Radar Fields:
Frequency-Space Neural Scene Representations for FMCW Radar

David Borts
dborts@princeton.edu
Princeton University

Erich Liang
erliang@princeton.edu
Princeton University

Tim Brödermann
timbr@vision.ee.ethz.ch
ETH Zürich

Andrea Ramazzina
andrea.ramazzina@daimler.com
Mercedes-Benz

Stefanie Walz
sefanie.walz@daimler.com
Mercedes-Benz

Edoardo Palladin
edoardo.palladin@torc.ai
Torc Robotics

Jipeng Sun
js2694@princeton.edu
Princeton University

David Bruggemann
brdavid@vision.ee.ethz.ch
ETH Zürich

Christos Sakaridis
csakarid@vision.ee.ethz.ch
ETH Zürich

Luc Van Gool
vangool@vision.ee.ethz.ch
ETH Zürich

Mario Bijelic
mario.bijelic@princeton.edu
Princeton University,
Torc Robotics

Felix Heide
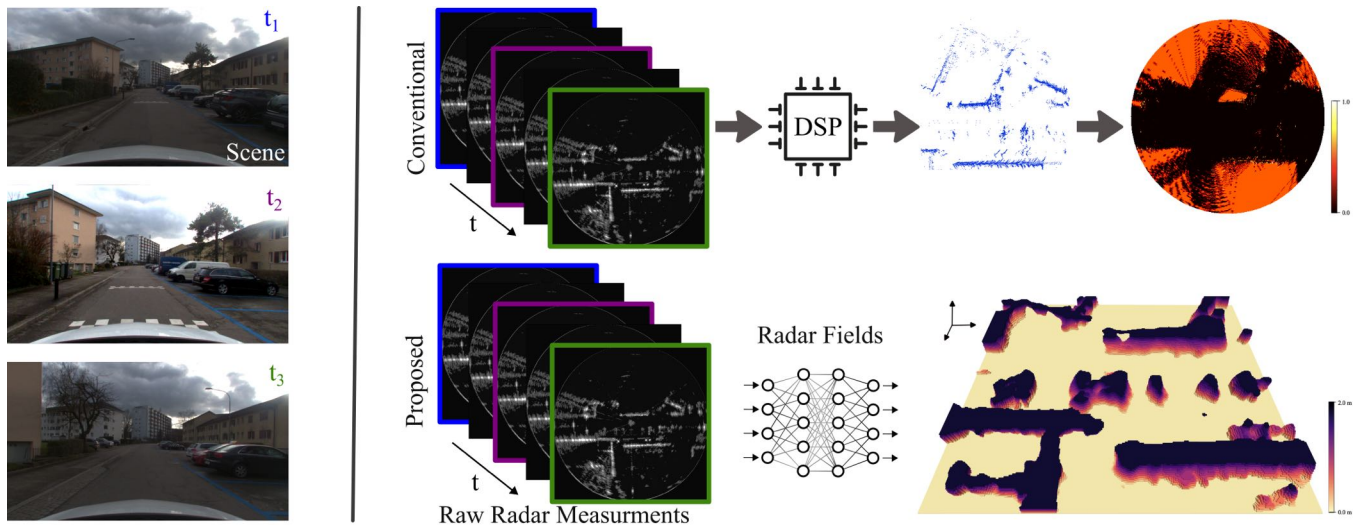fheide@princeton.edu
Princeton University,
Torc Robotics

Figure 1: Consider an urban scene with complex geometry, including vehicles and infrastructure. Existing methods threshold radar FFT measurements using a DSP pipeline, producing a sparse 2D disc of points that is fed into grid mapping algorithms to get bird's eye view drivable space. Radar Fields, in contrast, is a 3D neural scene representation for radar. Operating on raw frequency-domain measurements instead of processed points, Radar Fields recovers dense 3D occupancy from 2D radar scans.

ABSTRACT

Neural fields have been broadly investigated as scene representations for the reproduction and novel generation of diverse outdoor scenes, including those autonomous vehicles and robots must handle. While successful approaches for RGB and LiDAR data exist, neural reconstruction methods for radar as a sensing modality have been largely unexplored. Operating at millimeter wavelengths, radar sensors are robust to scattering in fog and rain, and, as such, offer a complementary modality to active and passive optical sensing techniques. Moreover, existing radar sensors are highly cost-effective and deployed broadly in robots and vehicles that operate outdoors. We introduce Radar Fields — a neural scene reconstruction method designed for active radar imagers. Our approach unites an explicit, physics-informed sensor model with an implicit neural geometry and reflectance model to directly synthesize raw radar measurements and extract scene occupancy. The proposed method *does not rely on volume rendering*. Instead, we learn fields in Fourier frequency space, supervised with raw radar data. We validate our method's effectiveness across diverse outdoor scenarios, including urban scenes with dense vehicles and infrastructure, and harsh weather scenarios, where mm-wavelength sensing is favorable.

CCS CONCEPTS

• **Computing methodologies** → *3D imaging*; **Vision for robotics**; **Computational photography**.

1 INTRODUCTION

Accurate reconstruction of large-scale outdoor scenes is essential for the development and validation of self-driving robots and drones. These lay the foundation for in-depth scene understanding, reliable navigation, and meticulous dataset generation and validation. While the large body of existing scene reconstruction methods relies on RGB and LiDAR sensors [Guo et al. 2023; Huang et al. 2023; Tao et al. 2023; Wang et al. 2023b; Zhang et al. 2023a], the potential of radar in this arena remains largely untapped. Typical radar sensors operate at around 77GHz - a modality complementary in both its exceptional ability to detect metallic objects and its resilience against adverse weather [Bijelic et al. 2020; Hwang et al. 2022], e.g. for snow or fog particle sizes which fall below the mm-wavelength. This makes radar sensing essential in cluttered urban settings, and in rain, fog, or snow, where optical modalities falter.

These capabilities have been instrumental in the adoption of radar sensors across multiple fields, including automotive [Caesar et al. 2020; Dickmann et al. 2014], aerospace [Pavlikov et al. 2021], robotics [Stetco et al. 2020], non-invasive medical imaging [Vilesov et al. 2022], and human-machine interfaces [Hayashi et al. 2021; Lien et al. 2016]. As an active sensor at lower cost than LiDAR – an order of magnitude for automotive systems – radar has emerged as a cornerstone for safety and efficiency in driver-assistance systems [Caesar et al. 2020; Dickmann et al. 2014], with ever-growing adoption rates [Burkacky et al. 2023] predicted to match 30 million units sold by 2030. Similarly, for robotics, the reliability and accuracy of radar data make it a vital sensor modality, both in indoor [Lee et al. 2021; Lu et al. 2020; Stetco et al. 2020] and outdoor [El Natour et al. 2015; Reina et al. 2011] settings.

Recent advances in neural rendering and scene reconstruction, epitomized by neural rendering methods [Mildenhall et al. 2020; Wang et al. 2021], have relied primarily on RGB camera images. As such, existing neural rendering methods can fail in challenging light conditions [Mildenhall et al. 2022; Wang et al. 2023c], or in the presence of challenging environmental factors, like fog [Levy et al. 2023; Ramazzina et al. 2023], as cameras fail to accurately capture the scene. Radar data, due to its inherent resiliency, offers a promising avenue for more versatile approaches to scene reconstruction, ensuring accuracy and reliability even in harsh weather conditions.

The potential of radar in 3D scene reconstruction is mostly unexplored. This is at least partly because recovering dense geometry from radar scans is not a straightforward task. Due to the relatively large wavelengths used – three orders of magnitude larger than LiDAR wavelengths – radar angular resolution is lower and point targets are drastically sparser than with optical sensors. Moreover, many common radar sensors are 2-dimensional, with no elevation resolution. Instead, as the beam diverges, signals from objects at different elevations are mixed into a single return at each range. This leads to a flattened 2D scan of the environment and makes it challenging to recover a 3D scene representation without a radar-specific signal formation model. Radar is also unique in its sensitivity to reflective objects, as angle-dependent sensor saturation artifacts are difficult to disentangle from occupancy. Altogether, these challenges render existing methods and their forward models ineffective for reconstructing scenes from radar data.

In this work, we propose a neural reconstruction method for raw radar data. Our approach bypasses both the resolution limitations of processed radar data and the computational cost of volume rendering by modeling in *frequency space*. We supervise our model with raw radar waveform data and recover relationships between detected power, scene geometry, distance from the sensor and scene reflectivity. This allows our method to access the high radial resolution of the correlated waveform, which is otherwise unavailable to existing methods that operate on post-processed range estimates.

To learn in frequency space, we introduce a Frequency-modulated continuous-wave (FMCW) radar signal formation model, which draws from the physics of these sensors to differentiably decompose the received radar power at a specific distance into occupancy and reflectance-dependent components that can be learned and predicted by neural networks. With geometry disentangled from view and material-dependent reflectance, our model can reconstruct dense scene geometry in addition to synthesizing radar waveforms, which would otherwise be impossible without such a decomposition. In order to leverage our sensor model effectively, we train on a new dataset of raw frequency-space radar measurements, allowing us to directly predict received power at specific ranges, instead of simply predicting depth-per-ray.

Specifically, we make the following contributions in this work:

- We present a neural reconstruction method for active radar sensors, which both recovers scene geometry from a sequence of radar scans and leverages its implicit scene model to synthesize raw radar data from novel views.
- We devise an optimization method which fits the model directly to raw frequency-space radar measurements without requiring volume rendering.
- We introduce a new dataset of automotive radar captures, and we validate the effectiveness of our approach on scene reconstruction and novel view synthesis on in-the-wild scenes.

2 RELATED WORK

Radar as a Sensing Modality. has become indispensable in object recognition and scene understanding, especially in robotic, maritime [Cheng et al. 2021], and autonomous driving domains [Bijelic et al. 2020; Hwang et al. 2022; Wang et al. 2023a]. The capability of mm-wave radiation in weather penetration [Bijelic et al. 2020; Hwang et al. 2022] sets it apart from optical sensors. Beyond this, radar contributes significantly to constructing detailed environmental models that can predict depth [Lin et al. 2020], semantics [Ouaknine et al. 2021; Zhang et al. 2023b], scene flow [Ding et al. 2023], object presence [Bijelic et al. 2020; Hwang et al. 2022; Kim et al. 2023; Li et al. 2022] and non-line-of-sight imaging [Scheiner et al. 2020]. When fused with camera [Ding et al. 2023; Lin et al.

2020] or LiDAR data [Hwang et al. 2022; Li et al. 2022], the inherent sparsity of radar can be mitigated. Other approaches for MIMO radar leverage beam steering techniques and neural networks to super-resolve measurements. [Farrell et al. 2023; Guan et al. 2020; Li et al. 2023a]. Progress in this field owes much to datasets [Caesar et al. 2020; Meyer and Kuschk 2019; Rebut et al. 2022], which are invaluable as benchmarks. We introduce a new dataset curated for radar neural rendering to further advance the state-of-the-art.

Radar Maps. provide a layer of information, in addition to LiDAR and camera reconstructions, allowing for improved understanding of the free space around an autonomous vehicle [Adams et al. 2012; Vivet et al. 2013]. To this end, existing methods accumulate radar measurements and interpret them to classify if each individual grid cell is occupied, thereby enhancing scene reasoning capabilities [Grebner et al. 2022; Lu et al. 2020]. This detailed grid mapping, combined with advanced radar perception, assists in translating raw radar data into interpretable occupancy information. Existing methods differentiate between static and dynamic entities and apply techniques like Doppler processing for object categorization based on relative velocity [Vahidpour and Sarabandi 2012]. With challenges such as multipath interference, beam divergence, and clutter from unwanted reflections, signal post-processing techniques like adaptive filtering [Werber et al. 2015a] and polarimetric backscatter analysis [Moallem and Sarabandi 2014] become crucial. Our proposed representation substantially expands the convex hull of existing methods, and allows us to model object geometry due to the separation of radar cross-section into occupancy and reflectance.

Neural Rendering. A large body of work uses single-sensor measurements to create detailed scene representations primarily from camera and LiDAR data [Huang et al. 2023; Tao et al. 2023; Zhang et al. 2023a]. The use of radar data remains unexplored in this domain. Recognizing this gap, we propose the first neural rendering approach to generate geometric representations from radar data, paving the way for navigation and control applications using this representation in autonomous robotics and beyond.

Learned representations have driven recent advancements in novel-view generation [Barron et al. 2021; Chen et al. 2022; Mildenhall et al. 2020; Müller et al. 2022] and depth estimation [Tosi et al. 2023]. Central to these advancements are neural radiance field methods [Barron et al. 2021; Chen et al. 2022; Mildenhall et al. 2020; Müller et al. 2022], which model scenes as continuous volumetric fields of radiance. These techniques utilize volumetric rendering as a forward model, facilitating smooth interpolation between sensor poses. Various scene representations have been investigated for this task: coordinate-based networks [Barron et al. 2021, 2022; Mildenhall et al. 2020; Zhang et al. 2021], 3D voxel-grids [Chen et al. 2022; Fridovich-Keil et al. 2022; Yu et al. 2021], and hybrid models [Barron et al. 2023; Müller et al. 2022; Tancik et al. 2023]. To enhance efficiency during both training and testing, approaches like [Barron et al. 2023; Chen et al. 2022; Müller et al. 2022; Yu et al. 2021] have been developed. Recent works have extended these techniques to capture vast outdoor scenes [Barron et al. 2022; Zhang et al. 2020]. Representing extensive terrains, especially when captured by moving vehicles, presents unique challenges. This is due to the restricted and aligned views stemming from a single trajectory [Guo et al. 2023; Kundu et al. 2022; Liu et al. 2023; Ost et al. 2022;

Rematas et al. 2022; Tancik et al. 2022; Turki et al. 2023; Wang et al. 2023b; Yang et al. 2023]. To overcome these challenges, researchers have incorporated supplemental supervisory cues, such as sparse LiDAR scans [Guo et al. 2023; Ost et al. 2022; Rematas et al. 2022; Turki et al. 2023], estimated depths [Deng et al. 2022; Guo et al. 2023; Roessle et al. 2022], optical flow data [Meuleman et al. 2023; Turki et al. 2023], and semantic segmentation [Kundu et al. 2022; Turki et al. 2023; Wang et al. 2023b].

Recent work aims to learn scene representations from LiDAR data, which presents challenges due to its sparsity — often at two orders of magnitude less dense than camera data. These methods incorporate precise LiDAR point registration [Huang et al. 2023], or apply a two-tiered approach, using weak semantic supervision to filter out points below a registration threshold [Zhang et al. 2023a]. Others directly predict point drop likelihoods [Tao et al. 2023].

Beyond optical sensors, non-optical sensors like imaging sonar [Qadri et al. 2023; Reed et al. 2023] have leveraged neural representations and physics-based rendering techniques to achieve state of the art 3D reconstructions of single objects in isolation. Applying neural radiance methods to radar data arguably poses an even bigger challenge due to its lower density and long range that enable it to capture large outdoor scenes. We harness raw radar waveforms, which offer improved estimates of empty space.

3 RADAR FIELDS

We introduce Radar Fields as a scene representation capable of recovering dense geometry and synthesizing radar signals at unseen views by fitting to a *single trajectory* of raw radar measurements. The method is illustrated in Fig. 2. It hinges on a physics-based forward model that allows us to disentangle occupancy and reflectance (Sec. 3.1). Tailored to this physical model, we introduce a novel implicit radar field representation (Sec. 3.2) and a physics-based importance sampling schema (Sec. 3.3). We fit the model by reconstructing raw radar signals in frequency space (Sec. 3.4).

3.1 Radar Signal Formation Model

Radar systems emit electromagnetic waves and analyze their reflections from objects to derive distance and velocity information. FMCW radar is distinct in that it emits a continuous radio waveform with a frequency that varies over time, often with a sawtooth-modulated 'chirp' pattern [Jankiraman 2018]. Assuming sawtooth modulation, the frequency of the transmitted signal, f_τ changes linearly over time t as

$$f_\tau(t) = \omega + \theta(t), \quad \text{with} \quad \theta(t) = 2\Delta f \cdot \text{mod}\left(\frac{t}{T_s}, 1\right), \quad (1)$$

where ω is the constant carrier frequency and $\theta(t)$ is a periodic sawtooth function with period T_s and half-amplitude Δf.

When the chirp reflects off an object and returns to the sensor, the time delay introduced by the distance traveled results in a phase offset and therefore a frequency difference between the transmitted waveform f_τ and the received waveform f_r. For a single object,

$$f_r(t) = \omega + \theta(t - t_d), \quad \text{with} \quad t_d = \frac{2R}{c}, \quad (2)$$
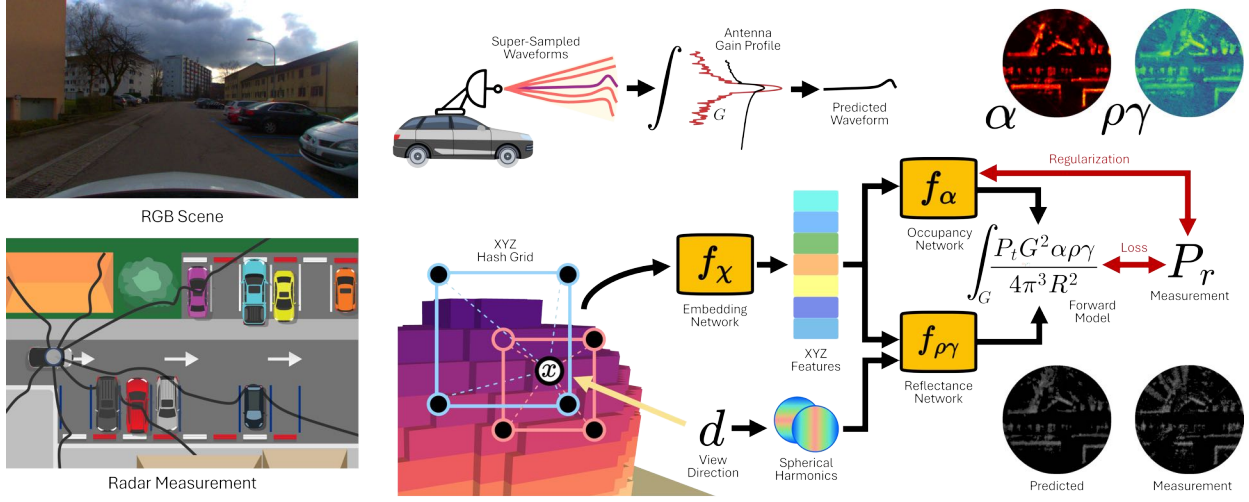
Figure 2: Radar Fields recovers 3D scene geometry from raw FFTs of 2D range-azimuth radar scans. Each radar frame integrates information per azimuth beam into a 2D disk of waveforms (bottom-left). Radar Fields models this process by sampling a cone of points within the divergence of each beam, and integrating these samples into a single 2D predicted measurement using the gain profile of the radar antenna (top-middle). Each cone of points is embedded by f_χ into a feature vector. These features are processed by f_α and $f_{\rho\gamma}$, which decompose the measurement intensity into point-wise occupancy α and reflectance $\rho\gamma$. We then apply our forward model and integrate the resulting beam of 3D predictions to reconstruct an FFT measurement.

where t_d is the two-way time delay for light reflecting off the object at range R and returning to the radar, assuming that the target is stationary and the Doppler frequency shift is zero.

Both signals are then processed through a mixer and low-pass filter to compute their instantaneous frequency and phase differences. These computed differences can be treated as a new signal, called the intermediate frequency (IF) waveform, whose frequency f_{IF} and phase are equal to the computed differences, respectively

$$f_{IF}(t) = f_\tau(t) - f_r(t) = \theta(t) - \theta(t - t_d). \quad (3)$$

In the case of a single target, the IF waveform is a sinusoid with frequency f_{IF}. But In practice, multiple objects are usually detected, meaning that the IF signal would be a sum of many different frequency sinusoids, each of which corresponds to the range and $f_r(t)$ of a different target [Jankiraman 2018]. Subsequently, a Fast Fourier Transform (FFT) yields the observable targets at any distance R_b,

$$P_r(b) = \sum_{n=0}^{N_b-1} IF(n) \cdot e^{\frac{-i2\pi bn}{N_b}}, \quad \text{with} \quad b = f_{IF_b}, \quad (4)$$

where IF is the IF waveform and N_b is the total number of frequency bins, such that each bin b corresponds to a different tone in the IF frequency f_{IF_b} and is correlated with a different range via

$$R_b = f_{IF_b} \frac{cT_s}{4\Delta f}. \quad (5)$$

We aim to fit the raw radar FFT signal described above. As such, our method needs to predict the detected power $P_r(b)$ at every bin b, thereby reconstructing a frequency-space waveform for every azimuth-resolved beam. To this end, we rely on the known physics of FMCW radar to formulate a signal formation model. Due to the correlation between range and IF frequency in Eq. 5, the detected

power in Fourier space at a given frequency bin b, $P_r(b)$, can be modeled as the returned power detected by the sensor at corresponding range R_b. Therefore, we can assume that for an FFT bin b, whose frequency corresponds to a distributed target at range R_b from the sensor, the received power $P_r(b)$ can be calculated as,

$$P_r(b) = \frac{P_t \cdot G^2 \cdot \sigma}{(4\pi)^3 R_b^2}, \quad (6)$$

where P_t is the transmitted power, G is the antenna gain and, σ is the total radar cross-section of the distributed target [Richards 2010; Skolnik 2001]. Note that every term in this equation is known except for σ. The radar cross-section considers object shape, material, and reflective properties, describing how detectable it is by a radar.

However, predicting σ directly is insufficient for geometry reconstruction. Instead, we aim to disentangle the geometry-dependent component of σ from its specular and material components. The radar cross section σ can be further decomposed into three terms, as $\sigma = \alpha \cdot \rho \cdot \gamma$, with α being the object size projected onto the cross-section of the radar beam, and with ρ being the reflectivity and γ being the directivity of the object. Here, α depends solely on object geometry, while ρ and γ also depend on how metallic an object is, as well as the incident angle of the radar beam. This defines an interpretable forward model as the foundation of this method, where separate neural fields can be learned for each term.

3.2 Implicit Neural Field Representation

We learn an implicit neural model to render our two-dimensional radar measurements $P_r \in \mathbb{R}^{N_\phi \times N_b}$ for N_b range bins and N_ϕ azimuth angles. The radar cross-section $\sigma \in \mathbb{R}$ is predicted for each measurement grid cell. σ is factorized point-wise into projected cross-sectional area $\alpha \in \mathbb{R}$, which represents scene occupancy, and
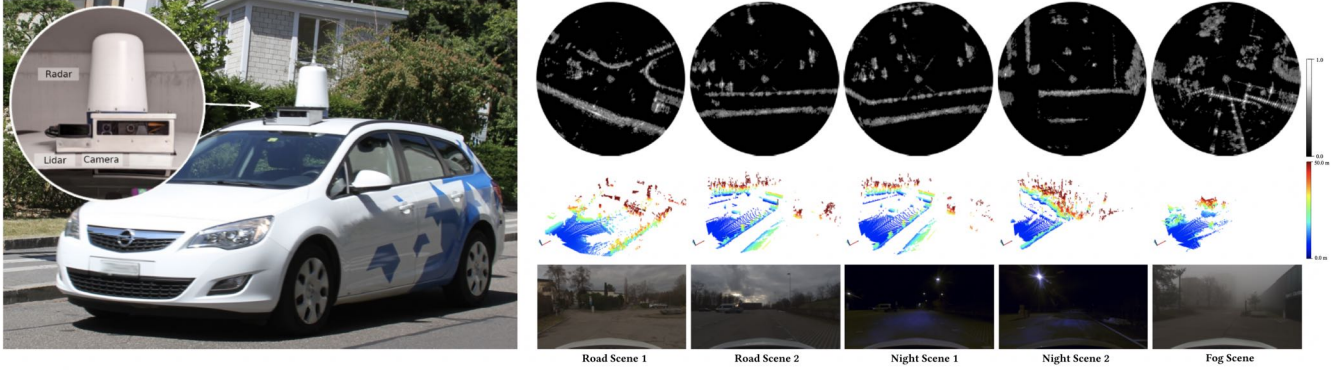
Figure 3: Multi-modal Dataset. We use a waterproof sensor rig (left) to collect data. On the right are 5 exemplary scenes, with (1st row) 40-meter-radius BEV radar returns with the car centered and facing right, (2nd row) point clouds from our forward-facing LiDAR, colored by height, with the car in the bottom-left, and (3rd row) images from our forward-facing RGB camera.

the product of reflectivity and directivity, $\rho\gamma \in \mathbb{R}$, which represents scene reflectance. We fit our model to a sequence of raw radar captures, and corresponding poses from an on-board Global Navigation Satellite System (GNSS), as in Fig. 2.

We reconstruct FFT measurements using two neural fields, f_α and $f_{\rho\gamma}$, which represent scene occupancy and reflectance, respectively. Both fields are conditioned on an embedding $\chi \in \mathbb{R}^d$ from the neural field f_χ. Given a point in space $\mathbf{x} \in \mathbb{R}^3$ and view direction $\mathbf{d} \in \mathbb{R}^3$, the total scene decomposition can be written as

$$f_\chi : \{\mathcal{H}(\mathbf{x})\} \longrightarrow \{\chi\} \text{ Feature Embedding}$$
$$f_\alpha : \{\chi\} \longrightarrow \{\alpha\} \text{ Occupancy}$$
$$f_{\rho\gamma} : \{\mathcal{S}(\mathbf{d}), \chi\} \longrightarrow \{\rho\gamma\} \text{ Reflectance}$$

where $\mathcal{H}$ and $\mathcal{S}$ are multi-resolution hash encodings as in Müller et al. [2022], and spherical harmonic positional encoding as in Fridovich-Keil and Yu et al. [2022], respectively.

The embedding field $f_\chi : \{\mathcal{H}(\mathbf{x})\} \rightarrow \{\chi\}$ predicts geometric and material features χ from hashgrid-encoded position. These scene latents are shared by both components of our representation. We predict occupancy with field $f_\alpha : \{\chi\} \rightarrow \{\alpha\}$, as projected object cross-sectional area can be decomposed into a point-wise measure of occupancy. We predict reflectance with field $f_{\rho\gamma} : \{\mathcal{S}(\mathbf{d}), \chi\} \rightarrow \{\rho\gamma\}$, which is also conditioned on view direction $\mathbf{d}$ encoded with spherical harmonics $\mathcal{S}$. This field directly predicts the product of surface reflectivity and directivity, $\rho \cdot \gamma$, as there is no clear way to disambiguate these two values and supervise them independently.

One advantage of this representation is its versatility, as α can be used alone to reconstruct scene geometry, but can also be multiplied with $\rho \cdot \gamma$ at unseen locations or novel views to *synthesize returned waveforms directly*. Note that this model does *not require volume rendering*. The radar signal used for supervision is a raw time-resolved waveform, and the ray samples themselves are the model predictions.

3.3 Beam Importance Sampling

Radar beams diverge as they travel through a scene. This divergence is not negligible in FMCW radar sensing, given both the long range capabilities of radio detection and the large elevation opening

angles in automotive settings. Most low-cost FMCW radars do not have any elevation resolution. However, these sensors still capture information in three dimensions, as they integrate signals across elevation and azimuth for each beam. We model these properties by super-sampling additional rays within the elliptical cone defined by our radar azimuth and elevation opening angles, as shown in Fig. 2 - this is the key to learning a 3D scene representation from a 2D range-azimuth supervision signal.

FMCW radar systems do not use isotropically-radiating antennas that would distribute and receive power evenly across the opening angles of the beam. Instead, they rely on antennas with a directional bias. These exhibit the highest signal gain in the direction of the beam center while dropping off steeply at increasingly divergent angles off-boresight. This can be represented as a radiation pattern, which measures signal gain as a function of angular offset relative to the beam center, and affects returned power as in Eq. 6. We consider two radiation patterns, $\mathcal{A}(a)$ and $\mathcal{E}(e)$, which map azimuth and elevation angular offsets a and e to signal gain, respectively.

To model these sensor properties, we super-sample additional rays, distributed uniformly within the elliptical cone of beam divergence around each beam center. For each of the N beam centers randomly sampled at any given training step, we sample an additional $S - 1$ rays to create a set $\mathbb{S} = \{s_1, s_2, ..., s_S\}$ such that $|\mathbb{S}| = S$. Each ray super-sample consists of angular offsets a_i and e_i from the beam center, such that $s_i = \{a_i, e_i\} \, \forall s_i \in \mathbb{S}$. The super-samples s_i are drawn from a pair of uniform distributions, such that $a_i \sim \mathcal{U}(-A, A)$ and $e_i \sim \mathcal{U}(-E, E)$, where $\mathcal{U}$ is a uniform distribution over all of the possible angular offsets from the beam center with azimuthal field of view $2A$ and elevation field of view $2E$. We query our model with S uniformly-distributed rays per beam and predict a separate radar cross-section $\hat{\sigma}_i = \hat{\alpha}_i \cdot \hat{\rho\gamma}_i$, for each super-sampled ray $s_i \in \mathbb{S}$ and average them proportional to our sensor radiation profiles $\mathcal{A}$ and $\mathcal{E}$ to compute the final predicted radar cross-section $\hat{\sigma}$ at each range bin,

$$\hat{\sigma} = \frac{\sum_{i=1}^{S} \hat{\sigma}_i \cdot \mathcal{A}(s_i[0]) \cdot \mathcal{E}(s_i[1])}{\sum_{i=1}^{S} \mathcal{A}(s_i[0]) \cdot \mathcal{E}(s_i[1])}, \tag{7}$$

from which we compute the predicted power return. This way, samples that lie further from each beam center contribute less to

the received signal intensity, mirroring the physics of the radar measurement process and establishing a mapping between our learned 3D scene representation and our 2D FFT data.

3.4 Training

We train our model to reconstruct a scene from a sequence of radar frames applying the total loss $\mathcal{L}$. The sole supervision signal is the raw radar waveform itself. Predicted occupancy is also regularized to enforce its geometric interpretation. The total loss $\mathcal{L}$ is a weighted sum of the reconstructed FFT loss $\mathcal{L}_W$, and the regularization terms $\mathcal{L}_R$ and $\mathcal{L}_P$ for the occupancy, leading to,

$$\mathcal{L} = \eta_W \mathcal{L}_W + \eta_R \mathcal{L}_R + \eta_P \mathcal{L}_P \tag{8}$$

$$\mathcal{L}_W = \frac{1}{N_\phi N_b} \sum_{\phi,b} \left\| \frac{P_t \cdot G^2 \cdot (\hat{\alpha}_{\phi,b} \cdot \hat{\rho}\gamma_{\phi,b})}{(4\pi)^3 R_b^2} - P_{r_{\phi,b}} \right\| \tag{9}$$

$$\mathcal{L}_R = \frac{1}{BatchSize} \sum_{\phi,b} O(P_r)_{\phi,b} \left(log(O(P_r)_{\phi,b}) - log(\hat{\alpha}_{\phi,b}) \right) \tag{10}$$

$$\mathcal{L}_P = \text{std} \left(\hat{\alpha}|_{O(P_r)>0.5} \right) + \text{std} \left(\hat{\alpha}|_{O(P_r)<0.5} \right) \tag{11}$$

where the weights are η_W, η_R and η_P. In detail, L_W assesses the quality of the reconstructed FFT signal. $\mathcal{L}_R$ ensures that the learned scene geometry aligns with the occupancy derived from the raw signal. Occupancy from a 2D raw radar signal can be estimated by applying an occupancy estimation algorithm O. Our algorithm outputs an occupancy probability estimate $O(P_r) \in \mathbb{R}^{N_\phi \times N_b}$, and follows previous work like [Werber et al. 2015b], scanning over measurement bins for each frame and estimating occupancy likelihood per-bin with a simple Bayesian update rule and occlusion model. More details are in the Supplementary Material. $\mathcal{L}_P$ enforces a bimodal distribution of occupancy probabilities, ensuring that empty space is correctly modeled and mitigating floater artifacts.

To prevent the model from over-fitting to noisy data and getting trapped in local minima, we employ a coarse-to-fine optimization procedure, as in [Li et al. 2023b]. We mask our multi-resolution hashgrid encodings $\mathcal{H}(\mathbf{x})$ as

$$\mathcal{H}(\mathbf{x}) = \begin{cases} \mathcal{H}(\mathbf{x}) & \text{if } \frac{i}{|\gamma|} < 0.4 + 0.6 \left(\sin\left(\frac{\text{epoch}}{\text{max. epoch}} \right) \right) \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

gradually providing the model with higher-resolution grid features over the course of training. This way, reconstruction noise cannot accumulate in higher frequency features during early training.

4 DATASET

We train and validate our method on a novel multi-modal dataset. The recorded modalities include raw radar, LiDAR, RGB camera, and GNSS.Our FMCW radar, which is a custom Navtech CIR-DEV, recording 360° around the vehicle and providing a long range of 330 m at a range resolution of ca. 4 cm. The camera, a TRI023S-C with 2 MPix, and the LiDAR both span a frontal field of view (FOV). We incorporate a state-of-the-art microelectromechanical-system (MEMS) LiDAR (RS-LiDAR-M1) with a horizontal FOV of 120°.

To ensure accurate sensor supervision, we performed both geometric calibration and synchronization of our sensors. We calibrate the RGB camera intrinsics with conventional checkerboard calibration [Bradski 2000], while the intrinsic calibration for both LiDAR

and radar is provided by the vendor. Our extrinsic calibration for LiDAR-camera uses mutual information maximization [Pandey et al. 2015]. For radar-LiDAR calibration, we follow [Burnett et al. 2023] and estimate the rotation via correlative scan matching using the Fourier Mellin transform [Checchin et al. 2010], with the translation directly measured. GNSS-LiDAR calibration uses u-center [u-blox AG 2023] and LiDAR point cloud consistency optimization. We synchronize all internal clocks and record each sensor independently.

We completed extensive in-the-wild driving sessions in Switzerland, capturing multiple hours of footage. We selected 15 sequences, emphasizing scene diversity and ensuring a high quality GNSS signal. Each sequence has a duration of 10-23 seconds, containing 40-90 radar frames. These authentically represent real-world driving scenarios, with the vehicle moving at speeds from 5 to 30 km/h.

The dataset includes diverse conditions, including day, night, and fog, and types of scenes, including urban streets and parking lots, with example scenes visualized in Fig. 3. This diversity allows us to investigate the impact of varying scenes and environmental conditions. We include recordings from both day and night in the same scene, enabling detailed examinations of ambient light effects. We investigate sensor disparities between radar and LiDAR by including adverse weather scenes recorded in strong fog. We anticipate the radar maintaining performance in these adverse conditions, while the LiDAR and camera are expected to be more affected.

5 ASSESSMENT

In this section, we assess the proposed method with the novel multimodal dataset described in the previous section. Specifically, we investigate scene reconstruction across day, night, and fog scenes, for bird's eye view (BEV) occupancy reconstruction and novel view synthesis in 2D, and with geometry reconstruction in 3D. Specifically, we validate that the method is capable of recovering 3D scene representations from conventional 2D radar scans by incorporating the angular-dependent antenna response in the forward model.

5.1 Experimental Setup

The collected outdoor scenes in our dataset fall into three main categories: parking lots featuring a high number of dielectric surfaces on parked vehicles that appear in radar returns, urban, where scene composition is complex, and foggy adverse weather scenes. We withhold 20 % of all frames to form a train-test split.

Evaluation Criteria. We measure the PSNR and RSME of the reconstructed radar FFT signal compared to the withheld groundtruth FFT data. To assess BEV occupancy reconstructions, we report the Chamfer Distance (CD) and Relative Chamfer Distance (RCD) between the 2D BEV point cloud estimated with our method and a LiDAR-based ground truth estimate. Note that we derive additional ground truth occupancy from LiDAR, which provides geometry with a resolution that is an order of magnitude higher. We accumulate LiDAR data to create a dense ground truth, which is filtered, projected into the radar frame, and combined with highly probable radar returns.

5.2 Scene Reconstruction

We report qualitative scene reconstruction results in Fig. 5. We observe that accumulated radar point clouds from conventional DSP

Table 1: Quantitative Assessment. We measure reconstruction accuracy of our method compared to a radar grid mapping baseline [Werber et al. 2015b] and LiDAR-NeRF [Tao et al. 2023]. We validate the effectiveness of importance sampling & regularization as an ablation. Radar Fields compares favorably in CD and RCD for scene reconstruction, and in terms of RMSE and PSNR for novel view synthesis. Radar Fields outperforms LiDAR-NeRF in adverse weather, as visibly proven in Fig. 6.

	Scene 1				Scene 2				Scene 3				Scene 4				Scene 5				Scene Fog			
	CD↓	RCD↓	RMSE↓	PSNR↑	CD↓	RCD↓	RMSE↓	PSNR↑	CD↓	RCD↓	RMSE↓	PSNR↑	CD↓	RCD↓	RMSE↓	PSNR↑	CD↓	RCD↓	RMSE↓	PSNR↑	CD↓	RCD↓	RMSE↓	PSNR↑
	[m]	-	-	[dB]	[m]	-	-	[dB]	[m]	-	-	[dB]	[m]	-	-	[dB]	[m]	-	-	[dB]	[m]	-	-	[dB]
Proposed	**0.163**	0.013	**0.185**	20.660	**0.296**	**0.007**	**0.190**	20.431	0.166	**0.022**	**0.169**	21.461	**0.227**	**0.025**	**0.198**	20.108	**0.193**	0.018	**0.184**	20.738	0.382	**0.015**	**0.185**	20.738
Ablation	0.260	0.014	0.231	18.740	0.482	0.012	0.262	17.647	0.243	0.023	0.199	20.030	0.284	0.026	0.278	17.132	0.194	0.018	0.261	17.691	**0.013**	0.356	0.217	19.32
Grid Mapping [2015b]	0.240	0.019	0.212	19.492	1.222	0.047	0.228	18.850	**0.164**	0.027	0.220	19.171	0.241	0.036	0.206	19.732	0.361	0.033	0.248	18.148	0.614	0.048	0.255	17.887
LiDAR-NeRF [2023]	0.189	**0.012**	-	-	0.312	0.008	-	-	0.187	0.022	-	-	0.325	0.093	-	-	0.399	0.083	-	-	0.797	0.087	-	-

processing are too sparse to provide dense scene information, and therefore the occupancy grid-maps recovered from these inputs are unreliable and minimally informative, only tracing out a rough contour of large structures in the immediate vicinity of the sensor. Row three of Fig. 5 shows that these types of methods completely fail to reconstruct any vehicles in the scene - which is problematic for safety-critical automotive settings - as the inconsistent multipath returns from these small objects average out instead of coinciding into a dense silhouette. These methods are largely unable to see behind occluding structures. Moreover, the 2D nature of processed radar point clouds confine occupancy to BEV, as there are no 3D cues available. Access to raw FFT data makes it possible for Radar Fields to reconstruct the scene, including all vehicles in test scenes. Crucially, our neural occupancy field is 3D. Our physics-based ray importance sampling, described in section 3.3, models radio beam divergence to extract 3D information from a sensor with no elevation resolution. In these 3D reconstructions, it is again possible to distinguish vehicles and walls that were not visible in previous approaches. The last row in Fig. 5 validates that the importance sampling and regularization terms are essential. Without them, our method struggles to disentangle occupancy from reflectance, as specular effects become more prevalent in the occupancy field.

In Table 1, we use CD and RCD as metrics to evaluate BEV scene reconstruction. Confirming our findings from qualitative evaluation, our method consistently improves on grid mapping [Werber et al. 2015b] with conventional radar post-processing. The metrics also validate the physics-based importance sampling and regularization.

5.3 Adverse Weather

Radar Fields retains reconstruction quality across weather and lighting conditions. Fig. 6 reports reconstructions in fog, validating that the proposed method is able to reliably recover both 2D occupancy and 3D geometry in extreme conditions where other methods using LiDAR or camera input fail. We compare Radar Fields against LiDAR-NeRF [Tao et al. 2023] using LiDAR input and Instant-NGP [Müller et al. 2022] relying on RGB camera input, trained on the same day-time and night-time foggy scenes from our dataset. In these extreme conditions, LiDAR-NeRF performance degrades due to limited range information in dense fog, while Instant-NGP fails to reconstruct any meaningful occupancy due to the severe scattering in the scene. In contrast Radar Fields recovers crisp outlines of buildings, steel road barriers, and even vehicles remain visible. The last scene in Table 1 shows metrics for fog. Our method achieves lower CD and RCD than LiDAR-NeRF for scene reconstructions.

5.4 Novel Radar View Synthesis

Radar Fields is grounded in reconstructing raw frequency waveforms, which is our only source of scene information. As such, it is capable of synthesizing raw radar returns at novel views. Fig. 4 confirms that these synthesized views can capture view-dependent reflective artifacts, like specular highlights and ray saturation. We validate the method with RMSE and PSNR as quantitative metrics. Table 1 confirms also that the absence of importance sampling and regularization terms in the model leads to less accurate signal prediction for novel views and inconsistent convergence.

6 CONCLUSION

We introduce a neural rendering method for raw radar data. While a large body of work on neural scene representations investigates the reconstruction and generation of novel views from RGB and LiDAR point cloud data, neural rendering for radar measurements has been unexplored. Operating at millimeter wavelengths, radar sensors provide a signal complementary to optical imaging techniques – radar signals penetrate fog and smoke with scattering particle sizes smaller than the wavelength. Unfortunately, the longer wavelength also inherently limits the angular resolution which existing neural rendering methods fundamentally rely on in multi-view consistent training. The proposed method tackles the resolution limitations of processed radar data and the computational cost of volume rendering by modeling scene parameters in frequency space. We supervise our models with raw radar waveform data in Fourier frequency space to recover relationships between detected power and distance from the sensor. We validate our method across diverse scenarios, especially in urban environments with dense vehicles and infrastructure, where mm-wavelength sensing is favorable. As a first step towards practical scene representations for radar, in the future, the proposed fusion methods could benefit from cross-modal input and supervision with LiDAR data, thereby bridging the resolution disparity, and optimizing geometry reconstruction despite the inherent angular resolution limitations of radar data.
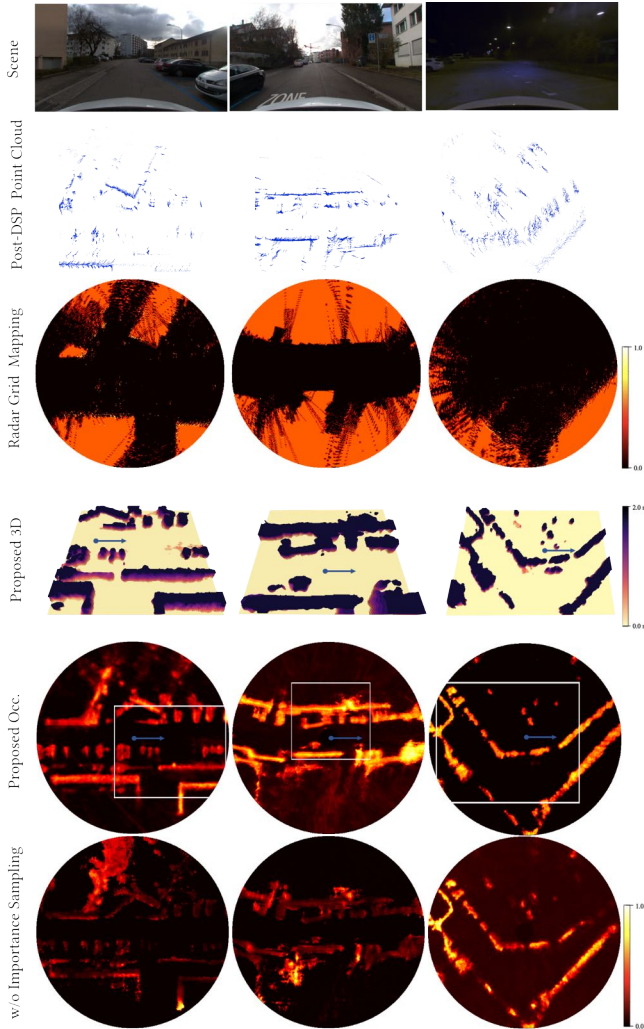
Figure 5: Radar Fields for Scene Reconstruction. Conventional post-processed radar point clouds (second row) are sparse, and, hence, conventional grid mapping methods [Werber et al. 2015b] (third row) fail to recover accurate geometry. Radar Fields relies on raw frequency-space radar measurements and recovers high-quality BEV occupancy (fourth row), and even accurate 3D geometry (third row) from the same 2D radar scans. Without physics-based ray importance sampling (last row), the predicted occupancy becomes inconsistent.
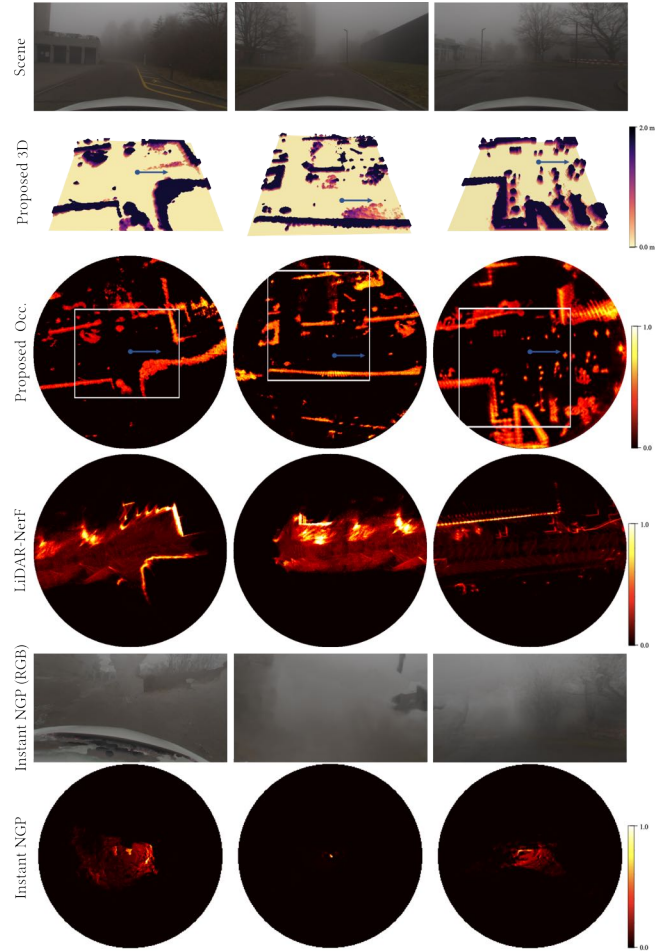

Figure 6: Radar Fields for adverse weather. Our method is robust to extreme weather and lighting conditions, including low-light and foggy scenes. RGB captures are reported in the top row. We compare our method to LiDAR-NeRF [Tao et al. 2023] for LiDAR and Instant-NGP [Müller et al. 2022] for camera inputs as alternative modalities. LiDAR reconstructions struggle to capture accurate scene geometry due to backscatter. Multi-view reconstruction via RGB fails for this monocular foggy trajectory, also indicated by the synthesized RGB frames (second to last row). Radar Fields exhibits minimal degradation compared to good weather conditions.
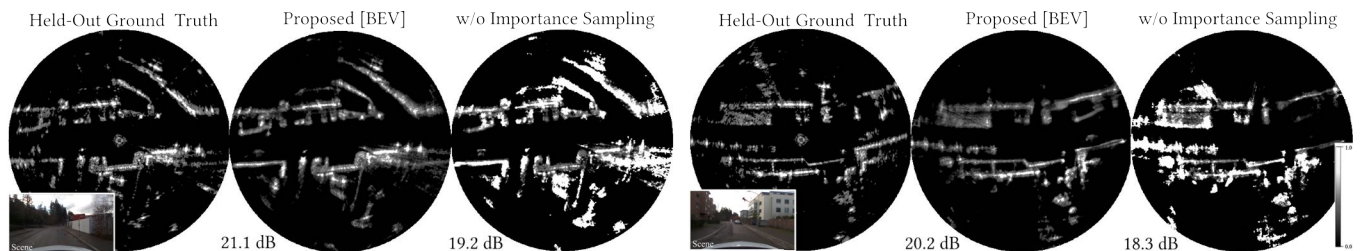

Figure 4: Novel View Synthesis. Radar Fields is capable of synthesizing high-quality raw radar FFT measurements at novel viewpoints. Without our proposed super-sampling procedure, predictions become noise prone and inaccurate in magnitude.

REFERENCES

Martin Adams, Ebi Jose, and Ba-Ngu Vo. 2012. *Robotic navigation and mapping with radar.* Artech House.

Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. 2021. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. *ICCV* (2021).

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5470–5479.

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2023. Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields. *ICCV* (2023).

Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. 2020. Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

Ondrey Burkacky, Johannes Deichmann, Michael Guggenheimer, and Martin Keller. 2023. *The automotive software and electronics market is poised for strong growth in the next decade. Our latest market projections offer a glimpse into the industry's future.* https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/mapping-the-automotive-software-and-electronics-landscape-through-2030

Keenan Burnett, David J Yoon, Yuchen Wu, Andrew Z Li, Haowei Zhang, Shichen Lu, Jingxing Qian, Wei-Kang Tseng, Andrew Lambert, Keith YK Leung, et al. 2023. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research* 42, 1-2 (2023), 33–42.

Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. 2020. nuScenes: A multimodal dataset for autonomous driving. In *CVPR*.

Paul Checchin, Franck Gérossier, Christophe Blanc, Roland Chapuis, and Laurent Trassoudaine. 2010. Radar scan matching slam using the Fourier-Mellin transform. In *Field and Service Robotics: Results of the 7th International Conference*.

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision*. Springer, 333–350.

Yuwei Cheng, Hu Xu, and Yimin Liu. 2021. Robust Small Object Detection on the Water Surface Through Fusion of Camera and Millimeter Wave Radar. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 15263–15272.

Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. 2022. Depth-supervised nerf: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12882–12891.

Juergen Dickmann, Nils Appenrodt, Hans-Ludwig Bloecher, C. Brenk, Thomas Hackbarth, Markus Hahn, Jens Klappstein, Marc Muntzinger, and Alfons Sailer. 2014. Radar contribution to highly automated driving. In *2014 44th European Microwave Conference*. 1715–1718. https://doi.org/10.1109/EuMC.2014.6986787

Fangqiang Ding, Andras Palffy, Dariu M. Gavrila, and Chris Xiaoxuan Lu. 2023. Hidden Gems: 4D Radar Scene Flow Learning Using Cross-Modal Supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 9340–9349.

Ghina El Natour, Omar Ait Aider, Raphael Rouveure, François Berry, and Patrice Faure. 2015. Radar and vision sensors calibration for outdoor 3D reconstruction. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*. 2084–2089. https://doi.org/10.1109/ICRA.2015.7139473

Sean M. Farrell, Vivek Boominathan, Nathaniel Raymondi, Ashutosh Sabharwal, and Ashok Veeraraghavan. 2023. CoIR: Compressive Implicit Radar. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023), 1–12. https://doi.org/10.1109/TPAMI.2023.3301553

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5501–5510.

Fridovich-Keil and Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance Fields without Neural Networks. In *CVPR*.

Timo Grebner, Pirmin Schoeder, Vinzenz Janoudi, and Christian Waldschmidt. 2022. Radar-Based Mapping of the Environment: Occupancy Grid-Map Versus SAR. *IEEE Microwave and Wireless Components Letters* 32, 3 (2022), 253–256. https://doi.org/10.1109/LMWC.2022.3145461

Junfeng Guan, Sohrab Madani, Suraj Jog, Saurabh Gupta, and Haitham Hassanieh. 2020. Through Fog High-Resolution Imaging Using Millimeter Wave Radar. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 11461–11470. https://doi.org/10.1109/CVPR42600.2020.01148

Jianfei Guo, Nianchen Deng, Xinyang Li, Yeqi Bai, Botian Shi, Chiyu Wang, Chenjing Ding, Dongliang Wang, and Yikang Li. 2023. StreetSurf: Extending Multi-view Implicit Surface Reconstruction to Street Views. *arXiv preprint arXiv:2306.04988* (2023).

Eiji Hayashi, Jaime Lien, Nicholas Gillian, Leonardo Giusti, Dave Weber, Jin Yamanaka, Lauren Bedal, and Ivan Poupyrev. 2021. RadarNet: Efficient Gesture Recognition Technique Utilizing a Miniature Radar Sensor. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 5, 14 pages. https://doi.org/10.1145/3411764.3445367

Shengyu Huang, Zan Gojcic, Zian Wang, Francis Williams, Yoni Kasten, Sanja Fidler, Konrad Schindler, and Or Litany. 2023. Neural LiDAR Fields for Novel View Synthesis. *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Jyh-Jing Hwang, Henrik Kretzschmar, Joshua Manela, Sean Rafferty, Nicholas Armstrong-Crews, Tiffany Chen, and Dragomir Anguelov. 2022. CramNet: Camera-Radar Fusion with Ray-Constrained Cross-Attention for Robust 3D Object Detection. https://doi.org/10.48550/ARXIV.2210.09267

M. Jankiraman. 2018. *FMCW Radar Design.* Artech House. https://books.google.com/books?id=TvX0tQEACAAJ

Youngseok Kim, Juyeb Shin, Sanmin Kim, In-Jae Lee, Jun Won Choi, and Dongsuk Kum. 2023. CRN: Camera Radar Net for Accurate, Robust, Efficient 3D Perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 17615–17626.

Abhijit Kundu, Kyle Genova, Xiaoqi Yin, Alireza Fathi, Caroline Pantofaru, Leonidas J Guibas, Andrea Tagliasacchi, Frank Dellaert, and Thomas Funkhouser. 2022. Panoptic neural fields: A semantic object-aware neural scene representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12871–12881.

Seongwook Lee, Song-Yi Kwon, Bong-Jun Kim, Hae-Seung Lim, and Jae-Eun Lee. 2021. Dual-Mode Radar Sensor for Indoor Environment Mapping. *Sensors* 21, 7 (2021). https://doi.org/10.3390/s21072469

Deborah Levy, Amit Peleg, Naama Pearl, Dan Rosenbaum, Derya Akkaynak, Simon Korman, and Tali Treibitz. 2023. SeaThru-NeRF: Neural Radiance Fields in Scattering Media. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 56–65.

Yu-Jhe Li, Shawn Hunt, Jinhyung Park, Matthew O'Toole, and Kris Kitani. 2023a. Azimuth Super-Resolution for FMCW Radar in Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 17504–17513.

Yu-Jhe Li, Jinhyung Park, Matthew O'Toole, and Kris Kitani. 2022. Modality-Agnostic Learning for Radar-Lidar Fusion in Vehicle Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 918–927.

Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H. Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. 2023b. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 8456–8465.

Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar. *ACM Trans. Graph.* 35, 4, Article 142 (jul 2016), 19 pages. https://doi.org/10.1145/2897824.2925953

Juan-Ting Lin, Dengxin Dai, and Luc Van Gool. 2020. Depth Estimation from Monocular Images and Sparse Radar Data. In *International Conference on Intelligent Robots and Systems (IROS)*.

Jeffrey Yunfan Liu, Yun Chen, Ze Yang, Jingkang Wang, Sivabalan Manivasagam, and Raquel Urtasun. 2023. Real-Time Neural Rasterization for Large Scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 8416–8427.

Chris Xiaoxuan Lu, Stefano Rosa, Peijun Zhao, Bing Wang, Changhao Chen, John A. Stankovic, Niki Trigoni, and Andrew Markham. 2020. See through Smoke: Robust Indoor Mapping with Low-Cost MmWave Radar. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services* (Toronto, Ontario, Canada) *(MobiSys '20)*. Association for Computing Machinery, New York, NY, USA, 14–27. https://doi.org/10.1145/3386901.3388945

Andreas Meuleman, Yu-Lun Liu, Chen Gao, Jia-Bin Huang, Changil Kim, Min H. Kim, and Johannes Kopf. 2023. Progressively Optimized Local Radiance Fields for Robust View Synthesis. In *CVPR*.

Michael Meyer and Georg Kuschk. 2019. Automotive Radar Dataset for Deep Learning Based 3D Object Detection. In *2019 16th European Radar Conference (EuRAD)*. 129–132.

Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. 2022. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16190–16199.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.

Meysam Moallem and Kamal Sarabandi. 2014. Polarimetric Study of MMW Imaging Radars for Indoor Navigation and Mapping. *IEEE Transactions on Antennas and Propagation* 62, 1 (2014), 500–504. https://doi.org/10.1109/TAP.2013.2289354

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.* 41, 4, Article 102 (July 2022), 15 pages. https://doi.org/10.1145/3528223.3530127

Julian Ost, Issam Laradji, Alejandro Newell, Yuval Bahat, and Felix Heide. 2022. Neural Point Light Fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2022).

Arthur Ouaknine, Alasdair Newson, Patrick Pérez, Florence Tupin, and Julien Rebut. 2021. Multi-View Radar Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 15671–15680.

Gaurav Pandey, James R McBride, Silvio Savarese, and Ryan M Eustice. 2015. Automatic extrinsic calibration of vision and lidar by maximizing mutual information. *Journal of Field Robotics* 32, 5 (2015), 696–722.

Volodymyr Pavlikov, Valeriy Volosyuk, Simeon Zhyla, Eduard Tserne, Olexandr Shmatko, and Anton Sobkolov. 2021. Active-Passive Radar for Radar Imaging from Aerospace Carriers. In *IEEE EUROCON 2021 - 19th International Conference on Smart Technologies*. 18–24. https://doi.org/10.1109/EUROCON52738.2021.9535619

Mohamad Qadri, Michael Kaess, and Ioannis Gkioulekas. 2023. Neural implicit surface reconstruction using imaging sonar. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1040–1047.

Andrea Ramazzina, Mario Bijelic, Stefanie Walz, Alessandro Sanvito, Dominik Scheuble, and Felix Heide. 2023. ScatterNeRF: Seeing Through Fog with Physically-Based Inverse Neural Rendering. *arXiv preprint arXiv:2305.02103* (2023).

Julien Rebut, Arthur Ouaknine, Waqas Malik, and Patrick Pérez. 2022. Raw High-Definition Radar for Multi-Task Learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 17000–17009. https://doi.org/10.1109/CVPR52688.2022.01651

Albert Reed, Juhyeon Kim, Thomas Blanford, Adithya Pediredla, Daniel Brown, and Suren Jayasuriya. 2023. Neural Volumetric Reconstruction for Coherent Synthetic Aperture Sonar. *ACM Trans. Graph.* 42, 4, Article 113 (jul 2023), 20 pages. https://doi.org/10.1145/3592141

Giulio Reina, James Underwood, Graham Brooker, and Hugh Durrant-Whyte. 2011. Radar-based perception for autonomous outdoor vehicles. *Journal of Field Robotics* 28, 6 (2011), 894–913. https://doi.org/10.1002/rob.20393 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/rob.20393

Konstantinos Rematas, Andrew Liu, Pratul P. Srinivasan, Jonathan T. Barron, Andrea Tagliasacchi, Tom Funkhouser, and Vittorio Ferrari. 2022. Urban Radiance Fields. *CVPR* (2022).

Mark A. Richards (Ed.). 2010. *Principles of Modern Radar: Basic principles.* Institution of Engineering and Technology. https://digital-library.theiet.org/content/books/ra/sbra021e

Barbara Roessle, Jonathan T Barron, Ben Mildenhall, Pratul P Srinivasan, and Matthias Nießner. 2022. Dense depth priors for neural radiance fields from sparse input views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12892–12901.

Nicolas Scheiner, Florian Kraus, Fangyin Wei, Buu Phan, Fahim Mannan, Nils Appenrodt, Werner Ritter, Jurgen Dickmann, Klaus Dietmayer, Bernhard Sick, and Felix Heide. 2020. Seeing Around Street Corners: Non-Line-of-Sight Detection and Tracking In-the-Wild Using Doppler Radar. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

M.I. Skolnik. 2001. *Introduction to Radar Systems.* McGraw-Hill. https://books.google.com/books?id=Y6-APwAACAAJ

Christian Stetco, Barnaba Ubezio, Stephan Mühlbacher-Karrer, and Hubert Zangl. 2020. Radar Sensors in Collaborative Robotics: Fast Simulation and Experimental Validation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 10452–10458. https://doi.org/10.1109/ICRA40945.2020.9197180

Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P Srinivasan, Jonathan T Barron, and Henrik Kretzschmar. 2022. Block-nerf: Scalable large scene neural view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8248–8258.

Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, et al. 2023. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–12.

Tang Tao, Longfei Gao, Guangrun Wang, Peng Chen, Dayang Hao, Xiaodan Liang, Mathieu Salzmann, and Kaicheng Yu. 2023. LiDAR-NeRF: Novel LiDAR View Synthesis via Neural Radiance Fields. *arXiv preprint arXiv:2304.10406* (2023).

Fabio Tosi, Alessio Tonioni, Daniele De Gregorio, and Matteo Poggi. 2023. NeRF-Supervised Deep Stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 855–866.

Haithem Turki, Jason Y Zhang, Francesco Ferroni, and Deva Ramanan. 2023. SUDS: Scalable Urban Dynamic Scenes. In *Computer Vision and Pattern Recognition (CVPR)*.

u-blox AG. 2023. u-center: GNSS evaluation software for Windows. https://www.u-blox.com/en/product/u-center.

Mehrnoosh Vahidpour and Kamal Sarabandi. 2012. Millimeter-Wave Doppler Spectrum and Polarimetric Response of Walking Bodies. *IEEE Transactions on Geoscience and Remote Sensing* 50, 7 (2012), 2866–2879. https://doi.org/10.1109/TGRS.2011.2176342

Alexander Vilesov, Pradyumna Chari, Adnan Armouti, Anirudh Bindiganavale Harish, Kimaya Kulkarni, Ananya Deoghare, Laleh Jalilian, and Achuta Kadambi. 2022. Blending Camera and 77 GHz Radar Sensing for Equitable, Robust Plethysmography. *ACM Trans. Graph.* 41, 4, Article 36 (jul 2022), 14 pages. https://doi.org/10.1145/3528223.3530161

Damien Vivet, Paul Checchin, and Roland Chapuis. 2013. Localization and Mapping Using Only a Rotating FMCW Radar Sensor. *Sensors* 13, 4 (2013), 4527–4552. https://doi.org/10.3390/s130404527

Haoyuan Wang, Xiaogang Xu, Ke Xu, and Rynson WH Lau. 2023c. Lighting up NeRF via Unsupervised Decomposition and Enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12632–12641.

Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689* (2021).

Yingjie Wang, Jiajun Deng, Yao Li, Jinshui Hu, Cong Liu, Yu Zhang, Jianmin Ji, Wanli Ouyang, and Yanyong Zhang. 2023a. Bi-LRFusion: Bi-Directional LiDAR-Radar Fusion for 3D Dynamic Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 13394–13403.

Zian Wang, Tianchang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. 2023b. Neural Fields meet Explicit Geometric Representations for Inverse Rendering of Urban Scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Klaudius Werber, Matthias Rapp, Jens Klappstein, Markus Hahn, Jürgen Dickmann, Klaus Dietmayer, and Christian Waldschmidt. 2015a. Automotive radar gridmap representations. In *2015 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*. 1–4. https://doi.org/10.1109/ICMIM.2015.7117922

Klaudius Werber, Matthias Rapp, Jens Klappstein, Markus Hahn, Jürgen Dickmann, Klaus Dietmayer, and Christian Waldschmidt. 2015b. Automotive radar gridmap representations. In *2015 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*. 1–4. https://doi.org/10.1109/ICMIM.2015.7117922

Ze Yang, Yun Chen, Jingkang Wang, Sivabalan Manivasagam, Wei-Chiu Ma, Anqi Joyce Yang, and Raquel Urtasun. 2023. UniSim: A Neural Closed-Loop Sensor Simulator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1389–1399.

Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. 2021. Plenoctrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5752–5761.

Junge Zhang, Feihu Zhang, Shaochen Kuang, and Li Zhang. 2023a. NeRF-LiDAR: Generating Realistic LiDAR Point Clouds with Neural Radiance Fields. *arXiv preprint arXiv:2304.14811* (2023).

Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. 2020. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492* (2020).

Liwen Zhang, Xinyan Zhang, Youcheng Zhang, Yufei Guo, Yuanpei Chen, Xuhui Huang, and Zhe Ma. 2023b. PeakConv: Learning Peak Receptive Field for Radar Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 17577–17586.

Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)* 40, 6 (2021), 1–18.