

Multivariate Gaussian NeRF for Wide Field-of-View Ultrasound Reconstruction

Patris Valera¹*, Magdalena Wysocki^{1,2}* **, Felix Duelmer^{1,2}, Mohammad Farid Azampour^{1,2}, Sebastian Herz³, Stefan Wörz³, and Nassir Navab^{1,2}

¹ Chair for Computer Aided Medical Procedures (CAMP), Technical University of Munich, Germany

{patris.valera, magdalena.wysocki, felix.duelmer, mf.azampour, nassir.navab}@tum.de

² Munich Center for Machine Learning (MCML), Munich, Germany

³ LUMA Vision, Beech Hill Road, Dublin, Ireland

<http://www.lumavision.com>
{sebastian.herz, stefan.woerz}@lumavision.com

Abstract. Wide Field-of-View (WFoV) reconstruction enhances 3D ultrasound imaging by providing valuable anatomical context for segmentation models and visualization. Clinical ultrasound volumes are predominantly acquired using convex probes, which generate expanding, diverging acoustic beams to maximize anatomical coverage. Stitching these sweeps together traditionally introduces significant compounding artifacts and aliasing due to depth-dependent resolution changes. Here, we introduce Ultra-Wide-NeRF, a Multivariate 3D Gaussian (MVG) NeRF-based method for WFoV ultrasound reconstruction. By explicitly modeling the complex beam geometry using distance-dependent convex volumetric sampling and anisotropic 3D Gaussians, our method inherently mitigates these compounding artifacts and provides anti-aliasing. Beyond simply reconstructing a static 3D grid, our NeRF-based approach yields a continuous neural representation of the tissue, enabling the synthesis of high-fidelity novel views from arbitrary virtual trajectories. We validate Ultra-Wide-NeRF for intracardiac echocardiography on phantom and porcine datasets, demonstrating that our method expands the spatial context important in intraoperative navigation. Code is available at <https://github.com/patrisvalera/ultra-wide-nerf>.

Keywords: Reconstruction · Ultrasound · Echocardiography

1 Introduction

Visualizing large, complex anatomical structures in their entirety is critical for accurate clinical diagnosis and intraoperative navigation. While three-dimensional (3D) ultrasound (US) provides a real-time, accessible, and radiation-free imaging modality [6, 10], individual acquisitions are physically constrained

* These authors contributed equally to this work.

** Corresponding author

to a limited Field-of-View (FoV). To overcome this restriction and achieve the necessary macroscopic anatomical context, multiple overlapping 3D US acquisitions can be stitched together using spatial compounding. This process yields a Wide FoV (WFOV) panoramic volume that captures extended anatomical regions. Conventional WFOV compounding methods align overlapping 3D sequences using feature- or intensity-based registration [7, 9, 11], which inevitably introduces interpolation artifacts. Recent Neural Radiance Fields (NeRFs) for ultrasound provide a continuous, physics-aware alternative. Approaches like Ultra-NeRF [14], NeRF-US [4], and UIRe-NeRF [8] leverage implicit representations mapped via ray tracing [2, 5, 13]. By evaluating the entire acoustic propagation path rather than assigning static voxel intensities, these models inherently avoid traditional interpolation artifacts. However, these models rely on equidistant point sampling along infinitesimally thin rays, which fundamentally fails for the diverging geometries of clinical convex probes, where the physical volume sampled by the acoustic pulse expands proportionally with depth [3, 12]. A concurrent approach [16] mitigates this with conical Gaussian sampling and integrated positional encoding, but models the beam as an axially symmetric cone over single 2D sweeps, leaving the cross-plane elevational resolution loss and far-field aliasing of multi-sweep 3D convex fusion panoramic unaddressed.

Here, we introduce Multivariate Gaussian NeRF for Wide Field-of-View Ultrasound Reconstruction (Ultra-Wide-NeRF), bridging NeRFs with convex geometry by aligning the sampling strategy directly with how the diverging ultrasound beam behaves in space. By replacing standard thin-ray approximations with distance-dependent convex volumetric sampling and multivariate Gaussians, we explicitly model the expanding spatial footprint of the acoustic beam. This continuous formulation naturally captures depth-dependent resolution, acting as an inherent anti-aliasing filter that eliminates traditional stitching artifacts. Consequently, Ultra-Wide-NeRF successfully fuses multiple sweeps into a seamless panoramic volume while enabling realistic novel view synthesis from unobserved trajectories. Validated on intracardiac echocardiography (ICE), our framework delivers comprehensive spatial context in two distinct ways: NeRF-based panoramic reconstruction effectively reduces stitching artifacts and enhances anatomical borders to extend context along the acquired trajectory, while high-fidelity novel view synthesis generates views from unobserved virtual perspectives. Ultimately, this combined, high-quality panoramic reconstruction establishes a robust foundation of visual context for seamless intraoperative cardiac navigation.

2 Methodology

Our objective is to reconstruct a unified, WFOV representation from a sequence of N overlapping, tracked 3D US volumes $\{\mathcal{V}_i \in \mathbb{R}^{D \times H \times W}\}_{i=1}^N$ with corresponding poses $\{\mathbf{P}_i \in SE(3)\}_{i=1}^N$, where D, H, W are the volume dimensions. As shown in Fig. 1, the proposed method optimizes a continuous physics-informed neural field and uses Multivariate 3D Gaussians (Sec. 2.3) to define a WFOV US vol-

ume $\mathcal{PAN} \in \mathbb{R}^{(D \times N) \times H \times W}$ that spans the entire catheter trajectory. We define the coordinate system such that the z -axis is aligned with the catheter length (elevational direction, or volume-depth D), the y -axis with the lateral direction of the ultrasound beam, and the x -axis with the axial (scan-depth) direction. We refer to z as the depth direction in the volumetric sense.

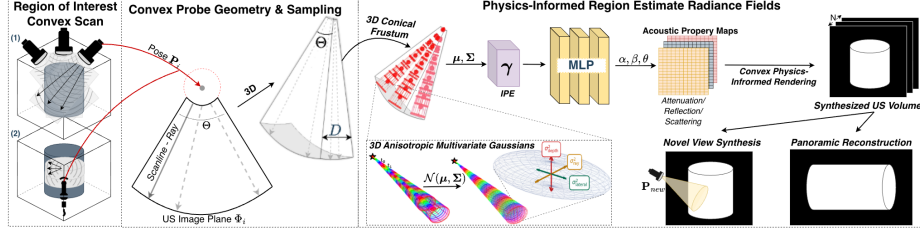


Fig. 1. Overview of the proposed method for unified, WFoV Volume Reconstruction. We model a convex transducer by radially shooting rays from an opening angle Θ which are then mapped to 3D Cartesian world space via their associated poses. Afterwards, **Ultra-Wide-NeRF** estimates a physics-informed continuous radiance field by fitting multivariate 3D Gaussians along the three resolution axes with an explicit attenuation model. Ultra-Wide-NeRF generates both WFoV Volume Reconstruction and novel view synthesis from unseen probe poses.

2.1 Physics-Based Neural Ultrasound Fields

In NeRF’s for ultrasound a coordinate-based network maps a 3D location $\mathbf{x} \in \mathbb{R}^3$ to a vector of acoustic parameters $\tau \in \mathbb{R}^d$, which serves as an input to an acoustic forward synthesis model. This model computes voxel intensity at the location $\mathbf{x}$ assuming a linear transducer casting parallel rays $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ sampled at discrete points t (Fig. 2). This sampling strategy is not optimal for convex and volumetric probes since the physical gaps between scanlines expand with depth, causing severe far-field aliasing.

2.2 Convex Probe Geometry

For the **2D convex FoV**, the domain is defined in polar coordinates (r, θ) where r defines range and θ defines azimuth on an image plane Φ_i corresponding to the pose $\mathbf{P}_i$. The convex geometry is parameterized by an inner radius R_{in} , an outer radius R_{out} , and an opening angle $\Theta \in (0^\circ, 360^\circ]$. In a simple case (Fig 2) the rays are cast uniformly over $\theta \in [-\frac{\Theta}{2}, \frac{\Theta}{2}]$ and sampled along $r \in [R_{in}, R_{out}]$. To evaluate these samples within the NeRF architecture, polar samples are transformed into Cartesian coordinates $(x, y, z) := (r \sin(\theta), r \cos(\theta) - R_{in}, z)$ and positioned in $\mathbb{R}^3$ using the pose $\mathbf{P}_i$.

To extend the convex probe geometry to 3D, we adopt a multi-planar acquisition model rather than an explicit spherical parameterization. Each volume

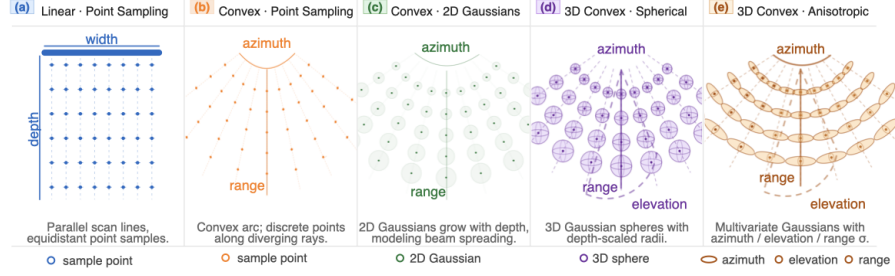


Fig. 2. Ultrasound beam sampling strategies. (a) Parallel point sampling for linear arrays. (b) Diverging point sampling for convex arrays, highlighting depth-dependent under-sampling. (c) 2D Gaussian sampling, capturing the in-plane profile. (d) Isotropic 3D Gaussian sampling, adding the elevation footprint. (e) Proposed multivariate 3D Gaussian sampling, accurately modeling the anisotropic volumetric expansion of diverging beams.

$\mathcal{V}_i$ is formed by stacking D fan-beam (convex) slices, each defined on its own image plane $\Phi_{i,d}$ with associated pose $\mathbf{P}_{i,d}$, where $d \in [0, D)$. Within each slice, rays are cast in polar coordinates (r, θ) as described above and lifted into 3D Cartesian world coordinates via a two-step mapping: first, the in-plane polar-to-Cartesian conversion and second, the rigid-body transform $\Phi_{i,d}$ that positions each slice plane in the shared world frame. The elevational dimension is implicitly defined by the inter-slice pose differences $\|\mathbf{P}_{i,d+1} - \mathbf{P}_{i,d}\|$, which encode the angular or translational sweep of the acquisition. All ray samples ultimately query the neural field in Cartesian world coordinates, enabling seamless integration of multi-planar acquisitions within a single continuous volumetric representation.

2.3 From Points to Multivariate 3D Gaussians

Approximating the interaction between an acoustic wave and a tissue scatterer as a discrete point estimate is fundamentally ill-posed. This approach disregards the finite volume each sample actually represents, inevitably leading to aliasing. Hence, we model each sampled region as a conical frustum parameterized by a 3D multivariate Gaussian to create a continuous space. Since 3D US imaging exhibits fundamental anisotropy in both lateral (in-plane) and elevational (cross-plane) directions, we extend the mip-NeRF’s [1] frustum parameterization to anisotropic Multivariate 3D Gaussians with independent radii for each spatial direction (Fig 1).

In Ultra-Wide-NeRF each sample at distance t along a ray $(\mathbf{o}, \mathbf{d})$ is represented as a frustum segment $[t_j, t_{j+1}]_{j \in \{0, \#samples\}}$ and encoded by a Gaussian $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean distance along the ray $\mu_t = \mu + \frac{2\mu\delta^2}{3\mu^2 + \delta^2}$ where $\mu = (t_j + t_{j+1})/2$ is the segment midpoint and $\delta = (t_{j+1} - t_j)/2$ is set to the segment’s half-width.

Variance decomposition. The covariance Σ is a full 3×3 matrix constructed from three distinct variances, encoding the spatial footprint of this region. Following [1], σ_{ray}^2 and $\sigma_{\text{lateral}}^2$ capture the along-ray and horizontally perpendicular-to-ray (lateral) variance. We additionally introduce σ_{depth}^2 to model the cross-plane elevational anisotropy of 3D US (vertically perpendicular-to-ray variance):

$$\sigma_{\text{ray}}^2 = \frac{\delta^2}{3} - \frac{4}{15} \cdot \frac{\delta^4(12\mu^2 - \delta^2)}{(3\mu^2 + \delta^2)^2} \quad (1)$$

$$\sigma_{\{\text{lateral}, \text{depth}\}}^2 = r_{\{\text{lat}, \text{dep}\}}^2 \left(\frac{\mu^2}{4} + \frac{5\delta^2}{12} - \frac{4\delta^4}{15(3\mu^2 + \delta^2)} \right) \quad (2)$$

where r_{lat} and r_{dep} are independent Gaussian base radii derived from the in-plane and elevational resolutions, s_{lat} and s_{dep} , respectively, as $r_{\{\text{lat}, \text{dep}\}} = \frac{2 \cdot s_{\{\text{lat}, \text{dep}\}}}{\sqrt{12}}$. They represent the voxel’s footprint in world coordinates, similar to pixels’ footprint in [1]. (Fig. 1)

Covariance construction. The 3×3 covariance matrix is assembled in a local orthonormal basis aligned with the ray geometry and then rotated to world coordinates:

$$\Sigma = \mathbf{B} \text{diag}(\sigma_{\text{ray}}^2, \sigma_{\text{lateral}}^2, \sigma_{\text{depth}}^2) \mathbf{B}^\top \quad (3)$$

where $\mathbf{B} = [\hat{\mathbf{d}}, \mathbf{v}_1, \mathbf{v}_2] \in \mathbb{R}^{3 \times 3}$ is an orthonormal basis with $\hat{\mathbf{d}}$ the ray direction, $\mathbf{v}_1$ a lateral vector (horizontally perpendicular to the ray, primarily in the imaging plane), and $\mathbf{v}_2$ a depth vector (perpendicular to both, aligned with the elevational direction).

Integrated positional encoding. To reduce aliasing, we encode each frustum segment as a Gaussian footprint $\mathcal{N}(\mu, \Sigma)$ and apply Integrated Positional Encoding (IPE, γ) [1]. Unlike standard positional encoding, which evaluates sine/cosine features at a point, IPE averages these features over the footprint. This yields the usual multi-frequency features, but with frequency-dependent attenuation: high frequencies are suppressed when the footprint is large, while low frequencies are preserved.

In Ultra-Wide-NeRF we derive Σ from the ray-, lateral-, and elevational (volumetric depth) variances, and use the per-axis variances (the diagonal of Σ) to control this attenuation. Because the footprint is anisotropic by construction, the attenuation is direction-dependent: frequencies are damped more along directions with lower resolution (larger variance) and preserved along directions with higher resolution. This embeds the ultrasound sampling footprint directly into the network input representation and mitigates aliasing artifacts.

2.4 3D-aware Loss Function

All rays query a neural field in Cartesian world coordinates, resulting in voxel footprint in 3D space. We introduce 3D-informed loss that enforces 3D structural consistency via 3D loss functions, and Ultra-NeRF regularization [14, 15], applied to each full-stacked volume during training. The final loss is defined as:

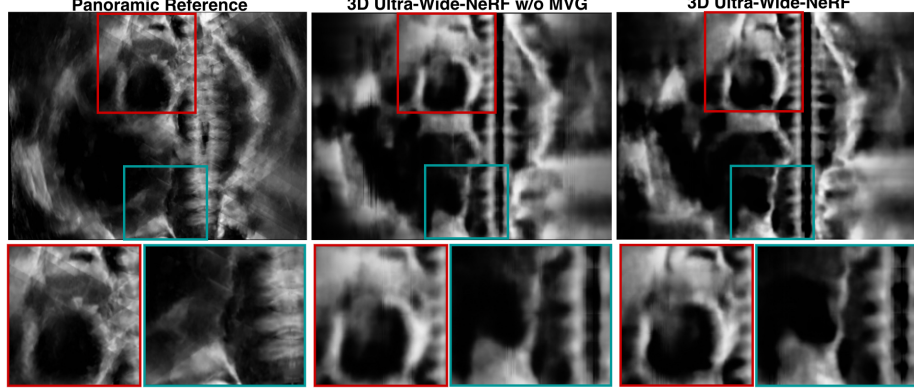


Fig. 3. Comparison of WFOV Panoramic ICE Volume Reconstruction using Ultra-Wide-NeRF. The first column shows the SOTA geometry-based reconstruction, the second one shows the voxel-based 3D Ultra-Wide-NeRF without MVG, and the last one shows the Ultra-Wide-NeRF using region estimates MVG. The second row illustrates close-up regions of stitching and aliasing artifacts removed by the region estimates.

$$\mathcal{L} = (1 - \lambda) \mathcal{L}_{\text{MSE}}(\hat{V}, V) + \lambda \mathcal{L}_{\text{SSIM}} + \lambda_r \mathcal{L}_{\text{reg}} \quad (4)$$

where $\hat{V}$ is the rendered volume, V is the ground truth volume, $\mathcal{L}_{\text{MSE}} = \|\hat{V} - V\|_2^2$ is the mean squared error, $\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}(\hat{V}, V)$ is the structural similarity loss with a 3×3 Gaussian kernel, and $\lambda = 0.2$ controls the balance. Additionally, after a warm-up period, we apply a finite-difference gradient magnitude loss along all three spatial directions to preserve edge sharpness and speckle texture:

$$\mathcal{L}_{\text{grad}} = \|\|\nabla \hat{V}\|_2 - \|\nabla V\|_2\|_1 \quad (5)$$

where ∇ denotes the finite-difference gradient computed along the x -, y -, and z -axes, and $\|\cdot\|_2$ is the per-voxel gradient magnitude.

3 Experiments and Results

Data. We evaluate Ultra-Wide-NeRF on two 3D ICE datasets, a static silicon phantom and an in vivo porcine acquisition. Each dataset contains 3D overlapping ICE volume datasets. We compare Ultra-Wide-NeRF against registration based, geometry-aware panoramic reconstruction for ICE [9]. Both datasets are acquired with a 360° FoV ICE catheter (20 mm length, 90 mm radius) that enters the inferior vena cava (IVC), passes through the right atrium (RA), extends into the superior vena cava (SVC) and is then pulled back approximately 100-130 mm. An acquisition produces 20 US volumes per second, creating a phantom dataset with 60 overlapping volumes and a porcine dataset with 17 volumes per

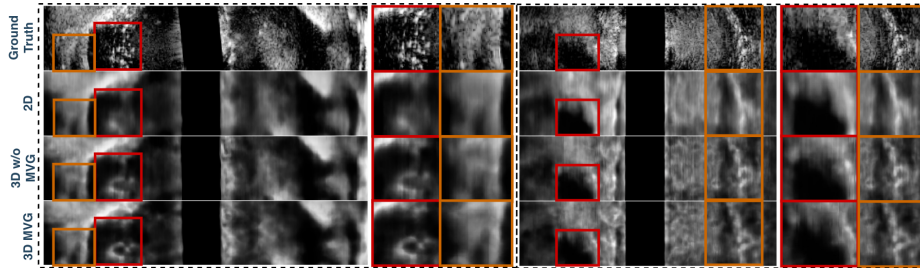


Fig. 4. Qualitative comparison of held-out ICE volume reconstruction across anatomical planes. Each column shows a distinct anatomical plane with close-up regions highlighting aliasing artifacts. Rows show, from top to bottom: ground truth, 2D Ultra-Wide-NeRF, 3D Ultra-Wide-NeRF w/o MVG, and **Ultra-Wide-NeRF with MVG**.

cardiac phase, both with $617 \times 617 \times 439$ voxels and isotropic spacing of 0.3mm . To assign a volume to a cardiac phase we use ECG gating.

Network & Training. The network is an 8-layer MLP (256 hidden units, ReLU, 10 IPE frequency bands) minimising the loss in Sec. 2.4. We cast 250 conical frustra per volumetric depth over the 360° convex definition, each integrated with 300 volumetric segments. The frustra start at the transducer and are estimated by MVGs whose principal axes encode the axial, lateral, and elevational resolutions of the ICE probe. The catheter FoV determines the lateral resolution, whereas the elevational resolution is scaled by the depth subsampling factor, directly embedding the transducer’s anisotropic characteristics into sampling. We trained on an NVIDIA RTX 6000 Ada (49 GB VRAM).

3.1 Panoramic Reconstruction

To validate the reconstruction fidelity and anti-aliasing capabilities of our approach, we benchmark Ultra-Wide-NeRF against the SOTA ICE panoramic compounding method [9]. In this evaluation, the SOTA baseline operates on full-resolution volumes (0.3 mm), whereas Ultra-Wide-NeRF is deliberately challenged to process volumes downsampled by a factor of 2 in the elevational direction (0.6 mm). Visualized cross-sections (Fig. 3) demonstrate that Ultra-Wide-NeRF recovers the identical anatomical context while producing sharper structural edges and eliminating the stitching artifacts inherent to the baseline. To isolate the contribution of our specific sampling strategy, we additionally compare Ultra-Wide-NeRF to an ablated model equipped with identical 3D spatial coherency and losses, but restricted to standard point sampling. This point-sampled ablation yields over-smoothed edges and washed-out structural details, explicitly validating that our continuous volumetric formulation is essential for preserving high-fidelity anatomical features, whilst being 1.5x faster to train for an accurate reconstruction, depending on the downsampling factor.

3.2 Novel View Synthesis

We evaluate Ultra-Wide-NeRF’s interpolation and novel-view synthesis capabilities across two scenarios: Hold-out Volume Synthesis and Novel View Synthesis. **Hold-out Volume Synthesis.** We evaluate Ultra-Wide-NeRF’s ability to synthesize unobserved spatial context. We train the panoramic representation on a subset of volumes and test its synthesis capabilities on a strictly held-out volume bordered by two adjacent, overlapping training sweeps. As shown in Fig. 4, the framework accurately reconstructs this unobserved volume directly from the learned panorama. The synthesized volume retains the anatomical details of the ground truth, confirming that our method delivers artifact-free novel view synthesis even when the acquisition steps along the physical acquisition trajectory are reduced.

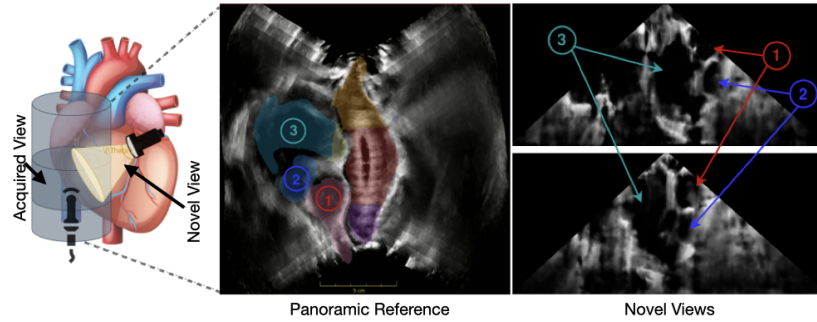


Fig. 5. Novel view B-mode synthesis using Ultra-Wide-NeRF. Centre: geometry-based panoramic reconstruction with overlaid segmentation masks. Right-top, bottom: synthesised novel views at 20° convex angle, rendered from unseen probe poses outside the training trajectory.

Novel Views Outside the Catheter Trajectory. Beyond interpolating within the acquisition path, Ultra-Wide-NeRF enables synthesis ultrasound volumes from virtual viewpoints strictly outside the original transducer trajectory. Fig. 5 shows that the model successfully synthesizes distinct off-trajectory B-mode images. By synthesising from entirely unobserved catheter positions Ultra-Wide-NeRF transforms the reconstructed volume into an interactive diagnostic tool. Clinicians can dynamically generate new perspectives, focus on specific anatomical regions of interest at higher resolutions, or utilize the virtual convex geometry to view the anatomy from optimal, unrecorded angles.

4 Conclusion

We proposed Ultra-Wide-NeRF, bridging continuous neural representations with the diverging geometry of convex ultrasound. By explicitly modeling the spa-

tial footprint of the acoustic beam using distance-dependent multivariate Gaussians, our framework acts as an inherent anti-aliasing filter to eliminate traditional stitching artifacts. Validations on phantom and porcine ICE datasets demonstrate superior artifact reduction and boundary enhancement compared to existing compounding methods. Delivering both artifact-free panoramic reconstructions and high-fidelity novel view synthesis, Ultra-Wide-NeRF expands critical spatial context from both acquired and virtual trajectories, establishing a robust foundation for seamless intraoperative navigation. While validated on intracardiac data, this geometric sampling strategy is inherently applicable to other clinical domains utilizing convex probes, including abdominal and obstetric imaging. By reliably expanding the observable field of view, this approach provides a practical tool to support more accurate spatial orientation during complex image-guided interventions.

Acknowledgments. This work was supported by the HINAV project funded by the Bavarian State Ministry for Economic Affairs, Regional Development and Energy.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5855–5864 (2021)
2. Burger, B., Bettinghausen, S., Radle, M., Hesser, J.: Real-time gpu-based ultrasound simulation using deformable mesh models. *IEEE transactions on medical imaging* **32**(3), 609–618 (2012)
3. Cobbold, R.S.C.: Foundations of biomedical ultrasound. Oxford University Press (2006)
4. Dagli, R., Hibi, A., Krishnan, R.G., Tyrrell, P.N.: Nerf us: Removing ultrasound imaging artifacts from neural radiance fields in the wild. In: Proceedings of Machine Learning Research. vol. 252, pp. 1–30 (2024)
5. Duellmer, F., Azampour, M.F., Wysocki, M., Navab, N.: Ultraray: Introducing full-path ray tracing in physics-based ultrasound simulation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 653–662. Springer (2025)
6. Fenster, A., Downey, D.B., Cardinal, H.N.: Three dimensional ultrasound imaging. *Physics in Medicine & Biology* **46**(5), R67 (2001)
7. Flach, B., Makhinya, M., Goksel, O.: Pure: panoramic ultrasound reconstruction by seamless stitching of volumes. In: International Workshop on Simulation and Synthesis in Medical Imaging. pp. 75–84. Springer (2016)
8. Guo, Z., Fang, Z., Fu, Z.: Ulre-nerf: 3d ultrasound imaging through neural rendering with ultrasound reflection direction parameterization. *arXiv preprint arXiv:2408.00860* (2024)

9. Hickler, J., Neary Zajiczek, L., Bandaru, R.S., Balle Sánchez, M., Hennemersperger, C., Wörz, S.: Panoramic anatomical context in 3d intracardiac echocardiography (ice) with 3d registration and geometry based image fusion. In: Statistical Atlases and Computational Models of the Heart (STACOM). pp. 87–97. Springer (2024)
10. Hsu, P.W., Prager, R.W., Gee, A.H., Treece, G.M.: Real time 4d ultrasound mosaicing and visualization. *Medical Image Analysis* **16**(3), 735–746 (2012)
11. Poon, T.C., Rohling, R.N.: Three dimensional ultrasound mosaicing. *IEEE Transactions on Medical Imaging* **25**(5), 588–598 (2006)
12. Prince, J.L., Links, J.M.: Medical imaging signals and systems. Pearson Prentice Hall (2006)
13. Salehi, M., Ahmadi, S.A., Prevost, R., Navab, N., Wein, W.: Patient-specific 3d ultrasound simulation based on convolutional ray-tracing and appearance optimization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 510–518. Springer (2015)
14. Wysocki, M., Azampour, M.F., Eilers, C., Busam, B., Salehi, M., Navab, N.: Ultra-nerf: Neural radiance fields for ultrasound imaging. In: Medical Imaging with Deep Learning (MIDL). vol. 227, pp. 382–401. PMLR (2024)
15. Yesilkaynak, V.B., Duque, V.G., Wysocki, M., Azampour, M.F., Navab, N., Mateus, D.: Ultramba neural bundle adjustment for pose refinement in 3d freehand ultrasound. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4363–4372 (2025)
16. Zhang, K., Gou, J., Sun, Y., Chen, L., Yu, B.: Physics-guided ultrasound neural radiance fields via conical acoustic sampling and integrated positional encoding. In: ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 6441–6445. IEEE (2026)