MammoNeRF: Neural Radiance Fields for Mammography Imaging

Yassine Ameskine[*1,2] YASSINE.AMESKINE@UM6P.MA
Youssef Alj [*1] YOUSSEF.ALJ@UM6P.MA

Editors: Under Review for MIDL 2026

Abstract

Reconstructing a 3D volumetric representation of breast tissue from 2D mammographic views remains a challenging problem. This is primarily due to extremely sparse imaging, the absence of geometric reasoning, and the lack of volumetric imaging in routine screening. To address these limitations, we propose MammoNeRF, a novel neural radiance field-based framework for 3D volumetric reconstruction of breast tissue from mammogram images. MammoNeRF is trained using craniocaudal (CC) and mediolateral oblique (MLO) views annotated with the corresponding binary lesion masks. To enforce cross-view alignment, we introduce a lesion consistency loss, encouraging the model to estimate consistent poses across both views. This enables the reconstruction of pseudo ground-truth 3D lesion bounding boxes (3DPB) and 3D volumetric breast representations. Without this constraint, traditional NeRF models tend to reconstruct the same lesion at inconsistent spatial locations across views, resulting in incoherent lesion geometry. MammoNeRF jointly optimizes NeRF parameters, camera poses, and lesion centroids within a unified framework. We compare MammoNeRF with existing reconstruction models using the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) metrics, demonstrating that MammoNeRF achieves superior quantitative performance while also improving anatomical fidelity. By transforming standard mammograms into a meaningful 3D representation, MammoNeRF provides a new pathway for enhanced structural understanding in breast imaging and establishes a robust foundation for future work in visualization and downstream clinical applications.

Keywords: Mammography, NeRF, 3D Reconstruction, Image Synthesis

1. Introduction

Mammography (MG) remains the cornerstone of breast cancer screening and diagnosis. MG typically provides two views per breast: craniocaudal (CC), taken above the breast, and mediolateral oblique (MLO), captured at an angle between 30° and 45° (Gupta and Markey, 2005). While these paired images offer essential diagnostic information, they also impose significant cognitive demands on radiologists, who must mentally reconstruct a three-dimensional (3D) understanding of breast anatomy and lesions from only two-dimensional (2D) inputs. This mental reconstruction is time-consuming and prone to interpretive variability. Increasing the number of views in mammography can alleviate some of this ambiguity. However, additional exposure to ionizing radiation, which carries health risks and is not recommended in routine screening. Consequently, 3D representation techniques that recover anatomical structure from the standard dual-view protocol hold promise for reducing diagnostic uncertainty while maintaining acceptable radiation levels.

[*] Contributed equally

Recent progress in neural scene representations, such as Neural Radiance Fields (NeRF) (Mildenhall et al., 2020) and Gaussian Splatting (Kerbl et al., 2023), has opened exciting opportunities for explicitly modeling 3D structure from 2D projections, including medical imaging (Wysocki et al., 2024; Corona-Figueroa et al., 2022; Cai et al., 2024; Yu et al., 2025). However, applying these methods to mammography presents four major challenges: (1) the need for accurately posed images to support geometric reasoning is especially important since pose errors can significantly degrade NeRF's ability to recover coherent 3D structure and lead to lower reconstruction quality, (2) mammography provides extremely sparse imaging, with only two views per breast in standard practice, (3) inconsistencies between views due to non-rigid breast compression, and (4) discrepancies between X-ray imaging physics and the pinhole camera model typically assumed in neural scene representations. In this work, we propose strategies to address these challenges for 3D-aware lesion understanding directly from standard mammographic projections. For (1), recent pose-free NeRF variants (Wang et al., 2021; Bian et al., 2023; Shi et al., 2025) motivate the joint estimation of pose and scene. In the context of mammography, traditional feature matching pipelines such as COLMAP (Schonberger and Frahm, 2016) fail because mammograms lack reliable natural image keypoints, making automatic pose estimation particularly challenging. For (2), few-shot NeRF approaches such as PixelNeRF (Yu et al., 2021) demonstrates that accurate volumetric representations can be learned from very limited views. To address (3), we introduce a lesion consistency loss during NeRF training to mitigate view inconsistencies from non-rigid deformation. This loss enforces spatial alignment of the 3D lesion points reconstructed from the CC and MLO views by a pose-consistent distance penalty that encourages the lesion centroid inferred from each view to converge to a single 3D location. Finally, for (4), we adopt the widely used pinhole camera approximation as a geometric model for X-ray projection, assuming that this representation sufficiently captures and preserves the relevant structural information present in mammograms. Our contributions are as follows,

- A novel framework for 3D breast reconstruction. MammoNeRF learns a NeRF-based 3D volumetric representation of breast tissue and lesion using only CC and MLO mammograms along with their corresponding masks.

- Within MammoNeRF, we incorporate a lesion consistency loss designed to enforce spatial coherence between the 3D lesion points reconstructed from the CC and MLO views. This loss encourages the two view-specific reconstructions to converge toward a unified representation.

- MammoNeRF produces a single pseudo–ground truth 3D lesion bounding box (3DPB) and a reconstructed 3D breast volume, while jointly optimizing camera parameters to improve geometric alignment across views.

In our experiments, we use CBIS-DDSM (Lee et al., 2017) and INBreast (Moreira et al., 2012) datasets, and evaluate our method quantitatively and qualitatively. To the best of our knowledge, this paper presents a new implicit neural representation for mammography that, for the first time, produces pseudo–ground truth 3D lesion bounding box (3DPB) and 3D volume representation of the breast tissue using CC and MLO mammograms, which can improve medical downstream tasks, including mass detection and segmentation. In general,

we demonstrate that enriching mammography with NeRF-based 3D volumes provides a promising direction for advancing computer-aided diagnosis.

2. Related Work

2.1. Neural Representations in Medical Imaging

Implicit Neural Representations (INRs), particularly Neural Radiance Fields (NeRFs) (Mildenhall et al., 2020), have emerged as a powerful paradigm for modeling continuous 3D structure using multilayer perceptrons. Unlike voxel or mesh based representations, INRs encode geometry and appearance implicitly, enabling volumetric reconstruction from sparse multi-view inputs without relying on explicit 3D supervision. These properties have driven rapid adoption of INR-based frameworks in medical imaging, where lack of data, radiation exposure, and annotation constraints are common challenges. Early research demonstrated the feasibility of adapting NeRFs to clinical modalities such as computed tomography (CT) (Cai et al., 2024; Zha et al., 2022) and ultrasounds (Wysocki et al., 2024). MedNeRF (Corona-Figueroa et al., 2022) demonstrate that NeRFs can synthesize volumetric CT scans from a single or few radiographs by incorporating domain-specific priors on X-ray attenuation supported by adversarial training to enhance realism and anatomical consistency. Beyond image reconstruction, INRs have also been applied to medical image registration, shape modeling, and segmentation. Works such as (Sitzmann et al., 2020; Wolterink et al., 2022) demonstrated that implicit fields can learn smooth deformation maps and anatomical priors that outperform discrete grid-based formulations. Despite these advances, no prior study has applied NeRFs to mammography-based 3D reconstruction. Screening mammography presents unique challenges, including extreme view-point sparsity (typically only CC and MLO views), complex tissue superposition, and modality-specific X-ray physics. Existing NeRF-based medical imaging methods generally assume access to multiple calibrated views or ground-truth volumetric supervision conditions, not available in standard breast imaging. To the best of our knowledge, our work is the first to leverage neural radiance fields for 3D reconstruction of breast tissue from routine mammographic views, integrating recent progress in neural scene modeling with the practical constraints of clinical mammography.

3. Methodology

3.1. Background

NeRF: Given a set of images $\mathcal{I} = \{I_1, I_2, \ldots, I_N\}$ captured from N sparse view points of a scene, with their associated camera parameters $\Pi = \{\pi_1, \pi_2, \ldots, \pi_N\}$, including intrinsic and extrincis. NeRF encodes a scene as a continuous volumetric radiance field F_Θ of color and density. Specifically, for a 3D point $\mathbf{x} = (x, y, z)$ and direction $\mathbf{d} = (\theta, \phi)$, $F_\Theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$ returns a differential density σ and color $\mathbf{c} = (r, g, b)$. To render an image from the NeRF model, the color at each pixel $p = (u, v)$ on the image plane $\hat{I}_i$ is obtained by a rendering function R, aggregating the radiance along a ray originating from the camera position o_i, passing through pixel p into the volume

$$\hat{I}_i(p) = R(p, \pi_i \mid \Theta) = \int_{h_n}^{h_f} T(h)\, \sigma(r(h))\, c(r(h), d)\, dh, \tag{1}$$

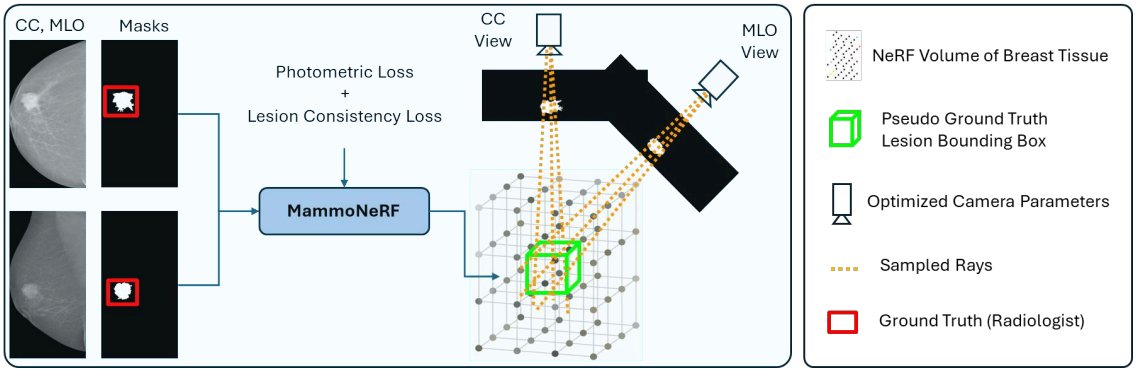

Figure 1: MammoNeRF jointly optimizes a NeRF model, camera parameters, and lesion point representations using paired CC and MLO mammograms. This process enables accurate 3D breast reconstruction and can be trained end-to-end using only the input images, without requiring prior camera calibration.

where the accumulated transmittance along the ray is defined as

$$T(h) = \exp\left(-\int_{h_n}^{h} \sigma(r(s))\, ds\right), \tag{2}$$

denoting the probability that the ray travels from h_n to h without hitting any other particle. The ray itself is parameterized as $r(h) = o + hd$, where o is the camera origin and d is the ray direction determined by the camera parameters π_i. The integration bounds h_n and h_f correspond to the near and far clipping planes. In practice, the integral in Eq. 1 is approximated by sampling a finite set of points along the ray and accumulating their radiance and density contributions. With this implicit scene representation $F_\Theta(x, d)$ and a differentiable renderer R.

Unposed-NeRF: Prior works (Wang et al., 2021; Jeong et al., 2021; Bian et al., 2023) show that we can jointly optimize both camera parameters and a NeRF at the same time by minimizing the photometric error under the same volumetric rendering process. The key lies in conditioning camera ray casting on variable camera parameters Π, as the camera ray r is a function of camera pose. Mathematically, this joint optimisation can be formulated as:

$$\Theta^*,\ \Pi^* = \arg\min_{\Theta,\,\Pi} \mathcal{L}_{\text{photo}}(\hat{\mathcal{I}}, \hat{\Pi} \mid \mathcal{I}).$$

where $\mathcal{L}_{\text{photo}} = \sum_{i=1}^{v} \left\| I_i - \hat{I}_i \right\|_2^2$, represent photometric loss between synthesised images $\hat{I}$ and reel images I

3.2. Problem Definition

We consider CC and MLO mammograms, denoted as I_{CC} and I_{MLO}, along with their binary lesion masks M_{CC} and M_{MLO}. For simplicity and generality in the context of NeRF, since we only have 2D images, the imaging process is modeled using a classical pinhole camera, assuming that this model can adequately capture and preserve the relevant information contained in mammograms.

Our objective is to reconstruct a 3D representation of the breast and obtain a 3D lesion using NeRF. To this end, we propose **MammoNeRF** that learns a neural radiance field from the input set $\{I_{\text{CC}}, I_{\text{MLO}}, M_{\text{CC}}, M_{\text{MLO}}\}$, jointly optimizing: (i) a 3D volumetric representation V, (ii) camera parameters Π, and (iii) a pseudo 3D lesion bounding box $\hat{B}_{\text{PGT}}$, derived from volumetric consistency of the lesion regions across views. Formally, MammoNeRF implements

$$f_\theta : \{I_{\text{CC}}, I_{\text{MLO}}, M_{\text{CC}}, M_{\text{MLO}}\} \mapsto \{V, \Pi, \hat{B}_{\text{PGT}}\},$$

where V denotes the reconstructed breast volume, Π the estimated camera parameters, $\hat{B}_{\text{PGT}}$ the pseudo ground truth bounding box. A key challenge arises from the breast compression applied during image acquisition. The deformation between the CC and MLO views introduces non-linear tissue displacement, making it difficult to consistently align both views within a single 3D volume. Consequently, optimization relying solely on photometric loss tends to reconstruct the same lesion at different spatial locations, leading to incoherent lesion representation and unrealistic estimation of the camera parameters.

3.3. MammoNeRF: Augmenting NeRF Optimization with a Lesion Consistency Loss

We introduce a Lesion Consistency Loss that enforces spatial alignment of the 3D lesion points reconstructed from the CC and MLO views. This results in a single coherent 3D lesion volume, with optimized camera parameters that encourage this alignment. For each view, the binary lesion mask is used to identify lesion pixels in the mammograms. Rays are then cast through these pixels, where N_{CC} and N_{MLO} denote the number of lesion pixels in the CC and MLO views, respectively. Along each ray, we sample a set of 3D points $\{\mathbf{x}_v^{i,j}\}_{j=1}^{k}$, where $v \in \{\text{CC}, \text{MLO}\}$. The centroid of the 3D lesion for each view is then computed as:

$$\mathbf{c}_v = \frac{1}{N_v k} \sum_{i=1}^{N_v} \sum_{j=1}^{k} \mathbf{x}_v^{i,j},$$

where N_v is the number of lesion rays per view and k is the number of sampled points along each ray. The lesion consistency loss is defined as:

$$\mathcal{L}_{\text{lesion}} = \|\mathbf{c}_{\text{CC}} - \mathbf{c}_{\text{MLO}}\|_2^2,$$

which minimizes the distance between the 3D lesion centroids reconstructed from the two views. Unlike conventional NeRF pipelines (Mildenhall et al., 2020; Yu et al., 2021), which require known camera parameters Π^*, mammography presents a unique challenge since the imaging geometry is not explicitly available. Following (Wang et al., 2021; Bian et al., 2023), we jointly optimize the NeRF model, camera poses, and lesion centroids as learnable parameters during training. The final objective of MammoNeRF combines the standard photometric loss with the proposed lesion consistency loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{photo}} + \mathcal{L}_{\text{lesion}}, \tag{3}$$

and the optimization problem is formulated as:

$$\hat{\Theta},\ \hat{\Pi},\ \hat{c} = \arg\min_{\Theta, \Pi, c}\ \mathcal{L}_{\text{total}}(\hat{I}, \hat{\Pi}, \hat{c} \mid I), \tag{4}$$

where $\hat{I}$ is the synthesized mammogram and I the real mammogram; $\hat{\Theta}$ and $\hat{\Pi}$ denote the optimized NeRF and camera parameters, respectively; and $\hat{c}$ represents the optimized lesion centroids. Since the camera parameters Π are unknown, they are initialized with Π^0 and treated as learnable variables during training. Without the lesion consistency loss, NeRF tends to reconstruct the same lesion at different positions across views, leading to incoherent lesion representation. By linking lesion centers, MammoNeRF encourages the formation of a consistent pseudo-3D lesion volume that is clinically more meaningful. The entire pipeline is fully differentiable, enabling joint optimization of NeRF parameters, camera poses, and lesion centroids by minimizing $\mathcal{L}_{\text{total}}$.

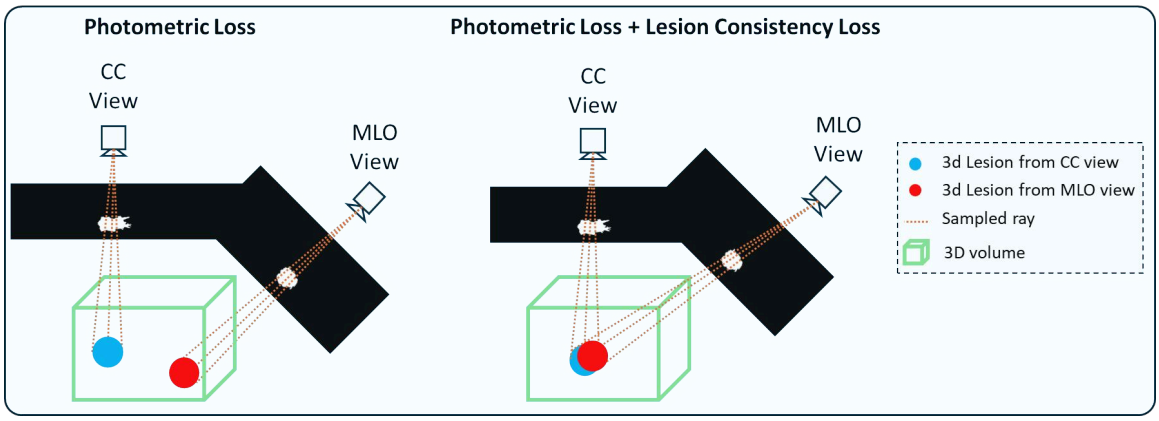

Figure 2: Improve lesion alignment across CC and MLO mammograms using Lesion Consistency Loss.

4. Experiments

4.1. Dataset

We evaluate our method on two public mammography datasets: CBIS-DDSM (Lee et al., 2017) and INbreast (Moreira et al., 2012). In this study, we focus only on lesion masses, as they represent our primary clinical target. Each patient includes two mammographic views: the craniocaudal (CC) and mediolateral oblique (MLO) projections, with their corresponding binary lesion masks. These mammograms are widely adopted in clinical practice because they provide complementary perspectives of breast tissue.

4.2. Implementation Details

We implement our framework in PyTorch following the same architecture as the original NeRFmm (Mildenhall et al., 2020). However, for computational efficiency, (1) we do not use the hierarchical sampling strategy, (2) reduce the hidden layer dimension from 256 to 128, and (3) sample 15 points along each ray for every pixel within the image in both CC and MLO mammograms. We use Kaiming initialization (He et al., 2015) for the MammoNeRF model. The initial camera origin for the CC view is placed at the origin and oriented to look in the negative y-direction. The MLO view is initialized by rotating the CC view by

45 degrees around the y-axis. We use three separate Adam optimisers for MammoNeRF, camera poses, and focal lengths, respectively, all with an initial learning rate of 0.001. The learning rate of the MammoNeRF model is decayed every 20 epochs by multiplying by 0.9954 (equivalent to a stair-cased exponential decay), and learning rates of the pose and focal length parameters are decayed every 100 epochs with a multiplier of 0.9. MammoNeRF is trained for 5,000 epochs per patient case using paired CC and MLO mammograms. We obtain pseudo ground-truth 3D lesion bounding boxes (3DPB) by aggregating the sampled points along the rays that intersect the lesion mask within the reconstructed volume. The 3DPB are then computed by identifying the minimal and maximal coordinates of these sampled points, thereby forming the smallest axis-aligned box that fully encloses the lesion. All experiments are run on NVIDIA A100-SXM4 GPUs with 80 GB of RAM.

4.3. Baselines

To evaluate the effectiveness of our proposed MammoNeRF model, we compare it against several baselines. NeRF (Mildenhall et al., 2020) serves as the foundational model, relying solely on photometric supervision and fixed camera poses to learn a continuous volumetric representation. NeRFmm (Wang et al., 2021) extends NeRF by jointly optimizing camera poses together with scene parameters, providing improved robustness in scenarios with uncertain imaging geometry. By comparing MammoNeRF with these baselines, we highlight the importance of our lesion consistency loss and demonstrate the importance of our approach to produce more stable and anatomically coherent 3D reconstructions from sparse dual-view mammography.

4.4. Evaluation metrics

We evaluate the MammoNeRF model based on its novel view synthesis performance, employing standard image-quality metrics, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) (Wang, 2004). To assess the fidelity of the pseudo ground-truth 3D lesion bounding boxes (3DPB) generated by MammoNeRF, we project the reconstructed 3D volumes onto the corresponding 2D CC and MLO mammographic views using the estimated camera poses. The resulting 2D projections are then quantitatively compared with the clinical annotations by computing the Intersection over Union (IoU), providing a rigorous measure of spatial alignment and lesion localization accuracy.

5. Ablation Study

To investigate the effect of image resolution on reconstruction quality and computational efficiency, we evaluate MammoNeRF under different sub-sampling factors. Table 1 demonstrates how PSNR and SSIM vary with the number of points sampled along rays, alongside the corresponding training efficiency measured in iterations per second (it/s). As expected, higher sample counts result in increased processing time. However, despite these variations in sampling density, PSNR and SSIM remain relatively stable, indicating consistent rendering quality across all configurations. Similarly, increasing the sub-sampling factor Table 2 accelerates training while still preserving competitive reconstruction quality. The highest

sub-sampling factor (4×) could not be processed due to memory limitations. These results indicate that lightweight sampling provides the most effective performance–efficiency trade-off for MammoNeRF.

Number of Samples	PSNR	SSIM	Time (it/s)
8	37.76	0.907	16.23
12	37.86	0.934	11.95
24	38.72	0.907	6.54
64	36.132	0.890	5.07
128	36.132	0.890	1.31

Table 1: Performance metrics (PSNR, SSIM) and processing speed (iterations per second) for different numbers of sample points along rays.

Sub-sampling Factor	Image size ($H \times W$)	PSNR	SSIM	Time (it/s)
8	(588×309)	37.87	0.907	3.40
16	(294×154)	37.86	0.934	11.76
24	(196×103)	38.72	0.905	23.46

Table 2: Quantitative evaluation of reconstruction quality and efficiency under different sub-sampling factors.

6. Results

After processing the datasets with several preprocessing steps, we evaluate MammoNeRF to determine whether the model can generate detailed reconstructions that preserve both global breast anatomy and subtle lesion characteristics. To this end, we report PSNR and SSIM, which measure pixel-level accuracy and perceptual quality, respectively. Figure 6

Dataset	Model	PSNR	SSIM
CBIS-DDSM (Lee et al., 2017)	NeRF (Mildenhall et al., 2020)	25.469	0.764
	Nerfmm (Wang et al., 2021)	36.469	0.9055
	Ours	36.132	0.8974
INbreast (Moreira et al., 2012)	NeRF (Mildenhall et al., 2020)	26.579	0.754
	Nerfmm (Wang et al., 2021)	36.052	0.8705
	Ours	35.629	0.8550

Table 3: Quantitative evaluation of novel view synthesis on CBIS-DDSM and INbreast datasets using PSNR and SSIM.

demonstrates the high intersection over union (IoU) between the projected pseudo ground-truth 3D lesion bounding boxes (3DPB) in 2D using estimated camera parametrs and the original ground-truth bounding boxes, further validating the effectiveness of our approach. This strong geometric agreement indicates that the reconstructed 3D lesion locations are consistent with their 2D annotations, demonstrating accurate cross-view alignment and reliable volumetric reasoning. Table 3 summarizes the synthesis performance across all patients

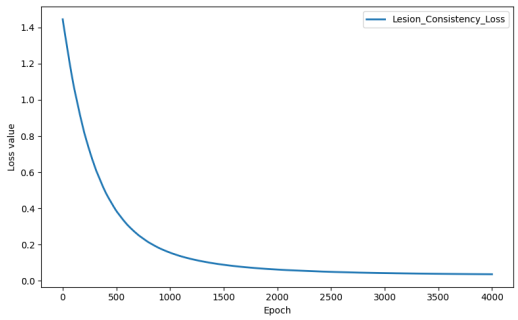

Figure 3: Training curve of the lesion consistency loss.

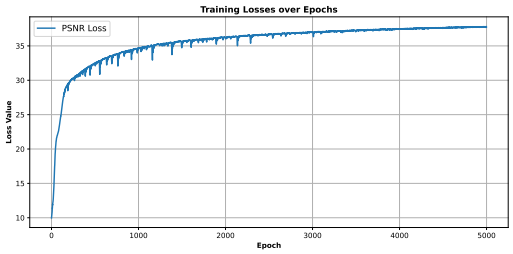

Figure 4: Training curve of PSNR loss.

in the datasets. Our method consistently achieves higher PSNR and SSIM scores, demonstrating both robustness and generalizability. Notably, MammoNeRF effectively preserves the same structures of the breast, including lesion boundaries, which is an essential factor for detection tasks. Furthermore, the incorporation of the lesion consistency loss does not degrade the synthesis quality, confirming the robustness and stability of the proposed training strategy. Finally, the projection of reconstructed pseudo ground truth-lesion 3D bounding boxes into 2D mammograms, using both learned intrinsic and extrinsic camera parameters $\hat{\Pi}$, remains accurate and robust across all patients in our datasets. Figure 3 and Figure 4

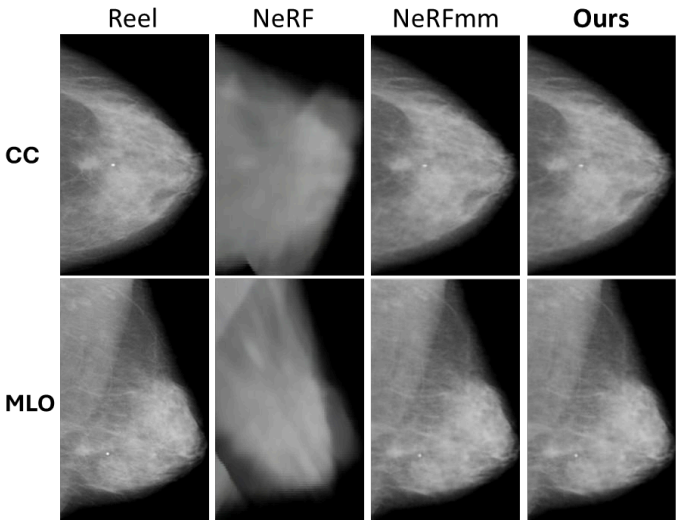

Figure 5: Qualitative evaluation of novel view synthesis.

present the training dynamics of the lesion consistency loss and the PSNR loss, respectively. The lesion consistency loss provides a rapid decrease during the early epochs, reflecting the ability of the model to quickly enforce cross-view alignment of lesion features, after which it gradually stabilizes at a low value as the reconstruction of lesion regions becomes more coherent. The PSNR curve increases in the early epochs as the model rapidly improves its ability to reconstruct image appearance. After this initial rise, the curve continues to grow

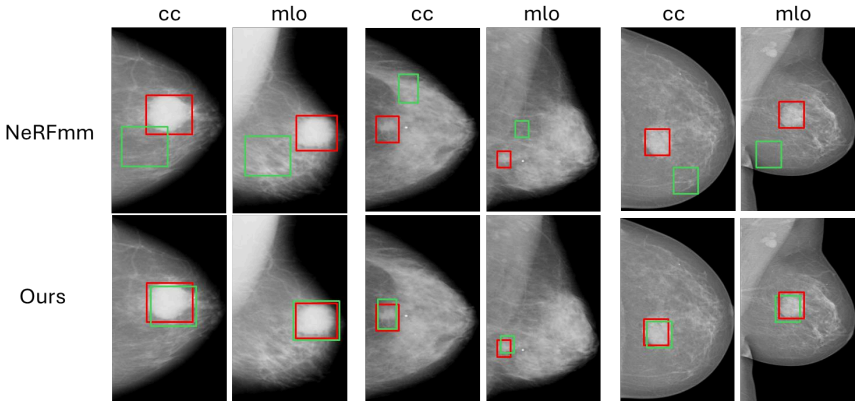

Figure 6: Comparison between Ground truth bounding boxes (red) and the projection of pseudo ground-truth 3D lesion bounding boxes (green) generated using learnable camera parameters with MammNeRF and NeRFmm.

more slowly, demonstrating steady refinement of fine-grained details and overall image fidelity. Together, these curves demonstrate that the model enhances anatomical consistency across views while refining image fidelity over the course of optimization.

7. Conclusion

In this study, we presented a novel framework for reconstructing 3D volumetric breast representations from paired CC and MLO mammograms using neural radiance fields. By introducing a lesion consistency loss that enforces cross-view alignment of lesion regions, our method produces coherent and anatomically plausible volumetric reconstructions that accurately capture true lesion morphology. This approach effectively bridges the gap between traditional 2D mammography and volumetric imaging, enabling richer spatial feature extraction and improved contextual reasoning. Experiments on the CBIS-DDSM and INbreast datasets demonstrate that our framework achieves state-of-the-art performance compared to publicly available methods, underscoring the potential of pseudo-3D reconstruction as a powerful augmentation strategy for breast imaging analysis. Future work will explore the incorporation of anatomical landmarks beyond lesions to further enhance cross-view geometric consistency, as well as the adoption of emerging rendering strategies such as Gaussian splatting to improve reconstruction fidelity and efficiency. Additionally, we plan to investigate the integration of the reconstructed 3D volumes into lesion detection and segmentation to assess their impact on clinical decision support and overall diagnostic accuracy.

References

Wenjing Bian, Zirui Wang, Kejie Li, Jia-Wang Bian, and Victor Adrian Prisacariu. Nope-nerf: Optimising neural radiance field with no pose prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4160–4169, 2023.

Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view x-ray 3d reconstruction. In *CVPR*, 2024.

Abril Corona-Figueroa, Jonathan Frawley, Sam Bond-Taylor, Sarath Bethapudi, Hubert PH Shum, and Chris G Willcocks. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. In *2022 44th annual international conference of the IEEE engineering in medicine & Biology society (EMBC)*, pages 3843–3848. IEEE, 2022.

Krzysztof J Geras, Stacey Wolfson, Yiqiu Shen, Nan Wu, S Kim, Eric Kim, Laura Heacock, Ujas Parikh, Linda Moy, and Kyunghyun Cho. High-resolution breast cancer screening with multi-view deep convolutional neural networks. *arXiv preprint arXiv:1703.07047*, 2017.

Shalini Gupta and Mia K Markey. Correspondence in texture features between two mammographic views. *Medical physics*, 32(6Part1):1598–1606, 2005.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

Yoonwoo Jeong, Seokjun Ahn, Christopher Choy, Anima Anandkumar, Minsu Cho, and Jaesik Park. Self-calibrating neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5846–5854, 2021.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), July 2023. URL https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

Rebecca Sawyer Lee, Francisco Gimenez, Assaf Hoogi, Kanae Kawai Miyake, Mia Gorovoy, and Daniel L Rubin. A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data*, 4(1):1–9, 2017.

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, 2020. URL https://arxiv.org/abs/2003.08934.

Inês C Moreira, Igor Amaral, Inês Domingues, António Cardoso, Maria Joao Cardoso, and Jaime S Cardoso. Inbreast: toward a full-field digital mammographic database. *Academic radiology*, 19(2):236–248, 2012.

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.

Dongbo Shi, Shen Cao, Bojian Wu, Jinhui Guo, Lubin Fan, Renjie Chen, Ligang Liu, and Jieping Ye. Nope-nerf++: Local-to-global optimization of nerf with no pose prior. In *Computer Graphics Forum*, page e70012. Wiley Online Library, 2025.

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.

Z Wang. Image quality assessment: Form error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):604–606, 2004.

Zirui Wang, Shangzhe Wu, Weidi Xie, Min Chen, and Victor Adrian Prisacariu. NeRF−−: Neural radiance fields without known camera parameters. *arXiv preprint arXiv:2102.07064*, 2021.

Jelmer M Wolterink, Jesse C Zwienenberg, and Christoph Brune. Implicit neural representations for deformable image registration. In *International Conference on medical imaging with deep learning*, pages 1349–1359. PMLR, 2022.

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.

Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4578–4587, 2021.

Weihao Yu, Yuanhao Cai, Ruyi Zha, Zhiwen Fan, Chenxin Li, and Yixuan Yuan. X2-gaussian: 4d radiative gaussian splatting for continuous-time tomographic reconstruction. In *ICCV*, 2025.

Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: neural attenuation fields for sparse-view cbct reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–452. Springer, 2022.

Appendix A. Datasets

We use the CBIS-DDSM (Lee et al., 2017) and INbreast (Moreira et al., 2012) mammography datasets, each containing two standard views: craniocaudal (CC) and mediolateral oblique (MLO). The raw mammograms often include artifacts, grid lines, and large black background regions that can negatively impact the performance of MammoNeRF during training. To mitigate these issues, we apply the preprocessing strategy described in (Geras et al., 2017) to both the mammograms and their corresponding binary masks before inputting them into the MammoNeRF.

Dataset	**CBIS-DDSM** (Lee et al., 2017)	**INbreast** (Moreira et al., 2012)
Patients	1324	115
Masses	600	50

Table 4: Summary of the datasets used in this study.

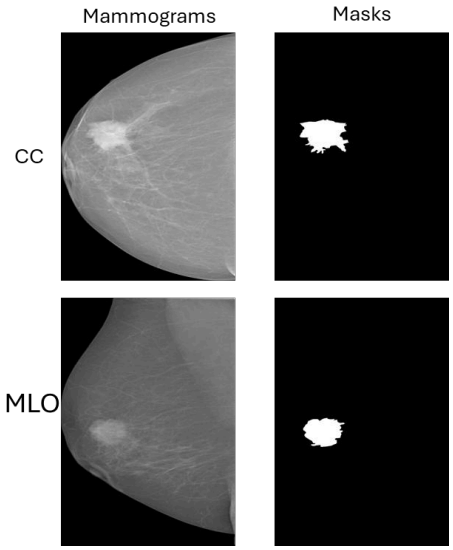

Figure 7: Sampled volumes generated using MammoNeRF across CC and MLO views

Appendix B. Additional results

We run MammoNeRF for 5,000 training iterations per patient and evaluate reconstruction quality using Peak Signal-to-Noise Ratio (PSNR). During training, the incorporation of the lesion consistency loss leads to a reduction in the total loss, indicating improved model stability. This stability is reflected both in the optimized alignment of lesion centroid points across views and in the enhanced quality of the rendered images. Figure 3 and Figure 4 shows the PSNR loss during training over 5000 epochs. Initially, the PSNR value increases sharply from around 10 to approximately 30 within the first few hundred epochs, indicating a rapid improvement in reconstruction quality. Thereafter, the curve continues to increase more gradually, with small oscillations (likely due to mini-batch variations or optimization noise). By around epoch 3000, the curve begins to plateau, stabilizing near 38, which suggests convergence of the model.

The model reconstructs coherent volumetric structures by leveraging information across both CC and MLO projections, producing anatomically plausible representations of breast tissue. These results demonstrate the capability of MammoNeRF to infer consistent 3D geometry from standard 2D screening views, highlighting its potential as a tool for volumetric breast analysis and visualization. Even though the model sees only 2D images, it is able to build a full 3D shape that resembles real breast anatomy, including tissue layers and density variations. Some visual artifacts appear, which is common for NeRF methods, but the

overall structure remains realistic. This demonstrates that MammoNeRF can turn standard mammograms into 3D volumes that may help with visualization and analysis.

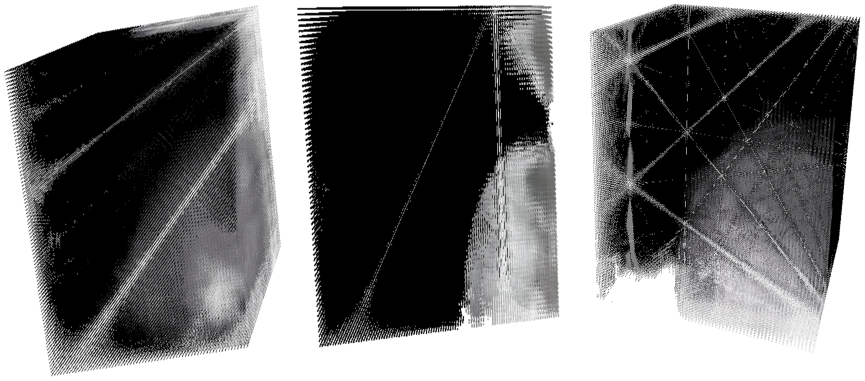

Figure 8: Sampled volumes generated using MammoNeRF across CC and MLO views