UltraG-Ray: Physics-Based Gaussian Ray Casting for Novel Ultrasound View Synthesis

Felix Duelmer[1,2,3,4]   iD   FELIX.DUELMER@TUM.DE
Jakob Klaushofer[1]   JAKOB.KLAUSHOFER@TUM.DE
Magdalena Wysocki[1,2]   iD   MAGDALENA.WYSOCKI@TUM.DE
Nassir Navab[1,2]   iD   NASSIR.NAVAB@TUM.DE
Mohammad Farid Azampour[1,2]   iD   MF.AZAMPOUR@TUM.DE

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

[2] *Munich Center for Machine Learning (MCML), Munich, Germany*

[3] *Institute of Biological and Medical Imaging, Bioengineering Center, Helmholtz Zentrum München, Neuherberg, Germany*

[4] *Chair of Biological Imaging, Central Institute for Translational Cancer Research (TranslaTUM), School of Medicine and Health & School of Computation, Information and Technology, Technical University of Munich, Munich, Germany*

Editors: Under Review for MIDL 2026

Abstract

Novel view synthesis (NVS) in ultrasound has gained attention as a technique for generating anatomically plausible views beyond the acquired frames, offering new capabilities for training clinicians or data augmentation. However, current methods struggle with complex tissue and view-dependent acoustic effects. Physics-based NVS aims to address these limitations by including the ultrasound image formation process into the simulation. Recent approaches combine a learnable implicit scene representation with an ultrasound-specific rendering module, yet a substantial gap between simulation and reality remains. In this work, we introduce UltraG-Ray, a novel ultrasound scene representation based on a learnable 3D Gaussian field, coupled to an efficient physics-based module for B-mode synthesis. We explicitly encode ultrasound-specific parameters, such as attenuation and reflection, into a Gaussian-based spatial representation and realize image synthesis within a novel ray casting scheme. In contrast to previous methods, this approach naturally captures view-dependent attenuation effects, thereby enabling the generation of physically informed B-mode images with increased realism. We compare our method to state-of-the-art and observe consistent gains in image quality metrics (up to 15% increase on MS-SSIM), demonstrating clear improvement in terms of realism of the synthesized ultrasound images.

Keywords: Ultrasound, Novel View Synthesis, Volumetric Representation, Gaussian Splatting, 3D Ultrasound Reconstruction

1. Introduction

Routine ultrasound examinations are based on 2D B-mode images (Adriaans et al., 2024), forcing clinicians to mentally reconstruct 3D anatomy from multiple slices, a process that depends heavily on spatial reasoning and experience (Kojcev et al., 2017; Krönke et al., 2022). This cognitively demanding task can be alleviated by 3D reconstruction methods, which provide spatially coherent representations that support more reliable anatomical interpretation. When probe poses or overlapping views are available, these methods combine

the acquired slices into a coherent representation of the scanned region and enable retrospective generation of views beyond the physically acquired plane, referred to as novel view synthesis (NVS). NVS has been gaining importance in clinical practice, as it can support simulators for training of clinicians (Blum et al., 2013; Ehricke, 1998), assist clinicians when reviewing data from robotic ultrasound systems (Jiang et al., 2023; Bi et al., 2024; Cao et al., 2025), and can potentially provide a training environment for emerging ultrasound world models (Yue et al., 2025; Qu et al., 2025).

A straightforward strategy for ultrasound NVS is to compound 2D frames into a 3D volume, either via freehand acquisition (Prevost et al., 2018; Wilson et al., 2025) or with robotic, optical, or electromagnetic tracking for higher accuracy (Jiang et al., 2023; Adriaans et al., 2024), and then re-slice the volume. However, ultrasound is highly anisotropic: intensity values corresponding to the same voxel but acquired from different orientations can disagree, producing inconsistent or blurred representations. Simple aggregation schemes such as mean or maximum compounding (Lasso et al., 2014) cannot resolve these orientation-dependent ambiguities and often fail to preserve the true underlying tissue structure.

View-dependent representations were introduced to model the strong dependence of ultrasound appearance on the probe position and orientation, which conventional view-independent reconstructions cannot capture. This motivation led to computational sonography (Hennersperger et al., 2015; Göbl et al., 2018) which models an ultrasound volume as a vector field in which each voxel stores an array of orientation-dependent intensity values, allowing view-specific appearance. However, this voxelized encoding of viewing directions imposes coarse angular resolution and interpolation artifacts, yielding blurred B-mode images. It also lacks physical consistency, reproducing only appearances seen during training.

Recent developments in NeRF-based methods have been proposed for ultrasound to mitigate these limitations (Wysocki et al., 2024; Dagli et al., 2024; Gaits et al., 2024; Grutman et al., 2025). These approaches use pose-annotated ultrasound images to learn an implicit continuous representation of ultrasound-specific parameters, coupled to a differentiable, ray-casting-based forward-synthesis module that models anisotropy in B-mode images. However, their sampling-based volumetric rendering inherently smooths out high-frequency structures, which still limits realism.

While the physical origin of image formation differs fundamentally between ultrasound and natural images, sampling-based volumetric integration in NeRF leads to analogous over-smoothing effects in both domains. In natural-image novel view synthesis, this limitation has been addressed by 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023). By replacing implicit volumetric fields with an explicit set of learnable 3D Gaussians and a GPU-efficient differentiable rasterizer, 3DGS enables significantly sharper reconstructions and real-time rendering performance. In the medical domain, 3DGS has already been applied to endoscopic scene representation (Liu et al., 2025), novel-view synthesis in X-ray (Cai et al., 2024), and intraoperative surgical navigation (Fehrentz et al., 2025).

In ultrasound, two methods using the Gaussian representation have been introduced in this context: UltraGS (Yang et al., 2025) and UltraGauss (Eid et al., 2025). UltraGS follows the original idea of projecting Gaussians onto a two-dimensional canvas. However, the approach assumes placing a virtual camera inside the tissue, directly in front of the acquired B-mode image. This setup would require an explicit selection mechanism to determine which Gaussians fall within the thin slab of tissue corresponding to the current ultrasound

slice, yet no such mechanism is specified in the original description (Yang et al., 2025). UltraGauss, on the other hand, follows a more promising direction by avoiding projection onto a distant plane and instead computing the intersection between each Gaussian and the image plane. Although this formulation is better aligned with the acquisition process, it does not include ultrasound-specific mechanisms for handling shadowing effects or resolving ambiguities caused by reflective structures (Eid et al., 2025).

In this paper, we introduce UltraG-Ray, a novel 3D Gaussian-based method that addresses these limitations by introducing a ray casting framework embedded in a Gaussian representation that follows the physics of ultrasound image formation (compare Fig. 1). Instead of relying on plane intersections or 2D projections, we follow the direction of acoustic wave propagation and incorporate attenuation directly into the 3D Gaussian field. This allows us to model depth-dependent signal decay, shadowing from strongly absorbing or reflecting structures, and the orientation-dependent visibility changes characteristic of real B-mode imaging. By capturing these ultrasound-specific mechanisms, our approach enables anatomically consistent, view-dependent novel view synthesis with enhanced visual realism. In summary, our contributions are as follows:

- A novel, high-fidelity NVS method for ultrasound that uses a learnable 3D Gaussian representation encoding ultrasound-specific parameters such as reflection and attenuation, combined with a physics-based, ray casting forward model for view-dependent real-time B-mode generation.
- An ultrasound-specific optimization strategy to prune, duplicate, and split Gaussians to enhance scene representation.
- Two open-source, pose-annotated datasets (in-silico and ex-vivo) containing overlapping views and various viewing angles.

2. Basics of Gaussian Splatting

3DGS (Kerbl et al., 2023) builds upon two established ideas in computer graphics: representing scenes using anisotropic 3D Gaussians with elliptical rasterization (Zwicker et al., 2002), and compositing their contributions using an emission–absorption model as in volumetric rendering (Mildenhall et al., 2021). Each Gaussian is defined by a mean $\boldsymbol{\mu}_i \in \mathbb{R}^3$, a covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{3\times3}$, an opacity α_i, and a color vector $\mathbf{c}_i \in \mathbb{R}^3$.

In practice, $\mathbf{c}_i$ is parameterized using low-order spherical harmonics (SH): for a viewing direction $\mathbf{v}$, the emitted color of Gaussian i is given by

$$\mathbf{c}_i(\mathbf{v}) = \sum_{b=1}^{B} \boldsymbol{\beta}_{i,b} Y_b(\mathbf{v}), \tag{1}$$

where Y_b are spherical harmonic basis functions and $\boldsymbol{\beta}_{i,b} \in \mathbb{R}^3$ are learned coefficients. The number of terms B is set by the SH degree. This enables a compact, view-dependent appearance per Gaussian. To render this representation from a given viewpoint, each Gaussian must be projected onto the image plane. Let $\Pi(\cdot)$ denote the camera projection and $\mathbf{J}_i$ the Jacobian of Π evaluated at $\boldsymbol{\mu}_i$. The projection yields a 2D Gaussian footprint with mean $\mathbf{u}_i = \Pi(\boldsymbol{\mu}_i)$ and covariance $\boldsymbol{\Sigma}_i^{\text{2D}} = \mathbf{J}_i \, \boldsymbol{\Sigma}_i \, \mathbf{J}_i^\top$.
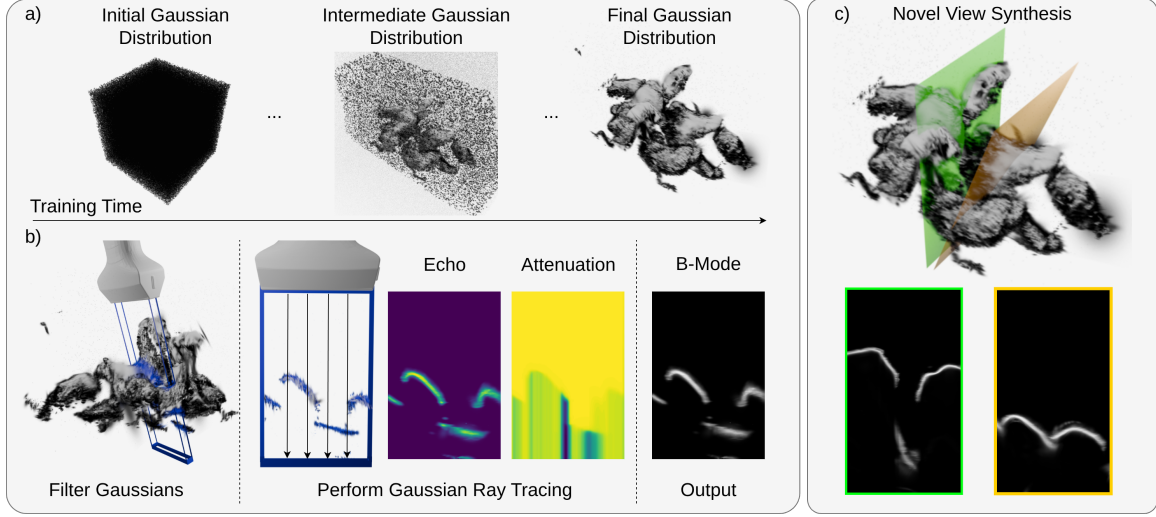
Figure 1: Overview of UltraG-Ray pipeline: a) Progressive adaptation of the learnable 3D Gaussian distribution to match the acquired data b) Image synthesis pipeline: First Gaussians are filtered based on the respective pose, second Gaussian ray-intersection is used to create the echo and transmittance maps and finally the resulting B-Mode image c) Downstream task evaluation based on novel view synthesis, where green and orange frames denote novel views

For a pixel location $\mathbf{p}$, the contribution of Gaussian i is determined by the Mahalanobis-weighted footprint

$$w_i(\mathbf{p}) = \exp\bigl(-\tfrac{1}{2}(\mathbf{p} - \mathbf{u}_i)^\top (\mathbf{\Sigma}_i^{\mathrm{2D}})^{-1}(\mathbf{p} - \mathbf{u}_i)\bigr), \qquad (2)$$

which smoothly decays with distance from the projected center. This footprint effectively modulates the effective opacity as $\alpha_i(\mathbf{p}) = \alpha_i\, w_i(\mathbf{p})$. Because multiple Gaussians can overlap along the viewing direction, their contributions must be composited in a physically meaningful order. All elements are therefore sorted by depth (front-to-back), and visibility is accumulated using transmittance. Starting with $T_0 = 1$, the transmittance after Gaussian i is given by

$$T_i(\mathbf{p}) = T_{i-1}(\mathbf{p})\,\bigl(1 - \alpha_i(\mathbf{p})\bigr), \qquad (3)$$

which accounts for the fraction of light that remains unoccluded after passing through the first i Gaussians. This recursion is equivalent to standard front-to-back alpha blending used in computer graphics. Finally, the pixel color follows the standard emission–absorption formulation

$$C(\mathbf{p}, \mathbf{v}) = \sum_{i=1}^{N} T_{i-1}(\mathbf{p})\,\alpha_i(\mathbf{p})\,\mathbf{c}_i(\mathbf{v}), \qquad (4)$$

where each Gaussian contributes according to its (view-dependent) color, its opacity at $\mathbf{p}$, and its visibility through previously encountered Gaussians. This formulation allows the method to efficiently approximate complex view-dependent appearance while remaining fully differentiable.

4

3. Methods

3.1. Ultrasound Physics

An ultrasound image is typically generated by a transducer positioned to face the tissue and emit acoustic waves into it. Due to the lenses and programmable focusing, the waves concentrate in a thin, but elongated, volume in front of the transducer. Tissue parameters, such as density or the speed of sound, determine how the wave propagates through the body. Due to inhomogeneities caused by variations in these tissues, sound is reflected back to the transducer. Next to this reflection, causing a weakening of the wave, the amplitude of the wave is being attenuated due to spherical decay and relaxation processes. Upon reception of the echo at the transducer surface, the signal is processed through a pipeline where it is beamformed, envelope detected, and log-compressed (Hoskins et al., 2019).

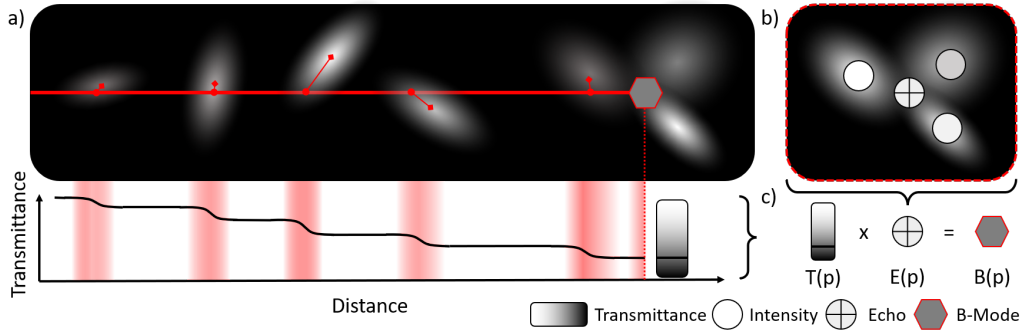### 3.2. Ultrasound-Adaptive Gaussian Forward Model

Figure 2: Depiction of the attenuation and intensity formation process: a) Transmittance accumulation, where each Gaussian gradually reduces the remaining energy along the ray. b) Computation of pixel-wise intensity contributions from view-dependent Gaussian intensity and their distance-weighted influence. c) Final pixel intensity obtained by combining accumulated attenuation with the weighted Gaussian contributions.

Acoustic attenuation differs fundamentally from light attenuation, making the standard RGB-based representation through color and opacity insufficient to capture the unique characteristics of ultrasound physics. To model ultrasound attenuation, we build on a ray casting scheme previously introduced by Salehi $et\ al.$ (Salehi et al., 2015), with important modifications enabled by the Gaussian formulation. In particular, we simplify the equations by leveraging the Gaussians' ability to approximate a point-spread function (PSF). Freed from a discretized sampling grid, speckle is represented directly with explicit Gaussians instead of as a distribution.

In UltraG-Ray, we define every 3D Gaussian by a mean $\boldsymbol{\mu}_i \in \mathbb{R}^3$, a covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{3\times 3}$, a transmittance value τ_i, and an intensity value I_i, which together parameterize the Gaussian used at rendering time. The parameters $\boldsymbol{\mu}_i$, τ_i, and I_i are directly optimized. To compute $\boldsymbol{\Sigma}_i$, we parameterize each Gaussian by a scale vector $\mathbf{s}_i$ and a quaternion $\mathbf{q}_i$,

from which the full covariance $\boldsymbol{\Sigma}_i$ is obtained analytically. We use this reparameterization because it ensures a stable and unconstrained optimization of anisotropic Gaussian shapes while guaranteeing that the resulting covariance matrices remain positive definite.

To ensure efficient computation, we cull Gaussians that cannot contribute to the final image. In this step, each Gaussian is projected onto the far plane, corresponding to the ultrasound imaging depth, to compute its 2D footprint. The projection follows an orthographic camera model defined by the transducer's pose and lateral width. For every potential ray endpoint, we define a small rectangular region on the far plane and test whether the projected Gaussian footprint intersects this rectangle. Only Gaussians whose projected support overlaps these rectangles are forwarded to the ray casting module. Note that the vertical extent is not set to the elevational height of the transducer but rather to a small value to approximate the 2D ultrasound imaging plane. Having restricted the set of Gaussians to those that can influence the image, we next characterize how each remaining Gaussian contributes to the intensity of a pixel (compare Fig. 2). We model the B-mode pixel value as:

$$B(p) = T(p) \cdot E(p), \tag{5}$$

where $T(p)$ is the transmission value defined for this specific pixel and $E(p)$ is the echo signal at this pixel, which can originate from both speckle and large-scale reflections. To model acoustic attenuation, we approximate wave interference with a ray casting model, treating each scan line as a ray emitted from the origin on the transducer surface $\mathbf{o}$ and direction $\mathbf{d}$, parametrized by $r(z) = \mathbf{o} + z\mathbf{d}$. To compute the contribution of a Gaussian G_i to local energy decay, we transform the ray into its canonical space following (Moenne-Loccoz et al., 2024; Yu et al., 2024):

$$\mathbf{o}_g = S_i^{-1} R_i^\top (\mathbf{o} - \boldsymbol{\mu}_i), \qquad \mathbf{d}_g = S_i^{-1} R_i^\top \mathbf{d}, \qquad \ell_g = \|\mathbf{d}_g\| \, \ell, \tag{6}$$

where R_i and S_i are the Gaussian's rotation and scale, and ℓ denotes the ray segment length in world space. In this canonical space, the covariance becomes the identity and the Gaussian is isotropic. We define the integral over the ray for a specific Gaussian as:

$$\psi_i = \int_0^\ell \exp\left(-\tfrac{1}{2} \|\mathbf{o}_g + t\,\mathbf{d}_g\|^2\right) dt, \tag{7}$$

which we approximate in practice using a 3-point Gauss–Legendre quadrature over an interval centered at the closest approach of the ray to the Gaussian. We map the ray–Gaussian integral ψ_i directly to a per-Gaussian transmittance and define the total transmission along the ray as:

$$T(z) = \prod_{i=1}^n T_i, \text{ where } T_i = \tau_i + (1 - \tau_i)\exp(-\psi_i). \tag{8}$$

Here, τ_i is the learnable transmittance of Gaussian G_i, $\exp(-\psi_i)$ represents the physically motivated attenuation induced by the Gaussian, and the product form provides a discrete approximation of Beer–Lambert exponential attenuation along the ray. This approximation is sufficient in practice because clinical ultrasound systems apply depth-dependent gain and log compression that largely compensate for global exponential decay.

For the echo signal $E(\mathbf{p})$, we compute the intensity contributed by each Gaussian based on its Mahalanobis-weighted distance to the pixel. Unlike splatting Gaussians in the 2D

image domain (see Eq. 2), we follow (Eid et al., 2025) and evaluate this contribution directly in 3D using the Mahalanobis distance: $w_i(\mathbf{p}) = \exp\left(-\frac{1}{2}(\mathbf{p} - \boldsymbol{\mu}_i)^\top (\boldsymbol{\Sigma}_i)^{-1} (\mathbf{p} - \boldsymbol{\mu}_i)\right)$.

To capture view-dependent backscattering, we parameterize the per-Gaussian echo amplitude as a low-order (degree-1) spherical harmonic (SH) expansion along ray direction $\mathbf{d}$ simplified to a single channel for intensity calculation (see Eq. 1). To ensure smoothness in regions where few or no Gaussians contribute to a pixel, we introduce a soft coverage mechanism that modulates the rendered intensity based on the strength of the pixel's support from the Gaussian field. Let $S(\mathbf{p}) = \sum_{i=1}^n w_i(\mathbf{p})$ denote the total Gaussian footprint at pixel $\mathbf{p}$. We define a soft coverage factor: $\gamma(\mathbf{p}) = 1 - \exp\left(-S(\mathbf{p})\right)$, which remains close to zero when the pixel is only weakly covered by Gaussians and approaches one as the local footprint mass increases. The final echo intensity is then given by

$$E(\mathbf{p}) = (1 - \gamma(\mathbf{p}))\, I_{\mathrm{bkg}} + \gamma(\mathbf{p})\, \frac{\sum_{i=1}^n I_i(\mathbf{d})\, w_i(\mathbf{p})}{S(\mathbf{p}) + \varepsilon}, \tag{9}$$

with a small $\varepsilon > 0$ for numerical stability and a constant background intensity $I_{\mathrm{bkg}} \in \mathbb{R}$ representing echo-free regions in ultrasound, which we set to black for the whole image.

3.3. Initialization and Optimization Strategy

Conventional 3DGS relies on structure-from-motion (Schonberger and Frahm, 2016) to initialize Gaussians and camera poses, which is incompatible with our data due to the fundamentally different image formation in ultrasound. We therefore initialize all Gaussians randomly within the local scene and set the remaining parameters to defaults (see App. A).

Like in 3DGS, we optimize all Gaussian parameters using stochastic gradient descent. To ensure agreement between the rendered slices and the ground truth, we employ a reconstruction objective composed of an L1 loss and a SSIM loss. Additionally, we regularize the Gaussian scales to encourage non-contributing or excessively large Gaussians to shrink. The complete training loss is therefore defined as:

$$\mathcal{L} = (1 - \lambda_{\mathrm{ssim}})\, \mathcal{L}_1 + \lambda_{\mathrm{ssim}}\, \mathcal{L}_{\mathrm{SSIM}} + \lambda_{\mathrm{scale}}\, \mathbb{E}[\exp(\mathbf{S})]. \tag{10}$$

Optimizing with this loss alone is insufficient because a refinement mechanism is required to add Gaussians where detail is missing and remove those that do not contribute. Since existing schemes are incompatible with our custom synthesis pipeline, we use a 3DGS-inspired refinement strategy (Kerbl et al., 2023) adapted to ultrasound data. After a brief warm-up, refinement is triggered every t iterations and applies pruning, duplication, and splitting to maintain a compact and expressive Gaussian set. Gaussians with negligible or excessively large scales, which either fail to contribute or oversmooth local structure, are pruned.

By tracking the gradients of each Gaussian over t iterations, we compute a mean importance score that determines whether a Gaussian should be modified during refinement. Gaussians whose importance exceeds a predefined threshold are either duplicated or split, depending on their scale. Small Gaussians are duplicated to increase local representational capacity, while larger Gaussians are split into two nearby Gaussians to refine structure. We cap the total number of Gaussians to balance computational cost and reconstruction

fidelity. Because real ultrasound acquires a thin elevational volume rather than infinitesimally thin rays (see Sec. 3.1), we approximate this effect by perturbing all ray origins with a small out-of-plane offset drawn from a cosine-weighted distribution. This perturbation also discourages the Gaussians from collapsing into thin 2D structures tied to individual frames.

3.4. Implementation

We implement UltraG-Ray[1] in PyTorch on top of the CUDA-accelerated 3D Gaussian splatting library gsplat (Ye et al., 2025). Our implementation adds custom CUDA kernels for selecting relevant Gaussians per pixel and for efficiently evaluating the ray–Gaussian equations described in Sec. 3.2. All experiments are run on a single desktop workstation with an Intel Core i7-12700 CPU and an NVIDIA RTX 4070 Ti GPU. For training, we use the Adam optimizer with parameter-specific learning rates. We set the loss weights to $\lambda_{\mathrm{ssim}} = 0.5$ and $\lambda_{\mathrm{scale}} = 10^{-3}$ and train for 30k epochs, which we found sufficient for convergence in all experiments. To approximate the elevational beam width, we perturb all ray origins with an out-of-plane offset sampled from a cosine-weighted distribution with maximum magnitude $\delta = 2\,\mathrm{mm}$. Additional information about parameters used in training and optimization strategy can be found in Appendix A.

4. Experiments and Results

Datasets To validate UltraG-Ray, we introduce two pose-annotated B-mode datasets. To demonstrate robustness across both strong reflectors and heterogeneous scattering, we evaluate on an in-silico spine phantom and an ex-vivo porcine muscle phantom. Data were acquired with a Siemens 12L3 probe connected to a Siemens Acuson Juniper system mounted on a KUKA LBR iiwa 14 R820 robot using a custom probe holder. Image and tracking data were recorded via the software ImFusion (ImFusion GmbH, Munich, Germany). We first applied a coarse calibration based on the probe geometry and the known offsets from the robot's last joint. This was then refined using a calibration procedure based on the BOBYQA optimization algorithm (Powell et al., 2009), implemented in the software ImFusion. For a detailed hardware setup overview, we refer the reader to (Jiang et al., 2023). Both datasets comprise multiple overlapping sweeps (six for the spine phantom dataset and three for the porcine muscle dataset), acquired at various tilting angles.

Novel View Synthesis We compare our method to a scene reconstruction using Ultra-NeRF (Wysocki et al., 2024). As both UltraGS (Yang et al., 2025) and UltraGauss (Eid et al., 2025) did not make their source code publicly available[2], we can not directly compare to their implementations. In particular, reproducing the original methods ourselves would risk not doing justice to the respective approaches. For quantitative evaluation, we report multi-scale structural similarity (MS-SSIM) and peak signal-to-noise ratio (PSNR). As these metrics can favor overly smooth reconstructions, we additionally include gradient-magnitude similarity (GMS) and gradient-magnitude similarity deviation (GMSD) (Xue et al., 2013), which better capture high-frequency detail and edge preservation.

---

1. Both datasets and the implementation will be released upon acceptance.
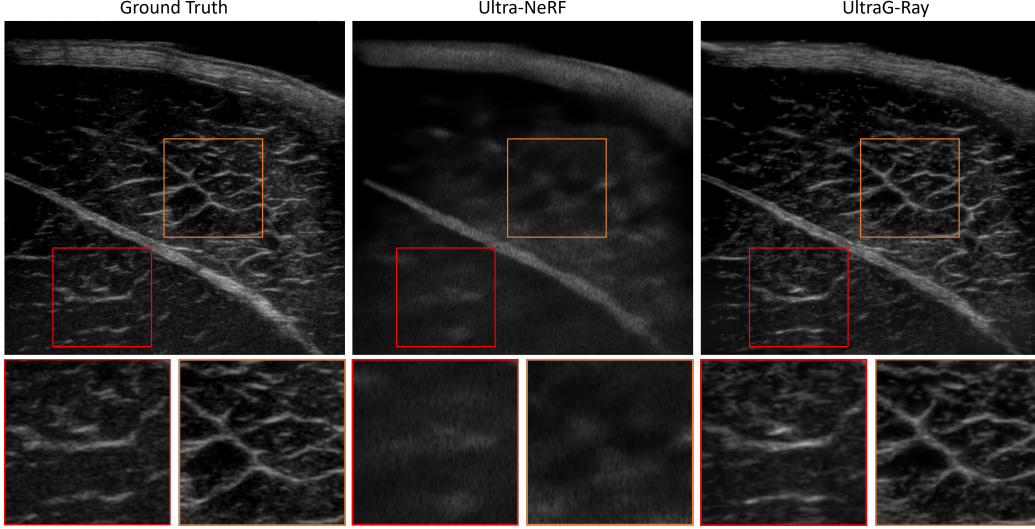2. At the time of submission of this paper.

Figure 3: NVS comparison on ex vivo porcine muscle using Ultra-NeRF and UltraG-Ray. Highlighted regions (red and orange) are enlarged as cut-outs in the second row.

Table 1: Evaluation results for NVS on both datasets. Best values are in bold. Arrows indicate whether higher (↑) or lower (↓) values correspond to better performance.

Dataset	Method	PSNR ↑	MS-SSIM ↑	GMS ↑	GMSD ↓
Porcine Muscle	Ultra-NeRF	$23.60_{\pm0.54}$	$0.62_{\pm0.03}$	$0.81_{\pm0.02}$	$0.23_{\pm0.01}$
	UltraG-Ray (w/o Trans.)	$24.31_{\pm0.76}$	$0.74_{\pm0.03}$	$0.86_{\pm0.01}$	$0.18_{\pm0.00}$
	UltraG-Ray	$\mathbf{25.49}_{\pm0.60}$	$\mathbf{0.77}_{\pm0.02}$	$\mathbf{0.88}_{\pm0.01}$	$\mathbf{0.16}_{\pm0.01}$
Spine Phantom	Ultra-NeRF	$24.84_{\pm2.48}$	$0.92_{\pm0.04}$	$0.96_{\pm0.02}$	$0.13_{\pm0.04}$
	UltraG-Ray (w/o Trans.)	$24.60_{\pm3.65}$	$0.91_{\pm0.05}$	$0.96_{\pm0.02}$	$0.14_{\pm0.04}$
	UltraG-Ray	$\mathbf{26.92}_{\pm2.35}$	$\mathbf{0.94}_{\pm0.03}$	$\mathbf{0.97}_{\pm0.01}$	$\mathbf{0.10}_{\pm0.03}$

As shown in Fig. 3, UltraG-Ray produces high-fidelity reconstructions in which individual muscle fibers are clearly distinguishable. Whereas Ultra-NeRF largely recovers only the dominant fiber orientation, UltraG-Ray additionally resolves fine structural variations and renders them with markedly higher contrast. This qualitative improvement is consistent with the quantitative results in Table 1, where the image-quality metrics reflect the enhanced structural detail. A similar trend appears in the spine phantom, although differences are less pronounced due to the simplicity of the dataset and the consequently high scores achieved by both methods. To assess the contribution of transmittance modeling, we conduct an ablation in which transmittance is removed from the rendering process. Both datasets exhibit a measurable drop in reconstruction quality, confirming the importance of separating transmittance from echo intensity. Representative echo and transmittance maps are shown in Fig. 4. Although we cap the scene at 500k Gaussians, the optimization does
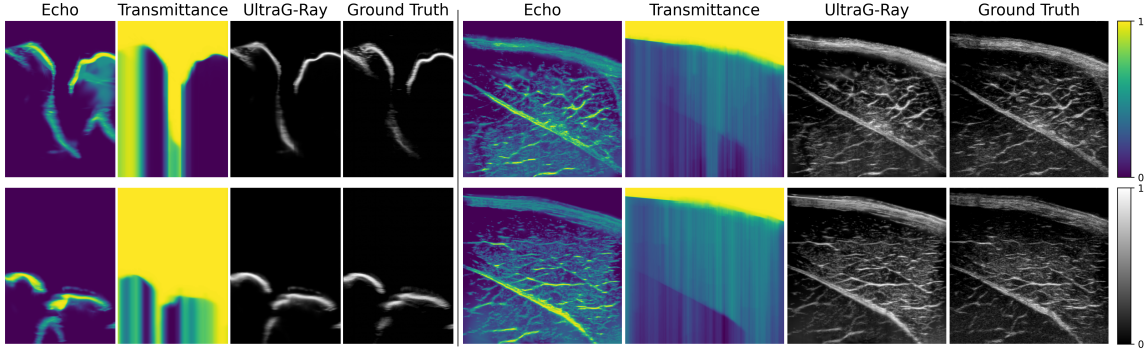
Figure 4: NVS examples from both datasets for the intermediate echo and transmittance maps, the synthesized B-mode output, and the corresponding ground truth image.

not always reach this limit: the porcine muscle dataset increases to the full 500k, whereas the spine phantom converges with only about 90k Gaussians. Further analysis of how the Gaussian count influences image quality is provided in App. B. At inference, this yields approximately 680 fps for the spine phantom and about 95 fps for the porcine muscle.

5. Discussion & Conclusion

In this paper, we present UltraG-Ray, a novel approach to representing and synthesizing ultrasound images that combines the flexibility of 3D Gaussian fields with the fidelity of a physically grounded ray casting model. Unlike previous Gaussian-splatting approaches for ultrasound, which rely on intensity–opacity blending, our formulation directly incorporates ultrasound phenomena such as attenuation, view-dependent backscattering, and ray-aligned energy decay. This separation of physics allows UltraG-Ray to capture shadowing behaviour and reflection amplitude, which is essential for realistic NVS. Our experiments on both in-silico and ex-vivo phantoms demonstrate that UltraG-Ray can synthesize realistic B-mode images with higher fidelity than state-of-the-art methods. The improvements in image quality metrics indicate that our ultrasound formation model helps resolve view-dependent ambiguities.

Despite its benefits, the method has limitations. First, it relies on a single straight-ray approximation that cannot model complex wave phenomena such as interference or multi-path propagation. Extending the model with secondary ray paths, as used in natural-image rendering, could improve physical fidelity. Second, the Gaussian field is an explicit representation. Although interpretable and easy to manipulate, it lacks inherent spatial continuity, which can leave sparsely sampled regions poorly represented. Incorporating anatomical or statistical priors may allow further generalization to unseen tissue and improve continuity across viewpoints. Overall, our results show that Gaussian-based ultrasound representations can combine physical interpretability, computational efficiency, and open new directions for realistic NVS in ultrasound.

Acknowledgments

References

Chrissy A Adriaans, Mark Wijkhuizen, Lennard M van Karnenbeek, Freija Geldof, and Behdad Dashtbozorg. Trackerless 3d freehand ultrasound reconstruction: a review. *Applied Sciences*, 14(17):7991, 2024.

Yuan Bi, Zhongliang Jiang, Felix Duelmer, Dianye Huang, and Nassir Navab. Machine learning in robotic ultrasound imaging: Challenges and perspectives. *Annual Review of Control, Robotics, and Autonomous Systems*, 7, 2024.

Tobias Blum, Andreas Rieger, Nassir Navab, Helmut Friess, and Marc Martignoni. A review of computer-based simulators for ultrasound training. *Simulation in Healthcare*, 8(2):98–108, 2013.

Yuanhao Cai, Yixun Liang, Jiahao Wang, Angtian Wang, Yulun Zhang, Xiaokang Yang, Zongwei Zhou, and Alan Yuille. Radiative gaussian splatting for efficient x-ray novel view synthesis. In *European Conference on Computer Vision*, pages 283–299. Springer, 2024.

Guanglin Cao, Danqian Cao, and Hongbin Liu. Respiratory motion-robust robotic ultrasound acquisitions via vision-haptic fusion control and 3d compensation. *IEEE Transactions on Instrumentation and Measurement*, 2025.

Rishit Dagli, Atsuhiro Hibi, Rahul G Krishnan, and Pascal N Tyrrell. Nerf-us: Removing ultrasound imaging artifacts from neural radiance fields in the wild. *arXiv preprint arXiv:2408.10258*, 2024.

Hans-Heino Ehricke. Sonosim3d: a multimedia system for sonography simulation and education with an extensible case database. *European journal of ultrasound*, 7(3):225–230, 1998.

Mark C Eid, Ana IL Namburete, and João F Henriques. Ultragauss: Ultrafast gaussian reconstruction of 3d ultrasound volumes. *arXiv preprint arXiv:2505.05643*, 2025.

Maximilian Fehrentz, Alexander Winkler, Thomas Heiliger, Nazim Haouchine, Christian Heiliger, and Nassir Navab. Bridgesplat: Bidirectionally coupled ct and non-rigid gaussian splatting for deformable intraoperative surgical navigation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 44–53. Springer, 2025.

François Gaits, Nicolas Mellado, and Adrian Basarab. Ultrasound volume reconstruction from 2d freehand acquisitions using neural implicit representations. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2024.

Rüdiger Göbl, Diana Mateus, Christoph Hennersperger, Maximilian Baust, and Nassir Navab. Redefining ultrasound compounding: Computational sonography. *arXiv preprint arXiv:1811.01534*, 2018.

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.

Christoph Hennersperger, Maximilian Baust, Diana Mateus, and Nassir Navab. Computational sonography. In *International conference on medical image computing and computer-assisted intervention*, pages 459–466. Springer, 2015.

Peter Hoskins, Kevin Martin, and Abigail Thrush. *Diagnostic ultrasound: physics and equipment*. CRC Press, 2019.

Zhongliang Jiang, Felix Duelmer, and Nassir Navab. Dopus-net: Quality-aware robotic ultrasound imaging based on doppler signal. *IEEE Transactions on Automation Science and Engineering*, 21(3):3229–3242, 2023.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

Risto Kojcev, Ashkan Khakzar, Bernhard Fuerst, Oliver Zettinig, Carole Fahkry, Robert DeJong, Jeremy Richmon, Russell Taylor, Edoardo Sinibaldi, and Nassir Navab. On the reproducibility of expert-operated and robotic ultrasound acquisitions. *International journal of computer assisted radiology and surgery*, 12(6):1003–1011, 2017.

Markus Krönke, Christine Eilers, Desislava Dimova, Melanie Köhler, Gabriel Buschner, Lilit Schweiger, Lemonia Konstantinidou, Marcus Makowski, James Nagarajah, Nassir Navab, et al. Tracked 3d ultrasound and deep neural network-based thyroid segmentation reduce interobserver variability in thyroid volumetry. *Plos one*, 17(7):e0268550, 2022.

Andras Lasso, Tamas Heffter, Adam Rankin, Csaba Pinter, Tamas Ungi, and Gabor Fichtinger. Plus: open-source toolkit for ultrasound-guided intervention systems. *IEEE transactions on biomedical engineering*, 61(10):2527–2537, 2014.

Yifan Liu, Chenxin Li, Hengyu Liu, Chen Yang, and Yixuan Yuan. Foundation model-guided gaussian splatting for 4d reconstruction of deformable tissues. *IEEE Transactions on Medical Imaging*, 2025.

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.

Nicolas Moenne-Loccoz, Ashkan Mirzaei, Or Perel, Riccardo de Lutio, Janick Martinez Esturo, Gavriel State, Sanja Fidler, Nicholas Sharp, and Zan Gojcic. 3d gaussian ray tracing: Fast tracing of particle scenes. *ACM Transactions on Graphics (TOG)*, 43(6):1–19, 2024.

Michael JD Powell et al. The bobyqa algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, 26: 26–46, 2009.

Raphael Prevost, Mehrdad Salehi, Simon Jagoda, Navneet Kumar, Julian Sprung, Alexander Ladikos, Robert Bauer, Oliver Zettinig, and Wolfgang Wein. 3d freehand ultrasound without external tracking using deep learning. *Medical image analysis*, 48:187–202, 2018.

Jingguo Qu, Xinyang Han, Tonghuan Xiao, Jia Ai, Juan Wu, Tong Zhao, Jing Qin, Ann Dorothy King, Winnie Chiu-Wing Chu, Jing Cai, et al. Adapting vision-language foundation model for next generation medical ultrasound image analysis. *arXiv preprint arXiv:2506.08849*, 2025.

Mehrdad Salehi, Seyed-Ahmad Ahmadi, Raphael Prevost, Nassir Navab, and Wolfgang Wein. Patient-specific 3d ultrasound simulation based on convolutional ray-tracing and appearance optimization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 510–518. Springer, 2015.

Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016.

Paul FR Wilson, Matteo Ronchetti, Rüdiger Göbl, Viktoria Markova, Sebastian Rosenzweig, Raphael Prevost, Parvin Mousavi, and Oliver Zettinig. Dualtrack: Sensorless 3d ultrasound needs local and global context. In *International Workshop on Advances in Simplifying Medical Ultrasound*, pages 3–12. Springer, 2025.

Magdalena Wysocki, Mohammad Farid Azampour, Christine Eilers, Benjamin Busam, Mehrdad Salehi, and Nassir Navab. Ultra-nerf: Neural radiance fields for ultrasound imaging. In *Medical Imaging with Deep Learning*, pages 382–401. PMLR, 2024.

Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE transactions on image processing*, 23(2):684–695, 2013.

Yuezhe Yang, Wenjie Cai, Dexin Yang, Yufang Dong, Xingbo Dong, and Zhe Jin. Ultrags: Gaussian splatting for ultrasound novel view synthesis. *arXiv preprint arXiv:2511.07743*, 2025.

Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, and Angjoo Kanazawa. gsplat: An open-source library for gaussian splatting. *Journal of Machine Learning Research*, 26 (34):1–17, 2025.

Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Transactions on Graphics (ToG)*, 43 (6):1–13, 2024.

Yang Yue, Yulin Wang, Haojun Jiang, Pan Liu, Shiji Song, and Gao Huang. Echoworld: Learning motion-aware world models for echocardiography probe guidance. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 25993–26003, 2025.

Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa splatting. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):223–238, 2002.

Appendix A. Training and implementation details

Optimization. We train UltraG-Ray using the Adam optimizer with parameter-specific learning rates. The learning rate for 3D Gaussian means is set to 1×10^{-4}, for Gaussian scales and quaternions to 5×10^{-3}, and for transmittances to 5×10^{-4}. For the spherical harmonic (SH) coefficients, we use 5×10^{-3} for the zero-order band and 1×10^{-5} for the first-order band. All Gaussians are initialized with a transmittance of 0.99 and an isotropic scale of 0.5 mm. The SH expansion starts at degree 0 and is increased to degree 1 after 1000 iterations, which progressively introduces higher-frequency detail in a stable manner. A learning-rate scheduler decays all rates smoothly to 10% of their initial value by the end of training. We use a batch size of 8. We found that this parameterization yields the most stable optimization and the highest reconstruction fidelity across datasets.
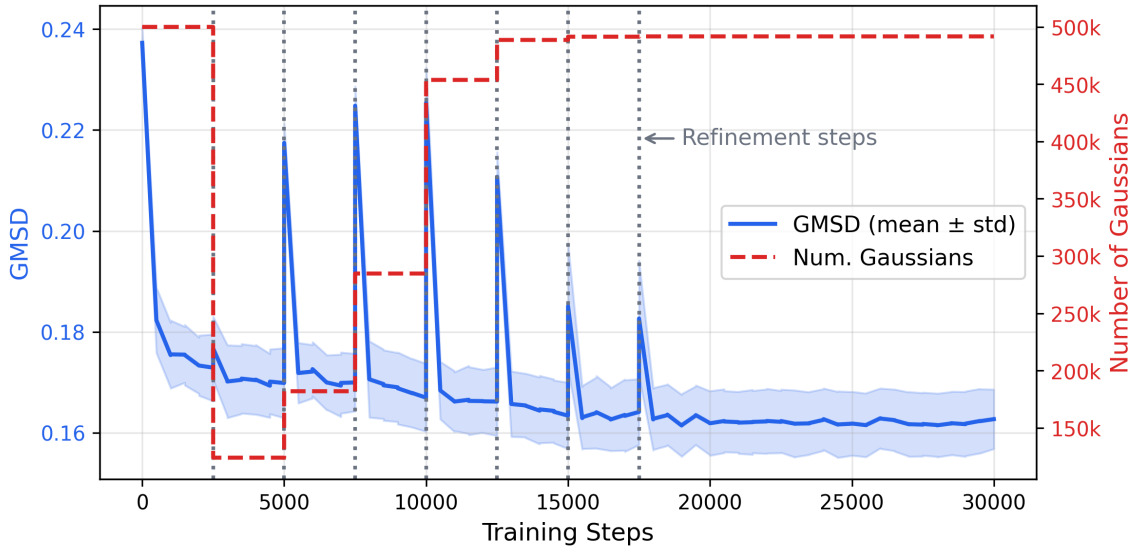

Figure 5: Evolution of GMSD scores of novel views for the porcine muscle dataset and number of Gaussians over training steps.

Refinement Strategy. For the ultrasound optimization strategy, we employ an importance-driven Gaussian refinement strategy. Refinement is triggered every $r = 2500$ iterations, with importance scores accumulated over the same interval. Refinement begins after 1k epochs and terminates at 20k epochs, while training continues until 30k epochs. Gaussians whose

scales fall below $s_{\min} = 5 \times 10^{-5}$ or exceed $s_{\max} = 5\,\mathrm{mm}$ are pruned, and the total number of Gaussians is capped at $N_{\max} = 500\mathrm{k}$ to balance reconstruction fidelity and computational efficiency.

Optimization Dynamics   In order to demonstrate our optimization strategy, we evaluate our datasets every 500 iterations and capture statistics of the Gaussian parameters. Fig. 5 clearly illustrates the refinement steps every 2500 iterations. Due to random initialization, a considerable number of Gaussians do not contribute meaningfully to reconstruction. Scale regularization continuously shrinks those Gaussians, and most are subsequently pruned during the first refinement cycle due to their small scales. This can be observed as a sharp reduction in Gaussian count after the first refinement step. In subsequent cycles, the Gaussian count increases again due to Gaussians being split/duplicated in order to model increasingly fine detail. We also observe a temporary degradation in metrics after applying refinement, since duplication/splitting adds Gaussians that require optimization before contributing in a positive manner. For this reason, we disable refinement for the last 10k Gaussians.

Appendix B. Impact of Gaussian Count on Image Quality

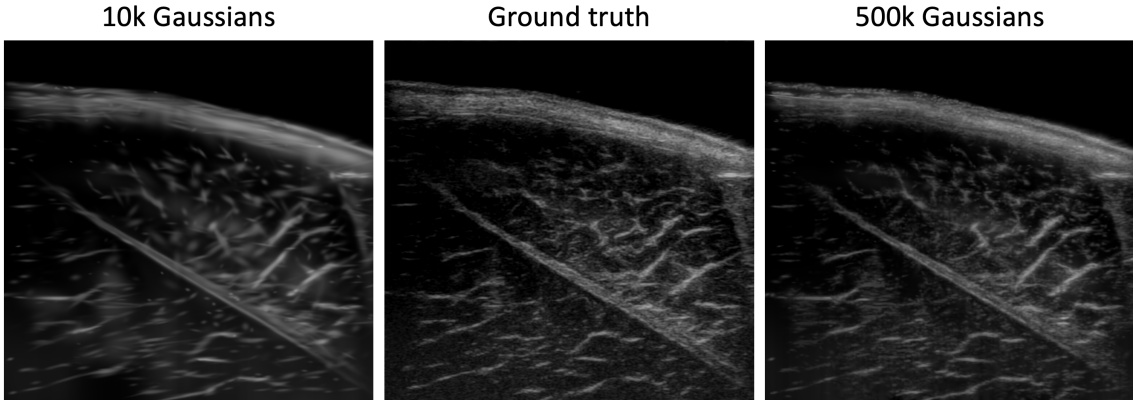10k Gaussians　　　　　　Ground truth　　　　　　500k Gaussians

Figure 6: Comparison of novel views with differing number of max. Gaussians in the porcine muscle dataset. The metrics are comparable with Table 2.

We evaluate how the number of Gaussians affects image quality by performing an ablation on the porcine dataset. As shown in Table 2, reducing the number of Gaussians results in only small changes in the image quality metrics. We only start seeing a noticeable reduction in these metrics below 10k Gaussians, demonstrating our method's ability to adapt to varying levels of geometric detail. This allows us to sacrifice fine detail preservation in favor of faster training and inference speeds.

These results highlight the flaws of standard evaluation metrics when applied to ultrasound imaging. Fig. 3 reveals a significant disparity in detail. The variant with 10k Gaussians manages to capture the overall anatomical structure but fails to model speckle

Table 2: Comparison of image quality metrics for different maximum numbers of Gaussians.

Dataset	#Gaussians	PSNR ↑	MS-SSIM ↑	GMS ↑	GMSD ↓
Porcine Muscle	500k	$25.49_{\pm0.60}$	$0.77_{\pm0.02}$	$0.88_{\pm0.01}$	$0.16_{\pm0.01}$
	300k	$25.39_{\pm0.62}$	$0.77_{\pm0.02}$	$0.87_{\pm0.01}$	$0.17_{\pm0.01}$
	100k	$25.46_{\pm0.66}$	$0.77_{\pm0.02}$	$0.87_{\pm0.01}$	$0.17_{\pm0.01}$
	10k	$25.39_{\pm0.54}$	$0.75_{\pm0.01}$	$0.85_{\pm0.01}$	$0.18_{\pm0.00}$

and other fine textures. Even so, the variant with 10k Gaussians manages to achieve quality metrics similar to higher Gaussian counts. Here, GMS and GMSD manage to detect the degradation, even if it is larger than the score difference might suggest.