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

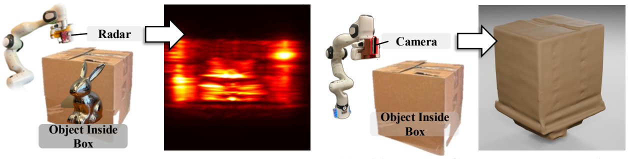
Jiachen Lu*, Hailan Shanbhag*, Haitham Al Hassanieh
École Polytechnique Fédérale de Lausanne (EPFL)

Abstract

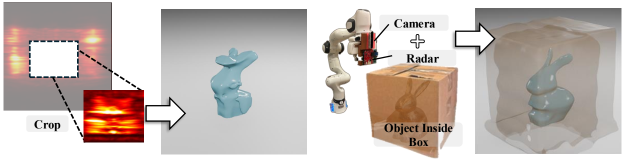
Reconstructing object geometry from radio frequency (RF) signals is fundamentally challenging due to the lensless imaging nature of RF sensing, which leads to low spatial resolution and high noise. Unlike light signals, RF signals can penetrate occlusions and thus capture information about hidden scenes. Existing Non-Line-of-Sight (NLoS) 3D neural reconstruction methods can recover coarse surfaces inside enclosed environments but often suffer from unstable optimization, noisy surface geometry, and surface ambiguity, failing to produce accurate zero-level sets from the signed distance field (SDF). These limitations largely stem from neglecting the role of Line-of-Sight (LoS) geometry outside the enclosed region, which provides valuable physical constraints for modeling signal propagation. In this paper, we introduce a Unified LoS and NLoS neural geometry reconstruction framework GeRaF 2.0 that leverages the outside LoS geometry to model and guide RF propagation from the LoS region into the NLoS region. By integrating visual LoS priors into the neural field formulation, GeRaF 2.0 achieves stable training and physically consistent reconstruction of both visible and hidden geometry, setting a new state-of-the-art in RF-based geometry reconstruction.

1. Introduction

Radio frequency (RF) reconstruction has exploded in recent years as a robust and versatile sensing modality due to its unique ability to see *through* occlusions and remain reliable under challenging visibility conditions. This unique ability to see through occlusions and operate in non-line-of-sight scenarios while being safe for humans [47], unlocks a large range of applications, such as allowing robots to see hidden objects inside boxes or behind clutter or allowing smart home devices to interact with occluded regions [1, 56].



(a) Radio Frequency Imaging (b) Vision-based Geometry Reconstruction



(c) Exsiting RF-based Geometry Reconstruction (d) Our Unified LoS & NLoS Reconstruction

Figure 1. This is the first work which uses line-of-sight modalities to *boost* non-line-of-sight 3D reconstruction for high resolution 3D reconstruction. **a)** Radar heatmap from combining multi-view images, **b)** vision based reconstruction of the line-of-sight surface, **c)** GeRaF [29] which crops the box out of reconstruction, and **d)** GeRaF 2.0 which combines line-of-sight and non-line-of-sight reconstruction.

However, directly reconstructing 3D objects from RF signals is challenging. Due to their *lensless* nature, the pinhole camera model commonly used in vision does not apply. Each antenna can receive signals from the whole scene, bringing challenges in high-cost sampling, low resolution, and high noise levels. Traditionally, it is common to use an array of antennas and combine the signals across antennas [50], producing more interpretable reconstructions. However, compared to vision reconstruction, the reconstruction from RF is corrupted by noise artifacts, missing surface patches due to specular RF reflections, and very low spatial resolution due to physical limitations of the antenna array apertures.

Recently, there has been a growing interest in neural reconstruction methods for radio frequency [4, 17, 22, 29, 57], which try to adapt and apply these methods to RF sensing to represent the geometry of a scene continuously, enabling smoother and more accurate 3D representations. Many of these works [4, 17, 22, 57] only reconstruct scenes in **Line-of-Sight (LoS)** and are geared towards mapping the environment for autonomous navigation rather than reconstructing detailed 3D models of complex objects. Our recent work GeRaF [29], reconstructs surfaces in **Non-Line-of-Sight (NLoS)** (e.g. behind paper or in a box) by simply cropping out the occlusion and treating as if it was not there (i.e. treating LoS and NLoS in the same way). However, the

---

[*] Co-primary first authors, indicates equal contribution.

LoS model incorrectly assumes that the wireless signal that reaches the NLoS surface after passing through the occluding surface, is completely unaffected by the LoS surface. In reality, interactions with the LoS occlusion is unavoidable, because the wireless signal that passes through the LoS occlusion is partially *reflected* and *attenuated* by the occluding surface [9, 37].

As a result, GeRaF [29] suffers from: (1) inaccurate surface reconstruction, since wireless reflections from the visible surfaces can "leak" into the hidden region, appearing as noise that alters the reconstruction, as shown in Fig. 1(c), where the bunny has a strange "hat" shape influenced by the box; (2) unstable training, because different LoS geometries (e.g., boxes of different shapes and sizes) can alter the optimization landscape, leading to failure to converge to a stable surface in some cases; (3) surface ambiguity, since the LoS geometry affects the power of the wireless signal reaching the NLoS regions, it becomes difficult to normalize the signal strength and, as a result, to determine the true surface (i.e., the zero-level set of the SDF). For example, the surface in Fig. 1(c) is not selected based on the SDF being equal to zero, but rather offset by a few centimeters. On the other hand, the vision-based reconstruction methods well known from [30, 46] are advantageous for stable training and accurate surface recovery. As shown in Fig. 1(b), vision successfully reconstructs a highly accurate LoS box, but it cannot see inside the box. This prompts us to ask the question: *can we leverage* ***visible*** *information outside the boxes to help see through them?*

To address this, we propose a **Unified Line-of-Sight and Non-Line-of-Sight (ULoS)** neural geometry reconstruction framework **GeRaF 2.0** that exploits stable and accurate information from line-of-sight modalities outside the box to guide low-resolution and noisy non-line-of-sight modalities inside the box. (1) We represent the combined LoS and NLoS regions as nested, closed, and compact sets, enabling a unified representation within a single field named the ULoS Signed Distance Field (ULoS SDF). This formulation supports consistent optimization across both regions and helps mitigate LoS-induced artifacts within the NLoS area. (2) We then propose a ULoS Rendering technique, which incorporates the vision-pretrained SDF to provide a stable initialization for training the ULoS SDF with RF signals. (3) To address the surface ambiguity problem and determine the correct zero-level set of the ULoS SDF, we introduce a second stage of training which aligns the vision-pretrained SDF and the RF-trained ULoS SDF ***outside*** the box, and prove that this alignment transfers to alignment ***inside*** the box. With these three components, Fig. 1(d) shows that GeRaF 2.0 achieves global surface reconstruction across both LoS and NLoS regions, producing a clean and accurate surface with the correct zero level of the SDF.

We evaluate GeRaF 2.0 using a 77 GHz mmWave radar mounted on a Franka Research 3 robotic arm, imaging a variety of real-world objects inside different boxes. Our results demonstrate that GeRaF 2.0 outperforms previous works and takes a significant step towards more robust and accurate 3D reconstruction from RF signals behind occlusions.

2. Related Work

Vision-Based Neural 3D Reconstruction: Neural Radiance Fields (NeRF) [30] and Gaussian Splatting [21] introduced using learnable parameters to reconstruct 3D scenes from multi-view images, which creates a neural implicit representation of the 3D scene. Motivated by this [8, 12, 16, 19, 24, 33, 46, 53, 54], more works further separates scene components into explicit geometric representations and reconstructs detailed surface meshes. [5, 14, 20, 28, 32, 36, 51, 52] take it a step further to predict lighting and material properties by learning complicated light interactions. They base the reconstruction on the explicit forward rendering equation adding in the reflectance distribution function. However, all of these 3D representations are based on optical signals, which are not easily translatable to radio frequency.

Further more there are works that do NLOS using optical sensors [26, 27, 34, 45], however, they tackle around-the corner imaging, requiring the signal to reflect off of surfaces, whereas our problem looks at wireless signals penetrating through surfaces.

3D Radar Reconstruction: Deep learning techniques have been used for imaging or point cloud completion in the context of self-driving cars. However, these works are geared towards street-level scenes and LoS (unoccluded) large objects like cars and pedestrians [15, 18, 23, 38–40].

mmNorm [10] performs non-line-of-sight surface reconstruction by estimating a normal field and optimizing over isosurfaces obtained by inverting the normal field. However, unlike our system, mmNorm focuses on one-sided 3D reconstruction (front view instead of 360°) and, similar to other works, it simply crops the occlusion out.

RF Neural Implicit Reconstruction: [4, 17, 22, 57] try to reconstruct self-driving car scenes by rendering a power distribution of the wireless signal and learning the occupancy of different locations in the scene. Another set of work [3, 42] apply neural implicit reconstruction to satellite images. However, all of these works perform reconstruction specifically tailored for reconstructing *large* scale scenes such as streets or satellite images and don't address close-range high-resolution object reconstruction, which requires different wireless propagation modeling as explained in the supplementary material.

For near range reconstruction, authors of [41], propose a method for 3D neural reconstruction of objects. However, their evaluation is limited to simulated data, which cannot represent the complexity of real world experiments with wireless signals. Most recently, GeRaF[29] propose a neural reconstruction method, tested on real world experiments, specifically for near-field objects, by using a physical rendering model of radio signals to learn a surface model. However, GeRaF avoids the need to model occlusions (eg. boxes, paper) by simply cropping the reflections from the occlusion out of the radar image, which introduces additional noise and degrades surface reconstruction.

Radar-Vision Joint Perception: A plethora of works have explored the benefits of combining radar and camera perception [2, 6, 11, 25, 48, 49, 55]. However, these papers are geared for self-driving car scenarios and bounding box detection, not reconstruction. Moreover, none of these works have used radar-camera fusion for neural implicit reconstruction for high-resolution 3D reconstruction. RadarSim [7] proposes a camera-radar joint reconstruction framework to create novel Doppler-range images from camera initialized 3D geometry for improvedd radar simulation. However, their goal is accurate radar synthesis from line-of-sight conditions, while our works tackles a different challenge: non-line-of-sight reconstruction from camera-radar joint observations.

3. Wireless Technical Background

3.1. Radio Frequency Background

Waveform A radar transmits a wireless waveform and receives reflections that come from the signal bouncing off of various objects in the environment. Our system uses Frequency Modulated Continuous Wave (FMCW) and antenna arrays to resolve range, azimuth and elevation ambiguity as a result of the lensless nature of wireless signals. The received signal is multiplied with the conjugate of the transmitted signal and is expressed as:

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

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, ν is the starting frequency, and k is the slope of the frequency change. For multiple reflectors in the scene, we receive the linear combination of Eq. 1.

Reflector Interaction Unlike light, whose short wavelength causes diffused reflections, RF signals have much longer wavelengths, making most surfaces appear smooth and produce primarily specular reflections [35]. In this paper we follow the reflection model as used in [29]. Given an input signal with amplitude A_{TX}, the received amplitude A_{RX} is expressed as:

$$A_{\text{RX}} \propto \frac{a}{(4\pi u)^2} A_{\text{TX}}\,(\omega_o \cdot \omega_r) \quad (2)$$

where a is the reflectivity, u is the propagation distance from the reflection point to the receiver, ω_r is the incoming vector of the RF signal, and ω_o is the outgoing vector.

3.2. Lensless Volumetric Rendering

Signal Tracing While vision-based rendering relies on optical lenses to filter and focus relevant light rays, wireless sensing operates through *lensless imaging*, capturing all incoming RF signals without directional filtering. The RF signal received at time t by an antenna located at $\mathbf{x}_{\text{ant}}$ can be expressed as:

$$s(\mathbf{x}_{\text{ant}}, t, u) = \sum_{\mathbf{x}\in\Omega_{\text{ULoS}}} A_{\text{rx}}(\mathbf{x})\, e^{-j2\pi k\tau_{\mathbf{x}} t}\, e^{-j2\pi\nu\tau_{\mathbf{x}}}, \quad (3)$$

where $\tau_{\mathbf{x}}$ denotes the propagation delay from point $\mathbf{x}$ to the antenna, and $A_{\text{rx}}(\mathbf{x})$ represents the received amplitude at $\mathbf{x} = \mathbf{x}_{\text{ant}} + \omega_r \cdot u$, which can be modeled using Eq. 2 as:

$$A_{\text{rx}}(\mathbf{x}_{\text{ant}}, \omega_r, u) = \frac{\mathbf{a}(u)}{(4\pi)^2}\,(\omega_o \cdot \omega_r)\,T(u)^2\,\rho(u)\,\mathbf{A}_{\text{tx}}\,\mathrm{d}t, \quad (4)$$

where $\mathbf{a}(u)$ is the reflectivity, ω_o is the outgoing vector computed as $\omega_o = \omega_i - 2(\mathbf{n}\cdot\omega_i)\mathbf{n}$, with ω_i as the incoming signal vector and $\mathbf{n}$ the surface normal. Here, $\mathbf{A}_{\text{tx}}$ denotes the equivalent transmitted power. The terms $T(u)$ and $\rho(u)$ are adopted from volumetric rendering in vision [46], where $T(u)$ represents accumulated transmittance and $\rho(u)$ denotes opacity. Since radar sensing involves two-way propagation, the transmittance term $T(u)$ is squared.

Lensless Sampling and Lensless Rendering Each antenna receives signals from all incoming directions, thus the lensless sampling strategy introduced in [29] is used to avoid exhaustive grid-based sampling for every antenna; the computation of opacity and accumulated transmittance is shared across antennas. Specifically, the quantities $\rho(u)$ and $T(u)$ only need to be computed once for each spatial position, as they are shared by all antenna rays that intersect the same voxel.

As shown in Fig. 3, lensless sampling begins by casting parallel rays aligned with the radar's primary direction ω_p from the antenna aperture. The opacity $\rho(u)$ and transmittance $T(u) = \exp\left(-\int_0^u \rho(v)dv\right)$ is computed along the primary ray using traditional rendering [46]. The true transmittance $T(u')$ along a real ray (orange ray in Fig. 3) pointing toward the antenna is avoided from being recomputed. Instead, the sigmoid-SDF values are adjusted at the scene boundary $\partial\Omega_{\text{ULoS}}$ (green and orange points in Fig. 3):

$$T(u') = T(u) - \Phi_s(f(\mathbf{x}(u_s))) + \Phi_s(f(\mathbf{x}(u'_s))), \quad (5)$$

where $\Phi_s(\cdot)$ denotes the sigmoid function, and $\mathbf{x}(u_s)$ and $\mathbf{x}(u'_s)$ represent the starting points of the primary and secondary rays.

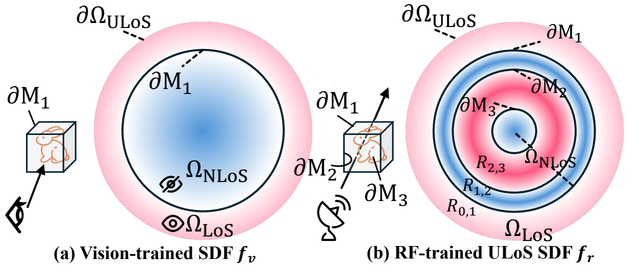


Figure 2. **a)** Vision-trained SDF uses negative values (blue) and positive values (red) to separate the inside and outside. **b)** RF-trained ULoS SDF models the scene as a series of nested closed and compact sets. Areas with a strong radio frequency interaction ((e.g., carton, metal) (blue) are assigned negative values, weak interactions ((e.g., air) regions (red) are assigned positive values.

4. Rendering from Radio Frequency Signals

4.1. ULoS Scene Representation

As shown in Fig. 2(a), according to visibility conditions, the entire scene can be split into a line-of-sight (LoS) region and a non-line-of-sight (NLoS) region, denoted by two closed and compact sets: $\Omega_{\text{LoS}} \subset \mathbb{R}^3$ and $\Omega_{\text{NLoS}} \subset \mathbb{R}^3$. The union of these two regions defines the Unified Line-of-Sight and Non-Line-of-Sight (ULoS) domain:

$$\Omega_{\text{ULoS}} = \Omega_{\text{LoS}} \cup \Omega_{\text{NLoS}}. \tag{6}$$

Fig. 2(b) illustrates the ULoS scene. We represent the scene as a series of nested, closed, and compact sets in three-dimensional Euclidean space. Let $M_1, \dots, M_n \subset \mathbb{R}^3$ be such that

$$M_n \subset \text{int}(M_{n-1}) \subset \dots \subset \text{int}(M_1) \subset \text{int}(\Omega_{\text{ULoS}}), \tag{7}$$

where $\text{int}(\cdot)$ denotes the interior of a set.

The spatial region between two consecutive layers M_i and M_{i+1} is defined as

$$R_{i,i+1} = \left\{ \mathbf{x} \in \mathbb{R}^3 \,\middle|\, \mathbf{x} \in \text{int}(M_i),\, \mathbf{x} \notin \text{int}(M_{i+1}) \right\}, \tag{8}$$

The outermost region is defined by

$$R_{0,1} = \{ \mathbf{x} \in \mathbb{R}^3 \mid \mathbf{x} \in \Omega_{\text{ULoS}},\, \mathbf{x} \notin \text{int}(M_1) \}.$$

The boundary of each intermediate region is the union of its two layer surfaces:

$$\partial R_{i,i+1} = \partial M_i \cup \partial M_{i+1}. \tag{9}$$

The observer is set outside the outermost domain Ω_{ULoS}. From Eq. (7), the outermost set M_1 fully contains all inner subsets. Therefore, its boundary ∂M_1 blocks visible light and defines the limit of optical visibility. Consequently, the line-of-sight region is given by $\Omega_{\text{LoS}} = R_{0,1}$, while the non-line-of-sight region corresponds to all nested layers within the box: $\Omega_{\text{NLoS}} = \bigcup_{i=1}^{n} M_i = M_1$.

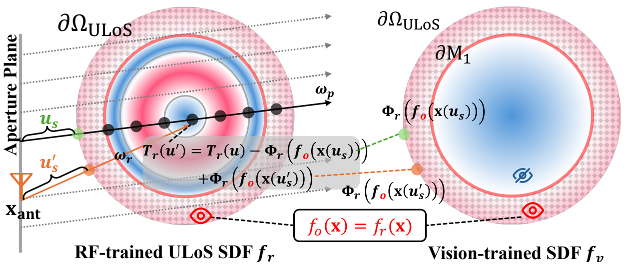


Figure 3. Lensless sampling and ULoS lensless rendering. Gray rays represent the primary sampling direction and orange ray indicates the actual ray pointing from the antenna. ULoS lensless rendering: we exploit the fact that vision-trained SDF and RF-trained ULoS SDF share identical values in the LoS region (shaded), allowing us to to adjust the accumulated transmittance.

ULoS SDF Vision-based neural geometry reconstruction methods [31, 46] represent scenes using a Signed Distance Function (SDF) f_v, which clearly distinguishes between the inside, outside, and surface of objects. However, RF signals interact with the environment in fundamentally different ways. The traditional binary notion of "inside" and "outside" becomes ambiguous, especially across multiple layers, as NLoS regions can still be partially transparent to RF signals. To address this, we introduce a unified Signed Distance Function tailored for both LoS and NLoS RF propagation, denoted as the *ULoS SDF*. We denote this RF-specific distance field as f_r throughout our formulation.

We define the "negative" (or interior) region for a given medium as the set of points that impose a strong radio frequency interaction against the propagating field, and the "positive" (or exterior) region as the set of points with a weak radio frequency interaction. For example, Fig. 2(b) shows a simple case of an object inside a box. The scene contains three surfaces: the outer surface of the box, the inner surface of the box, and the surface of the object. The region between the box surfaces and the interior of the object is assigned negative values, while all other regions are positive.

Formally, the sign of the ULoS SDF is determined by the relative interference coefficient:

$$\text{sign}(f(\mathbf{x})) = \begin{cases} -1, & \text{if } R_{i,i+1} \text{ is strong interaction,} \\ +1, & \text{if } R_{i,i+1} \text{ is weak interaction.} \end{cases}$$

To maintain continuity and geometric consistency across multi-layered boundaries, the signed distance for each region $R_{i,i+1}$ is defined as the *minimum Euclidean distance to its nearest bounding surfaces*:

$$f(\mathbf{x}) = \begin{cases} \text{sign}(f(\mathbf{x})) \min\big(d(\mathbf{x}, \partial M_i), & \\ \qquad d(\mathbf{x}, \partial M_{i+1})\big), & \mathbf{x} \in \text{int}(R_{i,i+1}), \\ 0, & \mathbf{x} \in \partial R_{i,i+1}, \end{cases} \tag{10}$$

where $d(\mathbf{x}, \partial M_i)$ is the Euclidean distance from $\mathbf{x}$ to the surface ∂M_i. This unified formulation preserves geometric continuity while maintaining physically meaningful sign semantics for each modality.

4.2. ULoS Lensless Rendering

It is important to note that the terms $\Phi_s(f_{\mathrm{NLoS}}(\mathbf{x}(u_s)))$ and $\Phi_s(f_{\mathrm{NLoS}}(\mathbf{x}(u'_s)))$ in Eq. 5 are excluded from the backpropagation process due to their high computational cost. This omission, however, introduces errors in gradient propagation, leading to suboptimal optimization. Moreover, these terms exhibit a significant bias during network initialization, resulting in skewed transmittance adjustment in the early stages of training.

This problem is particularly challenging in NLoS scenes, where limited signal visibility makes stable optimization difficult. However, in the ULoS setting, additional geometric information from the LoS domain can play a crucial role. As illustrated in the shaded region of Fig. 3, the *vision-trained SDF and the RF-trained ULoS SDF share identical values outside the box region* $R_{0,1}$:

$$f_v(\mathbf{x}) = f_r(\mathbf{x}), \qquad \mathbf{x} \in R_{0,1}. \tag{11}$$

Since the starting points of the primary rays lie in free space outside the enclosing box, the SDF and the ULoS SDF are equivalent at these locations. By initializing the ULoS SDF at these starting points using a well-converged, vision-pretrained neural reconstruction, we provide a stable prior for training with RF signals, leading to faster convergence and improved consistency.

4.3. Overall Pipeline

The overall pipeline is illustrated in Fig. 4. We perform reconstruction in the following way: **(1)** We train the SDF of the exterior of LoS surface using NeuS [46]. After training, the vision model is fixed for the remainder of the pipeline. **(2)** Then we move onto RF reconstruction, and begin with lensless sampling strategy to generate point samples in space. **(3)** Each sampled point is processed by three sub-networks: the SDF Network, the Reflectivity Network, and the Signal Power Prediction Network. The SDF Network is initialized with the vision-pretrained SDF and is used to predict the ULoS SDF, which is used to compute opacity and transmittance. The Reflectivity and Signal Power Networks estimate surface reflectance and received signal power, respectively. **(4)** From the outputs of the sub-networks, the ULoS Lenless Rendering module (Sec. 4.2) simulates the received antenna signal. Within this module, the vision-pretrained SDF is used to adjust the transmittance rather than relying on the model under training. **(5)** Finally, a differentiable matched filter is applied to render the matched filter heatmap, and the loss is computed between the rendered and ground-truth heatmaps.

However, after this pipeline, the reconstructed surface still suffers from inaccuracies and an incorrect zero-level set of the SDF. This issue, known as the surface ambiguity problem [10], arises from the inherent difficulty of normalizing radar signal strength. To address this, we introduce a second-stage training process, described in detail in Sec. 5.

5. The Surface Ambiguity Problem

The surface ambiguity problem [10] arises from the inherent difficulty of normalizing the radar signal strength. Unlike RGB images in LoS scenes, where light intensity is pre-normalized to the range $[0, 1]$, radar signals cannot be normalized in the same way due to their strong dependence on scene geometry, material properties, and signal attenuation. As a result, the reflectivity, predicted signal power, and the zero-level surface of the SDF become mutually entangled variables. Without additional constraints or external information, the network cannot uniquely determine the correct surface configuration, leading to an incorrect zero-level SDF definition and inaccurate surface geometry.

5.1. Relative Signed Distance Function

To facilitate analysis, we adopt the *relative signed distance function* (RSDF) [10], denoted by $g(\mathbf{x})$, which is defined as the SDF offset by some unknown constant. The gradient of the RSDF is identical to that of the conventional signed distance function (SDF) $f(\mathbf{x})$:

$$\nabla f(\mathbf{x}) = \nabla g(\mathbf{x}), \qquad \forall \mathbf{x} \in \Omega. \tag{12}$$

Our objective is for the RSDF to converge to the true SDF, i.e.,

$$f(\mathbf{x}) = g(\mathbf{x}), \qquad \forall \mathbf{x} \in \Omega.$$

To formalize this equivalence, we present the following proposition.

Proposition. Let $f, g : \Omega \to \mathbb{R}$ be two continuously differentiable scalar fields defined on a connected region $\Omega \subset \mathbb{R}^3$. If the following two conditions hold:

1. $\nabla f(\mathbf{x}) = \nabla g(\mathbf{x})$ for all $\mathbf{x} \in \Omega$, and
2. $f(\mathbf{x}) = g(\mathbf{x})$ for all $\mathbf{x}$ on a closed surface $S \subset \Omega$,

then

$$f(\mathbf{x}) \equiv g(\mathbf{x}), \qquad \forall \mathbf{x} \in \Omega.$$

Proof. Define $h = f - g$. Since $\nabla f = \nabla g$, it follows that $\nabla h = \mathbf{0}$ for all $\mathbf{x} \in \Omega$. A differentiable scalar field with zero gradient on a connected domain must be constant; hence $h = C$ for some $C \in \mathbb{R}$. From condition (2), $f(\mathbf{x}) = g(\mathbf{x})$ for all $\mathbf{x} \in S$, implying $C = 0$. Therefore, $h(\mathbf{x}) = 0$ for all $\mathbf{x} \in \Omega$, and thus $f \equiv g$ throughout Ω.

Condition (1) is directly satisfied by the definition of the RSDF. Therefore, it remains to ensure that the RSDF derived from radar signals, denoted by $g_r(\mathbf{x})$, matches the corresponding SDF, $f_r(\mathbf{x})$, on a closed reference surface $S \subset \Omega_{\mathrm{ULoS}}$.

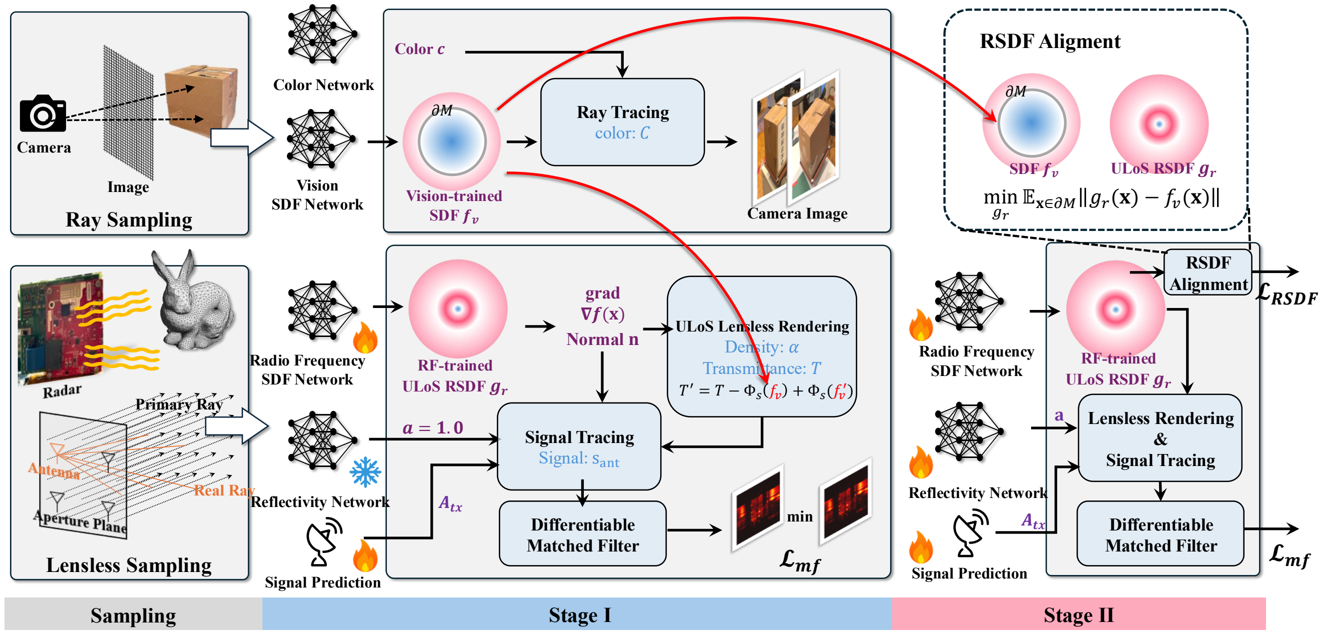


Figure 4. Overall pipeline of GeRaF 2.0. ***Top:*** The vision-pretrained SDF on the outside of the box. ***Bottom:*** The training pipeline for RF signals. The pipeline begins with lensless sampling. In the first stage of training, we freeze the Reflectivity Network and use the vision-pretrained SDF to adjust transmittance in the ULoS Lensless Rendering module. In the second stage, we use the vision-pretrained SDF to align the relative SDF, thereby addressing the surface ambiguity problem.

Observation. Within the LoS region $R_{0,1}$, the RF-trained SDF and vision-trained SDF coincide, i.e., $f_r(\mathbf{x}) = f_v(\mathbf{x})$.

5.2. Relative Signed Distance Function Alignment

Our objective is to enforce equality between the RF-trained RSDF and the vision-trained SDF on a selected closed surface $S_{\text{LoS}} \subset R_{0,1}$, such that

$$g_r(\mathbf{x}) = f_v(\mathbf{x}), \qquad \forall \mathbf{x} \in S_{\text{LoS}}.$$

As shown in Fig. 5, for implementation convenience, we choose $S_{\text{LoS}} = \partial M_1$, that is, the outer surface of the box, as the reference surface for alignment. The corresponding optimization objective is defined as

$$\mathcal{L}_{\text{RSDF}} = \min_{g_r} \mathbb{E}_{\mathbf{x}\in\partial M_1}\left[\left|g_r(\mathbf{x}) - f_v(\mathbf{x})\right|\right]. \tag{13}$$

However, in practice, directly sampling on the surface ∂M_1 and supervising the SDF values can be difficult and may lead to unstable training. Since the alignment only requires consistency of the reconstructed surfaces, a more stable alternative is to supervise the *depth* along primary rays.

Primary Ray Depth. Let $\mathbf{x}_{\text{ant}}$ denote the antenna position and ω_p the ray direction. The expected depth along the primary ray for vision and RF modalities is expressed as

$$d(\mathbf{x}_{\text{ant}}, \omega_p) = \int_0^\infty u\,\rho(u)\,T(u)\,du \tag{14}$$

where ρ_v and T_v are derived from the vision-trained SDF f_v, and ρ_r and T_r are derived from the RF-trained RSDF g_r. The final RSDF alignment loss is defined as

$$\mathcal{L}_{\text{RSDF}} = \mathbb{E}_{\mathbf{x}_{\text{ant}}}\left[\left|d_v(\mathbf{x}_{\text{ant}}, \omega_p) - d_r(\mathbf{x}_{\text{ant}}, \omega_p)\right|\right]. \tag{15}$$

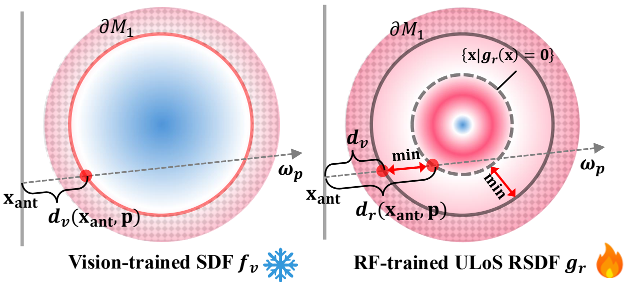


Figure 5. Illustration of RSDF alignment. In the LoS region (shaded area), the vision-pretrained SDF coincides with the RF-trained ULoS RSDF. Therefore, the target of RSDF alignment becomes aligning the outer surface of the box, which can further be reduced to aligning the depth along the primary ray.

5.3. Optimization Target

Directly training with RSDF alignment can make it difficult to reconstruct the NLoS geometry due to optimization instability. To address this, we disentangle the optimization process into two stages. In **Stage 1**, as illustrated in Fig. 4, we freeze the Reflectivity Network and initialize its output to a constant 1.0 across all spatial positions. We adopt the training objectives proposed in GeRaF [29], employing a Matched Filter (MF) on the predicted signals. MF effec-

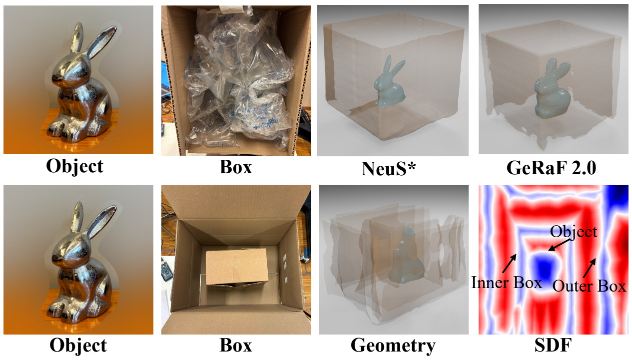


Figure 6. ***Top***: Reconstruction results for a box filled with bubble wrap.*For NeuS, the object and box are captured separately visualized together via post-processing. ***Bottom***: Reconstruction of the nested box object. `Geometry` denotes GeRaF 2.0's reconstructed 3D mesh, while `SDF` illustrates the 2D SDF values for a horizontal slice.

tively suppresses noise and irrelevant signal components while preserving the most informative features for geometric reconstruction. The resulting optimization objective is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{MF}} + \lambda_{\text{GRAD}} \mathcal{L}_{\text{GRAD}}, \tag{16}$$

where $\mathcal{L}_{\text{MF}}$ computes the loss between the predicted and ground-truth matched-filter power distributions, and $\mathcal{L}_{\text{GRAD}}$ denotes the Eikonal regularization term [13] used to enforce a valid signed distance field.

In **Stage 2**, we train all networks jointly and apply RSDF alignment. Thanks to the stable initialization from Stage 1, there is no need to use ULoS Lensless Rendering in this stage. The overall training objective combines the matched-filter loss, RSDF alignment loss, and gradient regularization:

$$\mathcal{L} = \mathcal{L}_{\text{MF}} + \lambda_{\text{GRAD}}\, \mathcal{L}_{\text{GRAD}} + \lambda_{\text{RSDF}}\, \mathcal{L}_{\text{RSDF}}. \tag{17}$$

The RSDF alignment loss $\mathcal{L}_{\text{RSDF}}$ is defined in Eq. (15).

6. Assessment

6.1. Experiment Setup

Dataset We train and evaluate our system on a multi-view radar and camera image dataset. Data was captured with a Franka Research 3 equipped with TI's AWR1843BOOST evaluation board [43].The object was placed on a 360° rotation plate, with 10° of rotation between each radar image. Ground truth comparison of the NLoS object for quantitative results was collected with Scaniverse [44].

Training Details Our model was trained using a NVIDIA H100 GPU for 100,000 iterations over 48 hours. More details will be included on the experimental setup and training parameters in the supplementary material.

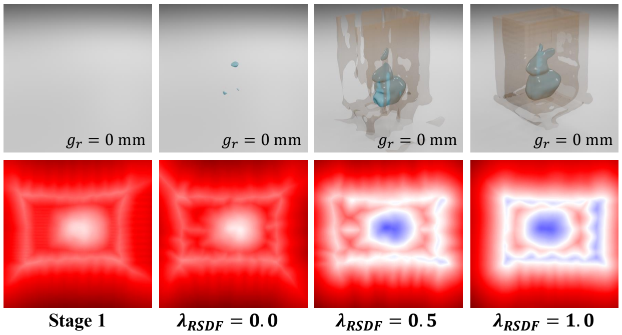


Figure 7. Ablation study on different values of λ_{RSDF}. The first column shows Stage 1 results. The top row visualizes surfaces at the zero-level set of the SDF, while the bottom row shows a cross-sectional slice of the SDF.

6.2. Results

We compare GeRaF 2.0 with three baselines: vision-based NeuS [46], Matched Filter (MF) imaging, and NLoS reconstruction GeRaF [29]. For NeuS, we train the object and box separately due to occlusion. For the Matched Filter, we sum outputs across views, threshold the heatmap, and apply Poisson surface reconstruction. We show results for both Stage 1 and Stage 2 of GeRaF 2.0. Quantitative results are calculated using the F1-Score($\tau = 0.015$) and Chamfer Distance (in millimeters), the box is cropped out of the point clouds for evaluation; the method is described in detail in the supplementary material.

Baseline Comparisons Qualitative results are shown in Fig. 8 and both stages of GeRaF 2.0 achieve the best RF reconstruction results, outperforming the Matched Filter and GeRaF [29], and closely matching the vision-based reconstruction. Unlike GeRaF, which often shows artifacts or fails to capture surfaces correctly, GeRaF 2.0 produces clean and accurate geometry due to the ULoS representation and ULoS rendering. Notably, Stage 2 benefits from RSDF alignment, enabling the surface to be extracted precisely at the SDF zero-level, something all baselines fail to do. This not only removes the need for manual thresholding but preserves fine details such as the elephant's tusks, the chicken's comb, the deer's antlers, and the ball's top.

Robustness to Fillers and Multilayer Occlusions. The first row of Fig. 6 illustrates the reconstruction of the target object when the container is filled with bubble wrap. Despite the introduction of additional scattering layers, the model maintains high-fidelity reconstruction with negligible degradation in geometric detail. Furthermore, the second row of Fig. 6 evaluates the system's performance in a nested-box configuration. The corresponding SDF visualization clearly resolves the zero-level sets for both the concentric boxes and the internal object albeit some degradation of the internal surface due to increased scattering.

Ablation Study on RSDF Alignment In Fig. 7, we show

Figure 8. Qualitative results between vision-based NeuS [46], point cloud-based reconstruction using matched filter, non-line-of-sight reconstruction with GeRaF [29], and GeRaF 2.0 in Stage 1 and Stage 2. *For NeuS, the object and box are captured separately visualized together via post-processing. For GeRaF and Stage 1, the surface level g_r is manually selected (indicated in the visualization).

the impact of RSDF alignment during Stage 2. All visualizations display surfaces extracted at the zero level set of the SDF. Without alignment (i.e., $\lambda_{\mathrm{RSDF}} = 0$), the reconstructed surface is significantly offset from the zero level. As we increase λ_{RSDF}, the surface gradually converges toward the correct zero level, demonstrating the effectiveness of RSDF alignment in resolving surface ambiguity.

7. Conclusion

In this paper, we present a unified neural reconstruction framework that bridges LoS and NLoS regions for 3D reconstruction. By incorporating LoS geometry, as a physical prior into the neural field formulation, our method establishes a consistent link between visible and hidden surfaces, stabilizing optimization and improving reconstruction.

Acknowledgments

We thank the anonymous CVPR reviewers and members of the SENS Lab for their valuable feedback. We would also like to thank Ralf Boehnke, Dymtro Rachkov and Daniel Ardila Palomino from Sony Research for their helpful feedback. This project is funded in part by the Sony Faculty Innovation Fellowship.

References

[1] Fadel Adib and Dina Katabi. See through walls with wifi! In *SIGCOMM*, 2013. 1

[2] Muhammad Kashif Ali, Asif Rajput, Muhammad Shahzad, Farhan Khan, Faheem Akhtar, and Anko Börner. Multi-sensor depth fusion framework for real-time 3d reconstruction. *Ieee Access*, 7:136471–136480, 2019. 3

[3] Emile Barbier-Renard, Florence Tupin, Nicolas Trouvé, and Loïc Denis. Multi-view 3d surface reconstruction from sar images by inverse rendering, 2025. 2

[4] David Borts, Erich Liang, Tim Broedermann, Andrea Ramazzina, Stefanie Walz, Edoardo Palladin, Jipeng Sun, David Brueggemann, Christos Sakaridis, Luc Van Gool, et al. Radar fields: Frequency-space neural scene representations for fmcw radar. In *SIGGRAPH*, 2024. 1, 2

[5] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. Nerd: Neural reflectance decomposition from image collections. In *ICCV*, 2021. 2

[6] Anjun Chen, Xiangyu Wang, Kun Shi, Yuchi Huo, Jiming Chen, and Qi Ye. Towards weather-robust 3d human body reconstruction: Millimeter-wave radar-based dataset, benchmark, and multi-modal fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024. 3

[7] Chuhan Chen, Tianshu Huang, Akarsh Prabhakara, Chaithanya Kumar Mummadi, Zhongxiao Cong, Anthony Rowe, Matthew O'Toole, and Deva Ramanan. Radarsim: Simulating single-chip radar via multimodal neural fields. In *Thirteenth International Conference on 3D Vision*. 3

[8] Hanlin Chen, Chen Li, and Gim Hee Lee. Neusg: Neural implicit surface reconstruction with 3d gaussian splatting guidance. *arXiv preprint*, 2023. 2

[9] Ashutosh Dhekne, Mahanth Gowda, Yixuan Zhao, Haitham Hassanieh, and Romit Roy Choudhury. Liquid: A wireless liquid identifier. In *Proceedings of the 16th annual international conference on mobile systems, applications, and services*, pages 442–454, 2018. 2

[10] Laura Dodds, Tara Boroushaki, Kaichen Zhou, and Fadel Adib. Non-line-of-sight 3d object reconstruction via mmwave surface normal estimation. In *MobiCom*, 2025. 2, 5

[11] Ghina El Natour, Omar Ait-Aider, Raphael Rouveure, François Berry, and Patrice Faure. Toward 3d reconstruction of outdoor scenes using an mmw radar and a monocular vision sensor. *Sensors*, 15(10):25937–25967, 2015. 3

[12] Jian Gao, Chun Gu, Youtian Lin, Zhihao Li, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing. In *ECCV*, 2024. 2

[13] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099*, 2020. 7

[14] Chun Gu, Xiaofei Wei, Zixuan Zeng, Yuxuan Yao, and Li Zhang. Irgs: Inter-reflective gaussian splatting with 2d gaussian ray tracing. In *CVPR*, 2025. 2

[15] Junfeng Guan, Sohrab Madani, Suraj Jog, Saurabh Gupta, and Haitham Hassanieh. Through fog high-resolution imaging using millimeter wave radar. In *CVPR*, 2020. 2

[16] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *SIGGRAPH*, 2024. 2

[17] Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana Laroia, Zico Kolter, and Anthony Rowe. Dart: Implicit doppler tomography for radar novel view synthesis. In *CVPR*, 2024. 1, 2

[18] Samah Hussein, Junfeng Guan, Swathi Narashiman, Saurabh Gupta, and Haitham Hassanieh. 3d object reconstruction with mmwave radars. *arXiv preprint arXiv:2504.12348*, 2025. 2

[19] Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In *CVPR*, 2024. 2

[20] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensoir: Tensorial inverse rendering. In *CVPR*, 2023. 2

[21] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (TOG)*, 2023. 2

[22] Pou-Chun Kung, Skanda Harisha, Ram Vasudevan, Aline Eid, and Katherine A Skinner. Radarsplat: Radar gaussian splatting for high-fidelity data synthesis and 3d reconstruction of autonomous driving scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 27596–27606, 2025. 1, 2

[23] Haowen Lai, Gaoxiang Luo, Yifei Liu, and Mingmin Zhao. Enabling visual recognition at radio frequency. In *MobiCom*, 2024. 2

[24] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering. In *CVPR*, 2024. 2

[25] Zhiwei Lin, Zhe Liu, Zhongyu Xia, Xinhao Wang, Yongtao Wang, Shengxiang Qi, Yang Dong, Nan Dong, Le Zhang, and Ce Zhu. Rcbevdet: Radar-camera fusion in bird's eye view for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14928–14937, 2024. 3

[26] David B Lindell, Gordon Wetzstein, and Matthew O'Toole. Wave-based non-line-of-sight imaging using fast fk migration. *ACM Transactions on Graphics (ToG)*, 38(4):1–13, 2019. 2

[27] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*, 572(7771): 620–623, 2019. 2

[28] Yuan Liu, Peng Wang, Cheng Lin, Xiaoxiao Long, Jiepeng Wang, Lingjie Liu, Taku Komura, and Wenping Wang. Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images. *ACM Transactions on Graphics (TOG)*, 2023. 2

[29] Jiachen Lu, Hailan Shanbhag, and Haitham Hassanieh. Geraf: Neural geometry reconstruction from radio frequency signals. In *NeurIPS*, 2025. 1, 2, 3, 6, 7, 8

[30] 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. 2

[31] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (TOG)*, 2022. 4

[32] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. Extracting triangular 3d models, materials, and lighting from images. In *CVPR*, 2022. 2

[33] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *CVPR*, 2020. 2

[34] Matthew O'Toole, David B Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338–341, 2018. 2

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

[36] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *CVPR*, 2021. 2

[37] William C Stone. Electromagnetic signal attenuation in construction materials. 1997. 2

[38] Yue Sun, Zhuoming Huang, Honggang Zhang, Zhi Cao, and Deqiang Xu. 3drimr: 3d reconstruction and imaging via mmwave radar based on deep learning. In *IEEE International Performance, Computing, and Communications Conference (IPCCC)*, 2021. 2

[39] Yue Sun, Zhuoming Huang, Honggang Zhang, and Xiaohui Liang. 3d reconstruction of multiple objects by mmwave radar on uav. In *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, 2022.

[40] Yue Sun, Honggang Zhang, Zhuoming Huang, and Benyuan Liu. R2p: A deep learning model from mmwave radar to point cloud. In *International Conference on Artificial Neural Networks*, 2022. 2

[41] Harshvardhan Takawale and Nirupam Roy. Spinr: Neural volumetric reconstruction for fmcw radars. *arXiv preprint*, 2025. 2

[42] Weiyi Tan, Yu Wang, Biao Tian, and Shiyou Xu. Fast 3d reconstruction of space targets from isar image sequences based on neural network. In *IEEE 8th International Conference on Vision, Image and Signal Processing (ICVISP)*, 2024. 2

[43] TI Inc. Texas instrument awr1843. `https://www.ti.com/product/AWR1843`, 2023. 7

[44] Toolbox AI, Inc. and Niantic, Inc. Scaniverse: 3d scanner app, 2025. Mobile application for iOS. 7

[45] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3 (1):745, 2012. 2

[46] 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. 2, 3, 4, 5, 7, 8

[47] Ting Wu, Theodore S Rappaport, and Christopher M Collins. Safe for generations to come: Considerations of safety for millimeter waves in wireless communications. *IEEE microwave magazine*, 2015. 1

[48] Zizhang Wu, Guilian Chen, Yuanzhu Gan, Lei Wang, and Jian Pu. Mvfusion: Multi-view 3d object detection with semantic-aligned radar and camera fusion. *arXiv preprint arXiv:2302.10511*, 2023. 3

[49] Sheng Yang, Tong Zhan, Shichen Qiao, Jicheng Gong, Qing Yang, Jian Wang, and Yanfeng Lu. Zfusion: An effective fuser of camera and 4d radar for 3d object perception in autonomous driving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 3768–3777, 2025. 3

[50] Muhammet Emin Yanik and Murat Torlak. Near-field 2-d sar imaging by millimeter-wave radar for concealed item detection. In *2019 IEEE radio and Wireless Symposium (RWS)*, 2019. 1

[51] Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Neilf: Neural incident light field for physically-based material estimation. In *ECCV*, 2022. 2

[52] Yuxuan Yao, Zixuan Zeng, Chun Gu, Xiatian Zhu, and Li Zhang. Reflective gaussian splatting. In *ICLR*, 2025. 2

[53] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. In *NeurIPS*, 2020. 2

[54] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In *NeurIPS*, 2021. 2

[55] Zedong Yu, Weibing Wan, Maiyu Ren, Xiuyuan Zheng, and Zhijun Fang. Sparsefusion3d: Sparse sensor fusion for 3d object detection by radar and camera in environmental perception. *IEEE Transactions on Intelligent Vehicles*, 9(1): 1524–1536, 2023. 3

[56] Shichao Yue, Yuzhe Yang, Hao Wang, Hariharan Rahul, and Dina Katabi. Bodycompass: Monitoring sleep posture with wireless signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2):1–25, 2020. 1

[57] Jiarui Zhang, Zhihao Li, Chong Wang, and Bihan Wen. Rf4d: Neural radar fields for novel view synthesis in outdoor dynamic scenes. *arXiv preprint arXiv:2505.20967*, 2025. 1, 2