Supplementary Material for Seeing through boxes: Non-Line-of-Sight 3D Reconstruction from Radar Signals

Jiachen Lu*, Hailan Shanbhag*, Haitham Al Hassanieh
École Polytechnique Fédérale de Lausanne (EPFL)

Supplementary Material Organization

We include the following items in the supplementary material:

1. Section 1: Training parameters and data collection setup.
2. Section 2: Additional results, ablation studies and novel view synthesis results.
3. Section 3: Detailed technical background information.
4. Section 4: Limitations and future research directions.

1. Experiment Details

Experiment Setup The antenna arrays are emulated using synthetic aperture radar (SAR), achieved by moving a radar sensor mounted on a robotic arm across multiple 2D planes around the object. We simulate 36 distinct scanning poses, covering angles from $0°$ to $350°$ in $10°$ increments. Each scan covers an area of $0.14\,\mathrm{m} \times 0.25\,\mathrm{m}$ with antenna spacing of approximately $\frac{\lambda}{4}$. The object is positioned approximately $0.3\,\mathrm{m}$ away from the radar. The radar operates with a total chirp bandwidth of approximately $4\,\mathrm{GHz}$.

Training Setup For vision, we follow all training settings from NeuS [6], including model parameters, resolution, and training strategy.

For radio frequency, the SDF Network is implemented as an MLP with 8 layers and a hidden dimension of 256. We apply sinusoidal positional encoding with 10 frequency levels as input. The Reflectivity Network is implemented as an MLP with 4 layers and a hidden dimension of 256. Signal power prediction is implemented as a single trainable scalar parameter.

We train our model for 100,000 iterations per stage (Stage 1 and Stage 2) over 48 hours on an NVIDIA H100 GPU, using `mmDetection3D` [2] as the codebase. In Stage 1, we freeze the Reflectivity Network and initialize it to output a constant value of 1.0. In Stage 2, all parameters are trained jointly.

We use an initial learning rate of 1×10^{-3}; however, due to the sparsity of the input, the learning rate for the SDF Network is reduced to 1×10^{-4}. Training is performed using the AdamW optimizer with cosine annealing learning rate scheduling.

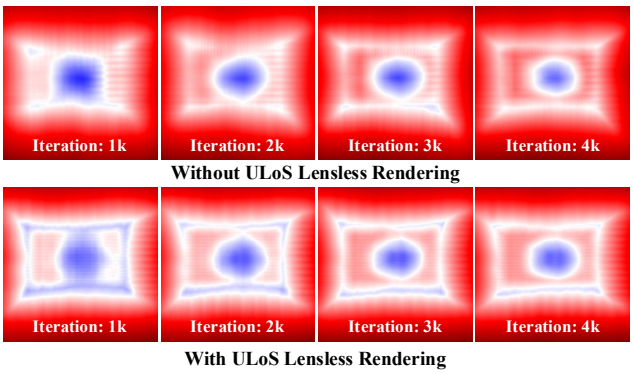

Figure 1. Ablation study on ULoS Lensless Rendering. We compare SDF slices of our Stage 1 baseline (***bottom***) with a variant that does not leverage the vision-trained SDF (***top***) during early training iterations on the *bunny* object. Red indicates positive values, blue indicates negative values, and white represents the surface (zero level set).

Metric Calculation We used the Chamfer distance and the F1-score (evaluated with a threshold of $\tau = 0.015$) for comparing our system to GeRaF and the Matched Filter. Both metrics evaluate how similar two 3D point clouds are. For mesh-to-point-cloud conversion, we uniformly sample 10,000 points from each mesh corresponding to the three candidate methods. We then rigidly align the matched-filter point clouds to the camera baseline, and align our outputs point clouds to the same reference.

The Chamfer distance is computed by, for each point in one cloud, finding its nearest neighbor in the other cloud and computing the squared distance. The final Chamfer distance is obtained by averaging these distances in both directions and summing the results.

The F1-score is computed by finding, for every point in one point cloud, the nearest point in the other cloud, and repeating this process in reverse. Precision (P) and Recall (R) are defined as the fractions of nearest-neighbor distances that fall below τ. The F1-score is then given by $F_1 = 2 \cdot (\mathrm{P} \cdot \mathrm{R})/(\mathrm{P} + \mathrm{R})$.

*Co-primary first authors, indicates equal contribution.

Table 1. Quantitative results (*different box).

Object	F1↑			CD (mm)↓		
	MF	GeRaF	Ours	MF	GeRaF	Ours
Bunny	0.329	0.822	**0.962**	4.24	0.28	**0.15**
Bunny*	0.790	0.859	**0.964**	2.87	0.28	**0.12**
Elephant	0.517	0.77	**0.845**	9.9	0.41	**0.24**
Elephant*	0.607	0.568	**0.734**	1.69	1.11	**0.47**
Deer	0.550	0.643	**0.786**	1.20	0.45	**0.24**
Chicken	0.376	0.869	**0.941**	6.19	0.21	**0.14**
Boat	0.595	0.684	**0.779**	0.93	0.47	**0.43**
Ball	0.389	0.560	**0.753**	1.42	0.94	**0.52**

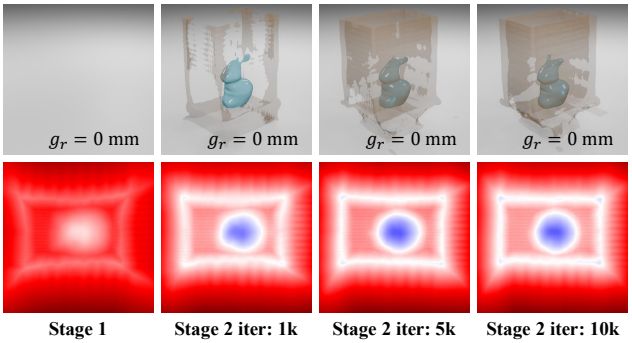

Figure 2. Effect of RSDF alignment during training, from Stage 1 (initialization) to 10,000 iterations. We show 3D reconstruction results (***top***) and SDF slices (***bottom***) of the *bunny* object, taken at Stage 1, and at 1,000, 5,000, and 10,000 iterations of Stage 2. Red indicates positive SDF values, blue indicates negative values, and white denotes the surface (zero level set).

2. Additional Experiments

2.1. Quantitative Comparisons

Quantitative results are shown in Tab. 1. UniRaF shows clear improvements in both F1-score and Chamfer distance (CD) over both the matched filter and GeRaF. It is notable that when the Matched Filter result has a high F1-score or CD both GeRaF and UniRaF report higher scores.

2.2. Ablation Studies on ULoS Lensless Rendering

In Fig. 1, we compare the effect of ULoS Lensless Rendering. This module leverages the vision-trained SDF as guidance to adjust the accumulated transmittance. We visualize a slice of the SDF during the early stage of Stage 1 training on *bunny* to observe the impact on convergence. With this guidance, the results shown at the bottom of Fig. 1 demonstrate faster and more accurate shape formation for both the box and the object.

2.3. Additional Ablation Studies on RSDF alignment

We include additional ablation studies. In Fig. 2, we visualize the reconstruction results and SDF slices during Stage 2 training on the *bunny* inside the box. With the help of primary depth supervision from the vision-trained SDF, RSDF alignment quickly brings the zero level set to the correct surface within only 10,000 iterations.

In Fig. 3, we show additional ablation studies on RSDF alignment. We compare the SDF slice for different objects. without RSDF aligmnent, the RF-trained SDF is not achieved the correct level while the shape is also incorrect. However, with help the RSDF alignment, it not only achieved zero level set, the shape is also correct.

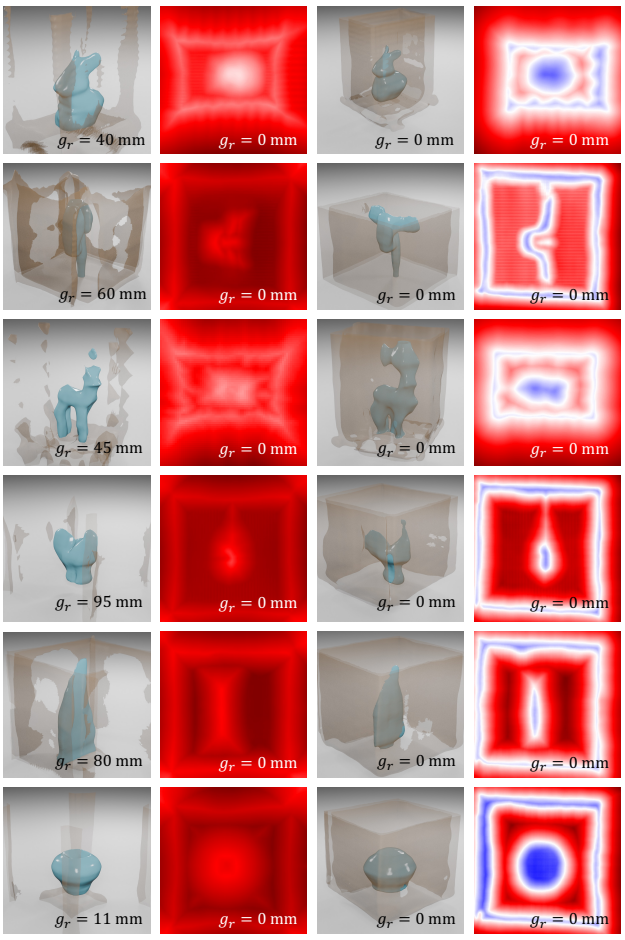

Figure 3. Ablation study on RSDF alignment. We show both 3D reconstruction results and SDF slices. (a) shows the results from Stage 1, and (b) shows the results after Stage 2 training.

In Fig. 3, we present additional ablation studies on RSDF alignment by comparing SDF slices for different objects. Without RSDF alignment, the RF-trained SDF fails to reach the correct zero level set, and the reconstructed shape is also

inaccurate. In contrast, with RSDF alignment, the surface converges to the correct zero level set and the overall shape reconstruction is significantly improved.

2.4. Novel View Synthesis

To evaluate novel view synthesis (NVS), we remove antenna planes at angles $0°$, $90°$, $180°$, and $270°$ from the training set and include them in the evaluation set. Fig. 4 shows qualitative and quantitative (PSNR [4]) NVS results for the *bunny* object. Unlike NVS in the visible light spectrum, mmWave synthesis typically lacks high-frequency spatial features. However, the reconstruction of low-frequency structures and the relatively high PSNR values demonstrate the model's potential capability to learn both the underlying geometry and the specific characteristics of mmWave imaging.

2.5. Number of Scanning Plane Impact

In the main paper, we use 36 measurements per object, with a $10°$ angular interval between viewing directions. We include an ablation study evaluating reconstruction quality relative to the number of scanning planes (images) used. Fig. 5 shows evaluation against varying measurement densities using $20°$, $30°$, $60°$, $90°$ and $180°$ intervals. The results show that UniRaF maintains strong reconstruction quality even with heavily reduced data. Using only 1/6 of the measurements ($60°$) still recovers the overall geometry, though some regions (such as the bunny's back) are missing, and using 1/9 of the measurements ($90°$) maintains the bunny's body, though loses the ear reconstruction. Given the minor performance drop, certain applications benefit from lower training times while preserving an acceptable reconstruction quality.

2.6. Non-Specular Objects

While our formulation proposed uses a specular-reflection model, it can handle some diffusion because our ray tracer uses a small angular spread around the specular direction. To demonstrate this, we provide an initial results for a 3D printed bunny (PLA) in Fig. 6, showing consistent performance for this material type. While this model is sufficient for the objects and materials tested, it doesn't necessarily capture all surface complexities. We acknowledge that future work will require power attenuation adjustment in the network or learning of the diffusion pattern based on material/texture to account for more diffused materials or a composite of materials.

3. Radio Frequency Background

3.1. Radar Basics

A mmWave radar works by transmitting a wireless signal (chirp) and receiving back the reflections that come from various reflectors in the scene. In this case, the transmitter and receiver are collocated, meaning they are side by side. It operates in the millimeter-wavelengths frequency bands at 77 GHz, and uses Frequency Modulated Continuous Wave (FMCW) and antenna arrays to help resolve spatial ambiguity. To resolve range ambiguity, the received chirp is multiplied with the conjugate of the transmitted chirp and can be expressed as a complex function:

$$s(t) = A \cdot e^{-j2\pi(f+kt)d/c} = A \cdot e^{-(j2\pi k\tau)t} \cdot e^{-j2\pi f\tau} \quad (1)$$

where A is the signal amplitude, d is the round-trip propagation distance, c is the speed of light, $\tau = d/c$ is the round-trip delay, f is the starting frequency, and k is the chirp slope. For multiple reflectors, we simply receive the linear combination of all the reflections.

3.2. Free-space Power Decay

In free space, the power of a radio frequency signal is inversely proportional to the square of the distance it travels due to spherical spreading. We consider the round-trip path, where the signal travels from the transmitter to a point in space and then reflects back to the receiver, then decay is even more pronounced [5]. This is known as *round-trip free-space path loss*, and the *power* decay factor reflected from distance u is given by:

$$P_{\text{rx}} \propto \frac{1}{(4\pi u)^4} P_{\text{tx}} \quad (2)$$

This expression accounts for two instances of inverse-square spreading: one during transmission to the point and another during reflection back to the receiver. As such, the received power decreases proportionally to $1/u^4$ (amplitude would be by a factor of $1/u^2$).

3.3. Distance

Radio frequency signals are modeled differently based on the distance the transmitter and receiver are relative to the reflectors in the scene. Commonly, this is referred to as near-field (when the objects are in much closer to the transmitter/receiver pair compared to the wavelength of the radio frequency signal) and far-field (when the objects are much further compared to the wavelength). In our system, the object is much closer to the transmitter and receiver, as compared to the wavelength, meaning when the signals from the antenna array meet the object, the waves cannot be modeled as parallel. In other words, the direction of the RF signal from different parts of the antenna array will have vastly different directions. As compared to when the object is very far, then the waves can be modeled as parallel to each other [5], which significantly simplifies the radar heatmap reconstruction. This is because when the waves are assumed to be parallel we can reuse filter weights, and use

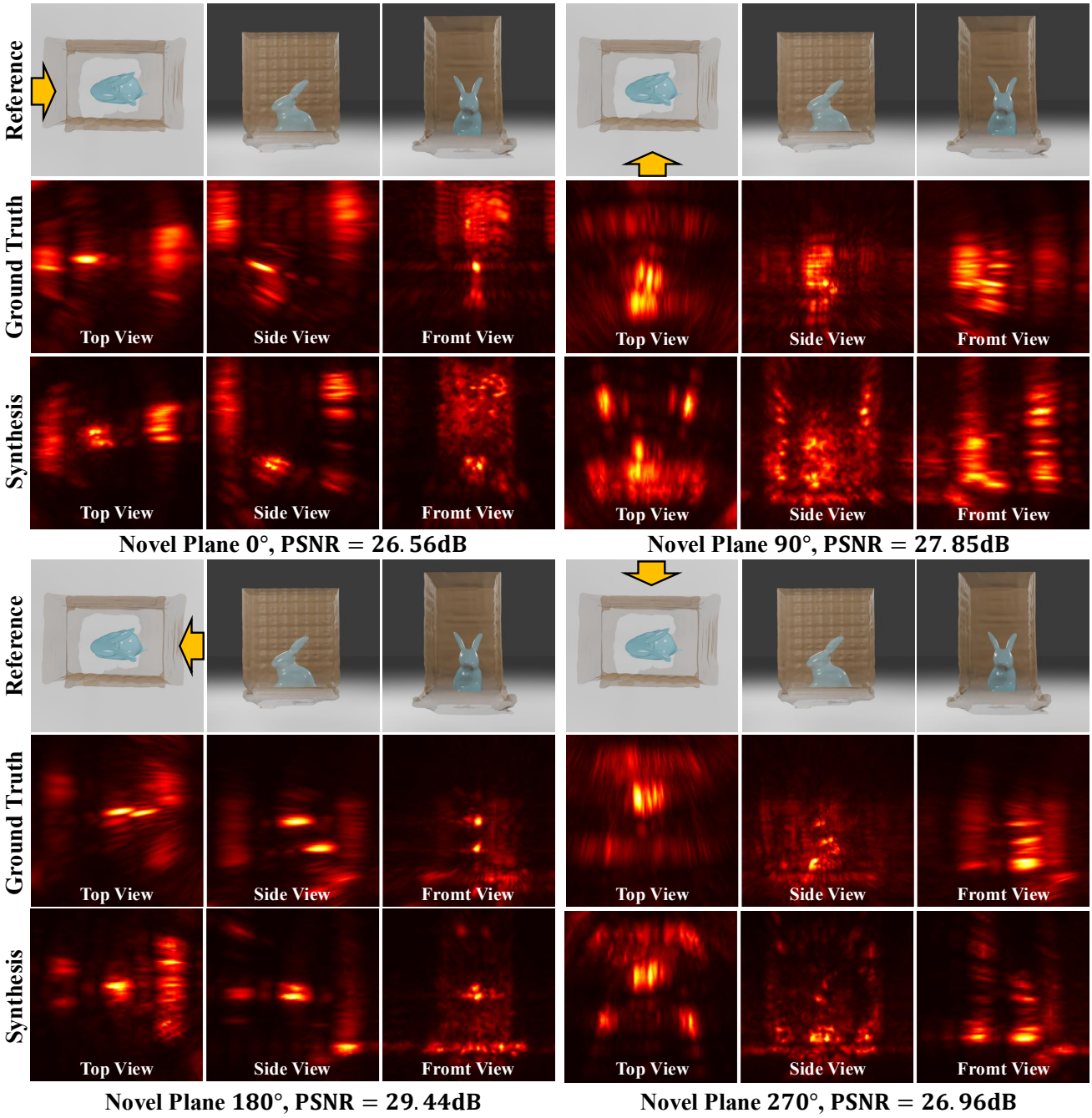

Figure 4. Novel view synthesis on antenna planes at 0°, 90°, 180°, and 270°. For each view, we show the corresponding reference image from the vision modality for comparison. The orange arrow indicates the incoming signal direction. The 3D matched filter results are projected onto the three axes using maximum intensity projection.

Fourier Transforms to speed up the computation time [1]. On the other hand, in the near-field, we must use more accurate reconstruction methods (e.g. Matched Filter) which significantly increases the computational complexity of the system.

3.4. Differentiable Matched Filter & Signal Tracing

We have the forward path of the Matched Filter:

$$P(\mathbf{x}_j) = \left\| \sum_{i=1}^{N_{\text{ant}}} \sum_{t} s(i,t) \cdot e^{j2\pi k \tau_i t} \cdot e^{j2\pi f \tau_i} \right\|, \mathbf{x}_j \in \Omega_{\text{pts}},$$

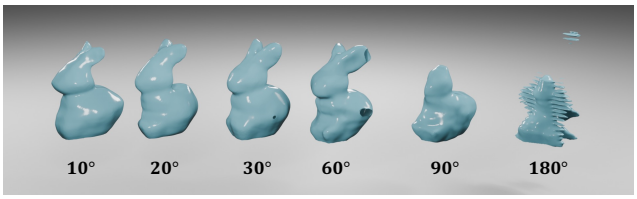

Figure 5. Ablation study on number of scanning planes used in training.

Figure 6. Reconstruction results for non-specular objects.

The backpropagated gradient to the signal $s(i,t)$ is given by:

$$\frac{\partial L}{\partial s(i,t)} = \sum_{\mathbf{x}_j \in \Omega_{\text{pts}}} \frac{1}{P(\mathbf{x})} \cdot \frac{\partial L}{\partial P(\mathbf{x})} \cdot e^{-j2\pi k \tau_i t} \cdot e^{-j2\pi f \tau_i} \quad (3)$$

Signal Tracing forward path is defined by,

$$s(i,t) = \sum_{\mathbf{x}_j \in \Omega_{\text{pts}}} A_{\text{rx}}(\mathbf{x}_j)\, e^{-j2\pi k \tau_j t}\, e^{-j2\pi f \tau_j},$$

The backpropagated gradient to the amplitude $A_{\text{rx}}(\mathbf{x}_j)$ is given by:

$$\frac{\partial L}{\partial A_{\text{rx}}(\mathbf{x}_j)} = \sum_{i=1}^{N_{\text{ant}}} \sum_{t} \frac{\partial L}{\partial s(i,t)} \cdot e^{j2\pi k \tau_i t} \cdot e^{j2\pi f \tau_i} \quad (4)$$

4. Future Work & Discussion

Materials While we present some initial results of non-metal reconstruction, our system currently shows results for objects that are primarily metal. This means the reflections we expect from the inner surface we are reconstructing reflect very clearly through the box. On the other hand, for objects which are made of a material which reflects much weaker than the box itself, may produce too noisy of matched filter heatmaps for our system to properly reconstruct the inside, because of the more complex RF interactions. Dealing with this requires adjusting the radar reflection model as well as the transmittance model. Additionally, future work will look into objects made of a composition of different materials.

Radar Scanning Methods If there was a point that existed on the surface of the object, which never reflect *at all* back to the radar in any of the scans taken, then the network has no way of truly recreating a point in that location, because it has never received any information from that location. Future work is required to deal with unseen surfaces, and more comprehensive scanning methods.

Computational Complexity The computational cost of RF rendering still remains significantly higher compared to vision-based rendering. For comparison to state-of-the-art radar processing, the matched-filter takes roughly 40 minutes to process each of the 36 ground-truth images at 1 mm resolution. Compared to existing neural RF imaging baseline GeRaF [3], which report training times near 32 hours per scene, our pipeline offers a similar efficiency. As neural RF representations are an emerging field, we believe that optimizing 3D space-sampling and training efficiency is a vital direction for future work.

References

[1] Fredrik Andersson, Randolph Moses, and Frank Natterer. Fast fourier methods for synthetic aperture radar imaging. *IEEE Transactions on Aerospace and Electronic Systems*, 2012. 4

[2] MMDetection3D Contributors. MMDetection3D: OpenMMLab next-generation platform for general 3D object detection. https://github.com/open-mmlab/mmdetection3d, 2020. 1

[3] Jiachen Lu, Hailan Shanbhag, and Haitham Hassanieh. Geraf: Neural geometry reconstruction from radio frequency signals. In *NeurIPS*, 2025. 5

[4] 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*, 2021. 3

[5] Mark A Richards, Jim Scheer, William A Holm, and William L Melvin. Principles of modern radar. 2010. 3

[6] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *NeurIPS*, 2021. 1