To provide geometric context, we compute an arc length embedding for each pixel row. Let h denote the vertical offset of the patch within the full slice, so that the patch spans rows $[h, h + p_h)$. For row index $u \in \{0, \ldots, p_h - 1\}$, the corresponding radial location is $(h + u) \cdot s + r_{\text{probe}}$, where s is the in-plane pixel spacing and r_{probe} is the probe radius. We then scale this radius by the angular separation between the two conditioning slices and embed the result using a small network E_θ:

$$\text{ArcEmbedding}(u) = E_\theta\big(\big((h + u) \cdot s + r_{\text{probe}}\big)\,(\theta_R - \theta_L)\big) \in \mathbb{R}^C. \tag{1}$$

This embedding provides each row with a learned representation of its radial position and the angular span between the two conditioning slices.

The projected patch features are fused with their arc embeddings and processed by a dual-path encoder based on the Hybrid Attention Transformer (HAT) (Chen et al., 2023) (Fig. 2). Each path processes one conditioning slice using a sequence of residual hybrid attention group (RHAG) blocks. After each RHAG block, a shifted-window multi-head cross-attention (SWMCA) module exchanges information between the left and right paths, enabling joint reasoning across slices. Residual connections are maintained throughout to preserve low-level information. The outputs of the two paths are then concatenated along the channel dimension. The concatenated features are refined by the Merge Encoder, which applies two additional RHAG blocks followed by a residual connection and a projection layer reducing the channel dimension to $K = 128$. The resulting latent representation $\mathbf{Z} \in \mathbb{R}^{p_h \times p_w \times K}$ contains K spatially organized "expert" channels, allowing the encoder to perform the bulk of spatial reasoning while leaving the decoder to determine how to combine these expert predictions.

3.2.2. Decoder

The Spatial Sigmoid Decoder (Fig. 2) operates on the latent volume $\mathbf{Z}$ using spatial offsets, inspired by recent INR decoders that modulate local codes with coordinate-dependent weights (Chen et al., 2024b). Given a query coordinate (x, r, θ_Q), we first identify the nearest spatial latent code $\mathbf{z}_i \in \mathbf{Z}$ in the (x, r) plane. Using the grid location of $\mathbf{z}_i$, we compute the offsets (dx, dr) between the query position and the latent grid point. The relative angular position of the query is encoded as $t = (\theta_Q - \theta_L)/(\theta_R - \theta_L)$, which normalizes the query angle between the two conditioning slices.

The three offset features (dx, dr, t) are concatenated and passed through a two-layer MLP that produces a weight vector $\mathbf{w}_i \in \mathbb{R}^K$. The latent vector $\mathbf{z}_i$ and the weight vector $\mathbf{w}_i$ are then combined via a scaled sigmoid operation. Specifically, we compute the dot product, scale it by $1/\sqrt{K}$ for numerical stability, apply a learned temperature T_θ, and

finally use the sigmoid function σ to obtain a bounded intensity prediction:

$$\hat{y}(x, r, \theta_Q) = \sigma\left(\frac{\mathbf{z}_i \cdot \mathbf{w}_i}{\sqrt{K}\, T_\theta}\right). \tag{2}$$

This decoder treats $\mathbf{Z}$ as a spatially coherent latent field and uses the offsets to modulate how each local code contributes to the final prediction.

3.3. Implementation Details

3.3.1. Dataset

Our dataset consists of micro-ultrasound scans from over 150 patients acquired using the ExactVu™ Micro-Ultrasound System (Exact Imaging, Markham, ON, Canada) (Ashouri et al., 2023). Each scan contains 100-300 oblique-plane slices of the prostate and surrounding anatomy with resolution 1372×962 pixels and an associated angle measured by an accelerometer in the scanner.

Because no ground-truth reconstructed volumes exist, we train the model on slice triplets. For each patient, we iterate through slices in groups of three (left conditioning slice, query slice, right conditioning slice) and filter based on angular spacing and span. Specifically, we require that adjacent slices satisfy minimum and maximum angular gap constraints, that the overall angular span falls within a specified range, and that the triplet achieves a minimum quality score $S_{\text{quality}} = 0.3$:

$$S_{\text{quality}} = 0.7\left(1 - \frac{\delta_{\max} - \delta_{\min}}{\delta_{\max}}\right) + 0.3\left(\frac{\theta_{\text{span}}}{\Theta_{\text{limit}}}\right) \tag{3}$$

where $\delta_{\max}$ and $\delta_{\min}$ denote the larger and smaller angular intervals between the query slice and its neighbors, θ_{span} represents the total angular width of the triplet, and Θ_{limit} is the maximum allowable span defined by our initial filter.

This quality metric penalizes both highly asymmetric triplets and triplets with very narrow angular spans, both of which provide poor signal as the ground truth and conditioning slices will be very similar. After filtering, we obtain approximately 30,000 triplets split 60/20/20 across training, validation, and test sets by patient to prevent data leakage. During training, patches are extracted at random spatial locations from the triplets.

3.3.2. Loss Function

Our loss function consists of $\mathcal{L}_1$ reconstruction loss ($\lambda_1 = 1$), LPIPS perceptual loss (Zhang et al., 2018) ($\lambda_{\text{perc}} = 0.1$), and patch-based adversarial loss (Isola et al., 2017) ($\lambda_{\text{adv}} = 0.005$). This is a standard image restoration loss formula balancing pixel-wise discrepancies, larger perceptual feature, and realism. To enforce self-consistency, we minimize error on the query slice as well as the conditioning slices (scaled by $\lambda_{\text{LR}} = 0.125$). We define the combined loss $\mathcal{L}_{\text{view}}$ for a predicted patch $\hat{\mathbf{I}}$ and ground truth $\mathbf{I}$ as:

$$\mathcal{L}_{\text{view}}(\hat{\mathbf{I}}, \mathbf{I}) = \lambda_1 \|\hat{\mathbf{I}} - \mathbf{I}\|_1 + \lambda_{\text{perc}} \mathcal{L}_{\text{perc}}(\hat{\mathbf{I}}, \mathbf{I}) + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}}(\hat{\mathbf{I}}, \mathbf{I}) \tag{4}$$

The total training objective sums the losses for the query (Q), left (L), and right (R) targets:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{view}}(\hat{\mathbf{I}}_Q, \mathbf{I}_Q) + \lambda_{\text{LR}} \left[\mathcal{L}_{\text{view}}(\hat{\mathbf{I}}_L, \mathbf{I}_L) + \mathcal{L}_{\text{view}}(\hat{\mathbf{I}}_R, \mathbf{I}_R) \right] \tag{5}$$

where $\hat{\mathbf{I}}_{\mathbf{x}} = f_\theta(\mathbf{x}, \mathbf{I}_{\text{L}}, \mathbf{I}_{\text{R}})$ represents the model prediction at coordinate $\mathbf{x}$. Further training details can be found in Appendix B.

4. Results

4.1. Quantitative Results

We exclude patches composed predominantly of homogeneous speckle, as standard image-quality metrics become unstable when both prediction and ground truth contain largely stochastic texture. This is a practical evaluation choice rather than a limitation of the method, since such regions carry minimal anatomical information.

We evaluate reconstruction quality using metrics that capture both general image fidelity and ultrasound-specific characteristics. Peak Signal-to-Noise Ratio (PSNR) measures pixel-wise reconstruction accuracy, while the Structural Similarity Index (SSIM) reflects perceptual fidelity and preservation of clinically relevant structure. The Speckle Similarity Index (SSI) quantifies retention of local texture patterns characteristic of ultrasound, with higher values indicating reduced over-smoothing. The Edge Preservation Index (EPI) assesses the sharpness and continuity of anatomical boundaries. Improvements across these metrics indicate stronger preservation of both structural detail and modality-specific speckle texture (Singh et al., 2009).

Table 1: Quantitative comparison of patch-level reconstruction quality.

Method	↑ PSNR	↑ SSIM	↑ SSI	↑ EPI
Nearest	20.83 ± 2.96	0.46 ± 0.15	0.97 ± 0.03	0.61 ± 0.12
Trilinear	23.26 ± 2.93	0.57 ± 0.15	0.97 ± 0.02	0.72 ± 0.10
Ours	**23.68 ± 2.98**	**0.62 ± 0.15**	**0.99 ± 0.02**	**0.73 ± 0.11**

We compare our model against a coordinate-based Trilinear baseline (the reconstruction method used in (Vassallo et al., 2023) applied to our data). Because of linear interpolation's quality of acting as a low pass filter, it fails to resolve complex structural transitions. This is evidenced in Table. 1 by the baseline's lower SSIM (0.57), indicating that averaging pixel intensities obscures anatomical boundaries. Replacing the linear interpolator with our INR yields a 9% relative improvement in SSIM (0.62). Crucially, this structural gain does not come at the cost of texture. Our high SSI (0.99) and EPI (0.73) scores confirm that the model preserves the statistical properties of the ultrasound signal slightly better than the baseline, but organizes them into more coherent anatomical structures.

4.2. Visual Results

Figure 3 illustrates these improvements visually. Our model achieves better accuracy in reconstructing anatomical structures, maintaining crisp boundaries between hypo-echoic and

hyper-echoic regions. In contrast, trilinear interpolation exhibits visible over-smoothing and intensity mismatches, particularly evident at tissue boundaries in the highlighted regions, corresponding to the lower SSIM values in Table 1.

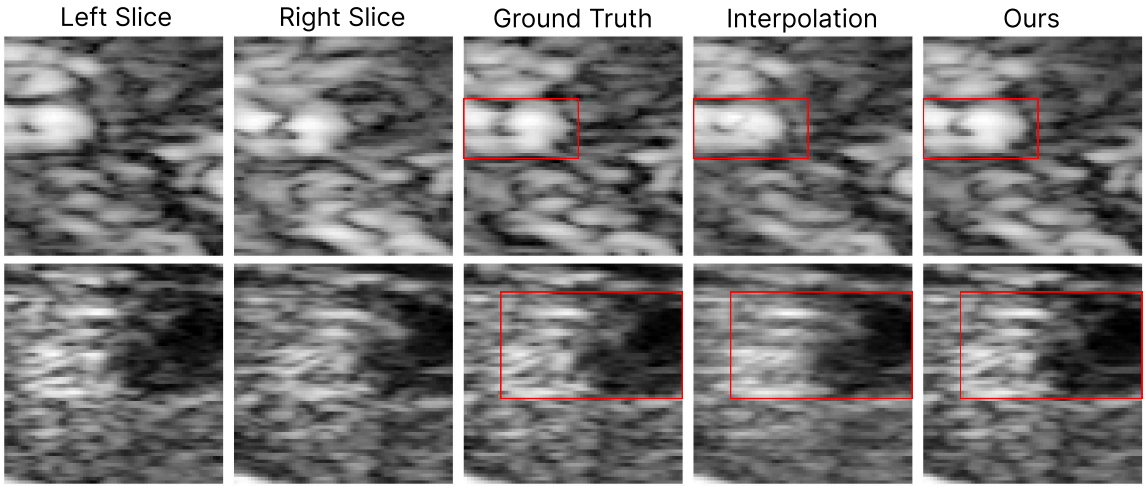

Figure 3: Visual comparison of patch level reconstruction quality. Our method recovers sharp structural boundaries and speckle texture that are otherwise smoothed out by linear interpolation.

Figure 4 presents the full volume reconstructions comparing our method against prior approaches. On the left, Imran et al.'s nearest-neighbor method (Imran et al., 2024) exhibits the blocky, banded artifacts discussed in the introduction; visible stair-stepping in textured regions obscures boundaries and limits clinical interpretability. The center panel demonstrates that while trilinear interpolation (Vassallo et al., 2023) creates a more coherent volume, it introduces over-smoothing and a streak-like blur. This is particularly evident in the highlighted regions where high-frequency details are averaged out. In contrast, our method on the right achieves superior visual fidelity, eliminating banding artifacts while preserving both structural boundaries and characteristic ultrasound speckle texture.

4.3. Ablation Study

We evaluate the impact of two key architectural components of the implicit neural representation (INR): the arc length embeddings and the shifted-window multi-head cross-attention (SWMCA) blocks. All quantitative results are reported at the patch level.

Effectiveness of Arc Embeddings The arc length embeddings explicitly encode the radial position of each patch row relative to the probe and the angular span between the conditioning slices. As shown in Table 2, removing these embeddings (*No Arc Embed*) leads to a small but consistent decrease in PSNR (approximately 0.07 dB) compared to the full model, while the other metrics remain effectively unchanged. Thus, the modest improvement in addition to the minimal overhead compelled us retain the addition in our final architecture.

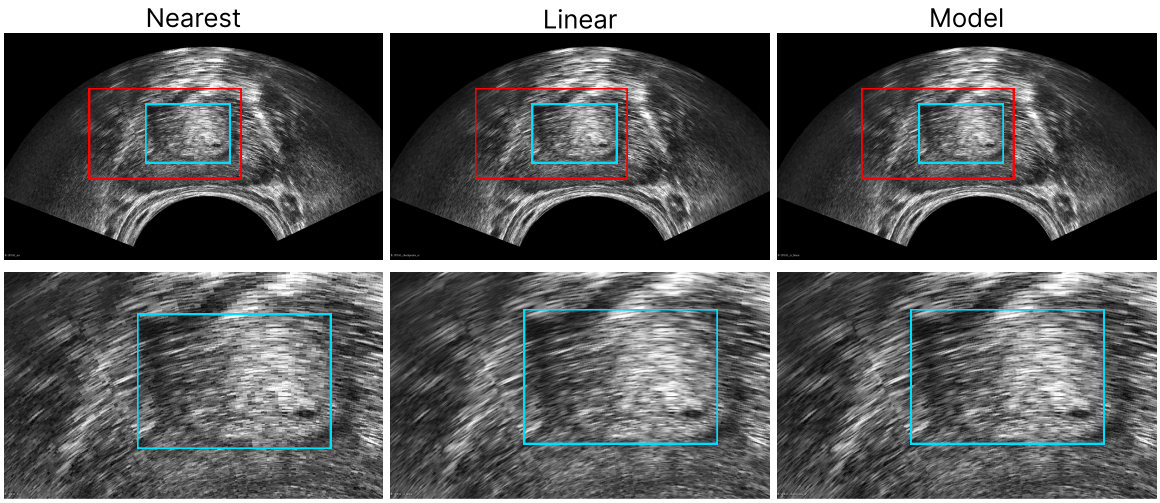

Figure 4: Axial cross-sections of the full volume reconstruction. Our approach eliminates the banding artifacts of nearest-neighbor selection and reduces blurring inherent to linear interpolation.

Effectiveness of SWMCA Blocks The SWMCA blocks promote interaction between the two conditioning slices within the dual-channel encoder. When we remove SW-MCA and compensate by adding an additional SW-MSA block to keep the parameter count comparable (*No SWMCA*), the patch-level metrics degrade slightly in both PSNR and SSIM (Table 2). More importantly, training becomes noticeably less stable. Under identical training settings, the model without SW-MCA exhibits intermittent regressions in validation performance, causing the performance degradation. This suggests that SW-MCA primarily benefits optimization stability and, secondarily, reconstruction quality. Given these stability gains and the modest quantitative improvements, we keep SW-MCA in the final model.

Table 2: Quantitative comparison of patch-level reconstruction quality.

Method	↑ PSNR	↑ SSIM	↑ SSI	↑ EPI
Ours	**23.68 ± 2.98**	**0.62 ± 0.15**	**0.99 ± 0.02**	**0.73 ± 0.11**
No Arc Embed	23.61 ± 3.00	**0.62 ± 0.16**	**0.99 ± 0.02**	**0.73 ± 0.11**
No SWMCA	23.48 ± 2.95	0.60 ± 0.16	**0.99 ± 0.02**	0.72 ± 0.11

5. Discussion

5.1. Interpretation of Results

The results illustrate clear trade-offs among the reconstruction strategies. Nearest-neighbor sampling preserves high-frequency speckle but produces blocky, angularly discontinuous

volumes. Trilinear interpolation improves smoothness yet oversmooths micro-ultrasound data due to its low-pass behavior. Our method balances these extremes: modest gains in PSNR, SSI, and EPI indicate statistical fidelity to the original acquisitions, while the larger improvement in SSIM reflects superior structural preservation. Together, these findings suggest that the INR maintains ultrasound-specific texture while producing geometrically coherent volumes suitable for downstream analysis.

5.2. Relation to Prior Approaches

Relative to existing micro-ultrasound reconstruction pipelines (Vassallo et al., 2023; Imran et al., 2024), our framework retains the same geometry-aware mapping but replaces the fixed interpolator with a learned, feed-forward implicit representation. Unlike ultrasound INR methods that require per-scene optimization (Grutman et al., 2025; Velikova et al., 2024; Chen et al., 2024a; Guo et al., 2024; Yeung et al., 2021; Chen et al., 2025), our approach scales to full clinical volumes without per-scan training. It can therefore serve as a drop-in, data-driven substitute for interpolation in current pipelines, preserving the physical correctness of the mapping while mitigating the blurring and banding artifacts inherent to classical interpolation.

5.3. Limitations

Our study has three key limitations. First, the dataset was acquired on a single micro-ultrasound system (ExactVu™), limiting assessment of generalization across scanners and probe geometries. Second, evaluation focuses on image-quality metrics; the impact on clinically relevant tasks such as lesion detection or biopsy guidance remains unknown. Third, due to the absence of full-volume ground truth, quantitative assessment is restricted to patch-level evaluation, preventing direct measurement of global volumetric consistency.

5.4. Future Work

As part of our future work, we will evaluate generalization across scanner platforms, incorporate task-based and clinical assessments, and explore end-to-end integration with downstream applications such as MRI or histopathology registration. Investigating lighter-weight or real-time variants of the INR can also be important for clinical deployment.

6. Conclusion

In this paper, we present a coordinate-based reconstruction framework combined with a generalizable implicit neural representation for micro-ultrasound volume reconstruction. We demonstrate that deep learning is not merely an enhancement but a necessity for preserving the high-resolution nature of micro-ultrasound during 3D reconstruction. While traditional trilinear interpolation degrades the image through low-pass filtering, our method achieves higher SSIM and PSNR while maintaining the high-frequency texture (SSI, EPI) characteristic of the modality. Our feed-forward INR approach provides geometrically consistent, high-fidelity volumes suitable for downstream clinical analysis and histopathology registration.

Appendix A. Reconstruction Algorithm

Algorithm 1: Conceptual 3D Micro-US Reconstruction

Input: Micro-US slices $I_{\theta_1}, \ldots, I_{\theta_S}$, geometry params $(H, W, \Delta_{xy}, r_{\text{probe}})$, output dims $(N_{\text{LR}}, N_{\text{AP}}, N_{\text{SI}})$, interpolation model f
Output: $V \in [0, 255]^{N_{\text{LR}} \times N_{\text{AP}} \times N_{\text{SI}}}$, the reconstructed 3D volume

$T_{\text{voxel}\to\text{phys}} \leftarrow$ GetVoxelToPhysicalTransform(Inputs...) `// Precompute coord maps`
$T_{\text{phys}\to\text{probe}} \leftarrow$ GetPhysicalToProbeTransform(Inputs...)
for $i \leftarrow 1$ **to** N_{LR} **do**
  for $j \leftarrow 1$ **to** N_{AP} **do**
    for $k \leftarrow 1$ **to** N_{SI} **do**
      $(x, y, z) \leftarrow T_{\text{voxel}\to\text{phys}}(i, j, k)$ `// Map voxel to probe coords`
      $(\theta, u, v) \leftarrow T_{\text{phys}\to\text{probe}}(x, y, z)$ `//` (θ, u, v) `= angle, lat-pixel, rad-pixel`
      if $u \notin [0, W)$ ***or*** $v \notin [0, H)$ **then** $V[i, j, k] \leftarrow 0$; **continue**
      Find s such that $\theta_s \le \theta \le \theta_{s+1}$ `// Generate normalized query point`
      $(x_n, r_n) \leftarrow (2(u + 0.5)/W - 1,\ 2(v + 0.5)/H - 1)$
      $t \leftarrow (\theta - \theta_s)/(\theta_{s+1} - \theta_s)$
      $z_n \leftarrow \text{clip}(t, 0, 1) - 0.5$
      $y_{\text{pred}} \leftarrow f(I_{\theta_s}, I_{\theta_{s+1}}, (x_n, r_n, z_n))$ `// Query model and store result`
      $V[i, j, k] \leftarrow \text{round}(255 \cdot \text{clip}(y_{\text{pred}}, 0, 1))$
    end
  end
end
return V

The algorithm as presented operates conceptually on individual voxels for clarity. In practice, the computation is vectorized and implemented with GPU-accelerated batch processing for efficiency.

Appendix B. Training Details

Our model was trained using two NVIDIA B200 GPUs.

B.1. Data Preprocessing

After analyzing the distribution of the adjacent angle distance over our dataset, we chose the following parameters:

```
MAX_ADJACENT_GAP = 1.3
MIN_ADJACENT_GAP = 0.1
MAX_OVERALL_SPAN = 1.8
MIN_OVERALL_SPAN = 0.4
MIN_QUALITY_SCORE = 0.2
```

B.2. Model Parameters

The following configuration was used for the model:

```
in_dim: 1
K: 128
embed_dim: 128
init_cab_weight: 0.1
cab_channel_reduction: 16
squeeze_factor: 4
H: 64
W: 64
num_heads: 4
window_size: 8
mlp_ratio: 4
num_rhag_blocks: 2
overlap_ratio: 0.5
num_hab_blocks: 2
```

B.3. Discriminator Parameters

The following configuration was used for the discriminator:

```
in_channels: 1
base_channels: 64
discriminator_lr: 1e-4
discriminator_min_lr: 1e-6
discriminator_warmup_epochs: 10
```

B.4. Training and Data Parameters

The following configuration was used for training and data loading:

```
patch_size: [64, 64]
batch_size: 128
lr: 1e-3
min_lr: 1e-5
warmup_epochs: 10
weight_decay: 0.01
num_epochs: 500
lpips_weight: 0.01
gan_weight: 0.005
lr_weight: 0.125
```

References

Rani Ashouri, Brianna Nguyen, Jeremy Archer, Paul Crispen, Padraic O'Malley, Li-Ming Su, Joseph Grajo, Sara M Falzarano, Yahya Acar, David Lizdas, et al. Micro-ultrasound

guided transperineal prostate biopsy: a clinic-based procedure. *J Vis Exp*, 10:64772, 2023.

Adriano Basso Dias and Sangeet Ghai. Micro-ultrasound: current role in prostate cancer diagnosis and future possibilities. *Cancers*, 15(4):1280, 2023.

Hongbo Chen, Logiraj Kumaralingam, Shuhang Zhang, Sheng Song, Fayi Zhang, Haibin Zhang, Thanh-Tu Pham, Kumaradevan Punithakumar, Edmond HM Lou, Yuyao Zhang, et al. Neural implicit surface reconstruction of freehand 3d ultrasound volume with geometric constraints. *Medical Image Analysis*, 98:103305, 2024a.

Xiangyu Chen, Xintao Wang, Wenlong Zhang, Xiangtao Kong, Yu Qiao, Jiantao Zhou, and Chao Dong. Hat: Hybrid attention transformer for image restoration. *arXiv preprint arXiv:2309.05239*, 2023.

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

Yinbo Chen, Oliver Wang, Richard Zhang, Eli Shechtman, Xiaolong Wang, and Michael Gharbi. Image neural field diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8007–8017, 2024b.

Zhurong Chen, Jinhua Chen, Wei Zhuo, Wufeng Xue, and Dong Ni. 3d heart reconstruction from sparse pose-agnostic 2d echocardiographic slices. In *International Workshop on Advances in Simplifying Medical Ultrasound*, pages 33–42. Springer, 2025.

Tal Grutman, Mike Bismuth, Bar Glickstein, and Tali Ilovitsh. Implicit neural representation for scalable 3d reconstruction from sparse ultrasound images. *npj Acoustics*, 1(1):14, 2025.

Ziwen Guo, Zi Fang, and Zhuang Fu. Ulre-nerf: 3d ultrasound imaging through neural rendering with ultrasound reflection direction parameterization. *arXiv preprint arXiv:2408.00860*, 2024.

Muhammad Imran, Brianna Nguyen, Jake Pensa, Sara M. Falzarano, Anthony E. Sisk, Muxuan Liang, John Michael DiBianco, Li-Ming Su, Yuyin Zhou, Jason P. Joseph, Wayne G. Brisbane, and Wei Shao. Image registration of in vivo micro-ultrasound and ex vivo pseudo-whole mount histopathology images of the prostate: A proof-of-concept study. *Biomedical Signal Processing and Control*, 96:106657, 2024. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2024.106657. URL https://www.sciencedirect.com/science/article/pii/S1746809424007158.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

Laurence Klotz, Giovanni Lughezzani, Davide Maffei, Andrea Sánchez, José Gregorio Pereira, Frédéric Staerman, Hannes Cash, Ferdinand Luger, Laurent Lopez, Rafael

Sanchez-Salas, et al. Comparison of micro-ultrasound and multiparametric magnetic resonance imaging for prostate cancer: A multicenter, prospective analysis. *Canadian Urological Association Journal*, 15(1):E11, 2020.

Giovanni Lughezzani, Alberto Saita, Massimo Lazzeri, Marco Paciotti, Davide Maffei, Giuliana Lista, Rodolfo Hurle, Nicolò Maria Buffi, Giorgio Guazzoni, and Paolo Casale. Comparison of the diagnostic accuracy of micro-ultrasound and magnetic resonance imaging/ultrasound fusion targeted biopsies for the diagnosis of clinically significant prostate cancer. *European urology oncology*, 2(3):329–332, 2019.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Jake Pensa, Wayne Brisbane, Adam Kinnaird, David Kuppermann, Griffith Hughes, Derrick Ushko, Alan Priester, Samantha Gonzalez, Robert Reiter, Arnold Chin, et al. Evaluation of prostate cancer detection using micro-ultrasound versus mri through co-registration to whole-mount pathology. *Scientific Reports*, 14(1):18910, 2024.

Mandeep Singh, Sukhwinder Singh, and Savita Kansal. Comparative analysis of spatial filters for speckle reduction in ultrasound images. In *2009 WRI World Congress on Computer Science and Information Engineering*, volume 6, pages 228–232. IEEE, 2009.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

Reid Vassallo, Tajwar Abrar Aleef, Qi Zeng, Brian Wodlinger, Peter C Black, and Septimiu E Salcudean. Robotically controlled three-dimensional micro-ultrasound for prostate biopsy guidance. *International Journal of Computer Assisted Radiology and Surgery*, 18 (6):1093–1099, 2023.

Yordanka Velikova, Mohammad Farid Azampour, Walter Simson, Marco Esposito, and Nassir Navab. Implicit neural representations for breathing-compensated volume reconstruction in robotic ultrasound. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1316–1322. IEEE, 2024.

Qing Wu, Yuwei Li, Yawen Sun, Yan Zhou, Hongjiang Wei, Jingyi Yu, and Yuyao Zhang. An arbitrary scale super-resolution approach for 3d mr images via implicit neural representation. *IEEE Journal of Biomedical and Health Informatics*, 27(2):1004–1015, 2022.

Pak-Hei Yeung, Linde Hesse, Moska Aliasi, Monique Haak, Weidi Xie, Ana IL Namburete, INTERGROWTH 21st Consortium, et al. Implicitvol: Sensorless 3d ultrasound reconstruction with deep implicit representation. *arXiv preprint arXiv:2109.12108*, 2021.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.